Oral Preliminary Examination

Topic 1: Optimal Predictors & Alternative Mean Square Error Estimators of General Small Area Parameters under an Informative Sample Design Using Parametric Sample Distribution Models

Yanghyeon Cho

Advisers: Dr. Emily Berg and Dr. Jae-Kwang Kim

Department of Statistics, Iowa State University

Dec 8, 2022

1. Introduction
2. Prediction
3. Special Case: Nested Error Linear Regression Model
4. Prediction Error Assessment
5. Simulation Study
6. Application to the Conservation Effects Assessment Project(CEAP) Data

Small Area and Informative Sample Design

- **Small Area**
  - A small area refers to any domain where a direct estimator has inadequate precision due to a small area-specific sample size.
  - A standard approach to small area estimation uses model-based estimators to improve the efficiency of direct estimators.
- **Informative Sample Design**
  - A design is said to be informative for the model if the selection probability is not independent of the model response variables after conditioning on the model covariate.
  - *Ex.* Suppose a sample design where the selection probability is proportional to the response variable.
  - Under an informative sample design, the population, sample, and sample-complement distributions naturally arise.

Motivation – Prediction

- **Pfeffermann & Sverchkov (2007)**

  - Showed how to predict small-area mean of both sampled and nonsampled areas under an informative sample design.

  - Limited to the linear parameter.

- **Molina & Rao (2010)**

  - Defined a procedure for obtaining the empirical best unbiased predictor of nonlinear parameters based on the nested error regression model.

  - Limited to a noninformative sample design and sampled areas.

- **Proposed Method**

  - Address prediction problems of general small area parameters for both sampled and non-sampled areas under an informative sample design.

Population Model and Sampling Procedure

- **Population - Two-Level Model**

$$y_{ij} \mid u_i \overset{\text{ind}}{\sim} f_p(y_{ij}|\mathbf{x}_{ij}, u_i),$$
$$u_i \overset{\text{ind}}{\sim} f_p(u_i),$$

for which $E_p(u_i) = 0$, $i = 1, ..., D$, and $j = 1, ..., N_i$.

- **Two-stage sampling design**

- Denote I_i and I_{ij} as the sample inclusion indicators for area i and unit j for area i.

- First stage : select d areas with probabilities $\pi_i = P(I_i = 1|u_i)$.

- Second stage : select n_i units in the selected area i with probabilities $\pi_{ij} = P(I_{ij} = 1|I_i = 1, y_{ij}, \mathbf{x}_{ij}, u_i I_i = 1)$.

- Further, let $s = \{i : I_i = 1\}$and $s_i = \{j : I_{ij} = 1, I_i = 1\}$.

Best Predictor

- **Population Parameter**

$$\theta_i = h(y_{i1}, \ldots, y_{iN_i}), \quad i = 1, \ldots, D.$$

- **Best Predictor(BP)**
 Under the squared error loss, the best predictor of the population parameter is

$$\hat{\theta}_i^B = E_p[\theta_i | D_s, I_i], \tag{1}$$

where $D_s = \{(y_{ij}, w_{ij}, w_i), i \in s, j \in s_i; \mathbf{x}_{kl}, (k,l) \in U\}$, $w_{ij} = \pi_{ij}^{-1}$, and $w_i = \pi_i^{-1}$.

- Due to the complexity of the function $h(\cdot)$, a closed-form expression for the BP may not exist.

MC Approximation of the Best Predictor

- For a sampled area i, rearrange $\boldsymbol{y}_{iN_i} = (y_{i1}, \ldots, y_{iN_i})$ as $\boldsymbol{y}_{iN_i} = (\boldsymbol{y}'_{si}, \boldsymbol{y}'_{ci})'$, where $\boldsymbol{y}_{si}$ and $\boldsymbol{y}_{ci}$ are the vectors of sampled and sample-complement units, respectively.
- **Sampled Areas**

$$\begin{aligned}\tilde{\theta}_i^B &= E_p[h(\boldsymbol{y}'_{si}, \boldsymbol{y}'_{ci})|D_s, I_i = 1] \\ &\approx R^{-1}\sum_{r=1}^{R} h((\boldsymbol{y}'_{si}, (\boldsymbol{y}_{ci}^{(r)})')') =: \hat{\theta}_i^B\end{aligned}$$

where $\boldsymbol{y}_{ci}^{(r)} \overset{iid}{\sim} f_p(\boldsymbol{y}_{ci} \mid D_s, I_i = 1, \boldsymbol{I}_{ci} = \boldsymbol{0})$, $\boldsymbol{I}_{ci} = \{I_{ij} = 0 : j \notin s_i, i \in s\}$.

- **Nonsampled Areas**

$$\begin{aligned}\tilde{\theta}_i^B &= E_p[h(\boldsymbol{y}_{iN_i})|D_s, I_i = 0] \\ &\approx R^{-1}\sum_{r=1}^{R} h(\boldsymbol{y}_{iN_i}^{(r)}) =: \hat{\theta}_i^B\end{aligned}$$

where $\boldsymbol{y}_{iN_i}^{(r)} \overset{iid}{\sim} f_p(\boldsymbol{y}_{N_i} \mid D_s, I_i = 0)$

Relationships among Distributions

- **Pfeffermann & Sverchkov (2007)**
  - **Sample distributions**

$$f_s(u_i) = f_p(u_i \mid I_i = 1)$$
$$f_{si}(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1) = f_p(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1, I_{ij} = 1)$$

  - **Sample-complement distributions**

$$f_c(u_i) = f_p(u_i \mid I_i = 0) = \frac{E_s[w_i - 1 \mid u_i] f_s(u_i)}{E_s[w_i - 1]} \tag{2}$$

$$\begin{aligned} f_{ci}(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1) &= f_p(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1, I_{ij} = 0) \\ &= \frac{E_{si}[w_{ij} - 1 \mid y_{ij}, \mathbf{x}_{ij}, u_i, I_i = 1] f_{si}(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1)}{E_{si}[w_{ij} - 1 \mid y_{ij}, \mathbf{x}_{ij}, I_i = 1]} \end{aligned} \tag{3}$$

- **Population distribution**

$$f_p(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1) = \frac{E_{si}[w_{ij} \mid y_{ij}, \mathbf{x}_{ij}, u_i, I_i = 1] f_{si}(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1)}{E_{si}[w_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1]} \tag{4}$$

Derivation of the Required Distributions

- **Required Distribution for Sampled Area**

$$
\begin{aligned}
& f_p(\boldsymbol{y}_{ci} \mid D_s, I_i = 1, \boldsymbol{I}_{ci} = \boldsymbol{0}) \\
&= \int_{-\infty}^{\infty} f_p(\boldsymbol{y}_{ci} \mid D_s, u_i, I_i = 1, \boldsymbol{I}_{ci} = \boldsymbol{0}) f_p(u_i \mid D_s, I_i = 1, \boldsymbol{I}_{ci} = \boldsymbol{0})\, du_i \\
&= \int_{-\infty}^{\infty} \prod_{j \notin s_i} f_p(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1, I_{ij} = 0) f_s(u_i \mid D_s)\, du_i \\
&= \prod_{j \notin s_i} \int_{-\infty}^{\infty} \frac{E_{si}[w_{ij} - 1 \mid y_{ij}, \mathbf{x}_{ij}, u_i, I_i = 1] f_{si}(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1)}{E_{si}[w_{ij} - 1 \mid y_{ij}, \mathbf{x}_{ij}, I_i = 1]} f_s(u_i \mid D_s)\, du_i \quad \because (3)
\end{aligned}
$$

- **Required Distribution for Non-sampled Area**

$$
\begin{aligned}
& f_p(\boldsymbol{y}_{iN_i} \mid D_s, I_i = 0) \\
&= \int_{-\infty}^{\infty} f_p(\boldsymbol{y}_{iN_i} \mid D_s, u_i, I_i = 0) f_p(u_i \mid D_s, I_i = 0)\, du_i \\
&= \prod_{j=1}^{N_i} \int_{-\infty}^{\infty} f_p(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 0) f_c(u_i)\, du_i \\
&= \prod_{j=1}^{N_i} \int_{-\infty}^{\infty} \frac{E_{si}[w_{ij} \mid y_{ij}, \mathbf{x}_{ij}, u_i, I_i = 1] f_{si}(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1)}{E_{si}[w_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1]} \frac{E_s[w_i - 1 \mid u_i]}{E_s[w_i - 1]} f_s(u_i)\, du_i \because (2),(4
\end{aligned}
$$

Prediction Algorithm Using Sampling Importance Resampling (SIR) Algorithm

- **Sampled Area**
  1. Generate $u_i^{(r)} \sim f_s(u_i \mid D_s)$.
  2. Generate $y_{ij}^{(k)} \sim f_{si}(y_{ij} \mid \mathbf{x}_{ij}, u_i^{(r)}, I_i = 1)$ for $k = 1, \ldots, K$ (sampling)
  3. Set $y_{ij}^{*(r)} = y_{ij}^{(k)}$ (resampling) with probability
  $$p_{ij}^{(k)} = \frac{E_{si}[w_{ij} - 1 \mid y_{ij}^{(k)}, u_i^{(r)}, \mathbf{x}_{ij}, I_i = 1]}{\sum_{k=1}^{K} E_{si}[w_{ij} - 1 \mid y_{ij}^{(k)}, u_i^{(r)}, \mathbf{x}_{ij}, I_i = 1]} \quad \text{(Importance)}.$$
  If (i, j) such that $I_{ij} = 1$, set $y_{ij}^{*(r)} = y_{ij}$.
  4. Define
  $$\hat{\theta}_i^{(r)} = h(y_{i1}^{*(r)}, \ldots, y_{iN_i}^{*(r)}).$$

Then, the best predictor can be approximated as $\hat{\theta}_i^B = R^{-1} \sum_{r=1}^{R} \hat{\theta}_i^{(r)}$.

- **Non-sampled Area**
  1. Generate $u_i^{(r,k_1)} \sim f_s(u_i)$, for $k_1 = 1, \dots, K_1$
  2. Set $u_i^{(r)} = u_i^{(r,k_1)}$ with probability
  $$p_i^{(r,k_1)} = \frac{E_s[w_i - 1 \mid u_i^{(r,k_1)}]}{\sum_{k_1=1}^{K_1} E_s[w_i - 1 \mid u_i^{(r,k_1)}]}.$$
  3. Generate $y_{ij}^{(r,k_2)} \sim f_{si}(y_{ij} \mid \mathbf{x}_{ij}, u_i^{(r)}, I_i = 1)$ for $k_2 = 1, \dots, K_2$
  4. Set $y_{ij}^{*(r)} = y_{ij}^{(r,k_2)}$ with probability
  $$p_{ij}^{(r,k_2)} = \frac{E_{si}[w_{ij} \mid y_{ij}^{(r,k_2)}, u_i^{(r)}, \mathbf{x}_{ij}]}{\sum_{k_2=1}^{K_2} E_{si}[w_{ij} \mid y_{ij}^{(r,k_2)}, u_i^{(r)}, \mathbf{x}_{ij}]}.$$
  5. Define
  $$\hat{\theta}_i^{(r)} = h(y_{i1}^{*(r)}, \dots, y_{iN_i}^{*(r)}).$$

Similar to sampled areas, the MC approximation of the best predictor is defined as $\hat{\theta}_i^B = R^{-1} \sum_{r=1}^{R} \hat{\theta}_i^{(r)}$.

Nested Error Linear Regression (NELR) Model

- **Sample - Nested Error Linear Regression Model**

$$y_{ij} = \beta_0 + \mathbf{x}'_{ij}\boldsymbol{\beta}_1 + u_i + e_{ij}, j \in s_i, i \in s, \tag{5}$$

where $u_i \overset{iid}{\sim} N(0, \sigma_u^2)$ and $e_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$.

- Assume that the sampling weight w_{ij} within the selected areas satisfies

$$E_{si}[w_{ij} \mid \mathbf{x}_{ij}, y_{ij}, u_i, I_i = 1] = \kappa_i \exp(\mathbf{x}'_{ij}\boldsymbol{\gamma}_1 + \gamma_2 y_{ij} + \mathbf{x}'_{ij}\boldsymbol{\gamma}_3 y_{ij}), \tag{6}$$

where $\kappa_i = N_i^{-1} \sum_{j=1}^{N_i} \exp(-\mathbf{x}'_{ij}\boldsymbol{\gamma}_1 - \gamma_2 y_{ij} - \mathbf{x}'_{ij}\boldsymbol{\gamma}_3 y_{ij})$.

- Further, assume that the area-level weight w_i satisfies a lognormal model given by

$$\log(w_i) \mid u_i, I_i = 1 \sim N(\lambda_1 + \lambda_2 u_i, \tau^2), \tag{7}$$

such that $E_s[w_i \mid u_i] = \exp(\lambda_1 + \lambda_2 u_i + \tau^2/2)$.

Required Distribution for Sampled Areas

- $f_s(u_i \mid D_s) = \frac{1}{\sqrt{\sigma_e^2 n_i^{-1}\gamma_i}}\phi\left(\frac{u_i - \hat{u}_i(\boldsymbol{\beta},\sigma_u^2,\sigma_e^2)}{\sqrt{\sigma_e^2 n_i^{-1}\gamma_i}}\right)$,

  where $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1')'$, $\hat{u}_i(\boldsymbol{\beta},\sigma_u^2,\sigma_e^2) = \gamma_i(\bar{y}_i - \beta_0 - \bar{\mathbf{x}}_i\boldsymbol{\beta}_1)$, $\gamma_i = \sigma_u^2(\sigma_u^2 + \sigma_e^2 n_i^{-1})^{-1}$, and $(\bar{\mathbf{x}}_i', \bar{y}_i) = n_i^{-1}\sum_{j=1}^{n_i}(\mathbf{x}_{ij}', y_{ij})$.

- $$\begin{aligned} f_{ci}(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1) &= \frac{E_{si}(w_{ij}|\mathbf{x}_{ij}, y_{ij}, u_i, I_i = 1) - 1}{E_{si}(w_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1) - 1} f_{si}(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1) \\ &= \frac{\lambda_{ij}}{\lambda_{ij} - 1} f_p(y_{ij}|\mathbf{x}_{ij}, u_i, I_i = 1) - \frac{1}{\lambda_{ij} - 1} f_{si}(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1) \end{aligned}$$

  where

  $$\begin{aligned} \lambda_{ij} &= E_{si}(w_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1) \\ &= \kappa_i \exp\left((\gamma_2 + \mathbf{x}_{ij}^T\boldsymbol{\gamma}_3)^2\sigma_e^2/2 + \mathbf{x}_{ij}^T\boldsymbol{\gamma}_1 + (\gamma_2 + \mathbf{x}_{ij}^T\boldsymbol{\gamma}_3)u_{ij}\right). \end{aligned}$$

- For small sampling fractions, the weights will be much larger than 1, such that $(\lambda_{ij}-1)^{-1}\lambda_{ij} \approx 1$. Therefore, an approximation for the complement distribution is

$$\begin{aligned} f_{ci}(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1) &\approx f_p(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1) \\ &= \frac{1}{\sigma_e}\phi\left(\frac{y_{ij} - u_{ij} - \gamma_2\sigma_e^2 - \mathbf{x}_{ij}'\boldsymbol{\gamma}_3\sigma_e^2}{\sigma_e}\right), \end{aligned} \tag{8}$$

where $u_{ij} = \beta_0 + \mathbf{x}_{ij}^T\boldsymbol{\beta}_1 + u_i$.

- $$\begin{aligned} f_p(y_{ij}|\mathbf{x}_{ij}, u_i, I_i = 1) &= \frac{E_{si}(w_{ij}|\mathbf{x}_{ij}, y_{ij}, u_i, I_i = 1)}{E_{si}(w_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1)} f_{si}(y_{ij} \mid \mathbf{x}_{ij}, u_i, I_i = 1) \\ &\propto \frac{1}{\sigma_e}\phi\left(\frac{y_{ij} - u_{ij} - \gamma_2\sigma_e^2 - \mathbf{x}_{ij}^T\boldsymbol{\gamma}_3\sigma_e^2}{\sigma_e}\right). \end{aligned}$$

Prediction Algorithm for Sampled Areas

For $r = 1, \dots, R$, repeat the following steps:

1. Generate $u_i^{(r)} \overset{ind}{\sim} N(\hat{u}_i(\boldsymbol{\beta}, \sigma_u^2, \sigma_e^2), \gamma_i \sigma_e^2 / n_i)$ for $i \in s$.
2. Generate $y_{ij}^{(r)} \overset{ind}{\sim} N(\beta_0 + \mathbf{x}_{ij}'\boldsymbol{\beta}_1 + u_i^{(r)} + \gamma_1 \sigma_e^2 + \mathbf{x}_{ij}'\boldsymbol{\gamma}_3 \sigma_e^2, \sigma_e^2)$ for $j = 1, \dots, N_i$
3. Set $y_{ij}^{(r)} = y_{ij}$ if $I_{ij} = 1$.
4. Define $\hat{\theta}_i^{(r)} = \hat{\theta}_i^{(r)}(\boldsymbol{\beta}, \sigma_u^2, \sigma_e^2, \gamma_2, \boldsymbol{\gamma}_3) = h(y_{i1}^{(r)}, \dots, y_{iN_i}^{(r)})$.

An approximation for the best predictor (1) is then

$$\hat{\theta}_{i,R}^{B} := \hat{\theta}_{i,R}(\boldsymbol{\beta}, \sigma_u^2, \sigma_e^2) = R^{-1} \sum_{r=1}^{R} \hat{\theta}_i^{(r)}. \tag{9}$$

Empirical Best Predictor (EBP)

- In practice, one must estimate $\boldsymbol{\psi}_s = (\boldsymbol{\beta}, \sigma_u^2, \sigma_e^2, \gamma_2, \boldsymbol{\gamma}_3)'$.
- We define $(\hat{\boldsymbol{\beta}}', \hat{\sigma}_u^2, \hat{\sigma}_e^2)$ to be the maximum likelihood estimators under the sample distribution (5).
- We obtain an estimate $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_1', \hat{\gamma}_2, \hat{\boldsymbol{\gamma}}_3')'$ by

$$\hat{\boldsymbol{\gamma}} = argmin_{\boldsymbol{\gamma}} \sum_{i=1}^{d} \sum_{j \in s_i} \left(w_{ij} - \kappa_i \exp(\mathbf{x}_{ij}' \boldsymbol{\gamma}_1 + \gamma_2 y_{ij} + \mathbf{x}_{ij}' \boldsymbol{\gamma}_3 y_{ij})\right)^2.$$

- Define an empirical best predictor as

$$\hat{\theta}_{i,R}^{EB} = \hat{\theta}_{i,R}(\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2, \hat{\sigma}_e^2, \hat{\gamma}_2, \hat{\boldsymbol{\gamma}}_3) = R^{-1} \sum_{r=1}^{R} \hat{\theta}_i^{(r)}. \quad (10)$$

Required Distribution for Non-sampled Areas

$$
\begin{aligned}
f_c(u_i) &= \frac{E_s(w_i - 1 \mid u_i)}{E_s(w_i) - 1} f_s(u_i) \\
&= \frac{1}{E_s(w_i) - 1} \left[exp(\lambda_1 + \lambda_2 u_i + \tau^2/2) * \frac{1}{\sigma_u} \phi\left(\frac{u_i}{\sigma_u}\right) - \frac{1}{\sigma_u} \phi\left(\frac{u_i}{\sigma_u}\right) \right] \\
&= \frac{1}{E_s(w_i) - 1} \left[E_s(w_i) * \frac{1}{\sigma_u} \phi\left(\frac{u_i - \sigma_u^2 \lambda_2}{\sigma_u}\right) - \frac{1}{\sigma_u} \phi\left(\frac{u_i}{\sigma_u}\right) \right],
\end{aligned}
$$

where $E_s(w_i) = \exp(\lambda_1 + \tau^2/2 + \sigma_u^2 \lambda_2^2/2)$.

Prediction Algorithm for Non-sampled Areas

For $r = 1, \ldots, R$, repeat the following steps:

1. Simulate u_i independently from $f_c(u_i)$ for $i \notin s$ through inversion sampling.
2. Generate $y_{ij}^{(r)} \overset{ind}{\sim} N(\beta_0 + \mathbf{x}'_{ij}\boldsymbol{\beta}_1 + u_i^{(r)} + \gamma_2\sigma_e^2 + \mathbf{x}'_{ij}\boldsymbol{\gamma}_3\sigma_e^2, \sigma_e^2)$ for $j = 1, \ldots, N_i$.
3. Define $\hat{\theta}_i^{(r)} = \hat{\theta}_i^{(r)}(\boldsymbol{\beta}, \sigma_u^2, \sigma_e^2, \gamma_2, \boldsymbol{\gamma}_3, \lambda_1, \lambda_2, \tau^2) = h(y_{i1}^{(r)}, \ldots, y_{iN_i}^{(r)})$.

Then, an MC approximation for the best predictor is given by

$$\hat{\theta}_{i,R}^B = \hat{\theta}_{i,R}(\boldsymbol{\beta}, \sigma_u^2, \sigma_e^2, \gamma_2, \boldsymbol{\gamma}_3, \lambda_1, \lambda_2, \tau^2) = R^{-1}\sum_{r=1}^{R} \tilde{\theta}_i^{(r)}. \tag{11}$$

Empirical Best Predictor (EBP) for Non-sampled Areas

- We define an estimator of $\boldsymbol{\psi}_{ns} = (\lambda_1, \lambda_2, \tau^2)'$ as

$$(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\tau}^2) = argmax_{\Theta} \prod_{i\in s} \int_{-\infty}^{\infty} \frac{1}{\tau}\phi\left(\frac{\log(w_i) - \lambda_1 - \lambda_2 u_i}{\tau}\right) \hat{f}_s(u_i \mid D_s) du_i$$

$$= argmax_{\Theta} \prod_{i\in s} \frac{1}{\sqrt{\lambda_2^2}} \frac{1}{\sqrt{2\pi\left(\frac{\tau^2}{\lambda_2^2} + \hat{v}_i^2\right)}} exp\left(-\frac{\left(\frac{log(w_i)-\lambda_1}{\lambda_2} - \hat{u}_i\right)^2}{2\left(\frac{\tau^2}{\lambda_2^2} + \hat{v}_i^2\right)}\right) \quad (12)$$

where $\Theta = (-\infty, \infty) \times (-\infty, \infty) \times (0, \infty)$, and $\hat{u}_i = \hat{u}_i(\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2, \hat{\sigma}_e^2)$ and $\hat{v}_i^2 = \hat{\sigma}_e^2 n_i^{-1} \hat{\sigma}_u^2 (\hat{\sigma}_u^2 + \hat{\sigma}_e^2/n_i)^{-1}$ are the mean and the variance of $\hat{f}_s(u_i \mid D_s)$.

- Define an empirical best predictor as

$$\hat{\theta}_{i,R}^{EB} = \hat{\theta}_{i,R}(\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2, \hat{\sigma}_e^2, \hat{\gamma}_2, \hat{\boldsymbol{\gamma}}_3, \hat{\lambda}_1, \hat{\lambda}_2, \hat{\tau}^2). \quad (13)$$

MSE Decomposition

- **MSE Components**

$$\begin{aligned}\mathrm{MSE}(\hat{\theta}_{i,R}^{EB}) &= E[(\hat{\theta}_i^{B} - \theta_i)^2] + E[(\hat{\theta}_i^{EB} - \hat{\theta}_i^{B})^2] + E[(\hat{\theta}_{i,R^{EB}} - \hat{\theta}_i^{EB})^2] \quad (14)\\ &= E[V(\theta_i \mid D_s, I_i; \boldsymbol{\psi})] + E[(\hat{\theta}_i^{EB} - \hat{\theta}_i^{B})^2] + R^{-1}E[V(\theta_i \mid D_s, I_i; \hat{\boldsymbol{\psi}})]\\ &= M_{1i} + M_{2i} + M_{3i},\end{aligned}$$

where $\hat{\boldsymbol{\psi}}$ are estimators of $\boldsymbol{\psi} = (\boldsymbol{\beta}', \sigma_u^2, \sigma_e^2, \gamma_2, \boldsymbol{\gamma}_3', \lambda_1, \lambda_2, \tau^2)'$.

- M_{1i}, called the leading term, is the MSE of the best predictor.
- M_{2i} accounts for the effect of the variance of $\hat{\boldsymbol{\psi}}$
- M_{3i} accounts for the variability due to the MC approximation of the EBP. It is $O(R^{-1})$, which is ignorable for sufficiently large R.

Estimation of M_{1i}

- Note that

$$\hat{M}_{1i} = V(\theta_i \mid D_s, I_i; \boldsymbol{\psi}) = V_i(\boldsymbol{\psi}) \tag{15}$$
$$\approx V(\theta_i \mid D_s, I_i; \hat{\boldsymbol{\psi}}) =: V_i(\hat{\boldsymbol{\psi}}) \tag{16}$$
$$\approx (R-1)^{-1} \sum_{r=1}^{R} (\hat{\theta}_i^{(r)} - \hat{\theta}_{i,R}^{EB})^2 \tag{17}$$

- Note that $\hat{\theta}_i^{(r)}$ for $r = 1, \ldots, R$ are *iid*, $E_R[\hat{\theta}_i^{(r)}] = E[\theta_i \mid D_s, I_i; \hat{\boldsymbol{\psi}}]$, and $V_R(\theta_i^{(r)}) = V_i(\hat{\boldsymbol{\psi}})$, where E_R and V_R, respectively, denote expectation and variance relative to the distribution used to generate $\hat{\theta}_i^{(r)}$.
- $\hat{M}_{1i}$ can be a consistent estimator of $V_i(\boldsymbol{\psi})$, if 1) $\hat{\boldsymbol{\psi}} \xrightarrow{p} \boldsymbol{\psi}$ as $D \to \infty$ and 2) $V_i(\boldsymbol{\psi})$ is a continuous function of $\boldsymbol{\psi}$.

$$\because\ |\hat{M}_{1i} - V_i(\boldsymbol{\psi})| \le \underbrace{|\hat{M}_{1i} - V_i(\hat{\boldsymbol{\psi}})|}_{O_p(R^{-0.5})} + \underbrace{|V_i(\hat{\boldsymbol{\psi}}) - V_i(\boldsymbol{\psi})|}_{CMT}$$

Estimation of M_{2i}

- Estimator of M_{2i} using a bootstrap procedure can be defined as

$$\hat{M}_{2i} = B^{-1} \sum_{b=1}^{B} \left\{ \hat{\theta}_{i,R}^{EB}\left(\hat{\boldsymbol{\psi}}^{(b)}\right) - \hat{\theta}_{i,R}^{EB} \right\}^2,$$

  where $\hat{\boldsymbol{\psi}} \rightarrow \{(y_{ij}^{(b)}, w_{ij}^{(b)}, w_i^{(b)}) : j \in s_i, i \in s\} \rightarrow \hat{\boldsymbol{\psi}}^{(b)} \rightarrow \hat{\theta}_{i,R}^{EB}\left(\hat{\boldsymbol{\psi}}^{(b)}\right)$.
- We cannot implement a fully parametric bootstrap procedure, because we do not specify the full distribution for the sampling weights w_{ij}.
- Instead, we employ a non-parametric estimate of the asymptotic normal distribution of $\hat{\boldsymbol{\psi}}$, using properties of the generalized estimating equation (GEE) estimator.

Generalized Estimating Equations (GEE) Estimator, Shao (2003)

Definition (GEE and GEE estimator)

Assume that $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_m$ are independent random vectors, where the dimension of $\boldsymbol{Y}_i$ is n_i and θ is a $k-$vector of unknown parameters, where $\Theta \subset R^k$.

Let η_i be a Borel function from $R^{n_i} \times \Theta$ to R^k, $i = 1, ..., m$ and

$$s_m(\gamma) = \sum_{i=1}^{m} \eta_i(\boldsymbol{Y}_i, \gamma), \ \gamma \in \Theta.$$

If θ is estimated by $\hat{\theta} \in \Theta$ satisfying $s_m(\hat{\theta}) = 0$, then $\hat{\theta}$ is called a GEE estimator. Moreover, the equation $s_m(\gamma) = 0$ is called a GEE.

Asymptotic Normality of GEE estimators

- If $\{\hat{\theta}_m\}$ is a consistent sequence of GEE estimators and η_i is suitably smooth, then
$$V_m^{-1/2}(\hat{\theta}_m - \theta) \xrightarrow{d} N_k(\mathbf{0}, I_k),$$
- The *jackknife* variance estimator for $\hat{\theta}_m$, Quenouille (1949) and Tukey(1958)
$$\hat{V}_J = \frac{m-1}{m}\sum_{i=1}^{m}\left(\hat{\boldsymbol{\theta}}_{-i} - \bar{\boldsymbol{\theta}}_m\right)\left(\hat{\boldsymbol{\theta}}_{-i} - \bar{\boldsymbol{\theta}}_m\right)^T,$$
where $\hat{\boldsymbol{\theta}}_{-i}$ is the estimator based on $\boldsymbol{y}_1, ..., \boldsymbol{y}_{i-1}, \boldsymbol{y}_{i+1}, ..., \boldsymbol{y}_m$ and $\bar{\boldsymbol{\theta}}_m$ is the average of $\hat{\boldsymbol{\theta}}_{-i}$'s.
- By the asymptotic normality of GEE estimators,
$$\hat{\psi}^{(b)} \sim N(\hat{\boldsymbol{\psi}}, \hat{V}_J)$$
- MSE estimator can be defined as
$$\widehat{MSE}_i^{\text{no\_BC}} = \hat{M}_{1i} + \hat{M}_{2i}.$$

Bias-Corrected MSE Eestimators

- A problem with $\widehat{MSE}_i^{\text{no\_BC}}$ is $E[\hat{M}_{1i} - M_{1i}] \neq 0$.
- While implementing the bootstrap procedure for estimating M_{2i}, we can calculate the estimate of the leading term from bootstrap sample b:

$$\hat{M}_{1i}^{(b)} = (R-1)^{-1} \sum_{r=1}^{R} \left\{ \hat{\theta}_i^{(r)}(\hat{\boldsymbol{\psi}}^{(b)}) - \hat{\theta}_{i,R}^{EB}(\hat{\boldsymbol{\psi}}^{(b)}) \right\}^2.$$

- **Bias Corrections for $\hat{M}_{1i}$**
  - $\hat{M}_{1i}^{\text{Add}} = \hat{M}_{1i} - \left\{ \bar{M}_{1i}^{*} - \hat{M}_{1i} \right\} = 2\hat{M}_{1i} - \bar{M}_{1i}^{*}$, where $\bar{M}_{1i}^{*} = B^{-1} \sum_{b=1}^{B} \hat{M}_{1i}^{(b)}$.
  - $\hat{M}_{1i}^{\text{Mult}} = \hat{M}_{1i} \frac{\hat{M}_{1i}}{\bar{M}_{1i}^{*}} = \hat{M}_{1i}^2 (\bar{M}_{1i}^{*})^{-1}$
  - $\hat{M}_{1i}^{\text{Comp}} = \begin{cases} \hat{M}_{1i}^{\text{Add}}, & \text{if } \hat{M}_{1i} \geq \bar{M}_{1i}^{*}, \\ \hat{M}_{1i}^{\text{Mult}}, & \text{if } \hat{M}_{1i} < \bar{M}_{1i}^{*}. \end{cases}$
  - $\hat{M}_{1i}^{\text{HM}}(\hat{\boldsymbol{\psi}}) = \begin{cases} 2\hat{M}_{1i} - \bar{M}_{1i}^{*}, & \text{if } \hat{M}_{1i} \geq \bar{M}_{1i}^{*}, \\ \hat{M}_{1i} \exp\left[- \left\{ \bar{M}_{1i}^{*} - \hat{M}_{1i} \right\} / \bar{M}_{1i}^{*} \right], & \text{if } \hat{M}_{1i} < \bar{M}_{1i}^{*}. \end{cases}$

Bias-Corrected MSE Estimators

$$\widehat{MSE}_i^{\text{Add}} = \hat{M}_{1i}^{Add}(\hat{\boldsymbol{\psi}}) + \hat{M}_{2i}, \tag{18}$$

$$\widehat{MSE}_i^{\text{Mult}} = \hat{M}_{1i}^{Mult}(\hat{\boldsymbol{\psi}}) + \hat{M}_{2i}, \tag{19}$$

$$\widehat{MSE}_i^{\text{Comp}} = \begin{cases} \widehat{MSE}_i^{\text{Add}}, & \text{if } \hat{M}_{1i} \geq \bar{M}_{1i}^* \\ \widehat{MSE}_i^{\text{Mult}}, & \text{if } \hat{M}_{1i} < \bar{M}_{1i}^*, \end{cases} \tag{20}$$

and

$$\widehat{MSE}_i^{\text{HM}} = \begin{cases} 2\hat{M}_{1i} - \bar{M}_{1i}^* + \hat{M}_{2i}, & \text{if } \hat{M}_{1i} \geq \bar{M}_{1i}^* \\ \hat{M}_{1i} exp\left[- \left\{ \bar{M}_{1i}^* - \hat{M}_{1i} \right\} / \bar{M}_{1i}^* \right] + \hat{M}_{2i}, & \text{if } \hat{M}_i < \bar{M}_{1i}^*. \end{cases} \tag{21}$$

Naive Confidence Interval

- Note that $\{\hat{\theta}_i^{(r)} : r = 1, \ldots, R\}$ is samples from $f(\theta_i | \boldsymbol{y}_{si}, I_i; \hat{\boldsymbol{\psi}})$.
- **Naive CI**

$$\begin{aligned}\widehat{CI}_i^{\text{naive}} &= \left(q_{\alpha/2}(\boldsymbol{y}_{si}, \hat{\boldsymbol{\psi}}), q_{1-\alpha/2}(\boldsymbol{y}_{si}, \hat{\boldsymbol{\psi}})\right) \\ &\approx (\hat{\theta}_{i,\alpha/2}^{(r)}, \hat{\theta}_{i,1-\alpha/2}^{(r)}),\end{aligned}$$

  where $q_\alpha(\boldsymbol{y}_{si}, \hat{\boldsymbol{\psi}})$ and $\hat{\theta}_{i,\alpha}^{(r)}$ are the αth quantile of $f(\theta_i | \boldsymbol{y}_{si}, I_i; \hat{\boldsymbol{\psi}})$ and $\left\{\hat{\theta}_i^{(r)} : r = 1, \ldots, R\right\}$, respectively.
- This CI could be too short (or even too long) because of ignoring the variability of $\hat{\boldsymbol{\psi}}$, failing to attain the nominal coverage probability.

Calibrated Naive CI, Carlin & Gelfand (1991)

- Define

$$r(\hat{\boldsymbol{\psi}}, \boldsymbol{\psi}, \boldsymbol{y}_{si}, \alpha) = E_{\theta_i|\boldsymbol{y}_{si},\psi} I\{q_{\alpha/2}(\boldsymbol{y}_{si}, \hat{\boldsymbol{\psi}}) \leq \theta_i \leq q_{1-\alpha/2}(\boldsymbol{y}_{si}, \hat{\boldsymbol{\psi}})\}$$

and

$$R(\boldsymbol{\psi}, \boldsymbol{y}_{si}, \alpha) = E_{\hat{\boldsymbol{\psi}}|\boldsymbol{y}_{si},\boldsymbol{\psi}}\{r(\hat{\boldsymbol{\psi}}, \boldsymbol{\psi}, \boldsymbol{y}_{si}, \alpha)\}.$$

- Find the α' such that

$$R(\boldsymbol{\psi}, \boldsymbol{y}_{si}, \alpha') = \alpha.$$

- Define the calibrated CI for θ_i as

$$\widehat{CI}_i^{\text{Cal}} = \left(q_{\alpha'/2}(\boldsymbol{y}_{si}, \hat{\boldsymbol{\psi}}), q_{1-\alpha'/2}(\boldsymbol{y}_{si}, \hat{\boldsymbol{\psi}})\right).$$

- In our framework, $\widehat{CI}_i^{\text{Cal}} \approx (\hat{\theta}^{(r)}_{i,\alpha'/2}, \hat{\theta}^{(r)}_{i,1-\alpha'/2})$ such that

$$\begin{aligned} R(\boldsymbol{\psi}, \boldsymbol{y}_{si}, \alpha') &= E_{\hat{\boldsymbol{\psi}}|\boldsymbol{y}_{si},\boldsymbol{\psi}}\{r(\hat{\boldsymbol{\psi}}, \boldsymbol{\psi}, \boldsymbol{y}_{si}, \alpha')\} \\ &\approx \frac{1}{B}\sum_b r(\hat{\boldsymbol{\psi}}_i^{(b)}, \hat{\boldsymbol{\psi}}, \boldsymbol{y}_{si}, \alpha') \\ &\approx \frac{1}{BL}\sum_b\sum_r I\{q_{\alpha'/2}(\boldsymbol{y}_{is}, \hat{\boldsymbol{\psi}}_i^{(b)}) \leq \theta_i^{(r)} \leq q_{\alpha'/2}(\boldsymbol{y}_{si}, \hat{\boldsymbol{\psi}}_i^{(b)})\} = 1-\alpha. \end{aligned}$$

Setup

- A slight modification of Pfeffermann & Sverchkov (2007)
- **Population - Nested Error Linear Regression Model**

$$y_{ij} = 5 + 0.1x_{ij} + u_i + e_{ij},\ j = 1, ..., N_i = 100,\ i = 1, ..., 150$$

$$u_i \overset{iid}{\sim} N(0, \sigma_u^2),\ e_{ij} \overset{iid}{\sim} N(0, 0.3^2),$$

where the area random effects u_i and the errors e_{ij} are independent.

- **Two-stage sampling design**

1. Stratify the areas into 3 strata: Stratum U_h, $h = 1, 2$, and 3.
2. Select 30 areas from each stratum with probabilities $\pi_i = 30 * z_i / \sum_{j_h} z_j$, where $z_i = Int[1000 \times exp(-u_i/8\sigma_u)]$.
3. Sample n_i units from selected area i with probabilities $\pi_{ij} = n_i z_{ij} / \sum_{k=1}^{N_i} z_{ik}$, where $z_{ij} = exp\{[-(y_{ij} - \mathbf{x}_{ij}\boldsymbol{\beta})/\sigma_e + \delta_{ij}/5]/3\}$, $\delta_{ij} \sim N(0, 1)$. Sample sizes are fixed in a given stratum;$n_i = 5$, 10, and 15 if $i \in U_1$, U_2 and U_3, respectively.

- Consider four scenarios by varying $R_\sigma = \sigma_e^{-1}\sigma_u \in \{0.5, 1, 2, 3\}$.
- **Parameters of Interest**
  - $\bar{Y}_i = N_i^{-1}\sum_{i=1}^{N_i} y_{ij}$
  - $\exp_i = N_i^{-1}\sum_{i=1}^{N_i}\exp(y_{ij})$
  - $Q_{i,p}(\mathbf{y}_{N_i})$: the 100$p$th quantile of $\{y_{i1}, \ldots, y_{iN_i}\}$, $p \in \{0.25, 0.75\}$.
  - $\mathrm{PG}_i = N_i^{-1}\sum_{j=1}^{N_i}\left(\frac{155 - exp(y_{ij})}{155}\right)$: Poverty Gap Indicator
  - $\mathrm{Gini}_i = (2N_i^2\bar{Y}_i)^{-1}\sum_{k=1}^{N_i}\sum_{\ell=1}^{N_i} \mid \exp(y_{ik}) - \exp(y_{i\ell}) \mid$: Gini coefficient

1. Performance of Proposed Predictor

Parameter	Method	R_σ	Bias 0.5	Bias 1	Bias 2	Bias 3	RMSE 0.5	RMSE 1	RMSE 2	RMSE 3
$\bar{Y}$	Proposed		-0.9745	-0.9696	-0.9526	-0.9653	0.0811	0.0915	0.0953	0.0962
	PS		-0.9749	-0.9694	-0.9533	-0.9658	0.0809	0.0914	0.0952	0.0960
	EB_MR		-9.6238	-9.6196	-9.6123	-9.6185	0.1257	0.1331	0.1358	0.1365
	PEB		-1.6541	-1.9280	-2.0265	-2.0658	0.0863	0.0995	0.1046	0.1057
exp	Proposed		-1.5794	-1.5920	-1.7108	-2.0746	13.4134	15.8545	20.5526	30.1148
	EB_MR		-14.9290	-15.1424	-16.6397	-19.9303	20.1072	22.3506	28.3924	41.2068
	PEB		-2.6377	-3.0935	-3.5528	-4.3203	14.2343	17.1511	22.3852	32.8262
$Q_{0.25}$	Proposed		-0.7330	-0.7283	-0.7139	-0.7185	0.0837	0.0936	0.0972	0.0980
	EB_MR		-9.3020	-9.2976	-9.2922	-9.2908	0.1253	0.1325	0.1352	0.1357
	PEB		-1.3804	-1.6434	-1.7401	-1.7706	0.0884	0.1010	0.1059	0.1068
$Q_{0.75}$	Proposed		-1.2176	-1.2192	-1.1988	-1.2181	0.0879	0.0982	0.1019	0.1028
	EB_MR		-9.9732	-9.9757	-9.9647	-9.9773	0.1325	0.1399	0.1426	0.1433
	PEB		-1.8755	-2.1485	-2.2404	-2.2855	0.0931	0.1061	0.1112	0.1122
PG	Proposed		0.3291	0.3293	0.2985	0.2607	0.0334	0.0386	0.0377	0.0344
	EB_MR		4.1117	3.8879	3.2890	2.7760	0.0539	0.0571	0.0539	0.0488
	PEB		0.6217	0.7227	0.6694	0.5822	0.0356	0.0420	0.0414	0.0377
$Gini$	Proposed		-0.0136	-0.0141	-0.0137	-0.0152	0.0023	0.0023	0.0024	0.0025
	EB_MR		0.0286	0.0288	0.0319	0.0351	0.0023	0.0024	0.0024	0.0026
	PEB		-0.0088	-0.0074	-0.0057	-0.0064	0.0023	0.0023	0.0024	0.0025

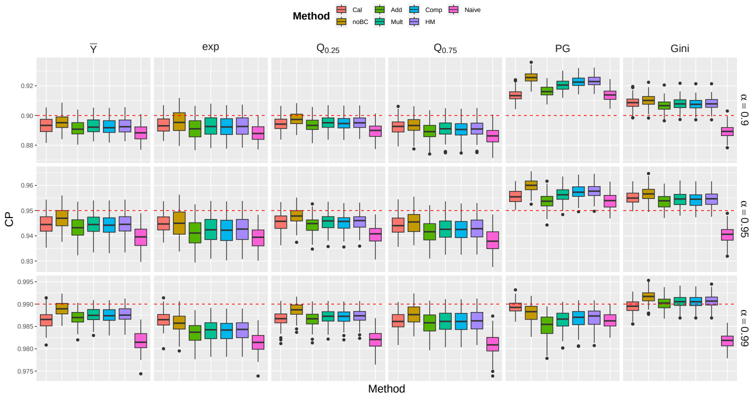
- PS is the EBP of $\bar{Y}_i$, calculated by an analytical formula of Pfeffermann & Sverchkov (2007).
- Prediction without considering an informative design (EP_MR) produces significant bias.
- PEB predictor incorporates the weights, thereby eliminating the bias of the MR predictor, but larger RMSE than that of the proposed predictor.

2. Relative Bias of Proposed MSE Estimators

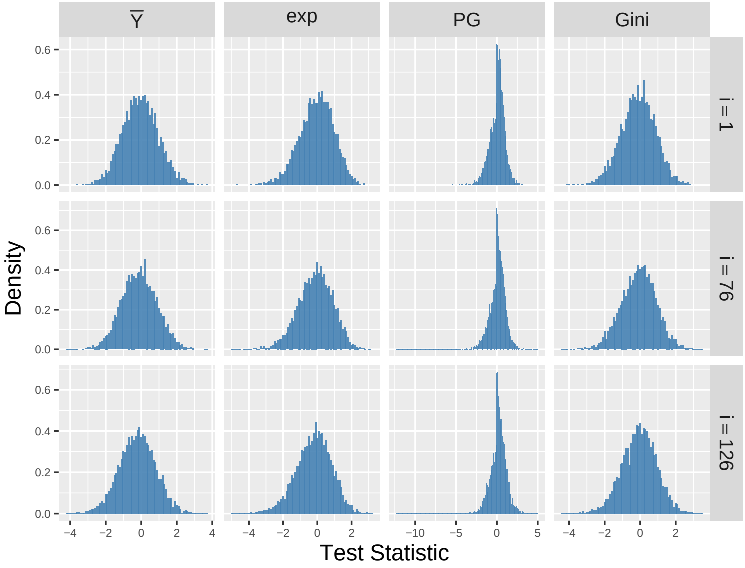
Parameter	Method	Scenario 0.5	1	2	3	Parameter	Method	Scenario 0.5	1	2	3
$\bar{Y}$	noBC	-3.2888	-1.6693	-0.9937	-0.7605	$Q_{0.75}$	noBC	-4.2644	-2.6178	-2.0532	-1.8496
	Add	-3.1283	-1.6218	-1.0680	-0.8543		Add	-4.1483	-2.5943	-2.1464	-1.9590
	Mult	-2.3594	-0.8318	-0.2733	-0.0556		Mult	-3.3858	-1.8097	-1.3565	-1.1654
	Comp	-2.7739	-1.2494	-0.6849	-0.4679		Comp	-3.7939	-2.2228	-1.7642	-1.5741
	HM	-2.6270	-1.0950	-0.5262	-0.3079		HM	-3.6471	-2.0690	-1.6060	-1.4149
exp	noBC	-4.8626	-2.8527	-1.9576	-1.4882	*PG*	noBC	1.2872	0.7887	0.2677	0.0462
	Add	-4.8290	-2.8991	-2.1051	-1.5930		Add	1.5305	0.9576	0.2876	0.0120
	Mult	-4.0221	-2.0448	-1.2328	-0.7103		Mult	2.4300	1.8406	Inf	Inf
	Comp	-4.4504	-2.4928	-1.6830	-1.1711		Comp	1.9361	1.3618	0.6932	0.4218
	HM	-4.2942	-2.3257	-1.5095	-0.9977		HM	2.1028	1.5277	0.8596	0.5899
$Q_{0.25}$	noBC	-2.4976	-1.0116	-0.3689	-0.1270	*Gini*	noBC	7.9220	8.0471	7.8479	7.5298
	Add	-2.3550	-0.9628	-0.4221	-0.2029		Add	7.6900	7.8008	7.6066	7.2867
	Mult	-1.5628	-0.1485	0.3980	0.6217		Mult	8.4909	8.6048	8.4075	8.0871
	Comp	-1.9886	-0.5793	-0.0288	0.1942		Comp	8.0870	8.2006	8.0051	7.6847
	HM	-1.8370	-0.4207	0.1338	0.3583		HM	8.2512	8.3659	8.1699	7.8493

- RBs of the proposed estimators are within 5 (%) in most parameters (uniformly within 10%).
- The relative performances of MSE estimators depend on a parameter.
- When R_σ is large, all units in a certain area could be smaller or larger than the poverty line $z = 155$.
- In that case, the denominator of the multiplicative factor in (19) could be 0, eventually making the relative bias infinity.

3. Empirical Coverage Probabilities of CIs for $R_\sigma = 3$



- The normal-theory confidence intervals almost attain the nominal coverage probability.
- However, these may be inappropriate for some nonlinear parameters since the statistic $T_i^{(m)} = \hat{\theta}_i^{(m)} - \theta_i^{(m)} / \sqrt{\widehat{\text{MSE}}_{i,t}^{(m)}}$ does not have an approximately normal distribution.


Ȳ
exp
PG
Gini
i = 1
i = 76
i = 126
Density
Test Statistic
0.0
0.2
0.4
0.6
−4
−2
0
2
4
−10
−5
5

1. Introduction
2. Prediction
3. Special Case: Nested Error Linear Regression Model
4. Prediction Error Assessment
5. Simulation Study
   - Simulation Results for Sampled Areas
   - Simluation Results for Nonsampled Areas
6. Application to the Conservation Effects Assessment Project(CEAP) Data

1. Performance of Proposed Predictor

Parameter	Method	R_σ	Bias 0.5	Bias 1	Bias 2	Bias 3	RMSE 0.5	RMSE 1	RMSE 2	RMSE 3
$\bar{Y}$	PS		-1.4056	-0.9040	-0.7044	-0.5307	0.1494	0.2926	0.5822	0.8727
	Proposed		-0.0843	0.0210	-0.1824	-0.2001	0.1490	0.2929	0.5832	0.8741
exp	–		-0.2257	-0.0249	0.0120	2.6399	25.3260	53.2528	131.5120	268.3979
$Q_{0.25}$	–		0.2777	0.3884	0.1870	0.1706	0.1515	0.2943	0.5839	0.8745
$Q_{0.75}$	–		-0.4461	-0.3452	-0.5498	-0.5715	0.1517	0.2944	0.5839	0.8746
PG	–		-0.0836	-0.1183	-0.0938	-0.1072	0.0560	0.1048	0.1855	0.2439
$Gini$	–		-0.0364	-0.0375	-0.0366	-0.0382	0.0026	0.0031	0.0046	0.0065

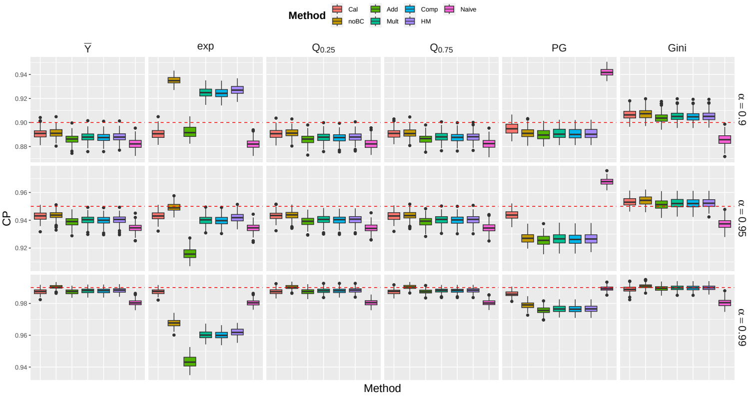
- $\text{Bias} = D^{-1}\sum_{i=1}^{D}\text{Bias}_i = D^{-1}\sum_{i=1}^{D}\frac{\sum_{m=1}^{10,000}(1-A_{im})\left\{\hat{\theta}_i^{(m)}-\theta_i^{(m)}\right\}}{\sum_{m=1}^{10,000}(1-A_{im})}$
- $\text{RMSE} = D^{-1}\sum_{i=1}^{D}\text{RMSE}_i = D^{-1}\sum_{i=1}^{D}\sqrt{\frac{\sum_{m=1}^{10,000}\left(1-A_{im}\right)\left\{\hat{\theta}_i^{(m)}-\theta_i^{(m)}\right\}^2}{\sum_{m=1}^{10,000}(1-A_{im})}}$

2. Relative Bias of Proposed MSE Estimators

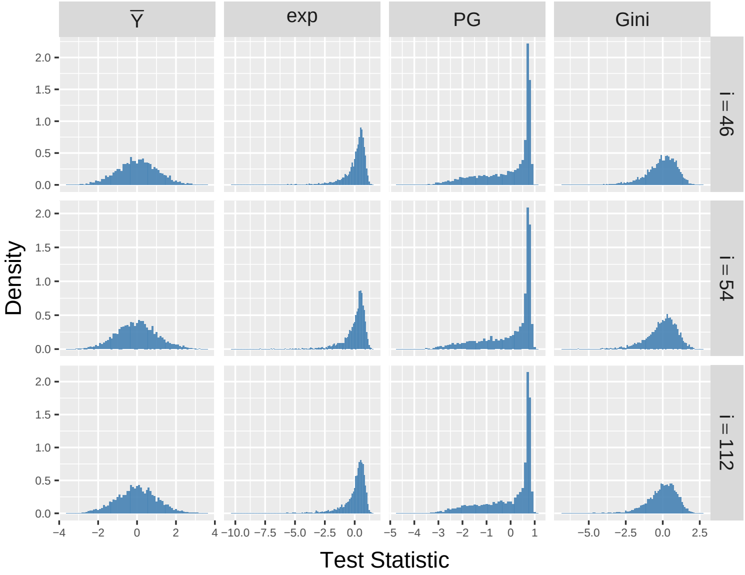
Parameter	Method	Scenario 0.5	1	2	3	Parameter	Method	Scenario 0.5	1	2	3
$\bar{Y}$	noBC	-1.3683	-1.4256	-1.1582	-0.6466	$Q_{0.75}$	noBC	-1.5885	-1.5482	-1.1727	-0.6586
	Add	-3.1419	-2.4329	-1.9914	-1.4671		Add	-3.3022	-2.5476	-2.0038	-1.4777
	Mult	-2.3158	-1.6337	-1.1966	-0.6673		Mult	-2.4832	-1.7502	-1.2093	-0.6783
	Comp	-2.6200	-1.9805	-1.5527	-1.0275		Comp	-2.7882	-2.0966	-1.5654	-1.0384
	HM	-2.4065	-1.7943	-1.3720	-0.8464		HM	-2.5778	-1.9110	-1.3848	-0.8574
exp	noBC	-0.2231	1.8354	10.8166	33.3885	PG	noBC	-1.9603	-2.4381	-2.2711	-1.6499
	Add	-2.9583	-0.7922	4.7869	17.1428		Add	-2.7721	-2.6968	-2.2157	-1.4249
	Mult	-1.9167	0.7558	10.2525	49.0299		Mult	-1.6633	-1.3688	-1.0854	-0.5383
	Comp	-2.2520	0.1543	7.7795	28.5723		Comp	-2.1856	-2.0457	-1.6693	-1.0012
	HM	-1.9693	0.5201	8.7725	31.4755		HM	-1.9516	-1.7919	-1.4539	-0.8302
$Q_{0.25}$	noBC	-1.5697	-1.4566	-1.1708	-0.6421	$Gini$	noBC	6.1099	5.6732	6.0536	5.6795
	Add	-3.2915	-2.4558	-2.0039	-1.4621		Add	5.8940	5.4628	5.8377	5.4697
	Mult	-2.4686	-1.6572	-1.2092	-0.6622		Mult	6.6738	6.2427	6.6168	6.2475
	Comp	-2.7749	-2.0043	-1.5653	-1.0225		Comp	6.2806	5.8480	6.2238	5.8548
	HM	-2.5636	-1.8185	-1.3846	-0.8414		HM	6.4406	6.0074	6.3837	6.0140

- RBs seem controlled well, except for exp.
- It is beneficial to use the bias-corrected MSE estimators for exp, PG, and Gini.

3. Empirical Coverage Probabilities of CIs for $R_\sigma = 3$



- The normal theory confidence intervals can suffer from over-coverage or under-coverage in exp and PG.
- Note that the statistics $T_i^{(m)}$ can be very left-skewed for non-sampled areas.



$\bar{Y}$
exp
PG
Gini
i = 46
i = 54
i = 112
Density
Test Statistic
2.0
1.5
1.0
0.5
0.0
−4 −2 0 2 4
−10.0 −7.5 −5.0 −2.5 0.0
−5 −4 −3 −2 −1 0 1
−5.0 −2.5 0.0 2.5

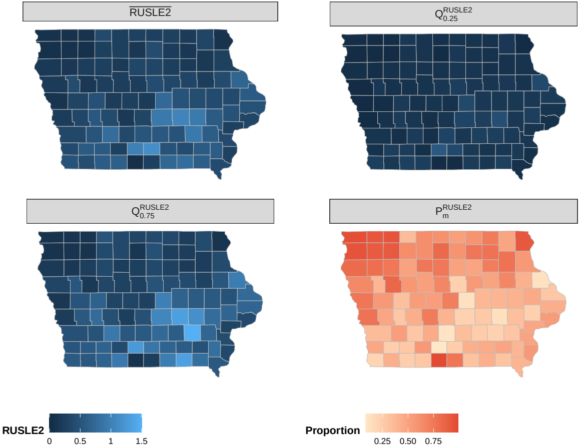
Conservation Effects Assessment Project(CEAP) Data

- The **CEAP** data is a nationwide survey of cropland to examine the environmental impacts of conventional efforts on cropland.
- The sample for the CEAP survey is a subset of a larger survey called the National Resources Inventory (NRI).
- Soil loss on cropland is an important variable of interest in the CEAP and NRI survey.
    - NRI: Universal Soil Loss Equation (USLE)
    - CEAP: Revised Universal Soil Loss Equation (RUSLE2)
- Our goal is to construct estimates of functions of RUSLE2, using the USLE as a covariate, for 99 Iowa counties.

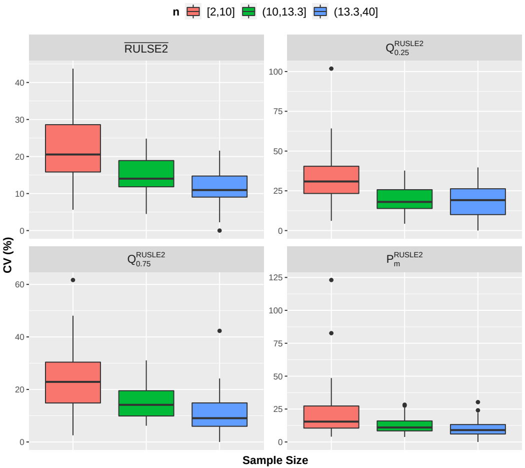
Sample Model and County-level Parameters of Interest

- Denote $RUSLE2_{ij}^{0.2}$ and $USLE_{ij}^{0.2}$ as the RUSLE2 and the USLE for unit j in county i transformed by a power of 0.2, Berg et al. (2016).
- Fit the NELR model (5) to transformed data, where $y_{ij} = RUSLE2_{ij}^{0.2}$, $x_{ij} = USLE_{ij}^{0.2}$, and $D = 99$.
- Considered county-level parameters
    - $\overline{RUSLE2}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}^5$
    - $Q_{i,0.25}^{RULSE2} = Q_{i,0.25}(y_{i1}^5, \ldots, y_{iN_i}^5)$
    - $Q_{i,0.75}^{RULSE2} = Q_{i,0.75}(y_{i1}^5, \ldots, y_{iN_i}^5)$
    - $P_{i,m}^{RUSLE2} = \frac{1}{N_i} \sum_{j=1}^{N_i} I(y_{ij}^5 < m)$, where $m = 0.232$ is the state sample median estimated from the observed $RUSLE2$.
- Here, the power of 5 converts the transformed RUSLE2 values to the original scale.

Predicted Values for the corresponding county-level Parameter in Iowa

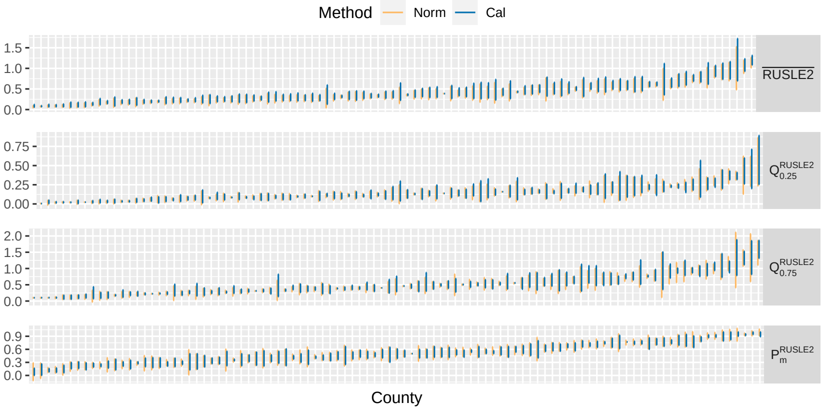


County's Percent Coefficients of Variation, $\sqrt{\widehat{MSE}_i^{HM}}/\hat{\theta}_i$



- The mean of CVs across all counties for each parameter, $\overline{RUSLE2}_i$, $Q_{i,0.25}^{RULSE2}$, $Q_{i,0.75}^{RULSE2}$, and $P_{i,m}^{RUSLE2}$ is 16.31%, 24.56%, 17.38%, and 16.27%, respectively.

Normal Theory CI with $\widehat{MSE}_i^{HM}$ (Norm) and Calibrated CI (Cal)



References I

Berg, E., Kim, J.-K. & Skinner, C. (2016), 'Imputation under informative sampling', *Journal of Survey Statistics and Methodology* **4**(4), 436–462.

Carlin, B. P. & Gelfand, A. E. (1991), 'A sample reuse method for accurate parametric empirical bayes confidence intervals', *Journal of the Royal Statistical Society: Series B (Methodological)* **53**(1), 189–200.

Chen, Y.-C. (2022), 'Pattern graphs: a graphical approach to nonmonotone missing data', *The Annals of Statistics* **50**(1), 129–146.

Graf, M., Marín, J. M. & Molina, I. (2019), 'A generalized mixed model for skewed distributions applied to small area estimation', *Test* **28**(2), 565–597.

Guadarrama, M., Molina, I. & Rao, J. (2018), 'Small area estimation of general parameters under complex sampling designs', *Computational Statistics & Data Analysis* **121**, 20–40.

Hall, P. & Maiti, T. (2006), 'On parametric bootstrap methods for small area prediction', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(2), 221–238.

References II

Hobza, T., Marhuenda, Y. & Morales, D. (2020), 'Small area estimation of additive parameters under unit-level generalized linear mixed models', *Sort* **44**, 3–38.

Kim, J. K. (2011), 'Parametric fractional imputation for missing data analysis', *Biometrika* **98**(1), 119–132.

Molina, I. & Rao, J. (2010), 'Small area estimation of poverty indicators', *Canadian Journal of Statistics* **38**(3), 369–385.

Pfeffermann, D. & Sverchkov, M. (2007), 'Small-area estimation under informative probability sampling of areas and within the selected areas', *Journal of the American Statistical Association* **102**(480), 1427–1439.

Shao, J. (2003), *Mathematical statistics*, Springer Science & Business Media.

Smith, A. F. & Gelfand, A. E. (1992), 'Bayesian statistics without tears: a sampling–resampling perspective', *The American Statistician* **46**(2), 84–88.

Thank You