Statistics and Probability Letters 83 (2013) 1624–1631

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A semiparametric Bayesian approach to joint mean and variance models

Dengke Xu, Zhongzhan Zhang *

College of Applied Sciences, Beijing University of Technology, Beijing, 100124, China

ARTICLE INFO

Article history:
Received 24 October 2012
Received in revised form 28 February 2013
Accepted 28 February 2013
Available online 28 March 2013

MSC:
62F15
62G08

Keywords:
Bayesian analysis
Joint mean and variance models
Gibbs sampler
Metropolis–Hastings algorithm
B-spline

ABSTRACT

We propose a fully Bayesian inference for semiparametric joint mean and variance models on the basis of B-spline approximations of nonparametric components. An efficient MCMC method which combines Gibbs sampler and Metropolis–Hastings algorithm is suggested for the inference, and the methodology is illustrated through a simulation study and a real example.

1. Introduction

In classical linear models, the vector of responses $Y = (y_1, y_2, \ldots, y_n)^T$ is related to the mean vector $\mu = (\mu_1, \mu_2, \ldots, \mu_n)^T$ by the relationship

$$Y = \mu + \varepsilon, \tag{1}$$

where $\mu = X\beta$, $X = (x_1, x_2, \ldots, x_n)^T$ is an $n \times p$ matrix, whose ith row $x_i^T = (x_{i1}, \ldots, x_{ip})$ is the observation of explanatory variables associated with the mean μ_i of y_i, and $\beta = (\beta_1, \ldots, \beta_p)^T$ is a $p \times 1$ vector of unknown regression coefficients. Furthermore, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^T$ is the vector of independent errors with mean zero and variance σ^2. Here, the superscript T denotes the transposed of a vector (or matrix).

If we consider $\mu_i = x_i^T\beta + g(u_i)$, $i = 1, 2, \ldots, n$, where u_i is a univariate observed covariate, $g(\cdot)$ is an arbitrary unknown smooth function in the mean model, we have a semiparametric linear normal model with constant variance. Furthermore, if we have variance heterogeneity, it is convenient to assume an explicit variance modeling related to some explanatory variables, that is:

$$\sigma_i^2 = h(z_i^T\gamma), \tag{2}$$

where $z_i = (z_{i1}, \ldots, z_{iq})^T$ is the observation of explanatory variables associated with the variance of y_i and $\gamma = (\gamma_1, \ldots, \gamma_q)^T$ is a $q \times 1$ vector of regression coefficients in the variance model. Furthermore, we let $Z = (z_1, z_2, \ldots, z_n)^T$.

* Correspondence to: College of Applied Sciences, Beijing University of Technology, 100 Pingleyuan, Chaoyang District, Beijing, 100124, China. Tel.: +86 10 67391733; fax: +86 10 67391738.
E-mail addresses: zzhang@bjut.edu.cn, xudengke1983@emails.bjut.edu.cn (Z. Zhang).

0167-7152/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.spl.2013.02.023

There might be some components of z which coincide with components of x. In addition, $h(\cdot) > 0$ is a known function, and assumed to be a monotone function for the identifiability of the model. So, in this article we consider the following semiparametric joint mean and variance models (SEJMVMs):

$$\begin{cases} y_i \sim N(\mu_i, \sigma_i^2), \\ \mu_i = x_i^T\beta + g(u_i), \\ \sigma_i^2 = h(z_i^T\gamma), \\ i = 1, 2, \ldots, n. \end{cases} \tag{3}$$

based on the independent observations (y_i, x_i, z_i, u_i), $i = 1, 2, \ldots, n$.

The joint mean and variance models have received a lot of attention in recent years. For example, Park (1966) proposed a log-linear model for the variance parameter and described the Gaussian model using a two-stage procedure to estimate the parameters. Harvey (1976) discussed the maximum likelihood (ML) estimation of the mean and variance effects and the subsequent likelihood ratio test under general conditions. Aitkin (1987) provided ML estimation for a joint mean and variance model and applied it to the commonly cited Minitab tree data. Lin and Wei (2003) provided several diagnostic techniques for the test of variance heterogeneity in the framework of nonlinear regression models. Xie et al. (2009) 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. Xu and Zhang (2011) proposed a regularized REML method for simultaneous variable selection in heteroscedastic regression models. Zhang and Wang (2011) proposed a new criterion to simultaneously select explanatory variables in the mean model and variance model based on the model structure in heteroscedastic linear models. Li et al. (2012) introduced a class of heterogeneous log-Birnbaum–Saunders regression models where both the location parameter as well as the shape parameter can be modeled in terms of covariates, and studied the diagnostic measures based on the case-deletion approach and the local influence approach. Wu and Li (2012) proposed a unified variable selection procedure which can simultaneously select significant variables in mean and dispersion models of the inverse Gaussian distribution. 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, Fan and Gijbels (1996) and Zhao and Xue (2010) 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 leads 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, 1981).

In this article, we extend the Bayesian methodology proposed in Cepeda and Gamerman (2001) to fit semiparametric joint mean and variance models. Many authors have considered Bayesian inference for the semiparametric models and the joint mean and variance models. For example, Cepeda and Gamerman (2001) summarized the Bayesian approach for modeling variance heterogeneity in normal regression analysis (Aitkin, 1987). Chen (2009) 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 (2010) 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 (2011) presented a fully Bayesian approach to multivariate regression models whose mean vector and scale covariance matrix are modeled jointly for analyzing longitudinal data. However, to the best of our knowledge, there is little work done for Bayesian analysis of semiparametric joint mean and variance models. Hence, a semiparametric Bayesian approach to SEJMVMs 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 article is as follows. A Bayesian procedure based on a data augmentation scheme, Gibbs sampler and the Metropolis–Hastings algorithm for obtaining estimates is developed in Section 2. The full conditional distributions for implementing the sampling-based methods are also derived. To illustrate the proposed methodology, results obtained from a simulation study are presented in Section 3. We further illustrate the proposed methodology through an analysis of the Ragweed Pollen Level data in Section 4. The article is concluded with a brief discussion in Section 5.

2. Bayesian analysis of SEJMVMs

2.1. B-splines for the nonparametric function

Without loss of generality, we assume that the covariate u_i is valued on $[0, 1]$. Let $U = (u_1, u_2, \ldots, u_n)^T$. From model (3), we obtain the likelihood function

$$L(\beta, \gamma|Y, X, Z, U) \propto |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{(y_i - x_i^T\beta - g(u_i))^2}{h(z_i^T\gamma)}\right\}, \tag{4}$$

where $\Sigma = \mathrm{diag}\{h(z_1^T\gamma), \ldots, h(z_n^T\gamma)\}$.

Since $g(\cdot)$ is nonparametric, (4) 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. (2005), the number of internal knots is taken to be the integer part of $n^{1/5}$. Thus $g(u)$ is approximated by $\pi^T(u)\alpha$, where $\pi(u) = (\pi_1(u), \ldots, \pi_K(u))^T$ is the vector of basis functions and $\alpha \in R^K$. With this notation, the mean model in (3) can be linearized as

$$\mu_i = x_i^T\beta + \pi^T(u_i)\alpha. \tag{5}$$

Hence, based on (5), the likelihood function (4) can be rewritten as follows:

$$L(\beta, \alpha, \gamma|Y, X, Z, U) \propto |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(Y - X\beta - B\alpha)^T\Sigma^{-1}(Y - X\beta - B\alpha)\right\}, \tag{6}$$

where $B = (\pi(u_1), \pi(u_2), \ldots, \pi(u_n))^T$.

2.2. Prior density of parameters

To implement a Bayesian approach to estimate the parameters of model (3), we need to specify a prior distribution for the parameters involved. For simplicity, we suppose that β, γ and α are independent and normally distributed in prior as $\beta \sim N(\beta_0, b_\beta)$, $\alpha \sim N(\alpha_0, \tau^2 I_K)$, and $\gamma \sim N(\gamma_0, B_\gamma)$, where the hyperparameters β_0, α_0, γ_0, b_β and B_γ are assumed known, and τ^2 is assumed to be distributed as $IG(a_\tau, b_\tau)$ with density function

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

where a_τ and b_τ are known positive constants.

2.3. Gibbs sampling and conditional distribution

Let $\theta = (\beta, \alpha, \gamma)$. Based on (6), we can sample from joint posterior distribution $p(\theta|Y, X, Z, U)$ by Gibbs sampling along the following process.

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

Step 2. Based on $\theta^{(l)} = (\beta^{(l)}, \alpha^{(l)}, \gamma^{(l)})$, compute $\Sigma^{(l)} = \mathrm{diag}\{h(z_1^T\gamma^{(l)}), \ldots, h(z_n^T\gamma^{(l)})\}$.

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

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

$$p(\tau^2|\alpha) \propto (\tau^2)^{-\frac{K}{2}-a_\tau-1}\exp\left\{-\frac{(\alpha^{(l)} - \alpha_0)^T(\alpha^{(l)} - \alpha_0) + 2b_\tau}{2\tau^2}\right\}. \tag{7}$$

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

$$p(\alpha|Y, X, Z, U, \beta, \gamma, \tau^2) \propto \exp\left\{-\frac{1}{2}(\alpha - \alpha_0^*)^T{b_\alpha^*}^{-1}(\alpha - \alpha_0^*)\right\}, \tag{8}$$

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

$$p(\beta|Y, X, Z, U, \alpha, \gamma) \propto \exp\left\{-\frac{1}{2}(\beta - \beta_0^*)^T{b_\beta^*}^{-1}(\beta - \beta_0^*)\right\}, \tag{9}$$

  where $\beta_0^* = b_\beta^*(b_\beta^{-1}\beta_0 + X^T{\Sigma^{(l)}}^{-1}(Y - B\alpha^{(l+1)}))$ and $b_\beta^* = (b_\beta^{-1} + X^T{\Sigma^{(l)}}^{-1}X)^{-1}$.
- Sampling $\gamma^{(l+1)}$:

$$p(\gamma|Y, X, Z, U, \beta, \alpha) \propto |\Sigma|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(Y - X\beta^{(l+1)} - B\alpha^{(l+1)})^T\Sigma^{-1}(Y - X\beta^{(l+1)} - B\alpha^{(l+1)})\right.$$
$$\left. - \frac{1}{2}(\gamma - \gamma_0)^TB_\gamma^{-1}(\gamma - \gamma_0)\right\}. \tag{10}$$

  Here, $\Sigma = \mathrm{diag}\{h(z_1^T\gamma), \ldots, h(z_n^T\gamma)\}$.

Step 4. Repeating Steps 2 and 3.

Then, we can generate sample series $(\beta^{(t)}, \alpha^{(t)}, \gamma^{(t)}, \tau^{2(t)}), t = 1, 2, \ldots$ by the above program. It is easily seen from (7)–(9) that conditional distributions $p(\tau^2|\alpha)$, $p(\alpha|Y, X, Z, U, \beta, \gamma, \tau)$ and $p(\beta|Y, X, Z, U, \alpha, \gamma)$ are some familiar distributions, such as the Inverse Gamma and normal distributions. Sampling observations from these standard distributions is straightforward and fast. But conditional distribution $p(\gamma|Y, X, Z, U, \beta, \alpha)$ is some unfamiliar and rather complicated, thus drawing observations from the distribution is rather difficult. Hence, the commonly used Metropolis–Hastings algorithm is employed to sample observations from it. To this end, we choose normal distribution $N(\gamma^{(l)}, \sigma_\gamma^2 \Omega_\gamma^{-1})$ as the proposal distribution (Roberts, 1996; Lee and Zhu, 2000), where σ_γ^2 is chosen such that the average acceptance rate is about between 0.25 and 0.45 (Gelman et al., 1995), and take

$$\Omega_\gamma = \frac{1}{2}\sum_{i=1}^{n} \frac{(y_i - x_i^T\beta^{(l+1)} - \pi(u_i)^T\alpha^{(l+1)})^2}{h(z_i^T\gamma^{(l)})} z_i z_i^T + B_\gamma^{-1}.$$

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

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

2.4. Bayesian inference

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

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

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

As is shown by Geyer (1992), $\hat\theta = (\hat\beta, \hat\alpha, \hat\gamma)$ is a consistent estimate of the corresponding posterior mean vector as J goes to infinity. Similarly, a consistent estimate of the posterior covariance matrix $\mathrm{Var}(\theta|Y, X, Z, U)$ can be obtained via the sample covariance matrix of the observations $\{\theta^{(j)} : j = 1, 2, \ldots, J\}$, that is

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

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

3. Simulation study

In this section, a simulation study is used to illustrate various aspects of the proposed Bayesian method. In this simulation, the structure of the mean model is $\mu_i = x_i^T\beta + 0.5\sin(2\pi u_i)$, where u_i follows uniform distribution $U(0, 1)$, x_i is a 3×1 vector with elements independently sampled from uniform distribution $U(-1, 1)$, and $\beta = (1, -0.5, 0.5)^T$. The structure of the variance model is $\log(\sigma_i^2) = z_i^T\gamma$ with $\gamma = (1, -0.5, 0.5)^T$ and z_i is a 3×1 vector with elements generated randomly from uniform distribution $U(-1, 1)$. In the following simulations, we use the cubic B-splines.

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

Type I: $\beta_0 = (0, 0, 0)^T, b_\beta = I_3, \gamma_0 = (0, 0, 0)^T, B_\gamma = I_3, \alpha_0 = (0, \ldots, 0)^T, a_\tau = 1, b_\tau = 1$. These hyperparameter values represent a situation with noninformative prior information.

Type II: $\beta_0 = (1, -0.5, 0.5)^T, b_\beta = I_3, \gamma_0 = (1, -0.5, 0.5)^T, B_\gamma = I_3, \alpha_0 = (0, \ldots, 0)^T, a_\tau = 1, b_\tau = 1$. This can be regarded as a situation with good prior information.

Type III: $\beta_0 = 3 \times (1, -0.5, 0.5)^T, b_\beta = I_3, \gamma_0 = 3 \times (1, -0.5, 0.5)^T, B_\gamma = I_3, \alpha_0 = (0, \ldots, 0)^T, a_\tau = 1, b_\tau = 1$. This can be regarded as a situation with inaccurate 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. $n = 70$ and $n = 150$ are employed in the simulations to show the effect of sample size. 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 (Gelman, 1996), and we observe that in all runs, the EPSR values are less than 1.2 after 5000 iterations. Observations are collected after 5000 iterations with $J = 5000$ in producing the Bayesian estimates for each replication. The summary of the simulation results for parameters is presented in Table 1. To investigate the accuracy of estimate of function $g(u)$, we plot the true value of function $g(u)$ against its estimates for three types of prior inputs under different sample sizes in Figs. 1 and 2.

Table 1
Bayesian estimates of parameters under different priors in simulation study.

Type	Parameters	$n = 70$			$n = 150$		
		EST	SD	RMS	EST	SD	RMS
I	β_1	0.9848	0.2021	0.1811	0.9937	0.1303	0.1418
	β_2	−0.4711	0.1990	0.1825	−0.4861	0.1312	0.1393
	β_3	0.4455	0.1991	0.2050	0.4865	0.1303	0.1399
	γ_1	0.8808	0.3322	0.3072	0.9579	0.2116	0.2335
	γ_2	−0.4474	0.3329	0.2753	−0.4645	0.2132	0.2029
	γ_3	0.5129	0.3317	0.2771	0.5034	0.2140	0.2080
II	β_1	1.0063	0.1937	0.2153	1.0116	0.1293	0.1336
	β_2	−0.5051	0.1936	0.2028	−0.4949	0.1296	0.1251
	β_3	0.5098	0.1930	0.1952	0.4750	0.1302	0.1416
	γ_1	0.9961	0.3302	0.3018	0.9945	0.2141	0.2134
	γ_2	−0.5119	0.3329	0.3139	−0.5418	0.2161	0.2001
	γ_3	0.4978	0.3268	0.3039	0.5100	0.2137	0.2308
III	β_1	1.0625	0.1883	0.2158	1.0177	0.1275	0.1305
	β_2	−0.5624	0.1924	0.1998	−0.5138	0.1271	0.1381
	β_3	0.5464	0.1875	0.1960	0.4918	0.1264	0.1477
	γ_1	1.2084	0.3357	0.3774	1.0893	0.2143	0.2442
	γ_2	−0.6243	0.3333	0.3305	−0.5246	0.2147	0.2136
	γ_3	0.6227	0.3335	0.3190	0.5342	0.2165	0.2091

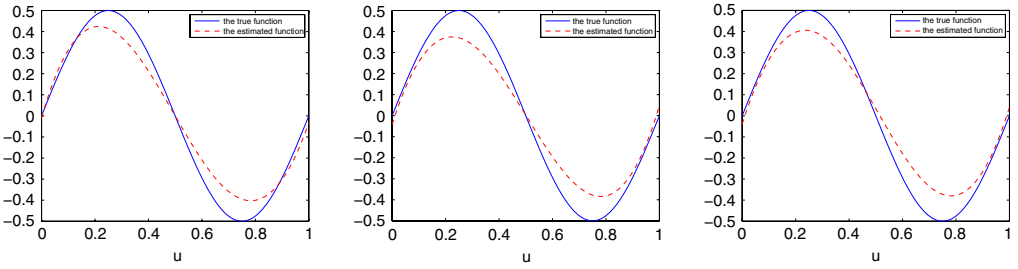

Fig. 1. The average of the estimates versus the true value of $g(u)$ under three priors: type I (left panel), type II (middle panel) and type III (right panel) when $n = 70$.

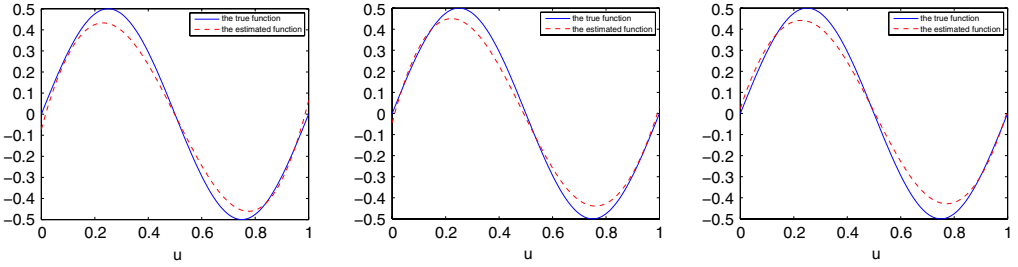

Fig. 2. The average of the estimates versus the true value of $g(u)$ under three priors: type I (left panel), type II (middle panel) and type III (right panel) when $n = 150$.

In Table 1, "EST" denotes 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 2.4, and "RMS" denotes the root of mean square errors of the Bayesian estimates based on 100 replications. Examination of Table 1 shows that (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 Figs. 1 and 2 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 SEJMVMs.

To investigate the sensitivity of the Bayesian estimate for $g(u)$ to the selection of the number of internal knots, we consider the other two choices of K, 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 $\lfloor s \rfloor$ denotes the largest integer not greater than s. Here, we only present the results of Bayesian estimates in

Table 2
Bayesian estimates of parameters for different choices of K when $n = 70$.

Type	Parameters	$(n = 70, K_1)$			$(n = 70, K_2)$		
		EST	SD	RMS	EST	SD	RMS
I	β_1	0.9450	0.1975	0.2297	0.9590	0.1966	0.1705
	β_2	−0.4857	0.1991	0.1817	−0.4952	0.1986	0.2069
	β_3	0.4909	0.1999	0.2110	0.4618	0.1962	0.1929
	γ_1	0.8286	0.3324	0.3857	0.9204	0.3280	0.2940
	γ_2	−0.4549	0.3251	0.3058	−0.4685	0.3361	0.3245
	γ_3	0.4230	0.3203	0.3010	0.4359	0.3365	0.2994
II	β_1	1.0191	0.1937	0.1733	1.0168	0.1956	0.1861
	β_2	−0.4777	0.1933	0.1924	−0.4960	0.1917	0.1917
	β_3	0.5023	0.1938	0.1905	0.4810	0.1911	0.1961
	γ_1	1.0835	0.3247	0.3298	1.0543	0.3394	0.3047
	γ_2	−0.5021	0.3303	0.2990	−0.5597	0.3347	0.2982
	γ_3	0.5398	0.3269	0.3051	0.4676	0.3371	0.3126
III	β_1	1.0699	0.1838	0.2242	1.0682	0.1863	0.1835
	β_2	−0.4922	0.1841	0.2085	−0.5524	0.1886	0.2133
	β_3	0.5267	0.1839	0.1781	0.5138	0.1856	0.1911
	γ_1	1.2961	0.3256	0.4270	1.2609	0.3336	0.3970
	γ_2	−0.6203	0.3304	0.3163	−0.6427	0.3421	0.3665
	γ_3	0.5892	0.3281	0.3267	0.6145	0.3416	0.3103

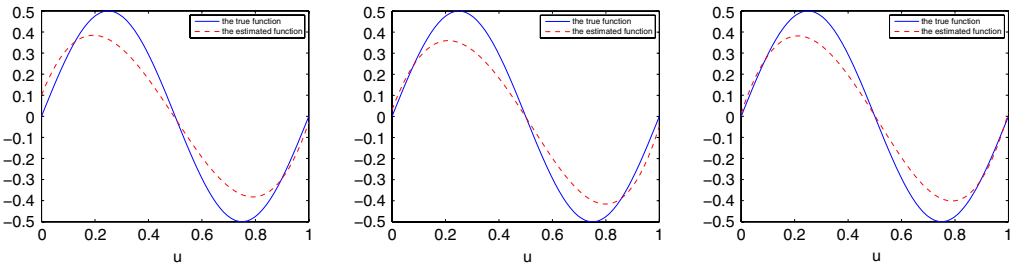

Fig. 3. The average of the estimates versus the true value of $g(u)$ under three priors: type I (left panel), type II (middle panel) and type III (right panel) for $n = 70$ and K_1.

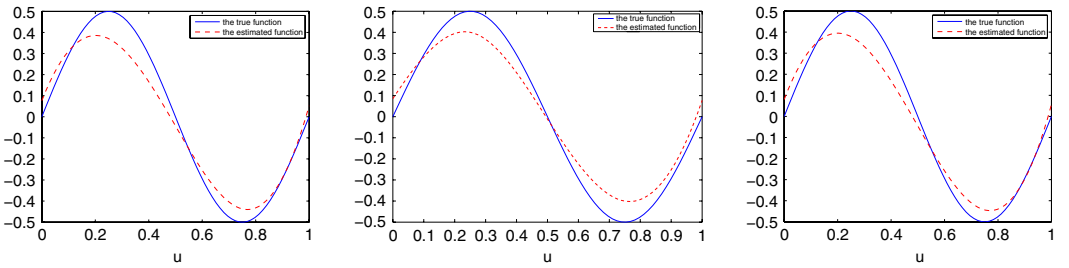

Fig. 4. The average of the estimates versus the true value of $g(u)$ under three priors: type I (left panel), type II (middle panel) and type III (right panel) for $n = 70$ and K_2.

Table 2 and Figs. 3 and 4 for $n = 70$ under different choices of K. By viewing Table 2 and comparing it with Table 1, 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 Figs. 3 and 4, we can obtain that the shapes of the estimated nonparametric function are very similar to those in Fig. 1.

4. Real data analysis

In this section, we apply the proposed method to the Ragweed Pollen Level data, which was analyzed in Ruppert et al. (2003). The data was recorded during the 1993 ragweed season in Kalamazoo, Michigan, and it consists of 87 daily observations of ragweed pollen level and relevant information. The main interest is to develop accurate models to forecast daily ragweed pollen level. The raw response ragweed is the daily ragweed pollen level (grains/m^3). Among the explanatory variables, x_1 is an indicator of significant rain, where $x_1 = 1$ if there is at least 3 h steady or brief but intense rain and $x_1 = 0$ otherwise; x_2 is temperature (°F); x_3 is wind speed (knots). The x-covariates are standardized first. Since the raw

Table 3
The real example: Bayesian estimates and their standard deviations.

Parameter	β_1	β_2	β_3	γ_1	γ_2	γ_3
EST	1.2196	0.6747	0.3445	1.2806	1.4307	0.1962
SD	0.2850	0.2787	0.1366	0.1870	0.2104	0.1816

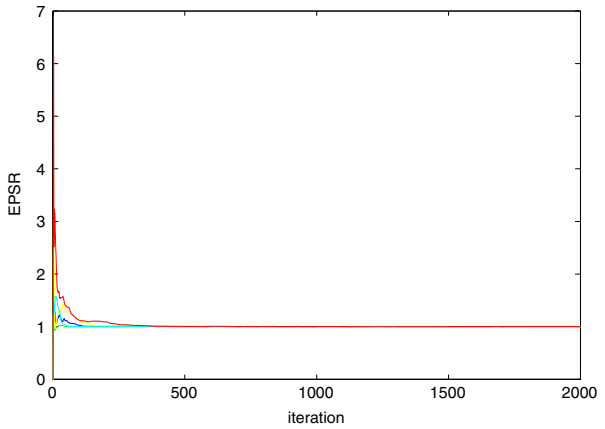

Fig. 5. EPSR values of all parameters against iterations in the real example.

response is rather skewed, Ruppert et al. (2003) suggested a square root transformation $y = \sqrt{ragweed}$. Marginal plots in Ruppert et al. (2003) suggest a strong nonlinear relationship between y and the day number in the current ragweed pollen season. Consequently, a semiparametric regression model with a nonparametric baseline $f(day)$ is reasonable. Therefore, we consider our semiparametric joint mean and variance models as follows:

$$\begin{cases} y_i \sim N(\mu_i, \sigma_i^2) \\ \mu_i = f(day) + \sum_{j=1}^{3} \beta_j x_{ij} \\ \log \sigma_i^2 = \sum_{j=1}^{3} \gamma_j z_{ij} \\ i = 1, \ldots, 87, \end{cases}$$

where $z_{ij} = x_{ij}, j = 1, \ldots, 3$.

The preceding proposed hybrid algorithm is used to obtain Bayesian estimates of β's and γ's, and the B-spline with $K = 2$ knots is employed to estimate the unknown nonparametric function. In the Metropolis–Hastings algorithm, we set $\sigma_\gamma^2 = 2.4^2/3$ in their corresponding proposal distributions, which gives approximate acceptance rates 30.91%. To test the convergence of the algorithm, plot of the EPSR values for all the unknown parameters against iterations is presented in Fig. 5, which indicates that the algorithm converges about 1000 iterations because EPSR values of all unknown parameters are less than 1.2 about 1000 iterations. We calculate Bayesian estimates (EST), standard deviation estimates (SD) of the Bayesian estimates of β's and γ's. Results are given in Table 3. Fig. 6 displays the Bayesian estimate of the nonparametric function $f(day)$.

From Table 3, we have the following observations. (i) All coefficients are estimated to be positive, suggesting that the ragweed pollen level increases as each of the covariates increases. (ii) There is a strong association with x_1 in the mean model and x_1, x_2 in the variance model, which also indicates that the significant rain has a strong effect on daily ragweed pollen level. In addition, Fig. 6 indicates a significant nonlinear trend with respect to the day number. Fig. 6 also depicts that the estimated nonparametric function $\hat{f}(day)$ climbs rapidly to the peak on around day 25 and plunges until day 70, and decreases steadily thereafter. These observations are very similar to those in Ruppert et al. (2003).

5. Conclusion and discussion

In this article, we extend the Bayesian methodology proposed in Cepeda and Gamerman (2001) to fit semiparametric joint mean and variance models. A fully Bayesian approach is developed to analyze this model via B-spline estimate of the nonparametric part by combining the Gibbs sampler and Metropolis–Hastings algorithm. A simulation study and 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.

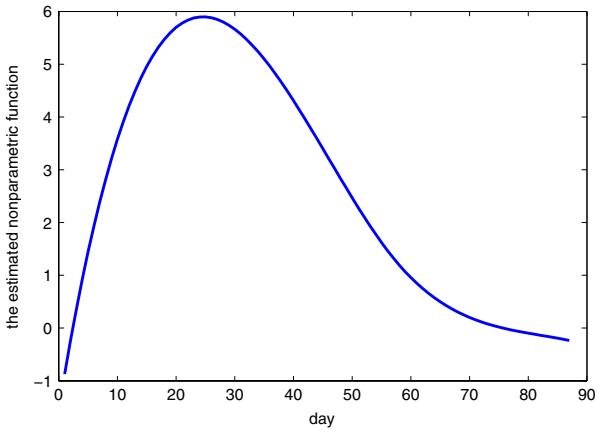

Fig. 6. The estimated function $f(day)$ based on Bayesian method.

Acknowledgments

The authors are grateful to the Editor, an Associate Editor and the referees for helpful comments and suggestions that greatly improved the paper. This work is supported by grants from the National Natural Science Foundation of China (11271039, 11101015); Funding Project of Science and Technology Research Plan of Beijing Education Committee (JC006790201001); the fund from the government of Beijing (2011D005015000007).

References

Aitkin, M., 1987. Modelling variance heterogeneity in normal regression using GLIM. Appl. Stat. 36, 332–339.
Cepeda, E., Gamerman, D., 2001. Bayesian modeling of variance heterogeneity in normal regression models. Braz. J. Probab. Stat. 14, 207–221.
Chen, X.D., 2009. Bayesian analysis of semiparametric mixed-effects models for zero-inflated count data. Comm. Statist. Theory Methods 38, 1815–1833.
Chen, X.D., Tang, N.S., 2010. Bayesian analysis of semiparametric reproductive dispersion mixed-effects models. Comput. Statist. Data Anal. 54, 2145–2158.
Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Application. Chapman and Hall, New York.
Gelman, A., 1996. Inference and Monitoring Convergence in Markov Chain Monte Carlo in Practice. Chapman and Hall, London.
Gelman, A., Roberts, G.O., Gilks, W.R., 1995. Efficient metropolis jumping rules. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, Vol. 5. Oxford University Press, Oxford, pp. 599–607.
Geyer, C.J., 1992. Practical Markov chain Monte Carlo. Statist. Sci. 7, 473–511.
Harvey, A.C., 1976. Estimating regression models with multiplicative heteroscedasticity. Econometrica 44, 460–465.
He, X., Fung, W.K., Zhu, Z.Y., 2005. Robust estimation in a generalized partial linear model for clustered data. J. Amer. Statist. Assoc. 100, 1176–1184.
Lee, S.Y., Zhu, H.T., 2000. Statistical analysis of nonlinear structural equation models with continuous and polytomous data. British J. Math. Statist. Psych. 53, 209–232.
Li, A.P., Chen, Z.X., Xie, F.C., 2012. Diagnostic analysis for heterogeneous log-Birnbaum–Saunders regression models. Statist. Probab. Lett. 82, 1690–1698.
Lin, T.I., Wang, W.L., 2011. Bayesian inference in joint modelling of location and scale parameters of the t distribution for longitudinal data. J. Statist. Plann. Inference 141, 1543–1553.
Lin, J.G., Wei, B.C., 2003. Testing for heteroscedasticity in nonlinear regression models. Comm. Statist. Theory Methods 32, 171–192.
Park, R.E., 1966. Estimation with heteroscedastic error terms. Econometrica 34, 888.
Roberts, G., 1996. Markov Chain Concepts Related to Sampling Algorithm. Chapman and Hall, London.
Ruppert, D., Wand, M.P., Carroll, R.J., 2003. Semiparametric Regression. Cambridge University Press, Cambridge, New York.
Schumaker, L.L., 1981. Spline Function. Wiley, New York.
Wu, L.C., Li, H.Q., 2012. Variable selection for joint mean and dispersion models of the inverse Gaussian distribution. Metrika 75, 795–808.
Xie, F.C., Wei, B.C., Lin, J.G., 2009. Homogeneity diagnostics for skew-normal nonlinear regression models. Statist. Probab. Lett. 79, 821–827.
Xu, D.K., Zhang, Z.Z., 2011. Regularized REML for estimation in heteroscedastic regression models. In: Li, S. (Ed.), Nonlinear Maths for Uncertainty and its Application. In: AISC, vol. 100. Springer Press, Berlin, Heidelberg, pp. 495–502.
Zhang, Z.Z., Wang, D.R., 2011. Simultaneous variable selection for heteroscedastic regression models. Sci. China Math. 54 (3), 515–530.
Zhao, P.X., Xue, L.G., 2010. Variable selection for semiparametric varying coefficient partially linear errors-in-variables models. J. Multivariate Anal. 101, 1872–1883.