

# Bayesian Estimation and Inference for the Generalized Partial Linear Model

Haitham M. Yousof<sup>1</sup>, Ahmed M. Gad<sup>2,\*</sup>

<sup>1</sup>Department of Statistics, Mathematics and Insurance, Benha University, Egypt

<sup>2</sup>Department of Statistics, Cairo University, Egypt

---

**Abstract** In this article we propose a Bayesian regression model called the Bayesian generalized partial linear model which extends the generalized partial linear model. We consider Bayesian estimation and inference of parameters for the generalized partial linear model (GPLM) using some multivariate conjugate prior distributions under the square error loss function. We propose an algorithm for estimating the GPLM parameters using Bayesian theorem in more detail. Finally, comparisons are made between the GPLM estimators using Bayesian approach and the classical approach via a simulation study.

**Keywords** Generalized Partial Linear Model, Profile Likelihood Method, Generalized Speckman Method, Back-Fitting Method, Bayesian Estimation

---

## 1. Introduction

The semi-parametric regression models are intermediate step between the fully parametric and nonparametric models. Many definitions of semi-parametric models are available in literature. The definition that will be adopted in this article is that the model is a semi-parametric model if it contains a nonparametric component in addition to a parametric component and they need to be estimated. The semi-parametric models are characterized by a finite-dimensional component,  $\theta \in \Theta \subseteq \mathbb{R}^k$  and an infinite-dimensional  $\gamma \in \Gamma$ .

Semi-parametric models try to combine the flexibility of a nonparametric model with the advantages of a parametric model. A fully nonparametric model will be more robust than semi-parametric and parametric models since it does not suffer from the risk of misspecification. On the other hand, nonparametric estimators suffer from low convergence rates, which deteriorate when considering higher order derivatives and multidimensional random variables. In contrast, the parametric model carries a risk of misspecification but if it is correctly specified it will normally enjoy  $\sqrt{n}$ -consistency with no deterioration caused by derivatives and multivariate data. The basic idea of a semi-parametric model is to take the best of both models. The semi parametric generalized linear model known as the generalized partial linear model (GPLM) is one of the semi-parametric regression models, See Powell

(1994); Ruppert et al. (2003); and Sperlich et al. (2006).

Many authors have tried to introduce new algorithms for estimating the semi-parametric regression models. Meyer et al. (2011) have introduced Bayesian estimation and inference for generalized partial linear models using shape-restricted splines. Zhang et al. (2014) have studied estimation and variable selection in partial linear single index models with error-prone linear covariates. Guo et al. (2015) have studied the empirical likelihood for single index model with missing covariates at random. Bouaziz et al. (2015) have studied semi-parametric inference for the recurrent events process by means of a single-index model.

The curse of dimensionality problem (COD) associated with nonparametric density and conditional mean function makes the nonparametric methods impractical in applications with many regressors and modest size samples. This problem limits the ability to examine data in a very flexible way for higher dimensional problems. As a result, the need for other methods became important. It is shown that semi parametric regression models can be of substantial value in solution of such complex problems.

Bayesian methods provide a joint posterior distribution for the parameters and hence allow for inference through various sampling methods. A number of methods for Bayesian monotone regression have been developed. Ramgopal et al (1993) introduced a Bayesian monotone regression approach using Dirichlet process priors. Perron and Mengersen (2001) proposed a mixture of triangular distributions where the dimension is estimated as part of the Bayesian analysis. Both Holmes and Heard (2003) and Wang (2008) model functions where the knot locations are free parameters with the former using a piecewise constant model and the latter imposing the

---

\* Corresponding author:

ahmed.gad@feps.edu.eg (Ahmed M. Gad)

Published online at <http://journal.sapub.org/ijps>

Copyright © 2015 Scientific & Academic Publishing. All Rights Reserved

monotone shape restriction using cubic splines and second-order cone programming with a truncated normal prior. Johnson (2007) estimates item response functions with free-knot regression splines restricted to be monotone by requiring spline coefficients to monotonically increasing. Neelon and Dunson (2004) proposed a piecewise linear model where the monotonicity is enforced via prior distributions; their model allows for flat spots in the regression function by using a prior that is a mixture of a continuous distribution and point mass at the origin. Bornkamp and Ickstadt (2009) applied their Bayesian monotonic regression model to dose-response curves. Lang and Brezger (2004) introduced Bayesian penalized splines for the additive regression model and Brezger and Steiner (2008) applied the Bayesian penalized splines model to monotone regression by imposing linear inequality constraints via truncated normal priors on the basis function coefficients to ensure monotonicity. Shively et al (2009) proposed two Bayesian approaches to monotone function estimation with one involving piecewise linear approximation and a Wiener process prior and the other involving regression spline estimation and a prior that is a mixture distribution of constrained normal distributions for the regression coefficients.

In this article, we propose a new method for estimating the GPLM based on Bayesian theorem using a new algorithm for estimation. The rest of the paper is organized as follows. In Section 2, we define the Generalized Partial Linear Model (GPLM). In Section 3, we present the Bayesian estimation and inference for the (GPLM). In Section 4, we provide Simulation Study. Finally, some concluding remarks and discussion are presented in Section 5.

## 2. Generalized Partial Linear Model (GPLM)

The GPLM model has the form

$$E[Y|X, W] = G[X\beta + m(W)], \quad (1)$$

where  $G(\cdot)$  is a known link function. It is a semi-parametric model since it contains both parametric and nonparametric components. This model can be reduced to the following model,

$$E[Y|X, W] = X\beta + m(W). \quad (2)$$

For a unit linear function. It is called partial linear model (PLM).

The GPLM is used by Severini and Staniswalis (1994); Chen (1995); Burda et al. (1998); Muller (2001); Peng (2003); and Peng and Wang (2004).

The estimation methods for the GPLM, in Eq. (1), are based on the idea that an estimator  $\hat{\beta}$  can be found for known  $m(\cdot)$ , and an estimator  $\hat{m}(\cdot)$  can be found for known  $\beta$ . The estimation methods that will be considered are based on kernel smoothing methods in the estimation of the nonparametric component of the model, therefore the

following estimation methods are presented in sequel.

### 2.1. Profile Likelihood Method

The profile likelihood method introduced by Severini and Wong (1992). It is based on assuming a parametric model for the conditional distribution of  $Y$  given  $X$  and  $W$ . The idea of this method is as follows:

**First:** Assume the parametric component of the model, i.e., the parameters vector,  $\beta$ .

**Second:** Estimate the nonparametric component of the model which depends on this fixed  $\beta$ , i.e.  $m_\beta(\cdot)$  by some type of smoothing method to obtain the estimator  $\hat{m}_\beta(\cdot)$ .

**Third:** Use the estimator  $\hat{m}_\beta(\cdot)$  to construct profile likelihood for the parametric component using either a true likelihood or quasi-likelihood function.

**Fourth:** The profile likelihood function is then used to obtain an estimator of the parametric component of the model using a maximum likelihood method.

Thus the profile likelihood method aims to separate the estimation problem into two parts, the parametric part which is estimated by a parametric method and the nonparametric part which is estimated by a nonparametric method.

Murphy and Vaart (2000) showed that the full likelihood method fails in semi-parametric models. In semi parametric models the observed information, if it exists, would be an infinite-dimensional operator. They used profile likelihood rather than a full likelihood to overcome the problem, the algorithm for profile likelihood method is derived as follows:

#### Derivation of different likelihood functions

For the parametric component of the model, the objective function is the parametric profile likelihood function which is maximized to obtain an estimator for  $\beta$  is given by

$$\ell(\beta) = \sum_{i=1}^n \ell(\mu_{i,\beta}, y_i) = \sum_{i=1}^n \ell_i, \quad (3)$$

where  $\ell(\cdot)$  denotes the log-likelihood or quasi-likelihood function,  $\mu_{i,\beta} = G[X_i^T \beta + m_\beta(W_i)]$ , and  $\ell_i = \ell(\mu_{i,\beta}, y_i)$ .

For the nonparametric component of the model, the objective function is a smoothed or a local likelihood function which is given by

$$\ell_H[m_\beta(w)] = \sum_{i=1}^n \ell(\mu_{i,\beta(w)}, y_i) K_H(w - W_i), \quad (4)$$

where  $\mu_{i,\beta(w)} = G[X_i^T \beta + m_\beta(W)] = G(\eta)$ , and the local weight  $K_H(w - W_i)$  is the kernel weight with  $\mathcal{K}$  denoting a multidimensional kernel function and  $H$  is a bandwidth matrix. The function in Eq. (4) is maximized to obtain an estimator for the smooth function  $m_\beta(w)$  at a point  $w$ .

#### Maximization of the likelihood functions

The maximization of the local likelihood in Eq. (4) requires solving

$$\mathbf{0} = \sum_{i=1}^n \ell(G(\eta), y_i) K_H(w - W_i), \quad (5)$$

with respect to  $m_\beta(t)$ . Note that  $\ell(\cdot)$  denotes the first derivative of the likelihood function  $\ell(\cdot)$ .

The maximization of the profile likelihood in Eq. (3) requires solving

$$\mathbf{0} = \sum_{i=1}^n \cdot \ell(G(\boldsymbol{\eta}), y_i) \left( \mathbf{U}_i + \nabla_{\mathbf{m}_\beta}(\mathbf{W}_i) \right), \quad (6)$$

with respect to the coefficient vector  $\boldsymbol{\beta}$ . The vector  $\nabla_{\mathbf{m}_\beta}(\mathbf{W}_i)$  denotes the vector of all partial derivatives of  $\mathbf{m}_\beta$  with respect to  $\boldsymbol{\beta}$ . A further differentiation of Eq. (5) with respect to  $p$  leads to an explicit expression for  $\nabla_{\mathbf{m}_\beta}$  as follows:

$$\nabla_{\mathbf{m}_\beta}(\mathbf{W}) = - \frac{\sum_{i=1}^n \cdot \ell(G(\boldsymbol{\eta}), y_i) \mathbf{K}_H(w - W_i) \mathbf{U}_i}{\sum_{i=1}^n \cdot \ell(G(\boldsymbol{\eta}), y_i) \mathbf{K}_H(w - W_i)}, \quad (7)$$

where  $\cdot \ell$  denotes the second derivative of  $\ell(G(\boldsymbol{\eta}), y_i)$ .

Equations (5) and (6) can only be solved iteratively. Severini and Saitniswalis (1994) presented a Newton-Raphson type algorithm for this problem as follows:

Let

$$\begin{aligned} \mathbf{m}_j &= \mathbf{m}_\beta(\mathbf{W}_j), \\ \boldsymbol{\eta}_i &= (\mathbf{X}_i^T \boldsymbol{\beta} + \mathbf{m}_i), \\ \boldsymbol{\eta}_{ij} &= (\mathbf{X}_i^T \boldsymbol{\beta} + \mathbf{m}_j), \\ \ell_i &= \ell(G(\boldsymbol{\eta}_i), y_i), \\ \ell_{ij} &= \ell(G(\boldsymbol{\eta}_{ij}), y_i). \end{aligned}$$

Further let  $\cdot \ell_i$ ,  $\cdot \ell_i$ ,  $\cdot \ell_{ij}$  and  $\cdot \ell_{ij}$  will be the first and second derivatives of  $\ell_i$  and  $\ell_{ij}$  with respect to their first argument. All values will be calculated at the observations  $W_i$  instead of the free parameter  $w$ . Then Equations (5) and (6) are transformed to

$$\mathbf{0} = \sum_{i=1}^n \cdot \ell_{ij} \mathcal{K}_H(\mathbf{W}_i - \mathbf{W}_j), \quad (8)$$

and

$$\mathbf{0} = \sum_{i=1}^n \cdot \ell_i (\mathbf{X}_i - \nabla_{\mathbf{m}_\beta}), \quad (9)$$

respectively.

The estimator of  $\nabla_{\mathbf{m}_\beta} = \nabla_{\mathbf{m}_\beta}(\mathbf{W}_j)$  based on Eq. (7) is necessary to estimate  $\boldsymbol{\beta}$

$$\nabla_{\mathbf{m}_\beta} = - \frac{\sum_{i=1}^n \cdot \ell_{ij} \mathcal{K}_H(\mathbf{W}_i - \mathbf{W}_j) \mathbf{U}_i}{\sum_{i=1}^n \cdot \ell_{ij} \mathcal{K}_H(\mathbf{W}_i - \mathbf{W}_j)}. \quad (10)$$

Equations (8) and (10) imply the following iterative Newton-Raphson Type algorithm

**First: Initial values**

Different strategies are found for obtaining start values of the estimators:

- Using  $\hat{\boldsymbol{\beta}}^{(0)}$  and  $\hat{\mathbf{m}}_j^{(0)}$  obtained from fitting a parametric generalized linear model (GLM).
- Using  $\boldsymbol{\beta} = \mathbf{0}$  and  $m_j = G^{-1}[(\mathbf{Y}_j + \bar{\mathbf{Y}})/2]$  with the adjustment for Binomial responses as  $m_j = G^{-1}[(\mathbf{Y}_j + 0.5)/2]$ .
- Using  $\boldsymbol{\beta} = \mathbf{0}$  and  $m_j = G^{-1}[\mathbf{Y}_j]$  with the adjustment for binomial responses as  $m_j = G^{-1}[(\mathbf{Y}_j + 0.5)/2]$ . (See Severini and Staniswalis, 1994).

**Second: The updating step for  $\boldsymbol{\beta}$**

$$\boldsymbol{\beta}^{\text{new}} = \boldsymbol{\beta} - \mathbf{B}^{-1} \sum_{i=1}^n \cdot \ell_i \tilde{\mathbf{X}}_i,$$

where  $\mathbf{B}$  is a Hessian type matrix defined as

$$\boldsymbol{\beta} = \sum_{i=1}^n \cdot \ell_i \tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_i^T,$$

and

$$\begin{aligned} \tilde{\mathbf{X}}_j &= \mathbf{X}_j + \nabla_{\mathbf{m}_j} = \mathbf{X}_j - \frac{\sum_{i=1}^n \cdot \ell_{ij} \mathcal{K}_H(\mathbf{W}_i - \mathbf{W}_j) \mathbf{X}_i}{\sum_{i=1}^n \cdot \ell_{ij} \mathcal{K}_H(\mathbf{W}_i - \mathbf{W}_j)} \\ & \quad j = 1, \dots, n \end{aligned}$$

The updating step for  $p$  can be summarized in a closed matrix form as follows:

$$\boldsymbol{\beta}^{\text{new}} = (\tilde{\mathbf{X}}^T \mathbf{D} \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \mathbf{D} \tilde{\mathbf{Z}},$$

where

$$\tilde{\mathbf{X}} = (\mathbf{I} - \mathbf{S}^p) \mathbf{U},$$

$$\tilde{\mathbf{Z}} = \tilde{\mathbf{X}} \boldsymbol{\beta} - \mathbf{D}^{-1} \mathbf{v}.$$

The matrix  $\mathbf{X}$  is the design matrix with rows  $\mathbf{X}_i^T$ ,  $\mathbf{I}$  is an  $(n \times n)$  identity matrix,

$$\mathbf{v} = (\cdot \ell_1, \dots, \cdot \ell_n), \quad (11)$$

$$\mathbf{D} = \text{diag}(\cdot \ell_1, \dots, \cdot \ell_n), \quad (12)$$

and  $\mathbf{S}^p$  is a smoother matrix with elements

$$(\mathbf{S}^p)_{ij} = \frac{\cdot \ell_{ij} \mathcal{K}_H(\mathbf{W}_i - \mathbf{W}_j)}{\sum_{i=1}^n \cdot \ell_{ij} \mathcal{K}_H(\mathbf{W}_i - \mathbf{W}_j)}. \quad (13)$$

**Third: The updating step for  $\mathbf{m}_j = \mathbf{m}_\beta(\mathbf{w}_j)$**

The function  $m_j = \mathbf{m}_\beta(\mathbf{w}_j)$  is updated by

$$\hat{\mathbf{m}}_j^{k+1} = \hat{\mathbf{m}}_j^k \frac{\sum_{i=1}^n \cdot \ell_{ij} \mathcal{K}_H(\mathbf{W}_i - \mathbf{W}_j)}{\sum_{i=1}^n \cdot \ell_{ij} \mathcal{K}_H(\mathbf{W}_i - \mathbf{W}_j)},$$

where  $k=0, 1, 2, \dots$  is the number of iteration.

It is noted that the function  $\cdot \ell_{ij}(\cdot)$  can be replaced by its expectation with respect to  $Y$  to obtain a Fisher scoring type algorithm (Severini and Staniswalis, 1994).

The previous procedure can be summarized as follows.

**Updating step for  $\boldsymbol{\beta}$**

$$\boldsymbol{\beta}^{\text{new}} = (\tilde{\mathbf{X}}^T \mathbf{D} \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \mathbf{D} \tilde{\mathbf{Z}}.$$

**Updating step for  $\mathbf{m}_j$**

$$\mathbf{m}_j^{\text{new}} = \mathbf{m}_j \frac{\sum_{i=1}^n \cdot \ell_{ij} \mathcal{K}_H(\mathbf{W}_i - \mathbf{W}_j)}{\sum_{i=1}^n \cdot \ell_{ij} \mathcal{K}_H(\mathbf{W}_i - \mathbf{W}_j)}.$$

Notes on the procedure:

1. The variable  $\tilde{\mathbf{Z}}$ , which is defined here, is a set of adjusted dependent variable.
2. The parameter  $\boldsymbol{\beta}$  is updated by a parametric method with a nonparametrically modified design matrix  $\tilde{\mathbf{X}}$ .
3. The function  $\cdot \ell_i$  can be replaced by its expectation, with respect to  $y$ , to obtain a Fisher scoring type procedure.
4. The updating step for  $\mathbf{m}_j$  is of quite complex structure and can be simplified in some models for identity and exponential link functions  $G$ .

**2.2. Generalized Speckman Method**

Generalized Speckman estimation method back to Speckman (1988). In the case of identity link function  $G$  and

normally distributed  $Y$ , the generalized Speckman and profile likelihood methods coincides with no updating steps are needed for the estimation of both  $\beta$  and  $m$ . This method can be summarized in the case of identity link (PLM) as follows:

- (1) Estimate  $p$  by:  $\hat{\beta} = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{Y}$
- (2) Estimate  $m$  by:  $\hat{m} = S(Y - X\hat{\beta})$ , where

$\tilde{X} = (1 - S)X$ ,  $\tilde{Y} = (1 - S)Y$ , and a smoother matrix  $S$  is defined by its elements as  $(S)_{ij} = \frac{\kappa_H(w_i - w_j)}{\sum_{i=1}^n \kappa_H(w_i - w_j)}$ .

This matrix is a simpler form of a smoother matrix, and differs from the one used in Eq. (13) where the matrix  $S$  yields  $\ell_i = -(Y_i - X_i^T \beta - m_j) / \sigma^2$  and  $\ell_i = -1 / \sigma^2$ , in the case of normally distributed  $Y$ .

For the GPLM, the Speckman estimator is combined with the IWLS method used in the estimation of GLM. As it is shown in IWLS each iteration step of GLM was obtained by WLS regression on the adjusted dependent variable. The same procedure will be used in the GPLM by replacing IWLS with a weighted partial linear fit on the adjusted dependent variable given by

$$Z = X\beta + m - D^{-1}v,$$

where  $v$  and  $D$  are defined as in Equations (11) and (12) respectively. The generalized Speckman algorithm for the GPLM can be summarized as:

**First: Initial values:**

The initial values used in this method are the same as in the previous profile likelihood algorithm.

**Second: Updating step for  $\beta$**

$$\beta^{\text{new}} = (\tilde{X}^T D \tilde{X})^{-1} \tilde{X}^T D \tilde{Z}.$$

**Third: Updating step for  $m$**

$$m^{\text{new}} = S(Z - X\beta),$$

Where

$$\begin{aligned} \tilde{U} &= (1 - S)X, \\ \tilde{Z} &= \tilde{X}\beta - D^{-1}v, \end{aligned}$$

The smoother matrix is used with elements:

$$(S)_{ij} = \frac{\ell_i \kappa_H(w_i - w_j)}{\sum_{i=1}^n \ell_i \kappa_H(w_i - w_j)} \quad (14)$$

There is a difference between the smoother matrix in Equation (14) and that in Equation (13). In Equation (13) we use  $\ell_i$  instead of  $\ell_{ij}$  that is used in Equation (14).

### 2.3. Back-fitting Method

Hastie and Tibshirani (1990) suggested the back-fitting method as an iterative algorithm to fit an additive model. The idea of this method is to regress the additive components separately on the partial residuals.

The back-fitting method will be presented in the case of identity link  $G$  (PLM) and non-monotone  $G$  (GPLM) as follows:

### Back-fitting algorithm for the GPLM

Back-fitting for the GPLM is an extension to that of PLM. The iterations in this method coincide with that in the Speckman method. This method differs from the Speckman method only in the parametric part. The back-fitting algorithm for the GPLM can be summarized as follows

**First: Initial values:**

This method often use

$$\beta = 0, \text{ and } m(w_j) = G^{-1}(\bar{Y})$$

**Second: Updating step for  $\beta$**

$$\beta^{\text{new}} = (\tilde{X}^T D \tilde{X})^{-1} \tilde{X}^T D \tilde{Z}.$$

**Third: Updating step for  $m$**

$$m^{\text{new}} = S(Z - X\beta),$$

where the matrices  $D$  and  $S$ , the vector  $v$ ,  $\tilde{X}$  and  $\tilde{Z}$  are defined as in the Speckman method for the GPLM (See Muller, 2001).

In practice, some of the predictor variables are correlated. Therefore, Hastie and Tibshirani (1990) proposed a modified back-fitting method which first search for a parametric solution and only fit the remaining parts non-parametrically.

## 2.4. Some Statistical Properties of the GPLM Estimators.

### (1) Statistical properties of the parametric component $\beta$

Under some regularity conditions the estimator  $\beta$  has the following properties:

1.  $\sqrt{n}$ -consistent estimator for  $\beta$ .
2. Asymptotically normal.
3. Its limiting covariance has a consistent estimator.
4. Asymptotically efficient; has asymptotically minimum variance (Severini and Staniswalis (1994).

### (2) Statistical properties of the non-parametric component $m$ :

The non-parametric function  $m$  can be estimated (in the univariate case) with the usual univariate rate of convergence. Severini and Staniswalis (1994) showed that the estimator  $\hat{m}_{\hat{\beta}}(\cdot)$  is consistent in supremum norm. They showed that the parametric and non-parametric estimators have the following asymptotic properties:

$$\hat{\beta} = \beta_0 + O_p\left(n^{-\frac{1}{2}}\right),$$

$$\|\hat{m}_{\hat{\beta}} - m_0\| = O_p\left(n^{-\frac{1}{4}}\right), \text{ as } n \rightarrow \infty,$$

$$\hat{\sigma}^2 = \sigma_0^2 + O_p(1),$$

where  $\beta_0, m_0, \sigma_0^2$  are the true parameter values so that

$$\text{var}(Y|X, W) = \hat{\sigma}_0^2 \text{var}[G(X\beta_0 + m_0(W))].$$

## 3. Bayesian Estimation and Inference for the GPLM

Bayesian inference derives the posterior distribution as a consequence of two antecedents; a prior probability and a

likelihood function derived from a probability model for the data to be observed. In Bayesian inference the posterior probability can be obtained according to Bayes theorem as follows

$$P(\boldsymbol{\beta}|\mathbf{y}) \propto \boldsymbol{\pi}(\boldsymbol{\beta})L(\mathbf{y}|\boldsymbol{\beta}), \tag{15}$$

where  $P(\boldsymbol{\beta}|\mathbf{y})$  is the posterior distribution,  $\boldsymbol{\pi}(\boldsymbol{\beta})$  is the prior distribution and  $L(\mathbf{y}|\boldsymbol{\beta})$  is the likelihood function.

**3.1. A proposed Algorithm for Estimating the GPLM Parameters**

1. Obtain the probability distribution of response variable,  $\underline{Y}$ .
2. Obtain the likelihood function of the probability distribution of response variable  $\underline{Y}$ .
3. Choose a suitable prior distribution of  $\boldsymbol{\beta}$ .
4. Use Eq. (15) to obtain the posterior distribution.
5. Obtain the Bayesian estimator under the square error loss function.
6. Replace the initial value of  $\boldsymbol{\beta}$  by the Bayesian estimator.
7. Use the profile likelihood method, generalized Speckman method and Back-fitting method with the new initial value of  $\boldsymbol{\beta}$  to estimate the GPLM

parameters.

**3.2. Bayesian approach for Estimating the GPLM Using Bayesian Estimator**

3.2.1.  $\sigma^2$  is assumed known

**Case 1**

Consider the GPLM in Eq. (1) and suppose that  $\underline{Y}$  belongs to the multivariate normal distribution with pdf

$$f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y_i - \boldsymbol{\beta}^T \mathbf{X}_i)^2}{2\sigma^2}\right\},$$

Then the likelihood function of the variable  $\underline{Y}$  can be written as

$$\prod_{i=1}^n f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}, \sigma^2) \propto \exp\left[\frac{-1}{2\sigma^2} \cdot \text{SSR}(\boldsymbol{\beta})\right], \tag{16}$$

where

$$\text{SSR}(\boldsymbol{\beta}) = (-2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}).$$

Let

$$\underline{\boldsymbol{\beta}} \sim \text{Multivariate Normal Distribution} (\underline{\boldsymbol{\mu}}, \underline{\boldsymbol{\Sigma}})$$

Then

$$f(\boldsymbol{\beta}) = (2\boldsymbol{\pi})^{-P/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \underline{\boldsymbol{\mu}})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \underline{\boldsymbol{\mu}})\right\}, \tag{17}$$

Combining (16) and (17) and using (15), we obtain the following posterior distribution (using some algebraic steps)

$$P(\boldsymbol{\beta}|\mathbf{Y}, \mathbf{X}, \sigma^2) \propto \exp\left[-\frac{1}{2}\left(\boldsymbol{\Sigma}^{-1} + \frac{\mathbf{X}^T \mathbf{X}}{\sigma^2}\right)^{-1} \left(\boldsymbol{\beta} - \left(\boldsymbol{\Sigma}^{-1} + \frac{\mathbf{X}^T \mathbf{X}}{\sigma^2}\right)^{-1} \left(\boldsymbol{\Sigma}^{-1} \underline{\boldsymbol{\mu}} + \frac{\mathbf{X}^T \mathbf{Y}}{\sigma^2}\right)\right)\right],$$

which is multivariate normal distribution with

$$\begin{aligned} \mathbf{E}(\boldsymbol{\beta}|\mathbf{Y}, \mathbf{X}, \sigma^2) &= \left(\boldsymbol{\Sigma}^{-1} + \mathbf{X}^T \mathbf{X} / \sigma^2\right)^{-1} \left(\boldsymbol{\Sigma}^{-1} \underline{\boldsymbol{\mu}} + \mathbf{X}^T \mathbf{Y} / \sigma^2\right), \\ \mathbf{Var}(\boldsymbol{\beta}|\mathbf{Y}, \mathbf{X}, \sigma^2) &= \left(\boldsymbol{\Sigma}^{-1} + \mathbf{X}^T \mathbf{X} / \sigma^2\right)^{-1}. \end{aligned}$$

Then, the Bayesian estimator under the square error loss function

$$\hat{\boldsymbol{\beta}}_{(\text{bayes})} = \left(\boldsymbol{\Sigma}^{-1} + \mathbf{X}^T \mathbf{X} / \sigma^2\right)^{-1} \left(\boldsymbol{\Sigma}^{-1} \underline{\boldsymbol{\mu}} + \mathbf{X}^T \mathbf{Y} / \sigma^2\right). \tag{18}$$

**Case 2**

Consider the GPLM in Eq. (1) and suppose that the pdf of  $\underline{Y}$  as follows

$$(2\boldsymbol{\pi})^{\frac{-p}{2}} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{X}_i - \boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1}(\mathbf{X}_i - \boldsymbol{\beta})\right\}.$$

Then the likelihood function of the probability distribution of  $\underline{Y}$  is

$$\begin{aligned} &((2\boldsymbol{\pi})^{\frac{-np}{2}} |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1}(\mathbf{X}_i - \boldsymbol{\beta})\right\}) \\ &= \exp\left\{-\frac{1}{2} \boldsymbol{\beta}^T \mathbf{n} \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{n} \boldsymbol{\Sigma}^{-1} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_{i,1}, \sum_{i=1}^n \mathbf{X}_{i,2}, \dots, \sum_{i=1}^n \mathbf{X}_{i,p}\right)^T\right\}. \end{aligned}$$

Let

$$\begin{aligned} \mathbf{A}_1 &= \mathbf{n}\boldsymbol{\Sigma}^{-1}, \\ \mathbf{b}_1 &= \mathbf{n}\boldsymbol{\Sigma}^{-1}\bar{\mathbf{X}}, \\ \bar{\mathbf{X}} &= \left( \frac{1}{\mathbf{n}} \sum_{i=1}^{\mathbf{n}} \mathbf{X}_{i,1}, \sum_{i=1}^{\mathbf{n}} \mathbf{X}_{i,2}, \dots, \sum_{i=1}^{\mathbf{n}} \mathbf{X}_{i,p} \right)^{\mathbf{T}}. \end{aligned}$$

Then, we can rewrite the likelihood function of as follows

$$\propto \exp\left\{-\frac{1}{2}\boldsymbol{\beta}^{\mathbf{T}}\mathbf{A}_1\boldsymbol{\beta} + \boldsymbol{\beta}^{\mathbf{T}}\mathbf{b}_1\right\}, \quad (19)$$

Let  $\underline{\boldsymbol{\beta}} \sim$  multivariate normal distribution  $(\boldsymbol{\mu}_0, \Lambda_0)$ ,

Then

$$\begin{aligned} \pi(\boldsymbol{\beta}) &= (2\boldsymbol{\pi})^{\frac{-\mathbf{p}}{2}} |\Lambda_0|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)^{\mathbf{T}}\Lambda_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)\right\} \\ &= (2\boldsymbol{\pi})^{\frac{-\mathbf{p}}{2}} |\Lambda_0|^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{\beta}^{\mathbf{T}}\Lambda_0^{-1}\underline{\boldsymbol{\beta}} + \boldsymbol{\beta}^{\mathbf{T}}\Lambda_0^{-1}\boldsymbol{\mu}_0 - \frac{1}{2}\boldsymbol{\mu}_0^{\mathbf{T}}\Lambda_0^{-1}\boldsymbol{\mu}_0\right\} \\ &\propto \exp\left\{-\frac{1}{2}\boldsymbol{\beta}^{\mathbf{T}}\Lambda_0^{-1}\underline{\boldsymbol{\beta}} + \underline{\boldsymbol{\beta}}^{\mathbf{T}}\Lambda_0^{-1}\boldsymbol{\mu}_0\right\} \\ &\propto \exp\left\{-\frac{1}{2}\boldsymbol{\beta}^{\mathbf{T}}\mathbf{A}_1\boldsymbol{\beta} + \boldsymbol{\beta}^{\mathbf{T}}\mathbf{b}_1\right\}, \end{aligned} \quad (20)$$

where

$$\begin{aligned} \mathbf{A}_1 &= \Lambda_0^{-1}, \\ \mathbf{b}_1 &= \Lambda_0^{-1}\boldsymbol{\mu}_0, \\ \mathbf{A}_0 &= \Lambda_0^{-1}, \\ \mathbf{b}_0 &= \Lambda_0^{-1}\boldsymbol{\mu}_0. \end{aligned}$$

Combining (19) and (20) and using (15), we obtain the following posterior distribution (using some algebraic steps)

$$\propto \exp\left\{-\frac{1}{2}\boldsymbol{\beta}^{\mathbf{T}}\mathbf{A}_1\boldsymbol{\beta} + \boldsymbol{\beta}^{\mathbf{T}}\mathbf{b}_1\right\} \times \exp\left\{-\frac{1}{2}\boldsymbol{\beta}^{\mathbf{T}}\mathbf{A}_1\boldsymbol{\beta} + \boldsymbol{\beta}^{\mathbf{T}}\mathbf{b}_1\right\} = \exp\left\{-\frac{1}{2}\boldsymbol{\beta}^{\mathbf{T}}\mathbf{A}_n\boldsymbol{\beta} + \boldsymbol{\beta}^{\mathbf{T}}\mathbf{b}_n\right\},$$

where

$$\begin{aligned} \mathbf{A}_n &= \mathbf{A}_0 + \mathbf{A}_1 = \Lambda_0^{-1} + \mathbf{n}\boldsymbol{\Sigma}^{-1}, \\ \mathbf{b}_n &= \mathbf{b}_0 + \mathbf{b}_1 = \Lambda_0^{-1}\boldsymbol{\mu}_0 + \mathbf{n}\boldsymbol{\Sigma}^{-1}\bar{\mathbf{X}}, \end{aligned}$$

which is multivariate normal distribution with

$$\mathbf{E}(\boldsymbol{\beta}|\mathbf{Y}, \mathbf{X}, \boldsymbol{\Sigma}) = \mathbf{A}_n^{-1} \mathbf{b}_n = \left(\Lambda_0^{-1} + \mathbf{n}\boldsymbol{\Sigma}^{-1}\right)^{-1} \left(\Lambda_0^{-1}\boldsymbol{\mu}_0 + \mathbf{n}\boldsymbol{\Sigma}^{-1}\bar{\mathbf{X}}\right)$$

and

$$\text{COV}(\boldsymbol{\beta}|\mathbf{Y}, \mathbf{X}, \boldsymbol{\Sigma}) = \Lambda_n = \left(\Lambda_0^{-1} + \mathbf{n}\boldsymbol{\Sigma}^{-1}\right)^{-1}.$$

Then, the Bayesian estimator under the square error loss function is

$$\hat{\boldsymbol{\beta}}_{(\text{bayes})} = \left(\Lambda_0^{-1} + \mathbf{n}\boldsymbol{\Sigma}^{-1}\right)^{-1} \left(\Lambda_0^{-1}\boldsymbol{\mu}_0 + \mathbf{n}\boldsymbol{\Sigma}^{-1}\bar{\mathbf{X}}\right) \quad (21)$$

3.2.2.  $\underline{\boldsymbol{\beta}}$  and  $\sigma^2$  is assumed unknown

### Case 1

Consider the GPLM in Eq. (1) and suppose that the marginal posterior distribution for  $\sigma^2$  is proportional with  $f(\sigma^2)$  and  $f(\mathbf{Y}|\mathbf{X}, \sigma^2)$  so that

$$\pi(\sigma^2|\mathbf{Y}, \mathbf{X}) \propto \pi(\sigma^2) \cdot \pi(\mathbf{Y}|\mathbf{X}, \sigma^2). \quad (22)$$

Hence

$$\pi(\mathbf{Y}|\mathbf{X}, \sigma^2) = \int \pi(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}, \sigma^2) \pi(\boldsymbol{\beta}|\mathbf{X}, \sigma^2) \, d\boldsymbol{\beta},$$

where

$$\begin{aligned} \pi(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}, \sigma^2) &= (2\pi\sigma^2)^{-n/2} \exp\left[\frac{-1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right] \\ \pi(\boldsymbol{\beta}|\mathbf{X}, \sigma^2) &= |2\pi\mathbf{g}\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}|^{-1} \exp\left[\frac{-1}{2\mathbf{g}\sigma^2} \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}\right]. \end{aligned}$$

Subsequently

$$\begin{aligned} \pi(\mathbf{Y}|\mathbf{X}, \sigma^2) &= \int \left\{ (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[\frac{-1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right] \right\} \left\{ |2\pi\mathbf{g}\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}|^{-1} \exp\left[\frac{-1}{2\mathbf{g}\sigma^2} \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}\right] \right\} d\boldsymbol{\beta} \\ \pi(\mathbf{Y}|\mathbf{X}, \sigma^2) &= \int \left\{ (2\pi\sigma^2)^{-\frac{n}{2}} \right\} \left\{ |2\pi\mathbf{g}\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}|^{-1} \right\} \exp\left\{ \frac{-1}{2\sigma^2} [(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} / \mathbf{g}] \right\} d\boldsymbol{\beta} \\ \pi(\mathbf{Y}|\mathbf{X}, \sigma^2) &= \int \left\{ (2\pi\sigma^2)^{-\frac{n}{2}} \right\} \left\{ |2\pi\mathbf{g}\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}|^{-1} \right\} \exp\left\{ \frac{-1}{2\sigma^2} \left[ \mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} \left(1 + \frac{1}{\mathbf{g}}\right) \right] \right\} d\boldsymbol{\beta} \\ \pi(\mathbf{Y}|\mathbf{X}, \sigma^2) &= \int \left\{ (2\pi\sigma^2)^{-\frac{n}{2}} \right\} \left\{ |2\pi\mathbf{g}\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}|^{-1} \right\} \exp\left\{ \frac{-1}{2\sigma^2} \mathbf{y}^T \mathbf{y} - \frac{1}{2} (\boldsymbol{\beta} - \mathbf{M})^T \mathbf{V}^{-1} (\boldsymbol{\beta} - \mathbf{M}) + \frac{1}{2} \mathbf{M}^T \mathbf{V}^{-1} \mathbf{M} \right\} d\boldsymbol{\beta}, \end{aligned}$$

where

$$\mathbf{M} = \frac{\mathbf{g}}{1+\mathbf{g}} (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \quad \mathbf{V} = \frac{\mathbf{g}}{1+\mathbf{g}} \sigma^2 (\mathbf{X}^T\mathbf{X})^{-1}, \quad \mathbf{g} = \frac{\mathbf{K}}{\sigma^2}.$$

$\mathbf{K}$  is any positive number, then we can rewrite the last expression as follows

$$\pi(\mathbf{Y}|\mathbf{X}, \sigma^2) = \left\{ (2\pi)^{-\frac{n}{2}} \right\} \times \left\{ (1 + \mathbf{g})^{-\frac{\mathbf{p}}{2}} \right\} \times \left\{ (\sigma^2)^{-\frac{n}{2}} \right\} \times \left\{ \exp\left[\frac{-1}{2\sigma^2} (\mathbf{SSR}_{\mathbf{g}})\right] \right\}. \tag{23}$$

Let

$$\gamma = \frac{1}{\sigma^2} \sim \text{Gamma}\left(\frac{\mathbf{v}_0}{2}, \frac{\mathbf{v}_0 \sigma_0^2}{2}\right)$$

Then

$$\pi(\gamma) \propto \left\{ \gamma^{\frac{\mathbf{v}_0}{2}-1} \exp\left[-\gamma \frac{\mathbf{v}_0 \sigma_0^2}{2}\right] \right\}, \tag{24}$$

Substituting form (24) and (25) into (23) we get

$$\pi(\gamma|\mathbf{Y}, \mathbf{X}) \propto \left\{ \gamma^{\frac{\mathbf{v}_0}{2}-1} \exp\left[-\gamma \frac{\mathbf{v}_0 \sigma_0^2}{2}\right] \right\} \left\{ \gamma^{\frac{\mathbf{n}}{2}} \exp\left[-\gamma \frac{(\mathbf{SSR}_{\mathbf{g}})}{2}\right] \right\}.$$

Therefore

$$\pi(\gamma|\mathbf{Y}, \mathbf{X}) \propto \left\{ \gamma^{\frac{\mathbf{v}_0+\mathbf{n}}{2}-1} \exp\left[-\frac{\gamma[\mathbf{v}_0 \sigma_0^2 + (\mathbf{SSR}_{\mathbf{g}})]}{2}\right] \right\}.$$

Then, the posterior distribution

$$\pi(\gamma|\mathbf{Y}, \mathbf{X}) \sim \text{Gamma}\left\{ \frac{\mathbf{v}_0 + \mathbf{n}}{2}, \frac{\mathbf{v}_0 \sigma_0^2 + (\mathbf{SSR}_{\mathbf{g}})}{2} \right\}$$

and the marginal posterior distribution for  $\sigma^2$  is

$$\pi(\sigma^2|\mathbf{Y}, \mathbf{X}) \sim \text{Inverse Gamma}\left\{ \frac{\mathbf{v}_0+\mathbf{n}}{2}, \frac{\mathbf{v}_0 \sigma_0^2 + (\mathbf{SSR}_{\mathbf{g}})}{2} \right\} \tag{25}$$

**Case 2**

Consider the GPLM in Eq. (1) and suppose that

$$f(y_i|\mathbf{X}_i, \boldsymbol{\beta}, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{(y_i - \boldsymbol{\beta}^T \mathbf{X}_i)^2}{2\sigma^2} \right\}.$$

Then the likelihood function of the pdf of response variable  $\underline{Y}$

$$\prod_{i=1}^n f(y_i|\mathbf{X}_i, \boldsymbol{\beta}, \sigma^2) \propto \exp\left[\frac{-1}{2\sigma^2} \cdot \mathbf{SSR}(\boldsymbol{\beta})\right], \tag{26}$$

where

$$\mathbf{SSR}(\boldsymbol{\beta}) = (-2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}).$$

Let

$$\underline{\beta} \sim \text{Multivariate Normal Distribution } (\mu_{\beta}, \sigma^2 V_{\beta}) \quad (27)$$

and

$$\sigma^2 \sim \text{Inverse Gamma } (\mathbf{a}, \mathbf{b}), \quad (28)$$

Then from (27) and (28) we get

$$\begin{aligned} \pi(\beta, \sigma^2) &= \mathbf{N}(\mu_{\beta}, \sigma^2 V_{\beta}) \times \mathbf{IG}(\mathbf{a}, \mathbf{b}) = \text{MNIG}(\mu_{\beta}, V_{\beta}, \mathbf{a}, \mathbf{b}), \\ \pi(\beta, \sigma^2) &= \frac{b^{\mathbf{a}}}{(2\pi)^{\frac{p}{2}} |V_{\beta}|^{\frac{1}{2}} \Gamma(\mathbf{a})} \left(\frac{1}{\sigma^2}\right)^{\frac{\mathbf{a}+\mathbf{p}}{2}+1} \times \exp\left[-\frac{1}{\sigma^2} \left\{ \mathbf{b} + \frac{1}{2} (\beta - \mu_{\beta})^T V_{\beta}^{-1} (\beta - \mu_{\beta}) \right\}\right] \\ \pi(\beta, \sigma^2) &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{\mathbf{a}+\mathbf{p}}{2}+1} \times \exp\left[-\frac{1}{\sigma^2} \left\{ \mathbf{b} + \frac{1}{2} (\beta - \mu_{\beta})^T V_{\beta}^{-1} (\beta - \mu_{\beta}) \right\}\right], \mathbf{a}, \mathbf{b}, \sigma^2 > 0 \end{aligned} \quad (29)$$

Combining (26) and (29), using (15) and after some algebraic steps, we obtain the following joint posterior distribution

$$\pi(\beta, \sigma^2 | \mathbf{Y}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{\mathbf{a}+\mathbf{p}}{2}+1} \exp\left[-\frac{1}{\sigma^2} \left\{ \mathbf{b}^* + \frac{1}{2} (\beta - \mu^*)^T V^{*-1} (\beta - \mu^*) \right\}\right], \quad (30)$$

where

$$\begin{aligned} \mu^* &= (V_{\beta}^{-1} + X^T X)^{-1} (V_{\beta}^{-1} \mu_{\beta} + X^T Y), \\ V^* &= (V^{-1} + X^T X)^{-1}, \\ \mathbf{a}^* &= \mathbf{a} + \mathbf{n}/2, \\ \mathbf{b}^* &= \mathbf{b} + \frac{1}{2} [\mu_{\beta}^T V_{\beta}^{-1} + Y^T Y + \mu^{*T} V^{*-1} \mu^*]. \end{aligned}$$

Then

$$\pi(\beta, \sigma^2 | \mathbf{Y}) \sim \text{NIG}(\mu^*, V^*, \mathbf{a}^*, \mathbf{b}^*),$$

which is normal inverse gamma distribution.

From (30), the marginal posterior distribution for  $\sigma^2$

$$\begin{aligned} \pi(\beta | \mathbf{Y}) &= \int \pi(\beta, \sigma^2 | \mathbf{Y}) d\beta = \int_{\beta} \text{NIG}(\mu^*, V^*, \mathbf{a}^*, \mathbf{b}^*) d\beta \\ \pi(\sigma^2 | \mathbf{Y}) &= \frac{b^{*\mathbf{a}^*}}{\Gamma(\mathbf{a}^*)} \left(\frac{1}{\sigma^2}\right)^{\mathbf{a}^*+1} \exp\left[-\frac{b^*}{\sigma^2}\right], \quad \pi(\sigma^2 | \mathbf{Y}) \sim \text{NIG}(\mathbf{a}^*, \mathbf{b}^*). \end{aligned} \quad (31)$$

From (30), the marginal posterior distribution for  $\beta$

$$\begin{aligned} \pi(\beta | \mathbf{Y}) &= \int \pi(\beta, \sigma^2 | \mathbf{Y}) d\sigma^2 = \int_{\sigma^2} \text{NIG}(\mu^*, V^*, \mathbf{a}^*, \mathbf{b}^*) d\sigma^2 \\ &\propto \int \left(\frac{1}{\sigma^2}\right)^{\frac{\mathbf{a}+\mathbf{n}}{2}+1} \times \exp\left[-\frac{1}{\sigma^2} \left\{ \mathbf{b}^* + \frac{1}{2} (\beta - \mu^*)^T V^{*-1} (\beta - \mu^*) \right\}\right] d\sigma^2 \\ &\propto \left\{ \mathbf{1} + \frac{(\beta - \mu^*)^T V^{*-1} (\beta - \mu^*)}{2b^*} \right\}^{-\left(\frac{\mathbf{a}^*+\mathbf{p}}{2}\right)}, \end{aligned} \quad (32)$$

which is multivariate t-distribution with the following pdf

$$\frac{\Gamma\left(\frac{\mathbf{v}^* + \mathbf{p}}{2}\right)}{\Gamma\left(\frac{\mathbf{v}^*}{2}\right) \pi^{\frac{\mathbf{p}}{2}} |\mathbf{v}^* \Sigma^*|^{\frac{1}{2}}} \left\{ \mathbf{1} + \frac{(\beta - \mu^*)^T \Sigma^{*-1} (\beta - \mu^*)}{\mathbf{v}^*} \right\}^{-\left(\frac{\mathbf{v}^* + \mathbf{p}}{2}\right)}$$

with expected value =  $\mathbf{0}$ , where

$$\begin{aligned} \mathbf{v}^* &= 2\mathbf{a}^* \\ \Sigma^* &= \left(\frac{\mathbf{b}^*}{\mathbf{a}^*}\right)^{\mathbf{v}^*}. \end{aligned}$$

Then, the Bayesian estimator under the square error loss function

$$\hat{\beta}_{(\text{bayes})} = \mathbf{0}. \quad (33)$$

3.2.3.  $\underline{\beta}$  is assumed known

**Case 1**

Consider the GPLM in Eq. (1) with likelihood function of the pdf of response variable  $Y$

$$\prod_{i=1}^n f(Y|X, \beta, \gamma) \propto \left\{ \gamma^{\frac{n}{2}} \exp \left[ -\gamma \frac{SSR(\beta)}{2} \right] \right\}. \tag{34}$$

Let

$$\gamma = \frac{1}{\sigma^2} \sim \text{Gamma} \left( \frac{v_0}{2}, \frac{v_0 \sigma_0^2}{2} \right).$$

Then

$$\pi(\gamma) \propto \left\{ \gamma^{\frac{v_0}{2}-1} \exp \left[ -\gamma \frac{v_0 \sigma_0^2}{2} \right] \right\}, \tag{35}$$

Combining (34) and (35) Using (15) and after some algebra, we obtain the following posterior distribution

$$\begin{aligned} \pi(\gamma|Y, X, \beta) &\propto \left\{ \gamma^{\frac{v_0+n}{2}-1} \exp \left[ -\gamma \frac{v_0 \sigma_0^2 + SSR(\beta)}{2} \right] \right\}, \\ (\sigma^2|Y, X, \beta) &\sim \text{Inverse Gamma} \left\{ \frac{v_0+n}{2}, \frac{v_0 \sigma_0^2 + SSR(\beta)}{2} \right\}. \end{aligned} \tag{36}$$

**Case 2**

Consider the GPLM in Eq. (1) with likelihood function of the pdf of response variable  $Y$

$$(2\pi)^{-\frac{np}{2}} |\Sigma|^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (X_i - \beta)^T \Sigma^{-1} (X_i - \beta) \right\}, \tag{37}$$

where

$$\begin{aligned} \sum_{i=1}^n (X_i - \beta)^T \Sigma^{-1} (X_i - \beta) &= \text{tr} \left( S_\beta \Sigma^{-1} \right), \\ S_\beta &= \sum_{i=1}^n (X_i - \beta)^T (X_i - \beta). \end{aligned}$$

Let

$$\pi(\Sigma) \sim \text{Inverse Wishert} \{ v_0, S_0^{-1} \}, \tag{38}$$

Combining (37) and (38) Using (15) and after some algebra, we obtain the following posterior distribution

$$\begin{aligned} &\propto \left[ |\Sigma|^{-\frac{(v_0+P+1)}{2}} \times \exp \left\{ -\text{tr} \left( \frac{S_0 \Sigma^{-1}}{2} \right) \right\} \right] \left[ |\Sigma|^{-n/2} \exp \left\{ -\frac{1}{2} \text{tr} (S_\beta \Sigma^{-1}) \right\} \right] \\ &= |\Sigma|^{-\frac{(v_0+P+n+1)}{2}} \exp \left\{ -\text{tr} [(S_0 + S_\beta) \Sigma^{-1}] / 2 \right\}, \end{aligned} \tag{39}$$

which is inverse Wishert distribution where

$$(\Sigma|Y, U, \beta) \sim \text{Inverse Wishert} \{ v_0 + n, (S_0 + S_\beta)^{-1} \}$$

with expected value

$$\frac{1}{v_0 - P + n + 1} (S_0 + S_\beta).$$

Then, the Bayesian estimator under the square error loss function

$$\hat{\Sigma}_{(bayes)} = \frac{v_0 - P - 1}{v_0 - P + n + 1} \frac{1}{v_0 - P - 1} S_0 + \frac{n}{v_0 - P + n + 1} \frac{1}{n} S_\beta$$

The previous results are summarized in Table (1).

**Table (1).** The Posterior Distribution Functions

Probability distribution of Y	Unknown parameter	Conjugate prior distribution	Posterior distribution	
Multivariate Normal Distribution	$\beta$	Multivariate Normal Distribution	Multivariate Normal Distribution	
Multivariate Normal Distribution	$(\sigma^2)^{-1}$	(GD)	(IGD)	
Multivariate Normal Distribution	$\beta$	(MND)	Joint distribution	(NIG)
	$\sigma^2$	(IG)	Marginal distribution for $\beta$	(MStD)
Multivariate Normal Distribution	$\sigma^2$	(IWD)	(IWD)	

(NIG)= Normal Inverse Gamma, (IGD) = Inverse Gamma, (MStD) = Multivariate Student t Distribution, (IWD) = Inverse Wishert Distribution, (GD) = Gamma Distribution.

### 4. Simulation Study

The aim of this simulation study is twofold. First, is to evaluate the performance of the proposed Bayesian estimators. Second, is to compare the proposed technique with classical approaches. The simulation study is based on 500 Monte Carlo replications. Different sample sizes have been used ranging from small, moderate, to large sizes. In specific sample sizes are fixed at n=50, n=100, n=200, n=500, n=1000, and n=2000. Also, different bandwidth parameters are used, namely, h=2, h=1, h=0.5, and h=0.2.

The estimation results and statistical analysis are obtained using statistical Package XploRe, 4.8 (XploRe, 2000). The methods are compared according to the following criteria.

First, the Average Mean of Squared Errors for ( $\hat{\beta}$ ),  $AMSE(\hat{\beta})$ , where

$$AMSE(\hat{\beta}) = n^{-1} \sum_{i=1}^n E\{\beta_i - \hat{\beta}_i\}^2.$$

The second is the Average Mean of Squared Errors for ( $\hat{m}$ ),  $AMSE(\hat{m})$ , where

$$AMSE(\hat{m}) = n^{-1} \sum_{i=1}^n E\{m_i - \hat{m}_i\}^2.$$

The third is the deviance where,

$$Deviance = -2 \text{ Log Likelihood.}$$

The results are shown in the following Tables (2) – (7).

**Table (2).** Simulation Results for n = 50

Criterion		n = 50					
		Classical Estimation			Bayesian Estimation		
		AMSE ( $\hat{\beta}$ )	AMSE ( $\hat{m}$ )	Mean Of Deviance	AMSE ( $\hat{\beta}$ )	AMSE ( $\hat{m}$ )	Mean Of Deviance
h=2	P- Likelihood	0.8271	0.8815	33.38	0.8215	0.8469	32.82
	G- Speckman	0.9473	0.7543	34.39	0.8849	0.7371	32.73
	Backfitting	0.7874	0.7332	33.32	0.71435	0.7276	30.37
h=1	P- Likelihood	0.8291	0.7803	45.12	0.8053	0.7049	43.64
	G- Speckman	0.9464	0.7161	45.39	0.8128	0.7088	43.11
	Backfitting	0.7859	0.7039	43.45	0.7544	0.6631	41.66
h=0.5	P- Likelihood	0.7712	0.6966	59.58	0.6649	0.6194	56.11
	G- Speckman	0.8981	0.6325	58.51	0.7146	0.6267	57.35
	Backfitting	0.7643	0.6120	57.81	0.6616	0.6120	56.01
h=0.2	P- Likelihood	0.7631	0.4929	68.35	0.6132	0.4052	64.98
	G- Speckman	0.7854	0.4341	68.59	0.6245	0.4181	66.11
	Backfitting	0.7562	0.4127	66.61	0.6087	0.3758	64.77

**Table (3).** Simulation Results for n = 100

Criterion  Bandwidth and model		n = 100					
		Classical Estimation			Bayesian Estimation		
		AMSE ( $\hat{\beta}$ )	AMSE ( $\hat{m}$ )	Mean Of Deviance	AMSE ( $\hat{\beta}$ )	AMSE ( $\hat{m}$ )	Mean Of Deviance
h=2	P- Likelihood	0.8043	0.8738	49.94	0.7313	0.7401	43.76
	G- Speckman	0.9144	0.8399	46.87	0.7377	0.7434	41.54
	Backfitting	0.7778	0.8232	44.35	0.7061	0.7322	40.12
h=1	P- Likelihood	0.8011	0.7268	55.89	0.6864	0.6746	54.41
	G- Speckman	0.9131	0.6351	57.69	0.6949	0.6459	54.36
	Backfitting	0.7749	0.6338	59.57	0.6533	0.5676	53.51
h=0.5	P- Likelihood	0.7810	0.5552	71.33	0.6155	0.4577	66.57
	G- Speckman	0.8612	0.5569	70.89	0.6147	0.4534	66.63
	Backfitting	0.7718	0.5435	70.85	0.6014	0.4293	64.22
h=0.2	P- Likelihood	0.7689	0.3346	78.66	0.5661	0.2489	75.91
	G- Speckman	0.7768	0.3473	79.75	0.5642	0.2433	75.11
	Backfitting	0.7661	0.3308	80.21	0.5457	0.2176	74.66

**Table (4).** Simulation Results for n = 200

Criterion  Bandwidth and model		n = 200					
		Classical Estimation			Bayesian Estimation		
		AMSE ( $\hat{\beta}$ )	AMSE ( $\hat{m}$ )	Mean Of Deviance	AMSE ( $\hat{\beta}$ )	AMSE ( $\hat{m}$ )	Mean Of Deviance
h=2	P- Likelihood	0.7523	0.8610	52.99	0.6553	0.6540	45.91
	G- Speckman	0.8643	0.7732	52.93	0.6611	0.6581	45.79
	Backfitting	0.7545	0.7954	52.76	0.6549	0.6521	45.76
h=1	P- Likelihood	0.7544	0.6961	70.81	0.5256	0.5133	66.43
	G- Speckman	0.8667	0.6016	73.01	0.5298	0.5216	67.51
	Backfitting	0.7528	0.6265	72.53	0.5332	0.5278	67.79
h=0.5	P- Likelihood	0.7465	0.4368	84.36	0.4989	0.3276	81.54
	G- Speckman	0.8389	0.5063	81.44	0.4959	0.3546	80.28
	Backfitting	0.7345	0.4829	80.75	0.4945	0.3258	80.21
h=0.2	P- Likelihood	0.7034	0.2342	90.21	0.4503	0.1839	88.11
	G- Speckman	0.8036	0.2973	91.33	0.4621	0.2059	89.38
	Backfitting	0.7043	0.2896	90.99	0.4635	0.2046	89.86

**Table (5).** Simulation Results for n = 500

Criterion		n = 500					
		Classical Estimation			Bayesian Estimation		
		AMSE ( $\hat{\beta}$ )	AMSE ( $\hat{m}$ )	Mean Of Deviance	AMSE ( $\hat{\beta}$ )	AMSE ( $\hat{m}$ )	Mean Of Deviance
h=2	P- Likelihood	0.5466	0.7067	163.92	0.4345	0.5822	160.34
	G- Speckman	0.6578	0.7149	164.88	0.4371	0.5876	162.37
	Backfitting	0.5756	0.7253	164.72	0.4356	0.5853	162.24
h=1	P- Likelihood	0.5538	0.5813	287.89	0.4059	0.3889	271.89
	G- Speckman	0.6736	0.5522	289.23	0.4036	0.3872	271.23
	Backfitting	0.5681	0.6265	289.57	0.4014	0.3855	271.13
h=0.5	P- Likelihood	0.5044	0.3332	402.63	0.2636	0.1812	388.34
	G- Speckman	0.5338	0.3861	401.69	0.2676	0.1852	388.58
	Backfitting	0.5143	0.3920	403.71	0.2634	0.1867	388.89
h=0.2	P- Likelihood	0.4734	0.1245	520.32	0.3136	0.0433	507.57
	G- Speckman	0.4912	0.1779	521.66	0.3255	0.0647	507.48
	Backfitting	0.4840	0.1511	520.97	0.3269	0.0689	507.61

**Table (6).** Simulation Results for n = 1000

Criterion		n = 1000					
		Classical Estimation			Bayesian Estimation		
		AMSE ( $\hat{\beta}$ )	AMSE ( $\hat{m}$ )	Mean Of Deviance	AMSE ( $\hat{\beta}$ )	AMSE ( $\hat{m}$ )	Mean Of Deviance
h=2	P- Likelihood	0.3354	0.4647	363.61	0.1943	0.3121	359.27
	G- Speckman	0.3360	0.4641	364.43	0.1977	0.3108	359.41
	Backfitting	0.3434	0.4699	364.24	0.2039	0.3183	359.71
h=1	P- Likelihood	0.2011	0.3011	587.70	0.0117	0.3011	584.76
	G- Speckman	0.2121	0.3206	589.40	0.0112	0.3216	584.43
	Backfitting	0.2162	0.3223	589.34	0.0132	0.3223	584.96
h=0.5	P- Likelihood	0.1429	0.1510	802.12	0.1029	0.1113	800.12
	G- Speckman	0.1456	0.1612	801.60	0.1056	0.1119	800.60
	Backfitting	0.1460	0.1633	803.11	0.1060	0.1143	800.91
h=0.2	P- Likelihood	0.0617	0.0715	1020.37	0.0358	0.0322	1018.37
	G- Speckman	0.0631	0.0734	1021.63	0.0326	0.0329	1018.63
	Backfitting	0.0635	0.0747	1022.97	0.0414	0.0330	1020.97

**Table (7).** Simulation Results for n=2000

Criterion		n = 2000					
		Classical Estimation			Bayesian Estimation		
		AMSE ( $\hat{\beta}$ )	AMSE ( $\hat{m}$ )	Mean Of Deviance	AMSE ( $\hat{\beta}$ )	AMSE ( $\hat{m}$ )	Mean Of Deviance
h=2	P- Likelihood	0.3354	0.4647	363.61	0.1943	0.3121	359.27
	G- Speckman	0.3360	0.4641	364.43	0.1977	0.3108	359.41
	Backfitting	0.3434	0.4699	364.24	0.2039	0.3183	359.71
h=1	P- Likelihood	0.2011	0.3011	587.70	0.0117	0.3011	584.76
	G- Speckman	0.2121	0.3206	589.40	0.0112	0.3216	584.43
	Backfitting	0.2162	0.3223	589.34	0.0132	0.3223	584.96
h=0.5	P- Likelihood	0.1429	0.1510	802.12	0.1029	0.1113	800.12
	G- Speckman	0.1456	0.1612	801.60	0.1056	0.1119	800.60
	Backfitting	0.1460	0.1633	803.11	0.1060	0.1143	800.91
h=0.2	P- Likelihood	0.0617	0.0715	1020.37	0.0358	0.0322	1018.37
	G- Speckman	0.0631	0.0734	1021.63	0.0326	0.0329	1018.63
	Backfitting	0.0635	0.0747	1022.97	0.0414	0.0330	1020.97

From the results we can conclude that

1. The Bayesian estimation for the GPLM outperforms the classical estimation for all sample sizes under the square error loss function. It gives high efficient estimators with smallest AMSE ( $\hat{\beta}$ ), AMSE ( $\hat{m}$ ) and Mean of Deviance.
2. The Bayesian estimation for the GPLM using the back-fitting method outperforms the profile likelihood method and the generalized Speckman method for  $n=50$  and  $n=100$ .
3. The Bayesian estimation for GPLM using the profile likelihood method and the generalized Speckman method outperforms the back-fitting method for  $n=1000$  and  $n=2000$ .

## 5. Discussions

In this article, we introduce a new Bayesian regression model called the Bayesian generalized partial linear model which extends the generalized partial linear model (GPLM). Bayesian estimation and inference of parameters for (GPLM) have been considered using some multivariate conjugate prior distributions under the square error loss function.

Simulation study is conducted to evaluate the performance of the proposed Bayesian estimators. Also, the simulation study is used to compare the proposed technique with classical approaches. The simulation study is based on 500 Monte Carlo replications. Different sample sizes have been used ranging from small, moderate, to large sizes. In specific sample sizes are fixed at  $n=50$ ,  $n=100$ ,  $n=200$ ,  $n=500$ ,  $n=1000$ , and  $n=2000$ . Also, different bandwidth parameters are used, namely,  $h=2$ ,  $h=1$ ,  $h=0.5$ , and  $h=0.2$ .

From the results of the simulation study, we can conclude the following. First, the Bayesian estimation for the GPLM outperforms the classical estimation for all sample sizes under the square error loss function. It gives high efficient estimators with smallest AMSE ( $\hat{\beta}$ ), AMSE ( $\hat{m}$ ) and Mean of Deviance. Second, the Bayesian estimation for the GPLM using the back-fitting method outperforms the profile likelihood method and the generalized Speckman method for  $n=50$  and  $n=100$ . The Bayesian estimation for GPLM using the profile likelihood method and the generalized Speckman method outperforms the back-fitting method for  $n=1000$  and  $n=2000$ . Finally, The Bayesian estimation of parameters for the GPLM gives small values of AMSE ( $\hat{\beta}$ ), AMSE ( $\hat{m}$ ) comparable to the classical estimation for all sample sizes.

---

## REFERENCES

- [1] Bornkamp, B., and Ickstadt, K. (2009) Bayesian nonparametric estimation of continuous monotone functions with applications to dose-response analysis, *Biometrics*, 65, 198–205.
- [2] Bouaziz, O., Geffray, S., and Lopez, O. (2015) Semiparametric inference for the recurrent events process by means of a single-index model, *Journal of Theoretical and Applied Statistics*, 49 (2), 361–385.
- [3] Burda, M., Hardel, W., Müller, M., and Weratz, A. (1994) Semiparametric analysis of German east-west migration: facts and theory, *Journal of Applied Econometrics*, 13 (5), 525–542.
- [4] Guo, X., Niu, G., Yang, Y., and Xu, W. (2015) Empirical likelihood for single index model with missing covariates at random, 49 (3), 588–601.
- [5] Holmes, C.C., and Heard, N. A. (2003) Generalized monotonic regression using random change points, *Statistics in Medicine*, 22, 623–638.
- [6] Johnson, M.S. (2007) Modeling dichotomous item responses with free-knot splines, *Journal of Computational Statistics & Data Analysis*, 51, 4178–4192.
- [7] Lang, S., and Brezger, A. (2004) Bayesian P-splines, *Journal of Computational and Graphical Statistics*, 13, 183–212.
- [8] Meyer, M.C., Hackstadt, A. J., and Hoeting, J. A. (2011) Bayesian estimation and inference for generalized partial linear models using shape-restricted splines, *Journal of Nonparametric Statistics*, 23, 867–884.
- [9] Müller, M. (2001) Estimation and testing in generalized partial linear models: a comparative study, *Statistics and Computing*, 11, 299–309.
- [10] Murphy, S.A., and Van Der Vaart, A.W. (2000) On profile likelihood, *Journal of the American Statistical Association*, 95, 449–485.
- [11] Neelon, B., and Dunson, D.B. (2004) Bayesian isotonic regression and trend analysis, *Biometrics*, 60, 498–506.
- [12] Peng, H. (2003) Efficient estimation in semi-parametric generalized linear models, *Proceedings of JSM 2003*.
- [13] Peng, H., and Wang, X., (2004) Moment estimation in semi-parametric generalized linear models, *Nonparametric Statistics*, 1–22.
- [14] Perron, F., and Mengersen, K. (2001) Bayesian nonparametric modeling using mixtures of triangular distributions, *Biometrics*, 57, 518–528.
- [15] Powel, L. (1994) Estimation of semi-parametric regression models, *Handbook of Econometrics*, 4, Ch. 41.
- [16] Brezger, A., and Steiner, W.J. (2008) Monotonic regression based on Bayesian P-splines: an application to estimating price response functions from store-level scanner data, *Journal of Business & Economic Statistics*, 26, 91–104.
- [17] Ramgopal, P., Laud, P.W., and Smith, A.F.M. (1993) Nonparametric Bayesian bioassay with prior constraints on the shape of the potency curve, *Biometrika*, 80, 489–498.
- [18] Ruppert, D., Wang, M.P., and Carroll, R.J., (2003) *Semi-parametric regression*, Cambridge University Press, London.
- [19] Severini, T.A., and Staniswalis, J.G., (1994) Quasi-likelihood estimation in semi-parametric models, *Journal of the American Statistical Association*, 89, 501–511.

- [20] Severini, T.A, and Wong, W.H., (1992) Profile likelihood and conditionally parametric models, *The Annals of Statistics*, 20, 1768-1802.
- [21] Shively, T.S., Sager, T.W., and Walker, S.G. (2009) A Bayesian approach to non-parametric monotone function estimation, *Journal of the Royal Statistical Society B*, 71, 159–175.
- [22] Sperlich, S., Hardle, W., and Aydinli, G., (2006) *The art of semi-parametric*, Physical-Verlag, Germany.
- [23] Wang, X. (2008) Bayesian free-knot monotone cubic spline regression, *Journal of Computational and Graphical Statistics*, 17, 373–387.
- [24] XploRe (2000), "XploRe-the interactive statistical computing environment" WWW: <http://www.xploRe-stat.de>.
- [25] Zhang, J., Wang, X., Yu, Y., and Gai, Y., (2010) Estimation and variable selection in partial linear single index models with error-prone linear covariates, *Journal of Theoretical and Applied Statistics*, 48 (5), 1048–1070.