Constrained Bayesian Optimization for Small Area Measurement Models

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

[a]Iowa State University [mosaferi@iastate.edu] [b]Duke University [rebecca.steorts@duke.edu]

Small Area Estimation Modeling

Small Area Estimation is a technique for capturing the strength from the surrounding areas (X_i's) of the i-th interested area to provide a reliable direct estimator y_i when the sample size in the i-th area is too small to support a direct estimator. However, there are cases in which X_i's themselves contain *errors*.

- Assume a situation where we have skewed variables or those containing outliers common in the statistical agencies.

Some Skewed Patterns from Census of Governments

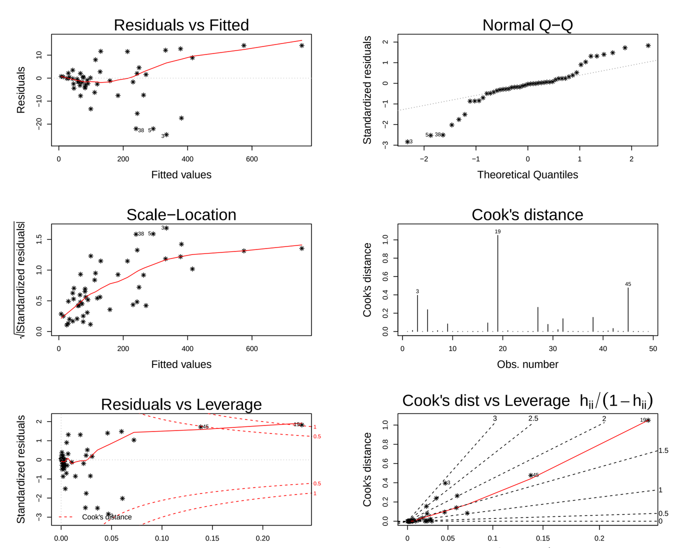


Figure: Plots correspond to the residuals of regressing y [average number of full-time employees] from 2012 Census of Governments on X [average number of full-time employees] from 2007 Census of Governments.

Area-Level Log Multiplicative Measurement Error Model

Define the sampling error $e_i := \log \xi_i$ and random effect $\nu_i := \log \lambda_i$ which are mutually independently distributed as $e_i \sim \mathcal{N}(0, \psi_i)$ and $\nu_i \sim \mathcal{N}(0, \sigma_\nu^2)$ for $i = 1, ..., m$. Consider the following multiplicative model for a positive response y_i ($y_i > 0$):

$$\begin{cases} y_i = Y_i \xi_i & \forall\, i = 1, ..., m \\ Y_i = \lambda_i \prod_{k=1}^{p} X_{ik}^{\beta_k} \end{cases}$$

- Logarithmic transformation of the area-level multiplicative model: $z_i := \log y_i = \log(Y_i \xi_i) = \theta_i + e_i$.
- Two-stage hierarchical model where the pairs (z_i, θ_i) for $i = 1, ..., m$ are independent:

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

  where $\boldsymbol{X}_i = (X_{i1}, ..., X_{ip})^\top$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^\top$.
- Multiplicative measurement error model: $\boldsymbol{x}_i = \boldsymbol{X}_i \odot \boldsymbol{\eta}_i$ [Hadamard Product], where $\boldsymbol{\eta}_i \sim$ log-normal$_p(\boldsymbol{0}, \Sigma_i)$; Mosaferi and Steorts (2020+).

Adjusted EB Predictor and Mean Squared Prediction Error (MSPE)

Interested parameter is $Y_i \equiv \exp(\theta_i)$. Therefore, the adjusted empirical Bayes (EB) predictor is

$$\hat{Y}_i^A := E(Y_i | z_i) = \exp\{\hat{\gamma}_i^* z_i + (1 - \hat{\gamma}_i^*) \sum_{k=1}^{p} \hat{\beta}_k w_{ik} + \psi_i \hat{\gamma}_i^* / 2\},$$

where $\hat{\gamma}_i^* = (\hat{\sigma}_\nu^2 + \hat{\boldsymbol{\beta}}^\top \Sigma_i \hat{\boldsymbol{\beta}}) / (\hat{\sigma}_\nu^2 + \hat{\boldsymbol{\beta}}^\top \Sigma_i \hat{\boldsymbol{\beta}} + \psi_i)$, and the vector of unknown parameter is $\phi = (\boldsymbol{\beta}, \sigma_\nu^2)^\top$.

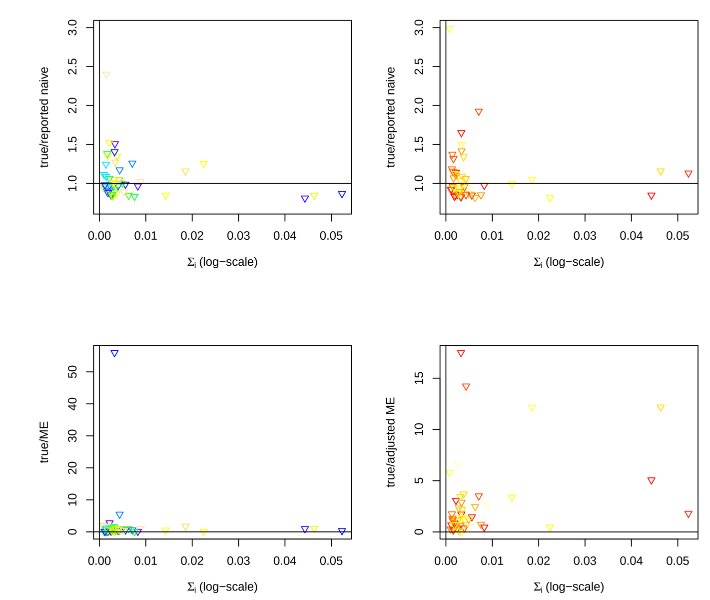


Figure: Up-Left) ratio of true to reported standard error for the naive model w/o covariate v.s. Σ_i in log-scale. Up-Right) ratio of true to reported standard error for the naive model w. covariate v.s. Σ_i in log-scale. Down-Left) ratio of true to predicted MSE for ME model w/o adjustment v.s. Σ_i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

Remark 1: Our proposed adjusted EB predictor does not uniformly beat the direct estimator in terms of the MSE and its bias-corrected version should be considered (see, Mosaferi and Steorts (2020+)).

Lemma) *Based on asymptotic optimality and Cauchy-Schwarz inequality:*

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

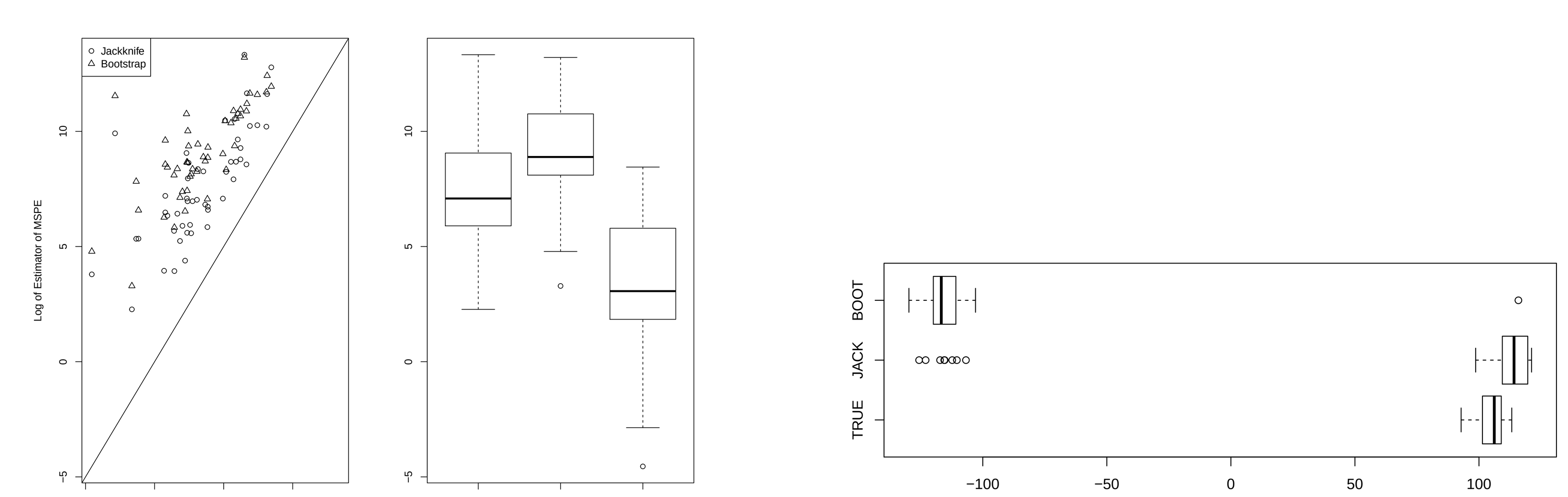
$$\text{MSPE}(\hat{Y}_i^A) \approx \mathbb{E}(\tilde{Y}_i^A - Y_i)^2 + \mathbb{E}(\hat{Y}_i^A - \tilde{Y}_i^A)^2 := M_{1i} + M_{2i},$$

where $\tilde{Y}_i^A$ is the *adjusted Bayes predictor* (Mosaferi and Steorts (2020+)).

Jackknife and Parametric Bootstrap Estimators of MSPE

Theorem) *Under our regularity conditions, we prove the order of the bias for Jackknife and parametric Bootstrap estimators of the MSPE are corrected up to the order* $O(m^{-1})$, *i.e.* $\mathbb{E}[mspe_J(\hat{Y}_i^A)] = MSPE(\hat{Y}_i^A) + O(m^{-1})$ *and* $\mathbb{E}[mspe_B(\hat{Y}_i^A)] = MSPE(\hat{Y}_i^A) + O(m^{-1})$*; Mosaferi and Steorts (2020+).*

Remark 2: Observe that the Jackknife is more accurate than the Bootstrap in practice.



(a) From Census of Governments (b) From Model-Based Simulation (in log-scale)

Figure: Plots of evaluating the estimators of MSPE for the adjusted EB predictor $\hat{Y}_i^A$.

Constrained EB Predictors with Their MSEs

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 $\Big[\sum_{i=1}^m w_i \hat{Y}_i = \sum_{i=1}^m w_i y_i$ where $\sum_{i=1}^m w_i = 1\Big]$ $(\star)$. 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.

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*:

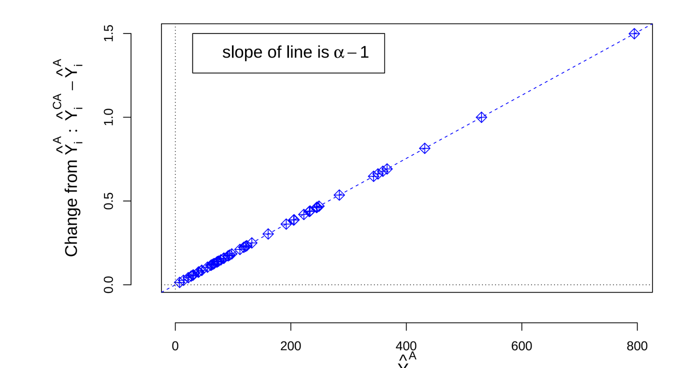
$$L_{KL}^w(\boldsymbol{Y}, \hat{\boldsymbol{Y}}) = \sum_{i=1}^{m} w_i \Big\{\hat{Y}_i - Y_i - Y_i \log(\hat{Y}_i / Y_i)\Big\}.$$

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

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

Empirical Measure of Uncertainties [in log-scale]

EMSE$\big(\hat{Y}^A\big)$ = 5.467	EMSE$\big(\hat{Y}^{CA}\big)$ = 5.465
mspe$_{\text{JACK}}\big(\hat{Y}^A\big)$= 6.041	mspe$_{\text{JACK}}\big(\hat{Y}^{CA}\big)$= 6.044
mspe$_{\text{BOOT}}\big(\hat{Y}^A\big)$= 9.7096	mspe$_{\text{BOOT}}\big(\hat{Y}^{CA}\big)$= 9.713



Remark 3: It is challenging to find a loss function with a convenient solution for the positive parameters. We observe the *weighted Kullback-Leibler loss* and its *ratio estimator* solution does not have much benefits in terms of MSE. Therefore, it is interesting to investigate other loss functions rather than Kullback-Leibler such as Pareto loss function and other benchmark constraints such as a geometric mean (This is under our further investigation).

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 (see *Remark 1*).
- 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, $\hat{\phi}$ should be calculated to as close to machine precision as possible; otherwise, Jackknife and Bootstrap may result in a decreased accuracy (see *Remark 2*).