Journal of the Korean Data & Information Science Society
2015, **26**(3), 749–754

http://dx.doi.org/10.7465/jkdi.2015.26.3.749
한국데이터정보과학회지

Bayesian curve-fitting with radial basis functions under functional measurement error model

Jinseub Hwang[1] · Dal Ho Kim[2]

[1]National Evidence-based Healthcare Collaborating Agency
[2]Department of Statistics, Kyungpook National University

Received 23 March 2015, revised 13 April 2015, accepted 27 April 2015

Abstract

This article presents Bayesian approach to regression splines with knots on a grid of equally spaced sample quantiles of the independent variables under functional measurement error model. We consider small area model by using penalized splines of non-linear pattern. Specifically, in a basis functions of the regression spline, we use radial basis functions. To fit the model and estimate parameters we suggest a hierarchical Bayesian framework using Markov Chain Monte Carlo methodology. Furthermore, we illustrate the method in an application data. We check the convergence by a potential scale reduction factor and we use the posterior predictive $p-$value and the mean logarithmic conditional predictive ordinate to compar models.

Keywords: Functional, hierarchical Bayes, measurement error, radial basis, semiparametric.

1. Introduction

We developed the semiparametric small area models with measurement errors models in our previous paper (Hwang and Kim, 2010). Specifically, we considered small area model by using penalized splines of non-linear pattern based on truncated polynomial basis functions and knots on a grid of equally spaced sample quantiles under functional measurement error model. Measurement error modeling is also related to Goo and Kim (2013).

The truncated polynomial basis functions (TPBF) is simple, but not always numerically stable when the number of knots is large and the smoothing parameter close to zero. In this case the computation has to be organized carefully and numerically superior alternatives are available, like B-splines and radial basis functions (Ruppert *et al.*, 2003).

The objective of this article is to develop alternative estimators of small area means by using radial basis functions (RBF) with functional measurement error model. RBF is defined with degree p from Ruppert *et al.* (2003) as follows.

$$1, x, \cdots, x^{p-1}, |x-\tau_1|^{2p-1}, \cdots, |x-\tau_k|^{2p-1}. \tag{1.1}$$

[1] Research fellow, Ph.D, National Evidence-based Healthcare Collaborating Agency, Seoul 100-705, Korea.
[2] Corresponding author: Professor, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea. E-mail: dalkim@knu.ac.kr

Here k is the number of knots, $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_k)^T$ is the vector of knot location ($\tau_1 < \cdots < \tau_k$) and p is the degree of polynomial pieces.

For our model, we have conducted a hierarchical Bayesian (HB) approach using Markov Chain Monte Carlo (MCMC) methodology, specifically Gibbs sampling. Before the computations, we have proved the propriety of the posterior since we have used non-informative improper priors for some parameters. But we have skipped this proof in this article because the proof is similar with our previous model.

Section 2 gives a brief overview of the model specification and we discuss the MCMC implementation of the proposed hierarchical Bayes procedure in Section 3. In Section 4, we conduct the analysis and compare two models based on a real data. Finally, we discuss several possible refinements and extensions of our model in Section 5.

2. Semiparametric model under functional measurement errors

In small area estimation, we often use one of two common models, so-called the area level "Fay-Herriot model" and the unit level "nested error regression model". In this paper we deal with the unit level nested error regression model. Also we consider semiparametric regression using RBF with p=1 and knots on a grid of equally spaced sample quantiles under functional measurement error model assumed that the covariate is measured with error and non-stochastic.

Suppose there are m small areas (labelled $1, \cdots, m$) and N_i (the known pupulation size) for the ith area. And let y_{ij} and X_{ij} denote the observed response and covariate of the jth unit in the ith area ($j = 1, \cdots, N_i; i = 1, \cdots, m$), respectively. Then the superpopulation model with semiparametric regression using RBF under functional measurement error can be expressed as follows.

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

$$X_{ij} = x_i + \eta_{ij} \tag{2.2}$$

where $\boldsymbol{x}_i = (1, x_i)^T, \boldsymbol{b} = (b_0, b_1)^T, \boldsymbol{z}_i = \{|x_i - \tau_1|, \cdots, |x_i - \tau_k|\}^T$, and $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_k)^T$. We assume that u_i, e_{ij} and η_{ij} are mutually independent with normal distribution having mean and variance are 0 and σ_u^2, σ_e^2 and σ_η^2, respectively. In this paper, we substitute x_i by $\bar{X}_i = N_i^{-1} \sum_{j=1}^{N_i} X_{ij}$, and we can express an alternative way as follows.

$$y_{ij} = \theta_i + e_{ij}; \theta_i = \bar{\boldsymbol{X}}_i^T \boldsymbol{b} + \bar{\boldsymbol{Z}}_i^T \boldsymbol{\gamma} + u_i \tag{2.3}$$

where $\bar{\boldsymbol{X}}_i = (1, \bar{X}_i)^T$ and $\bar{\boldsymbol{Z}}_i = \{|\bar{X}_i - \tau_1|, \cdots, |\bar{X}_i - \tau_k|\}^T$. Here, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)^T$ is the goal that we want to estimate.

3. Bayesian approach to adaptive model

To fit the model and estimate small area means based on sample (n_i is drawn from the ith area and $\sum_{i=1}^m n_i = n_t$), we conduct a hierarchical Bayesian framework based on equation (2.3):

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

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

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

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

Stage 4. $b_0, b_1, \sigma_e^2, \sigma_u^2, \sigma_\eta^2$ and σ_γ^2 are mutually independent with

b_0 & $b_1 \sim Uniform(-\infty, \infty)$, $(\sigma_e^2)^{-1} \sim G(a_e, b_e)$, $(\sigma_u^2)^{-1} \sim G(a_u, b_u)$, $(\sigma_u^2)^{-1} \sim G(a_u, b_u)$, $(\sigma_\eta^2)^{-1} \sim G(a_\eta, b_\eta)$, $(\sigma_\gamma^2)^{-1} \sim G(a_\gamma, b_\gamma)$

where $G(\alpha, \beta)$ denotes an gamma distribution with shape parameter α and rate parameter β having the expression $f(x) \propto x^{\alpha-1} exp(-\beta x)$.

First, we check the priporiety of the joint posterior since we use non-informative improper priors for the regression parameters b_0 and b_1. We can factorize the full posterior by the conditional independence properties as follows.

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

The detail proof of the propriety of the posterior is referred at the appendix in Hwang and Kim (2010).

We use the MCMC numerical integration technique for the implementation of the Bayesian procedure, in particular the Gibbs sampler. To generate samples from the full conditions of $\boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\gamma}, \sigma_e^2, \sigma_u^2, \sigma_\eta^2$ and σ_γ^2 given the remaining parameters and the observed data (y_{ij}, X_{ij}), we calculate the full conditional distribution and generate samples based on these distributions.

Full conditional distributions

(i) $\left[\theta_i | \boldsymbol{b}, \boldsymbol{\gamma}, \sigma_e^2, \sigma_u^2, \sigma_\gamma^2, \sigma_\eta^2, \boldsymbol{X}, \boldsymbol{y}\right] \overset{iid}{\sim} N\left[(1 - D_i)\, \bar{y}_i + D_i \left(\bar{\boldsymbol{X}}_i^T \boldsymbol{b} + \bar{\boldsymbol{Z}}_i^T \boldsymbol{\gamma}\right), \sigma_e^2 / n_i\, (1 - D_i)\right]$
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 $\mathbf{X}_* = \left(\bar{\boldsymbol{X}}_1^T, \cdots, \bar{\boldsymbol{X}}_m^T\right)^T$, $\mathbf{w} = (w_1, \cdots, w_m)^T$, $w_i = \theta_i - \bar{\boldsymbol{Z}}_i^T \boldsymbol{\gamma}$

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

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

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

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

(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 run $L(\geq 2)$ chains and $2d$ iteration for each chain. After sampling from the full conditional distribution, we burn out the first half and use the averaging principle and take the average of the HB estimates over all the remaining sets to obtain the final HB estimates. The HB estimator for small area means is approximated as follows.

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

and the posterior variance is estimated as:

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

4. Numerical analysis

The key feacutre of our implementation is that we use a semiparametric regression model with a radial basis functions instead of truncated polynomial basis functions. We conduct the analysis using a real data and compare the result between two models. The analysis is conducted using R 3.1.3.

The LANDSAT (Land observatory satellites) data is a compilation (by Battese et.al, 1988) of survey and satellite data from the U.S. Department of Agriculture (USDA).This data 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) have been determined by USDA Statistical Reporting Service staff, who interviewed farm operators. This data consists of 37 observations and 10 variables (number of segments per county, hectares of corn and soybean for each sample segment, number of pixels classified as corn and soybean for each sample segment and so on). In this article we consider prediction of the areas under corn (OUTCOME) using corn pixels (PixelsCorn) only as a measurement error covariate. Figure 4.1 is the scatter plot for LANDSAT data and the real line (——) indicates the fitted line based on locally weighted scatterplot smoothing. There is a little non-linear pattern.

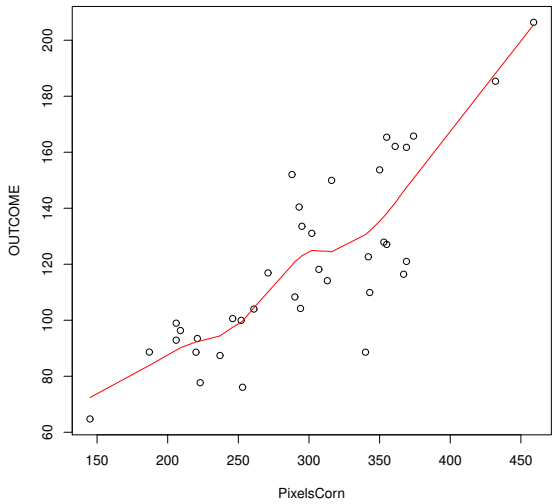

Figure 4.1 Figure scatter plot for LANDSAT data

We conduct five independent chains with runs of length 5,000 following burn-ins of 2,500 and we set all 1.0 for all hyperparameters $a_e, b_e, a_u, b_u, a_\eta, b_\eta, a_\gamma$ and b_γ. The small area means and standard error are estimated using (3.2) and (3.3). We check the convergence and model adequacy based on $\sqrt{\hat{R}_i}$ (Gelman and Rubin, 1992), the posterior predictive p-value (Meang, 1994) and the mean logarithmic conditional predictive ordinate (Carlin and Louis, 2009), respectively.

In our all case, $\sqrt{\hat{R}} \simeq 1$ for all θ_i and the detailed results are reported in Table 4.1. We report the sample size, estimates, standard error (s.e.), the posterior predictive p-value (p-value) and the mean logarithmic conditional predictive ordinate ($\overline{LCPO}_1$) for each case. It can be seen that the model with three knots is better for all of two basis functions based on p-value and $\overline{LCPO}_1$. Also, TPBS and RBF models with three knots have same model adequacy by p-value, which is 0.44. And truncated polynomial basis function is better than radial basis function by $\overline{LCPO}_1$, which are 5.44 and 5.45, respectively, but the difference is very small.

Table 4.1 Model fitting results for LANDSAT data

counties	n_i	Truncated polynomial basis						Radial basis					
		$K=1$		$K=3$		$K=7$		$K=1$		$K=3$		$K=7$	
		Est.	s.e.	Est.	s.e.	Est.	s.e.	Est.	s.e.	Est.	s.e.	Est.	s.e.
Cerro Gordo	1	160.22	8.67	150.83	8.28	152.45	9.94	159.66	9.50	152.61	10.51	149.16	13.84
Hamilton	1	97.71	11.14	94.82	10.65	104.03	13.77	102.12	10.97	95.28	11.31	101.07	14.46
Worth	1	106.47	6.40	108.19	8.49	114.35	8.34	107.60	4.97	109.38	12.96	111.44	9.52
Humboldt	2	174.51	11.44	156.46	11.19	155.63	14.22	174.39	12.76	149.87	15.98	154.69	15.62
Franklin	3	140.71	6.12	139.87	8.89	138.09	8.28	139.54	6.49	147.66	7.64	137.24	8.46
Pocahontas	3	100.44	9.27	99.35	9.30	109.38	11.47	103.86	8.53	100.30	9.53	104.07	12.27
Winnebago	3	114.26	7.13	112.38	7.88	114.87	6.04	112.56	7.78	108.66	9.35	115.33	8.53
Wright	3	147.65	6.75	145.73	9.76	145.64	9.66	146.70	7.20	154.69	8.62	144.22	10.34
Webster	4	108.82	5.94	109.60	6.80	114.69	6.41	109.19	4.76	109.45	8.84	113.35	6.78
Hancock	5	112.80	6.52	111.57	6.77	114.67	5.03	111.61	6.68	108.78	7.13	115.25	6.86
Kossuth	5	125.66	6.20	123.94	6.73	119.82	8.86	124.03	6.64	124.93	7.55	118.67	11.14
Hardin	6	113.20	6.60	111.87	6.94	114.82	4.98	111.91	6.87	108.90	7.50	115.39	6.88
$\overline{LCPO}_1$		5.50		5.44		5.56		5.50		5.45		5.56	
p-value		0.43		0.44		0.43		0.43		0.44		0.43	

5. Discuss

We have developed a semiparametric regression model with radial basis functions instead of truncated polynomial basis functions under functional measurement error model with knots on a grid of equally spaced sample quantiles of the independent variables. We can extend our model in several situations. First, we can consider small area models with structural measurement errors. Second, free knots and another basis functions like B-spline could be considered. Also, in this paper, we have considered only the normal outcome and covariate with measurement error. But we can consider the generalized linear model to accommodate binomial or Poisson data.

References

Battese, G., Harter, R. and Fuller, W. (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.

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.

Goo, Y. M. and Kim, D. H. (2013). Bayesian small area estimations with measurement errors. *Journal of the Korean Data & Information Science Society*, **24**, 885-893.

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

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

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