Hacettepe Journal of Mathematics and Statistics
Volume 45 (1) (2016), 279 – 296

Bayesian analysis for semiparametric mixed-effects double regression models

Dengke Xu*†‡, Zhongzhan Zhang§ and Liucang Wu¶

Abstract

In recent years, based on jointly modeling the mean and variance, double regression models are widely used in practice. In order to assess the effects of continuous covariates or of time scales in a flexible way, a class of semiparametric mixed-effects double regression models(SMMEDRMs) is considered, in which we model the variance of the mixed effects directly as a function of the explanatory variables. In this paper, we propose a fully Bayesian inference for SMMEDRMs on the basis of B-spline estimates of nonparametric components. A computational efficient MCMC method which combines the Gibbs sampler and Metropolis-Hastings algorithm is implemented to simultaneously obtain the Bayesian estimates of unknown parameters and the smoothing function, as well as their standard deviation estimates. Finally, some simulation studies and a real example are used to illustrate the proposed methodology.

Keywords: Bayesian analysis, Semiparametric mixed-effects double regression models, Gibbs sampler, Metropolis-Hastings algorithm, B-spline.

2000 AMS Classification: 62F15, 62G05.

Received : 20.05.2014 *Accepted* : 13.12.2014 *Doi* : 10.15672/HJMS.2014458271

---

*Department of Statistics, Yunnan University, Kunming, People's Republic of China
†Department of Statistics, Zhejiang Agriculture and Forestry University, Hangzhou, People's Republic of China, Email: `xudengke1983@163.com`
‡Corresponding Author.
§College of Applied Sciences, Beijing University of Technology, Beijing, People's Republic of China, Email: `zzhang@bjut.edu.cn`
¶Faculty of Science, Kunming University of Science and Technology, Kunming, People's Republic of China, Email: `wuliucang@163.com`

1. Introduction

Many different approaches have been suggested to the problem of flexibly modeling of the mean. In statistical literature, compared with that of the mean, modeling of the variance has often been neglected. In many applications, particularly in the econometric area and industrial quality improvement experiments, modeling the variance will be of direct interest in its own right, to identify the source of variability in the observations, such as Taguchi-type experiments for robust design. On the other hand, modeling the variance itself may be of scientific interest. Thus, modeling of the variance can be as important as that of the mean. Furthermore, it is well known that efficient estimation of mean parameters in regression depends on correct modeling of the variance. The loss of efficiency may be substantial using constant variance models when the variance is varying. In addition, modeling of the variance is also necessary to obtain correct standard errors and confidence intervals, as well as for many other applications such as prediction and so on. Recently, the joint mean and variance models have been receiving a lot of attention. For example, Aitkin [1] provided maximum likelihood (ML) estimation for a joint mean and variance model and applied it to the commonly cited Minitab tree data. Xie et al. [22] investigated the score tests for homogeneity of a scalar parameter and a skewness parameter in skew-normal nonlinear regression models, which are included in the variance. Wu and Li [21] proposed a unified variable selection procedure which can simultaneously select significant variables in mean and dispersion models of the inverse Gaussian distribution. Zhao et al. [25] considered the issue of variable selection for beta regression models with varying dispersion, in which both the mean and the dispersion depend upon predictor variables. Wu [20] investigated the simultaneous variable selection in joint location and scale models of the skew-t-normal distribution when the dataset under consideration involves heavy tail and asymmetric outcomes. The similar works can be also seen from [12, 13, 24] and so on. On the other hand, semiparametric mixed models are useful extensions to linear mixed models and provide a flexible framework for analyzing longitudinal data. Many authors have studied semiparametric mixed models for longitudinal data (e.g., Ni et al. [15]). But, there is little work about the case in which the variance is additionally modelled. Therefore, in this paper we are interested in jointly modelling mean and variance of semiparametric mixed models.

Bayesian inference for the semiparametric mixed-effects models and the joint mean and variance models have also receiving a lot of attention in recent years. For example, Cepeda and Gamerman [2] summarized the Bayesian approach for modeling variance heterogeneity in normal regression analysis. Chen [3] proposed a fully Bayesian inference for semiparametric mixed-effects models of zero-inflated count data based on a data augmentation scheme that reflects both random effects of covariates and mixture of zero-inflated distribution. Chen and Tang [4] developed a Bayesian procedure for analyzing semiparametric reproductive dispersion mixed-effects models on the basis of P-spline estimates of nonparametric components. Lin and Wang [14] presented a fully Bayesian approach to multivariate regression models whose mean vector and scale covariance matrix are modelled jointly for analyzing longitudinal data. Tang and Duan [18] proposed a semiparametric Bayesian approach to generalized partial linear mixed models for longitudinal data. Xu and Zhang [23] proposed a fully Bayesian inference for semiparametric joint mean and variance models on the basis of B-spline approximations of nonparametric components. However, to the best of our knowledge, there is little work done for Bayesian analysis of semiparametric mixed-effects double regression models with longitudinal data, in which we model the variance of the mixed effects directly as a function of the explanatory variables.

On the other hand, various methods are available for fitting the semiparametric models, such as, the kernel smoothing method and the spline method. See for example,[5, 19, 26] and so on. Recently, the B-spline method is widely used to fit semiparametric models because of its advantages. Firstly, it does not need to estimate the nonparametric component of model point by point, that is, instead of concerning the local quality, the global quality is taken into consideration, which lead to the reduction of the computational complexity. Secondly, there are no boundary effects so that the splines can fit polynomial data exactly. Thirdly, the B-spline base functions have bounded supports and are numerically stable (Schumaker [17]).

Therefore, in this paper we extend the Bayesian methodology proposed in [2, 23] to fit semiparametric mixed-effects double regression models. Hence, a semiparametric Bayesian approach to SMMEDRMs is developed based on the B-spline approximation of nonparametric function and the hybrid algorithm combining the Gibbs sampler and Metropolis-Hastings algorithm in this article.

The outline of the paper is as follows. In Section 2 we first describe semiparametric mixed-effects double regression models. A Bayesian procedure based on a data augmentation scheme, Gibbs sampler and the Metropolis-Hastings algorithm for obtaining estimates is developed in Section 3. The full conditional distributions for implementing the sampling-based methods are also derived. To illustrate the proposed methodology, results obtained from some simulation studies are presented in Section 4. We further illustrate the proposed methodology through an analysis of the CD4 data in Section 5. The article is concluded with a brief discussion in Section 6.

2. Semiparametric Mixed-Effects Double Regression Models

Suppose that there are n independent subjects and the ith subject has m_i repeated measurements. Specifically, denote the response vector $Y_i = (Y_{i1}, \cdots, Y_{im_i})^T$ for the ith subject, $i = 1, \cdots, n$, which are observed at time $t_i = (t_{i1}, \cdots, t_{im_i})^T$. We assume that the response is normally distributed as $Y_{ij}|(X_{ij}, v_i, t_{ij}) \sim N(\mu_{ij}, \sigma^2)$. Here, the superscript T denotes the transposed of a vector (or matrix).

In this paper we consider

$$\left\{ \begin{array}{l} \mu_{ij} = X_{ij}^T\beta + v_i + g(t_{ij}), \\ i = 1, 2, \cdots, n, \\ j = 1, 2, \cdots, m_i, \end{array} \right. \tag{2.1}$$

where t_{ij} is a univariate observed covariate, $g(\cdot)$ is an unknown smooth function in the mean model, v_i is a random effect with $v_i \sim N(0, \sigma_i^2)$. Furthermore, if we have variance heterogeneity of the random effect, it is convenient to assume an explicit variance modeling related to some explanatory variables, that is:

$$\sigma_i^2 = h(Z_i, \gamma), \tag{2.2}$$

where $Z_i = (Z_{i1}, \cdots, Z_{iq})^T$ is the observation of explanatory variables associated with the variance of v_i and $\gamma = (\gamma_1, \cdots, \gamma_q)^T$ is a $q \times 1$ vector of regression coefficients in the variance model. Furthermore, we let $Z = (Z_1, Z_2, \cdots, Z_n)^T$. In addition, $h(\cdot, \cdot) > 0$ is a known function. Here two specific forms of $h(\cdot, \cdot)$ are usually taken to model varying variance: (i) log-linear model: $h(Z_i, \gamma) = \exp(\sum_{j=1}^q Z_{ij}\gamma_j)$; (ii) power product model: $h(Z_i, \gamma) = \prod_{j=1}^q Z_{ij}^{\gamma_j} = \exp(\sum_{j=1}^q \gamma_j \log Z_{ij})$. Of course, (ii) requires that the Z_{ij} is strictly positive, while no such restriction is needed for (i). In practice, one may make a choice of the variance weight $h(\cdot, \cdot)$, even a choice of the explanatory variables Z_i, according to the domain knowledge or modeling convenience. Therefore, in this article we consider the following semiparametric mixed-effects double regression models (SMMEDRMs):

$$
\begin{cases}
Y_{ij} = X_{ij}^T\beta + v_i + g(t_{ij}) + \varepsilon_{ij}, \\
\varepsilon_{ij} \sim N(0, \sigma^2), \\
v_i|Z_i \sim N(0, \sigma_i^2), \\
\sigma_i^2 = h(Z_i, \gamma), \\
i = 1, 2, \cdots, n, \\
j = 1, 2, \cdots, m_i,
\end{cases}
\tag{2.3}
$$

based on the independent observations $(Y_{ij}, X_{ij}, Z_i, t_{ij}), i = 1, 2, \cdots, n, j = 1, 2, \cdots, m_i$.

3. Bayesian Analysis of SMMEDRMs

3.1. B-splines for the Nonparametric Function. Without loss of generality, we assume that the covariate t_{ij} is valued on $[0, 1]$. Let $T = (t_1^T, t_2^T, \cdots, t_n^T)^T$. From the model (2.3), we obtain the likelihood function

$$
\begin{aligned}
L(\beta, \gamma, \phi^2, v|Y, X, Z, T) &= \prod_{i=1}^{n} \left\{ f(v_i|Z_i, \gamma) \prod_{j=1}^{m_i} f(Y_{ij}|X_{ij}, v_i, t_{ij}, \beta) \right\} \\
&\propto \left\{ \prod_{i=1}^{n} \sigma_i \right\}^{-1} (\phi^2)^{\frac{N}{2}} \exp\left\{ -\frac{\phi^2}{2} \sum_{i=1}^{n}\sum_{j=1}^{m_i} (Y_{ij} - X_{ij}^T\beta - v_i - g(t_{ij}))^2 - \sum_{i=1}^{n} \frac{v_i^2}{2\sigma_i^2} \right\},
\end{aligned}
\tag{3.1}
$$

where $\phi^2 = 1/\sigma^2$, $N = \sum_{i=1}^{n} m_i$, $v = (v_1, \cdots, v_n)^T$, $Y = (Y_1^T, \cdots, Y_n^T)^T$, $X = (X_1^T, \cdots, X_n^T)^T$, $X_i = (X_{i1}, \cdots, X_{im_i})^T$.

Since $g(\cdot)$ is nonparametric, (3.1) is not yet ready for optimization. So, we first use B-splines to approximate the nonparametric function $g(\cdot)$. Any computational algorithm developed for generalized linear models (GLM) can be used for fitting a semiparametric extension of GLM, since one can treat a nonparametric function as a linear function with the basis functions as covariates. For simplicity, let $0 = s_0 < s_1 < \cdots < s_{k_n} < s_{k_n+1} = 1$ be a partition of the interval [0, 1]. Using $\{s_i\}$ as the internal knots, we have $K = k_n + M$ normalized B-spline basis functions of order M that form a basis for the linear spline space. Selection of knots is generally an important aspect of spline smoothing. In this paper, similar to He et al. [10], the number of internal knots is taken to be the integer part of $N^{1/5}$. Thus $g(t)$ is approximated by $\pi^T(t)\alpha$, where $\pi(t) = (\pi_1(t), ..., \pi_K(t))^T$ is the vector of basis functions and $\alpha \in R^K$. With this notation, the mean model in (2.3) can be linearized as

$$
\mu_{ij} = x_{ij}^T\beta + v_i + \pi^T(t_{ij})\alpha. \tag{3.2}
$$

Hence, based on (3.2), the likelihood function (3.1) can be rewritten as follows:

$$
\begin{aligned}
L(\beta, \alpha, \gamma, \phi^2, v|Y, X, Z, T) &= \prod_{i=1}^{n} \left\{ f(v_i|Z_i, \gamma) \prod_{j=1}^{m_i} f(Y_{ij}|X_{ij}, v_i, t_{ij}, \beta) \right\} \\
&\propto \left\{ \prod_{i=1}^{n} \sigma_i \right\}^{-1} (\phi^2)^{\frac{N}{2}} \exp\left\{ -\frac{\phi^2}{2} \sum_{i=1}^{n}\sum_{j=1}^{m_i} (Y_{ij} - X_{ij}^T\beta - v_i - \pi^T(t_{ij})\alpha)^2 - \sum_{i=1}^{n} \frac{v_i^2}{2\sigma_i^2} \right\}.
\end{aligned}
\tag{3.3}
$$

3.2. Prior Density of Parameters. To implement a Bayesian approach to estimate the parameters of the models (2.3), we need to specify a prior distribution for the parameters involved. For simplicity, we suppose that β, α and γ are independent and normally distributed in prior as $\beta|\phi^2 \sim N(\beta_0, \phi^{-2}b_\beta)$, $\alpha \sim N(\alpha_0, \tau^2 I_K)$ and $\gamma \sim N(\gamma_0, B_\gamma)$, where the hyperparameters $\beta_0, \alpha_0, \gamma_0$, b_β and B_γ are assumed known, and τ^2 is assumed to be distributed as Gamma(a_τ, b_τ) with density function

$$p(\tau^2|a_\tau, b_\tau) \propto (\tau^2)^{a_\tau - 1} \exp\left(-b_\tau \tau^2\right),$$

where a_τ and b_τ are known positive constants. In addition, we also suppose that ϕ^2 is distributed in prior as Gamma(a_{ϕ^2}, b_{ϕ^2}), where a_{ϕ^2} and b_{ϕ^2} are known positive constants.

3.3. Gibbs Sampling and Conditional Distribution. Let $\theta = (\beta, \alpha, \gamma, \phi^2)$, $B_i = (\pi(t_{i1}), \cdots, \pi(t_{im_i}))^T$ and $B = (B_1^T, \cdots, B_n^T)^T$. Based on (3.3), we can sample from joint posterior distribution $p(\theta, v|Y, X, Z, T)$ by Gibbs sampling along the following process.

Step 1. Setting initial values of parameters as $\theta^{(0)} = (\beta^{(0)}, \alpha^{(0)}, \gamma^{(0)}, \phi^{2(0)})$.

Step 2. Based on $\theta^{(l)} = (\beta^{(l)}, \alpha^{(l)}, \gamma^{(l)}, \phi^{2(l)})$, compute $\Sigma^{(l)} = \mathrm{diag}\{h(Z_1, \gamma^{(l)}), \cdots, h(Z_n, \gamma^{(l)})\}$, $\tilde{v}_i^{(l)} = v_i^{(l)} \bigotimes 1_{m_i}$ and $\tilde{v}^{(l)} = ((\tilde{v}_1^{(l)})^T, \cdots, (\tilde{v}_n^{(l)})^T)^T$.

Step 3. Based on $\theta^{(l)} = (\beta^{(l)}, \alpha^{(l)}, \gamma^{(l)}, \phi^{2(l)})$, sample $\theta^{(l+1)} = (\beta^{(l+1)}, \alpha^{(l+1)}, \gamma^{(l+1)}, \phi^{2(l+1)})$, $v^{(l+1)}$ and $\tau^{2(l+1)}$ as follows:

- Sampling $\phi^{2(l+1)}$:

(3.4)
$$p(\phi^2|Y, X, v, \beta, \gamma, \alpha) \approx (\phi^2)^{\frac{N+p}{2} + a_{\phi^2} - 1} \exp\Bigg\{ -\phi^2 \Bigg[\frac{1}{2}(Y - X\beta^{(l)} - \tilde{v}^{(l)} - B\alpha^{(l)})^T (Y - X\beta^{(l)} - \tilde{v}^{(l)} - B\alpha^{(l)}) + \tfrac{1}{2}(\beta^{(l)} - \beta_0)^T(\beta^{(l)} - \beta_0) + b_{\phi^2} \Bigg] \Bigg\}.$$

- Sampling $\tau^{2(l+1)}$:

$$p(\tau^2|\alpha) \propto (\tau^2)^{-\frac{K}{2} - a_\tau - 1} \exp\left\{ -\frac{(\alpha^{(l)} - \alpha_0)^T(\alpha^{(l)} - \alpha_0) + 2b_\tau}{2\tau^2} \right\}. \tag{3.5}$$

- Sampling $\alpha^{(l+1)}$:

$$p(\alpha|Y, X, Z, T, \beta, \gamma, \tau^2, \phi^2) \propto \exp\left\{ -\frac{1}{2}(\alpha - \alpha_0^*)^T {b_\alpha^*}^{-1} (\alpha - \alpha_0^*) \right\}, \tag{3.6}$$

where $\alpha_0^* = b_\alpha^*({\tau^{2(l+1)}}^{-1} I_K \alpha_0 + \phi^{2(l+1)} B^T(Y - X\beta^{(l)} - \tilde{v}^{(l)}))$ and $b_\alpha^* = ({\tau^{2(l+1)}}^{-1} I_K + \phi^{2(l+1)} B^T B)^{-1}$, I_K is the identity matrix.

- Sampling $\beta^{(l+1)}$:

$$p(\beta|Y, X, Z, T, \alpha, \gamma, \phi^2) \propto \exp\left\{ -\frac{1}{2}(\beta - \beta_0^*)^T {b_\beta^*}^{-1} (\beta - \beta_0^*) \right\}, \tag{3.7}$$

where $\beta_0^* = b_\beta^*((\phi^{-2(l+1)} b_\beta)^{-1}\beta_0 + \phi^{2(l+1)} X^T(Y - \tilde{v}^{(l)} - B\alpha^{(l+1)}))$ and $b_\beta^* = ((\phi^{-2(l+1)} b_\beta)^{-1} + \phi^{2(l+1)} X^T X)^{-1}$.

- Sampling $v^{(l+1)}$:

$$\begin{aligned} p(v|Y, X, T, Z, \beta, \gamma, \phi^2) \quad \propto \quad & \exp\Bigg\{ -\frac{\phi^{2(l+1)}}{2} \sum_{i=1}^{n} \sum_{j=1}^{m_i} (Y_{ij} - X_{ij}^T \beta^{(l+1)} \\ & -\pi(t_{ij})^T \alpha^{(l+1)} - v_i)^2 - \sum_{i=1}^{n} \frac{v_i^2}{2\sigma_i^{2(l)}} \Bigg\}, \end{aligned} \tag{3.8}$$

where $\sigma_i^{2(l)} = h(Z_i, \gamma^{(l)})$.

• Sampling $\gamma^{(l+1)}$:

(3.9)
$$p(\gamma|Y,X,Z,\beta,\phi^2) \propto |\Sigma_1|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}v^{(l+1)^T}\Sigma_1 v^{(l+1)} - \frac{1}{2}(\gamma-\gamma_0)^T B_\gamma^{-1}(\gamma-\gamma_0)\right\}.$$

Here, $\Sigma_1 = \mathrm{diag}\{h(Z_1,\gamma),\cdots,h(Z_n,\gamma)\}$.

Step 4. Repeating Steps 2 and 3.

Then, we can generate sample series $(\beta^{(t)},\alpha^{(t)},\gamma^{(t)},\phi^{2(t)},\tau^{2(t)}), t=1,2,\cdots$ by the above program. It is easily seen from (3.4), (3.5), (3.6) and (3.7) that conditional distributions $p(\tau^2|\alpha), p(\alpha|Y,X,Z,T,\beta,\gamma,\tau^2,\phi^2), p(\beta|Y,X,Z,T,\alpha,\gamma,\phi^2)$ and $p(\phi^2|Y,X,T,v,\beta,\gamma,\alpha)$ are some familiar distributions, such as the Gamma and normal distributions. Sampling observations from these standard distributions is straightforward and fast. But conditional distributions $p(v|Y,X,Z,T,\beta,\gamma,\phi^2)$ and $p(\gamma|Y,X,Z,\beta,\phi^2)$ are some unfamiliar and rather complicated, thus drawing observations from the distributions are rather difficult. Hence, the commonly used Metropolis-Hastings algorithm is employed to sample observations from them. To this end, we choose normal distribution $N(v^{(l)},\sigma_v^2\Omega_v^{-1})$ and $N(\gamma^{(l)},\sigma_\gamma^2\Omega_\gamma^{-1})$ as the proposal distribution [11, 16], where σ_v^2 and σ_γ^2 are chosen such that the average acceptance rate is about between 0.25 and 0.45 (Gelman et al. [8]), and take

$$\Omega_v = E\left(-\frac{\partial^2 \log p(v|Y,X,T,Z,\beta^{(l+1)},\gamma^{(l)},\phi^{2(l+1)})}{\partial v\partial v^T}\right),$$
$$\Omega_\gamma = E\left(-\frac{\partial^2 \log p(\gamma|Y,X,Z,\beta^{(l+1)},\phi^{2(l+1)})}{\partial \gamma\partial \gamma^T}\right).$$

The Metropolis-Hastings algorithm is implemented as follows: at the $(l+1)$th iteration with the current value $v^{(l)}$, $\gamma^{(l)}$, new candidates v^* and γ^* are generated from $N(v^{(l)},\sigma_v^2\Omega_v^{-1})$, $N(\gamma^{(l)},\sigma_\gamma^2\Omega_\gamma^{-1})$ and are accepted respectively with probability

$$\min\left\{1,\frac{p(v^*|Y,X,Z,\beta,\gamma,\phi^2)}{p(v^{(l)}|Y,X,Z,\beta,\gamma,\phi^2)}\right\}$$

and

$$\min\left\{1,\frac{p(\gamma^*|Y,X,Z,\beta,\phi^2)}{p(\gamma^{(l)}|Y,X,Z,\beta,\phi^2)}\right\}.$$

3.4. Bayesian Inference. Observations generated from the above proposed computational procedure are used to obtain Bayesian estimates of parameters β,α,γ and ϕ^2 and their standard deviations.

Let $\{\theta^{(j)}=(\beta^{(j)},\alpha^{(j)},\gamma^{(j)},\phi^{2(j)}) : j=1,2,\cdots,J\}$ be the observations of $(\beta,\alpha,\gamma,\phi^2)$ generated from the joint conditional distribution $p(\beta,\alpha,\gamma,\phi^2|Y,X,Z,T)$ via the proposed hybrid algorithm. The Bayesian estimates of β,α,γ and ϕ^2 are given as:

$$\hat\beta=\frac{1}{J}\sum_{j=1}^J\beta^{(j)},\quad \hat\alpha=\frac{1}{J}\sum_{j=1}^J\alpha^{(j)},$$
$$\hat\gamma=\frac{1}{J}\sum_{j=1}^J\gamma^{(j)},\quad \hat\phi^2=\frac{1}{J}\sum_{j=1}^J\phi^{2(j)}.$$

As is shown by Geyer [9], $\hat\theta=(\hat\beta,\hat\alpha,\hat\gamma,\hat\phi^2)$ is a consistent estimate of the corresponding posterior mean vector as J goes to infinity. Similarly, a consistent estimate of the posterior covariance matrix $\text{Var}(\theta|Y, X, Z, T)$ can be obtained via the sample covariance matrix of the observations $\{\theta^{(j)} : j = 1, 2, \cdots, J\}$, that is

$$\widehat{\text{Var}}(\theta|Y, X, Z, T) = (J-1)^{-1} \sum_{j=1}^{J} (\theta^{(j)} - \hat{\theta})(\theta^{(j)} - \hat{\theta})^T.$$

Thus, the posterior standard deviations for the components can be obtained from the diagonal elements of the matrix.

4. Simulation Studies

In this section, some simulation studies are used to illustrate various aspects of the proposed Bayesian method. In the following simulations, $\sigma^2 = 0.5$ and the structure of the mean model is $\mu_{ij} = X_{ij}^T\beta + v_i + 0.5\sin(2\pi t_{ij}), i = 1, 2, \cdots, n, j = 1, 2, \cdots, m$, where $m = 4$, t_{ij} follows uniform distribution $U(0, 1)$, X_{ij} is a 3×1 vector with elements independently sampled from normal distribution $N(0, 1)$, and $\beta = (1, -0.8, 1)^T$. The structure of the variance model of the random effect v_i will be taken to be different models in the following examples.

To investigate sensitivity of Bayesian estimates to prior inputs, we consider the following three types of hyperparameter values for unknown parameters $\beta, \alpha, \gamma, \tau^2, \phi^2$:

Type I: $\beta_0 = (1, -0.8, 1)^T, b_\beta = I_3, \gamma_0 = (1, -0.5)^T, B_\gamma = I_2,\ a_\tau = 1, b_\tau = 1, a_{\phi^2} = 1, b_{\phi^2} = 1$. This can be regarded as a situation with good prior information.

Type II: $\beta_0 = 1.5 \times (1, -0.8, 1)^T, b_\beta = I_3, \gamma_0 = 1.5 \times (1, -0.5)^T, B_\gamma = I_2,\ a_\tau = 1, b_\tau = 1, a_{\phi^2} = 1, b_{\phi^2} = 1$. This can be regarded as a situation with inaccurate prior information.

Type III: $\beta_0 = (0, 0, 0)^T, b_\beta = I_3, \gamma_0 = (0, 0)^T, B_\gamma = I_2,\ a_\tau = 1, b_\tau = 1, a_{\phi^2} = 1, b_{\phi^2} = 1$. These hyperparameter values represent a situation with noninformative prior information.

For the above various settings, the preceding proposed hybrid algorithm combining the Gibbs sampler and the Metropolis-Hastings algorithm is used to evaluate the Bayesian estimates of unknown parameters and the smoothing function. In the following simulations, we use the cubic B-splines. Different sample sizes are employed in the simulations to show the effect of sample sizes. For each setting, 100 replications are carried out. For each data set generated in a replication, the convergence of the MCMC sampler is checked by estimated potential scale reduction (EPSR) value [7], and we observe that in all runs, the EPSR values are less than 1.2 after 4000 iterations. Observations are collected after 4000 iterations with $J = 4000$ in producing the Bayesian estimates for each replication.

4.1. Example 1: Comparisons for different prior inputs and sample sizes. In this example, we take the log-linear model as the structure of the variance model of the random effect v_i,

$$\log(\sigma_i^2) = Z_i^T\gamma, \quad i = 1, 2, \cdots, n$$

with $\gamma = (1, -0.5)^T$ and Z_i is a 2×1 vector with elements generated randomly from normal distribution $N(0, 1)$, n is the sample size ranging from n=30, 50, 100, 150. The summary of the simulation results for parameters is presented in Tables 1 and 2. To investigate accuracy of estimate of function $g(t)$, we plot the true value of function $g(t)$ against its estimates for three types of prior inputs under different sample sizes in Figures 1-4.

Table 1. Bayesian estimates of parameters under different priors when $n = 30$ and $n = 50$ in Example 1

Type	Parameters	$n = 30$ BIAS	$n = 30$ RMS	$n = 30$ SD	$n = 50$ BIAS	$n = 50$ RMS	$n = 50$ SD
I	β_1	0.0103	0.0769	0.0748	0.0028	0.0553	0.0577
	β_2	0.0041	0.0693	0.0743	0.0028	0.0542	0.0577
	β_3	0.0100	0.0731	0.0743	0.0055	0.0559	0.0573
	γ_1	0.0711	0.3301	0.3532	0.0093	0.2727	0.2649
	γ_2	0.0154	0.3218	0.3520	0.0141	0.2160	0.2576
	σ^2	0.0058	0.0704	0.0769	0.0065	0.0553	0.0588
II	β_1	0.0002	0.0799	0.0747	0.0037	0.0660	0.0579
	β_2	0.0019	0.0752	0.0749	0.0029	0.0627	0.0585
	β_3	0.0102	0.0717	0.0741	0.0038	0.0606	0.0579
	γ_1	0.0744	0.3294	0.3448	0.0227	0.2620	0.2727
	γ_2	0.0271	0.3318	0.3453	0.0526	0.2770	0.2641
	σ^2	0.0028	0.0695	0.0764	0.0107	0.0603	0.0595
III	β_1	0.0181	0.0708	0.0775	0.0009	0.0613	0.0578
	β_2	0.0163	0.0778	0.0765	0.0029	0.0481	0.0581
	β_3	0.0139	0.0786	0.0778	0.0053	0.0567	0.0580
	γ_1	0.1190	0.3099	0.3403	0.0139	0.2387	0.2518
	γ_2	0.0358	0.2739	0.3264	0.0132	0.2323	0.2467
	σ^2	0.0409	0.0851	0.0818	0.0194	0.0595	0.0605

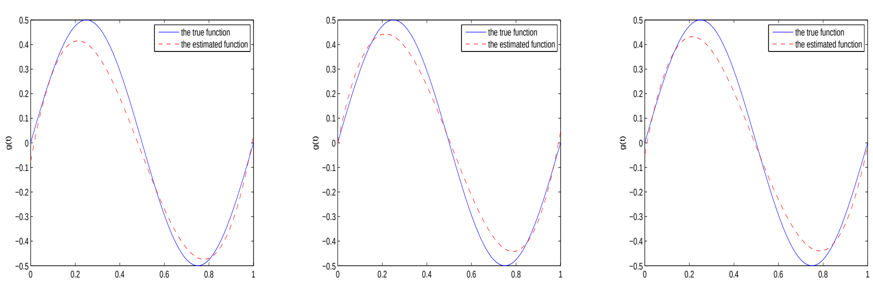

Figure 1. The average of the estimates versus the true value of $g(t)$ under three priors in Example 1: type I (left panel), type II (middle panel) and type III(right panel) when $n = 30$.

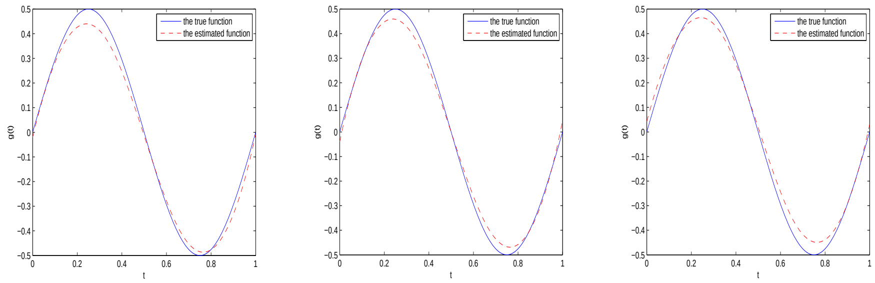

Figure 2. The average of the estimates versus the true value of $g(t)$ under three priors in Example 1: type I (left panel), type II (middle panel) and type III(right panel) when $n = 50$.

Table 2. Bayesian estimates of parameters under different priors when $n = 100$ and $n = 150$ in Example 1

		$n = 100$			$n = 150$		
Type	Parameters	BIAS	RMS	SD	BIAS	RMS	SD
I	β_1	0.0028	0.0394	0.0398	0.0019	0.0349	0.0323
	β_2	0.0005	0.0393	0.0398	0.0017	0.0340	0.0325
	β_3	0.0005	0.0394	0.0393	0.0012	0.0346	0.0323
	γ_1	0.0074	0.2050	0.1809	0.0290	0.1384	0.1449
	γ_2	0.0059	0.1642	0.1723	0.0057	0.1465	0.1394
	σ^2	0.0026	0.0405	0.0403	0.0036	0.0324	0.0332
II	β_1	0.0012	0.0450	0.0400	0.0005	0.0341	0.0323
	β_2	0.0055	0.0368	0.0398	0.0013	0.0319	0.0324
	β_3	0.0001	0.0389	0.0398	0.0014	0.0351	0.0323
	γ_1	0.0016	0.1507	0.1743	0.0281	0.1449	0.1434
	γ_2	0.0303	0.1777	0.1729	0.0113	0.1583	0.1366
	σ^2	0.0037	0.0409	0.0409	0.0020	0.0302	0.0328
III	β_1	0.0002	0.0379	0.0402	0.0036	0.0345	0.0322
	β_2	0.0008	0.0427	0.0401	0.0009	0.0323	0.0326
	β_3	0.0000	0.0398	0.0399	0.0028	0.0348	0.0323
	γ_1	0.0587	0.1764	0.1729	0.0018	0.1398	0.1400
	γ_2	0.0318	0.1857	0.1734	0.0016	0.1531	0.1367
	σ^2	0.0089	0.0380	0.0414	0.0060	0.0302	0.0332

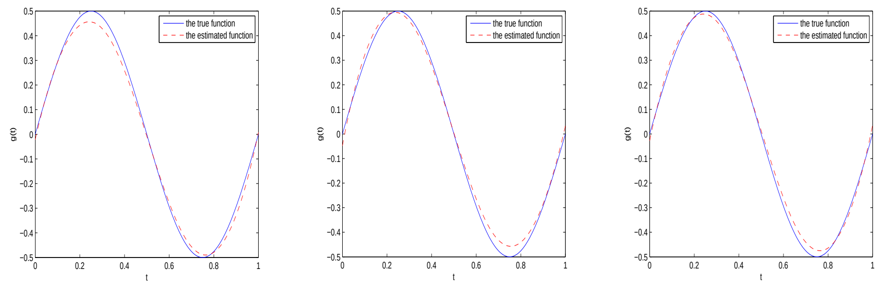

Figure 3. The average of the estimates versus the true value of $g(t)$ under three priors in Example 1: type I (left panel), type II (middle panel) and type III (right panel) when $n = 100$.

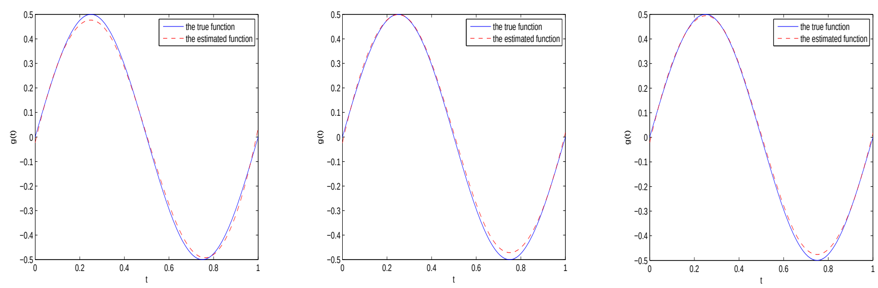

Figure 4. The average of the estimates versus the true value of $g(t)$ under three priors in Example 1: type I (left panel), type II (middle panel) and type III (right panel) when $n = 150$.

In Tables 1 and 2, “BIAS" denotes the absolute difference between the true value and the average of the Bayesian estimates of the parameters based on 100 replications, “SD" denotes the average of the estimated posterior standard deviation obtained from the formula in Section 3.4, and “RMS" denotes the root of mean square errors of the Bayesian estimates based on 100 replications. From Tables 1-2, we can make the following observations:

(i) the Bayesian estimates are reasonably accurate regardless of prior inputs in the sense of bias values of the estimates and their RMS values and SD values;

(ii) the estimates are mild sensitive to prior inputs for smaller sample size, but the infection clear away rapidly as the sample size goes large;

(iii) the estimates become better as the sample size increases, especially for the estimates of the parameters in the variance model.

Examination of Figures 1-4 shows that the shapes of the estimated nonparametric function are very close to the corresponding true line regardless of prior inputs. All in all, all the above findings show that the preceding proposed estimation procedures can well recover the true information in SMMEDRMs.

4.2. Example 2: Comparisons for different prior inputs and the different number of internal knots. To investigate the sensitivity of the Bayesian estimate for $g(t)$ to the selection of the number of internal knots, we consider the other two different choices of K in this example, i.e. $K_1 = \lfloor K_0/1.5 \rfloor$ and $K_2 = \lfloor 1.5K_0 \rfloor$, where K_0 is the optimal number of interior knots and $\lfloor u \rfloor$ denotes the largest integer not greater than u. To save space, here we only present the results of Bayesian estimates in Table 3 and Figures 5-6 for $n = 50$ under different choices of K.

Table 3. Bayesian estimates of parameters for different choices of K when $n = 50$ in Example 2

		$(n=50, K_1)$			$(n=50, K_2)$		
Type	Parameters	BIAS	RMS	SD	BIAS	RMS	SD
I	β_1	0.0091	0.0543	0.0572	0.0031	0.0606	0.0575
	β_2	0.0054	0.0612	0.0572	0.0022	0.0612	0.0573
	β_3	0.0001	0.0570	0.0577	0.0030	0.0522	0.0572
	γ_1	0.0198	0.2432	0.2624	0.0314	0.2589	0.2631
	γ_2	0.0145	0.2677	0.2579	0.0244	0.2377	0.2509
	σ^2	0.0102	0.0605	0.0598	0.0053	0.0643	0.0589
II	β_1	0.0052	0.0551	0.0570	0.0161	0.0566	0.0572
	β_2	0.0117	0.0523	0.0576	0.0020	0.0585	0.0573
	β_3	0.0023	0.0608	0.0567	0.0003	0.0540	0.0575
	γ_1	0.0719	0.2630	0.2681	0.0913	0.2571	0.2711
	γ_2	0.0467	0.2382	0.2529	0.0398	0.2303	0.2565
	σ^2	0.0077	0.0674	0.0586	0.0016	0.0590	0.0587
III	β_1	0.0092	0.0578	0.0585	0.0030	0.0568	0.0581
	β_2	0.0121	0.0575	0.0580	0.0042	0.0584	0.0581
	β_3	0.0014	0.0566	0.0586	0.0024	0.0524	0.0585
	γ_1	0.0476	0.2737	0.2540	0.0669	0.2323	0.2510
	γ_2	0.0267	0.2554	0.2634	0.0643	0.2392	0.2507
	σ^2	0.0249	0.0589	0.0610	0.0190	0.0624	0.0604

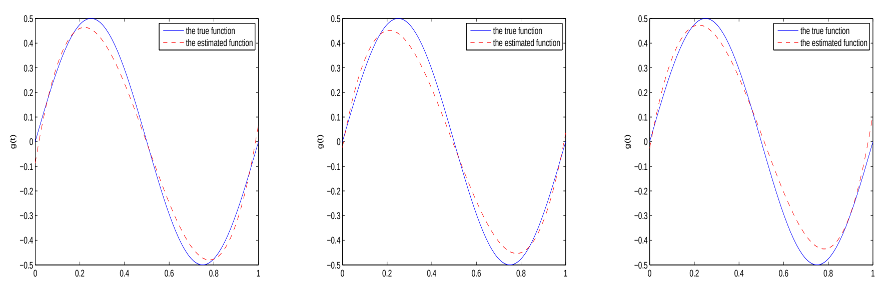

Figure 5. The average of the estimates versus the true value of $g(t)$ under three priors in Example 2: type I (left panel), type II (middle panel) and type III (right panel) for $n = 50$ and K_1.

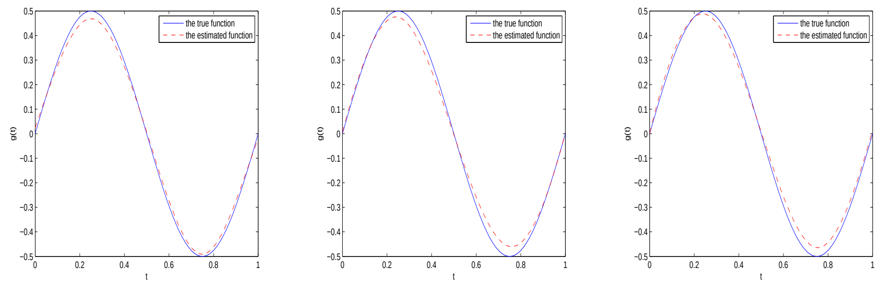

Figure 6. The average of the estimates versus the true value of $g(t)$ under three priors in Example 2: type I (left panel), type II (middle panel) and type III (right panel) for $n = 50$ and K_2.

By viewing Table 3 and comparing the results with Tables 1-2, we can see that the Bayesian estimates are reasonably accurate regardless of the values of K in the sense of their SD values and RMS values. From Figures 5-6, we can obtain that the shapes of the estimated nonparametric function are very similar to those in Figure 2. Therefore, the Bayesian estimates for parameter estimates and the nonparametric function $g(t)$ are not very sensitive to the selection of the number of internal knots.

4.3. Example 3: Comparisons for different prior inputs and different variance model. To investigate the sensitivity of the proposed Bayesian method to the structure of the variance model in SMMEDRMs, we consider the other common structure of the variance model of the random effect v_i (i.e. power product model), which can be defined as

$$\sigma_i^2 = \prod_{j=1}^{q} Z_{ij}^{\gamma_j}$$

with $\gamma = (1, -0.5)^T$ and Z_i is a 2×1 vector with elements generated randomly from uniform distribution $U(0, 2)$. The simulation results for the parameters and the nonparametric function are reported in Table 4 and Figures 7-8.

The results in Table 4 show that with using power product model as the variance structure, which is different with the variance model in example 1, the proposed Bayesian method also has the desired performance, which is substantively similar to the results in example 1.

In addition, to consider the effect of variance structure misspecification on parameter estimates, here we do some simulations with n=50 and n=100 under Type I. The main measurements for comparison are differences between the fitted mean parameters $\hat{\beta}$ and the true mean parameters β, the fitted variances $\hat{\sigma}_i^2 (i = 1, 2, \cdots, n)$ to the true variances $\sigma_i^2 (i = 1, 2, \cdots, n)$, and the fitted error variance $\hat{\sigma}^2$ to the true error variance σ^2. In particular, we define three relative errors:

$$RERR(\hat{\beta}) = \left| \frac{\sum_{j=1}^{p} (\hat{\beta}_j - \beta_j)}{\sum_{j=1}^{p} \beta_j} \right|; RERR(\hat{\sigma}_i^2) = \left| \frac{\sum_{i=1}^{n} (\hat{\sigma}_i^2 - \sigma_i^2)}{\sum_{i=1}^{n} \sigma_i^2} \right|; RERR(\hat{\sigma}^2) = \left| \frac{\hat{\sigma}^2 - \sigma^2}{\sigma^2} \right|.$$

Here variance structure misspecification means we use the variance structure in example 1 to model the variance of random effect. The results are reported in Table 5. From Table 5 we can find that when the true variance structure follows power product model, the errors in estimating $\hat{\beta}$, $\hat{\sigma}_i^2$ and $\hat{\sigma}^2$ increase when incorrectly modeling the variance using log-linear model. However, for this simulation study, variance model misspecification seems to affect the fitted results not larger, especially for the mean parameters and the error variance.

Table 4. Bayesian estimates of parameters under different priors in Example 3

		$n = 50$			$n = 100$		
Type	Parameters	BIAS	RMS	SD	BIAS	RMS	SD
I	β_1	0.0005	0.0566	0.0579	0.0025	0.0411	0.0399
	β_2	0.0029	0.0530	0.0572	0.0014	0.0431	0.0395
	β_3	0.0000	0.0494	0.0576	0.0055	0.0421	0.0398
	γ_1	0.0614	0.3186	0.3712	0.0368	0.2915	0.2398
	γ_2	0.0287	0.2471	0.2383	0.0001	0.1540	0.1611
	σ^2	0.0071	0.0592	0.0587	0.0021	0.0487	0.0405
II	β_1	0.0024	0.0572	0.0580	0.0004	0.0411	0.0399
	β_2	0.0063	0.0539	0.0574	0.0017	0.0432	0.0396
	β_3	0.0033	0.0494	0.0576	0.0067	0.0402	0.0398
	γ_1	0.1226	0.3470	0.3748	0.0722	0.3187	0.2375
	γ_2	0.0507	0.2515	0.2422	0.0062	0.1550	0.1623
	σ^2	0.0116	0.0597	0.0593	0.0032	0.0483	0.0405
III	β_1	0.0042	0.0589	0.0585	0.0050	0.0432	0.0405
	β_2	0.0005	0.0563	0.0580	0.0110	0.0469	0.0402
	β_3	0.0085	0.0512	0.0583	0.0068	0.0479	0.0401
	γ_1	0.0583	0.2874	0.3503	0.0493	0.2521	0.2416
	γ_2	0.0130	0.2423	0.2365	0.0663	0.1766	0.1581
	σ^2	0.0221	0.0627	0.0606	0.0154	0.0462	0.0416

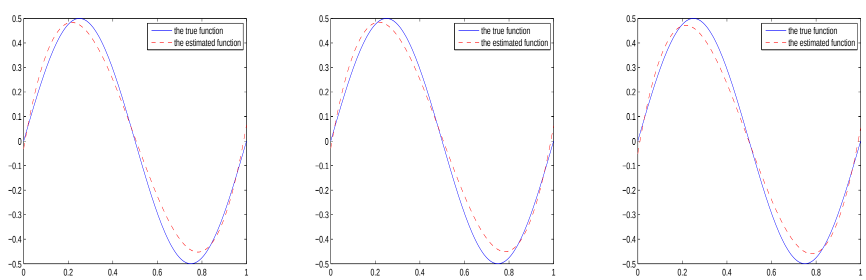

Figure 7. The average of the estimates versus the true value of $g(t)$ under three priors in Example 3: type I (left panel), type II (middle panel) and type III(right panel) when $n = 50$.

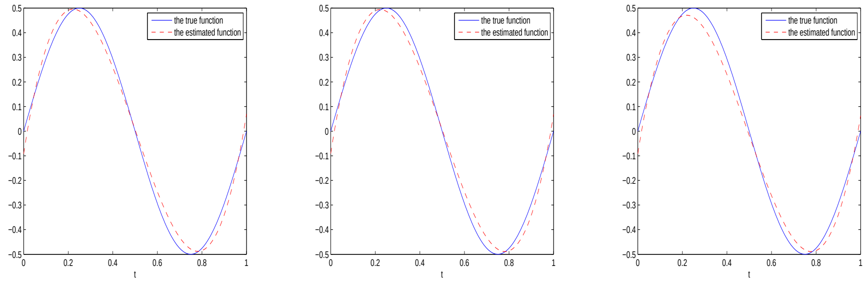

Figure 8. The average of the estimates versus the true value of $g(t)$ under three priors in Example 3: type I (left panel), type II (middle panel) and type III (right panel) when $n = 100$.

Table 5. Average of relative errors using different variance structures and sample size under Type I in Example 3

		$n = 50$	$n = 100$
correct specification	$RERR(\hat{\beta})$	0.0048	0.0013
	$RERR(\hat{\sigma}_i^2)$	0.9226	0.6868
	$RERR(\hat{\sigma}^2)$	0.0068	0.0003
misspecification	$RERR(\hat{\beta})$	0.0067	0.0024
	$RERR(\hat{\sigma}_i^2)$	2.4665	1.3371
	$RERR(\hat{\sigma}^2)$	0.0069	0.0007

5. Application to Real Data

In this section, we illustrate the proposed method through analysis of a data set from the MultiCenter AIDS Cohort study. The dataset contains the human immunodeficiency virus (HIV) status of 283 homosexual men who were infected with HIV during a follow-up period between 1984 and 1991. This dataset has been used by many authors to illustrate semiparametric linear regression models, such as [6, 26]. The objective of their analysis is to describe the trend of the mean CD4 percentage depletion over time and evaluate the effects of smoking, the pre-HIV infection CD4 percentage, and age at HIV infection on the mean CD4 percentage after infection. This motivates us to use the semiparametric models for this dataset.

Let Y be the individual's CD4 percentage, X_1 be the smoking status:(1 for a smoker and 0 for a nonsmoker), X_2 be the centered age at HIV infection, X_3 be the centered preCD4 percentage. To model jointly the mean for the CD4 cell data and the variance of random effect in the model, we use the following semiparametric mixed-effects double regression models:

$$\begin{cases} Y_{ij} = \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + v_i + g(t_{ij}) + \varepsilon_{ij}, \\ \varepsilon_{ij} \sim N(0, \sigma^2), \\ v_i \sim N(0, \sigma_i^2), \\ \log(\sigma_i^2) = \gamma_1 Z_{1i} + \gamma_2 Z_{2i}, \\ i = 1, 2, \cdots, 283, \end{cases}$$

where $Z_1 = X_1, Z_2 = X_3$, $g(t)$, the baseline CD4 percentage, represents the mean CD4 percentage t years after the infection.

The preceding proposed hybrid algorithm is used to obtain Bayesian estimates of β's, γ's and σ^2, where we use noninformative prior information for all unknown parameters. In the Metropolis-Hastings algorithm, we set $\sigma_\gamma^2 = 1.8$ and $\sigma_v^2 = 0.015$ in their corresponding proposal distributions, which give approximate acceptance rates 43.76% and 31.37%. To test the convergence of the algorithm, plot of the EPSR values for all the unknown parameters against iterations is presented in Figure 9, which indicates that the algorithm converges about 5000 iterations because EPSR values of all unknown parameters are less than 1.2 about 5000 iterations. We calculate Bayesian estimates (EST), standard deviation estimates (SD) of the Bayesian estimates of β's, γ's and σ^2. Results are given in Table 6, which indicate that X_3 has significant impact on the mean of Y and is somehow consistent with the results of variable selection seen in Fan and Li [6]. In addition, the curve of the estimated baseline function is shown in Figure 10. From Figure 10, we find that the mean baseline CD4 percentage decreases very quickly at the beginning of HIV infection, and the rate of decrease somewhat slows down four years after infection. The findings basically agree with that which was discovered by the local linear fitting method of Fan and Li [6].

Table 6. The real example: Bayesian estimates and their standard deviations

Parameter	EST	SD
β_1	0.4431	0.5775
β_2	-0.1955	0.2648
β_3	3.2706	0.2696
γ_1	2.0399	0.3593
γ_2	2.3333	0.2186
σ^2	80.9993	2.7830

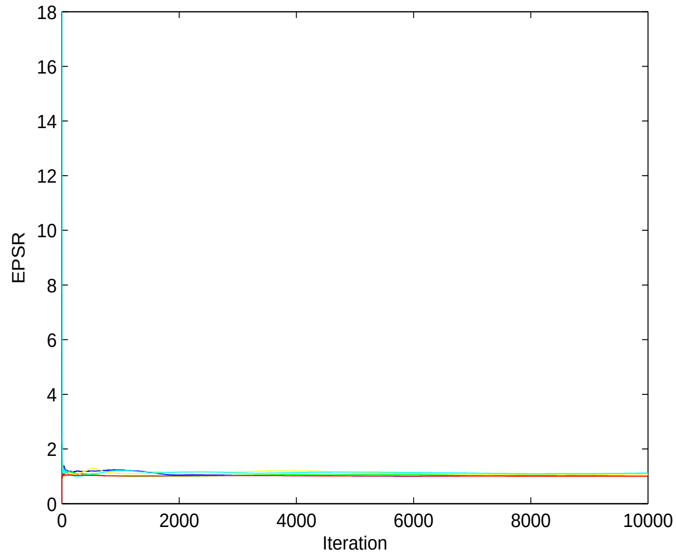

Figure 9. EPSR values of all parameters against iterations in the real example

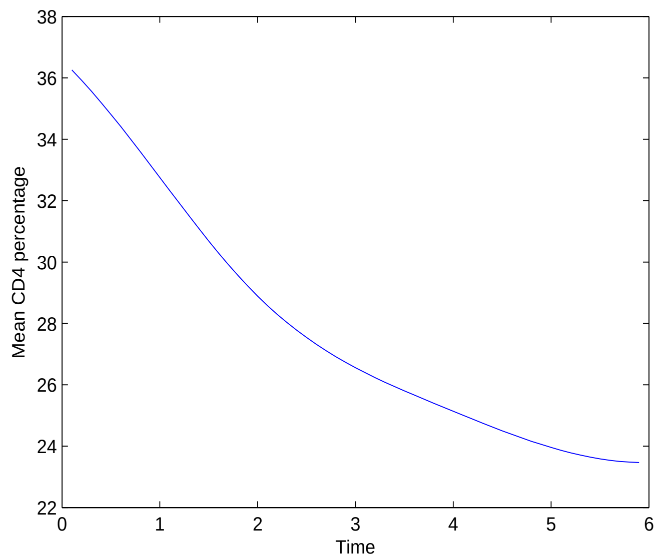

Figure 10. Application to AIDS data. The Bayesian estimate of the mean CD4 percentage $g(t)$. The solid line represents the estimated function.

6. Conclusion and Discussion

In this article, based on jointly modeling the mean and variance, we propose semi-parametric mixed-effects double regression models, in which we model the variance of the mixed effects directly as a function of the explanatory variables. Then we extend the Bayesian methodology proposed in [2, 23] to fit SMMEDRMs. A fully Bayesian approach is developed to analyze this models via B-spline estimate of the nonparametric part by combining the Gibbs sampler and Metropolis-Hastings algorithm. Some simulation studies and a real data are used to show the efficiency of the proposed Bayesian approach. The results show that the developed Bayesian method is highly efficient and computationally fast. A possible extension of the current model is being considered when covariates are missing under different missingness mechanisms.

Acknowledgements

The authors would like to thank the editor and the referees for helpful comments, which lead to improvement of an earlier version of the paper. This work is supported by grants from the National Natural Science Foundation of China (11301485, 11271039, 11261025, 11526188), Collaborative Innovation Center on Capital Social Construction and Social Management (006000546615539), China Postdoctoral Science Foundation funded project (2014M562348), National Statistical Science Research Project(2015LY70), and Zhejiang Provincial Natural Science Foundation of China(LQ15A010008).

References

[1] Aitkin, M. *Modelling variance heterogeneity in normal regression using GLIM*, Applied Statistics **36**, 332-339, 1987.

[2] Cepeda, E. and Gamerman, D. *Bayesian modeling of variance heterogeneity in normal regression models*, Brazilian Journal of Probability and Statistics **14**, 207-221, 2001.

[3] Chen, X. D. *Bayesian analysis of semiparametric mixed-effects models for zero-inflated count data*, Communications in Statistics−Theory and Methods **38**, 1815-1833, 2009.

[4] Chen, X. D. and Tang, N. S. *Bayesian analysis of semiparametric reproductive dispersion mixed-effects models*, Computational Statistics & Data Analysis **54**, 2145-2158, 2010.

[5] Fan, J. and Gijbels, I. *Local polynomial modelling and its application*, Chapman and Hall, New York, 1996.

[6] Fan, J. and Li, R. *New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis*, Journal of the American Statistical Association **99**, 710-723, 2004.

[7] Gelman, A. *Inference and monitoring convergence in markov chain monte carlo in practice*, Chapman and Hall, London, 1996.

[8] Gelman, A., Roberts, G. O. and Gilks, W. R. *Efficient metropolis jumping rules*, In:J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, (Eds), *Bayesian statistics*, vol. 5. Oxford University Press, Oxford, 599-607, 1995.

[9] Geyer, C. J. *Practical markov chain monte carlo*, Statistical Science **7**, 473-511, 1992.

[10] He, X., Fung, W. K. and Zhu, Z. Y. *Robust estimation in a generalized partial linear model for clustered data*, Journal of the American Statistical Association **100**, 1176-1184, 2005.

[11] Lee, S. Y. and Zhu, H. T. *Statistical analysis of nonlinear structural equation models with continuous and polytomous data*, British Journal of Mathematical and Statistical Psychology **53**, 209-232, 2000.

[12] Li, A. P., Chen, Z. X. and Xie, F. C. *Diagnostic analysis for heterogeneous log-Birnbaum-Saunders regression models*, Statistics & Probability Letters **82**, 1690-1698, 2012.

[13] Lin, J. G. and Wei, B. C. *Testing for heteroscedasticity in nonlinear regression models*, Communications in Statistics−Theory and Methods **32**, 171-192, 2003.

[14] Lin, T. I. and Wang, W. L. *Bayesian inference in joint modelling of location and scale parameters of the t distribution for longitudinal data*, Journal of Statistical Planning and Inference **141**, 1543-1553, 2011.

[15] Ni, X., Zhang, D. and Zhang, H. H. *Variable selection for semiparametric mixed models in longitudinal studies*, Biometrics **66**, 79-88, 2010.

[16] Roberts, G. *Markov chain concepts related to sampling algorithm*, Chapman and Hall, London, 1996.

[17] Schumaker, L. L. *Spline function*, Wiley, New York, 1981.

[18] Tang, N. S. and Duan, X. D. 2012. *A semiparametric bayesian approach to generalized partial linear mixed models for longitudinal data*, Computational Statistics & Data Analysis **56**, 4348-4365.

[19] Tian, R. Q. and Xue, L. G. *Variable selection for semiparametric errors-in-variables regression model with longitudinal data*, Journal of Statistical Computation and Simulation **84**, 1654-1669, 2014.

[20] Wu, L. C. *Variable selection in joint location and scale models of the skew-t-normal distribution*, Communications in Statistics −Simulation and Computation **43**, 615-630, 2014.

[21] Wu, L. C. and Li, H. Q. *Variable selection for joint mean and dispersion models of the inverse Gaussian distribution*, Metrika **75**, 795-808, 2012.

[22] Xie, F. C., Wei, B. C. and Lin, J. G. *Homogeneity diagnostics for skew-normal nonlinear regression models*, Statistics & Probability Letters **79**, 821-827, 2009.

[23] Xu, D. K. and Zhang, Z. Z. *A semiparametric Bayesian approach to joint mean and variance models*, Statistics & Probability Letters **83**, 1624-1631, 2013.

[24] Xu, D. K., Zhang, Z. Z. and Du, J. *Skew-normal semiparametric varying coefficient model and score test*, Journal of Statistical Computation and Simulation **85**, 216-234, 2015.

[25] Zhao, W. H., Zhang, R. Q., Lv, Y. Z. and Liu, J. C. *Variable selection for varying dispersion beta regression model*, Journal of Applied Statistics **41**, 95-108, 2014.

[26] Zhao, P. X. and Xue, L. G. *Variable selection for semiparametric varying coefficient partially linear errors-in-variables models*, Journal of Multivariate Analysis **101**, 1872-1883, 2010.