Bayesian Curve-Fitting in Semiparametric Small Area Models with Measurement Errors

Jinseub Hwang[a], Dal Ho Kim[1,b]

[a]National Evidence-based Healthcare Collaborating Agency, Korea;
[b]Department of Statistics, Kyungpook National University, Korea

Abstract

We study a semiparametric Bayesian approach to small area estimation under a nested error linear regression model with area level covariate subject to measurement error. Consideration is given to radial basis functions for the regression spline and knots on a grid of equally spaced sample quantiles of covariate with measurement errors in the nested error linear regression model setup. We conduct a hierarchical Bayesian structural measurement error model for small areas and prove the propriety of the joint posterior based on a given hierarchical Bayesian framework since some priors are defined non-informative improper priors that uses Markov Chain Monte Carlo methods to fit it. Our methodology is illustrated using numerical examples to compare possible models based on model adequacy criteria; in addition, analysis is conducted based on real data.

Keywords: Hierarchical Bayesian, nested error regression, penalized spline, radial basis, semi-parametric, small areas, structural measurement

1. Introduction

Sample surveys are designed to provide estimates of totals and means of items of interest for large subpopulations (or domains). These estimates are "direct" because they use only domain-specific data with domain sample sizes large enough to support reliable direct "design-based" estimates. However, it is seldom possible to procure a large enough sample size to support direct estimates for small local areas such as counties, and school districts. The term "small area" to denote any domain for which direct estimates of adequate precision cannot be produced due to a small domain-specific sample size. It is often necessary to employ "indirect" estimates for small areas that can increase the "effective" domain sample size by "borrowing strength" from related areas through linking models and using some auxiliary data associated with the small areas. A comprehensive account of model-based small area estimation is given by Rao (2003).

Census and administrative data are often taken for auxiliary data in small area estimation with the assumption that auxiliary data are available for all areas and are measured without error. However, sometimes we can use auxiliary data from another survey and in this case the auxiliary data may be measured with error. This measurements error complicates the statistical analysis and this problem is commonly called measurement error problem (Fuller, 1987). This paper considers the situation in which auxiliary data are available for use in small area estimation, but where auxiliary data may be measured with error.

---

[1] Corresponding author: Department of Statistics, Kyungpook National University, 80 Daehakro, Bukgu, Daegu 702-701, Korea. E-mail: dalkim@knu.ac.kr

Semiparametric regression methods have not been used in small area estimation contexts until recently. This was mainly due to methodological difficulties in combining different smoothing techniques with estimation tools generally used in small area estimation. The pioneering contribution was due to work by Opsomer *et al.* (2008) that combined small area random effects with a smooth non-parametrically specified trend using penalized splines. Bhadra *et al.* (2013) considered a Bayesian semiparametric modeling procedure that used penalized splines to estimate the median household income for all of the United States.

In our previous model (Hwang and Kim, 2010), we developed a Bayesian semiparametric model to deal with the small area problem with measurement errors. We considered penalized splines for non-linear pattern that uses truncated polynomial basis functions (TPBF) with knots of equally spaced sample quantiles under a structural measurement error model. It follows from Hwang and Kim (2010) that the semiparametric model was superior to the model without nonparametric component (Ghosh *et al.*, 2006).

This paper studies a Bayesian semiparametric model with radial basis functions (RBF) under a structural measurement error model for the small area problem with measurement errors. TPBF are not always numerically stable when the number of knots is large and the smoothing parameter is close to zero. Instead of TPBF, B-splines and RBF can deal with this problem (Ruppert *et al.*, 2003). To fit our model and estimate parameters, we have conducted a hierarchical Bayesian (HB) approach with a Markov Chain Monte Carlo (MCMC) methodology that uses a Gibbs sampling and Metropolis-Hastings (M-H) algorithm.

We begin with a brief overview of the model specification and our notations in Section 2. In Section 3, we prove the propriety of the posterior and established the MCMC implementation of the proposed hierarchical Bayes procedure. In Section 4, we exhibit the performance of the proposed model based on simulation studies. We have checked the convergence and model adequacy using a potential scale reduction factor (PSRF) and the posterior predictive p-value (Meng, 1994); in addition, we also used root mean squared errors to compare models. We conduct the analysis based on real data in Section 5 and Section 6 discusses possible extensions of our modeling.

2. Model Specification and Notations

We consider a nested error regression model setup to estimate small area means originally considered by Battese *et al.* (1988). Suppose there are m strata and each strata consists of the known population size N_i. Let (y_{ij}, X_{ij}) denote the observed response and covariate of the jth unit in the ith stratum $(j = 1, \ldots, N_i; i = 1, \ldots, m)$, respectively. The basis model can be expressed as

$$y_{ij} = f(x_i) + u_i + e_{ij}, \tag{2.1}$$

where u_i and e_{ij} are the random effects and sampling errors with independent and identically distributed (i.i.d.) normal random variables, respectively. Here $f(x_i)$ is a smoothing function of x_i reflecting the unknown response-covariate relationship; in addition, we use penalized splines with RBF. RBF is defined as

$$1, x, \ldots, x^{p-1}, |x - \tau_1|^{2p-1}, \ldots, |x - \tau_K|^{2p-1},$$

where $|\cdot|$ is the function of absolute value. Here K is the number of knots and we called this as the tuning parameter in our model. And $\tau = (\tau_1, \ldots, \tau_K)^T$ is the vector of knot locations on a grid equally spaced sample quantiles of covariate x_i $(\tau_1 < \cdots < \tau_K)$. In our paper we consider $p = 1$ case, then

we can rewrite (2.1) as:

$$y_{ij} = b_0 + b_1 x_i + \sum_{k=1}^{K} |x_i - \tau_k| + u_i + e_{ij}.$$

Our superpopulation model with structural measurement error covariates is assumed as:

$$y_{ij} = \boldsymbol{x}_i^T \boldsymbol{b} + \boldsymbol{z}_i^T \boldsymbol{\gamma} + u_i + e_{ij} \tag{2.2}$$
$$= \theta_i + e_{ij};$$
$$X_{ij} = x_i + \eta_{ij},$$

where $\boldsymbol{x}_i = (1, x_i)^T$, $\boldsymbol{z} = (|x_i - \tau_1|, \ldots, |x_i - \tau_K|)^T$ and η_{ij} is the measurement error with i.i.d. normal distribution. We consider structural measurement error model, so let x_i is a normal random variable. Here $\boldsymbol{b} = (b_0, b_1)^T$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)^T$ are the vector of regression and spline coefficients, respectively. Specifically, $\theta_i (= \boldsymbol{x}_i^T \boldsymbol{b} + \boldsymbol{z}_i^T \boldsymbol{\gamma} + u_i)$ is the small area means that we want to estimate. Finally, we assumed that x_i, u_i, e_{ij} and η_{ij} are mutually independent with $x_i \sim N(\mu, \sigma_x^2)$, $u_i \sim N(0, \sigma_u^2)$, $e_{ij} \sim N(0, \sigma_e^2)$ and $\eta_{ij} \sim N(0, \sigma_\eta^2)$, respectively.

3. Hierarchical Bayesian Estimation Procedure

3.1. Hierarchical Bayesian framework

To fit the model and estimate parameters based on sample (n_i is drawn from the ith stratum and $\sum_{i=1}^{m} n_i = n_t$), we conduct a hierarchical Bayesian framework based on (2.2):

Stage 1. $y_{ij} = \theta_i + e_{ij}$ $(j = 1, \ldots, n_i; i = 1, \ldots, m)$

Stage 2. $\theta_i = \boldsymbol{x}_i^T \boldsymbol{b} + \boldsymbol{z}_i^T \boldsymbol{\gamma} + u_i$ $(i = 1, \ldots, m)$

Stage 3. $X_{ij} = x_i + \eta_{ij}$ $(j = 1, \ldots, n_i; i = 1, \ldots, m)$

Stage 4. $x_i \sim N(\mu_x, \sigma_x^2)$.

Stage 5. $\boldsymbol{\gamma} \sim N(0, \sigma_\gamma^2 I)$.

Stage 6. $b_0, b_1, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \sigma_\eta^2$ and σ_γ^2 are mutually independent with b_0, b_1 and μ_x i.i.d. uniform $(-\infty, \infty)$; $(\sigma_e^2)^{-1} \sim G(a_e, b_e)$, $(\sigma_u^2)^{-1} \sim G(a_u, b_u)$, $(\sigma_\eta^2)^{-1} \sim G(a_\eta, b_\eta)$, $(\sigma_\gamma^2)^{-1} \sim G(a_\gamma, b_\gamma)$, $(\sigma_x^2)^{-1} \sim G(a_x, b_x)$, where $G(\alpha, \beta)$ denotes a gamma distribution shape parameter α and rate parameter β having the expression $f(x) \propto x^{\alpha-1} \exp(-\beta x)$.

3.2. Propriety of the joint posterior

Before the implementation, we check the propriety of the joint posterior based on given a hierarchical Bayesian framework since some priors are defined non-informative improper priors for regression parameters b_0 and b_1, and mean parameter μ_x of x_i. We can factorize the full posterior by conditional independence properties as

$$\left[\boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \mu_x, \sigma_x^2, \sigma_e^2, \sigma_u^2, \sigma_\eta^2, \sigma_\gamma^2 \,\middle|\, \mathbf{X}, \mathbf{y}\right]$$
$$\propto \left[\mathbf{y} \,\middle|\, \boldsymbol{\theta}, \sigma_e^2\right]\left[\boldsymbol{\theta} \,\middle|\, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \sigma_u^2, \mathbf{X}\right]\left[\mathbf{X} \,\middle|\, \boldsymbol{x}, \sigma_\eta^2\right]\left[\boldsymbol{x} \,\middle|\, \mu_x, \sigma_x^2\right]\left[\boldsymbol{\gamma} \,\middle|\, \sigma_\gamma^2\right][\boldsymbol{b}]\,[\mu_x]\left[\sigma_x^2\right]\left[\sigma_e^2\right]\left[\sigma_u^2\right]\left[\sigma_\eta^2\right]\left[\sigma_\gamma^2\right].$$

Theorem 1. *Assume that a_e, a_u, a_x, a_η and a_γ are all positive. Also, let $a_u + m/2 - p/2 > 0, a_x + m/2 - 1 > 0$ and $a_\eta + n_t/2 - m/2 > 0$ where $p = rank(X_*)$ and $X_* = (x_1^T, \ldots, x_m^T)^T$. Then the joint posterior is proper.*

Proof: The basic full parameter space is $\boldsymbol{\Omega} = \{\boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \mu_x, \sigma_x^2, \sigma_e^2, \sigma_u^2, \sigma_\eta^2, \sigma_\gamma^2\}$. Let

$$
\begin{aligned}
I &= \int \cdots \int p(\boldsymbol{\Omega} \,|\, \mathbf{y}, \mathbf{X}) d\boldsymbol{\Omega} \\
&= \int \cdots \int \left[\mathbf{Y} \,|\, \boldsymbol{\theta}, \sigma_e^2\right]\left[\boldsymbol{\theta} \,|\, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \sigma_u^2, \mathbf{X}\right]\left[\mathbf{X} \,|\, \boldsymbol{x}, \sigma_\eta^2\right]\left[\boldsymbol{x} \,|\, \mu_x, \sigma_x^2\right] \\
&\quad \times \left[\boldsymbol{\gamma} \,|\, \sigma_\gamma^2\right][\boldsymbol{b}][\mu_x]\left[\sigma_e^2\right]\left[\sigma_u^2\right]\left[\sigma_\eta^2\right]\left[\sigma_\gamma^2\right]\left[\sigma_x^2\right] d\boldsymbol{\Omega}.
\end{aligned}
\tag{3.1}
$$

To show the propriety of the joint posterior, we have to prove that $I \le M$ where M is any finite positive constant.

First, integrating with respect to μ_x based on $\exp[-1/(2\sigma_x^2) \sum_{i=1}^m (x_i - \bar{x})^2] \le 1$,

$$
\begin{aligned}
I_{\mu_x} &= \int \left[\boldsymbol{x} \,|\, \mu_x, \sigma_x^2\right][\mu_x]d\mu_x \\
&= \left(\sigma_x^2\right)^{-\frac{m}{2}} \int \exp\left[-\frac{1}{2\sigma_x^2} \sum_{i=1}^m (x_i - \mu_x)^2\right] d\mu_x \\
&= \left(\sigma_x^2\right)^{-\frac{m}{2}} \exp\left[-\frac{1}{2\sigma_x^2} \sum_{i=1}^m (x_i - \bar{x})^2\right] \int \exp\left[-\frac{1}{2\sigma_x^2} m (\mu_x - \bar{x})^2\right] d\mu_x \\
&\le K_1 \cdot \left(\sigma_x^2\right)^{-\frac{m-1}{2}},
\end{aligned}
\tag{3.2}
$$

where K_1 is a constant. Second, integrating with respect to $\boldsymbol{b}$ with $\mathbf{w}^T(I - P_{X_*})\mathbf{w} \ge 0$,

$$
\begin{aligned}
I_{\boldsymbol{b}} &= \int \left[\boldsymbol{\theta} \,|\, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \sigma_u^2, \mathbf{X}\right][b]d\boldsymbol{b} \\
&= \left(\sigma_u^2\right)^{-\frac{m}{2}} \int \exp\left[-\frac{1}{2\sigma_u^2} \sum_{i=1}^m \left(\theta_i - x_i^T \boldsymbol{b} - z_i^T \boldsymbol{\gamma}\right)^2\right] d\boldsymbol{b} \\
&= \left(\sigma_u^2\right)^{-\frac{m}{2}} \int \exp\left[-\frac{1}{2\sigma_u^2} \sum_{i=1}^m \left(w_i - x_i^T \boldsymbol{b}\right)^2\right] d\boldsymbol{b} \\
&= \left(\sigma_u^2\right)^{-\frac{m}{2}} \int \exp\left\{-\frac{1}{2\sigma_u^2} \mathbf{w}^T (I - P_{X_*}) \mathbf{w}\right\} d\boldsymbol{b} \left(\sigma_u^2\right)^{\frac{2}{p}} \left|\mathbf{X}_*^T \mathbf{X}_*\right|^{-\frac{1}{2}} (2\pi)^{\frac{m}{2}} \\
&\le K_2 \cdot \left(\sigma_u^2\right)^{-\frac{(m-p)}{2}} \cdot \left|\mathbf{X}_*^T \mathbf{X}_*\right|^{-\frac{1}{2}},
\end{aligned}
\tag{3.3}
$$

where K_2 is a constant and $P_{\mathbf{X}_*} = \mathbf{X}_*(\mathbf{X}_*^T \mathbf{X}_*)^{-1} \mathbf{X}_*^T$. Next, we integrate with respect to $\boldsymbol{x}$ based on the method by Ghosh *et al.* (2006).

$$
\begin{aligned}
I_{\boldsymbol{x}} &= \int \left[\mathbf{X}|\boldsymbol{x}, \sigma_\eta^2\right]\left|\mathbf{X}_*^T \mathbf{X}_*\right|^{-\frac{1}{2}} d\boldsymbol{x} \\
&= \left(\sigma_\eta^2\right)^{-\frac{n_t}{2}} \exp\left[-\frac{1}{2\sigma_\eta^2} \sum_{i=1}^m \sum_{j=1}^{n_i} \left(X_{ij} - \bar{X}_i\right)^2\right] \int \left|\mathbf{X}_*^T \mathbf{X}_*\right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma_\eta^2} \sum_{i=1}^m n_i \left(\bar{X}_i - x_i\right)^2\right] d\boldsymbol{x}
\end{aligned}
$$

$$\leq K_3' \cdot \left(\sigma_\eta^2\right)^{-\frac{n_t-m}{2}} \exp\left[-\frac{1}{2\sigma_\eta^2} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left(X_{ij} - \bar{X}_i\right)^2\right]$$

$$\leq K_3 \cdot \left(\sigma_\eta^2\right)^{-\frac{n_t-m}{2}}, \tag{3.4}$$

where K_3' and K_3 are some constants. Now, integrating with respect to σ_u^2, σ_x^2 and σ_η^2, using gamma distribution,

$$I_{\sigma_u^2} = \int \left(\sigma_u^2\right)^{-\frac{m-p}{2}} \left[\sigma_u^2\right] d\sigma_u^2 = \int \left(\sigma_u^2\right)^{-(a_u+\frac{m}{2}-\frac{p}{2})-1} \exp\left(-\frac{b_u}{\sigma_u^2}\right) d\sigma_u^2 = K_4, \tag{3.5}$$

$$I_{\sigma_x^2} = \int \left(\sigma_x^2\right)^{-\frac{m-1}{2}} \left[\sigma_x^2\right] d\sigma_x^2 = \int \left(\sigma_x^2\right)^{-(a_x+\frac{m}{2}-1)-1} \exp\left(-\frac{b_x}{\sigma_x^2}\right) d\sigma_x^2 = K_5, \tag{3.6}$$

$$I_{\sigma_\eta^2} = \int \left(\sigma_\eta^2\right)^{-\frac{n_t-m}{2}} \left[\sigma_\eta^2\right] d\sigma_\eta^2 = \int \left(\sigma_\eta^2\right)^{-(a_\eta+\frac{n_t}{2}-\frac{m}{2})-1} \exp\left(-\frac{b_\eta}{\sigma_\eta^2}\right) d\sigma_\eta^2 = K_6, \tag{3.7}$$

where K_4, K_5 and K_6 are some constants.

Combining (3.1)–(3.7), we have

$$I \leq K_1 K_2 K_3 K_4 K_5 K_6 \int \cdots \int \left[\mathbf{y} \,|\, \boldsymbol{\theta}, \sigma_e^2\right] \left[\boldsymbol{\gamma} \,|\, \sigma_\gamma^2\right] \left[\sigma_e^2\right] \left[\sigma_\gamma^2\right] d\boldsymbol{\Omega}^*, \tag{3.8}$$

where $\boldsymbol{\Omega}^* = (\boldsymbol{\Omega} - \mu_x - \boldsymbol{b} - \boldsymbol{x} - \sigma_u^2 - \sigma_x^2 - \sigma_\eta^2)$. The posterior propriety is proved since the above integral would be finite due to all components of the integrand in (3.8) having proper distributions. □

3.3. Full conditional distributions

To fit the model, we use MCMC integration technique for the implementation of the Bayesian procedure. The M-H algorithm and Gibbs sampler are used because x_i has not full conditional distribution of closed form. We generate samples from the full conditional distribution of each $\boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \mu_x, \sigma_x^2, \sigma_e^2, \sigma_u^2, \sigma_\eta^2$ and σ_γ^2 given the remaining parameters and the data. Gibbs sampling and M-H algorithm are based on the following full conditional distributions:

(1) $\left[\theta_i \,\Big|\, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \mu_x, \sigma_x^2, \sigma_e^2, \sigma_u^2, \sigma_\gamma^2, \sigma_\eta^2, X, \boldsymbol{y}\right] \overset{iid}{\sim} N\left[(1-C_i)\,\bar{y}_i + C_i\left(x_i^T \boldsymbol{b} + z_i^T \boldsymbol{\gamma}\right), \frac{\sigma_e^2}{n_i}\,(1-C_i)\right]$,

   where $C_i = \sigma_e^2/(\sigma_e^2 + n_i \sigma_u^2)$;

(2) $\left[\boldsymbol{b} \,\Big|\, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{x}, \mu_x, \sigma_x^2, \sigma_e^2, \sigma_u^2, \sigma_\gamma^2, \sigma_\eta^2, X, \boldsymbol{y}\right] \sim N\left[\left(X_*^T X_*\right)^{-1} X_*^T \mathbf{w}, \sigma_u^2 \left(X_*^T X_*\right)^{-1}\right]$,

   where $X_* = (x_1^T, \ldots, x_m^T)^T$, $\mathbf{w} = (w_1, \ldots, w_m)^T$, $w_i = \theta_i - z_i^T \boldsymbol{\gamma}$;

(3) $\left[\boldsymbol{\gamma} \,\Big|\, \boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{x}, \mu_x, \sigma_x^2, \sigma_e^2, \sigma_u^2, \sigma_\gamma^2, \sigma_\eta^2, X, \boldsymbol{y}\right] \sim N\left[\left(\frac{\mathbf{Z}_*^T \mathbf{Z}_*}{\sigma_u^2} + \frac{I}{\sigma_\gamma^2}\right)^{-1} \frac{\mathbf{Z}_*^T}{\sigma_u^2} \mathbf{t}, \left(\frac{\mathbf{Z}_*^T \mathbf{Z}_*}{\sigma_u^2} + \frac{I}{\sigma_\gamma^2}\right)^{-1}\right]$,

   where $\mathbf{Z}_* = \begin{pmatrix} |x_1 - \tau_1| & \cdots & |x_1 - \tau_k| \\ \vdots & \ddots & \vdots \\ |x_m - \tau_1| & \cdots & |x_m - \tau_k| \end{pmatrix}$, $\mathbf{t} = (t_1, \ldots, t_m)^T$, $t_i = \theta_i - x_i^T \boldsymbol{b}$;

(4) $\left[x_i \,\middle|\, \boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\gamma}, \mu_x, \sigma_x^2, \sigma_e^2, \sigma_u^2, \sigma_\gamma^2, \sigma_\eta^2, \boldsymbol{X}, \boldsymbol{y} \right]$

$$\stackrel{iid}{\sim} \exp\left\{ -\frac{1}{2\sigma_u^2} \left(\theta_i - \boldsymbol{x}_i^T \boldsymbol{b} - \boldsymbol{z}_i^T \boldsymbol{\gamma} \right)^2 \right\} N\left[\left(\sigma_\eta^{-2} n_i + \sigma_x^{-2} \right)^{-1} \left(\sigma_\eta^{-2} n_i \bar{X}_i + \sigma_x^{-2} \mu_x \right), \left(\sigma_\eta^{-2} n_i + \sigma_x^{-2} \right)^{-1} \right];$$

(5) $\left[\mu_x \,\middle|\, \boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \sigma_x^2, \sigma_e^2, \sigma_u^2, \sigma_\gamma^2, \sigma_\eta^2, \boldsymbol{X}, \boldsymbol{y} \right] \sim N\left(\bar{x}, \frac{\sigma_x^2}{m} \right);$

(6) $\left[\sigma_e^{-2} \,\middle|\, \boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \mu_x, \sigma_x^2, \sigma_u^2, \sigma_\gamma^2, \sigma_\eta^2, \boldsymbol{X}, \boldsymbol{y} \right] \sim G\left[\frac{n_t}{2} + a_e, \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left(y_{ij} - \theta_i \right)^2 + b_e \right];$

(7) $\left[\sigma_u^{-2} \,\middle|\, \boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \mu_x, \sigma_x^2, \sigma_e^2, \sigma_\gamma^2, \sigma_\eta^2, \boldsymbol{X}, \boldsymbol{y} \right] \sim G\left[\frac{m}{2} + a_u, \frac{1}{2} \sum_{i=1}^{m} \left(\theta_i - \boldsymbol{x}_i^T \boldsymbol{b} - \boldsymbol{z}_i^T \boldsymbol{\gamma} \right)^2 + b_u \right];$

(8) $\left[\sigma_\eta^{-2} \,\middle|\, \boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \mu_x, \sigma_x^2, \sigma_e^2, \sigma_\gamma^2, \sigma_u^2, \boldsymbol{X}, \boldsymbol{y} \right] \sim G\left[\frac{n_t}{2} + a_\eta, \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left(X_{ij} - x_i \right)^2 + b_\eta \right];$

(9) $\left[\sigma_\gamma^{-2} \,\middle|\, \boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \mu_x, \sigma_x^2, \sigma_e^2, \sigma_u^2, \sigma_\eta^2, \boldsymbol{X}, \boldsymbol{y} \right] \sim G\left[\frac{k}{2} + a_\gamma, \frac{1}{2} \boldsymbol{\gamma}^T \boldsymbol{\gamma} + b_\gamma \right];$

(10) $\left[\sigma_x^{-2} \,\middle|\, \boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \mu_x, \sigma_\gamma^2, \sigma_e^2, \sigma_u^2, \sigma_\eta^2, \boldsymbol{X}, \boldsymbol{y} \right] \sim G\left[\frac{m}{2} + a_x, \frac{1}{2} \sum_{i=1}^{m} (x_i - \mu_x)^2 + b_x \right].$

Our target of inference is to estimate $\{\theta_i, i = 1, \dots, m\}$. We run $L(\geq 2)$ chains and $2d$ iterations for each chain. To reduce the effect of the starting distributions, the first d iterations of each chain are discarded and posterior summaries are calculated based on remaining d iterates. The HB estimators for small area means is approximated by

$$E\left(\theta_i \,\middle|\, \boldsymbol{X}, \boldsymbol{y} \right) = E\left[E\left(\theta_i \,\middle|\, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \mu_x, \sigma_x^2, \sigma_e^2, \sigma_u^2, \sigma_\gamma^2, \sigma_\eta^2, \boldsymbol{X}, \boldsymbol{y} \right) \right] \tag{3.9}$$

$$\simeq \left(Ld^{-1} \right) \sum_{l=1}^{L} \sum_{r=d+1}^{2d} \left[\left(1 - C_i^{(lr)} \right) \bar{y}_i + C_i^{(lr)} \left(\boldsymbol{x}_i^{T(lr)} \boldsymbol{b}^{(lr)} + \boldsymbol{z}_i^{T(lr)} \boldsymbol{\gamma}^{(lr)} \right) \right],$$

and the posterior variance is estimated by

$$V\left(\theta_i \,\middle|\, \boldsymbol{X}, \boldsymbol{y} \right) = E\left[V\left(\theta_i \,\middle|\, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \mu_x, \sigma_x^2, \sigma_e^2, \sigma_u^2, \sigma_\gamma^2, \sigma_\eta^2, \boldsymbol{X}, \boldsymbol{y} \right) \right]$$

$$+ V\left[E\left(\theta_i \,\middle|\, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \mu_x, \sigma_x^2, \sigma_e^2, \sigma_u^2, \sigma_\gamma^2, \sigma_\eta^2, \boldsymbol{X}, \boldsymbol{y} \right) \right]$$

$$\simeq \left(Ld^{-1} \right) \sum_{l=1}^{L} \sum_{r=d+1}^{2d} \left(\frac{\sigma_e^{2(lr)}}{n_i} \left(1 - C_i^{(lr)} \right) \right)$$

$$+ \left(Ld^{-1} \right) \sum_{l=1}^{L} \sum_{r=d+1}^{2d} \left[\left(1 - C_i^{(lr)} \right) \bar{y}_i + C_i^{(lr)} \left(\boldsymbol{x}_i^{T(lr)} \boldsymbol{b}^{(lr)} + \boldsymbol{z}_i^{T(lr)} \boldsymbol{\gamma}^{(lr)} \right) \right]^2$$

$$- \left[E(\theta_i | \boldsymbol{X}, \boldsymbol{y}) \right]^2 .$$

3.4. Convergence and model adequacy

We implement MCMC chains and monitored the convergence of the MCMC following general guidelines based on $\sqrt{\hat{R}_i}$ (Gelman and Rubin, 1992) as the estimator of PSRF. This is given by

$$\sqrt{\hat{R}_i} = \sqrt{\frac{\widehat{var(\theta_i)}}{W_i}},$$

where $\widehat{var(\theta_i)} = (1-1/n)W_i + (1/n)B_i$, W_i and B_i are within chain variance and between chain variance, respectively. If $\sqrt{\hat{R}_i}$ is close to 1 for all θ_i, it implies that the MCMC chains are convergent.

We calculate the posterior predictive p-value (Meng, 1994) to check the model adequacy. The model is better supported by the data if the posterior predictive p-value (p) under the model is close to 0.5.

$$p = (Ld)^{-1} \sum_{l=1}^{L} \sum_{r=d+1}^{2d} I\left\{d\left(\mathbf{y}^{(lr)}, \boldsymbol{\theta}^{(lr)}\right) \geq d\left(\mathbf{y}_{obs}, \boldsymbol{\theta}^{(lr)}\right)\right\},$$

where $I(\cdot)$ is indicator function and $d(\mathbf{y}, \boldsymbol{\theta})$ is the posterior discrepancy measure.

4. Simulation Studies

4.1. Computational details

The key feature of our implementation is the use of a semiparametric regression model with a RBF instead of TPBF. To compare the performance between these two choices in the model, we consider two functions (a slowly-varying smooth function and a function with a sharp peak).

Example 1. The true function is

$$f(x) = 5 + 3x + 4x^2, \quad x \in [-2, 2].$$

Example 2. The true function is

$$f(x) = \sin(x) + 0.5\exp\left(-3x^2\right), \quad x \in [-2, 2].$$

To take simulated data, at first, we sequentially generate x_i ($i = 1, \ldots, 12$) on $[-2, 2]$ and then we generate X_{ij} from x_i with error $\eta_{ij} \sim N(0, 0.3^2)$. And θ_i are generated from x_i with random effect $u_i \sim N(0, 0.1^2)$ for each function. Lastly, y_{ij} are generated from θ_i with errors $e_{ij} \sim N(0, 2^2)$ and $e_{ij} \sim N(0, 0.3^2)$ for each examples. We run three independent chains with runs of length 5,000 following burn-ins of 2,500. We set all 1.0 for all hyperparameters $a_e, b_e, a_u, b_u, a_\eta, b_\eta, a_\eta, b_\eta, a_x$ and b_x. Notice that the results are not sensitive to other values for the hyperparameters. We then take the average of the squared differences of the estimators from the true mean (TM) over the 100 independent simulations and take their squared roots to obtain the root mean squared errors (RMSE) to compare the performance, where

$$\text{RMSE}_i = \sqrt{\sum_{k=1}^{100} \frac{\left(\theta_i^{(k)} - \hat{\theta}_i^{(k)}\right)^2}{100}}.$$

Table 1: Estimates of small area means for Example 1

i	n_i	TM	Mean with TPBF				Mean with RBF			
			$K = 1$	$K = 3$	$K = 5$	$K = 7$	$K = 1$	$K = 3$	$K = 5$	$K = 7$
1	5	15.015	14.441	14.903	14.890	14.884	14.443	14.909	14.910	14.908
2	10	10.847	10.829	10.804	10.778	10.784	10.831	10.805	10.780	10.782
3	5	7.656	8.183	7.744	7.818	7.822	8.184	7.744	7.806	7.815
4	7	5.587	6.006	5.570	5.668	5.652	6.006	5.569	5.661	5.645
5	8	4.563	4.727	4.742	4.594	4.606	4.726	4.739	4.590	4.603
6	8	4.580	4.231	4.599	4.635	4.632	4.229	4.597	4.633	4.629
7	7	5.658	5.385	5.635	5.692	5.693	5.384	5.634	5.692	5.692
8	9	7.854	8.046	8.059	7.998	8.003	8.045	8.058	7.999	8.002
9	8	11.040	11.292	11.074	11.145	11.121	11.292	11.077	11.145	11.122
10	7	15.264	15.382	15.234	15.213	15.238	15.383	15.235	15.215	15.237
11	7	20.674	20.602	20.604	20.591	20.590	20.599	20.605	20.595	20.591
12	5	26.987	26.648	26.845	26.849	26.848	26.651	26.844	26.853	26.855

TM = true mean; TPBF = truncated polynomial basis functions; RBF = radial basis functions.

Table 2: RMSE for Example 1

i	n_i	RMSE with TPBF				RMSE with RBF			
		$K = 1$	$K = 3$	$K = 5$	$K = 7$	$K = 1$	$K = 3$	$K = 5$	$K = 7$
1	5	1.022	0.881	0.920	0.942	1.023	0.871	0.910	0.918
2	10	0.548	0.560	0.580	0.579	0.548	0.560	0.580	0.580
3	5	0.828	0.708	0.678	0.700	0.828	0.708	0.684	0.702
4	7	0.731	0.593	0.602	0.590	0.730	0.592	0.600	0.590
5	8	0.504	0.454	0.445	0.444	0.505	0.453	0.444	0.444
6	8	0.612	0.441	0.446	0.448	0.612	0.442	0.445	0.449
7	7	0.701	0.574	0.561	0.559	0.702	0.576	0.561	0.559
8	9	0.656	0.617	0.609	0.621	0.656	0.620	0.609	0.621
9	8	0.704	0.659	0.665	0.662	0.704	0.657	0.665	0.662
10	7	0.754	0.767	0.771	0.762	0.754	0.771	0.772	0.766
11	7	0.774	0.784	0.793	0.802	0.773	0.784	0.795	0.804
12	5	0.887	0.853	0.863	0.859	0.885	0.853	0.858	0.856
Overall		8.721	7.891	7.933	7.968	8.720	7.887	7.923	7.951

RMSE = root mean square error; TPBF = truncated polynomial basis functions; RBF = radial basis functions.

4.2. Results

In all cases, $\hat{R}_i \simeq 1$ for all θ_i and $p \simeq 0.5$ for all simulation data and models. The detailed results are reported in Tables 1–4. We report the sample sizes, TM and small area means estimates with two basis functions as well as RMSE for each strata in Examples 1 and 2. In Example 1, we can see that the model with RBF provides minimal improvement for all number of knots based on the overall RMSE; however, the difference is only a third decimal point. The performance of the model with RBF is also very similar with TPBF in Example 2. Additionally, the case of three knots has the smallest RMSE for both basis functions. Figure 1 shows the one result ($K = 3$ case) for Examples 1 and 2 that the real line (——), dashed line (- - - -) , and dotted line ($\cdots\cdots$) are the true function and the estimates with TPBS and RBF, respectively.

5. Application

We conducted the analysis based on a real data. We used LANDSAT (Land observatory satellites) data by Battese *et al.* (1988), hereafter BHF, for analysis. BHF was made available by Tobias Schoch with the R package "rsae". This data on the areas under corn and soybeans (reported in hectares) in the 37 segments of the 12 counties (north-central Iowa) were determined by USDA Statistical Reporting

Table 3: Estimates of small area means for Example 2

i	n_i	TM	Mean with TPBF				Mean with RBF			
			$K = 1$	$K = 3$	$K = 5$	$K = 7$	$K = 1$	$K = 3$	$K = 5$	$K = 7$
1	5	−0.895	−0.896	−0.883	−0.882	−0.881	−0.899	−0.877	−0.876	−0.874
2	10	−0.989	−1.000	−0.998	−0.999	−0.999	−1.001	−0.999	−1.000	−1.000
3	5	−0.941	−0.903	−0.911	−0.910	−0.912	−0.904	−0.918	−0.916	−0.918
4	7	−0.785	−0.766	−0.775	−0.774	−0.776	−0.766	−0.781	−0.780	−0.782
5	8	−0.520	−0.497	−0.502	−0.504	−0.504	−0.496	−0.504	−0.507	−0.507
6	8	0.568	0.550	0.549	0.548	0.548	0.552	0.552	0.551	0.552
7	7	0.924	0.890	0.892	0.893	0.894	0.893	0.898	0.899	0.900
8	9	0.517	0.541	0.545	0.546	0.547	0.542	0.548	0.549	0.549
9	8	0.787	0.790	0.796	0.796	0.797	0.791	0.797	0.797	0.797
10	7	0.957	0.940	0.943	0.943	0.943	0.939	0.943	0.943	0.943
11	7	0.993	0.983	0.982	0.982	0.982	0.982	0.981	0.981	0.981
12	5	0.897	0.931	0.924	0.922	0.921	0.928	0.920	0.918	0.918

TM = true mean; TPBF = truncated polynomial basis functions; RBF = radial basis functions.

Table 4: RMSE for Example 2

i	n_i	RMSE with TPBF				RMSE with RBF			
		$K = 1$	$K = 3$	$K = 5$	$K = 7$	$K = 1$	$K = 3$	$K = 5$	$K = 7$
1	5	0.138	0.139	0.140	0.141	0.138	0.140	0.141	0.142
2	10	0.095	0.095	0.095	0.095	0.096	0.095	0.096	0.096
3	5	0.127	0.125	0.125	0.125	0.127	0.124	0.125	0.124
4	7	0.111	0.110	0.111	0.111	0.111	0.110	0.110	0.110
5	8	0.089	0.088	0.088	0.088	0.089	0.087	0.087	0.088
6	8	0.091	0.092	0.092	0.092	0.091	0.092	0.092	0.092
7	7	0.111	0.111	0.111	0.11	0.111	0.109	0.109	0.109
8	9	0.104	0.105	0.106	0.107	0.105	0.107	0.107	0.108
9	8	0.108	0.109	0.109	0.109	0.108	0.109	0.109	0.110
10	7	0.122	0.122	0.123	0.123	0.123	0.123	0.124	0.123
11	7	0.123	0.123	0.124	0.124	0.123	0.124	0.124	0.125
12	5	0.131	0.131	0.131	0.131	0.130	0.130	0.131	0.131
Overall		1.350	1.350	1.355	1.356	1.352	1.350	1.355	1.358

RMSE = root mean square error; TPBF = truncated polynomial basis functions; RBF = radial basis functions.
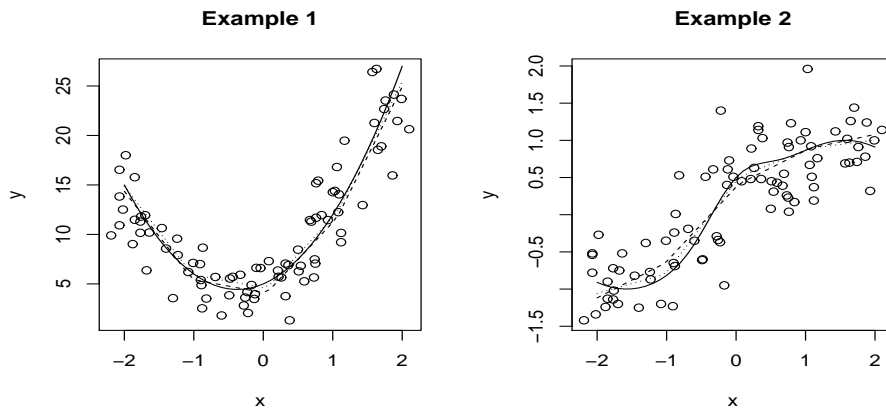

Figure 1: *For K = 3, True :* ——*, TPBF : - - - -, RBF :* ······ *(TPBF = truncated polynomial basis functions; RBF = radial basis functions).*
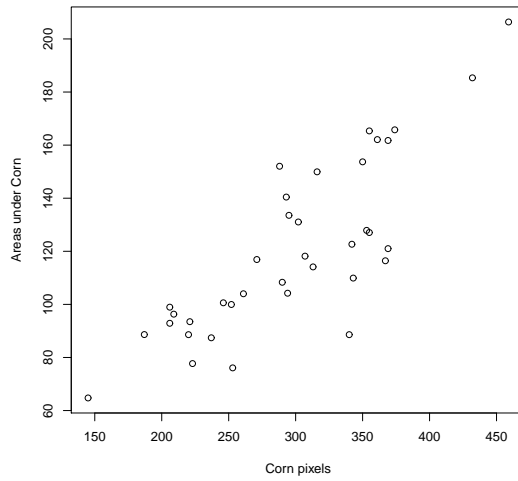
Figure 2: *Scatter plot for areas under corn and corn pixels.*

Table 5: Comparative data analyses for the LANDSAT data

Counties	n_i	TPBF						RBF					
		$K = 1$		$K = 3$		$K = 5$		$K = 1$		$K = 3$		$K = 5$	
		Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.
Cerro Gordo	1	151.039	9.779	142.155	14.020	142.738	14.477	159.396	9.188	155.336	10.114	149.823	12.915
Hamilton	1	98.541	10.431	79.712	36.185	93.086	16.925	104.270	12.058	95.364	18.822	90.516	44.626
Worth	1	107.205	6.457	124.905	19.296	111.662	12.022	107.216	5.570	106.394	11.703	105.727	11.598
Humboldt	2	162.104	13.115	142.849	19.889	141.193	19.516	174.935	12.647	166.024	16.310	157.566	18.405
Franklin	3	135.980	6.150	136.407	11.033	139.456	10.942	138.206	5.754	139.369	7.814	140.704	10.729
Pocahontas	3	101.258	8.810	95.917	20.708	106.723	12.125	105.209	9.663	98.983	12.783	105.379	14.144
Winnebago	3	115.003	6.852	113.916	10.999	115.501	9.395	110.005	7.006	112.438	7.386	110.107	7.626
Wright	3	141.251	7.130	140.075	13.581	141.530	11.196	145.761	6.684	145.891	8.795	143.483	10.436
Webster	4	109.644	5.966	121.833	13.821	115.458	11.474	108.218	4.954	108.286	8.621	106.516	12.537
Hancock	5	113.533	6.372	115.107	8.265	114.080	7.645	109.478	6.068	111.308	6.278	109.121	6.961
Kossuth	5	123.925	5.763	120.675	8.176	121.354	9.577	121.904	5.806	123.941	6.233	123.089	6.105
Hardin	6	113.980	6.398	115.231	8.915	115.162	8.506	109.705	6.240	111.664	6.398	109.471	6.935
p-value		0.440		0.435		0.433		0.432		0.435		0.434	

LANDSAT = land observatory satellites; TPBF = truncated polynomial basis functions; RBF = radial basis functions.

Service staff consists of 37 observations and 10 variables. This paper considers a prediction of areas under corn that uses soya corn pixels only as a measurement error covariate. Figure 2 is the scatter plot for areas under corn and corn pixels; consequently, we can view it is a little non-linear pattern.

We use same settings such as simulation studies. In our all case, $\sqrt{\hat{R}} \simeq 1$ for all θ_i and we report the sample sizes, estimates, standard errors (s.e.) and the posterior predictive *p*-value for each case in Table 5. The models with five knots TPBF and one knot RBF are better based on *p*-value and these two models have similar model adequacy.

6. Concluding Remarks

We developed a semiparametric small area regression model with radial basis functions instead of truncated polynomial basis functions under a structural measurement error model with knots on a grid of equally spaced sample quantiles of the covariate. Numerical studies show that estimates of small

area means seem similar to estimates based on truncated polynomial basis functions.

We can extend our model in various way. First, we considered normal outcome and covariate with the measurement error in this paper. We develop generalized versions for the measurement error problem in binomial or Poisson outcome cases. Second, we only considered measurement error in covariate without measurement error in outcome variable. We can consider measurement error in both covariate and outcome variables. Third, the results depend on the number and location of knots. The selection of knots is always a subjective and tricky issue in these kinds of problems. The complete underlying pattern may not be captured properly and result in a biased fit if too few or too many knots are used; therefore, we will handle semiparametric Bayesian estimation with free knots in future research.

References

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

Bhadra, D., Ghosh, M. and Kim, D. (2013). Estimation of median household income for small areas: A Bayesian semiparametric approach, *Calcutta Statistical Association Bulletin*, **64**, 115–142.

Fuller, W. A. (1987). *Measurement Error Models*, John Wiley & Sons, New York.

Ghosh, M., Sinha, K. and Kim, D. (2006). Empirical and hierarchical Bayesian estimation in finite population sampling under structural measurement error models, *Scandinavian Journal of Statistics*, **33**, 591–608.

Gelman, A. E. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, **7**, 457–472.

Hwang, J. S. and Kim, D. H. (2010). Semiparametric Bayesian estimation under structural measurement error model, *Communications of the Korean Statistical Society*, **17**, 551–560.

Meng, X. L. (1994). Posterior predictive p-values. *Annals of Statistics*, **22**, 1142–1160.

Opsomer, J. D., Claeskens, G., Ranalli, M. G., Kauermann, G. and Breidt, F. J. (2008). Nonparametric small area estimation using penalized spline regression, *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, **70**, 265–286.

Rao, J. N. K. (2003). *Small Area Estimation*, John Wiley & Sons, New York.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression*, Cambridge University Press, Cambridge.