Journal of the Korean Data & Information Science Society
2016, **27**(6), 1645–1651

http://dx.doi.org/10.7465/jkdi.2016.27.6.1645
한국데이터정보과학회지

Multivariable Bayesian curve-fitting under functional measurement error model

Jinseub Hwang[1] · Dal Ho Kim[2]

[1]Department of Computer Science and Statistics, Daegu University
[2]Department of Statistics, Kyungpook National University

Received 6 October 2016, revised 26 October 2016, accepted 8 November 2016

Abstract

A lot of data, particularly in the medical field, contain variables that have a measurement error such as blood pressure and body mass index. On the other hand, recently smoothing methods are often used to solve a complex scientific problem. In this paper, we study a Bayesian curve-fitting under functional measurement error model. Especially, we extend our previous model by incorporating covariates free of measurement error. In this paper, we consider penalized splines for non-linear pattern. We employ a hierarchical Bayesian framework based on Markov Chain Monte Carlo methodology for fitting the model and estimating parameters. For application we use the data from the fifth wave (2012) of the Korea National Health and Nutrition Examination Survey data, a national population-based data. To examine the convergence of MCMC sampling, potential scale reduction factors are used and we also confirm a model selection criteria to check the performance.

Keywords: Functional measurement error, hierarchical Bayes, multivariable, penalized spline.

1. Introduction

In recent years, the demand for small area estimation and for solving measurement error problem have greatly increased worldwide. The term "small area" broadly refers to a small geographical area such as a county. It may also refer to a "small domain". And statistically models are established in terms of variables $\boldsymbol{x}$ that for some reason are not directly observable such as blood pressure (BP) and body mass index (BMI). Problems of this nature are commonly called "measurement error" problem and the statistical models and methods for analyzing such data are called measurement error model. In the general terminology it is called structural measurement error model where $\boldsymbol{x}$ is considered as a random variable. This is in contrast to the functional measurement error model where $\boldsymbol{x}$ is considered non-random variable. Ghosh and Meeden (1986) conducted empirical Bayesian estimation of small area

[1] Assistant professor, Department of Computer Science and Statistics, Daegu University, Gyeongsangbuk-do 38453, Korea.
[2] Corresponding author: Professor, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea. E-mail: dalkim@knu.ac.kr

means using a simple one-way ANOVA model. The model could be extended from including of covariates, and that procedures have been commented in Ghosh and Meeden (1996). However, this model is not able to consider the measurement error covariate. Ghosh and Sinha (2007) developed the small area model with functinoal measurement error covariate. Recently, Arima, Datta and Liseo (2015) conducted Bayesian estimators for small area models when auziliary information is measured with error. We developed Bayesian curve-fitting based on penalized splines with functinoal measurement errors model (Hwang and Kim, 2010; Hwang and Kim, 2015). In our previous paper we considered only one covariate having measurement errors. But if there are other useful auxiliary variables that does not have measurement error, then we need to add this variables as covariates in the model to estimate. In this paper, our purpose is to develop the multivariable Bayesian curving-fitting model under functional measurement error. Especially, we consider p covariates and we assume one covariate has measurement error and others have not measurement error. For non-linear pattern of measurement error covariate, we use truncated polynomial basis functions (TPBF), this functions are one of penalized splines. Also, we apply fixed knots based on a equally spaced sample quantiles.

For our model, we carry out a hierarchical Bayesian (HB) approach based on Markov Chain Monte Carlo (MCMC) methodology. First, we prove the propriety of posterior because we consider non-informative improper priors for regression coefficients. Section 2 provides a overview of the model specification and we explain the MCMC application of the HB procedure in Section 3. In Section 4, we conduct a real data analysis and compare models based on model selection criteria. Finally, we suggest some possible extensions of our model in Section 5.

2. Model specification

In this paper, we consider the unit level nested error regression model for estimating small area means. Also we consider smoothing based on TPBF with one dimension and fixed knots for measurement error covariate (x_1). And other covariates ($x_2, x_3, \cdots, x_p$) are considered that have a linear relation with outcome variable without measurement error.

Assume m (labelled $1, \cdots, m$) small areas are in the data. Let N_i (the known pupulation number) for the ith area and X_{1ij} and y_{ij} denote the observed covariate and response of the jth subject in the ith area ($j = 1, \cdots, N_i; i = 1, \cdots, m$), respetively. And let $x_{2i}, x_{3i}, \cdots, x_{pi}$ denote observed other covariates that have not measurement error of the ith area ($i = 1, \cdots, m$). Then the superpopulation model with all covariates and smoothing can be expressed as follows.

$$y_{ij} = \boldsymbol{x}_i^T \boldsymbol{b} + \boldsymbol{z}_i^T \boldsymbol{\gamma} + u_i + e_{ij} \tag{2.1}$$

$$X_{1ij} = x_{1i} + \eta_{ij} \tag{2.2}$$

where $\boldsymbol{x}_i = (1, x_{1i}, x_{2i}, \cdots, x_{pi})^T, \boldsymbol{b} = (b_0, b_1, b_2, \cdots, b_p)^T, \boldsymbol{z}_i = \{(x_{1i} - \tau_1)_+, \cdots, (x_{1i} - \tau_k)_+\}^T$, and $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_k)^T$. Here, $\boldsymbol{z}_i$ presents truncated polynomial basis associated with the measurement error covariate x_1 with k-knots. We presume that e_{ij}, η_{ij} and u_i are mutually independent with normal distribution with mean and variance are 0 and $\sigma_\eta^2, \sigma_e^2$ and σ_u^2, respectively. In this paper, we replace x_{1i} by $\overline{X}_{1i} = N_i^{-1} \sum_{j=1}^{N_i} X_{1ij}$, and then equation (2.1) can express an alternative way as follows.

$$y_{ij} = \theta_i + e_{ij} \tag{2.3}$$

where $\theta_i = \overline{\boldsymbol{X}}_i^T \boldsymbol{b} + \overline{\boldsymbol{Z}}_i^T \boldsymbol{\gamma} + u_i$, $\overline{\boldsymbol{X}}_i = (1, \overline{X}_{1i}, x_{2i}, \cdots, x_{pi})^T$ and $\overline{\boldsymbol{Z}}_i = \{(\overline{X}_{1i} - \tau_1)_+, \cdots, (\overline{X}_{1i} - \tau_k)_+\}^T$. Our goal is to estimate small area means $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)^T$.

3. Hierarchical Bayesian framework

We execute a HB framework based on equation (2.3) for fitting the model and estimating small area means as follows.

Stage 1. $y_{ij} = \theta_i + e_{ij}$ $(i = 1, \cdots, m; j = 1, \cdots, n_i)$ where $e_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$.

Stage 2. $\theta_i = \overline{\boldsymbol{X}}_i^T \boldsymbol{b} + \overline{\boldsymbol{Z}}_i^T \boldsymbol{\gamma} + u_i$ $(i = 1, \cdots, m)$ where $u_i \overset{iid}{\sim} N(0, \sigma_u^2)$.

$X_{1ij} = x_{1i} + \eta_{ij}$ $(i = 1, \cdots, m; j = 1, \cdots, n_i)$ where $\eta_{ij} \overset{iid}{\sim} N(0, \sigma_\eta^2)$.

Stage 3. $\boldsymbol{\gamma} \sim N(\boldsymbol{0}, \sigma_\gamma^2 I)$ where I is the identity matrix of dimension k.

Stage 4. $\boldsymbol{b} = (b_0, b_1, b_2, \cdots, b_p)^T$, $\sigma_\gamma^2, \sigma_\eta^2, \sigma_u^2$, and σ_e^2 are mutually independent with

$\boldsymbol{b} \overset{iid}{\sim} Uniform(-\infty, \infty)$, $(\sigma_\gamma^2)^{-1} \sim G(a_\gamma, b_\gamma)$, $(\sigma_\eta^2)^{-1} \sim G(a_\eta, b_\eta)$,

$(\sigma_u^2)^{-1} \sim G(a_u, b_u)$ and $(\sigma_e^2)^{-1} \sim G(a_e, b_e)$

where $G(\alpha, \beta)$ is a gamma distribution with shape α and rate β parameters

where the function $f(x) \propto x^{\alpha-1} exp(-\beta x)$.

Before proceeding with the computations, we confirm the propriety of joint the posterior because we consider non-informative improper priors for regression coefficients $\boldsymbol{b} = (b_0, b_1, b_2, \cdots, b_p)$. By the conditional independence properties we factorize the full posterior as follows.

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

Our previous paper (Hwang and Kim, 2010) showed the propriety of the posterior.

The Gibbs sampler that is one of the MCMC numerical integration technique is conducted for the implementation of the Bayesian procedure. To generate samples from the full conditions of each parameter given the observed data $(y_{ij}, X_{1ij}, x_{2i}, \cdots, x_{pi})$ and the remaining parameters, we find the full conditional distribution for each parameter.

Full conditional distributions for each parameter

(i) $\left[\theta_i|\boldsymbol{b}, \boldsymbol{\gamma}, \sigma_e^2, \sigma_u^2, \sigma_\gamma^2, \sigma_\eta^2, \boldsymbol{X}, \boldsymbol{y}\right] \overset{iid}{\sim} N\left[(1-D_i)\,\overline{y}_i + D_i\left(\overline{\boldsymbol{X}}_i^T \boldsymbol{b} + \overline{\boldsymbol{Z}}_i^T \boldsymbol{\gamma}\right), \sigma_e^2/n_i\,(1-D_i)\right]$
where $D_i = \sigma_e^2 / \left(\sigma_e^2 + n_i \sigma_u^2\right)$

(ii) $\left[\boldsymbol{b}|\boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma_e^2, \sigma_u^2, \sigma_\gamma^2, \sigma_\eta^2, \boldsymbol{X}, \boldsymbol{y}\right] \sim N\left[\left(\boldsymbol{X}_*^T \boldsymbol{X}_*\right)^{-1} \boldsymbol{X}_*^T \boldsymbol{w}, \sigma_u^2 \left(\boldsymbol{X}_*^T \boldsymbol{X}_*\right)^{-1}\right]$
where $\mathbf{X}_* = \left(\overline{\boldsymbol{X}}_1^T, \cdots, \overline{\boldsymbol{X}}_m^T\right)^T$, $\mathbf{w} = (w_1, \cdots, w_m)^T$, $w_i = \theta_i - \overline{\boldsymbol{Z}}_i^T \boldsymbol{\gamma}$

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

where $\mathbf{Z}_* = \begin{pmatrix} (\overline{X}_1 - \tau_1)_+ & \cdots & (\overline{X}_1 - \tau_k)_+ \\ \vdots & \vdots & \vdots \\ (\overline{X}_m - \tau_1)_+ & \cdots & (\overline{X}_m - \tau_k)_+ \end{pmatrix}$, $\mathbf{t} = (t_1, \cdots, t_m)^T$, $t_i = \theta_i - \overline{\mathbf{X}}_i^T \boldsymbol{b}$;

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

where $n_t = \sum_{i=1}^m n_i$

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

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

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

We generate several sets of samples by L chains and $2d$ iteration for each chain from the full conditional distribution of each parameter. After sampling, we burn out the first half sampling d. We take the averaging principle and the mean of the HB estimates over all d sets.

$$E(\theta_i|\mathbf{X}, \mathbf{y}) = E\left[E\left(\theta_i|\boldsymbol{b}, \boldsymbol{\gamma}, \sigma_e^2, \sigma_u^2, \sigma_\gamma^2, \sigma_\eta^2, \mathbf{X}, \mathbf{y}\right)\right] \simeq (Ld)^{-1} \sum_{l=1}^{L} \sum_{r=d+1}^{2d} \left[\left(1 - D_i^{(lr)}\right) \overline{y}_i + D_i^{(lr)} \left(\overline{\boldsymbol{X}}_i^T \boldsymbol{b}^{(lr)} + \overline{\boldsymbol{Z}}_i^T \boldsymbol{\gamma}^{(lr)}\right)\right] \tag{3.2}$$

and the posterior variance is estimated as

$$\begin{aligned} V(\theta_i|\mathbf{X}, \mathbf{y}) &= E\left[V\left(\theta_i|\boldsymbol{b}, \boldsymbol{\gamma}, \sigma_e^2, \sigma_u^2, \sigma_\gamma^2, \sigma_\eta^2, \mathbf{X}, \mathbf{y}\right)\right] + V\left[E\left(\theta_i|\boldsymbol{b}, \boldsymbol{\gamma}, \sigma_e^2, \sigma_u^2, \sigma_\gamma^2, \sigma_\eta^2, \mathbf{X}, \mathbf{y}\right)\right] \\ &\simeq (Ld)^{-1} \sum_{l=1}^{L} \sum_{r=d+1}^{2d} \left(\frac{\sigma_e^{2(lr)}}{n_i}(1 - D_i^{(lr)})\right) \qquad (3.3) \\ &+ (Ld)^{-1} \sum_{l=1}^{L} \sum_{r=d+1}^{2d} \left[\left(1 - D_i^{(lr)}\right) \overline{y}_i + D_i^{(lr)} \left(\overline{\boldsymbol{X}}_i^T \boldsymbol{b}^{d(lr)} + \overline{\boldsymbol{Z}}_i^T \boldsymbol{\gamma}^{(lr)}\right)\right]^2 \\ &- [E(\theta_i|\mathbf{X}, \mathbf{y})]^2 \end{aligned}$$

4. Data analysis

We studied about the requirement of measurement error model in our previous paper (Hwang, 2015) based on a real data. In this study, the key feature of our implementation is that we use additional covariates under functional measurement error model to estimate small area means based on the same data. We use the data from the fifth wave (2012) of the KNHANES that is a nationally representative cross-section. This survey has been systematically executed since 1988 by the Korean Centre for Disease Control and Prevention (KCDC). Even if this data was collected by multistage probability sampling design we don't consider survey design such as sampling weight. This data include many information such as demographic (age, gender), lifestyle, personal medical history, family history, lab (BP, weight, height) information and so on (Korea Centers for Disease Control and Prevention, 2013).

The risk factors of blood pressure are known as age, gender, obesity, consumption of sodium, potassium vitamin D, tobacco and so on. In this paper, we want to estimate blood pressure at each small group stratified by age and gender. So, we consider diastolic blood pressure (DBP) and systolic blood pressure (SBP) as outcome variable and BMI ($kg/\sqrt{m^2}$) as a measurement error covariate. BMI is an attempt to quantify the amount of tissue mass (muscle, fat, and bone) in an individual, and then categorize that person as underweight, normal weight, overweight, or obese based on that value. Also we use amount of sodium (mg/day), potassium (mg/day) and vitamin D (ng/mL)as other covariates without measurement errors.

The number of total subjects for 2012 was 8,058. We take 446 subjects by excluding under aged 19 years, who had hypertension by taking medication, DBP above 90 $mmHG$ or SBP above 140 $mmHG$. Figure 4.1 indicates scatter plots between (SBP and DBP) and BMI. The real line (—-) is the fitted line from locally weighted scatterplot smoothing. We can see a non-linear pattern.

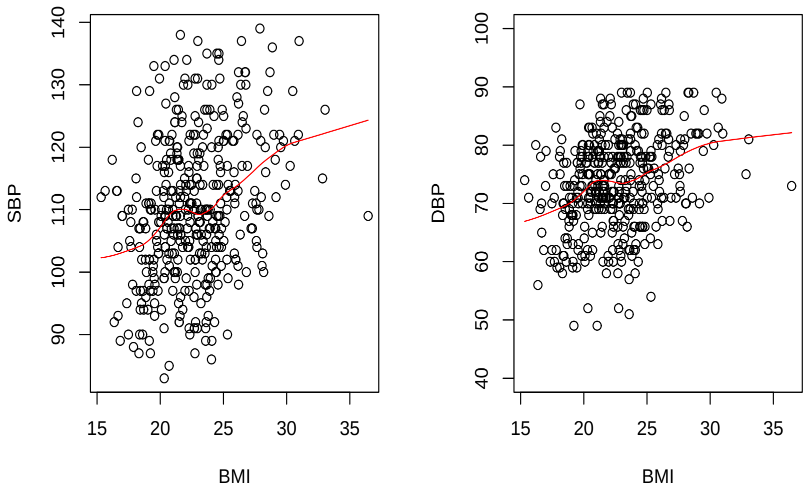

Figure 4.1 Scatter plot between (SBP, DBP) and BMI

To fit models, we run five independent chains ($L = 5$) with runs of length 10,000 ($d = 5,000$) following burn-ins of 5,000 and we give 1.0 for all hyperparameters $a_\gamma, b_\gamma, a_\eta, b_\eta, a_u, b_u$, a_e and b_e and 3 for k. We conduct sensitivity analysis by changing value of hyperparameters. By equation (3.2) and (3.3) we estimate the small area means and standard error based on sampled data. We confirm the convergence by $\sqrt{\hat{R}_i}$ (Gelman and Rubin, 1992). And the mean logarithmic conditional predictive ordinate ($\overline{LCPO}_1$) of Carlin and Louis (2009) and the posterior predictive p-value (Meang, 1994) are used for checking the model adequacy. If the $p-$values is an extreme value (close 0 to or 1), we can conclude that the model is not fit well. Otherwise, when the model fit the data well, $p-$value close to 0.5.

In this data analysis, $\sqrt{\hat{R}} \simeq 1$ for all θ_i and 4 models (Covariates; Model 1: BMI, Model 2: BMI + vitman D, Model 3: BMI + vitman D + sodium, Model 4: BMI + vitman D + sodium + potassium). We show the sample size, small area means, standard error (s.e.) for each strata and ($\overline{LCPO}_1$) and the posterior predictive p-value are in Table 4.1 and 4.2 for each outcome variable and models.

In Table 4.1, we can see that SBP of female is generally higher than male for all ages and we can colclude that Model 2 with BMI and vitamin D for SBP is better than other models based on both $\overline{LCPO}_1$ and p-value as 4.527 and 0.458, respectively, even though the difference is so small. Also we can see that both male and female have the highest SBP in 20's based on Model 2.

In Table 4.2, DBP of female is also higher than male for all ages but it can be seen that Model 1 with BMI only for DBP is better than other models based on both $\overline{LCPO}_1$ and p-value as 4.306 and 0.459, respectively. And both male and female have the highest DPB in 50's based on Model 2.

Table 4.1 Results for SBP

Category	n_i	Model 1		Model 2		Model 3		Model 3	
		Means	s.e.	Means	s.e.	Means	s.e.	Means	s.e.
20's Male	26	114.203	1.687	114.639	1.422	114.443	1.445	113.542	1.783
30's Male	41	103.922	1.565	104.096	1.539	103.475	1.646	103.598	1.587
40's Male	39	108.959	1.524	108.677	1.367	108.275	1.520	108.416	1.542
50's Male	28	113.468	1.733	113.836	1.497	114.685	1.735	114.277	1.716
60's Male	16	106.646	2.196	106.678	1.838	106.545	1.815	106.537	1.813
20's Female	57	117.417	1.325	118.833	1.325	118.830	1.338	119.044	1.360
30's Female	68	113.801	1.178	112.841	1.227	113.664	1.215	113.860	1.236
40's Female	87	118.218	1.064	117.892	1.029	117.513	1.076	117.752	1.090
50's Female	62	114.380	1.311	114.274	1.214	114.059	1.208	113.802	1.203
60's Female	22	117.584	2.068	117.509	1.770	118.067	1.716	117.701	1.815
$\overline{LCPO}_1$		4.553		4.527		4.557		4.564	
PPP		0.436		0.458		0.448		0.439	

Table 4.2 Results for DBP

Category	n_i	Model 1		Model 2		Model 3		Model 3	
		Means	s.e.	Means	s.e.	Means	s.e.	Means	s.e.
20's Male	26	74.805	1.093	74.583	1.050	74.334	1.143	74.432	1.390
30's Male	41	68.946	1.182	68.902	1.165	68.817	1.290	68.908	1.246
40's Male	39	72.576	1.018	72.956	1.038	72.403	1.218	72.373	1.210
50's Male	28	75.726	1.140	75.975	1.107	76.418	1.328	76.277	1.421
60's Male	16	71.172	1.344	71.671	1.347	71.295	1.431	71.196	1.496
20's Female	57	73.734	1.009	73.034	1.029	72.957	1.051	72.922	1.061
30's Female	68	75.201	0.894	75.332	0.864	75.819	0.937	75.810	0.964
40's Female	87	78.348	0.789	78.416	0.770	78.281	0.794	78.255	0.838
50's Female	62	78.584	0.903	78.863	0.887	78.776	0.911	78.781	0.953
60's Female	22	75.983	1.249	75.294	1.352	75.750	1.420	75.962	1.491
$\overline{LCPO}_1$		4.306		4.332		4.321		4.318	
PPP		0.459		0.446		0.444		0.439	

5. Discussion

We develop multivariable Bayesian curve-fitting regression under functional measurement error model. We demonstrated the availability of measurement error models in our previous paper (Hwang, 2015). In this paper we show the availability with additional auxiliary covariates under measurement error model based on the real data. But we don't consider some posibility that additional auxiliary covariates (vitiman D, sodium and potassium) also could have a measurement error, so we will extend multivariable model with $p-$dimensional measurement error and $q-$dimensional non-measurement error covariates. And, outcome variable (SBP and DBP) also have an easurement error, so we have plan to extend measurement error model with measurement error of covariate and outcome variable. Next, we consider functional measurement error in this paper and we will develop structural measurement error case. Finally, we consider fixed knots in this paper, so we can extend our model with random knots.

References

Arima S., Datta, G. S. and Liseo B. (2014). Bayesian estimators for small area models when auxiliary information is measured with error. *Scandinavian Journal of Statistics*, **42**, 518-529.

Carlin, B. O. and Louis, T. A. (2009). *Bayesian methods for data analysis,* 3rd Ed., Chapman & Hall/CRC, Boca Raton.

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.

Hwang, J. (2015). Statistical analysis of KNHANES data with measurement error models. *Journal of the Korean Data and Information Science Society*, **26**, 773-779.

Hwang, J. and Kim, D. (2015). Bayesian curve-fitting with radial basis functions under functional measurement error model. *Journal of the Korean Data and Information Science Society*, **26**, 749-754.

Hwang, J. and Kim, D. (2010). Semiparametric Bayesian estimation under functional measurement error model. *Journal of the Korean Data and Information Science Society*, **21**, 379-385.

Korea Centers for Disease Control & Prevention. (2013). *Statistics on the fourth Korea National Health and Nutrition Examination Survey (KNHANES IV)*, Korea Centers for Disease Control & Prevention, Osong, Korea.

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