Communications of the Korean Statistical Society
2010, Vol. 17, No. 4, 551–560

Semiparametric Bayesian Estimation under Structural Measurement Error Model

Jin Seub Hwang[a], Dal Ho Kim[1,a]

[a]Department of Statistics, Kyungpook National University

Abstract

This paper considers a Bayesian approach to modeling a flexible regression function under structural measurement error model. The regression function is modeled based on semiparametric regression with penalized splines. Model fitting and parameter estimation are carried out in a hierarchical Bayesian framework using Markov chain Monte Carlo methodology. Their performances are compared with those of the estimators under structural measurement error model without a semiparametric component.

Keywords: Gibbs sampling, hierarchical Bayes, Metropolis-Hastings, penalized spline, semiparametric, structural measurement error.

1. Introduction

Sample survey methodologies are widely used for collecting relevant information about a population of interest. Apart from providing population level estimates, surveys are also designed to estimate various features of subpopulations or domains. Domains may be geographic areas like states or provinces or county school districts or can even be identified by a particular social-demographic characteristic like a specific age-gender group. Sometimes, the domain-specific sample size may be too small to yield direct estimates of adequate precision. This led to the development of small area estimation procedures which specifically deal with the estimation of various features of small domains. Rao (2003) gives a comprehensive account of model-based methods that lead to efficient estimators of small area means when the area-specific sample sizes are small.

Ghosh and Meeden (1986) considered empirical Bayesian(EB) estimation in a stratified finite population context using a simple one-way ANOVA model. The results can be extended by inclusion of covariates, and such procedures have been discussed in Ghosh and Meeden (1996). Often, however, it is not possible to obtain exact measurements of covariates. Ghosh *et al.* (2006), abbreviated GSK, assumed that the covariates are measured with error and stochastic. This is the so-called structural measurement error model.

Semiparametric regression methods have not been used in small area estimation contexts until recently. This was mainly due to methodological difficulties in combining the different smoothing techniques with the estimation tools generally used in small area estimation. The pioneering contribution in this regard is the work by Opsomer *et al.* (2008) in which they combined small area random effects with a smooth.

The objective of this article is to develop efficient estimators of small area means by using flexible smoothing of non-linear pattern with structural measurement error model. In doing so, we have modeled the small area means using penalized spline (Eilers and Marx, 1996) which is a commonly used

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

but powerful function estimation tool in non-parametric inference. We have used truncated polynomial basis functions with varying degrees and number of knots, although other types of basis functions like B-splines or thinplate splines can also be used. For our semiparametric model, the analysis has been carried out using a hierarchical Bayesian(HB) approach. Since we chose non-informative improper priors for the regression parameters and one hyper parameter, propriety of the posterior has been proved before proceeding with the computations. Markov chain Monte Carlo(MCMC) methodologies, specifically Gibbs sampling and Metropolis- Hastings(M-H) algorithm, has been used to obtain the parameter estimates. We develop HB procedures for semiparametric small area estimation under structural measurement error model. Following the general convention, we have placed the knots on a grid of equally spaced sample quantiles of the independent variable.

The remaining sections are arranged as follows. The model specification are given in Section 2. In Section 3, we have established the propriety of the posteriors, and have discussed the MCMC implementation of the proposed hierarchical Bayes procedure. We illustrate the performance of the approach for a simulation study in Section 4. Finally, we present a discussion of the results in Section 5. The proofs of certain technical results are deferred to the Appendix.

2. Model Specification

Suppose there are m strata labelled $1, \ldots, m$ and let N_i denote the known population size for the ith stratum. We denote by y_{ij} the response and by X_{ij} the covariate of the j^{th} unit in the i^{th} stratum ($j = 1, \ldots, N_i$; $i = 1, \ldots, m$). A sample of size n_i is drawn from the ith stratum ($\sum_{i=1}^{m} n_i = n_t$).

The basic semiparametric model can be expressed as

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

where $f(x_i)$ is an unspecified function of x_i reflecting the unknown response-covariate relationship. We approximate $f(x_i)$ using a P-spline and rewrite (2.1) as

$$y_{ij} = b_0 + b_1 x_i + \cdots + b_p x_i^p + \sum_{a=1}^{k} \gamma_a (x_i - \tau_a)_+{}^p + u_i + e_{ij}, \tag{2.2}$$

where, for any number u, u_+ is equal to u if u is positive and is equal to 0 otherwise. The above spline model with degree p can adequately approximate any unspecified smooth function. Typically, linear($p = 1$) or quadratic($p = 2$) splines serves most practical purposes since they ensure adequate smoothness in the fitted curve.

For illustration, we consider $p = 1$ case with structural measurement errors. In this case, the superpopulation model is assumed as follows:

$$\begin{aligned} y_{ij} &= b_0 + b_1 x_i + \sum_{a=1}^{k} \gamma_a (x_i - \tau_a)_+ + u_i + e_{ij} \\ &= \boldsymbol{x}_i^T \boldsymbol{b} + \boldsymbol{z}_i^T \boldsymbol{\gamma} + u_i + e_{ij} \\ &= \theta_i + e_{ij} \quad (j = 1, \ldots, N_i;\ i = 1, \ldots, m); \\ X_{ij} &= x_i + \eta_{ij}, \end{aligned} \tag{2.3}$$

where $\theta_i = \boldsymbol{x}_i^T \boldsymbol{b} + \boldsymbol{z}_i^T \boldsymbol{\gamma} + u_i$ is our target of inference. Here $\boldsymbol{x}_i = (1, x_i)^T$ and $\boldsymbol{z}_i = \{(x_i - \tau_1)_+, \ldots, (x_i - \tau_k)_+\}^T$. And $\boldsymbol{b} = (b_0, b_1)^T$ is the vector of regression coefficients, while $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)^T$ is the

vector of spline coefficients. Also, k is the number of knots and $\tau = (\tau_1, \ldots, \tau_k)$ is the vector of knot location($\tau_1 < \cdots < \tau_k$). It is assumed that the x_i, u_i, e_{ij} and η_{ij} are mutually independent with $x_i \sim N(\mu_x, \sigma_x^2)$, $u_i \sim N(0, \sigma_u^2)$, $e_{ij} \sim N(0, \sigma_e^2)$ and $\eta_{ij} \sim N(0, \sigma_\eta^2)$. The available data consist of (y_{ij}, X_{ij}).

3. Bayesian Inference

3.1. Hierarchical Bayesian framework

We consider a HB framework to predict the small area means $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$. Using expression (2.3), we begin with the following HB model:

Stage 1. $y_{ij} = \theta_i + e_{ij}$ $(j = 1, \ldots, n_i;\ i = 1, \ldots, m)$ where e_{ij} are *i.i.d.* $N(0, \sigma_e^2)$.

Stage 2. $\theta_i = \boldsymbol{x}_i^T \boldsymbol{b} + \boldsymbol{z}_i^T \boldsymbol{\gamma} + u_i$ $(i = 1, \ldots, m)$ where u_i are *i.i.d.* $N(0, \sigma_u^2)$.
$X_{ij} = x_i + \eta_{ij}$ $(j = 1, \ldots, n_i;\ i = 1, \ldots, m)$ where η_{ij} are *i.i.d.* $N(0, \sigma_\eta^2)$.

Stage 3. $x_i \sim N(\mu_x, \sigma_x^2)$.

Stage 4. $\boldsymbol{\gamma} \sim N(0, \sigma_\gamma^2 I)$.

Stage 5. b_0, b_1, μ_x, σ_e^2, σ_u^2, σ_x^2, σ_η^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 an gamma distribution shape parameter α and rate parameter β having the expression $f(x) \propto x^{\alpha-1} \exp(-\beta x)$.

First check the propriety of the posterior under the given prior. By the conditional independence properties, we can factorize the full posterior as

$$\begin{aligned} &\left[\boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \mu_x, \sigma_x^2, \sigma_e^2, \sigma_u^2, \sigma_\eta^2, \sigma_\gamma^2 | \mathbf{X}, \mathbf{y}\right] \\ &\propto \left[\mathbf{y}|\boldsymbol{\theta}, \sigma_e^2\right]\left[\boldsymbol{\theta}|\boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \sigma_u^2, \mathbf{X}\right]\left[\mathbf{X}|\boldsymbol{x}, \sigma_\eta^2\right]\left[\boldsymbol{x}|\mu_x, \sigma_x^2\right]\left[\boldsymbol{\gamma}|\sigma_\gamma^2\right] \times [\boldsymbol{b}]\,[\mu_x]\left[\sigma_x^2\right]\left[\sigma_e^2\right]\left[\sigma_u^2\right]\left[\sigma_\eta^2\right]\left[\sigma_\gamma^2\right]. \end{aligned} \tag{3.1}$$

The proof of the propriety of the posterior is deferred to the appendix.

Our target of inference is $\{\theta_i,\ i = 1, \ldots, m\}$. Since the marginal posterior distribution of θ_i is analytically intractable, high dimensional integration needs to be carried out in a theoretical framework. However, this task can be easily accomplished in an MCMC framework by using Gibbs sampler and M-H algorithm to sample from the full conditionals of θ_i and other relevant parameters. In implementing the Gibbs sampler and M-H algorithm, we follow the recommendation of Gelman and Rubin (1992) and run $L(\geq 2)$ chains. For each chain, we run $2d$ iterations with starting points drawn from an overdispersed distribution. To diminish the effects of the starting distributions, the first d iterates of each chain are discarded and posterior summaries are calculated based on the rest of the d iterates.

3.2. Full conditional distribution and inference

The Gibbs sampling and M-H algorithm analysis are based on the following full conditional distribution:

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

where $C_i = \sigma_e^2 / \left(\sigma_e^2 + n_i \sigma_u^2\right)$.

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

where $\mathbf{X}_* = \left(\boldsymbol{x}_1^T,\ldots,\boldsymbol{x}_m^T\right)^T$, $\mathbf{w} = (w_1,\ldots,w_m)^T$, $w_i = \theta_i - \mathbf{z}_i^T\boldsymbol{\gamma}$.

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

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

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

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

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

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

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

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

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

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

We may use the M-H algorithm in (iv). We generate several sets of these samples. After burning out the first half, we use the averaging principle and take the average of the HB estimates over all the remaining sets to obtain the final HB estimate. The HB estimators for small area means is approximated as:

$$\begin{aligned} E\left(\theta_i|\mathbf{X},\mathbf{y}\right) &= E\left[E\left(\theta_i|\boldsymbol{b},\boldsymbol{\gamma},\boldsymbol{x},\mu_x,\sigma_x^2,\sigma_e^2,\sigma_u^2,\sigma_\gamma^2,\sigma_\eta^2,\mathbf{X},\mathbf{y}\right)\right] \\ &\simeq \left(Ld^{-1}\right)\sum_{l=1}^{L}\sum_{r=d+1}^{2d}\left[\left(1-C_i^{(lr)}\right)\bar{y}_i + C_i^{(lr)}\left(\boldsymbol{x}_i^{T(lr)}\boldsymbol{b}^{(lr)} + \mathbf{z}_i^{T(lr)}\boldsymbol{\gamma}^{(lr)}\right)\right] \end{aligned} \tag{3.2}$$

and the posterior variance is estimated as:

$$
\begin{aligned}
V(\theta_i|\mathbf{X},\mathbf{y}) &= E\left[V\left(\theta_i|\boldsymbol{b},\boldsymbol{\gamma},\boldsymbol{x},\mu_x,\sigma_x^2,\sigma_e^2,\sigma_u^2,\sigma_\gamma^2,\sigma_\eta^2,\mathbf{X},\mathbf{y}\right)\right] \\
&\quad + V\left[E\left(\theta_i|\boldsymbol{b},\boldsymbol{\gamma},\boldsymbol{x},\mu_x,\sigma_x^2,\sigma_e^2,\sigma_u^2,\sigma_\gamma^2,\sigma_\eta^2,\mathbf{X},\mathbf{y}\right)\right] \\
&\simeq \left(Ld^{-1}\right)\sum_{l=1}^{L}\sum_{r=d+1}^{2d}\left(\frac{\sigma_e^{2(lr)}}{n_i}\left(1-C_i^{(lr)}\right)\right) \\
&\quad + \left(Ld^{-1}\right)\sum_{l=1}^{L}\sum_{r=d+1}^{2d}\left[1-C_i^{(lr)}\bar{y}_i + C_i^{(lr)}\left(\boldsymbol{x}_i^{T(lr)}\boldsymbol{b}^{(lr)} + \mathbf{z}_i^{T(lr)}\boldsymbol{\gamma}^{(lr)}\right)\right]^2 - \left[E(\theta_i|\mathbf{X},\mathbf{y})\right]^2 .
\end{aligned}
\tag{3.3}
$$

3.3. Convergence diagnostics and model adequacy

Gelman and Rubin (1992) suggested simulating multiple chains. They constructed a potential scale reduction factor(PSRF), interpreted as the factor by which the estimated variance of θ_i could be reduced if the chain were simulated for more iterations. The estimator of PSRF $\sqrt{\hat{R}_i}$ is calculated as follows.

$$
\sqrt{\hat{R}_i} = \sqrt{\frac{\widehat{\mathrm{var}}(\theta_i)}{W_i}}, \quad (i = 1,\ldots,m), \tag{3.4}
$$

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

The posterior predictive p-value is useful in model adequacy checking. The principle of posterior predictive checking is that the realized results should look plausible under a posterior predictive distribution. The posterior predictive p-value is computed by the expression

$$
p = (Ld)^{-1}\sum_{l=1}^{L}\sum_{r=d+1}^{2d} I\left\{d\left(\mathbf{y}^{(lr)},\,\boldsymbol{\theta}^{(lr)}\right) \geq d\left(\mathbf{y}_{obs},\,\boldsymbol{\theta}^{(lr)}\right)\right\}, \tag{3.5}
$$

where $I(\cdot)$ is indicate function and $d(\mathbf{y},\boldsymbol{\theta})$ is the posterior discrepancy measure. If the p-value is an extreme value(close to 0 or 1), we can conclude that the model does not fit well. On the other hand, if the model fit the data well, the p-value is close to 0.5.

4. Numerical Studies

4.1. Simulation studies

We consider two functions as follows

(a) $y = 5 + 3x + 4x^2$, $\qquad x \in [-2,2]$.

(b) $y = 5 + \cos(2x) + 2\exp(-16x^2)$, $\qquad x \in [-2,2]$.

We conduct a simulation study to compare the performance of our model estimators in comparison with the GSK estimators. Simulated data are created as follows: The $x_i\,(i = 1,\ldots,12)$ are generated sequentially on [−2,2] and then X_{ij} are generated from x_i with errors η_{ij} independent $N(0,0.1^2)$ in (a) and $N(0,0.2^2)$ in (b). For each functions (a) and (b), θ_i are generated from x_i with random effect u_i

Table 1: Means of example (a)

i	n_i	TM	GS	mean(fixed)				mean(random)		
				$k=3$	$k=4$	$k=5$	$k=6$	$\lambda=1$	$\lambda=3$	$\lambda=9$
1	5	15.004	14.621	14.609	14.655	14.632	14.663	14.377	14.431	14.520
2	10	10.838	10.815	10.968	10.925	10.903	10.899	10.890	10.894	10.907
3	5	7.633	7.784	7.850	7.771	7.918	7.941	7.963	7.942	7.913
4	7	5.592	5.780	5.632	5.865	5.830	5.737	5.875	5.855	5.816
5	8	4.590	4.710	4.805	4.610	4.600	4.683	4.726	4.718	4.685
6	8	4.578	4.715	4.586	4.622	4.623	4.589	4.666	4.660	4.639
7	7	5.672	5.884	5.670	5.773	5.733	5.723	5.779	5.770	5.752
8	9	7.873	8.035	8.004	7.883	7.894	7.939	7.970	7.963	7.953
9	8	11.013	10.913	10.808	10.966	10.982	10.917	10.933	10.935	10.943
10	7	15.265	15.182	15.290	15.202	15.213	15.256	15.204	15.210	15.228
11	7	20.668	20.493	20.682	20.649	20.637	20.627	20.540	20.547	20.569
12	5	27.002	26.809	26.811	26.857	26.849	26.860	26.813	26.827	26.836

Table 2: RMSEs of example (a)

i	n_i	RMSE(GS)	RMSE(fixed)				RMSE(random)		
			$k=3$	$k=4$	$k=5$	$k=6$	$\lambda=1$	$\lambda=3$	$\lambda=9$
1	5	1.108	0.952	0.986	1.013	1.025	1.105	1.075	1.027
2	10	0.605	0.514	0.510	0.517	0.533	0.524	0.520	0.518
3	5	0.954	0.743	0.763	0.734	0.752	0.806	0.777	0.759
4	7	0.706	0.568	0.604	0.612	0.596	0.637	0.618	0.605
5	8	0.699	0.562	0.564	0.553	0.555	0.591	0.573	0.561
6	8	0.727	0.584	0.548	0.556	0.567	0.598	0.579	0.563
7	7	0.780	0.559	0.566	0.558	0.570	0.602	0.586	0.565
8	9	0.686	0.535	0.547	0.551	0.543	0.566	0.554	0.545
9	8	0.642	0.569	0.502	0.505	0.531	0.534	0.527	0.516
10	7	0.772	0.633	0.662	0.653	0.635	0.658	0.645	0.640
11	7	0.770	0.610	0.619	0.628	0.636	0.673	0.664	0.653
12	5	0.836	0.761	0.754	0.762	0.757	0.777	0.773	0.768
overall		9.284	7.589	7.626	7.641	7.699	8.071	7.893	7.720

independent $N(0, 0.1^2)$. Lastly, y_{ij} are generated from θ_i with errors e_{ij} independent $N(0, 2^2)$ in (a) and $N(0, 0.3^2)$ in (b). We draw 50 independent samples (y_{ij}, X_{ij}). We set all 1.0 for all hyperparameters a_e, b_e, a_u, b_u, a_η, b_η, a_γ, b_γ, a_x and b_x. Notice that the results are not sensitive to other values for those hyperparameters.

4.2. Computational details and results

We implement and monitor the convergence of the MCMC following the general guidelines given in Gelman and Rubin (1992). We run ten independent chains each with a sample size of 5,000 and a burn-in sample of another 2,500. Using the equation (3.2), we estimate the small area means. In our all case, $\hat{R} \simeq 1$ for all θ_i and $p \simeq 0.5$. Moreover, we take the average of the squared differences of the estimators from the true means(TM) over the 50 simulations and take their squared roots to obtain the root mean squared errors(RMSE).

$$\text{RMSE}_i = \sqrt{\frac{1}{50}\sum_{k=1}^{50}\left(\theta_i^{(k)} - \hat{\theta}_i^{(k)}\right)^2}. \tag{4.1}$$

In Table $1 \sim 4$, we report the sample sizes, the TM, GSK, proposal model estimates as well as RMSE for the 12 strata in examples (a) and (b). It follows from the above tables that our estimates are doing

Table 3: Means of example (b)

i	n_i	TM	GS	mean(fixed)				mean(random)		
				$k=3$	$k=4$	$k=5$	$k=6$	$\lambda=1$	$\lambda=3$	$\lambda=9$
1	5	4.350	4.355	4.328	4.326	4.326	4.327	4.329	4.328	4.326
2	10	4.010	4.029	4.024	4.023	4.024	4.023	4.025	4.025	4.025
3	5	4.167	4.208	4.210	4.212	4.212	4.212	4.214	4.214	4.214
4	7	4.763	4.783	4.794	4.792	4.794	4.792	4.793	4.793	4.793
5	8	5.500	5.494	5.510	5.509	5.508	5.509	5.505	5.505	5.507
6	8	7.115	7.086	7.085	7.086	7.085	7.087	7.085	7.084	7.085
7	7	7.130	7.105	7.101	7.105	7.106	7.103	7.103	7.103	7.103
8	9	5.483	5.486	5.499	5.500	5.499	5.500	5.496	5.496	5.497
9	8	4.724	4.702	4.711	4.711	4.711	4.711	4.711	4.711	4.711
10	7	4.179	4.179	4.181	4.181	4.180	4.181	4.183	4.183	4.183
11	7	3.999	4.004	4.000	3.997	3.999	3.998	3.999	3.999	3.998
12	5	4.349	4.383	4.354	4.353	4.353	4.353	4.358	4.357	4.355

Table 4: RMSEs of example (b)

i	n_i	RMSE(GS)	RMSE(fixed)				RMSE(random)		
			$k=3$	$k=4$	$k=5$	$k=6$	$\lambda=1$	$\lambda=3$	$\lambda=9$
1	5	0.155	0.157	0.157	0.157	0.157	0.156	0.156	0.156
2	10	0.094	0.092	0.092	0.092	0.092	0.092	0.092	0.092
3	5	0.149	0.148	0.148	0.148	0.149	0.150	0.150	0.149
4	7	0.105	0.106	0.106	0.107	0.106	0.106	0.106	0.106
5	8	0.104	0.104	0.104	0.103	0.103	0.103	0.103	0.103
6	8	0.112	0.112	0.112	0.111	0.111	0.112	0.112	0.111
7	7	0.116	0.116	0.116	0.115	0.116	0.116	0.116	0.116
8	9	0.100	0.101	0.101	0.101	0.101	0.100	0.101	0.101
9	8	0.098	0.096	0.096	0.095	0.096	0.096	0.096	0.096
10	7	0.116	0.115	0.115	0.115	0.115	0.115	0.115	0.115
11	7	0.113	0.112	0.112	0.111	0.111	0.112	0.112	0.112
12	5	0.126	0.121	0.121	0.121	0.121	0.122	0.121	0.121
overall		11.389	1.380	1.380	1.376	1.379	1.380	1.380	1.379

better than the GSK estimates according to the RMSE criterion. Thus, based on our simulation, it appears that the our method has better overall performance than the GSK estimates.

In Figure 1 and 2, we display the one case with $k = 5$. The real line(——), dashed line(- - - -) and dotted line(· · · · ·) are the true functions, GSK estimates and proposal model estimates, respectively. It can be seen from the above figures that the proposal model estimates are slightly superior to the GSK estimates in view of the closeness to the true curve.

5. Discussion

In this paper, we have proposed HB procedures for semiparametric small area estimation under structure measurement error model with fixed knots and our estimates proved to be superior to the GSK estimates. The main advantage of our modeling procedure is that it can be used for any possible patterns in the response observations of small areas. Our model can be extended in various ways. (1) We have used the truncated polynomial basis function but other types of bases like B-splines radial or basis functions could also be used. (2) Although we have used a parametric normal distributional assumption for the random state, a broader class of distributions like the Dirichlet process or Polya trees may be tested. Lastly, we have noted that the result depends on the number and location of knots. The selection of knots is always a subjective but tricky issue in these kind of problems. Sometimes experience on the subject matter may be a guiding force in placing the knots in the “optimum” loca-

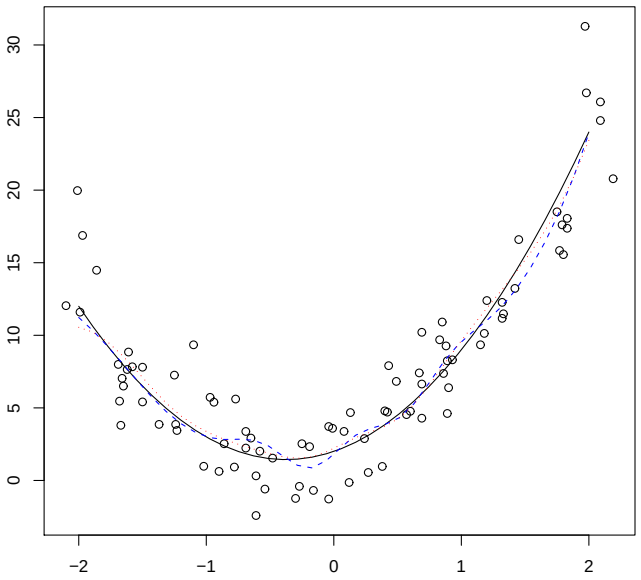

Figure 1: *Example (a), true curve:* ——-, *GSK:* - - - -, *Proposal:* · · · · · ·

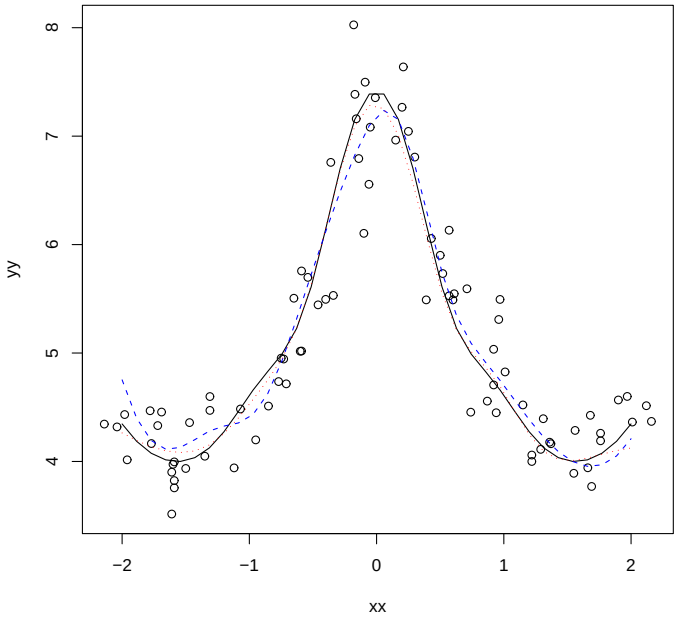

Figure 2: *Example (b), true curve:* ——-, *GSK:* - - - -, *Proposal:*· · · · · ·

tions where a sharp change in the curve pattern can be expected. Too few or too many knots are used the complete underlying pattern may not be captured properly, thus resulting in a biased fit. So, we will pursue the semiparametric Bayesian estimation with random knots in future research.

Appendix: Proof of Posterior Propriety

The basic 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

$$\begin{aligned} I &= \int \cdots \int p\left(\boldsymbol{\Omega}\middle|\mathbf{y}, \mathbf{X}\right) d\boldsymbol{\Omega} \\ &= \int \cdots \int \left[\mathbf{Y}\middle|\boldsymbol{\theta}, \sigma_e^2\right]\left[\boldsymbol{\theta}\middle|\boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \sigma_u^2, \mathbf{X}\right]\left[\mathbf{X}\middle|\boldsymbol{x}, \sigma_\eta^2\right]\left[\boldsymbol{x}\middle|\mu_x, \sigma_x^2\right] \\ &\quad \times \left[\boldsymbol{\gamma}\middle|\sigma_\gamma^2\right][\boldsymbol{b}][\mu_x]\left[\sigma_e^2\right]\left[\sigma_u^2\right]\left[\sigma_\eta^2\right]\left[\sigma_\gamma^2\right]\left[\sigma_x^2\right] d\boldsymbol{\Omega}. \end{aligned} \tag{A.1}$$

We have to show that $I \leq M$ where M is any finite positive constant.

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

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

where K_1 is a constant. Now integrating with respect to $\boldsymbol{b}$ and using $\mathbf{w}^T(I - P_{\mathbf{X}_*})\mathbf{w} \geq 0$,

$$\begin{aligned} I_{\boldsymbol{b}} &= \int \left[\boldsymbol{\theta}\middle|\boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{x}, \sigma_u^2, \mathbf{X}\right][\boldsymbol{b}] d\boldsymbol{b} \\ &= \left(\sigma_u^2\right)^{-\frac{m}{2}} \int \exp\left[-\frac{1}{2\sigma_u^2}\sum_{i=1}^m \left(\theta_i - \boldsymbol{x}_i^T\boldsymbol{b} - \mathbf{z}_i^T\boldsymbol{\gamma}\right)^2\right] d\boldsymbol{b} \\ &= \left(\sigma_u^2\right)^{-\frac{m}{2}} \int \exp\left[-\frac{1}{2\sigma_u^2}\sum_{i=1}^m \left(w_i - \boldsymbol{x}_i^T\boldsymbol{b}\right)^2\right] d\boldsymbol{b} \\ &= \left(\sigma_u^2\right)^{-\frac{m}{2}} \int \exp\left\{-\frac{1}{2\sigma_u^2}\mathbf{w}^T\left(I - P_{\mathbf{X}_*}\right)\mathbf{w}\right\} d\boldsymbol{b} \left(\sigma_u^2\right)^{\frac{2}{p}} |\mathbf{X}_*^T\mathbf{X}_*|^{-\frac{1}{2}} (2\pi)^{\frac{m}{2}} \\ &\leq K_2 \cdot \left(\sigma_u^2\right)^{-\frac{(m-p)}{2}} \cdot |\mathbf{X}_*^T\mathbf{X}_*|^{-\frac{1}{2}}, \end{aligned} \tag{A.3}$$

where $P_{\mathbf{X}_*} = \mathbf{X}_*(\mathbf{X}_*^T\mathbf{X}_*)^{-1}\mathbf{X}_*^T$, $\operatorname{rank}(\mathbf{X}_*) = p$ and K_2 is a constant.

Next, we consider integration with respect to $\boldsymbol{x}$. Let us refer to the appendix of Ghosh *et al.* (2006).

$$\begin{aligned} I_{\boldsymbol{x}} &= \int \left[\mathbf{X}\middle|\boldsymbol{x}, \sigma_\eta^2\right]\left|\mathbf{X}_*^T\mathbf{X}_*\right|^{-\frac{1}{2}} d\boldsymbol{x} \\ &= \left(\sigma_\eta^2\right)^{-\frac{n_t}{2}} \exp\left[-\frac{1}{2\sigma_\eta^2}\sum_{i=1}^m\sum_{j=1}^{n_i}\left(X_{ij} - \bar{X}_i\right)^2\right] \int \left|\mathbf{X}_*^T\mathbf{X}_*\right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma_\eta^2}\sum_{i=1}^m n_i\left(\bar{X}_i - x_i\right)^2\right] d\boldsymbol{x} \end{aligned} \tag{A.4}$$

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

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

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,

$$I_{\sigma_u^2} = \int \left(\sigma_u^2\right)^{-\frac{(m-p)}{2}} \left[\sigma_u^2\right] d\sigma_u^2 = \int \left(\sigma_u^2\right)^{-\left(a_u+\frac{m}{2}-\frac{p}{2}\right)-1} \exp\left(-\frac{b_u}{\sigma_u^2}\right) d\sigma_u^2 = K_4, \tag{A.5}$$

$$I_{\sigma_x^2} = \int \left(\sigma_x^2\right)^{-\frac{(m-1)}{2}} \left[\sigma_x^2\right] d\sigma_x^2 = \int \left(\sigma_x^2\right)^{-\left(a_x+\frac{m}{2}-1\right)-1} \exp\left(-\frac{b_x}{\sigma_x^2}\right) d\sigma_x^2 = K_5, \tag{A.6}$$

$$I_{\sigma_\eta^2} = \int \left(\sigma_\eta^2\right)^{-\frac{(n_t-m)}{2}} \left[\sigma_\eta^2\right] d\sigma_\eta^2 = \int \left(\sigma_\eta^2\right)^{-\left(a_\eta+\frac{n_t}{2}-\frac{m}{2}\right)-1} \exp\left(-\frac{b_\eta}{\sigma_\eta^2}\right) d\sigma_\eta^2 = K_6, \tag{A.7}$$

where K_4, K_5 and K_6 are some constants.

Combining (A.1)~(A.7), we have

$$I \leq K_1K_2K_3K_4K_5K_6 \int \cdots \int \left[\mathbf{y}|\boldsymbol{\theta}, \sigma_e^2\right]\left[\boldsymbol{\gamma}|\sigma_\gamma^2\right]\left[\sigma_e^2\right]\left[\sigma_\gamma^2\right] d\boldsymbol{\Omega}^*, \tag{A.8}$$

where $\boldsymbol{\Omega}^* = (\boldsymbol{\Omega} - \mu_x - \boldsymbol{b} - \boldsymbol{x} - \sigma_u^2 - \sigma_x^2 - \sigma_\eta^2)$. Since all the components of the integrand in (A.8) have proper distributions, the above integral would be finite thus proving posterior propriety.

References

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties (with Discussion), *Statistical Science,* **11**, 89–121.

Gelman, A. E. and Rubin, D. (1992). Inference from iterative simulation (with discussion), *Statistical Science,* **7**, 457–511.

Ghosh, M. and Meeden, G. (1986). Empirical Bayes estimation in finite population sampling, *Journal of the American Statistical Association,* **81**, 1058–1062.

Ghosh, M. and Meeden, G. (1996). *Bayesian Methods for Finite Population Sampling,* Chapman & Hall, New York.

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

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

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

Received June 2010; Accepted July 2010