

Two-fold Subarea Models for Small Area Estimation: Some Theory and Model Selection

Song Cai* Golshid Chatrchi† Laura Dumitrescu‡ J.N.K. Rao*

Abstract

We study some inference problems under two-fold nested subarea models with random effects, which generalize the celebrated Fay-Herriot area model. We develop the theory of empirical best prediction for a subarea parameter under unmatched two-fold subarea models. In addition, we propose an effective variable selection method under matched two-fold subarea models based on transformation. Simulation results are provided to illustrate the strengths of the proposed methods.

Key Words: Empirical best prediction, Fay-Herriot model, Importance sampling, Information criterion, Transformation, Variable selection

1. Introduction

Small area estimation (SAE) aims to provide reliable estimates of some parameters of interest, such as means or totals, of subpopulations (areas). Sample surveys are usually carried out in some or all areas to collect unit-level data and design-based “direct” estimators of the parameters are obtained. A common issue in SAE is that the sample sizes of some sampled areas are small, which yield unreliable direct estimates for those areas. A more pertinent issue is that direct estimators are not available for areas where no samples are collected (non-sampled areas).

When an area-level auxiliary variable is available for all areas, it could be used to improve the accuracy of direct estimators and to provide useful estimates for non-sampled areas. The Fay-Herriot (FH) area model (Fay and Herriot, 1979) implements this idea. Let θ_i denote the parameter of interest of a sampled area $i = 1, \dots, m$, y_i be a direct estimator of θ_i , and \mathbf{x}_i be a covariate vector. The FH model assumes that

$$y_i = \theta_i + e_i, \quad (1)$$

$$\theta_i = \mathbf{x}_i^\top \boldsymbol{\beta} + u_i, \quad (2)$$

where $\boldsymbol{\beta}$ is a parameter vector, $u_i \stackrel{i.i.d.}{\sim} N(0, \sigma_u^2)$ with unknown σ_u , $e_i \stackrel{i.n.d.}{\sim} N(0, \Psi_i)$ with known sampling variance Ψ_i , and u_i is independent of e_i . In practice, Ψ_i is obtained by smoothing the direct estimates of the sampling variances, based on the unit level data, and then treating the smoothed estimates as the true sampling variances. Model (1) is called the “sampling model” and model (2) is referred to as the “linking model”. The empirical best linear unbiased prediction (EBLUP) estimator of θ_i for a sampled area is given by $\hat{\theta}_i = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i) \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}$, where $\hat{\gamma}_i = \hat{\sigma}_u^2 / (\Psi_i + \hat{\sigma}_u^2)$, $\hat{\boldsymbol{\beta}}$ is the best linear unbiased estimator of $\boldsymbol{\beta}$, and $\hat{\sigma}_u^2$ is the maximum likelihood estimator (MLE) or a method of moments estimator of σ_u^2 (Rao and Molina, 2015, Chapter 6). The EBLUP estimator is a weighted sum of the direct estimator y_i and the so-called “synthetic estimator” $\mathbf{x}_i^\top \hat{\boldsymbol{\beta}}$. For a non-sampled area l , the estimator of θ_l is taken as the synthetic estimator $\hat{\theta}_l = \mathbf{x}_l^\top \hat{\boldsymbol{\beta}}$.

*Carleton University, 1125 Colonel By Drive Ottawa, ON K1S 5B6, Canada

†Statistics Canada, 100 Tunney’s Pasture Driveway, Ottawa, ON K1A 0T6, Canada

‡Victoria University of Wellington, Wellington 6140, New Zealand

An extension of the FH model is composed of the sampling model (1) and a linking model of the form

$$h(\theta_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + u_i,$$

where $h(\cdot)$ is a specified monotonic function. When $h(\cdot)$ is a nonlinear function, this linking model, along with (1), is referred to as an unmatched one-fold area-level model. You and Rao (2002) used a hierarchical Bayes (HB) approach for the estimation of small area parameters θ_i under the unmatched one-fold area-level model. Sugawara et al. (2018) studied empirical best prediction (EBP) estimators of small area means under the unmatched one-fold model. Unlike the EBLUP estimator under the FH model, in general, the EBP estimator of θ_i under the unmatched model does not have a closed-form expression because the presence of the nonlinear linking function $h(\cdot)$. The EBP estimator for a sampled area i is given by

$$\hat{\theta}_i^{(EBP)} = E(\theta_i | y_i; \hat{\boldsymbol{\beta}}, \hat{\sigma}_u) = \frac{E_z[\theta_i^* \exp\{-(2\Psi_i)^{-1}(y_i - \theta_i^*)^2\}]}{E_z[\exp\{-(2\Psi_i)^{-1}(y_i - \theta_i^*)^2\}]},$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_u$ are the MLEs of $\boldsymbol{\beta}$ and σ_u , respectively, $\theta_i^* = h^{-1}(\hat{\sigma}_u z + \mathbf{x}_i^\top \hat{\boldsymbol{\beta}})$ and $z \sim N(0, 1)$. When an analytical expression is not available, the expectations in the expression for $\hat{\theta}_i^{(EBP)}$ need to be evaluated using numerical integration, for example, Monte Carlo (MC) integration or Laplace approximation.

In many applications, some areas, e.g. states, are sampled; in each sampled area, a sample of subareas, e.g. counties, is further selected. Unit-level data then are collected from the sampled subareas. The goal is to estimate a subarea parameter θ_{ij} where i denotes an area and j denotes a subarea. An example of this nested two-fold setup is given in Mohadjer et al. (2012). In this case, subareas within an area are likely to share some common characteristics and hence the variables of interest are correlated among those subareas. This correlation may be used to improve estimation accuracy of non-sampled subareas within a sampled area over simple synthetic estimators. Naively applying the FH model to the subarea-level data will not capture the correlation.

The two-fold subarea model generalizes the FH model and is tailored for the above two-fold setup. Suppose that m areas, labelled as $i = 1, \dots, m$, are sampled from M areas, and for i th sampled area, n_i subareas, labelled as $j = 1, \dots, n_i$, are further sampled from N_i subareas. Let y_{ij} , $i = 1, \dots, m$ and $j = 1, \dots, n_i$, be design-unbiased director estimators of θ_{ij} , and \mathbf{x}_{ij} be associated covariate vectors. We assume the following sampling model:

$$y_{ij} = \theta_{ij} + e_{ij}, \tag{3}$$

where $e_{ij} \stackrel{i.i.d.}{\sim} N(0, \Psi_{ij})$ with known sampling variances Ψ_{ij} . We further connect θ_{ij} to \mathbf{x}_{ij} with linking model

$$h(\theta_{ij}) = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + v_i + u_{ij}, \tag{4}$$

where $h(\cdot)$ is a specified monotonic function, $\boldsymbol{\beta}$ is a regression parameter vector, $v_i \stackrel{i.i.d.}{\sim} f_v(\boldsymbol{\eta}_v)$ is a zero-mean area-level random effect having a parametric density f_v and parameter $\boldsymbol{\eta}_v$, and $u_{ij} \stackrel{i.i.d.}{\sim} f_u(\boldsymbol{\eta}_u)$ is a zero-mean subarea-level random effect with parametric density f_u and parameter $\boldsymbol{\eta}_u$. Sampling model (3) and linking model (4) constitute the two-fold subarea model. When $h(\cdot)$ is the identity function, the model is called a matched two-fold model; when $h(\cdot)$ is a nonlinear function, the two-fold model is said to be unmatched. For modelling of certain variables of interest, an unmatched two-fold model is more appropriate than a matched one. For example, when estimating a subarea proportion, a logit

transformation on the mean would be suitable, and for income data a log transformation on small area mean is apt.

Research on two-fold subarea models has been scarce. In this paper, we present the theory of EBP estimators under unmatched two-fold models (Section 2) and a simple and effective method for variable selection under matched two-fold models (Section 3). Relevant literature reviews are provided in the respective sections. An outline of some future work is given in Section 4.

2. EBP estimators under unmatched two-fold subarea models

Different from a FH model at subarea level where all the direct estimators y_{ij} are independent, in the two-fold subarea model they are correlated across subareas within a given area because of the presence of v_i in (3). Moreover, when $h(\cdot)$ is a nonlinear link function, y_{ij} are not normally distributed even if v_i and u_{ij} are normally distributed. Mohadjer et al. (2012) used a hierarchical Bayes approach to estimate subarea parameters θ_{ij} under an unmatched two-fold model with logit link assuming normal random effects for adult literacy data. Torabi and Rao (2014) developed the theory of EBLUP estimators under a matched two-fold model. In this section, we present the theory of EBP estimators under the unmatched two-fold subarea model defined by (3) and (4).

For the given model parameter vector $\boldsymbol{\eta} = (\boldsymbol{\beta}^\top, \boldsymbol{\eta}_v^\top, \boldsymbol{\eta}_u^\top)^\top$, the best prediction (BP) estimator of the subarea parameter θ_{ij} is $\tilde{\theta}_{ij}(\boldsymbol{\eta}) = E(\theta_{ij}|\mathbf{y}; \boldsymbol{\eta})$, where $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_m^\top)^\top$ and $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})$. Let $\hat{\boldsymbol{\eta}}$ be a sensible estimator, e.g. the MLE, of $\boldsymbol{\eta}$. Then the EBP estimator of θ_{ij} is given by

$$\hat{\theta}_{ij} = \tilde{\theta}_{ij}(\hat{\boldsymbol{\eta}}) = E(\theta_{ij}|\mathbf{y}; \hat{\boldsymbol{\eta}}) \quad (5)$$

for all areas $i = 1, \dots, M$ and all subareas $j = 1, \dots, N_i$. Although conceptually simple, the EBP estimator $\hat{\theta}_{ij}$ does not have analytical solution in general unless $h(\cdot)$ is a linear function and v_i and u_{ij} are normally distributed. The computation of $E(\theta_{ij}|\mathbf{y}; \hat{\boldsymbol{\eta}})$ requires a computationally intensive method. In the following subsections, we propose an efficient method for computing $\hat{\theta}_{ij}$ based on importance sampling. Section 2.1 presents the proposed method for computing BP estimators; section 2.2 describes the computation of the MLE of model parameters and EBP estimators; section 2.3 reports the results of a simulation study.

2.1 Computing BP estimators

2.1.1 BP estimators for sampled areas

We first consider the BP estimator of θ_{ij} for a sampled subarea $j = 1, \dots, n_i$, where $i = 1, \dots, m$. Note that $\theta_{ij} = h^{-1}(\mathbf{x}_{ij}^\top \boldsymbol{\beta} + v_i + u_{ij})$, which is a function of v_i and u_{ij} . By independence of \mathbf{y}_i across $i = 1, \dots, m$, the BP estimator of θ_{ij} is given by

$$\tilde{\theta}_{ij}(\boldsymbol{\eta}) = \int \theta_{ij} f(v_i, u_{ij}|\mathbf{y}_i; \boldsymbol{\eta}) dv_i du_{ij}, \quad (6)$$

where f denotes generally a probability density function throughout the paper. In general, $f(v_i, u_{ij}|\mathbf{y}_i; \boldsymbol{\eta})$ has a complicated expression because the components of \mathbf{y}_i are correlated, which can result in inefficient computation of $\tilde{\theta}_{ij}(\boldsymbol{\eta})$.

To avoid evaluating $f(v_i, u_{ij}|\mathbf{y}_i; \boldsymbol{\eta})$, we propose to use an alternative expression

$$\tilde{\theta}_{ij}(\boldsymbol{\eta}) = \int \theta_{ij} f(v_i, \mathbf{u}_i|\mathbf{y}_i; \boldsymbol{\eta}) dv_i d\mathbf{u}_i, \quad (7)$$

where $\mathbf{u}_i = (u_{i1}, \dots, u_{in_i})^\top$. Although (7) involves integration of a higher dimension than (6), $f(v_i, \mathbf{u}_i | \mathbf{y}_i; \boldsymbol{\eta})$ has a relatively simple expression, hence (7) is easier to evaluate. By (7),

$$\tilde{\theta}_{ij}(\boldsymbol{\eta}) = \frac{\int \theta_{ij} f(v_i, \mathbf{u}_i, \mathbf{y}_i; \boldsymbol{\eta}) dv_i d\mathbf{u}_i}{\int f(v_i, \mathbf{u}_i, \mathbf{y}_i; \boldsymbol{\eta}) dv_i d\mathbf{u}_i}.$$

Both the numerator and denominator of the above expression are integrals of the following general form:

$$G(\boldsymbol{\eta}) := \int g(v_i, \mathbf{u}_i) f(v_i, \mathbf{u}_i, \mathbf{y}_i; \boldsymbol{\eta}) dv_i d\mathbf{u}_i \tag{8}$$

for some specified measurable function $g(\cdot, \cdot)$. To evaluate $G(\boldsymbol{\eta})$, we may take the advantage of the fact that v_i and \mathbf{u}_i are independent and use importance sampling with importance density function $f(v_i, \mathbf{u}_i; \boldsymbol{\eta})$. Rewrite $G(\boldsymbol{\eta})$ as

$$\begin{aligned} G(\boldsymbol{\eta}) &= \int g(v_i, \mathbf{u}_i) \frac{f(v_i, \mathbf{u}_i, \mathbf{y}_i; \boldsymbol{\eta})}{f(v_i, \mathbf{u}_i; \boldsymbol{\eta})} f(v_i, \mathbf{u}_i; \boldsymbol{\eta}) dv_i d\mathbf{u}_i \\ &= \int \{g(v_i, \mathbf{u}_i) f(\mathbf{y}_i | v_i, \mathbf{u}_i; \boldsymbol{\eta})\} f(v_i, \mathbf{u}_i; \boldsymbol{\eta}) dv_i d\mathbf{u}_i. \end{aligned}$$

To implement importance sampling, we generate a large number, K , of random vector values, $(v_{i,k}^*, \mathbf{u}_{i,k}^*)$, $k = 1, \dots, K$, from $f(v_i, \mathbf{u}_i; \boldsymbol{\eta})$, then use Monte Carlo integration to approximate $G(\boldsymbol{\eta})$ by

$$\hat{G}(\boldsymbol{\eta}) = \frac{1}{K} \sum_{k=1}^K g(v_{i,k}^*, \mathbf{u}_{i,k}^*) f(\mathbf{y}_i | v_{i,k}^*, \mathbf{u}_{i,k}^*; \boldsymbol{\eta}).$$

Since v_i and all components of \mathbf{u}_i are independent, generating random numbers from the joint distribution $f(v_i, \mathbf{u}_i; \boldsymbol{\eta})$ is equivalent to independently generating random numbers from the marginal distributions $f(v_i; \boldsymbol{\eta})$ and $f(u_{ij}; \boldsymbol{\eta})$, $j = 1, \dots, n_i$. In addition, $f(\mathbf{y}_i | v_{i,k}^*, \mathbf{u}_{i,k}^*; \boldsymbol{\eta})$ is simply the normal density with mean $h^{-1}(\mathbf{X}_i \boldsymbol{\beta} + v_{i,k}^* + \mathbf{u}_{i,k}^*)$, where $\mathbf{X}_i = (\mathbf{x}_{i1} \dots \mathbf{x}_{in_i})^\top$, and covariance matrix $\text{diag}(\Psi_{i1}, \dots, \Psi_{in_i})$. Because of the simplicity of evaluating $\hat{G}(\boldsymbol{\eta})$, computationally we can afford to use a very large K to attain a high accuracy for the MC approximation.

2.1.2 BP estimators for non-sampled subareas of sampled areas

We now turn to the BP estimator of θ_{il} for a non-sampled subarea $l = n_i + 1, \dots, N_i$ of a sampled area $i = 1, \dots, m$. The BP estimator of θ_{il} is given by $\tilde{\theta}_{il}(\boldsymbol{\eta}) = \int \theta_{il} f(v_i, u_{il} | \mathbf{y}_i; \boldsymbol{\eta}) dv_i du_{il}$. Similar to the case for sampled subareas, the density function $f(v_i, u_{il} | \mathbf{y}_i; \boldsymbol{\eta})$ has a complex expression due to the dependence among the components of \mathbf{y}_i . Following the same idea as used in (7), the BP estimator of θ_{il} can also be obtained by introducing an augmented variable \mathbf{u}_i in the integration,

$$\begin{aligned} \tilde{\theta}_{il}(\boldsymbol{\eta}) &= \int \theta_{il} f(v_i, \mathbf{u}_i, u_{il} | \mathbf{y}_i; \boldsymbol{\eta}) dv_i d\mathbf{u}_i du_{il}, \\ &= \frac{\int \theta_{il} f(v_i, \mathbf{u}_i, \mathbf{y}_i; \boldsymbol{\eta}) f(u_{il}; \boldsymbol{\eta}) dv_i d\mathbf{u}_i du_{il}}{\int f(v_i, \mathbf{u}_i, \mathbf{y}_i; \boldsymbol{\eta}) dv_i d\mathbf{u}_i}, \end{aligned} \tag{9}$$

where the second equality holds because u_{il} is independent of v_i , \mathbf{u}_i , and \mathbf{y}_i . Observing that

$$\int \theta_{il} f(v_i, \mathbf{u}_i, \mathbf{y}_i; \boldsymbol{\eta}) f(u_{il}; \boldsymbol{\eta}) dv_i d\mathbf{u}_i du_{il} = \int \{\theta_{il} f(\mathbf{y}_i | v_i, \mathbf{u}_i; \boldsymbol{\eta})\} f(v_i, \mathbf{u}_i, u_{il}; \boldsymbol{\eta}) dv_i d\mathbf{u}_i du_{il},$$

we can use importance sampling to compute the numerator in (9). Generate a large number of random values $v_{i,k}^*$, $\mathbf{u}_{i,k}^*$ and $u_{il,k}^*$ for $k = 1, \dots, K$ from $f(v_i; \boldsymbol{\eta})$, $f(\mathbf{u}_i; \boldsymbol{\eta})$ and $f(u_{il}; \boldsymbol{\eta})$, respectively. Then above integral can be approximated by

$$\frac{1}{K} \sum_{k=1}^K h^{-1}(\mathbf{x}_{ij}^\top \boldsymbol{\beta} + v_{i,k}^* + u_{il,k}^*) f(\mathbf{y}_i | v_{i,k}^*, \mathbf{u}_{i,k}^*; \boldsymbol{\eta})$$

Although there is no direct estimator for a non-sampled subarea l in a sampled area i , in the presence of the area-level random effect v_i , the direct estimators \mathbf{y}_i of the sampled subareas still carry information about θ_{il} . Clearly, the BP estimator given in (9) uses this information to aid the estimation of θ_{il} . Presumably, $\hat{\theta}_{il}(\boldsymbol{\eta})$ will be more efficient than the BP estimator under the FH model at subarea level which does not borrow strength from \mathbf{y}_i using an area-level random effect v_i .

2.1.3 BP estimators for non-sampled subareas of non-sampled areas

For subarea $l = 1, \dots, N_k$ of a non-sampled area $k = m + 1, \dots, M$, both random effects v_k and u_{kl} are independent of the data \mathbf{y} . The BP estimator of θ_{kl} hence is given by

$$\tilde{\theta}_{kl}(\boldsymbol{\eta}) = \int \theta_{kl} f(v_k; \boldsymbol{\eta}) f(u_{kl}; \boldsymbol{\eta}) dv_k du_{kl},$$

which is straightforward to compute using MC integration. The BP estimator $\tilde{\theta}_{kl}(\boldsymbol{\eta})$ is a synthetic-type estimator which only uses the available covariate information but not \mathbf{y}_i .

When $h(\cdot) = \log(\cdot)$ and the random effects v_k and u_{kl} are normally distributed, θ_{kl} follows a log-normal distribution and the analytical solution of $\tilde{\theta}_{kl}(\boldsymbol{\eta})$ is available:

$$\tilde{\theta}_{kl}(\boldsymbol{\eta}) = \exp\{\mathbf{x}_{ij}^\top \boldsymbol{\beta} + 0.5(\sigma_v^2 + \sigma_u^2)\}.$$

2.2 Parameter estimation and EBP estimators

The EBP estimator $\hat{\theta}_{ij}$ for $i = 1, \dots, M$ and $j = 1, \dots, N_i$, as given by (5), is obtained by plugging the MLE, $\hat{\boldsymbol{\eta}}$, of $\boldsymbol{\eta}$ into the expression for the BP estimator. The likelihood function for the two-fold model is given by $L(\boldsymbol{\eta}) = \prod_{i=1}^m f(\mathbf{y}_i; \boldsymbol{\eta})$, where $f(\mathbf{y}_i; \boldsymbol{\eta})$ does not have a closed-form expression in general. Observing that

$$f(\mathbf{y}_i; \boldsymbol{\eta}) = \int f(v_i, \mathbf{u}_i, \mathbf{y}_i; \boldsymbol{\eta}) dv_i d\mathbf{u}_i,$$

which is in the integral form (8) with $g(\cdot, \cdot) \equiv 1$, we can use the important sampling method discussed in Section 2.1.1 to compute $f(\mathbf{y}_i; \boldsymbol{\eta})$ for a given value of $\boldsymbol{\eta}$. That is, generate random values $(v_{i,k}^*, \mathbf{u}_{i,k}^*)$, $k = 1, \dots, K$, from $f(v_i, \mathbf{u}_i; \boldsymbol{\eta})$, then approximate $f(\mathbf{y}_i; \boldsymbol{\eta})$ by $K^{-1} \sum_{k=1}^K f(\mathbf{y}_i | v_{i,k}^*, \mathbf{u}_{i,k}^*; \boldsymbol{\eta})$.

Numerical maximization of $L(\boldsymbol{\eta})$ is computationally intensive since each evaluation of $L(\boldsymbol{\eta})$ requires MC integration. A Newton or quasi-Newton type optimization method can be used to maximize $L(\boldsymbol{\eta})$. However, such methods require numerical evaluations of the gradient of $L(\boldsymbol{\eta})$ which itself needs multiple evaluations of $L(\boldsymbol{\eta})$ for each given $\boldsymbol{\eta}$. To avoid evaluating the gradient function, we use the ‘‘BOBYQA’’ method by Powell (2009), a derivative-free bound-constrained optimization method using an iteratively constructed quadratic approximation, to maximize $L(\boldsymbol{\eta})$ in our simulation study. This method is implemented in the R package `nloptr`.

2.3 Simulation study

We now present the results of a simulation study for assessing the performance of the proposed EBP estimators under unmatched two-fold models. In the simulation, the total number of areas is set to $M = 50$. The number of subareas is set to 20 for the first 15 areas, 30 for the next 20 areas, and 15 for the last 15 areas. Without loss of generality, the first $m = 30$ areas are taken to be sampled areas, and within each sampled area i , the first n_i subareas are taken to be sampled subareas. The number of sampled subareas, n_i , is set to 8 for the first 10 sampled areas, 5 for the next 15 sampled areas, and 10 for the last 5 sampled areas. The number of simulation replications is set to 5000. The number of random draws in Monte Carlo integration is set to 30000 for computing EBP estimators and 15000 for computing the MLE of the model parameter η .

We generate subarea data y_{ij} , $i = 1, \dots, m$ and $j = 1, \dots, n_i$, according to the unmatched two-fold model defined by (3) and (4), and assess the EBP estimators of subarea means for three different classes of subareas separately: sampled subareas in sampled areas (S-S), non-sampled subareas in sampled areas (N-S), and non-sampled subareas in non-sampled areas (N-N). For each class of subareas, we consider two performance measures, average absolute bias (AABIAS) and average root mean-squared error (ARMSE), defined by

$$\begin{aligned} \text{AABIAS} &= \frac{1}{n} \sum_i \sum_j \left| \frac{1}{L} \sum_{l=1}^L (\hat{\theta}_{ij}^{(l)} - \theta_{ij}^{(l)}) \right|, \\ \text{ARMSE} &= \frac{1}{n} \sum_i \sum_j \left\{ \frac{1}{L} \sum_{l=1}^L (\hat{\theta}_{ij}^{(l)} - \theta_{ij}^{(l)})^2 \right\}^{1/2}, \end{aligned}$$

where L is the number of simulation replications, $\theta_{ij}^{(l)}$ and $\hat{\theta}_{ij}^{(l)}$ are the subarea mean and its EBP estimator, respectively, for the l th simulation replication. The summation indices i and j range over all the areas and subareas, respectively, in the class under consideration, for example, S-S, and n is the total number of subareas in the class.

2.3.1 Unmatched two-fold model with normal random effects versus one-fold model

We first evaluate the performance of EBP estimators under two-fold models with normally distributed random effects v_i and u_{ij} . In the simulation, we adopt a linking model with an intercept and a covariate,

$$h(\theta_{ij}) = \beta_0 + \beta_1 x_{ij} + v_i + u_{ij}, \quad (10)$$

where $v_i \stackrel{i.i.d.}{\sim} N(0, \sigma_v^2)$ and $u_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma_u^2)$. We consider two link functions, $h(t) = \text{logit}(t) = \log\{t/(1-t)\}$ and $h(t) = \log t$. To generate data, under each setting of the link function, we fix the model parameters $\eta = (\beta_0, \beta_1, \sigma_v, \sigma_u)$, the covariate x_{ij} and the sampling variances Ψ_{ij} . Then, for each simulation replication, we generate v_i , u_{ij} and e_{ij} from their respective distributions, and obtain y_{ij} using (3) and (10).

For the logit-link case, η is set to $(1, 1.2, 2, 1)$, x_{ij} are generated from centered Gamma(4, 3), where Gamma(α, β) denotes a gamma distribution with shape parameter α and rate parameter β , and Ψ_{ij} are generated from Unif(0.2, 0.25), where Unif(a, b) represents a uniform distribution on the interval (a, b) . For the log-link case, η is set to $(-4.5, 1.5, 1.8, 0.8)$, x_{ij} are generated from centered Gamma(4, 3), and Ψ_{ij} are generated from Unif(1.5, 2.5).

We compare the proposed EBP estimator under the unmatched two-fold model to the EBP estimator under an unmatched one-fold model (Sugasawa et al., 2018) at subarea

level. The unmatched one-fold subarea model assumes the same sampling model (3) as the two-fold subarea model, but with linking model $h(\theta_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + u_{ij}$, where there is no area-level random effect. The simulation results are reported in Table 1. Under

Table 1: Performance of EBP estimators under unmatched two-fold model and unmatched one-fold subarea model with normal random effects

Link	Model	S-S		N-S		N-N	
		AABIAS	ARMSE	AABIAS	ARMSE	AABIAS	ARMSE
logit	two-fold	0.00176	0.133	0.00245	0.181	0.00625	0.312
	one-fold	0.00219	0.169	0.00469	0.306	0.00442	0.313
log	two-fold	0.0128	0.305	0.0205	0.885	0.1485	1.696
	one-fold	0.0150	0.308	0.0988	1.380	0.1196	1.462

both settings of link function, the EBP estimators of subarea means under the two-fold model significantly outperform those under the one-fold model in N-S case, with 36%–41% reduction in ARMSE and 48%–80% reduction in AABIAS. This is because under the two-fold model, the direct estimators \mathbf{y}_i provide information about non-sampled subareas in a sampled area i through the area-level random effect v_i , while under the one-fold model \mathbf{y}_i is not used in the estimation of a non-sampled subarea. In the S-S and N-N cases, the EBP estimators under the two-fold model perform closely to those under the one-fold model. This is likely because both models use \mathbf{y} for estimation in the S-S case, and both do not directly use \mathbf{y} for computing EBP in the N-N case.

2.3.2 Unmatched two-fold model with skewed random effects

When the random effects v_i and u_{ij} are symmetrically distributed, assuming normally is usually a good practice. However, it is difficult to check symmetry assumption on the underlying distributions of the random effects. To increase the flexibility of the two-fold model, one idea is to assume a more flexible parametric family, which can capture possible skewness, for the random effects. To achieve this, we propose to use the skew-normal (SN) distribution family (Azzalini, 1985) with density function

$$f(t; \xi, \tau, \gamma) = \frac{2}{\tau} \phi\left(\frac{t - \xi}{\tau}\right) \Phi\left(\gamma \frac{t - \xi}{\tau}\right), t \in (-\infty, \infty),$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density function and cumulative distribution function of the standard normal distribution, respectively, and $\xi \in (-\infty, \infty)$, $\tau \in (0, \infty)$ and $\gamma \in (-\infty, \infty)$ are the location parameter, scale parameter and shape parameter of the distribution, respectively. The SN family, denoted $SN(\xi, \tau, \gamma)$, encompasses the normal family as a special case; when $\gamma = 0$, the SN distribution reduces to $N(\xi, \tau^2)$. It is more flexible than the normal family by allowing for moderate skewness in distribution: when $\gamma > 0$, the SN distribution is right skewed, and when $\gamma < 0$, it is left skewed. The skewness of a SN distribution takes value in the range of $(-c_0, c_0)$ with $c_0 \approx 0.995$. The mean of $SN(\xi, \tau, \gamma)$ is $\xi + \sqrt{2/\pi} \tau \gamma / \sqrt{1 + \gamma^2}$, so a centered SN distribution can be obtained by setting $\xi = -\sqrt{2/\pi} \tau \gamma / \sqrt{1 + \gamma^2}$, which will be denoted as $CSN(\tau, \gamma)$. Diallo and Rao (2018) used SN random effects under the unit-level nested-error regression model for estimating complex parameters of small areas.

We now use simulation to compare the EBP estimators under an unmatched two-fold model assuming SN random effects v_i and u_{ij} (skew-normal model fit) to those assuming normal random effects (normal model fit). For the skew-normal model fit, we use the linking model (10) with $v_i \stackrel{i.i.d}{\sim} CSN(\tau_v, \gamma_v)$ and $u_{ij} \stackrel{i.i.d}{\sim} CSN(\tau_u, \gamma_u)$.

We first compute the EBP estimators under a skew-normal model fit based on the data generated from unmatched two-fold models with normal random effects as described in Section 2.3.1. The results are shown in Table 2. Clearly, these estimators have equally good performance in terms of both AABIAS and ARMSE compared to the EBP estimators under the two-fold normal model fit reported in Table 1. This shows that using a skew-normal fit on data from a normal model does not result in notable loss of estimation efficiency.

Table 2: Performance of EBP estimators under a skew-normal model fit based on data from unmatched two-fold models with normal random effects

link	S-S		N-S		N-N	
	AABIAS	ARMSE	AABIAS	ARMSE	AABIAS	ARMSE
logit	0.00168	0.133	0.00262	0.181	0.00493	0.312
log	0.0136	0.309	0.0251	0.931	0.1225	1.607

We then compare the EBP estimators under the skew-normal model fit and the normal model fit when data are generated from an unmatched model with skew-normal random effects. When generating data, we use the linking model (10) with $h(\cdot) = \log(\cdot)$, $\beta_0 = -1.5$, $\beta_1 = 1$, $v_i \stackrel{i.i.d}{\sim} CSN(3.2, -10)$ and $u_{ij} \stackrel{i.i.d}{\sim} CSN(2, -15)$. Both random effects are left skewed. Covariate values x_{ij} are generated from centered Gamma(3, 6), and sampling variances Ψ_{ij} are generated from Unif(2, 3). The simulation results are given in Table 3. The skew-normal model fit clearly produces substantially smaller AABIAS and ARMSE than the normal model fit for the N-S and N-N cases. The particular poor performance of the normal model fit for the N-N case is likely due to the failure of taking into account the skewness in the random effects.

Table 3: Skew-normal model fit versus normal model fit for data from unmatched two-fold models with skewed random effects

Model fit	S-S		N-S		N-N	
	AABIAS	ARMSE	AABIAS	ARMSE	AABIAS	ARMSE
Skew-normal	0.0164	1.141	0.0264	1.873	0.2752	4.700
Normal	0.1194	1.199	0.0924	2.675	3.4272	39.767

Unreported results show that, when $h(\cdot)$ is a logit link, the skew-normal model fit and the normal model fit have similar performances even if data are generated from a model with skewed random effects.

3. Variable selection under two-fold subarea model

In small area estimation, it is common that data have large noise and the area or subarea sample size is small. In this context, selecting a parsimonious model that fits data well is especially important for attaining high estimation accuracy for area or subarea parameters. Meza and Lahiri (2005) proposed a variable selection method based on a transformation of the unit-level nested-error regression model. Han (2013) used a conditional Akaike information criterion (cAIC) for selecting variables under the FH model. Lahiri and Sun-tornchost (2015) proposed a variable selection method for the FH model by estimating information criteria under the linking model (2). Lombardía et al. (2017) introduced generalized Akaike information criteria (GAIC) for selecting small area models that follow a linear-mixed-model structure. In principle, since many small area models are variants of linear mixed models, a model selection method for linear mixed models may be adapted for use with small area models.

Here we propose a variable selection method especially tailored for the matched two-fold subarea model with normal random effects. Our goal is to provide a method that is simple, effective and easy to implement. The matched two-fold subarea model with normal random effects can be written in vector form as

$$\mathbf{y}_i = \boldsymbol{\theta}_i + \mathbf{e}_i, \tag{11}$$

$$\boldsymbol{\theta}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\tau}_i \tag{12}$$

for $i = 1, \dots, m$, where $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i})^\top$, $\boldsymbol{\theta}_i = (\theta_{i1}, \dots, \theta_{in_i})^\top$, $\mathbf{e}_i = (e_{i1}, \dots, e_{in_i})^\top$, and $\boldsymbol{\tau}_i = v_i \mathbb{1}_{n_i} + \mathbf{u}_i$ with $\mathbb{1}_k$ denoting a k -vector of 1s. We have $\boldsymbol{\tau}_i \sim N(0, \Sigma_i)$, where

$$\Sigma_i = \sigma_v^2 \mathbb{1}_{n_i} \mathbb{1}_{n_i}^\top + \sigma_u^2 I_{n_i} \tag{13}$$

and I_k denotes a $k \times k$ identity matrix. Note that, if $\boldsymbol{\theta}_i$ are observed and Σ_i are diagonal matrices with equal diagonal entries for all $i = 1, \dots, m$, then the linking model (12) would be a regular regression model and a information criterion (IC) such as Akaike information criterion (AIC) or Bayesian information criterion (BIC) can be used to perform variable selection. Our method is based on this simple observation and is outlined in two steps as follows.

First, we linearly transform (12) into a model with independent and identically distributed (i.i.d.) random errors. Specifically, for each $i = 1, \dots, m$, we find a matrix A_i such that $\boldsymbol{\tau}_i^* := A_i \boldsymbol{\tau}_i$ has a diagonal covariance matrix with constant diagonal entries for all i , and then transform (12) into

$$\boldsymbol{\theta}_i^* = \mathbf{X}_i^* \boldsymbol{\beta} + \boldsymbol{\tau}_i^*, \tag{14}$$

where $\boldsymbol{\theta}_i^* = A_i \boldsymbol{\theta}_i$ and $\mathbf{X}_i^* = A_i \mathbf{X}_i$. Model (14) takes the form of a regular regression model, but with unknown $\boldsymbol{\theta}_i^*$. Hence, an IC for (14), although conceptually simple, cannot be obtained. Second, we use the observed \mathbf{y}_i to obtain an estimator of an IC for (14). Variable selection is then carried out using the estimated IC.

In what follows, we present two transformation methods in section 3.1, and describe the proposed method of estimating IC in section 3.2. The results of a simulation study are given in section 3.3.

3.1 Transformation

3.1.1 Fuller-Battese transformation

The purpose of the linear transformation A_i is to make $\text{Var}(\boldsymbol{\tau}_i^*) = A_i \Sigma_i A_i^\top$ a diagonal matrix with constant diagonal entries. A straightforward idea is to take $A_i = c \Sigma_i^{-1/2}$, where $\Sigma_i^{-1/2}$ is the positive definite square-root matrix of Σ_i^{-1} and c is a non-zero constant. Choosing $c = \sigma_u^2$ and working out $\Sigma_i^{-1/2}$, we get

$$A_i = I_{n_i} - \frac{1}{n_i} \left(1 - \sqrt{\frac{1 - \rho}{1 + (n_i - 1)\rho}} \right) \mathbb{1}_{n_i} \mathbb{1}_{n_i}^\top,$$

where $\rho = \sigma_v^2 / (\sigma_v^2 + \sigma_u^2)$. This is the same as the transformation used by Fuller and Battese (1973). Under the transformation, $\text{Var}(\boldsymbol{\tau}_i^*) = \sigma_u^2 I_{n_i}$.

In practice, ρ has to be estimated. One can use the estimating equation method by Torabi and Rao (2014) or the maximum likelihood (ML) method to estimate ρ .

3.1.2 A parameter-free transformation

It is possible to avoid inclusion of an unknown parameter in the transformation matrix A_i . By (13),

$$\text{Var}(\boldsymbol{\tau}_i^*) = A_i \Sigma_i A_i^\top = \sigma_v^2 (A_i \mathbb{1}_{n_i})(A_i \mathbb{1}_{n_i})^\top + \sigma_u^2 A_i A_i^\top.$$

Hence, to make a diagonal structure for $\text{Var}(\boldsymbol{\tau}_i^*)$, it suffices to find an A_i such that

- (a) $A_i \mathbb{1}_{n_i} = 0$,
- (b) $A_i A_i^\top$ is a diagonal matrix with constant diagonal entries.

Note that the rank of such an A_i is at most $n_i - 1$ because of the linear constraint (a).

This transformation was proposed by Lahiri and Li (2009), and Li and Lahiri (2018) used the transformation for variable selection under a unit-level nested-error regression model. Particular examples of A_i that satisfy the conditions (a) and (b) were given therein but no general method for finding A_i was suggested.

Here we give a general method to construct a desired A_i as follows.

Step 1: Fix a set of $n_i - 1$ linearly independent vectors of length n_i , denoted b_1, \dots, b_{n_i-1} , which satisfies $b_k^\top \mathbb{1}_{n_i} = 0$ for $k = 1, \dots, n_i - 1$. For example, one can take b_k to be the vector with k th entry being 1, the last entry being -1 and all the other entries being 0, or, the vector with k th entry being 1, the $(k + 1)$ th entry being -1 and all the other entries being 0.

Step 2: Apply the Gram-Schmidt process to b_1, \dots, b_{n_i-1} to obtain a set of orthogonal vectors a_1, \dots, a_{n_i-1} with $a_1 = b_1$ and $a_k = b_k - \sum_{l=1}^{k-1} \text{Proj}_{a_l}(b_k)$ for $k = 2, \dots, n_i - 1$, where $\text{Proj}_y(x) := \frac{x^\top y}{y^\top y} y$ is the projection of vector x on vector y . Take $A_i = [a_1 \ \dots \ a_{n_i-1}]^\top$.

The A_i constructed this way satisfies the requirements (a) and (b).

In spite of being parameter free, this transformation has two drawbacks: (1) Since the rank of A_i is $n_i - 1$ instead of n_i , each area i loses one observation after transformation, which is undesirable when the number of sampled areas, m , is large. (2) After transformation, the intercept term, if included in the original model, will be removed because of the requirement (a). Hence, if the intercept is to be selected, this transformation cannot be used. Moreover, a transformation matrix that satisfies (a) and (b) is not unique, although we do not find that using different parameter-free transformation matrices affects variable selection results significantly.

3.2 Estimating information criteria of transformed linking model

The transformed linking model (12) is a regular regression model with unobserved responses $\boldsymbol{\theta}_i^*$. Define the mean sum of squares of errors (MSE) of (12) as

$$\text{MSE}_{\boldsymbol{\theta}^*} = \frac{1}{n^* - p} \boldsymbol{\theta}^{*\top} (I_{n^*} - P^*) \boldsymbol{\theta}^*,$$

where $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^{*\top} \ \dots \ \boldsymbol{\theta}_m^{*\top})^\top$ and $P^* = \mathbf{X}^* (\mathbf{X}^{*\top} \mathbf{X}^*)^{-1} \mathbf{X}^{*\top}$ with $\mathbf{X}^* = (\mathbf{X}_1^{*\top} \ \dots \ \mathbf{X}_m^{*\top})^\top$, n^* is the length of $\boldsymbol{\theta}^*$, and p is the dimension of $\boldsymbol{\beta}$.

For a sub-model of (12) with p_s covariates, the AIC, BIC and Mallorw's C_p are given, respectively, as

$$\begin{aligned} \text{AIC}^{(s)} &= n^* \log\{(n^* - p_s) \text{MSE}_{\theta^*}^{(s)} / n^*\} + 2p_s, \\ \text{BIC}^{(s)} &= n^* \log\{(n^* - p_s) \text{MSE}_{\theta^*}^{(s)} / n^*\} + p_s \log(n^*), \\ C_p^{(s)} &= (n^* - p_s) \text{MSE}_{\theta^*}^{(s)} / \text{MSE}_{\theta^*} + p_s - 2n^*, \end{aligned}$$

where $\text{MSE}_{\theta^*}^{(s)}$ is the MSE of the sub-model. If θ^* is known, we can choose one of the above information criteria and compute its value for a set of sub-models under consideration; the sub-model with the smallest IC value then is selected as the final model.

Given that θ^* is unknown, we aim to find an estimator, denoted $\widehat{\text{MSE}}_{\theta^*}$, of MSE_{θ^*} . Observing that each of the above IC is a function of MSE_{θ^*} , estimators of IC can be obtained by plugging in $\widehat{\text{MSE}}_{\theta^*}$. Variable selection then can be carried out using the estimated IC. Lahiri and Suntornchost (2015) used this idea for variable selection under the one-fold FH model.

Let $\mathbf{y}_i^* = A_i \mathbf{y}_i$ and $\mathbf{y}^* = (\mathbf{y}_1^{*\top} \dots \mathbf{y}_m^{*\top})^\top$. Define $\text{MSE}_{\mathbf{y}^*} = \frac{1}{n^* - p} \mathbf{y}^{*\top} (I_{n^*} - P^*) \mathbf{y}^*$. We propose to estimate MSE_{θ^*} by

$$\widehat{\text{MSE}}_{\theta^*} = \text{MSE}_{\mathbf{y}^*} - \frac{1}{n^* - p} \text{tr}\{(I_{n^*} - P^*) A V_e A^\top\},$$

where $A = \text{diag}(A_1, \dots, A_m)$ and $V_e = \text{diag}(\Psi_{11}, \dots, \Psi_{mn_m})$. The second term on the right hand side of the above equation can be viewed as a bias-correction term. It can be shown that, if Ψ_{ij} is bounded for all i and j and n_i is bounded for all i , then as the number of areas $m \rightarrow \infty$,

$$\widehat{\text{MSE}}_{\theta^*} = \text{MSE}_{\theta^*} + o_p(1).$$

Estimates of AIC, BIC and Mallorw's C_p are then obtained by plugging $\widehat{\text{MSE}}_{\theta^*}$ into their corresponding expressions. By the continuous mapping theorem (van der Vaart, 1998, Theorem 2.3), the errors of the estimated IC are also of $o_p(1)$.

3.3 Simulation study

We conducted a small simulation study to assess the performance of the proposed variable selection method. We consider a matched two-fold subarea model with normal random effects $v_i \stackrel{i.i.d}{\sim} N(0, \sigma_v^2)$ and $u_{ij} \stackrel{i.i.d}{\sim} N(0, \sigma_u^2)$. The number of sampled areas m is set to 30. The number of sampled subareas is set to 8 for the first 10 sampled areas, 5 for the next 15 sampled areas, and 10 for the last 5 sampled areas. The sampling standard deviation $\sqrt{\Psi_{ij}}$ are generated from $\text{Unif}(0.5, 1.5)$. We set $\sigma_u = 2$ and consider a few settings for the standard deviation of the area-level random effect with $\sigma_v = 2, 3.5, 5, 6.5$ and 8. In the linking model, we consider an intercept and five covariates with

$$\begin{aligned} \mathbf{x}_{ij,1} &\sim \text{log-normal}(0.3, 0.5), \quad \mathbf{x}_{ij,2} \sim \text{gamma}(1.5, 2), \quad \mathbf{x}_{ij,3} \sim N(0, 0.8), \\ \mathbf{x}_{ij,4} &\sim \text{gamma}(0.6, 10), \quad \mathbf{x}_{ij,5} \sim \text{beta}(0.5, 0.5), \end{aligned}$$

where $\mathbf{x}_{ij,k}$ represents the value of the k th covariate for the i th area and j th subarea, $\text{log-normal}(\mu, \sigma)$ denotes a log-normal distribution with mean μ and standard deviation σ on log scale, and $\text{beta}(\kappa, \gamma)$ stands for a beta distribution with shape parameter κ and scale parameter γ . The regression parameter is set to $\beta = (2, 0, 0, 4, 8, 0)^\top$; the corresponding true model is the sub-model with an intercept and covariates $(\mathbf{x}_{ij,3}, \mathbf{x}_{ij,4})$. The number of simulation replications is set to 3000.

In each simulation replication, data $(\mathbf{y}_i, \mathbf{X}_i)$, $i = 1, \dots, m$, are generated from the matched two-fold subarea model using the above setting. We use the proposed method to select covariates by comparing all sub-models defined by the subsets of the five covariates. We consider the proposed method using the Fuller-Battese transformation with the true ρ value ($\text{TWOF}_1(\rho_0)$), that with the MLE of ρ ($\text{TWOF}_1(\hat{\rho}_{mle})$), and that with the estimated ρ based on the estimating equation of Torabi and Rao (2014) ($\text{TWOF}_1(\hat{\rho}_{ee})$). We also apply the proposed method with the parameter-free transformation (TWOF_2). For comparison, we consider two naive competitors, the method of Lahiri and Suntornchost (2015) for FH model (Naive 1) and information criterion approach for regular linear regression model fitted naively to the data (Naive 2).

The simulation results using BIC for variable selection are reported in Table 4. The pro-

Table 4: Percentage (%) of selecting the true model $\beta = (2, 0, 0, 4, 8, 0)^T$ using BIC

Method	σ_v				
	2	3.5	5	6.5	8
$\text{TWOF}_1(\rho_0)$	86.63	85.90	86.47	86.20	85.63
$\text{TWOF}_1(\hat{\rho}_{mle})$	86.43	85.50	86.43	86.43	85.60
$\text{TWOF}_1(\hat{\rho}_{ee})$	86.47	85.53	86.43	86.50	85.47
TWOF_2	85.77	85.10	85.87	85.67	84.97
Naive 1	76.67	47.37	25.53	13.63	7.73
Naive 2	76.40	45.40	24.40	12.70	7.40

posed methods have significantly higher percentage of selecting the true model in all cases. When the standard deviation σ_v of the area-level random effect increases, the proposed methods exhibit stable rate of selecting the true model at approximately 85% level, while both naive methods show dramatic decay in performance to nearly 7% when $\sigma_v = 8$. This suggests that when there is a strong area-level effect, as it commonly happens in practice, the proposed methods are clear choices over the naive ones. The proposed methods based on the Fuller-Battese transformation and that based on the parameter-free method perform equally well. Moreover, using an estimated ρ instead of the true value of ρ in the Fuller-Battese transformation does not adversely affect the performance of variable selection in this case.

The simulation results using AIC for variable selection are given in Table 5. Compared

Table 5: Percentage (%) of selecting the true model using AIC

Method	σ_v				
	2	3.5	5	6.5	8
$\text{TWOF}_1(\rho_0)$	54.80	53.63	53.13	52.40	52.87
$\text{TWOF}_1(\hat{\rho}_{mle})$	53.87	53.37	52.63	52.33	52.87
$\text{TWOF}_1(\hat{\rho}_{ee})$	54.07	53.50	52.70	52.47	52.97
TWOF_2	54.90	53.57	53.30	52.23	52.23
Naive 1	53.07	47.07	37.30	29.70	21.40
Naive 2	55.53	48.00	37.43	29.80	21.40

to BIC, AIC yields lower percentage of selecting the true model for all the methods. However, the comparison between the proposed methods and the naive methods are similar to the case using BIC. The proposed methods perform similarly and give stable results for different values of σ_v . The naive methods, on the other hand, have poorer performance, and their performance drop considerably as σ_v increases.

The simulation results using Mallows's C_p for variable selection are similar to those using AIC, and hence are omitted for brevity.

4. Concluding remarks

Our immediate future work on EBP estimation under unmatched two-fold subarea models includes using parametric bootstrap to estimate mean squared prediction error (MSPE) of the EBP estimator of a subarea parameter and extending the EBP theory to an unmatched three-fold sub-subarea model.

Additionally, we are working on extending the proposed variable selection method for matched two-fold model and the cAIC method for FH model by Han (2013) to the unmatched two-fold model case.

This work was supported by research grants to Song Cai and J. N. K. Rao from the Natural Sciences and Engineering Research Council of Canada.

References

- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2):171–178.
- Diallo, M. S. and Rao, J.N.K.. (2018). Small area estimation of complex parameters under unit-level models with skew-normal errors. *Scandinavian Journal of Statistics*, 45(4):1092–1116.
- Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: An application of james-stein procedures to census data. *Journal of the American Statistical Association*, 74(366a):269–277.
- Fuller, W. A. and Battese, G. E. (1973). Transformation for estimation of linear models with nested-error structure. *Journal of the American Statistical Association*, 68(343):626–632.
- Han, B. (2013). Conditional Akaike information criterion in the Fay-Herriot model. *Statistical Methodology*, 11:53–67.
- Lahiri, P. and Li, Y. (2009). A new alternative to the standard f test for clustered data. *Journal of Statistical Planning and Inference*, 139:3430–3441.
- Lahiri, P. and Suntornchost, J. (2015). Variable selection for linear mixed models with applications in small area estimation. *Sankhyā*, 77-B(2):312–320.
- Li, Y. and Lahiri, P. (2018). A simple adaptation of variable selection software for regression models to select variables in nested error regression models. *Sankhyā B*.
- Lombardía, M. J., López-Vizcaíno, E., and Rueda, C. (2017). Mixed generalized akaike information criterion for small area models. *Journal of the Royal Statistical Society, Series A*, 180(4):1229–1252.
- Meza, J. L. and Lahiri, P. (2005). A note on the C_p statistic under the nested error regression model. *Survey Methodology*, 31(1):105–109.
- Mohadjer, L., Rao, J.N.K., Liu, B., Krenzke, T., and Van de Kerckhove, W. (2012). Hierarchical Bayes small area estimates of adult literacy using unmatched sampling and linking models. *Journal of the Indian Society of Agricultural Statistics*, 66(1):55–63.

- Powell, M.J.D. (2009). The BOBYQA algorithm for bounded constrained optimization without derivatives. Technical report, University of Cambridge, Cambridge, UK.
- Rao, J.N.K. and Molina, I. (2015). *Small Area Estimation*. Wiley, Hoboken, New Jersey, second edition.
- Sugasawa, S., Kubokawa, T., and Rao, J.N.K. (2018). Small area estimation via unmatched sampling and linking models. *TEST*, 27(2):407–427.
- Torabi, M. and Rao, J.N.K. (2014). One small area estimation under a sub-area level model. *Journal of Multivariate Analysis*, 127:36–55.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge, New York.
- You, Y. and Rao, J.N.K.. (2002). Small area estimation using unmatched sampling and linking models. *The Canadian Journal of Statistics*, 30(1):3–15.