Journal of the Korean Data & Information Science Society
2016, **27**(5), 1433–1441

http://dx.doi.org/10.7465/jkdi.2016.27.5.1433
한국데이터정보과학회지

A comparative study in Bayesian semiparametric approach to small area estimation[†]

Simyoung Heo[1] · Dal Ho Kim[2]

[1]Dongbuk Regional Office of Statistics
[2]Department of Statistics, Kyungpook National University

Received 10 July 2016, revised 9 September 2016, accepted 13 September 2016

Abstract

Small area model provides reliable and accurate estimations when the sample size is not sufficient. Our dataset has an inherent nonlinear pattern which significantly affects our inference. In this case, we could consider semiparametric models such as truncated polynomial basis function and radial basis function. In this paper, we study four Bayesian semiparametric models for small areas to handle this point. Four small area models are based on two kinds of basis function and different knots positions. To evaluate the different estimates, four comparison measurements have been employed as criteria. In these comparison measurements, the truncated polynomial basis function with equal quantile knots has shown the best result. In Bayesian calculation, we use Gibbs sampler to solve the numerical problems.

Keywords: Basis function, Gibbs sampler, knots, small area model.

1. Introduction

Sometimes, there is a case where the sample size in specific domain is too small to provide good estimations. In this case, we usually consider the models including small area random effects which are called small area models. These small area estimations deal with the problem of providing reliable estimates when the information available on those variables is not sufficient to provide accurate direct estimate. In this sense, the demand for small area estimation has greatly increased. A study on the small area estimation is related to Ghosh and Rao (1994), Ghosh, Nangia and Kim (1996) and Lee and Kim (2015). These estimates play an important role in formulating policies and programs, in the allocation of government funds and in regional planning.

Area-level models based on area direct survey estimators are popular in model based small area inference. The Fay-Herriot (1979) model is a popular area-level model. The Fay-Herriot model produces reliable small area estimates by combining the design model and the regression model and then borrowing strength from other domains. It is assumed that the direct survey estimators are linear function of the covariates. When this assumption fails

[†] This paper is based on part of the first author's master thesis.
[1] Local statistics division, Dongbuk Regional Office of Statistics, Daegu 41422, Korea.
[2] Corresponding author: Professor, Department of Statistics, Kyungpook National University, Daegu 41566, Korea. E-mail: dalkim@knu.ac.kr

down, the Fay-Herriot model can lead to biased estimators of the small area parameters. A semiparametric specification of the Fay-Herriot model, which allows nonlinearities in the relationship between Y and the auxiliary variables X, can be obtained by penalized-splines (or P-spline) (Eilers and Marx, 1996). Current work in small area model by using penalized splines of nonlinear pattern has been studied in Bhadra, Ghosh and Kim (2012) and Hwang and Kim (2015). P-spline, expressed using truncated polynomial basis functions with varying degrees and number of knots, is a commonly used but powerful function estimation tool in nonparametric approaches.

In this paper, the estimation of median income for four-person families was our interest. We set a semiparametric modeling procedure for estimating the median household income for all the U.S states. Figure 1.1 shows the plot of the Current Population Survey (CPS) median income against the Internal Revenue Service (IRS) mean income for all the states for the years 1995 through 1999. It is apparent that CPS median income may have an underlying nonlinear pattern with respect to IRS mean income, especially for large values of the latter.

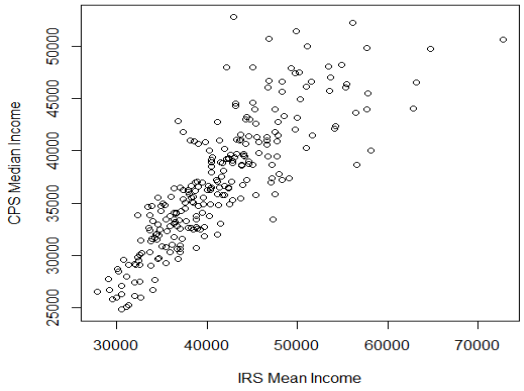

Figure 1.1 IRS mean income vs CPS median income

The organization of our paper is as follows. In Section 2, we introduce the two types of Bayesian semiparametric models. In Section 3, we provide the result of the numerical study. It includes a real example and numerical analysis using two types of method of positioning knots. Then we give our concluding remarks in Section 4.

2. Bayesian semiparametric models

2.1. Semiparametric income trajectory models

Let Y_{ij} and X_{ij} denote the CPS median household income and the IRS mean income recorded for the i^{th} state and j^{th} year. The basic semiparametric model can be expressed as

$$Y_{ij} = f(x_{ij}) + b_i + u_{ij} + e_{ij},$$

where $f(x_{ij})$ is an unspecified function of x_{ij} reflecting the unknown response-covariate relationship. b_i is a state-specific random effect while u_{ij} shows an interaction effect between the

i^{th} state and the j^{th} year. Further, it is assumed that u_{ij} and e_{ij} are mutually independent with $u_{ij} \sim N(0, \psi_j^2)$ and $e_{ij} \sim N(0, \sigma_{ij}^2)$.

We approximate $f(x_{ij})$ using the truncated polynomial basis function (TPBF),

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 x_{ij} + \cdots + \beta_p x_{ij}^p + \sum_{k=1}^{K} \gamma_k (x_{ij} - \tau_k)_+^p + b_i + u_{ij} + e_{ij} \\ &= X_{ij}'\boldsymbol{\beta} + Z_{ij}'\boldsymbol{\gamma} + b_i + u_{ij} + e_{ij} \\ &= \theta_{ij} + e_{ij}, \end{aligned}$$

where $\theta_{ij} = X_{ij}'\boldsymbol{\beta} + Z_{ij}'\boldsymbol{\gamma} + b_i + u_{ij}$ is our target of interest. Here $X_{ij} = (1, x_{ij}, \cdots, x_{ij}^p)'$, $Z_{ij} = \{(x_{ij} - \tau_1)_+^p, \cdots, (x_{ij} - \tau_K)_+^p\}'$, $\boldsymbol{\beta} = (\beta_0, \cdots, \beta_p)'$ is the vector of regression coefficients while $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_K)'$ is the vector of spline coefficients. We assume $b_i \sim^{iid} N(0, \sigma_b^2)$ and $\boldsymbol{\gamma} \sim N(\mathbf{0}, \sigma_\gamma^2 \boldsymbol{I}_K)$. Commonly, linear or quadratic splines provides most practical purposes since they ensure adequate smoothness in the fitted curve.

When a real-valued function depending only on the distance from the origin, we call this a radial basis function (RBF), $\phi(\boldsymbol{x}) = \phi(\|\boldsymbol{x}\|)$. Alternatively, the function depending on the distance from some other point c, called a center, is expressed as $\phi(\boldsymbol{x}, \boldsymbol{c}) = \phi(\|\boldsymbol{x} - \boldsymbol{c}\|)$. The radial function, $Z_{ij}'\boldsymbol{\gamma}$, is used to approximate $f(x_{ij})$ in the model,

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 x_{ij} + \cdots + \beta_p x_{ij}^p + \sum_{k=1}^{K} \gamma_k |x_{ij} - \tau_k| + b_i + u_{ij} + e_{ij} \\ &= X_{ij}'\boldsymbol{\beta} + Z_{ij}'\boldsymbol{\gamma} + b_i + u_{ij} + e_{ij} \\ &= \theta_{ij} + e_{ij}. \end{aligned}$$

Here $X_{ij} = (1, x_{ij}, \cdots, x_{ij}^p)'$, $Z_{ij} = \{|x_{ij} - \tau_1|, \cdots, |x_{ij} - \tau_K|\}'$, $\boldsymbol{\beta} = (\beta_0, \cdots, \beta_p)'$ is the vector of regression coefficients while $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_K)'$ is the vector of spline coefficients. We assume $b_i \sim^{iid} N(0, \sigma_b^2)$ and $\boldsymbol{\gamma} \sim N(\mathbf{0}, \sigma_\gamma^2 \boldsymbol{I}_K)$.

Here m and t respectively denote the number of small areas and time points at which the response and covariates are taken. In this paper, $m = 51$ for the 50 U.S states and the District of Columbia and $t = 5$ for the years 1995-1999.

2.2. Hierarchical Bayesian inference

Let $\boldsymbol{Y}_i = (Y_{i1}, \cdots, Y_{it})'$ be the response and $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \cdots, \boldsymbol{X}_{it})'$ and $\boldsymbol{Z}_i = (\boldsymbol{Z}_{i1}, \cdots, \boldsymbol{Z}_{it})'$ be the covariates for the i^{th} state. Let $\boldsymbol{\Omega}_i = (\boldsymbol{\theta}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, b_i, \boldsymbol{\psi}, \sigma_b^2, \sigma_\gamma^2)$ be the parameter space. The full parameter space is $\boldsymbol{\Omega} = \boldsymbol{\Omega}_1 \times \cdots \times \boldsymbol{\Omega}_m$. Thus, the likelihood function is as below.

$$\begin{aligned} L(\boldsymbol{Y}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i | \boldsymbol{\Omega}_i) &\propto L(\boldsymbol{Y}_i | \boldsymbol{\beta}, \boldsymbol{\gamma}, b_i, \boldsymbol{\psi}^2, \boldsymbol{X}_i, \boldsymbol{Z}_i) L(b_i | \sigma_b^2) L(\boldsymbol{\gamma} | \sigma_\gamma^2) \\ &\propto \prod_{j=1}^{t} L(Y_{ij} | \theta_{ij}, \sigma_{ij}^2) L(\theta_{ij} | \boldsymbol{X}_{ij}'\boldsymbol{\beta} + \boldsymbol{Z}_{ij}'\boldsymbol{\gamma} + b_i, \psi_j^2) L(b_i | \sigma_b^2) L(\boldsymbol{\gamma} | \sigma_\gamma^2). \end{aligned}$$

Here, $L(U|a, b)$ denotes a normal distribution with mean a and variance b while $L(V|a)$ denotes a normal density with mean 0 and variance a.

We assign noninformative improper uniform prior for the polynomial coefficients and proper conjugate gamma prior on the inverse of the variance components. The prior distributions are assumed to be mutually independent. We have the following priors: $\boldsymbol{\beta} \sim uniform(R^{p+1})$, $(\boldsymbol{\psi}_j^2)^{-1} \sim Gamma(c_j, d_j)$ $(j = 1, \cdots, t)$, $(\sigma_b^2)^{-1} \sim Gamma(c, d)$ and $(\sigma_\gamma^2)^{-1} \sim Gamma(c_\gamma, d_\gamma)$. Here $X \sim G(a, b)$ denotes a gamma distribution with shape parameter a and rate parameter b having the expression $f(x) \propto x^{a-1}exp(-bx), x \geqq 0$.

The full posterior of the parameters given the data is as follows.

$$p(\boldsymbol{\Omega}|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}) \propto \prod_{i=1}^{m} L(\boldsymbol{Y}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i|\boldsymbol{\Omega}_i)\pi(\boldsymbol{\beta})\pi(\sigma_b^2)\pi(\sigma_\gamma^2) \prod_{j=1}^{t} \pi(\psi_j^2),$$

which is the product of the likelihood functions and prior distributions.

The conditional density of θ_{ij} is given by

$$\pi(\theta_{ij}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\psi}^2, \boldsymbol{b}, \boldsymbol{X}, \boldsymbol{Z}) \propto exp\left(-\frac{\left(\theta_{ij} - \frac{y_{ij}\psi_j^2 + (X_{ij}'\boldsymbol{\beta} + Z_{ij}'\boldsymbol{\gamma} + b_i)\sigma_{ij}^2}{\psi_j^2 + \sigma_{ij}^2}\right)^2}{2\frac{\sigma_{ij}^2\psi_j^2}{\psi_j^2 + \sigma_{ij}^2}} \right).$$

Therefore the conditional density of θ_{ij} is as follows.

$$[\theta_{ij}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\psi}^2, \boldsymbol{b}, \boldsymbol{X}, \boldsymbol{Z}]$$
$$\sim N\left[\left(\frac{1}{\sigma_{ij}^2} + \frac{1}{\psi_j^2}\right)^{-1}\left(\frac{y_{ij}}{\sigma_{ij}^2} + \frac{(\boldsymbol{X}_{ij}'\boldsymbol{\beta} + \boldsymbol{Z}_{ij}'\boldsymbol{\gamma} + b_i)}{\psi_j^2}\right), \left(\frac{1}{\sigma_{ij}^2} + \frac{1}{\psi_j^2}\right)^{-1}\right].$$

The conditional density of b_i is given by

$$\pi(b_i|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\psi}^2, \sigma_b^2, \boldsymbol{X}, \boldsymbol{Z}) \propto exp\left\{ -\frac{1}{2}\frac{b_i^2 - 2\left(\sum_j \frac{\theta_{ij} - X_{ij}'\boldsymbol{\beta} - Z_{ij}'\boldsymbol{\gamma}}{\psi_j^2}\right) / \left(\sum_j \frac{1}{\psi_j^2} + \frac{1}{\sigma_b^2}\right) b_i}{\left(\sum_j \frac{1}{\psi_j^2} + \frac{1}{\sigma_b^2}\right)^{-1}} \right\}.$$

So the conditional density of b_i is as follows.

$$[b_i|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\psi}^2, \sigma_b^2, \boldsymbol{X}, \boldsymbol{Z}]$$
$$\sim N\left[\left(\frac{1}{\sigma_b^2} + \sum_{j=1}^{t}\frac{1}{\psi_j^2}\right)^{-1}\left(\sum_{j=1}^{t}\frac{1}{\psi_j^2}(\theta_{ij} - \boldsymbol{X}_{ij}'\boldsymbol{\beta} - \boldsymbol{Z}_{ij}'\boldsymbol{\gamma})\right), \left(\frac{1}{\sigma_b^2} + \sum_{j=1}^{t}\frac{1}{\psi_j^2}\right)^{-1}\right].$$

The conditional density of $\boldsymbol{\beta}$ is given by

$$\pi(\boldsymbol{\beta}|\boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\psi}^2, \boldsymbol{X}, \boldsymbol{Z})$$
$$\propto exp\left[-\frac{1}{2}\left(\boldsymbol{\beta}'\left(\sum_i\sum_j \frac{X_{ij}X_{ij}'}{\psi_j^2}\right)\boldsymbol{\beta} - 2\sum_i\sum_j \boldsymbol{\beta}'\frac{X_{ij}(\theta_{ij} - Z_{ij}'\boldsymbol{\gamma} - b_i)}{\psi_j^2}\right)\right].$$

Thus the conditional density of $\boldsymbol{\beta}$ is as follows.

$$[\boldsymbol{\beta}|\boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\psi}^2, \boldsymbol{X}, \boldsymbol{Z}]$$
$$\sim N\left[\left(\sum_{i=1}^{m}\sum_{j=1}^{t}\frac{\boldsymbol{X}_{ij}\boldsymbol{X}_{ij}'}{\psi_j^2}\right)^{-1}\left(\sum_{i=1}^{m}\sum_{j=1}^{t}\frac{\boldsymbol{X}_{ij}}{\psi_j^2}(\theta_{ij} - \boldsymbol{Z}_{ij}'\boldsymbol{\gamma} - b_i)\right), \left(\sum_{i=1}^{m}\sum_{j=1}^{t}\frac{\boldsymbol{X}_{ij}\boldsymbol{X}_{ij}'}{\psi_j^2}\right)^{-1}\right].$$

The conditional density of $\boldsymbol{\gamma}$ is given by

$$\pi(\boldsymbol{\gamma}|\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{b},\boldsymbol{\psi}^2,\sigma_\gamma^2,\boldsymbol{X},\boldsymbol{Z}) \propto exp\left[-\frac{\left\{\boldsymbol{\gamma}'\left(\sum_{ij}\frac{Z_{ij}Z'_{ij}}{\psi_j^2}+\frac{1}{\sigma_\gamma^2}\boldsymbol{I}\right)\boldsymbol{\gamma}-2\boldsymbol{\gamma}'\sum_{ij}\frac{Z_{ij}(\theta_{ij}-X'_{ij}\boldsymbol{\beta}-b_i)}{\psi_j^2}\right\}}{2}\right].$$

Therefore the conditional density of $\boldsymbol{\gamma}$ is as below.

$$[\boldsymbol{\gamma}|\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{b},\boldsymbol{\psi}^2,\sigma_\gamma^2,\boldsymbol{X},\boldsymbol{Z}] \sim N\left[\left(\sum_{i,j}\frac{\boldsymbol{Z}_{ij}\boldsymbol{Z}'_{ij}}{\psi_j^2}+\frac{1}{\sigma_\gamma^2}\boldsymbol{I}\right)^{-1}\left(\sum_{i,j}\frac{\boldsymbol{Z}_{ij}}{\psi_j^2}(\theta_{ij}-\boldsymbol{X}'_{ij}\boldsymbol{\beta}-b_i)\right),\left(\sum_{i,j}\frac{\boldsymbol{Z}_{ij}\boldsymbol{Z}'_{ij}}{\psi_j^2}+\frac{1}{\sigma_\gamma^2}\boldsymbol{I}\right)^{-1}\right].$$

The conditional density of $(\sigma_\gamma^2)^{-1}$ is given by

$$\pi((\sigma_\gamma^2)^{-1}|\boldsymbol{\gamma}) \propto \left(\frac{1}{\sigma_\gamma^2}\right)^{\frac{K}{2}+c_r-1} exp\left(-\frac{\frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{\gamma}+d_r}{\sigma_\gamma^2}\right).$$

So the conditional density of $(\sigma_\gamma^2)^{-1}$ is as below.

$$[(\sigma_\gamma^2)^{-1}|\boldsymbol{\gamma}] \sim Gamma\left[\frac{K}{2}+c_\gamma,\frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{\gamma}+d_\gamma\right].$$

The conditional density of $(\psi_j^2)^{-1}$ is given by

$$\pi((\psi_j^2)^{-1}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\theta},\boldsymbol{b},\boldsymbol{X},\boldsymbol{Z}) \propto \left(\frac{1}{\psi_j^2}\right)^{\frac{m}{2}+c_j-1} exp\left(-\frac{\frac{1}{2}\sum_i(\theta_{ij}-X'_{ij}\boldsymbol{\beta}-Z'_{ij}\boldsymbol{\gamma}-b_i)^2+d_j}{\psi_j^2}\right).$$

Thus the conditional density of $(\psi_j^2)^{-1}$ is as below.

$$[(\psi_j^2)^{-1}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\theta},\boldsymbol{b},\boldsymbol{X},\boldsymbol{Z}] \sim Gamma\left[c_j+\frac{m}{2},\frac{1}{2}\sum_{i=1}^{m}(\theta_{ij}-\boldsymbol{X}'_{ij}\boldsymbol{\beta}-\boldsymbol{Z}'_{ij}\boldsymbol{\gamma}-b_i)^2+d_j\right].$$

Finally, the conditional density of $(\sigma_b^2)^{-1}$ is given by

$$\pi((\sigma_b^2)^{-1}|\boldsymbol{b}) \propto \left(\frac{1}{\sigma_b^2}\right)^{\frac{m}{2}+c-1} exp\left(-\frac{\frac{1}{2}\sum_i b_i^2+d}{\sigma_b^2}\right).$$

So the conditional density of $(\sigma_b^2)^{-1}$ is as below.

$$[(\sigma_b^2)^{-1}|\boldsymbol{b}] \sim Gamma\left[\frac{m}{2}+c,\frac{1}{2}\sum_{i=1}^{m}b_i^2+d\right].$$

Each conditional densities has standard distribution. So we can infer the parameters by using Gibbs sampler to sample from the full conditional relevant parameters.

3. Numerical studies

We applied the semiparametric models to analyze the median household income dataset. For this study, we have used IRS mean income as covariate. This is because it seems to possess an underlying nonlinear relationship with the CPS median income in Figure 1.1. Our dataset includes the median household income of all the U.S states and the District of Columbia for the year from 1995 to 1999. Our targets of inference are the state specific median household incomes for 1999. Estimates are compared to the corresponding census figures for 1999 since in small area estimation problems, the census estimates are often treated as gold standard against which all other estimates are compared.

To check the performance of our estimates, we use the following four criteria to compare the different estimates.

- Average Relative Bias (ARB) $= (51)^{-1}\sum_{i=1}^{51}\frac{|c_i-e_i|}{c_i}$
- Average Squared Relative Bias (ASRB) $= (51)^{-1}\sum_{i=1}^{51}\frac{|c_i-e_i|^2}{c_i^2}$
- Average Absolute Bias (AAB) $= (51)^{-1}\sum_{i=1}^{51}|c_i-e_i|$
- Average Squared Deviation (ASD) $= (51)^{-1}\sum_{i=1}^{51}(c_i-e_i)^2$

Here c_i and e_i respectively denote the census and model based estimates for the i^{th} state $(i = 1, ...51)$. The lower values of these measures would imply a better model based estimate.

We would like to compare the four different models such as TPBF with equal quantile knots (model 1), TPBF with equal distance knots (model 2), RBF with equal quantile knots (model 3) and RBF with equal distance (model 4). Equal quantile knots mean that the knots $(\tau_1, \cdots, \tau_K)$ are placed on a grid of equally spaced sample quantiles of $x'_{ij}s$. On the other hands, equal distance knots mean that the knots $(\tau_1, \cdots, \tau_K)$ are placed on a grid which has equal distance of range of $x'_{ij}s$. Figure 3.1 shows the exact location of knots. Figure 3.1 (a) depicts the equal quantile knots and Figure 3.1 (b) depicts the equal distance knots. The general models are identical to those in Section 2.1. Considering the pattern the data possess in Figure 3.1, we only choose linear coefficient (p=1) in mean structure.

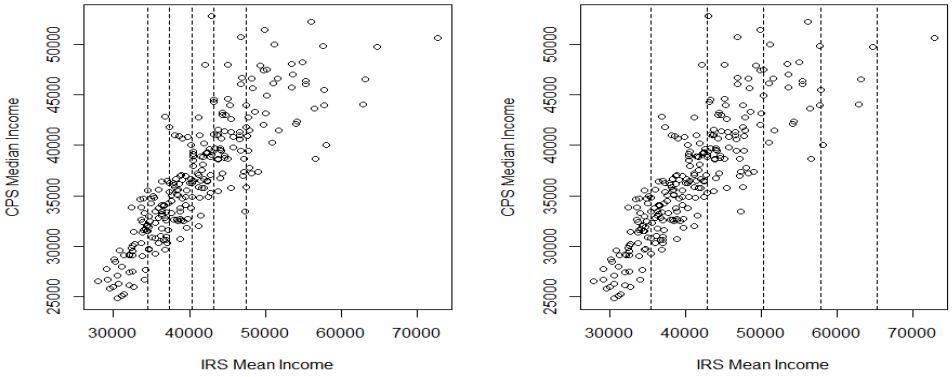

Figure 3.1 Knots location

CPS provides direct estimates which are drawn using only data for specific state and time period. We are interested in comparing our four models to the simple model without knots and CPS direct estimates. We monitor for the 10,000 iterations in each methods and discard 5,000 burn-in samples. In Table 3.1, the results of each models are shown. The value of ARB of the without knots model is 0.0362, model 1 is 0.0323 and the ASD value of model 4 is 3,614,617. The comparative measures for the model 1 is the lowest among these models.

Table 3.1 Comparative measures

model	ARB	ASRB	AAB	ASD
w/o knots	0.0362	0.0019	1509.84	3,629,741
model 1	0.0323	0.0016	1395.15	3,365,540
model 2	0.0326	0.0016	1409.34	3,522,738
model 3	0.0325	0.0016	1402.60	3,427,372
model 4	0.0330	0.0017	1429.58	3,614,617
CPS	0.0415	0.0027	1753.33	5,300,023

Table 3.2 describes the percentage improvement of the model 1 over the CPS and without knots model. It is obvious that the model 1 is better than the CPS estimation. Table 3.3 depicts the posterior mean, median and 95% CI for the model 1. The each 95% credible intervals for parameters $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5$ doesn't contain 0 indicating the significance of the knots.

Table 3.2 Percentage improvements with regard to model 1

method	ARB	ASRB	AAB	ASD
CPS	22.14%	38.89%	20.43%	36.50%
w/o knots	10.82%	17.09%	7.60%	7.28%

Table 3.3 Parameter estimates for model 1

Parameter	Mean	Median	95% CI
β_0	4908.50837	4908.50836	(4908.50730, 4908.50947)
β_1	0.8102296	0.8102296	(0.8102296, 0.8102297)
γ_1	-0.1806231	-0.1806231	(-0.1806233, -0.1806229)
γ_2	0.0212493	0.0212493	(0.0212488, 0.0212497)
γ_3	0.0868506	0.0868506	(0.0868501, 0.0868511)
γ_4	-0.1503352	-0.1503352	(-0.1503356, -0.1503347)
γ_5	-0.1816672	-0.1816672	(-0.1816675, -0.1816670)

We need to check the convergence of chains to guarantee adequate results. Figures 3.2-3.5 describe the convergence of each parameters. The trace plot is the plots of the iterations versus the generated values. If all values are within a zone without strong periodicities and tendencies, then we can assume convergence. Figure 3.2 and 3.4 are the trace plots for 5,000 iterations after discarding the first 5,000. It is shown that the generated sampled values are stabilized within a zone. Moreover, the ergodic mean refers to the mean value until the current iteration. If the ergodic mean is stabilized after some iterations, then this is an indication of the convergence of the algorithm. Figure 3.3 and 3.5 are ergodic mean plots for 10,000 iterations. The algorithm has reached convergence since the ergodic means have been stabilized.

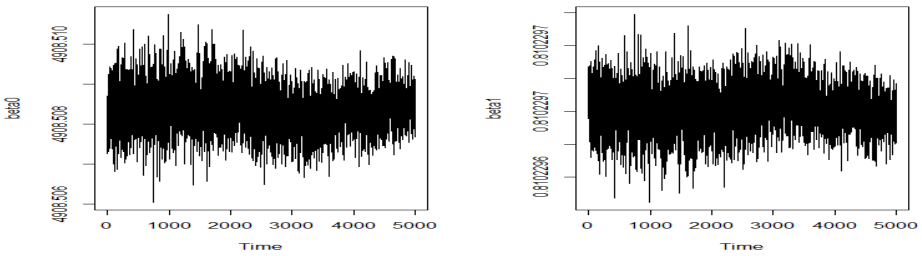

Figure 3.2 Trace plot of $\boldsymbol{\beta}$ for model 1

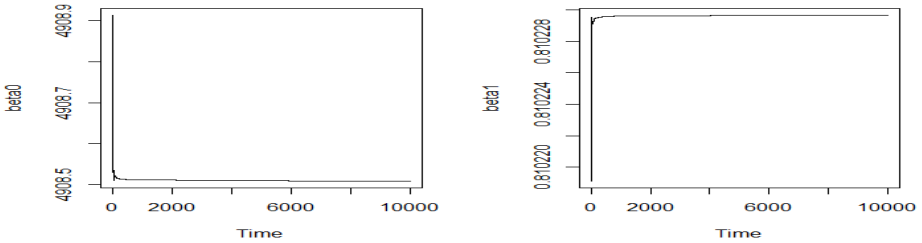

Figure 3.3 Ergodic mean plot of $\boldsymbol{\beta}$ for model 1

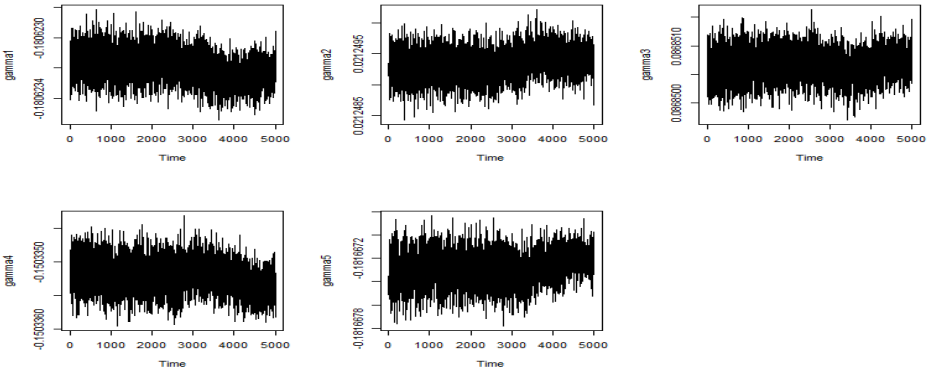

Figure 3.4 Trace plot of $\boldsymbol{\gamma}$ for model 1

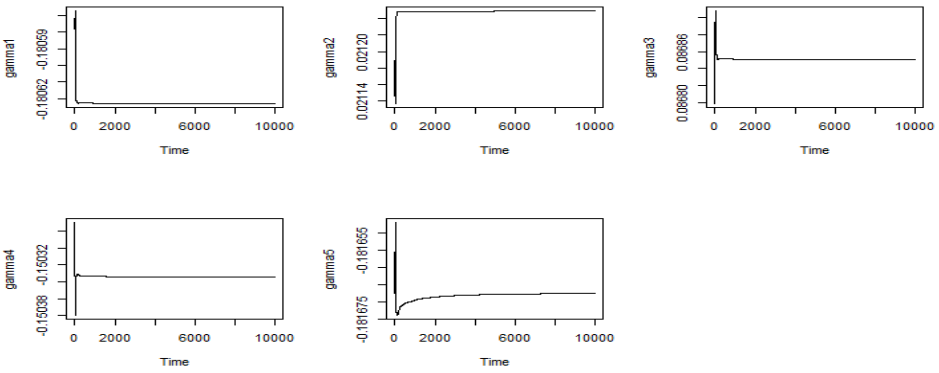

Figure 3.5 Ergodic mean plot of $\boldsymbol{\gamma}$ for model 1

Monitoring the Monte Carlo error (MC error) is another way to check convergence of the algorithm. Small MC errors indicate we have the quantity of interest with precision. We use the batch mean method to estimate MC error with 50 batches. MC errors of our interest parameters, θ_{ij} (2.3, 3.3, 0.7, $\cdots$), are considerably smaller than one third of the posterior standard deviation (32.6, 25.8, 24.7, $\cdots$) respectively.

4. Concluding remarks

The proper estimation of median household income for different small areas is one of the principal goals of the U.S Census Bureau. In this paper, we have proposed a semiparametric class of models which exploits the longitudinal trend in the state-specific income observations. In doing so, we have modeled the CPS median income observations as an income trajectory using P-splines. In detail, we have shown that estimation of the median household income for all the U.S states using different basis functions and locations of knots (truncated polynomial basis function, radial basis function, equal quantile knots and equal distance knots). Also, analysis has been carried out in a hierarchical Bayesian framework. As seen in Table 3.1, the model with knots works better than one without knots. The truncated polynomial basis function is slightly better than radial basis function. Similarly, comparative measures of models with the equal quantile knots are smaller than those of models with the equal distance knots.

The models in our study can be extended to those with other locations of knots and types of bases like B-splines, L-splines, etc. Although we used a parametric normal distributional assumption for the random state effect, a broader class of distributions like the Dirichlet process or Polya trees could be considered.

References

Bhadra, D., Ghosh, M. and Kim, D. (2012). Estimation of median household income for small areas: A Bayesian semiparametric approach. *Calcutta Statistical Association Bulletin*, **64**, 115-142.

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

Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: An application of James-Stein procedure to Census data. *Journal of the American Statistical Association*, **74**, 269-277.

Ghosh, M., Nangia, N. and Kim, D. H. (1996). Estimation of median income of four-person families : A bayesian time series approach. *Journal of the American Statistical Association*, **91**, 1423-1431.

Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal. *Statistical Science*, **9**, 55-76.

Lee, J. and Kim, D. H. (2015). Bayesian estimation of median household income for small area with some longitudinal pattern. *Journal of the Korean Data & Information Science Society*, **26**, 755-762.

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