Mathematical Theory and Modeling
ISSN (Paper)2224-5804 ISSN (Online)2225-0522
Vol 3, No. 12, 2013

A Note on Bayes semiparametric Regression

Ameera Jaber Mohaisen and Ammar Muslim Abdulhussain
Mathematics Department College of Education for Pure Science AL-Basrah University-Iraq
E-mail: ammar.muslim1@yahoo.com

Abstract

In the Bayesian approach to inference, all unknown quantities contained in a probability model for the observed data are treated as random variables. Specifically, the fixed but unknown parameters are viewed as random variables under the Bayesian approach. In this paper, Bayesian approach is employed to making inferences on the semiparametric regression model as mixed model , and we prove some theorems about posterior.

Keywords

Mixed models, Semiparametric regression, Penalized spline, Bayesian inference, Prior density, Posterior density.

1. Introduction

Consider the model:

$$y_i = \sum_{j=0}^{p} \beta_j x_{ji} + m(x_{p+1,i}) + \epsilon_i \qquad , i = 1,2, \dots, n \tag{1}$$

Where $y_1 , \dots , y_n$ response variables and the unobserved errors are $\epsilon_1 , \dots , \epsilon_n$ are known to be i.i.d. normal with mean 0 and covariance $\sigma_\epsilon^2 I$ with σ_ϵ^2 unknown.

The mean function of the regression model in (1) has two parts. The parametric (first part) is assumed to be linear function of p-dimensional covariates x_{ji} and nonparametric (second part) $m(x_{p+1,i})$ is function defined on some index set $T \subset R^1$. Inferences a bout model (1) such as its estimation as well as model checking are of interest.

A Bayesian approach to (fully) semiparametric regression problems typically requires specifying prior distributions on function spaces which is rather difficult to handle. The extent of the complexity of this approach can be gauged from sources such as Angers and Delampady [1], and Lenk [7], and so on.

In this paper, a simple Bayesian approach to semiparametric regression. By using penalized spline for the nonparametric function (second part) of the model (1) we can representation semiparametric regression model (1) as mixed model and Bayesian approach is employed to making inferences on the resulting mixed model coefficients, and we prove some theorems about posterior.

2. Mixed Models

The general form of a linear mixed model for the ith subject (i = 1,…, n) is given as follows [9],[12],[13],

$$Y_i = X_i \beta + \sum_{j=1}^{r} Z_{ij} u_{ij} + \epsilon_i \quad , \quad u_{ij} \sim N(0, G_j) \, , \; \epsilon_i \sim N(0, R_i), \tag{2}$$

where the vector Y_i has length m_i, X_i and Z_{ij} are, respectively, a $m_i \times p$ design matrix and a $m_i \times q_i$ design matrix of fixed and random effects. β is a p-vector of fixed effects and u_{ij} are the q_i-vectors of random effects. The variance matrix G_j is a $q_i \times q_i$ matrix and R_i is a $m_i \times m_i$ matrix.

We assume that the random effects $\{ u_{ij} \, ; \, i = 1, \dots , n \, ; \, j = 1, \dots , r\}$ and the set of error terms $\{\epsilon_1, \dots , \epsilon_n\}$ are independent. In matrix notation,

$$Y = X\beta + Zu + \epsilon \, . \tag{3}$$

Here $Y = (Y_1, \dots , Y_n)^T$ has length $N = \sum_{i=1}^{n} m_i$, $X = (X_1^T, \dots , X_n^T)^T$ is a $N \times p$ design matrix of fixed effects, Z is a $N \times q$ block diagonal design matrix of random effects, $q = \sum_{j=1}^{r} q_j$, $u = (u_1^T, \dots , u_r^T)^T$ is a q-vector of random effects, $R = diag(R_1, \dots , R_n)$ is a $N \times N$ matrix and $G = diag(G_1, \dots , G_r)$ is a $q \times q$ block diagonal matrix.

3. Spline Semiparametric regression and Prior

The model (1) can be expressed as a smooth penalized spline with q degree, then it's become as (see [12]):

$$y_i = \sum_{j=0}^{p} \beta_j x_{ji} + \sum_{j=1}^{q} \beta_{p+j} x_{p+1,i}^{j} + \sum_{k=1}^{K} \{ u_k (x_{p+1,i} - k_k)_+^q + \epsilon_i \, , \tag{4}$$

where $k_1, \dots, k_K$ are inner knots $a < k_1 <,,,< k_K < b$.

By using a convenient connection between penalized splines and mixed models. Model (4) is rewritten as follows (see [9,12,13])

$$Y = X\beta + Zu + \epsilon \tag{5}$$

where

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \\ \beta_{p+1} \\ \vdots \\ \beta_{p+q} \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \vdots \\ u_K \end{bmatrix}, Z = \begin{bmatrix} (x_{p+1,1} - k_1)_+^q & \cdots & (x_{p+1,1} - k_K)_+^q \\ \vdots & \ddots & \vdots \\ (x_{p+1,n} - k_1)_+^q & \cdots & (x_{p+1,n} - k_K)_+^q \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{p1} & x_{p+1,1} & \dots & x_{p+1,1}^q \\ 1 & x_{12} & \dots & x_{p2} & x_{p+1,2} & \dots & x_{p+1,2}^q \\ \vdots & \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \dots & x_{pn} & x_{p+1,n} & \dots & x_{p+1,n}^q \end{bmatrix}$$

We assume that the function g is:

$$\mathrm{g} = X\beta + Zu \tag{6}$$

And its prior guess g^o can be written as:

$$\mathrm{g}^o = X\beta \tag{7}$$

Further, some of the a priori information penalized spline coefficients can be translated into:

$$\begin{aligned} E(\epsilon) &= 0; & var(\epsilon) &= \sigma_\epsilon^2 I \\ E(\beta) &= 0; & var(\beta) &= \sigma_\beta^2 I \\ E(u) &= 0; & var(u) &= \sigma_u^2 I \end{aligned} \tag{8}$$

The term $X\beta$ in (5) is the pure polynomial component of the spline, and Zu is the component with spline truncated functions with covariance $\sigma_u^2\, Q$, where $Q = ZZ^T$. Letting $(\beta, u, \sigma_u^2, \sigma_\epsilon^2)$ be the parameter vector, the mixed model specifies a $N(0, \sigma_u^2 I)$ prior on u as well as the likelihood, $f(Y|\beta, u, \sigma_u^2, \sigma_\epsilon^2)$. To specify a complete Bayesian model, we also need a prior distribution on $(\beta, \sigma_u^2, \sigma_\epsilon^2)$. Assuming that little is known about β, it makes sense to put an improper uniform prior on β. Or, if a proper prior is desired, one could use a $N(0, \sigma_\beta^2 I)$ prior with σ_β^2 so large that, for all intents and purposes, the normal distribution is uniform on the range of β. Therefore, we will use $\pi_0(\beta) \equiv 1$. We will assume that the prior on σ_ϵ^2 is inverse gamma with parameters A_ϵ and B_ϵ – denoted $IG(A_\epsilon, B_\epsilon)$ – so that its density is

$$\pi_0(\sigma_\epsilon^2) = \frac{B_\epsilon^{A_\epsilon}}{\Gamma(A_\epsilon)} (\sigma_\epsilon^2)^{-(A_\epsilon+1)} exp\left(-\frac{B_\epsilon}{\sigma_\epsilon^2}\right) \tag{9}$$

Also, we assume that:

$$\sigma_u^2 \sim IG(A_u, B_u)$$

Here $A_\epsilon, B_\epsilon, A_u$ and B_u are “hyperparameters” that determine the priors and must be chosen by the statistician. These hyperparameters must be strictly positive in order for the priors to be proper. If A_ϵ and B_ϵ were zero, then $\pi_0(\sigma_\epsilon^2)$ would be proportional to the improper prior $\frac{1}{\sigma_\epsilon^2}$, which is equivalent to $\log(\sigma_\epsilon)$ having an improper uniform prior. Therefore, choosing A_ϵ and B_ϵ both close to zero (say, both equal to 0.1) gives an essentially noninformative, but proper, prior. The same reasoning applies to A_u and B_u. The model we have constructed is a hierarchical Bayes model, where the random variables are arranged in a hierarchy such that distributions at each level are determined by the random variables in the previous levels. At the bottom of the hierarchy are the known hyperparameters. At the next level are the fixed effects parameters and variance components whose distributions are determined by the hyperparameters. At the level above this are the random effects, u and ϵ, whose distributions are determined by the variance components. The top level contains the data, y. (see [13])

4. Posterior calculations

We have the model

$$Y|F, \sigma_u^2, \sigma_\epsilon^2 \sim N(CF, \sigma_\epsilon^2 I_n + \sigma_u^2 Q). \tag{10}$$

where $C = [X \;\; Z]$.

Unless F has a normal prior distribution or a hierarchical prior with a conditionally normal prior distribution, analytical simplifications in the computation of posterior quantities are not expected. For such cases, we have the joint posterior density of the penalized spline coefficients F and the error variances σ_ϵ^2 and σ_u^2 given by the expression.

$$\pi(F, \sigma_u^2, \sigma_\epsilon^2|Y) \propto f(Y|F, \sigma_u^2, \sigma_\epsilon^2)\, \pi_0(F, \sigma_u^2, \sigma_\epsilon^2)$$

Where f is the likelihood. From (10), f can be expressed as

$$f(Y|F, \sigma_u^2, \sigma_\epsilon^2) \propto |\sigma_\epsilon^2 I_n + \sigma_u^2 Q|^{-1/2} exp\{\frac{-1}{2}(Y - CF)^T (\sigma_\epsilon^2 I_n + \sigma_u^2 Q)^{-1}(Y - CF)\}$$

Proceeding further, suppose π_0 of the form

$$\pi_0(F, \sigma_u^2, \sigma_\epsilon^2) = \pi_1(\sigma_u^2, \sigma_\epsilon^2) \tag{11}$$

which is constant in F, is chosen.

Markov Chain Monte Carlo (MCMC) based approaches to posterior computations are now readily available. For example, Gibbs sampling is straightforward (see [1,13]).

Note that $Q = CDC^T = ZZ^T$

where $D = \begin{bmatrix} 0_{p+q+1} & 0 \\ 0 & I_{n-(p+q+1)} \end{bmatrix}$

and $\sigma_u^2 D = \begin{bmatrix} 0_{p+q+1} & 0 \\ 0 & \sigma_u^2 I_{n-(p+q+1)} \end{bmatrix}$

we see

$$Y|F, \sigma_u^2, \sigma_\epsilon^2 \sim N(CF, \sigma_\epsilon^2 I_n + \sigma_u^2 Q) \tag{12}$$

However, the prior of F given σ_u^2 specified that $F|\sigma_u^2 \sim N(0, \sigma_u^2 D)$

Therefore, it follows that

$$Y|\sigma_u^2, \sigma_\epsilon^2 \sim N(0, \sigma_\epsilon^2 I_n + C\sigma_u^2 DC^T) \tag{13}$$

where $\sigma_u^2 Q = C\sigma_u^2 DC^T$

$$F|Y, \sigma_u^2, \sigma_\epsilon^2 \sim N(A_1 Y, A_2) \tag{14}$$

where

$$A_1 = \sigma_u^2 D\, C^T (\sigma_\epsilon^2 I_n + C\sigma_u^2 DC^T)^{-1} \tag{15}$$

$$A_2 = \sigma_u^2 D - \sigma_u^4 D\, C^T (\sigma_\epsilon^2 I_n + C\sigma_u^2 DC^T)^{-1} CD \tag{16}$$

We can rewrite covariance of Y given F, σ_u^2 and σ_ϵ^2 as

$$\sigma_\epsilon^2 I_n + \sigma_u^2 Q = \sigma_\epsilon^2 I_n + C\sigma_u^2 DC^T = C\sigma_\epsilon^2 \left(C^{-1} I_n C^{T^{-1}} + \frac{\sigma_u^2}{\sigma_\epsilon^2} D \right) C^T$$

$$= C\sigma_\epsilon^2 \left(C^T C + \frac{\sigma_\epsilon^2}{\sigma_u^2} D \right)^{-1} C^T, \text{ where } D^{-1} = D.$$

Result 1: $F|Y, \sigma_u^2, \sigma_\epsilon^2 \sim N\left\{\left(C^T C + \frac{\sigma_\epsilon^2}{\sigma_u^2} D\right)^{-1} C^T Y, \sigma_\epsilon^2 \left(C^T C + \frac{\sigma_\epsilon^2}{\sigma_u^2} D\right)^{-1}\right\}$. (17)

Proof:

Since $E(F|Y) = A_1 Y$

$$= \sigma_u^2 D\, C^T (\sigma_\epsilon^2 I_n + C\sigma_u^2 DC^T)^{-1} Y$$

$$= \sigma_u^2 D\, C^T \left(C\sigma_\epsilon^2 \left(C^T C + \frac{\sigma_\epsilon^2}{\sigma_u^2} D \right)^{-1} C^T \right)^{-1} Y$$

$$= \sigma_u^2 D\, C^T \left(C^{T^{-1}} \left(C^T C + \frac{\sigma_\epsilon^2}{\sigma_u^2} D \right) \frac{C^{-1}}{\sigma_\epsilon^2} \right) Y$$

$$= \frac{\sigma_u^2}{\sigma_\epsilon^2} D \left(C^T C^{T^{-1}} C^T C C^{-1} + C^T C^{T^{-1}} \frac{\sigma_\epsilon^2}{\sigma_u^2} D C^{-1} \right) C^{T^{-1}} C^T Y$$

$$= \frac{\sigma_u^2}{\sigma_\epsilon^2} D \left(C^T C^{T^{-1}} + \frac{\sigma_\epsilon^2}{\sigma_u^2} D C^{-1} C^{T^{-1}} \right) C^{T^{-1}} C^T Y$$

$$= \frac{\sigma_u^2}{\sigma_\epsilon^2} D \left(I_n + \frac{\sigma_\epsilon^2}{\sigma_u^2} D C^{-1} C^{T^{-1}} \right) C^T Y$$

$$= \frac{\sigma_u^2}{\sigma_\epsilon^2} D \left(\frac{\sigma_\epsilon^2}{\sigma_u^2} D C^{-1} C^{T^{-1}} + I_n \right) C^T Y$$

$$= \left(C^{-1} C^{T^{-1}} + \frac{\sigma_u^2}{\sigma_\epsilon^2} D \right) C^T Y$$

$$= \left(C^T C + \frac{\sigma_\epsilon^2}{\sigma_u^2} D \right)^{-1} C^T Y$$

By same way we can prove the covariance is $\sigma_\epsilon^2\left(C^TC+\frac{\sigma_\epsilon^2}{\sigma_u^2}D\right)^{-1}$.

Now proceeding as in [3], we employ spectral decomposition to obtain $CDC^T = BHB^T$, where $H = \text{diag}(h_1,\dots,h_n)$ is the matrix of eigenvalues and B is the orthogonal matrix of eigenvectors. Thus,

$$\sigma_\epsilon^2 I_n + [C\sigma_u^2DC^T] = \sigma_\epsilon^2 I_n + B\sigma_u^2HB^T = B\sigma_\epsilon^2 I_nB^T + B\sigma_u^2HB^T = B\sigma_\epsilon^2\left(I_n + \frac{\sigma_u^2}{\sigma_\epsilon^2}H\right)B^T$$
$$= \sigma_\epsilon^2B(I_n+\delta H)B^T$$

where $\delta = \sigma_u^2/\sigma_\epsilon^2$. Then, the first stage (conditional) marginal density of Y given σ_ϵ^2 and δ can be written as

$$m(Y|\sigma_\epsilon^2,\delta) = \frac{1}{(2\pi\sigma_\epsilon^2)^{n/2}}\frac{1}{\det[I_n+\delta H]^{1/2}}\exp\{-\frac{1}{2\sigma_\epsilon^2}Y^TB(I_n+\delta H)B^TY$$
$$= \frac{1}{(2\pi\sigma_\epsilon^2)^{n/2}}\frac{1}{[\prod_{i=1}^n[1+\delta h_i]]^{1/2}}\exp\{-\frac{1}{2\sigma_\epsilon^2}(\sum_{i=1}^n\frac{s_i^2}{1+\delta h_i})\}, \quad (18)$$

where $s = (s_1,\dots,s_n)^T = B^TY$. We choose the prior on σ_ϵ^2, $\delta = \sigma_u^2/\sigma_\epsilon^2$, qualitatively similar to the used in [1]. Specifically, we take $\pi_1(\sigma_\epsilon^2,\delta)$ to be proportional to the product of an inverse gamma density $\{B_\epsilon{}^{A_\epsilon}/\Gamma(A_\epsilon)\}\,exp(-B_\epsilon/\sigma_\epsilon^2)(\sigma_\epsilon^2)^{-(A_\epsilon+1)}$ for σ_ϵ^2 and the density of a $F(b,a)$ distribution for δ (for suitable choice of B_ϵ, A_ϵ, b and a). Conditions apply on a and b such that (see [1]):

1- The prior covariance of $\delta(=\frac{2b^2(a+b-2)}{a(b-4)(b-2)^2})$ is infinite.
2- The fisher information number $= (\frac{a^2(b+2)(b+6)}{2(a-4)(a+b+2)})$ is minimum.
3- The prior mode $= (\frac{b(a-2)}{a(b+2)})$ is greater than 0.

This can be done by choosing $2 < b \le 4$ and $a = 8(b+2)/(b-2)$

Once $\pi_1(\sigma_\epsilon^2,\delta)$ is chosen as above, we obtain the posterior density of δ given Y, the posterior mean and covariance matrix of F as in the following theorems.

Theorem1: the posterior density of δ given Y is:

$$\pi_{22}(\delta|Y) \propto \frac{\delta^{(b/2)-1}}{(a+b\delta)^{-(a+b)/2}}\left(\prod_{i=1}^n(1+\delta h_i)\right)^{-1/2}\left(2\text{B}_\epsilon+\sum_{i=1}^n\frac{s_i^2}{1+\delta h_i}\right)^{-(\text{n}+2\text{A}_\epsilon+2)/2} \quad (19)$$

Proof:

$$\pi_{22}(\delta|Y) = \int m(Y|\sigma_\epsilon^2,\delta)\ f(\delta,b,a)\ f(\sigma_\epsilon^2,A_\epsilon,B_\epsilon)d\sigma_\epsilon^2$$
$$= \int\frac{1}{(2\pi\sigma_\epsilon^2)^{n/2}}\left(\prod_{i=1}^n(1+\delta h_i)\right)^{-1/2}\frac{b^{b/2}a^{a/2}}{B(b,a)}\frac{\delta^{(b/2)-1}}{(a+b\delta)^{-(a+b)/2}}\exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(\sum_{i=1}^n\frac{s_i^2}{1+\delta h_i}\right)\right\}$$
$$\frac{\text{B}_\epsilon^{\text{A}_\epsilon}}{\Gamma(\text{A}_\epsilon)}(\sigma_\epsilon^2)^{-(\text{A}_\epsilon+1)}\exp\left(-\frac{\text{B}_\epsilon}{\sigma_\epsilon^2}\right)d\sigma_\epsilon^2$$
$$= \frac{(2\pi)^{-\frac{n}{2}}}{\Gamma(\text{A}_\epsilon)}\frac{b^{b/2}a^{a/2}}{B(b,a)}\frac{\delta^{(b/2)-1}}{(a+b\delta)^{-(a+b)/2}}\int\left(\prod_{i=1}^n(1+\delta h_i)\right)^{-1/2}\exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(2\text{B}_\epsilon+\sum_{i=1}^n\frac{s_i^2}{1+\delta h_i}\right)\right\}$$
$$(\sigma_\epsilon^2)^{-(\text{n}+2\text{A}_\epsilon+2)/2}d\sigma_\epsilon^2$$
$$= \frac{(2\pi)^{-n/2}}{\Gamma(\text{A}_\epsilon)}\frac{b^{b/2}a^{a/2}}{B(b,a)}\frac{\delta^{(b/2)-1}}{(a+b\delta)^{-(a+b)/2}}(2)^{(\text{n}+2\text{A}_\epsilon+2)/2}\int\left(\prod_{i=1}^n(1+\delta h_i)\right)^{-1/2}\exp\left\{-\frac{\left(2\text{B}_\epsilon+\sum_{i=1}^n\frac{s_i^2}{1+\delta h_i}\right)}{2\sigma_\epsilon^2}\right\}$$
$$\left(\frac{2\text{B}_\epsilon+\sum_{i=1}^n\frac{s_i^2}{1+\delta h_i}}{2\sigma_\epsilon^2}\right)^{(\text{n}+2\text{A}_\epsilon+2)/2}\left(2\text{B}_\epsilon+\sum_{i=1}^n\frac{s_i^2}{1+hd_i}\right)^{-(\text{n}+2\text{A}_\epsilon+2)/2}d\sigma_\epsilon^2$$
$$\propto \frac{\delta^{(b/2)-1}}{(a+b\delta)^{-(a+b)/2}}\int\left(\prod_{i=1}^n(1+\delta h_i)\right)^{-1/2}\exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(2\text{B}_\epsilon+\sum_{i=1}^n\frac{s_i^2}{1+\delta h_i}\right)\right\}$$
$$\left(\frac{2\text{B}_\epsilon+\sum_{i=1}^n\frac{s_i^2}{1+\delta h_i}}{2\sigma_\epsilon^2}\right)^{[(\text{n}+2\text{A}_\epsilon+4)/2]-1}\left(2\text{B}_\epsilon+\sum_{i=1}^n\frac{s_i^2}{1+\delta h_i}\right)^{-(\text{n}+2\text{A}_\epsilon+2)/2}d\sigma_\epsilon^2$$
$$\propto \frac{\delta^{(b/2)-1}}{(a+b\delta)^{-(a+b)/2}}\Gamma((\text{n}+2\text{A}_\epsilon+4)/2)\left(2\text{B}_\epsilon+\sum_{i=1}^n\frac{s_i^2}{1+\delta h_i}\right)^{-(\text{n}+2\text{A}_\epsilon+2)/2}\left(\prod_{i=1}^n(1+\delta h_i)\right)^{-1/2}$$
$$\propto \frac{\delta^{(b/2)-1}}{(a+b\delta)^{-(a+b)/2}}\left(\prod_{i=1}^n(1+\delta h_i)\right)^{-1/2}\left(2\text{B}_\epsilon+\sum_{i=1}^n\frac{s_i^2}{1+\delta h_i}\right)^{-(\text{n}+2\text{A}_\epsilon+2)/2}$$

Theorem2: The posterior mean and covariance matrix of F are:

$$E(F|Y) = DC^T BE\{(I_n + \delta H)^{-1} \mid Y\}s \tag{20}$$

And

$$var(F|Y) = \frac{1}{n+2A_\epsilon+2} E\left[\left(2B_\epsilon + \left(\sum_{i=1}^n \frac{s_i^2}{1+\delta h_i}\right)\right)|Y\right] D - \frac{1}{n+2A_\epsilon+2} HC^T BE\left[\left(2B_\epsilon + \left(\sum_{i=1}^n \frac{s_i^2}{1+\delta h_i}\right)\right)[I_n + \delta H]^{-1}|Y\right] B^T CD + E[R(\delta)R(\delta)^T|Y], \tag{21}$$

where $R(\delta) = DC^T B(I_n + \delta H)^{-1}s$

Proof:

From (14):

$$\begin{aligned} E(F|Y) &= A_1 Y \\ &= \sigma_u^2 D\, C^T\, (\sigma_\epsilon^2 I_n + C\sigma_u^2 DC^T)^{-1} Y \\ &= \sigma_u^2 DC^T\, \{\sigma_\epsilon^2 B(I_n + \delta H)B^T\}^{-1} Y \\ &= \frac{\sigma_u^2}{\sigma_\epsilon^2} DC^T B^{T^{-1}} (I_n + \delta H)^{-1}\, B^{-1} Y \end{aligned}$$

Since B is the orthogonal matrix of eigenvectors, then $B^{-1} = B^T$ and $B^{T^{-1}} = B$.

Therefore

$$\begin{aligned} E(F|Y) &= DC^T B\, \delta\, (I_n + \delta H)^{-1} B^T Y \\ &= DC^T B\, E((I_n + \delta H)^{-1}|Y)s, \end{aligned}$$

where the expectation $E((I_n + \delta H)^{-1}|Y)$ is taken with respect to $\pi_{22}(\delta|Y)$ (see theorem 1 above). And by the same way we can prove the variance of F given Y.

5. Model checking and Bayes factors

An important and useful model checking problem in the present setup is checking the two models

$$H_o : \mathrm{g} = X\beta = \mathrm{g}^o \text{ versus } H_1 : \mathrm{g} = X\beta + Zu \neq \mathrm{g}^o.$$

Under H_1, ($\mathrm{g} = \mathrm{g}(F), \sigma_u^2, \sigma_\epsilon^2$) is given the prior $\pi_0(F, \sigma_u^2, \sigma_\epsilon^2) I(\mathrm{g} \neq \mathrm{g}^o)$, whereas under H_o, $\pi_0(\sigma_\epsilon^2)$ induced by $\pi_0(F, \sigma_u^2, \sigma_\epsilon^2)$ is the only part needed. In order to conduct the model checking, we compute the Bayes factor, B_{01}, of H_o relative to H_1:

$$B_{01}(Y) = \frac{m(Y|H_O)}{m(Y|H_1)} \tag{22}$$

where $m(Y|H_i)$ is the predictive (marginal) density of Y under model $H_i, i = 0, 1$. We have

$$m(Y|H_o) = \int f(Y|\mathrm{g}^o, \sigma_\epsilon^2)\pi_0(\sigma_\epsilon^2)\; d\sigma_\epsilon^2$$

and

$$m(Y|H_1) = \int f(Y|F, \sigma_u^2, \sigma_\epsilon^2)\pi_0(F, \sigma_u^2, \sigma_\epsilon^2)\; dF d\sigma_u^2 d\sigma_\epsilon^2$$

As in the previous section $\pi_0(\sigma_u^2, \sigma_\epsilon^2)$ will be constant in F, while σ_ϵ^2 is inverse gamma and is independent of $v_1 = \sigma_\epsilon^2/\sigma_u^2$ which is given the $F_{a,b}$ prior distribution. (Equivalently, $\delta = \sigma_u^2/\sigma_\epsilon^2$ is given the $F_{b,a}$, Specifically, $\pi_0(\sigma_\epsilon^2) = \frac{B_\epsilon^{A_\epsilon}}{\Gamma(A_\epsilon)} (\sigma_\epsilon^2)^{-(A_\epsilon+1)} exp\left(-\frac{B_\epsilon}{\sigma_\epsilon^2}\right)$, where A_ϵ and B_ϵ (small) are suitably chosen. Therefore,

$$\begin{aligned} m(Y|H_o) &= \int f(Y|\mathrm{g}^o, \sigma_\epsilon^2)\, \pi_0(\sigma_\epsilon^2)\; d\sigma_\epsilon^2 \\ &= (2\pi)^{-n/2} \frac{B_\epsilon^{A_\epsilon}}{\Gamma(A_\epsilon)} \int (\sigma_\epsilon^2)^{-n/2} exp\left(-\frac{B_\epsilon}{\sigma_\epsilon^2}\right) (\sigma_\epsilon^2)^{-(A_\epsilon+1)}\; exp\left(-\frac{(Y-g^o(x))^2}{2\sigma_\epsilon^2}\right) d\sigma_\epsilon^2 \\ &= (2\pi)^{-n/2} \frac{B_\epsilon^{A_\epsilon}}{\Gamma(A_\epsilon)} \int (\sigma_\epsilon^2)^{-(n/2+A_\epsilon+1)} exp\left(-\frac{B_\epsilon + \frac{1}{2}(y_i - g^o(x_i))^2}{\sigma_\epsilon^2}\right) d\sigma_\epsilon^2 \\ &= (2\pi)^{-\frac{n}{2}} \frac{\mathrm{B}_\epsilon^{\mathrm{A}_\epsilon}}{\Gamma(\mathrm{A}_\epsilon)} \int (\sigma_\epsilon^2)^{-\left(\frac{n}{2}+\mathrm{A}_\epsilon+1\right)} (\mathrm{B}_\epsilon + \tfrac{1}{2}(y_i - \mathrm{g}^o(x_i))^2)^{\frac{n}{2}+\mathrm{A}_\epsilon+1} (\mathrm{B}_\epsilon + \tfrac{1}{2}(y_i - \mathrm{g}^o(x_i))^2)^{-\left(\frac{n}{2}+\mathrm{A}_\epsilon+1\right)} \\ &\quad \exp\left(-\frac{\mathrm{B}_\epsilon + \frac{1}{2}(y_i - \mathrm{g}^o(x_i))^2}{\sigma_\epsilon^2}\right) d\sigma_\epsilon^2 \\ &= (2\pi)^{-n/2} \frac{\mathrm{B}_\epsilon^{\mathrm{A}_\epsilon}}{\Gamma(\mathrm{A}_\epsilon)} \int \frac{(\mathrm{B}_\epsilon + \frac{1}{2}(y_i - \mathrm{g}^o(x_i))^2)^{\frac{n}{2}+\mathrm{A}_\epsilon+1}}{(\sigma_\epsilon^2)^{\left(\frac{n}{2}+\mathrm{A}_\epsilon+1\right)}} \exp\left(-\frac{\mathrm{B}_\epsilon + \frac{1}{2}(y_i - \mathrm{g}^o(x_i))^2}{\sigma_\epsilon^2}\right) \\ &\quad (\mathrm{B}_\epsilon + \tfrac{1}{2}(y_i - \mathrm{g}^o(x_i))^2)^{-\left(\frac{n}{2}+\mathrm{A}_\epsilon+1\right)}\; d\sigma_\epsilon^2 \end{aligned}$$

$$= (2\pi)^{-n/2} \frac{B_\epsilon^{A_\epsilon}}{\Gamma(A_\epsilon)} \int \left(\frac{B_\epsilon + \frac{1}{2}(y_i - g^o(x_i))^2}{\sigma_\epsilon^2} \right)^{(\frac{n}{2}+A_\epsilon+2)-1} \exp\left(-\frac{B_\epsilon + \frac{1}{2}(y_i - g^o(x_i))^2}{\sigma_\epsilon^2} \right) (B_\epsilon + \frac{1}{2}(y_i - g^o(x_i))^2)^{-(\frac{n}{2}+A_\epsilon+1)} \, d\sigma_\epsilon^2$$

$$= (2\pi)^{-n/2} \frac{B_\epsilon^{A_\epsilon}}{\Gamma(A_\epsilon)} \Gamma(\frac{n}{2} + A_\epsilon + 1)(B_\epsilon + \frac{1}{2}(y_i - g^o(x_i))^2)^{-(\frac{n}{2}+A_\epsilon+1)} \, d\sigma_\epsilon^2 \quad (23)$$

Further, using (11) it follows that:

$$m(Y|H_1, \sigma_\epsilon^2, \delta) = (2\pi\sigma_\epsilon^2)^{-\frac{n}{2}} (\prod_{i=1}^n (1 + \delta h_i))^{-1/2} \exp\left\{ -\frac{1}{2\sigma_\epsilon^2} \left(\sum_{i=1}^n \frac{s_i^2}{1+\delta h_i} \right) \right\} \quad (24)$$

Therefore,

$$m(Y|H_1) = \int m(Y|M_1, \sigma_\epsilon^2, \delta)\, \pi_0(\sigma_\epsilon^2, \delta)\, d\sigma_\epsilon^2\, d\delta$$

$$= \int \frac{B_\epsilon^{A_\epsilon}}{\Gamma(A_\epsilon)} (\sigma_\epsilon^2)^{-(A_\epsilon+1)} \exp\left(-\frac{B_\epsilon}{\sigma_\epsilon^2} \right) (2\pi\sigma_\epsilon^2)^{-n/2} \left(\prod_{i=1}^n (1 + \delta h_i) \right)^{-1/2} \exp\left\{ -\frac{1}{2\sigma_\epsilon^2} \left(\sum_{i=1}^n \frac{s_i^2}{1+hd_i} \right) \right\} \pi_0(\delta)\, d\sigma_\epsilon^2\, d\delta$$

$$= \frac{B_\epsilon^{A_\epsilon}}{\Gamma(A_\epsilon)} (2\pi)^{-n/2} \int \left(\prod_{i=1}^n (1 + \delta h_i) \right)^{-1/2} \pi_0(\delta) \left\{ \int \exp\left\{ -\frac{1}{\sigma_\epsilon^2} \left(B_\epsilon + \frac{1}{2} \sum_{i=1}^n \frac{s_i^2}{1+\delta h_i} \right) \right\} d\sigma_\epsilon^2 \right\}$$

$$= \frac{B_\epsilon^{A_\epsilon}}{\Gamma(A_\epsilon)} (2\pi)^{-n/2} \Gamma(n/2 + A_\epsilon) \int \left(\prod_{i=1}^n (1 + \delta h_i) \right)^{-1/2} \left(B_\epsilon + \frac{1}{2} \sum_{i=1}^n \frac{s_i^2}{1+\delta h_i} \right)^{-(n/2+c-1)} \pi_0(\delta)\, d\delta \quad (25)$$

6- Simulation results

In this section, we illustrate the effectiveness of the our methodology. We generated observations from the model (1) with the following regression functions which represent a variety of shapes:

(i) $y_1 = 1 - 3x_1 + e^{cos(\pi x_2 + 2x_2)}$, (26)

(ii) $y_2 = 2x_1 - sin(2\pi x_2) + 0.3(x_2 - 0.75)^2 - \frac{1}{2} x_2^3$. (27)

The settings for the simulation study are as follows. The observations for the design variable are generated from uniform distribution on the interval [-1,1], for various sample sizes. These values are kept fixed for all settings to reduce simulation variability. The sample size taken is n=150.

For the error distribution we used normal distribution $N(0, \sigma_\epsilon^2)$, where $\sigma = 0.125$, 0.25 and 0.5. We have tried with different choices of K as well. The penalty parameter λ is chosen by minimizing the generalized cross validation (GCV) criterion.

To give an impression on the variability of the obtained estimators, we plot in figure (1) a scatter plot of the randomly generated data sets together with the fitted values from the penalized LS. regression spline estimation method. Table (1) presents summary values of the (*AMSE*) and (*AMAE*) for the estimation method. From this table we can see that the values of (*AMSE*) and (*AMAE*) when ($\sigma = 0.5$) are smaller than that ($\sigma = 0.125$ and 0.25), which were (0.01656268) and (0.008447995) respectively. While the values of (*AMSE*) and (*AMAE*) are smaller when ($\sigma = 0.125$) and ($\sigma = 0.5$) respectively for second test function were (0.009150507) and (0.004085605) respectively.

Table (1) result of the AMSE and AMAE for Bayesian semiparametric regression

Test function	σ	AMSE	AMAE
y_1	0.125	0.04299587	0.01400477
	0.25	0.03975184	0.01345516
	0.5	0.01656268	0.008447995
y_2	0.125	0.009150507	0.00641508
	0.25	0.009218403	0.006443437
	0.5	0.009684469	0.004085605

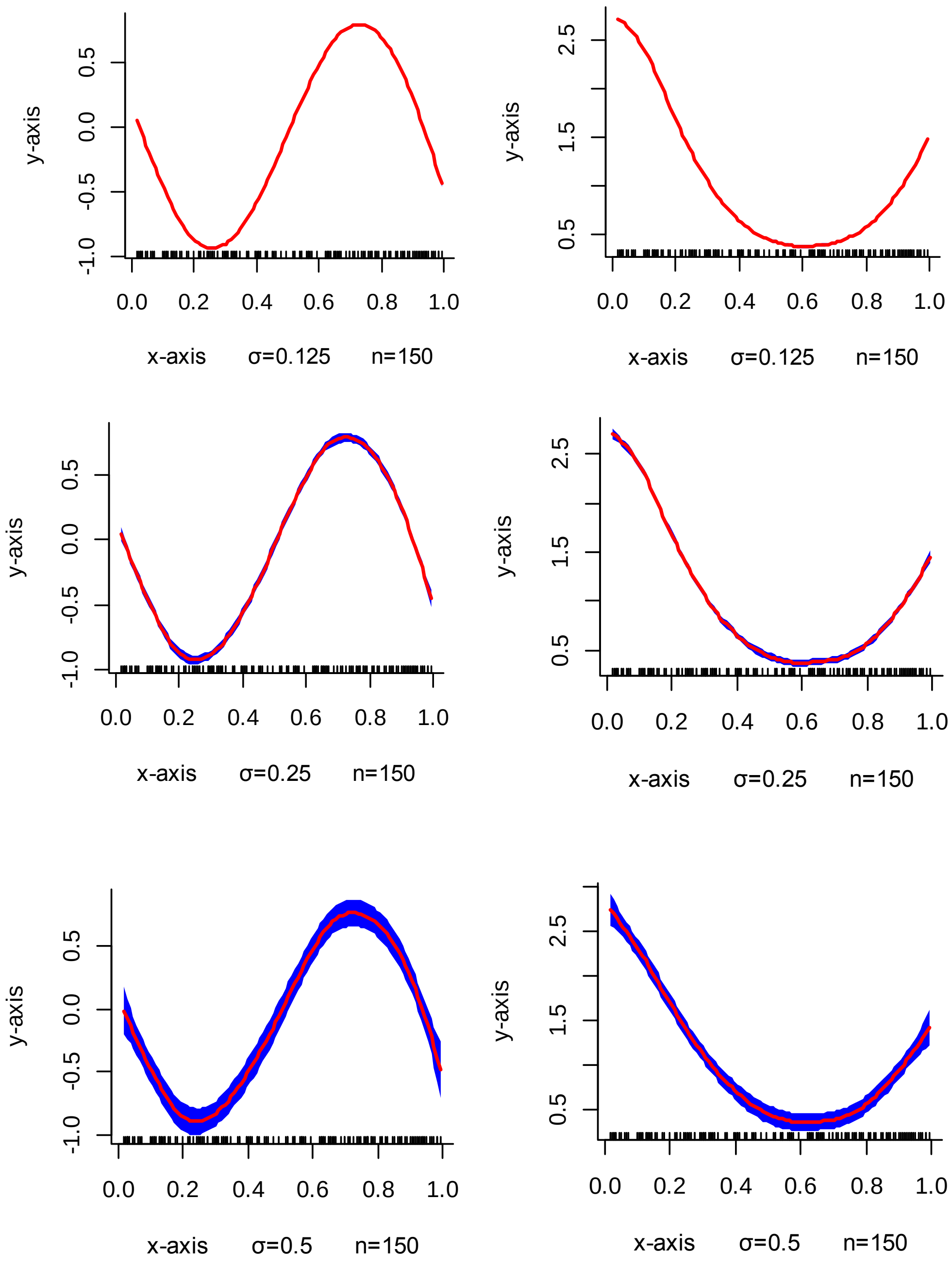

Figure (1) fitted curves from penalized regression spline estimation of first (right side) and second test function (left side) with design variable X distributed uniform distribution [-1,1] with the error distributed normal distribution $(0,\sigma^2)$, $\sigma = 0.125, 0.25$ and 0.5 , and sample size n=150

Figure (2) shows the posterior of β and u given Y (equation (14)) for above test functions((22) and (23)), where red curve represents the posterior of the first test function (22) while blue curve represents the posterior of the second test function (23)**.**

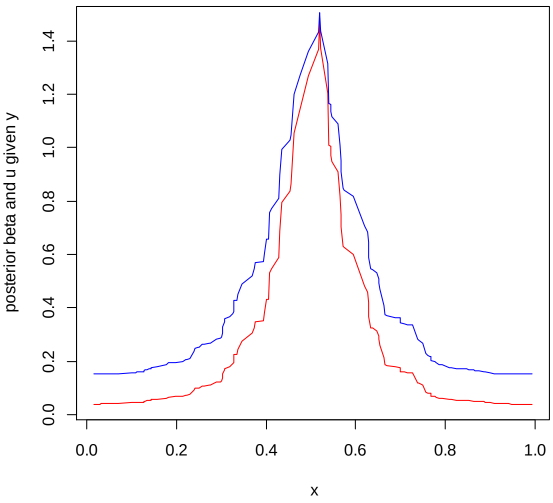

Figure (2) posterior of beta (β) and u given Y

Figure (3) below shows the number for iterations of the Gibbs sampler which used in this study. Which was 600 iterations for this data. While Figure (4) shows density estimates based on 600 iterations of σ_ϵ^2 and σ_u^2.

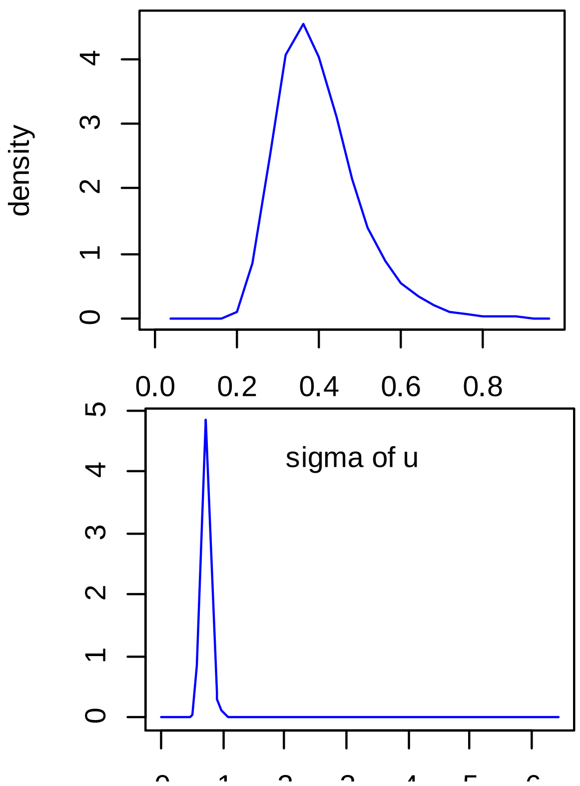

Figure (3) shows density estimates based on 600 iterations of σ_ϵ^2 and σ_u^2

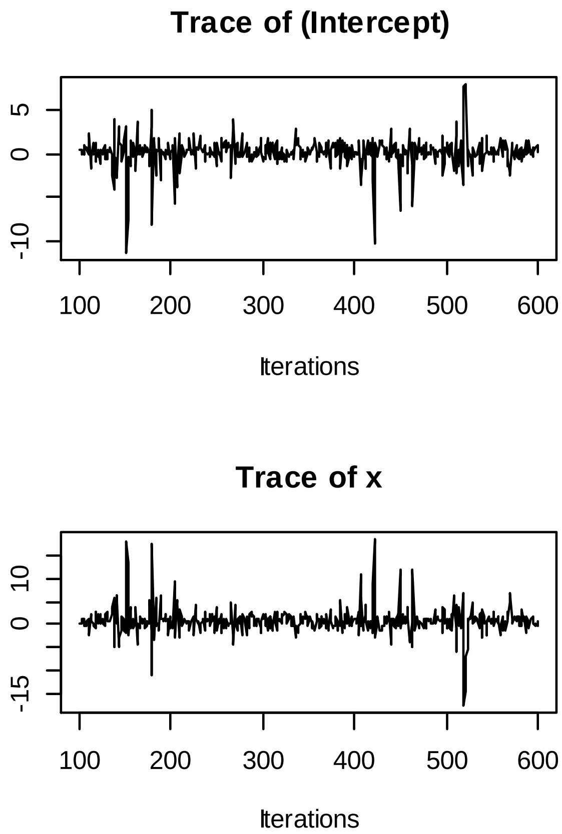

Figure (4) shows 600 iterations of the Gibbs sampler for the this data

7- Conclusions

1- The posterior density of δ given Y in Bayesian semiparametric regression is:

$$\pi_{22}(\delta|Y) \propto \frac{\delta^{(b/2)-1}}{(a+b\delta)^{-(a+b)/2}} \left(\prod_{i=1}^{n}(1+\delta d_i)\right)^{-1/2} \left(2\mathrm{B}_{\epsilon} + \sum_{i=1}^{n} \frac{s_i^2}{1+\delta d_i}\right)^{-(\mathrm{n}+2\mathrm{A}_{\epsilon}+2)/2}$$

2- The posterior mean of F in Bayesian semiparametric regression is:

$$E(F|Y) = \Gamma C^T BE\{(I_n + \delta D)^{-1} \mid Y\}s$$

3- The posterior covariance matrix of F in Bayesian semiparametric regression is:

$$var(F|Y) = \frac{1}{n+2A_{\epsilon}+2} E\left[\left(2B_{\epsilon} + \left(\sum_{i=1}^{n}\frac{s_i^2}{1+\delta d_i}\right)\right)|Y\right]\Gamma - \frac{1}{n+2A_{\epsilon}+2}\Gamma C^T BE\left[\left(2B_{\epsilon} + \left(\sum_{i=1}^{n}\frac{s_i^2}{1+\delta d_i}\right)\right)[I_n + \delta D]^{-1}|Y\right] B^T C\Gamma + E[R(\delta)R(\delta)^T|Y]$$

4- The Bayes factor in Bayesian semiparametric regression for testing the two models H_o : $\mathrm{g} = X\beta = \mathrm{g}^o$ versus H_1 : $\mathrm{g} = X\beta + Zu \neq \mathrm{g}^o$ is:

$$B_{01}(Y) = \frac{m(Y|H_o)}{m(Y|H_1)}$$

where $m(Y|H_i)$ is the predictive (marginal) density of Y under model $H_i, i = 0, 1$. We have

$$m(Y|H_o) = \int f(Y|\mathrm{g}^o, \sigma_\epsilon^2)\pi_0(\sigma_\epsilon^2)\, d\sigma_\epsilon^2$$

and

$$m(Y|H_1) = \int f(Y|F, \sigma_u^2, \sigma_\epsilon^2)\pi_0(F, \sigma_u^2, \sigma_\epsilon^2)\, dF d\sigma_u^2 d\sigma_\epsilon^2.$$

5- In a simulation study of Bayesian semiparametric regression we observe that :

(i) The values of (AMSE) and (AMAE) when ($\sigma = 0.5$) are smaller than that ($\sigma = 0.125$ and 0.25), which were (0.01656268) and (0.008447995) respectively.

(ii) The values of (AMSE) and (AMAE) are smaller when ($\sigma = 0.125$) and ($\sigma = 0.5$) respectively for second test function which were (0.009150507) and (0.004085605) respectively.

References

[1] Angers, J.-F. and Delampady, M. (2001). Bayesian nonparametric regression using wavelets. Sankhy$\tilde{a}$ Ser. B 63, 287 – 308.

[2] B. Walsh 2002, Introduction to Bayesian Analysis 1-3.

[3] Berger, J. O. (1985). Statistical Decision Theory and Bayesian Analysis. Springer, New York.

[4] DeRobertis, L. and Hartigan, J. A. (1981). Bayesian inference using intervals of measures. Annals of Statistics. Vol. 9, No. 2, 235 – 244.

[5] Ghosh, J. K., Delampady, M. and Samanta, T. (2006). Introduction to Bayesian Analysis: Theory and Methods. Springer, New York.

[6] Ghosh, J. K. and Ramamoorthi, R. V. (2003). Bayesian Nonparametrics. Springer, New York.

[7] Lenk, P. J, (1988). The logistic Normal distribution for Bayesian, Nonparametric, Predictive Densities. Journal of the American Statistical Association, Vol. 83, No. 402, 509-516.

[8] Muirhead, R. J. (1982). Aspects of Multivariate Statistical Theory. Wiley, New York.

[9] Natio, k. (2002) " Semiparametric regression with multiplicative adjustment " Communications in statistics, Theory and methods 31, 2289-2309.

[10] Pelenis, J (2012) Bayesian Semiparametric Regression. Institute for Advanced Studies, Vienna 1-39.

[11] Rashid, D. H. and Nasir, A. M. (1988). Statistical Inference. University of Baghdad, Baghdad - Iraq.

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

[13] Tarmaratram, K. (2011) Robust Estimation and Model Selection in Semiparametric Regression Models, proefschrift voorgedragen tot het behalen van de grad van Doctor.