Research Article ISSN: 2319-507X
Ammar Muslim Abdulhussein, IJPRET, 2013; Volume 2 (3): 1-18 IJPRET

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

BAYESIAN SEMIPARAMETRIC REGRESSION WITH FUZZY SETS

AMEERA JABER MOHAISEN, AMMAR MUSLIM ABDULHUSSEIN

1. Mathematics Department College of Education for Pure Science AL-Basrah University-Iraq

Accepted Date: 14/10/2013 ; Published Date: 01/11/2013

Abstract: In this paper, Bayesian approach to semiparametric regression is described using fuzzy sets and membership functions. The membership functions are interpreted as likelihood functions for the model, and we prove some theorems about posterior and Bayes factor.

Keywords: Mixed models, Semiparametric regression, Penalized spline, Fuzzy sets, Membership functions, Bayesian inference, Prior density, Posterior density, Bayes factor.

Corresponding Author: AMMAR MUSLIM ABDULHUSSEIN

Access Online On:

www.ijpret.com

How to Cite This Article:

Ammar Muslim Abdulhussein, IJPRET, 2013; Volume 2 (3): 1-18

PAPER-QR CODE

INTRODUCTION

Consider the model:

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

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 (see [1]), Ghosh and Ramamoorthi (see [8]) , and Lenk (see[9]), and so on. Furthermore, quantifying useful prior information of model (1) such as “g is close to (a specified function) g^o” (we will define this function in section 4) is difficult probabilistically, whereas this seems quite straightforward if instead an appropriate metric on the concerned function space is used. This is where fuzzy sets or membership functions can be made use of.

In this paper, a simple Bayesian approach to semiparametric regression is described using fuzzy sets and membership functions. The membership functions are interpreted as likelihood functions for the model, so that with the help of a reference prior they can be transformed to prior density functions. 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 and Bayes factor.

2. Fuzzy sets and membership functions

A fuzzy subset A of a space G (or just a fuzzy set A) is defined by a membership function:

$$h_A : G \rightarrow [0, 1].$$

The membership function, h_A(g), is supposed to express the degree of compatibility of g with A. For example, if G is the real line and A is the set of points “close to 0”, then $h_A(0) = 1$ indicates that 0 is certainly included in A, but $h_A(0.07) = 0.03$ says that 0.07 is not really "close" to 0 in this context. Similarly, if G is a set of functions and $A \subset G$ is a set of functions "close" to a given function g^0, then $h_A(g^0) = 1$ indicates that g^0 is certainly included in A; however, if $h_A(g^1) = 0.03$ with $g^1(x) = 4g^0(x) + 24$ then g^1 is not really "close" to g^0 in this case (See [2,3,5,15,16]). Note that even when $G = \Theta$ is the parameter space, a membership function $h_A(\theta)$ is not a probability density or mass function defined on Θ, and hence cannot be used to obtain a prior distribution directly. Angers and Delampady (see [3]) propose that a reasonable interpretation for a fuzzy subset A of Θ is that it is a likelihood function for θ given A. Another important question is how to define $h_{A\cap B}$ from h_A and h_B for incorporating h_A and h_B in Bayesian inference. If A and B are independent, then interpreting h_A and h_B as likelihood functions leads to the result that $h_{A\cap B} = h_A h_B$, for this purpose. Further, the qualitative ordering that underlies a membership function can also be investigated with this interpretation, in conjunction with a prior distribution, (see [2,3,5,15,16]).

3. Mixed Models

The general form of a linear mixed model for the ith subject (i = 1,..., n) is given as follows (see [14,17]),

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

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,

$$Y = X\beta + Zu + \epsilon \qquad (3)$$

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.

4. Semiparametric regression and spline

The model (1) can be expressed as a smooth penalized spline with q degree, then it's become as(see [14]):

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

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 [11,14])

$$Y = X\beta + Zu + \epsilon \qquad (5)$$

where

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

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

We assume that the function g is:

$$g = X\beta + Zu \qquad (6)$$

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

$$g^o = X\beta \qquad (7)$$

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

$$E(\epsilon) = 0; \qquad var(\epsilon) = \sigma_\epsilon^2 I$$

$$E(\beta) = 0; \qquad var(\beta) = \sigma_\beta^2 I \qquad (8)$$

$$E(u) = 0; \qquad var(u) = \sigma_u^2 I$$

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

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

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_u 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 [14])

5. Prior information and Membership functions

We have explained in the previous section that we would like to make use of imprecise prior information such as "g is close to g^o" by using a membership function (see [3,7,12]) which translates this into a measure of distance between the corresponding penalized spline coefficients. Let us examine the implications of assuming that the available prior information is quantified in terms of a membership function

$$h_A(g) = \varphi(d(g, g^o)),$$

where d is a measure of distance in L_2. Due to the penalized spline decomposition assumed on g as well as g^o (see section 4), a natural choice for d is the distance given by

$$d^2(g, g^o) = \|g - g^o\|^2 = \|X\beta + Zu - X\beta\|^2 = \|Zu\|^2 = \sum_k^K u_k^2.$$

We will use a membership function that will depend only on $d^2(g, g^o)$. Some possibilities for h_A are the following:

(i) The Gaussian membership function given by:

$$h_A(g) = exp(-d^2(g, g^o)) = exp(-\alpha \sum_k^K u_k^2) \quad (10)$$

This membership function can be explained as follows. Suppose we have available some past data of the form

$$y^* = x^*\beta + z^*u + \epsilon$$

where

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

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

Suppose $\mathrm{g} = X\beta + Zu$ is estimated from this data by $\hat{\mathrm{g}}$. Then the information in this data may be quantified using a membership function of the type

$$h_A(\mathrm{g}) = exp\left(-d^2(\mathrm{g},\hat{\mathrm{g}})\right) = exp\left(-\|\mathrm{g} - \hat{\mathrm{g}}\|^2\right) = \exp\left(-\left\|X\beta + Zu - X\hat{\beta} - Z\hat{u}\right\|^2\right) =$$
$$\exp\left(-\left\|X(\beta - \hat{\beta}) + Z(u - \hat{u})\right\|^2\right) = \exp\left(-\alpha\left(\sum_{j=0}^{p+q+1}\left(\beta_j - \hat{\beta}_j\right)^2 + \sum_k^K (u_k - \hat{u}_k)^2 - \mu\right)\right)$$

Where constant $\mu \geq 0$, g^o may then be identified with $\hat{\mathrm{g}}$. If we have multiple past data sets, we may then have available $h_{A_1}(\mathrm{g}) = exp(-d^2(\mathrm{g},\hat{\mathrm{g}}_1))$, $h_{A_2}(\mathrm{g}) = exp(-d^2(\mathrm{g},\hat{\mathrm{g}}_2))$, and so on, which may be combined into $h_A(\mathrm{g}) = h_{A_1 \cap A_2}(\mathrm{g}) = h_{A_1}(\mathrm{g})\, h_{A_2}(\mathrm{g})$

$$= exp(-d^2(\mathrm{g},\hat{\mathrm{g}}_1))\ exp(-d^2(\mathrm{g},\hat{\mathrm{g}}_2))$$

$$= \exp\left(-\left\|X\beta + Zu - X\hat{\beta}_1 - Z\hat{u}_1\right\|^2 - \left\|X\beta + Zu - X\hat{\beta}_2 - Z\hat{u}_2\right\|^2\right)$$

$$= \exp\left(-\left\|X(\beta - \hat{\beta}_1) + Z(u - \hat{u}_1)\right\|^2 - \left\|X(\beta - \hat{\beta}_2) + Z(u - \hat{u}_2)\right\|^2\right)$$

$$= \exp\left(-\alpha_1\left(\sum_{j=o}^{p+q+1}\left(\beta_j - \hat{\beta}_{1j}\right)^2 + \sum_k^K (u_k - \hat{u}_{1k})^2 - \mu_1\right)\right) +$$

$$\exp\left(-\alpha_2\left(\sum_{j=o}^{p+q+1}\left(\beta_j - \hat{\beta}_{2j}\right)^2 + \sum_k^K (u_k - \hat{u}_{2k})^2 - \mu_2\right)\right),$$

Where constants $\mu_1, \mu_2 \geq 0$, as an example one could consider fitting regression lines to two (or more) sets of past data with possibly different error variances and use the fitted regression lines along with the estimated variances for constructing the membership functions. The constants α_1 and α_2 provide additional scope for assigning different weights to the two sources of information, which is another appealing feature of this approach.

(ii) The multivariate t membership function

$$h_A(\mathrm{g}) = (1 + d^2(\mathrm{g},\mathrm{g}^o))^{-(K+q)/2} = (1 + u^T V^{-1} u/q)^{-(K+q)/2} \quad \ldots\ldots\ldots\ldots (11)$$

Where $q > 2$ is the degrees of freedom and K denotes the dimension of u. This is a continuous scale mixture of Gaussian membership functions with the same g^{o} for each of the membership functions. Since this vanishes more slowly than Gaussian membership function, one could expect better robustness with this (see [2,3]).

(iii) The uniform function

$$h_A(\mathrm{g}) = \begin{cases} 1, & if\ d(\mathrm{g},\mathrm{g}^o) \le \delta \\ 0, & otherwise \end{cases} \quad \ldots\ldots\ldots\ldots (12)$$

This is an extreme case where g is restricted to a neighborhood of g^{o}(see [2,3]). In order to proceed with Bayesian inference on g, we need to convert the membership function into a prior density. Thus we obtain the prior density

$$\pi(\mathrm{g}) \propto h_A(\mathrm{g})\, \pi_0(\mathrm{g}),$$

or, upon utilizing the spline decomposition for g, we have an equivalent prior density

$$\pi(F, \sigma_\epsilon^2) \propto h_A(F)\, \pi_0(F, \sigma_\epsilon^2), \quad \ldots\ldots\ldots(13)$$

where $F = [\beta, u]$.

5. Posterior calculations

We have the model

$$Y|F, \sigma_\epsilon^2, \sigma_u^2 \sim N(CF, \sigma_\epsilon^2 I_n + \sigma_u^2 Q). \quad (14)$$

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.

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

where f is the likelihood. From (14), f can be expressed as

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

Proceeding further, suppose π_0 of the form

$$\pi_0(F,\sigma_\epsilon^2,\sigma_u^2) = \pi_1(\sigma_\epsilon^2,\sigma_u^2) \qquad (15)$$

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 [3,14]).

Note that the conditional posterior densities are given by

$$\pi(F|Y,\sigma_\epsilon^2,\sigma_u^2) \propto \exp\left\{\frac{-1}{2}\,(Y-CF)^T\,(\sigma_\epsilon^2 I_n+\sigma_u^2 Q)^{-1}(Y-CF)\right\}\,h_A(F)$$
(16)

$$\pi(\sigma_\epsilon^2|Y,F,\sigma_u^2) \propto |\,\sigma_\epsilon^2 I_n+\sigma_u^2 Q|^{-1/2}\exp\left\{\frac{-1}{2}\,(Y-CF)^T\,(\sigma_\epsilon^2 I_n+\sigma_u^2 Q)^{-1}(Y-CF\quad \pi 1(\sigma\epsilon 2,\sigma u 2)\right. \qquad (17)$$

$$\pi(\sigma_u^2|Y,F,\sigma_\epsilon^2) \propto |\,\sigma_\epsilon^2 I_n+\sigma_u^2 Q|^{-1/2}\exp\left\{\frac{-1}{2}\,(Y-CF)^T\,(\sigma_\epsilon^2 I_n+\sigma_u^2 Q)^{-1}\,(Y-CF\,\pi 1\sigma\epsilon 2,\sigma u 2\right. \qquad (18)$$

However, major simplifications are possible with the Gaussian h_A as in (i) (see section 4). Specifically, assuming that $h_A(F)$ is proportional to the density of $N(F_o,\sigma_u^2\Gamma\,)$ with

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

$$Y|F,\sigma_\epsilon^2,\sigma_u^2 \sim N(\,CF\,,\sigma_\epsilon^2 I_n+\sigma_u^2 Q\,) \qquad (19)$$

$$F|\sigma_u^2 \sim N(F_o,\sigma_u^2\Gamma\,)$$

Therefore, it follows that

$$Y|\sigma_\epsilon^2,\sigma_u^2 \sim N(CF_o,\sigma_\epsilon^2 I_n + C\sigma_u^2\Gamma C^T\,) \qquad (20)$$

where $C\sigma_u^2\Gamma C^T = \sigma_u^2 Q$.

$$F|Y,\sigma_\epsilon^2,\sigma_u^2 \sim N(F_o + A_1(Y-CF_o),A_2) \qquad (21)$$

where

$$A_1 = \sigma_u^2\Gamma\,C^T\,\{\,\sigma_\epsilon^2 I_n + C\sigma_u^2\Gamma C^T\}^{-1} \qquad (22)$$

$$A_2 = \sigma_u^2\Gamma - \sigma_u^4\Gamma\,C^T\,\{\,\sigma_\epsilon^2 I_n + C\sigma_u^2\Gamma C^T\}^{-1}\{C\Gamma\} \qquad (23)$$

Now proceeding as in [3], we employ spectral decomposition to obtain $C\Gamma C^T = BDB^T$, where $D = diag(d_1, \ldots, d_n)$ is the matrix of eigenvalues and B is the orthogonal matrix of eigenvectors. Thus,

$$\sigma_\epsilon^2 I_n + [C\sigma_u^2 \Gamma C^T] = \sigma_\epsilon^2 I_n + B\sigma_u^2 DB^T = B\sigma_\epsilon^2 I_n B^T + B\sigma_u^2 DB^T = B\sigma_\epsilon^2 \left(I_n + \frac{\sigma_u^2}{\sigma_\epsilon^2} D \right) B^T$$

$$= \sigma_\epsilon^2 B(I_n + \delta D)B^T$$

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

$$m(Y|\sigma_\epsilon^2, \delta) = \frac{1}{(2\pi\sigma_\epsilon^2)^{n/2}} \frac{1}{\det[I_n + \delta D]^{1/2}} \exp\{ -\frac{1}{2\sigma_\epsilon^2} (Y - CF_o)^T B(I_n + \delta D)B^T(Y - CF_o)$$

$$= \frac{1}{(2\pi\sigma_\epsilon^2)^{n/2}} \frac{1}{[\prod_{i=1}^n [1+\delta d_i]]^{1/2}} \exp\{ -\frac{1}{2\sigma_\epsilon^2} (\sum_{i=1}^n \frac{s_i^2}{1+\delta d_i})\} \quad (24)$$

where $s = (s_1, \ldots, s_n)^T = B^T(Y - CF_o)$. We choose the prior on σ_ϵ^2, $\delta = \sigma_u^2 / \sigma_\epsilon^2$, qualitatively similar to the used in [3]. 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 A_ϵ, B_ϵ, *b* and *a*). Conditions apply on *a* and *b* such that (see[2,3]):

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 $2 < b \leq 4$ 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:

$$\pi_{22}(\delta|Y) \propto \frac{\delta^{(b/2)-1}}{(a+b\delta)^{-(a+b)/2}} (\prod_{i=1}^n (1 + \delta d_i))^{-1/2} \left(2B_\epsilon + \sum_{i=1}^n \frac{s_i^2}{1+\delta d_i} \right)^{-(n+2A_\epsilon+2)/2} \quad (25)$$

Proof:

$$\pi_{22}(\delta|Y) = \int m(Y|\sigma_\epsilon^2, \delta)\ f(\delta, b, a)\, f(\sigma_\epsilon^2, A_\epsilon, B_\epsilon) d\sigma_\epsilon^2$$

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

$$\exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(\sum_{i=1}^n \frac{s_i^2}{1+\delta d_i}\right)\right\} \frac{b^{b/2}a^{a/2}}{B(b,a)} \frac{\delta^{(b/2)-1}}{(a+b\delta)^{-(a+b)/2}} \, d\sigma_\epsilon^2$$

$$= \frac{(2\pi)^{-n/2}}{\Gamma(A_\epsilon)} \frac{b^{b/2}a^{a/2}}{B(b,a)} \frac{\delta^{(b/2)-1}}{(a+b\delta)^{-(a+b)/2}} \int \left(\prod_{i=1}^n (1+\delta d_i)\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(2B_\epsilon + \sum_{i=1}^n \frac{s_i^2}{1+\delta d_i}\right)\right\}$$

$$(\sigma_\epsilon^2)^{-(n+2A_\epsilon+2)/2} d\sigma_\epsilon^2$$

$$= \frac{(2\pi)^{-n/2}}{\Gamma(A_\epsilon)} \frac{b^{b/2}a^{a/2}}{B(b,a)} \frac{\delta^{(b/2)-1}}{(a+b\delta)^{-(a+b)/2}} (2)^{(n+2A_\epsilon+2)/2} \int \left(\prod_{i=1}^n (1+\delta d_i)\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(2B_\epsilon + \right.\right.$$

$i=1nsi21+\delta di$

$$\left(\frac{2B_\epsilon + \sum_{i=1}^n \frac{s_i^2}{1+\delta d_i}}{2\sigma_\epsilon^2}\right)^{(n+2A_\epsilon+2)/2} \left(2B_\epsilon + \sum_{i=1}^n \frac{s_i^2}{1+\delta d_i}\right)^{-(n+2A_\epsilon+2)/2} d\sigma_\epsilon^2$$

$$\propto \frac{\delta^{(b/2)-1}}{(a+b\delta)^{-(a+b)/2}} \int \left(\prod_{i=1}^n (1+\delta d_i)\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(2B_\epsilon + \sum_{i=1}^n \frac{s_i^2}{1+\delta d_i}\right)\right\}$$

$$\left(\frac{2B_\epsilon \sum_{i=1}^n \frac{s_i^2}{1+\delta d_i}}{2\sigma_\epsilon^2}\right)^{[(n+2A_\epsilon+4)/2]-1} \left(2B_\epsilon + \sum_{i=1}^n \frac{s_i^2}{1+\delta d_i}\right)^{-(n+2A_\epsilon+2)/2} d\sigma_\epsilon^2$$

$$\propto \frac{\delta^{(b/2)-1}\Gamma((n+2A_\epsilon+4)/2)}{(a+b\delta)^{-(a+b)/2}} \left(\prod_{i=1}^n (1+\delta d_i)\right)^{-1/2} \left(2B_\epsilon + \sum_{i=1}^n \frac{s_i^2}{1+\delta d_i}\right)^{-(n+2A_\epsilon+2)/2}$$

$$\propto \frac{\delta^{(b/2)-1}}{(a+b\delta)^{-(a+b)/2}} \left(\prod_{i=1}^n (1+\delta d_i)\right)^{-1/2} \left(2B_\epsilon + \sum_{i=1}^n \frac{s_i^2}{1+\delta d_i}\right)^{-(n+2A_\epsilon+2)/2}$$

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

$$E(F|Y) = F_o + \Gamma C^T B E\{(I_n + \delta D)^{-1}|Y\}s \tag{26}$$

and

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

$i=1nsi21+\delta di\ In+\delta D-1|Y\ BTC\Gamma+E[R(\delta)R\delta T|Y],$ (27)

where $R(\delta) = \Gamma C^T B(I_n + \delta D)^{-1}s$

Proof:

From (21):

$$E(F|Y) = F_o + A_1(Y - CF^o)$$

$$= F_o + \{\sigma_u^2 \Gamma C^T\}\{\sigma_\epsilon^2 I_n + C\sigma_u^2 \Gamma C^T\}^{-1}(Y - CF_o)$$

$$= F_o + \sigma_u^2 \Gamma C^T \{\sigma_\epsilon^2 B(I_n + \delta D)B^T\}^{-1}(Y - CF_o)$$

$$= F_o + \frac{\sigma_u^2}{\sigma_\epsilon^2}\Gamma C^T {B^T}^{-1}(I_n + \delta D)^{-1} B^{-1}(Y - CF_o)$$

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

Therefore

$$E(F|Y) = F_o + \Gamma C^T B\, \delta\, (I_n + \delta D)^{-1} B^T (Y - CF_o))$$

$$= F_o + \Gamma C^T B\, E((I_n + \delta D)^{-1}|Y)s,$$

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

6. Model checking and Bayes factors

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

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

Under H_1, (g = g(F), $\sigma_u^2, \sigma_\epsilon^2$) is given the prior $h_A(F)\pi_0(F, \sigma_u^2, \sigma_\epsilon^2)I(\text{g} \neq \text{g}^o)$, 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:

$$B_{01}(Y) = \frac{m(Y|H_o)}{m(Y|H_1)} \qquad (28)$$

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

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

and

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

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,

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

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

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

$= (2\pi)^{-n/2} \frac{B_\epsilon^{A_\epsilon}}{\Gamma(A_\epsilon)} \int (\sigma_\epsilon^2)^{-\left(\frac{n}{2}+A_\epsilon+1\right)} (B_\epsilon + \frac{1}{2}(y_i - g^o(x_i))^2)^{\frac{n}{2}+A_\epsilon+1} (B_\epsilon + \frac{1}{2}(y_i -$
go(xi))2)−(n2+Aϵ+1)exp−Bϵ+12(yi−go(xi))2σϵ2 dσϵ2

$= (2\pi)^{-n/2} \frac{B_\epsilon^{A_\epsilon}}{\Gamma(A_\epsilon)} \int \frac{(B_\epsilon+\frac{1}{2}(y_i-g^o(x_i))^2)^{\frac{n}{2}+A_\epsilon+1}}{(\sigma_\epsilon^2)^{\left(\frac{n}{2}+A_\epsilon+1\right)}} \exp\left(-\frac{B_\epsilon+\frac{1}{2}(y_i-g^o(x_i))^2}{\sigma_\epsilon^2}\right) (B_\epsilon + \frac{1}{2}(y_i -$
go(xi))2)−(n2+Aϵ+1) dσϵ2

$= (2\pi)^{-n/2} \frac{B_\epsilon^{A_\epsilon}}{\Gamma(A_\epsilon)} \int \left(\frac{B_\epsilon+\frac{1}{2}(y_i-g^o(x_i))^2}{\sigma_\epsilon^2}\right)^{\left(\frac{n}{2}+A_\epsilon+2\right)-1} \exp\left(-\frac{B_\epsilon+\frac{1}{2}(y_i-g^o(x_i))^2}{\sigma_\epsilon^2}\right) (B_\epsilon + \frac{1}{2}(y_i -$
go(xi))2)−(n2+Aϵ+1) dσϵ2

$$= (2\pi)^{-n/2} \frac{B_\epsilon^{A_\epsilon}}{\Gamma(A_\epsilon)} \Gamma(\frac{n}{2} + A_\epsilon + 1)(B_\epsilon + \frac{1}{2}(y_i - g^o(x_i))^2)^{-\left(\frac{n}{2}+A_\epsilon+1\right)}\ d\sigma_\epsilon^2 \quad (29)$$

Further, using (20) it follows that:

$$m(Y|H_1, \sigma_\epsilon^2, \delta) = (2\pi\sigma_\epsilon^2)^{-\frac{n}{2}} \left(\prod_{i=1}^n (1 + \delta d_i)\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(\sum_{i=1}^n \frac{s_i^2}{1+\delta d_i}\right)\right\} \quad (30)$$

Therefore,

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

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

$$\exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(\sum_{i=1}^n \frac{s_i^2}{1+\delta d_i}\right)\right\} \pi_0(\delta)\ d\sigma_\epsilon^2\ d\delta$$

$$= \frac{B_\epsilon^{A_\epsilon}}{\Gamma(A_\epsilon)}(2\pi)^{-n/2}\int \left(\prod_{i=1}^{n}(1+\delta d_i)\right)^{-1/2}\pi_0(\delta)$$

$$\left\{\int \exp\left\{-\frac{1}{\sigma_\epsilon^2}\left(B_\epsilon + \frac{1}{2}\sum_{i=1}^{n}\frac{s_i^2}{1+\delta d_i}\right)\right\} d\sigma_\epsilon^2\right\} d\delta$$

$$= \frac{B_\epsilon^{A_\epsilon}}{\Gamma(A_\epsilon)}(2\pi)^{-n/2}\Gamma(n/2 + A_\epsilon)\int \left(\prod_{i=1}^{n}(1+\delta d_i)\right)^{-1/2}$$

$$\pi_0(\delta)\left(B_\epsilon + \frac{1}{2}\sum_{i=1}^{n}\frac{s_i^2}{1+\delta d_i}\right)^{-(n/2+c-1)} d\delta \qquad (31)$$

6.1. Prior robustness of Bayes factors

Note that the most informative part of the prior density that we have used is contained in the membership function h_A. Since a membership function h_A(F) is to be treated only as a likelihood for F, any constant multiple ch_A(F) also contributes the same prior information about F. Therefore, a study of the robustness of the Bayes factor that we obtained above with respect to a class of priors compatible with h_A is of interest. Here we consider a sensitivity study using the density ratio class defined as follows. Since the prior π that we use has the form

$$\pi(F,\sigma_u^2,\sigma_\epsilon^2) \propto h_A(F)\pi_0(F,\sigma_u^2,\sigma_\epsilon^2),$$

we consider the class of priors

$$C_A = \{\pi : c_1 h_A(F)\pi_0(F,\sigma_u^2,\sigma_\epsilon^2) \le \alpha\pi(F,\sigma_u^2,\sigma_\epsilon^2) \le c_2 h_A(F)\pi_0(F,\sigma_u^2,\sigma_\epsilon^2), \alpha > 0\}$$

For specified $0 < c_1 < c_2$. We would like to investigate how the Bayes factor (28) behaves as the prior π varies in C_A. We note that for any $\pi \in C_A$, the Bayes factor B_{01} has the form

$$B_{01} = \frac{\int f(Y|g^o,\sigma_\epsilon^2)\,\pi(F,\sigma_u^2,\sigma_\epsilon^2)\,dF\,d\sigma_u^2\,d\sigma_\epsilon^2}{\int f(Y|F,\sigma_u^2,\sigma_\epsilon^2)\,\pi(F,\sigma_u^2,\sigma_\epsilon^2)\,dF\,d\sigma_u^2\,d\sigma_\epsilon^2}$$

Even though the integration in the numerator above need not involve F, σ_u^2, we do so to apply the following result(see[1,2,3,6,8]).

Consider the density-ratio class

$$\Gamma_{DR} = \{\pi : L(\eta) \le \alpha\pi(\eta) \le U(\eta) \text{ for some } \alpha > 0\}$$

, for specified non-negative functions L and U. Further, let $q \equiv q^+ + q^-$ be the usual decomposition of q into its positive and negative parts, i.e., $q^+(u) = max\{q(u), 0\}$ and $q^-(u) = -max\{-q(u), 0\}$. Then we have the following theorem.

Theorem 3: For functions q_1 and q_2 such that $\int q_i(\eta)|U(\eta)\, d\eta \; < \; \infty$, for i = 1, 2, and with q_2 positive a.s. with respect to all $\pi \in \Gamma_{DR}$,

$$\inf_{\pi\in\Gamma_{DR}} \frac{\int q_1(\eta)\,\pi(\eta)\,d\eta}{\int q_2(\eta)\,\pi(\eta)\,d\eta}$$

is the unique solution ϑ of

$$\int(q_1(\eta)-\vartheta q_2(\eta))^{-}U(\eta)d\eta + \int(q_1(\eta)-\vartheta q_2(\eta))^{+}L(\eta)d\eta = 0 \qquad (32)$$

$$\sup_{\pi\in\Gamma_{DR}} \frac{\int q_1(\eta)\,\pi(\eta)\,d\eta}{\int q_2(\eta)\,\pi(\eta)\,d\eta}$$

is the unique solution ϑ of

$$\int(q_1(\eta)-\vartheta q_2(\eta))^{+}U(\eta)d\eta + \int(q_1(\eta)-\vartheta q_2(\eta))^{-}L(\eta)d\eta = 0 \qquad (33)$$

Proof:

To prove part one

$$\int q_1(\eta)^{-}\,U(\eta)d\eta \;+\; \int q_1(\eta)^{+}\,L(\eta)d\eta \;-\; \vartheta\int q_2(\eta)^{-}\,U(\eta)d\eta \;-\; \vartheta\int q_2(\eta)^{+}\,L(\eta)d\eta = 0$$

$$\Rightarrow \int (q_1(\eta)^{-}\,U(\eta) + q_1(\eta)^{+}\,L(\eta))d\eta \;-\; \vartheta\int (q_2(\eta)^{-}\,U(\eta) + q_2(\eta)^{+}\,L(\eta))d\eta \;=0$$

$$\Rightarrow \vartheta = \frac{\int (q_1(\eta)^{-}\,U(\eta) + q_1(\eta)^{+}\,L(\eta))d\eta}{\int (q_2(\eta)^{-}\,U(\eta) + q_2(\eta)^{+}\,L(\eta))d\eta}$$

By theorem 4.1. in DeRobertis and Hartigan (1981) (see [6]),

$(q_1(\eta)^{-}\,U(\eta) + q_1(\eta)^{+}\,L(\eta)) = inf_{\pi\in\Gamma_{DR}}\, Kq_1(\eta)\,, where\ K \in I(L,U)$, then

$$\Rightarrow \vartheta = \frac{\int \inf_{\pi\in\Gamma_{DR}} Kq_1(\eta)\,\pi(\eta)\,d\eta}{\int \inf_{\pi\in\Gamma_{DR}} Kq_2(\eta)\,\pi(\eta)\,d\eta}$$

$$\Rightarrow \vartheta = \inf_{\pi\in\Gamma_{DR}} \frac{\int q_1(\eta)\,\pi(\eta)\,d\eta}{\int q_2(\eta)\,\pi(\eta)\,d\eta}$$

Then the $\inf_{\pi\in\Gamma_{DR}} \frac{\int q_1(\eta)\,\pi(\eta)\,d\eta}{\int q_2(\eta)\,\pi(\eta)\,d\eta}$ is the sol99oiuophpkiution ϑ, now to prove unique solution suppose

$\vartheta_0 = \inf_{\pi\in\Gamma_{DR}} \frac{\int q_1(\eta)\,\pi(\eta)\,d\eta}{\int q_2(\eta)\,\pi(\eta)\,d\eta}$, $c_1 = \inf_{\pi\in\Gamma_{DR}} \int q_2(\eta)\,\pi(\eta)\,d\eta$ and $c_2 = \sup_{\pi\in\Gamma_{DR}} \int q_2(\eta)\,\pi(\eta)\,d\eta$. Then $0 < c_1 < c_2 < \infty$ and $|\vartheta_0| < \infty$ it follows that $\vartheta_0 \geq \vartheta$ if and only if $\int(q_1(\eta) -$

$\vartheta q_2(\eta))^- U(\eta)d\eta + \int(q_1(\eta) - \vartheta q_2(\eta))^+ L(\eta)d\eta \geq 0$.Moreover, for any $\epsilon \geq 0$, $\vartheta + \epsilon / c_1 \leq \vartheta_0$ implies $\int(q_1(\eta) - \vartheta q_2(\eta))^- U(\eta)d\eta + \int(q_1(\eta) - \vartheta q_2(\eta))^+ L(\eta)d\eta \geq \epsilon$ which in turn implies $\vartheta + \epsilon / c_2 \leq \vartheta_0$; thus ; $\vartheta_0 > \vartheta$ if and only if $\int(q_1(\eta) - \vartheta q_2(\eta))^- U(\eta)d\eta + \int(q_1(\eta) - \vartheta q_2(\eta))^+ L(\eta)d\eta > 0$. Hence, then ϑ is the unique solution.

Now to prove part two

$$\int q_1(\eta)^+ U(\eta)d\eta + \int q_1(\eta)^- L(\eta)d\eta - \vartheta \int q_2(\eta)^+ U(\eta)d\eta - \vartheta \int q_2(\eta)^- L(\eta)d\eta = 0$$

$$\Rightarrow \int (q_1(\eta)^+ U(\eta) + q_1(\eta)^- L(\eta))d\eta - \vartheta \int (q_2(\eta)^+ U(\eta) + q_2(\eta)^- L(\eta))d\eta = 0$$

$$\Rightarrow \vartheta = \frac{\int (q_1(\eta)^+ U(\eta) + q_1(\eta)^- L(\eta))d\eta}{\int (q_2(\eta)^+ U(\eta) + q_2(\eta)^- L(\eta))d\eta}$$

Also by theorem 4.1. in DeRobertis and Hartigan (1981) (see [6]),

$\left(q_1(\eta)^+ U(\eta) + q_1(\eta)^- L(\eta)\right) = \sup_{\pi\in\Gamma_{DR}} K q_1(\eta)$, where $K \in I(L, U)$, then

$$\Rightarrow \vartheta = \frac{\int \sup_{\pi\in\Gamma_{DR}} K q_1(\eta)\, \pi(\eta)\, d\eta}{\int \sup_{\pi\in\Gamma_{DR}} K q_2(\eta)\, \pi(\eta)\, d\eta}$$

$$\Rightarrow \vartheta = \sup_{\pi\in\Gamma_{DR}} \frac{\int q_1(\eta)\, \pi(\eta)\, d\eta}{\int q_2(\eta)\, \pi(\eta)\, d\eta}$$

By same way of proof the unique of first part above (the proof complete) .

Now we shall discuss this result for the Gaussian membership function only. Then, since the prior π that we use has the form $\pi_0(F, \sigma_u^2, \sigma_\epsilon^2) \propto h_A(\theta)\, \pi_0(\sigma_u^2, \sigma_\epsilon^2)$, and we don't intend to vary $\pi_0(\sigma_u^2, \sigma_\epsilon^2)$ in our analysis, we redefine C_A as

$$C_A = \{\pi(F) : c_1 h_A(\mathrm{F}) \leq \alpha\pi(F) \leq c_2 h_A(\mathrm{F}), \quad \alpha > 0\}$$

For specified $0 < c_1 < c_2$. Now, were express B_{01} as

$$B_{01} = \frac{\int\{\int f(Y|\mathrm{g}^o, \sigma_\epsilon^2)\, \pi_0(\sigma_\epsilon^2)\, d\sigma_\epsilon^2\}\pi(F)dF}{\int\{\int f(Y|F, \sigma_u^2, \sigma_\epsilon^2)\, \pi_0(\sigma_u^2, \sigma_\epsilon^2) d\sigma_u^2 d\sigma_\epsilon^2\}\pi(F)dF} = \frac{\int q_1(F)\, \pi(F)\, dF}{\int q_2(F)\, \pi(F)\, dF}$$

where

$$q_1(F) = \int f(Y|\mathrm{g}^o, \sigma_\epsilon^2)\, \pi_0(\sigma_\epsilon^2)\, d\sigma_\epsilon^2$$

$$q_2(F) = \int f(Y|F, \sigma_u^2, \sigma_\epsilon^2)\, \pi_0(\sigma_u^2, \sigma_\epsilon^2) d\sigma_u^2 d\sigma_\epsilon^2$$

Then by theorem 3 is readily applicable, and we obtain the following theorem:

Theorem 4:

$\inf_{\pi \in C_A} B_{01}(\pi)$ is the unique solution ϑ of

$$c_2 \int (q_1(F) - \vartheta q_2(F))^- h_A(F) dF + c_1 \int (q_1(F) - \vartheta q_2(F))^+ h_A(F) dF = 0 \quad (34)$$

and $\sup_{\pi \in C_A} B_{01}(\pi)$ is the unique solution ϑ of

$$c_2 \int (q_1(F) - \vartheta q_2(F))^+ h_A(F) dF + c_1 \int (q_1(F) - \vartheta q_2(F))^- h_A(F) dF = 0 \quad (35)$$

Proof:

To prove the first part

$$c_2 \int q_1(\mathrm{F})^- U(F) dF + c_1 \int q_1(\mathrm{F})^+ L(F) dF - \vartheta c_2 \int q_2(\mathrm{F})^- U(F) dF - \vartheta c_1 \int q_2(\mathrm{F})^+ L(F) dF = 0$$

$$\Rightarrow \int (c_2 q_1(F)^- U(F) + c_1 q_1(F)^+ L(F)) dF - \vartheta \int (c_2 q_2(F)^- U(F) + c_1 q_2(F)^+ L(F)) dF = 0$$

$$\Rightarrow \vartheta = \frac{\int (c_2 q_1(\mathrm{F})^- U(F) + c_1 q_1(\mathrm{F})^+ L(F)) dF}{\int (c_2 q_2(\mathrm{F})^- U(F) + c_1 q_2(\mathrm{F})^+ L(F)) dF}$$

Then,

$\left(c_2 q_1(\mathrm{F})^- U(F) + c_1 q_1(\mathrm{F})^+ L(F)\right) = \inf_{\pi \in \Gamma_{DR}} cKq_1(\mathrm{F})$ $, where\ K \in I(L, U), c \le c_1 + c_2,$ then

$$\Rightarrow \vartheta = \frac{\int \inf_{\pi \in C_A} cKq_1(\mathrm{F})\, h_A(\mathrm{F})\, dF}{\int \inf_{\pi \in C_A} cKq_2(\mathrm{F})\, h_A(\mathrm{F})\, dF}$$

$$\Rightarrow \vartheta = \inf_{\pi \in C_A} \frac{\int q_1(\mathrm{F})\, h_A(\mathrm{F})\, dF}{\int q_2(\mathrm{F})\, h_A(\mathrm{F})\, dF}$$

$$\Rightarrow \vartheta = \inf_{\pi \in C_A} B_{01}(\pi)$$

To prove the second part

$$c_2 \int q_1(\mathrm{F})^+ \, U(F)dF \;+\; c_1 \int q_1(\mathrm{F})^- \, L(F)dF \;-\; \vartheta c_2 \int q_2(\mathrm{F})^+ \, U(F)dF$$
$$- \; \vartheta c_1 \int q_2(\mathrm{F})^- \, L(F)dF \;= 0$$

$$\Longrightarrow \int (c_2 q_1(\mathrm{F})^+ \, U(F) + c_1 q_1(\mathrm{F})^- \, L(F))dF \;-\; \vartheta \int (c_2 q_2(\mathrm{F})^+ \, U(F) + c_1 q_2(\mathrm{F})^- \, L(F))dF$$
$$= 0$$

$$\Longrightarrow \vartheta = \frac{\int (c_2 q_1(\mathrm{F})^+ \, U(F) + c_1 q_1(\mathrm{F})^- \, L(F))dF}{\int (c_2 q_2(\mathrm{F})^+ \, U(F) + c_1 q_2(\mathrm{F})^- \, L(F))dF}$$

Then,

$(c_2 q_1(\mathrm{F})^+ \, U(F) + c_1 q_1(\mathrm{F})^- \, L(F) = \sup_{\pi \in \Gamma_{DR}} cKq_1(\mathrm{F})$, $where\ K \in I(L, U), c \le c_1 + c_2,$ then

$$\Longrightarrow \vartheta = \frac{\int \sup_{\pi \in \Gamma_{DR}} cKq_1(\mathrm{F}) \, h_A(\mathrm{F}) \, dF}{\int \sup_{\pi \in \Gamma_{DR}} cKq_2(\mathrm{F}) \, h_A(\mathrm{F}) \, dF}$$

$$\Longrightarrow \vartheta = \sup_{\pi \in \Gamma_{DR}} \frac{\int q_1(\mathrm{F}) \, h_A(\mathrm{F}) \, dF}{\int q_2(\mathrm{F}) \, h_A(\mathrm{F}) \, dF}$$

$$\Longrightarrow \vartheta = \sup_{\pi \in \Gamma_{DR}} B_{01}(\pi)$$

By same as the unique prove to first part in theorem 3.

7. CONCLUSIONS

In this paper we suggest approach to semiparametric regression by proposing an alternative to dealing with complicated analyses on function spaces. The proposed technique uses fuzzy sets to quantify the available prior information on a function space by starting with a “prior guess” baseline regression function g^o. First the penalized spline is used for the model and by using a convenient connection between penalized splines and mixed models, we can representation semiparametric regression model as mixed model. The penalized spline assumed on g and pure polynomial on prior g^o. Then prior of g relative to distance from g^o specified in the form of a membership function which translates this distance into a measure of distance between the corresponding mixed model coefficients. Furthermore we obtain the posterior density of δ given Y, the posterior mean and covariance matrix of F (theorem 1, 2), and a Bayesian test is proposed to check whether the baseline function g^o is compatible with the data or not and we proved the prior robustness of Bayes factors (theorem 3, 4).

REFERENCES

1. Angers, J.-F. and Delampady, M. (2001). Bayesian nonparametric regression using wavelets. Sankhyã Ser. B 63, 287–308.
2. Angers, J.-F. and Delampady, M. (2004). Fuzzy sets in hierarchical Bayes and robust Bayes inference. J. Statist. Research 38, 1–11.
3. Angers, J.-F. and Delampady, M. (2008) Fuzzy set in nonparametric Bayes regression. IMS Collections Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh Vol 3, 89-104.
4. Berger, J. O. (1985). Statistical Decision Theory and Bayesian Analysis. Springer, New York.
5. Chen, G and Pham, T. T., (2001). Introduction to fuzzy sets, fuzzy logic, and fuzzy control systems CRC Press Boca Raton London New York Washington, D.C.1-36.
6. DeRobertis, L. and Hartigan, J. A. (1981). Bayesian inference using intervals of measures. Annals of Statistics. Vol. 9, No. 2, 235–244.
7. Ghosh, J. K., Delampady, M. and Samanta, T. (2006). Introduction to Bayesian Analysis: Theory and Methods. Springer, New York.
8. Ghosh, J. K. and Ramamoorthi, R. V. (2003). Bayesian Nonparametrics. Springer, New York.
9. 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.
10. Muirhead, R. J. (1982). Aspects of Multivariate Statistical Theory. Wiley, New York.
11. Natio, k. (2002) " Semiparametric regression with multiplicative adjustment " Communications in statistics, Theory and methods 31, 2289-2309.
12. Pelenis, J (2012) Bayesian Semiparametric Regression. Institute for Advanced Studies, Vienna 1-39.
13. Rashid, D. H. and Nasir, A. M. (1988). Statistical Inference. University of Baghdad, Baghdad - Iraq.
14. Ruppert, D, Wand, M. P. and Carroll, R. J.(2003) Semiparametric Regression. Cambridge University Press.
15. Singpurwalla, N., Booker, J. M. and Bement, T. R. (2002). Probability theory. In Fuzzy Logic and Probability Applications (T. J. Ross et al., eds.) 55–71. SIAM, Philadelphia.
16. Singpurwalla, N. and Booker, J. M. (2004). Membership functions and probability measures of fuzzy sets (with discussion). Journal of the American Statistical Association. Assoc. 99, 867–889.
17. Tarmaratram, K. (2011) Robust Estimation and Model Selection in Semiparametric Regression Models, proefschrift voorgedragen tot het behalen van de grad van Doctor.