Stat Papers (2017) 58:1091–1113
DOI 10.1007/s00362-016-0739-x

REGULAR ARTICLE

Semiparametric Bayesian inference on generalized linear measurement error models

Nian-Sheng Tang[1] · **De-Wang Li**[1] · **An-Min Tang**[1]

Received: 11 October 2014 / Revised: 9 December 2015 / Published online: 1 February 2016

Abstract The classical assumption in generalized linear measurement error models (GLMEMs) is that measurement errors (MEs) for covariates are distributed as a fully parametric distribution such as the multivariate normal distribution. This paper uses a centered Dirichlet process mixture model to relax the fully parametric distributional assumption of MEs, and develops a semiparametric Bayesian approach to simultaneously obtain Bayesian estimations of parameters and covariates subject to MEs by combining the stick-breaking prior and the Gibbs sampler together with the Metropolis–Hastings algorithm. Two Bayesian case-deletion diagnostics are proposed to identify influential observations in GLMEMs via the Kullback–Leibler divergence and Cook's distance. Computationally feasible formulae for evaluating Bayesian case-deletion diagnostics are presented. Several simulation studies and a real example are used to illustrate our proposed methodologies.

Keywords Cook's distance · Dirichlet process prior · Generalized linear models · Kullback–Leibler divergence · Measurement error models

1 Introduction

Generalized linear models (GLMs) are widely used to fit responses that do not satisfy the usual requirements of least-squares methods in biostatistics, epidemiology, and many other areas. However, the real data fitted via a GLM often involve covariates subject to measurement errors (Carroll et al. 2006; Singh et al. 2014). GLMs with covariates having measurement errors (MEs), which are often referred to as gener-

✉ Nian-Sheng Tang
nstang@ynu.edu.cn

[1] Department of Statistics, Yunnan University, Kunming 650091, People's Republic of China

alized linear measurement error models (GLMEMs), have received a lot of attention in past years. For example, Stefanski and Carroll (1985) developed a bias-adjusted estimator, a functional maximum likelihood estimator and an estimator exploiting the consequences of sufficiency for a logistic regression when covariates were subject to MEs; Stefanski and Carroll (1987) studied parameter estimation in GLM with canonical form when the explanatory vector was measured with an independent normal error; Buzas and Stefanski (1996) investigated instrumental variable estimation in GLMEMs with canonical link functions; Aitkin and Rocci (2002) presented an EM algorithm for maximum likelihood estimation in GLMs with continuous MEs in the explanatory variables; Battauz (2011) developed a Laplace-based estimator for GLMEMs; Battauz and Bellio (2011) proposed a structural analysis for GLMs when some explanatory variables were measured with error and the ME variance was a function of the true variables.

All the above mentioned studies assume that the covariate MEs in GLMEMs are distributed as a fully parametric distribution such as the multivariate normal distribution. However, in some applications, the covariate MEs in GLMEMs may not follow a fully parametric distribution but follow a non-normal distribution such as the skew-normal (Cancho et al. 2010) and skew-t (Lachos et al. 2010) and bimodal and heavy-tailed distributions (Lachos et al. 2011). Moreover, the violation of the parametric assumption on the covariate MEs may lead to unreasonable or even misleading conclusions. Therefore, it is of practical interest to consider a flexible distributional assumption on the covariate MEs in GLMEMs. The nonparametric method is one of the most widely adopted approaches to specify a flexible probability distribution for random variables or MEs in the Bayesian framework.

The Dirichlet process (DP) prior (Ferguson 1973) is the most popular nonparametric approach to specify a probability distribution for random variables or MEs in the Bayesian framework due to the availability of some efficient computational techniques. The nonparametric method has been successfully used to make statistical inference on various random effects models. For example, see Kleinman and Ibrahim (1998), Dunson (2006), Guha (2008), Lee et al. (2008), Chow et al. (2011), Tang and Duan (2012) and Tang et al. (2014). However, to the best of our knowledge, little work is done on Bayesian analysis of GLMEMs with the covariate MEs following a nonparametric distribution. Hence, this paper develops a semiparametric Bayesian approach to simultaneously obtain Bayesian estimations of parameters and covariates subject to MEs, and proposes two Bayesian case deletion diagnostics to detect the potential influential observations under the centered DP mixture model specification of the covariate MEs in GLMEMs. In this paper, a hybrid algorithm is presented to generate observations required for a Bayesian inference from the posterior distributions of parameters and covariates subject to MEs by combining the stick-breaking prior and the Gibbs sampler (Geman and Geman 1984) together with the Metropolis–Hastings algorithm.

Bayesian case deletion approaches to detect influential observations (or sets of observations) have been proposed for some statistical models such as linear regression models (Carlin and Polson 1991), GLMs (Jackson et al. 2012) and generalized linear mixed models (Fong et al. 2010) based on the Kullback–Leibler divergence (K–L divergence) and the conditional predictive ordinate. But, extending these existing

Bayesian case deletion diagnostics to our considered GLMEMs is computationally challenging because of the complexity of the considered models and the unknown distribution of the covariate MEs. To this end, the well-known Markov chain Monte Carlo (MCMC) algorithm is employed to develop two computationally feasible Bayesian case deletion diagnostics to assess the effect of cases (or sets of observations) on posterior distributions or estimations of parameters based on the K–L divergence and Cook's distance in this paper.

The rest of this paper is organized as follows. Section 2 introduces GLMEMs by using the centered DP mixture model to specify the distribution of covaraite MEs. Section 3 develops a Bayesian MCMC algorithm to make Bayesian inference on GLMEMs by using the Gibbs sampler together with the Metropolis–Hastings algorithm. Two Bayesian case deletion diagnostic measures are presented to detect influential observations based on the K–L divergence and Cook's distance in Sect. 3. Several simulation studies and a real example are used to illustrate our proposed methodologies in Sect. 4. Some concluding remarks are given in Sect. 5. Technical details are presented in the Appendix.

2 Generalized linear measurement error models

For $i = 1, \ldots, n$, let y_i denote the observed outcome variable, $\boldsymbol{x}_i$ be a $r \times 1$ vector of the unobserved covariate variables, and $\boldsymbol{v}_i$ be a $p \times 1$ vector of the observed covariate variables for the ith individual. Generally, the unobserved components of covariates may vary across different individuals. For simplicity, we assume that the unobserved components of covariates have the same components for $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$, where $\boldsymbol{z}_i = (\boldsymbol{x}_i^\top, \boldsymbol{v}_i^\top)^\top$ for $i = 1, \ldots, n$. Given $\boldsymbol{z}_i$, we assume that y_i's are conditionally independent of each other, and the conditional distribution of y_i is a one-parameter exponential family with a canonical parameter θ_i and a mean that is a function of $\boldsymbol{z}_i$. That is, for $i = 1, \ldots, n$, the conditional probability density function of y_i given $\boldsymbol{z}_i$ is given by

$$p(y_i|\boldsymbol{z}_i, \phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi)\right\} \tag{2.1}$$

with $\mu_i = \mathrm{E}(y_i|\boldsymbol{z}_i) = \dot{b}(\theta_i)$ and $V_i = \mathrm{var}(y_i|\boldsymbol{z}_i) = \phi\ddot{b}(\theta_i)$, where ϕ is a scale parameter, $b(\cdot)$ and $c(\cdot, \cdot)$ are specific differentiable functions, $\dot{b}(\theta_i) = \partial b(\theta_i)/\partial\theta_i$ and $\ddot{b}(\theta_i) = \partial^2 b(\theta_i)/\partial\theta_i^2$. The conditional mean μ_i is assumed to satisfy

$$\eta_i = h(\mu_i) = \boldsymbol{x}_i^\top\boldsymbol{\beta}_x + \boldsymbol{v}_i^\top\boldsymbol{\beta}_v = \boldsymbol{z}_i^\top\boldsymbol{\beta}, \tag{2.2}$$

where $h(\cdot)$ is a monotonic differentiable link function, $\boldsymbol{\beta} = (\boldsymbol{\beta}_x^\top, \boldsymbol{\beta}_v^\top)^\top$ is a $(r + p) \times 1$ vector of unknown regression coefficients. Generally, there are two approaches to specify the ME structure. One is the structural ME model, and the other is the functional ME model. In a structural ME model, the special assumption is made on the distributional structure of the unobserved covariates, whilst nothing is assumed on the unobserved covariates in a functional ME model. Following Carroll et al. (2006), if

the true covariate $\boldsymbol{x}_i$ is measured m times for individual i, giving outcomes $\boldsymbol{w}_{ij}$ for $j = 1, \ldots, m$, the structural ME model can be expressed as

$$\boldsymbol{w}_{ij} = \boldsymbol{x}_i + \boldsymbol{u}_{ij}, \tag{2.3}$$

where the MEs $\boldsymbol{u}_{ij}$'s are assumed to follow an unknown distribution, and are independent of $\boldsymbol{x}_i$.

Following Lee et al. (2008), we use the DP mixture model to specify the distribution of $\boldsymbol{u}_{ij}$. That is, $\boldsymbol{u}_{ij} \overset{\text{i.i.d.}}{\sim} \sum_{g=1}^{\infty} \pi_g N_r(\boldsymbol{\alpha}_g, \boldsymbol{\Omega}_g)$ with $(\boldsymbol{\alpha}_g, \boldsymbol{\Omega}_g) \sim \mathscr{P}$ and $\mathscr{P} \sim \text{DP}(\tau P_0)$, where π_g is a random probability weight between 0 and 1 such that $0 \le \pi_g \le 1$ and $\sum_{g=1}^{\infty} \pi_g = 1$, $N_r(\boldsymbol{\alpha}_g, \boldsymbol{\Omega}_g)$ denotes the multivariate normal distribution with mean $\boldsymbol{\alpha}_g$ and covariance matrix $\boldsymbol{\Omega}_g$, $\mathscr{P}$ is a random probability with an unknown form, P_0 is a base distribution that serves as a starting-point for constructing the nonparametric distribution, and τ is a weight assigning a priori to the base distribution and represents the certainty of P_0 as the distribution of $\mathscr{P}$. The widely used distribution for P_0 is the multivariate normal distribution. The DP prior with the stick-breaking representation may yield non-zero mean of MEs (Yang et al. 2010), which is inconsistent with the classic assumption that mean of $\boldsymbol{u}_{ij}$ is zero (Carroll et al. 2006).

Inspired by Yang et al. (2010), we consider the following truncated and centered DP (TCDP) mixture model for $\boldsymbol{u}_{ij}$:

$$\boldsymbol{u}_{ij} \overset{\text{i.i.d.}}{\sim} \sum_{g=1}^{G} \pi_g N_r(\boldsymbol{\alpha}_g, \boldsymbol{\Omega}_g) \text{ with } \boldsymbol{\alpha}_g = \boldsymbol{\alpha}_g^* - \sum_{l=1}^{G} \pi_l \boldsymbol{\alpha}_l^* \text{ and } (\boldsymbol{\alpha}_g^*, \boldsymbol{\Omega}_g) \overset{\text{i.i.d.}}{\sim} P_0, \tag{2.4}$$

where G is the number of the truncated mixture components, π_g is taken to be the following stick-breaking procedure:

$$\pi_1 = \nu_1 \text{ and } \pi_g = \nu_g \prod_{l=1}^{g-1} (1 - \nu_l) \text{ for } g = 2, \ldots, G \tag{2.5}$$

with $\nu_g \overset{\text{i.i.d.}}{\sim} \text{Beta}(1, \tau)$ for $g = 1, \ldots, G-1$, and $\nu_G = 1$ so that $\sum_{g=1}^{G} \pi_g = 1$.

Based on the above specified TCDP mixture model, it is quite difficult to sample observations from posterior distribution of $\boldsymbol{u}_{ij}$ via the MCMC algorithm because of the complicated posterior distribution involved. To this end, we generate $\boldsymbol{u}_{ij}$ from $N_r(\boldsymbol{\alpha}_{L_{ij}}, \boldsymbol{\Omega}_{L_{ij}})$ in terms of a latent variable $L_{ij} \in \{1, \ldots, G\}$, where $\boldsymbol{\alpha}_{L_{ij}} = \boldsymbol{\alpha}_{L_{ij}}^* - \sum_{g=1}^{G} \pi_g \boldsymbol{\alpha}_g^*$ in which $\boldsymbol{\alpha}_{L_{ij}}^*$ is sampled from the following reformulated model. Let $\boldsymbol{\pi} = \{\pi_1, \ldots, \pi_G\}$, $\boldsymbol{\alpha}^* = \{\boldsymbol{\alpha}_1^*, \ldots, \boldsymbol{\alpha}_G^*\}$ and $\boldsymbol{\Omega} = \{\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_G\}$ in which $\boldsymbol{\Omega}_g = \text{diag}(\omega_{g1}, \ldots, \omega_{gr})$ for $g = 1, \ldots, G$. It follows from Lee et al. (2008) that Equation (2.4) can be rewritten as

$$L_{ij}|\boldsymbol{\pi} \overset{\text{i.i.d.}}{\sim} \sum_{g=1}^{G} \pi_g \delta_g(\cdot) \text{ and } (\boldsymbol{\pi}, \boldsymbol{\alpha}^*, \boldsymbol{\Omega}) \sim f_1(\boldsymbol{\pi}) f_2(\boldsymbol{\alpha}^*) f_3(\boldsymbol{\Omega}), \tag{2.6}$$

where $\delta_g(\cdot)$ is a discrete probability measure concentrated at g, $f_1(\boldsymbol{\pi})$ is specified by the stick-breaking prior as given in Eq. (2.5), the prior distribution of $\boldsymbol{\alpha}_g^*$ involved in $f_2(\boldsymbol{\alpha}^*) = \prod_{g=1}^{G} p(\boldsymbol{\alpha}_g^*)$ is given by

$$\boldsymbol{\alpha}_g^* | \boldsymbol{\xi}, \boldsymbol{\Psi} \overset{\text{i.i.d.}}{\sim} N_r(\boldsymbol{\xi}, \boldsymbol{\Psi}), \quad \boldsymbol{\xi} | \boldsymbol{\xi}^0, \boldsymbol{\Psi}^0 \sim N_r(\boldsymbol{\xi}^0, \boldsymbol{\Psi}^0), \quad \psi_j^{-1} | c_1, c_2 \sim \Gamma(c_1, c_2) \quad \text{for } j = 1, \ldots, r, \tag{2.7}$$

where $\boldsymbol{\Psi} = \text{diag}(\psi_1, \ldots, \psi_r)$, $\boldsymbol{\xi}^0$, $\boldsymbol{\Psi}^0$, c_1 and c_2 are hyperparameters whose values are assumed to be known, $\Gamma(c_1, c_2)$ denotes the Gamma distribution with parameters c_1 and c_2, and the prior distribution for ω_{gj} involved in $f_3(\boldsymbol{\Omega}) = \prod_{g=1}^{G} \prod_{j=1}^{r} p(\omega_{gj})$ is given by

$$\omega_{gj}^{-1} | \varphi_j^a, \varphi_j^b \sim \Gamma(\varphi_j^a, \varphi_j^b) \text{ and } \varphi_j^b | \varphi_j^c, \varphi_j^d \sim \Gamma(\varphi_j^c, \varphi_j^d), \tag{2.8}$$

where φ_j^a, φ_j^c and φ_j^d are the pre-specified hyperparameters.

To complete specification of the covariate ME model, we require defining a true covariate model. Following Aitkin and Rocci (2002) and Gustafson (2004), the true covariate model for x_{ki} $(k = 1, \ldots, r)$ can be defined as

$$x_{ki} = \gamma_{k0} + \boldsymbol{\gamma}_{kv}^\top \boldsymbol{v}_i + \varepsilon_{ki}, \quad \varepsilon_{ki} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_x^2), \tag{2.9}$$

where γ_{k0} is an intercept, $\boldsymbol{\gamma}_{kv} = (\gamma_{k1}, \ldots, \gamma_{kp})^\top$ is a $p \times 1$ vector of unknown regression parameters, and ε_{ki}'s are residuals and assumed to be independent of the covariates $\boldsymbol{v}_i$ and MEs $\boldsymbol{u}_{ij}$'s. The model defined in Eqs. (2.1)–(2.3) together with (2.9) is referred to as a GLMEM.

The above defined model is not identifiable when there are no replicate measurements, i.e., $m = 1$. In this case, some identification conditions on parameters are required (Lee and Tang 2006), for example, we may set σ_x^2 and γ_{k0} to be some pre-specified values.

Let $\boldsymbol{y} = \{y_1, \ldots, y_n\}$, $\boldsymbol{x} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, $\boldsymbol{v} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$, $\boldsymbol{u} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ and $\boldsymbol{w} = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$ in which $\boldsymbol{x}_i = (x_{1i}, \ldots, x_{ri})^\top$, $\boldsymbol{u}_i = \{\boldsymbol{u}_{i1}, \ldots, \boldsymbol{u}_{im}\}$ and $\boldsymbol{w}_i = \{\boldsymbol{w}_{i1}, \ldots, \boldsymbol{w}_{im}\}$ for $i = 1, \ldots, n$. Denote $\boldsymbol{\theta}_y = \{\boldsymbol{\beta}, \phi\}$, $\boldsymbol{\theta}_u = \{\tau, \boldsymbol{\pi}, \boldsymbol{\alpha}^*, \boldsymbol{\Omega}\}$, $\boldsymbol{\theta}_\gamma = \{\gamma_{10}, \ldots, \gamma_{r0}, \boldsymbol{\gamma}_{1v}, \ldots, \boldsymbol{\gamma}_{rv}, \sigma_x^2\}$ and $\boldsymbol{\theta} = \{\boldsymbol{\theta}_y, \boldsymbol{\theta}_u, \boldsymbol{\theta}_\gamma\}$. Under the above assumptions, the joint probability density function for $\{\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{u}, \boldsymbol{x}\}$ is given by

$$P(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{u}, \boldsymbol{x} | \boldsymbol{v}, \boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{ p(y_i | \boldsymbol{x}_i, \boldsymbol{v}_i; \boldsymbol{\theta}_y) p(\boldsymbol{w}_i | \boldsymbol{x}_i; \boldsymbol{\theta}_u) p(\boldsymbol{x}_i | \boldsymbol{v}_i; \boldsymbol{\theta}_\gamma) \right\}. \tag{2.10}$$

To make Bayesian inference on parameters in $\{\tau, \boldsymbol{\theta}_y, \boldsymbol{\theta}_\gamma\}$, it is necessary to specify their corresponding priors. Similar to Lee and Tang (2006), we consider the following priors for parameters τ, $\boldsymbol{\beta}$, ϕ, $\boldsymbol{\gamma}_k^* = (\gamma_{k0}, \boldsymbol{\gamma}_{kv}^\top)^\top$ for $k = 1, \ldots, r$, and σ_x^2:

$$\tau | a_1, a_2 \sim \Gamma(a_1, a_2), \quad \boldsymbol{\beta} | \phi, \boldsymbol{\beta}^0, \boldsymbol{H}_\beta^0 \sim N_{r+p}(\boldsymbol{\beta}^0, \phi^{-1} \boldsymbol{H}_\beta^0),$$
$$\phi^{-1} | a_3, a_4 \sim \Gamma(a_3, a_4), \quad \boldsymbol{\gamma}_k^* | \boldsymbol{\gamma}_k^0, \boldsymbol{H}_{\gamma k}^0 \sim N_{p+1}(\boldsymbol{\gamma}_k^0, \boldsymbol{H}_{\gamma k}^0), \quad \sigma_x^{-2} | c_3, c_4 \sim \Gamma(c_3, c_4),$$

where $a_1, a_2, a_3, a_4, \boldsymbol{\beta}^0, \boldsymbol{H}_{\beta}^0, \boldsymbol{\gamma}_k^0, \boldsymbol{H}_{\gamma k}^0, c_3$ and c_4 are hyperparameters whose values are assumed to be given by the prior information. Thus, we specify standard conjugate priors for parameters τ, ϕ^{-1}, $\boldsymbol{\gamma}_k^*$ and σ_x^{-2}. The associated hyperparameters can be determined in a relatively straightforward manner based on previous applications or data-dependent inputs (Raftery 1996; Richarson and Green 1997; Lee and Tang 2006). Particularly, the values of a_1 and a_2 should be selected carefully because they directly affect the value of τ which controls the behavior of $\boldsymbol{u}_{ij}$. Detailed discussions on selection of a_1 and a_2 can refer to Chow et al. (2011). Detailed comments on selection of c_3 and c_4 are given in simulation studies and real example analysis.

Based on the above presented joint probability density function and priors, a Bayesian approach is developed to make statistical inference on parameters in $\{\tau, \boldsymbol{\theta}_y, \boldsymbol{\theta}_\gamma\}$ by utilizing the Gibbs sampler together with the Metropolis–Hastings algorithm for our considered GLMEMs.

3 Bayesian inference on GLMEMs

From Eq. (2.10) and the above defined priors, it is easily seen that it is rather difficult to directly make Bayesian inference on parameters in $\{\tau, \boldsymbol{\theta}_y, \boldsymbol{\theta}_\gamma\}$ because of the intractable high-dimensional integrals involved. Owing to recent development in statistical computing, the Gibbs sampler is employed to generate a sequence of random observations from the joint posterior distribution $p(\boldsymbol{\xi}, \boldsymbol{\Psi}, \tau, \boldsymbol{\varphi}^b, \boldsymbol{\pi}, \boldsymbol{L}, \boldsymbol{\alpha}^*, \boldsymbol{\Omega}, \boldsymbol{u}, \boldsymbol{\beta}, \phi, \boldsymbol{\gamma}, \sigma_x^2, \boldsymbol{x}|\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{v})$, and Bayesian estimates of unknown parameters and covariates subject to MEs are obtained from the generated sequence of random observations, where $\boldsymbol{\varphi}^b = \{\varphi_1^b, \ldots, \varphi_r^b\}$, $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}_1^*, \ldots, \boldsymbol{\gamma}_r^*\}$ and $\boldsymbol{L} = \{L_{ij} : i = 1, \ldots, n, j = 1, \ldots, m\}$. In this algorithm, observations $\{\boldsymbol{\xi}, \boldsymbol{\Psi}, \tau, \boldsymbol{\varphi}^b, \boldsymbol{\pi}, \boldsymbol{L}, \boldsymbol{\alpha}^*, \boldsymbol{\Omega}, \boldsymbol{u}, \boldsymbol{\beta}, \phi, \boldsymbol{\gamma}, \sigma_x^2, \boldsymbol{x}\}$ are iteratively drawn from the following conditional distributions: $p(\boldsymbol{\xi}|\boldsymbol{\alpha}^*, \boldsymbol{\Psi})$, $p(\boldsymbol{\Psi}|\boldsymbol{\alpha}^*, \boldsymbol{\xi})$, $p(\tau|\boldsymbol{\pi})$, $p(\boldsymbol{\varphi}^b|\boldsymbol{\Omega})$, $p(\boldsymbol{\pi}|\boldsymbol{L}, \tau)$, $p(\boldsymbol{L}|\boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{u})$, $p(\boldsymbol{\alpha}^*|\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\Omega}, \boldsymbol{L}, \boldsymbol{u})$, $p(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \phi)$, $p(\boldsymbol{\Omega}|\boldsymbol{\alpha}, \boldsymbol{\varphi}^b, \boldsymbol{L}, \boldsymbol{u})$, $p(\phi|\boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v})$, $p(\boldsymbol{\gamma}|\boldsymbol{x}, \boldsymbol{v}, \sigma_x^2)$, $p(\sigma_x^2|\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{\gamma})$, $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{u}, \boldsymbol{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi, \sigma_x^2, \boldsymbol{\gamma}, \boldsymbol{L})$ and $p(\boldsymbol{u}|\boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{L}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{\theta}_u)$. These conditional distributions are presented in the Appendix.

3.1 Bayesian estimates

Let $\{(\boldsymbol{\beta}^{(\ell)}, \phi^{(\ell)}, \boldsymbol{\gamma}^{(\ell)}, \boldsymbol{x}^{(\ell)}, \sigma_x^{2(\ell)}) : \ell = 1, \ldots, \mathfrak{L}\}$ be observations of $\{\boldsymbol{\beta}, \phi, \boldsymbol{\gamma}, \boldsymbol{x}, \sigma_x^2\}$ generated from $p(\boldsymbol{\xi}, \boldsymbol{\Psi}, \tau, \boldsymbol{\varphi}^b, \boldsymbol{\pi}, \boldsymbol{L}, \boldsymbol{\alpha}^*, \boldsymbol{\Omega}, \boldsymbol{u}, \boldsymbol{\beta}, \phi, \boldsymbol{\gamma}, \sigma_x^2, \boldsymbol{x}|\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{w})$ via the preceding presented algorithm. Thus, Bayesian estimates of $\{\boldsymbol{\beta}, \phi, \boldsymbol{\gamma}, \boldsymbol{x}, \sigma_x^2\}$ can be obtained by

$$\hat{\boldsymbol{\beta}} = \frac{1}{\mathfrak{L}} \sum_{\ell=1}^{\mathfrak{L}} \boldsymbol{\beta}^{(\ell)}, \quad \hat{\phi} = \frac{1}{\mathfrak{L}} \sum_{\ell=1}^{\mathfrak{L}} \phi^{(\ell)}, \quad \hat{\boldsymbol{\gamma}} = \frac{1}{\mathfrak{L}} \sum_{\ell=1}^{\mathfrak{L}} \boldsymbol{\gamma}^{(\ell)}, \quad \hat{\boldsymbol{x}} = \frac{1}{\mathfrak{L}} \sum_{\ell=1}^{\mathfrak{L}} \boldsymbol{x}^{(\ell)}, \quad \hat{\sigma}_x^2 = \frac{1}{\mathfrak{L}} \sum_{\ell=1}^{\mathfrak{L}} \sigma_x^{2(\ell)}. \tag{3.1}$$

The consistent estimates of the posterior covariance matrices $\mathrm{var}(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{w})$, $\mathrm{var}(\phi|\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{w})$, $\mathrm{var}(\boldsymbol{\gamma}|\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{w})$ and $\mathrm{var}(\sigma_x^2|\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{w})$ can be obtained from the sample covari-

ance matrices of the above generated observations. For example, $\widehat{\text{var}}(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{v},\boldsymbol{w}) = \frac{1}{\mathfrak{L}-1}\Sigma_{\ell=1}^{\mathfrak{L}}(\boldsymbol{\beta}^{(\ell)}-\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}^{(\ell)}-\hat{\boldsymbol{\beta}})^{\top}$. The standard errors for components of $\boldsymbol{\beta}$ can be obtained from the diagonal elements of the sample covariance matrix.

3.2 Bayesian case influence analysis

In this subsection, we consider Bayesian case-deletion influence analysis based on the case-deletion method given in Cook and Weisberg (1982). For notational simplicity, let $\boldsymbol{R} = \{\boldsymbol{y},\boldsymbol{v},\boldsymbol{w}\}$ be a full data set, and $\boldsymbol{R}_{(i)} = \{(y_j,\boldsymbol{v}_j,\boldsymbol{w}_j) : j = 1,\ldots,n,\ j \neq i\}$ be a subset of $\boldsymbol{R}$ with the ith individual deleted. To assess the effect of the ith individual on the posterior distribution of parameter vector $\boldsymbol{\vartheta} = \{\tau,\boldsymbol{\theta}_y,\boldsymbol{\theta}_\gamma\}$, we use the following Kullback–Leibler (K–L) divergence

$$KL(i) = \int \log\left\{\frac{p(\boldsymbol{\vartheta}|\boldsymbol{R})}{p(\boldsymbol{\vartheta}|\boldsymbol{R}_{(i)})}\right\} p(\boldsymbol{\vartheta}|\boldsymbol{R})d\boldsymbol{\vartheta},$$

as a Bayesian case influence measure, where $p(\boldsymbol{\vartheta}|\boldsymbol{R})$ and $p(\boldsymbol{\vartheta}|\boldsymbol{R}_{(i)})$ are the posterior distributions of $\boldsymbol{\vartheta}$ for the full data set $\boldsymbol{R}$ and the reduced data set $\boldsymbol{R}_{(i)}$, respectively. Thus, $KL(i)$ measures the distance (discrepancy) between two posterior distributions $p(\boldsymbol{\vartheta}|\boldsymbol{R}_{(i)})$ and $p(\boldsymbol{\vartheta}|\boldsymbol{R})$, which can be regarded as a Bayesian analogue of the likelihood displacement (Cook and Weisberg 1982).

To measure the effect of the ith individual on the posterior mean of $\boldsymbol{\vartheta}$, we use the following Cook's distance

$$CD(i) = (\hat{\boldsymbol{\vartheta}} - \hat{\boldsymbol{\vartheta}}_{(i)})^{\top}\boldsymbol{W}_{\vartheta}^{-1}(\hat{\boldsymbol{\vartheta}} - \hat{\boldsymbol{\vartheta}}_{(i)})$$

as another Bayesian case influence measure, where $\hat{\boldsymbol{\vartheta}} = \int \boldsymbol{\vartheta}\, p(\boldsymbol{\vartheta}|\boldsymbol{R})d\boldsymbol{\vartheta}$ and $\hat{\boldsymbol{\vartheta}}_{(i)} = \int \boldsymbol{\vartheta}\, p(\boldsymbol{\vartheta}|\boldsymbol{R}_{(i)})d\boldsymbol{\vartheta}$ are the posterior means of $\boldsymbol{\vartheta}$ for the full data set $\boldsymbol{R}$ and the deleted data set $\boldsymbol{R}_{(i)}$, respectively, and $\boldsymbol{W}_{\vartheta}$ is selected to be the posterior covariance matrix of $\boldsymbol{\vartheta}$. A large value of $CD(i)$ corresponds to an influential observation with respect to the posterior mean. Generally, we can use $\bar{D} + d \times SM$ as a benchmark (e.g., see Lee and Tang 2006), where $\bar{D}$ and SM are the mean and standard error of $\{CD(i) : i = 1,\ldots,n\}$, and d is a selected constant depending on the problem-by-problem. Specifically, we set $d = 3.0$ in our conducted simulation studies and $d = 5.0$ in real example analysis.

To compute $KL(i)$, we require calculating the marginal posterior distributions $p(\boldsymbol{\vartheta}|\boldsymbol{R})$ and $p(\boldsymbol{\vartheta}|\boldsymbol{R}_{(i)})$. It is rather difficult to directly compute $KL(i)$ because of the MEs involved. It is desirable to develop a computationally feasible formula to reduce computational burden. It is easily shown from independence of individuals that

$$p(\boldsymbol{\vartheta}|\boldsymbol{R}_{(i)}) = \frac{\{p_i(\boldsymbol{\vartheta})\}^{-1}p(\boldsymbol{\vartheta}|\boldsymbol{R})}{\int\{p_i(\boldsymbol{\vartheta})\}^{-1}p(\boldsymbol{\vartheta}|\boldsymbol{R})d\boldsymbol{\vartheta}}, \tag{3.2}$$

where $p_i(\boldsymbol{\vartheta}) = p(y_i,\boldsymbol{v}_i,\boldsymbol{w}_i|\boldsymbol{\vartheta})$. Substituting Eq. (3.2) into the definition of $KL(i)$ yields

$$KL(i) = \log E_{\vartheta|R}\{p_i(\boldsymbol{\vartheta})\}^{-1} + E_{\vartheta|R}\{\log p_i(\boldsymbol{\vartheta})\},$$

which indicates that computation of $KL(i)$ can be done using MCMC samples from the full data posterior distribution $p(\boldsymbol{\vartheta}|\boldsymbol{R})$ via the above developed Gibbs sampler. Specifically, if $\boldsymbol{\vartheta}^{(\ell)}$ is the ℓth Gibbs sample after $\mathcal{J}$ burn-in iterations for $\ell = 1, \ldots, \mathcal{L}$, thus we get the MCMC approximation of $KL(i)$ as

$$KL(i) \approx \log\left[\frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}}\{p_i(\boldsymbol{\vartheta}^{(\ell)})\}^{-1}\right] + \frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}}\log\{p_i(\boldsymbol{\vartheta}^{(\ell)})\}, \tag{3.3}$$

where $p_i(\boldsymbol{\vartheta}^{(\ell)}) = p(y_i, \boldsymbol{v}_i, \boldsymbol{w}_i|\boldsymbol{\vartheta}^{(\ell)})$.

On the other hand, to compute $CD(i)$, we need evaluating $\hat{\boldsymbol{\vartheta}}$, $\hat{\boldsymbol{\vartheta}}_{(i)}$ and $\boldsymbol{W}_{\vartheta}$. It follows from Eq. (3.2) and the definitions of $\hat{\boldsymbol{\vartheta}}$, $\hat{\boldsymbol{\vartheta}}_{(i)}$ and $\boldsymbol{W}_{\vartheta}$ that $\hat{\boldsymbol{\vartheta}} = E_{\vartheta|R}(\boldsymbol{\vartheta})$, $\hat{\boldsymbol{\vartheta}}_{(i)} = E_{\vartheta|R}[\boldsymbol{\vartheta}\{p_i(\boldsymbol{\vartheta})\}^{-1}]/E_{\vartheta|Y}\{p_i(\boldsymbol{\vartheta})\}^{-1}$ and $\boldsymbol{W}_{\vartheta} = \mathrm{var}_{\vartheta|R}(\boldsymbol{\vartheta})$. Similarly, the MCMC approximations of $\hat{\boldsymbol{\vartheta}}$, $\hat{\boldsymbol{\vartheta}}_{(i)}$ and $\boldsymbol{W}_{\vartheta}$ are given by

$$\hat{\boldsymbol{\vartheta}} = \frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}}\boldsymbol{\vartheta}^{(\ell)}, \quad \hat{\boldsymbol{\vartheta}}_{(i)} = \frac{\frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}}\boldsymbol{\vartheta}^{(\ell)}\{p_i(\boldsymbol{\vartheta}^{(\ell)})\}^{-1}}{\frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}}\{p_i(\boldsymbol{\vartheta}^{(\ell)})\}^{-1}}, \quad \boldsymbol{W}_{\vartheta} = \frac{1}{\mathcal{L}-1}\sum_{\ell=1}^{\mathcal{L}}(\boldsymbol{\vartheta}^{(\ell)} - \hat{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta}^{(\ell)} - \hat{\boldsymbol{\vartheta}})^{\top}, \tag{3.4}$$

respectively.

Regardless of $KL(i)$ or $CD(i)$, we need computing $p_i(\boldsymbol{\vartheta}^{(\ell)})$. From the definition of our considered model, we obtain

$$p_i(\boldsymbol{\vartheta}) = \int p(y_i, \boldsymbol{v}_i, \boldsymbol{w}_i, \boldsymbol{u}_i, \boldsymbol{x}_i|\boldsymbol{\vartheta})d\boldsymbol{u}_i d\boldsymbol{x}_i = E_{u_i,x_i}\{p(y_i, \boldsymbol{v}_i, \boldsymbol{w}_i|\boldsymbol{\vartheta}, \boldsymbol{x}_i, \boldsymbol{u}_i)\}, \tag{3.5}$$

where E_{u_i,x_i} denotes the expectation taken with respect to the joint distribution of $\boldsymbol{u}_i$ and $\boldsymbol{x}_i$ (denoted by $p(\boldsymbol{u}_i, \boldsymbol{x}_i|\boldsymbol{\theta}_u)$). Monte Carlo approximation of $p_i(\boldsymbol{\vartheta}^{(\ell)})$ in Eq. (3.5) can be implemented by using the following steps:

- *Step 0* Specifying the initial value $\boldsymbol{x}_i^{(0)}$ of $\boldsymbol{x}_i$. Generally, we can take $\boldsymbol{x}_i^{(0)}$ to be the mean of $\boldsymbol{w}_{ij}$'s.
- *Step 1* Sampling Gibbs sample $\boldsymbol{u}_i^{(t)}$ from the conditional distribution $p(\boldsymbol{u}_i|\boldsymbol{x}_i^{(t-1)}, \boldsymbol{\theta}_u^{(\ell)})$.
- *Step 2* Drawing Gibbs sample $\boldsymbol{x}_i^{(t)}$ from the conditional distribution $p(\boldsymbol{x}_i|\boldsymbol{u}_i^{(t)}, \boldsymbol{\theta}_u^{(\ell)})$.
- *Step 3* Repeating Steps 1 and 2 for $\mathcal{T}$ times.
- *Step 4* Getting the nested Gibbs samples $\{(\boldsymbol{u}_i^{(t)}, \boldsymbol{x}_i^{(t)}) : t = 1, \ldots, \mathcal{T}\}$ from $p(\boldsymbol{u}_i, \boldsymbol{x}_i|\boldsymbol{\vartheta}_u)$. Thus, the MCMC approximation of $p_i(\boldsymbol{\vartheta}^{(\ell)})$ is given by

$$p_i(\boldsymbol{\vartheta}^{(\ell)}) \approx \frac{1}{\mathcal{T}}\sum_{t=1}^{\mathcal{T}} p\big(y_i, \boldsymbol{v}_i, \boldsymbol{w}_i|\boldsymbol{\vartheta}^{(\ell)}, \boldsymbol{x}_i^{(t)}, \boldsymbol{u}_i^{(t)}\big). \tag{3.6}$$

Combining Eqs. (3.2)–(3.6) yields the values of $KL(i)$ and $CD(i)$ for all individuals. Index plots of $KL(i)$ and $CD(i)$ can be used to identify influential cases.

4 Numerical examples

4.1 Simulation studies

To investigate the finite performance of the preceding proposed Bayesian approaches under various distributional assumptions of the MEs $\boldsymbol{u}_{ij}$ and prior specifications, we conducted the following simulation studies by generating 100 replicated data sets from our defined GLMEMs with sample size $n = 200$ together with $m = 5$.

In the first simulation study, each of 100 replicated data sets $\{(y_i, \boldsymbol{v}_i, \boldsymbol{w}_i, \boldsymbol{x}_i) : i = 1, \ldots, n\}$ was generated from a Poisson distribution with the probability density $p(y_i|\mu_i) = \mu_i^{y_i} \exp(-\mu_i)/y_i!$ and $\eta_i = \log(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}_x + \boldsymbol{v}_i^\top \boldsymbol{\beta}_v$, where $\boldsymbol{v}_i$'s were generated from a multivariate normal distribution with mean zero and covariance matrix $0.25\boldsymbol{I}_3$ and components x_{1i} and x_{2i} in $\boldsymbol{x}_i$ were generated via Eq. (2.9). In this case, $\phi = 1$ relating to Eq. (2.1) is a constant. The true values of $\boldsymbol{\beta}_x = (\beta_0, \beta_1)^\top$, $\boldsymbol{\beta}_v = (\beta_2, \beta_3, \beta_4)^\top$, $\boldsymbol{\gamma}_k^* = (\gamma_{k0}, \gamma_{k1}, \gamma_{k2}, \gamma_{k3})^\top$ and σ_x^2 were taken to be $\boldsymbol{\beta}_x = (0.4, -0.3)^\top$, $\boldsymbol{\beta}_v = (0.4, 0.3, 0.4)^\top$, $\boldsymbol{\gamma}_k^* = (0.2, 0.2, 0.2, 0.5)^\top$ for $k = 1$ and 2, and $\sigma_x^2 = 1$, respectively. To test the effectiveness of using the TCDP prior to approximate distributions of MEs $\boldsymbol{u}_{ij} = (u_{ij1}, u_{ij2})^\top$, we considered the following eight distributional assumptions for u_{ijk}.

Assumption 1 We assumed the distribution of u_{ijk} to be $u_{ijk} \overset{\text{i.i.d.}}{\sim} N(0, 1.2^2)$ for $k = 1$ and 2.

Assumption 2 We assumed the distribution of u_{ijk} to be bimodal: $u_{ijk} \overset{\text{i.i.d.}}{\sim} 0.6N(-0.4, 0.2^2) + 0.4N(0.6, 0.2^2)$ for $k = 1$ and 2.

Assumption 3 We took the distribution of u_{ijk} to be trimodal: $u_{ijk} \overset{\text{i.i.d.}}{\sim} 0.3N(0.5, 0.1) + 0.2N(3, 0.1) + 0.5N(-1.5, 0.1)$ for $k = 1$ and 2.

Assumption 4 We set the distribution of u_{ijk} to be multimodal: $u_{ijk} \overset{\text{i.i.d}}{\sim} 0.3N(0.5, 0.1) + 0.2N(3, 0.1) + 0.1N(-3.5, 0.1) + 0.4N(-1, 0.1)$ for $k = 1$ and 2.

The above four assumptions were used to illustrate that even when the assumed distribution is multimodal, our presented TCDP prior can still capture their characteristics.

Assumption 5 We took $u_{ijk} = u_{ijk}^* - 1$ with $u_{ijk}^* \overset{\text{i.i.d.}}{\sim} \Gamma(1, 1)$.

Assumption 6 We specified the distribution of u_{ijk} to be $u_{ijk} \overset{\text{i.i.d.}}{\sim} 0.5N(-0.5, 0.1) + 0.5\Gamma(1, 2)$.

Assumption 7 We set $u_{ijk} = 0.75(u_{ij1}^* - 0.5) + 0.25(u_{ij2}^* - 1)$ with $u_{ij1}^* \overset{\text{i.i.d.}}{\sim} \Gamma(4, 8)$ and $u_{ij2}^* \overset{\text{i.i.d.}}{\sim} \Gamma(1, 1)$.

Assumption 8 We took $u_{ijk} = u^*_{ijk} - 0.5$ with $u^*_{ijk} \overset{\text{i.i.d.}}{\sim} \text{Beta}(4, 4)$.

The above presented four assumptions were designed to generate the skewed distribution for u_{ijk}. Based on the generated u_{ijk}'s, we obtained data set $\{w_{ijk} : i = 1, \ldots, n, j = 1, \ldots, m, k = 1, 2\}$ via Eq. (2.3).

The hyperparameter values of the prior distributions for τ and σ_x^2 were specified as follows. To ensure that our presented DP mixture approximations were not biased with respect to the selection of our hyperparameters, we set $\varphi_j^a = 3$, $\varphi_j^c = 150$ and $\varphi_j^d = 20$ for $j = 1, \ldots, r$, $c_1 = 5$ and allowed c_2 to be generated randomly from a uniform distribution $U(2, 6)$. For the hyperparameters relating to the prior distribution of τ, we set $a_1 = 200$ and $a_2 = 10$ to generate large values of τ which lead to more unique covariate MEs. For the conjugate prior of σ_x^{-2}, we set $c_3 = 10$ and randomly generated c_4 from a uniform distribution $U(9, 10)$. For the hyperparameters $\boldsymbol{\xi}^0$ and $\boldsymbol{\Psi}^0$, we took $\boldsymbol{\xi}^0 = \mathbf{0}$ and $\boldsymbol{\Psi}^0 = \boldsymbol{I}_r$ to satisfy the condition of the centered DP procedure. Also, to investigate sensitivity of Bayesian estimates to prior inputs, we considered the following three types of priors for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_k$.

- *Type A* The hyperparameters corresponding to the priors of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_k$ were taken to be $\boldsymbol{\beta}^0 = (0.4, -0.3, 0.4, 0.3, 0.4)^\top$, $\boldsymbol{H}_\beta^0 = 0.25\boldsymbol{I}_5$, $\boldsymbol{\gamma}_k^0 = (0.2, 0.2, 0.2, 0.5)^\top$ and $\boldsymbol{H}_{k\gamma}^0 = 0.25\boldsymbol{I}_4$. This can be regarded as a situation with good prior information.
- *Type B* The hyperparameters corresponding to the priors of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_k$ were taken to be $\boldsymbol{\beta}^0 = 1.5 \times (0.4, -0.3, 0.4, 0.3, 0.4)^\top$, $\boldsymbol{H}_\beta^0 = 0.75\boldsymbol{I}_5$, $\boldsymbol{\gamma}_k^0 = 1.5 \times (0.2, 0.2, 0.2, 0.5)^\top$ and $\boldsymbol{H}_{k\gamma}^0 = 0.75\boldsymbol{I}_4$. This can be regarded as a situation with inaccurate prior information.
- *Type C* The hyperparameters corresponding to the priors of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_k$ were taken to be $\boldsymbol{\beta}^0 = (0, 0, 0, 0, 0)^\top$, $\boldsymbol{H}_\beta^0 = 10\boldsymbol{I}_5$, $\boldsymbol{\gamma}_k^0 = (0, 0, 0, 0)^\top$ and $\boldsymbol{H}_{k\gamma}^0 = 10\boldsymbol{I}_4$. This can be regarded as a situation with noninformative prior information.

For each of the generated 100 data sets, the preceding proposed MCMC algorithm with $G = 50$ was used to evaluate Bayesian estimates of unknown parameters and covariates subject to MEs for each of three types of priors based on three different starting values of unknown parameters. The estimated potential scale reduction (EPSR) values (Gelman et al. 1996) for all unknown parameters were computed. For the first five test runs, we observed that the EPSR values of all unknown parameters were less than 1.2 after 10,000 iterations. Hence, $\mathfrak{L} = 5000$ observations after 10,000 burn-in iterations were collected to evaluate Bayesian estimates via Eq. (3.1). Results under eight assumptions together with three types of prior inputs were presented in Table 1, where 'Bias' was the absolute difference between the true value and the mean of the estimates based on 100 replications and 'RMS' was the root mean square between the estimates based on 100 replications and its true value. Also, for comparison, we calculated Bayesian estimates of $\boldsymbol{\beta}$ for each of the above generated 100 data sets under eight distributional assumptions of u_{ijk} on the basis of a GLM without error modelling. The corresponding results were given in Table 2.

Examination of Tables 1 and 2 indicated that (i) Bayesian estimates were reasonably accurate regardless of distributional assumptions of u_{ij} and prior inputs of unknown parameters because their Bias values were less than 0.10 and their RMS values were less than 0.20; (ii) Bayesian estimates were not sensitive to prior inputs of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_k$

Table 1 Performance of parameter estimates in the first simulation study

Par.	Type A		Type B		Type C		Type A		Type B		Type C	
	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS
	Assumption 1						Assumption 2					
γ_{10}	0.003	0.079	0.001	0.085	0.003	0.087	0.002	0.055	0.003	0.071	0.003	0.078
γ_{11}	0.000	0.144	0.004	0.126	0.017	0.167	0.010	0.116	0.003	0.130	0.007	0.141
γ_{12}	0.013	0.154	0.043	0.160	0.021	0.162	0.001	0.122	0.010	0.140	0.001	0.129
γ_{13}	0.025	0.145	0.002	0.159	0.001	0.164	0.012	0.123	0.010	0.147	0.004	0.133
γ_{20}	0.003	0.089	0.011	0.085	0.003	0.081	0.000	0.073	0.009	0.067	0.012	0.080
γ_{21}	0.011	0.142	0.010	0.165	0.018	0.164	0.010	0.133	0.001	0.124	0.007	0.136
γ_{22}	0.015	0.154	0.021	0.175	0.008	0.154	0.002	0.126	0.007	0.159	0.002	0.140
γ_{23}	0.001	0.135	0.015	0.156	0.010	0.170	0.023	0.133	0.015	0.127	0.019	0.143
β_0	0.007	0.064	0.006	0.068	0.008	0.066	0.003	0.055	0.007	0.059	0.006	0.065
β_1	0.012	0.080	0.011	0.077	0.005	0.081	0.010	0.065	0.000	0.060	0.002	0.063
β_2	0.017	0.107	0.087	0.154	0.007	0.147	0.004	0.125	0.096	0.156	0.007	0.133
β_3	0.003	0.118	0.023	0.134	0.007	0.142	0.020	0.142	0.004	0.123	0.012	0.132
β_4	0.022	0.122	0.021	0.127	0.002	0.137	0.001	0.108	0.017	0.134	0.017	0.138
σ_x^2	0.011	0.077	0.027	0.092	0.015	0.090	0.003	0.071	0.004	0.071	0.007	0.070
	Assumption 3						Assumption 4					
γ_{10}	0.000	0.070	0.003	0.076	0.001	0.071	0.001	0.064	0.010	0.068	0.020	0.070
γ_{11}	0.009	0.133	0.032	0.150	0.017	0.143	0.015	0.151	0.000	0.148	0.005	0.149
γ_{12}	0.008	0.126	0.006	0.133	0.012	0.157	0.000	0.140	0.000	0.134	0.018	0.153
γ_{13}	0.012	0.147	0.013	0.142	0.009	0.144	0.002	0.125	0.011	0.125	0.003	0.149
γ_{20}	0.004	0.075	0.000	0.078	0.009	0.080	0.009	0.068	0.018	0.077	0.006	0.069
γ_{21}	0.015	0.134	0.004	0.148	0.008	0.156	0.010	0.126	0.020	0.134	0.005	0.132
γ_{22}	0.001	0.124	0.019	0.136	0.001	0.142	0.002	0.150	0.009	0.121	0.016	0.136
γ_{23}	0.014	0.135	0.024	0.146	0.017	0.141	0.009	0.121	0.005	0.152	0.007	0.129
β_0	0.005	0.056	0.001	0.054	0.001	0.055	0.008	0.063	0.005	0.063	0.001	0.055
β_1	0.015	0.061	0.004	0.074	0.004	0.058	0.003	0.060	0.012	0.060	0.003	0.064
β_2	0.017	0.114	0.085	0.154	0.008	0.128	0.002	0.120	0.020	0.131	0.015	0.128
β_3	0.009	0.119	0.007	0.117	0.017	0.133	0.004	0.117	0.012	0.132	0.006	0.147
β_4	0.009	0.129	0.010	0.140	0.010	0.141	0.018	0.132	0.022	0.133	0.006	0.147
σ_x^2	0.004	0.068	0.008	0.075	0.013	0.067	0.012	0.072	0.011	0.064	0.019	0.069
	Assumption 5						Assumption 6					
γ_{10}	0.011	0.071	0.006	0.081	0.005	0.074	0.002	0.061	0.007	0.073	0.014	0.075
γ_{11}	0.002	0.137	0.014	0.135	0.002	0.144	0.015	0.130	0.015	0.144	0.019	0.137
γ_{12}	0.014	0.126	0.005	0.157	0.013	0.130	0.016	0.121	0.000	0.136	0.014	0.132
γ_{13}	0.004	0.136	0.003	0.131	0.007	0.138	0.029	0.127	0.014	0.139	0.009	0.127
γ_{20}	0.001	0.069	0.002	0.080	0.008	0.073	0.008	0.061	0.001	0.063	0.010	0.086
γ_{21}	0.018	0.138	0.008	0.158	0.010	0.135	0.002	0.116	0.005	0.141	0.019	0.120
γ_{22}	0.013	0.148	0.009	0.125	0.017	0.134	0.001	0.128	0.010	0.153	0.014	0.152
γ_{23}	0.009	0.148	0.015	0.148	0.003	0.149	0.005	0.126	0.032	0.151	0.026	0.141

Table 1 continued

Par.	Type A		Type B		Type C		Type A		Type B		Type C	
	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS
β_0	0.006	0.063	0.000	0.065	0.008	0.065	0.012	0.058	0.008	0.051	0.013	0.058
β_1	0.000	0.064	0.000	0.067	0.004	0.064	0.008	0.059	0.003	0.066	0.019	0.068
β_2	0.010	0.123	0.113	0.173	0.004	0.120	0.011	0.112	0.082	0.141	0.010	0.137
β_3	0.006	0.107	0.013	0.137	0.001	0.135	0.011	0.120	0.001	0.119	0.002	0.109
β_4	0.003	0.115	0.017	0.148	0.018	0.157	0.003	0.128	0.013	0.142	0.020	0.122
σ_x^2	0.019	0.067	0.015	0.074	0.007	0.077	0.014	0.067	0.005	0.066	0.010	0.057
	Assumption 7						Assumption 8					
γ_{10}	0.000	0.062	0.009	0.082	0.015	0.067	0.000	0.074	0.014	0.065	0.012	0.069
γ_{11}	0.008	0.134	0.021	0.120	0.007	0.142	0.014	0.133	0.000	0.156	0.040	0.154
γ_{12}	0.006	0.112	0.005	0.114	0.002	0.113	0.007	0.110	0.006	0.102	0.026	0.126
γ_{13}	0.008	0.119	0.019	0.126	0.030	0.140	0.004	0.121	0.009	0.102	0.007	0.127
γ_{20}	0.001	0.081	0.002	0.072	0.012	0.067	0.007	0.066	0.002	0.064	0.005	0.074
γ_{21}	0.000	0.134	0.014	0.122	0.001	0.133	0.022	0.113	0.044	0.128	0.007	0.137
γ_{22}	0.000	0.107	0.027	0.129	0.007	0.127	0.008	0.124	0.007	0.109	0.011	0.128
γ_{23}	0.002	0.144	0.034	0.159	0.011	0.135	0.016	0.135	0.036	0.144	0.016	0.136
β_0	0.000	0.056	0.002	0.049	0.012	0.060	0.008	0.057	0.001	0.060	0.001	0.054
β_1	0.013	0.067	0.008	0.066	0.010	0.062	0.002	0.068	0.006	0.060	0.000	0.083
β_2	0.011	0.128	0.122	0.158	0.022	0.118	0.011	0.111	0.079	0.143	0.024	0.125
β_3	0.011	0.109	0.024	0.114	0.016	0.115	0.000	0.117	0.018	0.111	0.036	0.115
β_4	0.000	0.112	0.003	0.135	0.015	0.124	0.004	0.112	0.021	0.113	0.011	0.114
σ_x^2	0.002	0.075	0.008	0.069	0.008	0.066	0.000	0.062	0.018	0.061	0.001	0.066

under our considered three prior inputs; (iii) Bayesian estimates obtained from the type A prior input behaved better than those obtained from the type B and type C prior inputs in terms of Bias and RMS, but their differences were minor; (iv) Bayesian estimates obtained from the type A and type B prior inputs were slightly better than those obtained from the type C prior input, but their differences were negligible; (iv) our proposed semiparametric Bayesian method produced smaller bias and RMS values than a Bayesian approach to a GLM without error modelling.

To investigate the accuracy of using TCDP prior to approximate distribution of u_{ijk}, we calculated means and standard deviations of $\hat{u}_{ijk}$'s across individuals and plotted the true densities of u_{ijk}'s against their corresponding approximated densities for a randomly selected replication. Table 3 presented the estimated means and standard deviations of u_{ijk}'s for our considered eight assumptions. To save space, we only plotted densities of u_{ijk} and $\hat{u}_{ijk}$ for Assumption 4 in Fig. 1. Examination of Table 3 and Fig. 1 implied that (i) the TCDP prior approximations to distributions of u_{ijk}'s were flexible enough to recover the shapes of u_{ijk}'s distributions for our considered eight distributional assumptions of u_{ijk}; (ii) the mean and standard deviation of the true distribution of u_{ijk} can be estimated well by our proposed method.

Table 2 Performance of parameter estimates for a GLM without error modelling in the first simulation study

Par.	Type A		Type B		Type C		Type A		Type B		Type C	
	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS
	Assumption 1						Assumption 2					
β_0	0.009	0.108	0.016	0.118	0.007	0.115	0.022	0.222	0.044	0.224	0.099	0.220
β_1	0.014	0.128	0.010	0.135	0.016	0.136	0.009	0.232	0.075	0.240	0.018	0.245
β_2	0.005	0.113	0.005	0.118	0.005	0.120	0.018	0.133	0.013	0.138	0.021	0.144
β_3	0.012	0.112	0.017	0.119	0.014	0.121	0.008	0.125	0.011	0.130	0.007	0.135
β_4	0.020	0.112	0.021	0.116	0.024	0.120	0.001	0.132	0.005	0.137	0.001	0.141
	Assumption 3						Assumption 4					
β_0	0.013	0.200	0.066	0.283	0.015	0.348	0.018	0.156	0.005	0.224	0.052	0.285
β_1	0.003	0.251	0.048	0.342	0.001	0.416	0.017	0.178	0.014	0.244	0.047	0.308
β_2	0.003	0.134	0.008	0.140	0.002	0.144	0.004	0.128	0.003	0.133	0.008	0.138
β_3	0.011	0.117	0.014	0.121	0.009	0.124	0.001	0.142	0.006	0.147	0.001	0.152
β_4	0.005	0.115	0.002	0.121	0.008	0.125	0.012	0.144	0.003	0.151	0.011	0.155
	Assumption 5						Assumption 6					
β_0	0.029	0.132	0.025	0.139	0.032	0.144	0.019	0.225	0.001	0.283	0.041	0.327
β_1	0.008	0.148	0.013	0.158	0.008	0.163	0.004	0.234	0.028	0.297	0.001	0.341
β_2	0.007	0.120	0.010	0.126	0.006	0.129	0.004	0.112	0.007	0.118	0.003	0.120
β_3	0.003	0.119	0.007	0.125	0.002	0.128	0.001	0.127	0.001	0.133	0.001	0.137
β_4	0.011	0.142	0.015	0.148	0.010	0.153	0.018	0.126	0.015	0.129	0.022	0.136
	Assumption 7						Assumption 8					
β_0	0.013	0.125	0.019	0.133	0.011	0.138	0.014	0.211	0.075	0.238	0.096	0.260
β_1	0.047	0.165	0.058	0.182	0.051	0.188	0.023	0.209	0.128	0.242	0.069	0.242
β_2	0.020	0.133	0.015	0.139	0.022	0.143	0.006	0.145	0.001	0.152	0.008	0.155
β_3	0.006	0.114	0.010	0.120	0.006	0.122	0.001	0.116	0.003	0.122	0.002	0.126
β_4	0.012	0.125	0.015	0.132	0.011	0.134	0.021	0.128	0.025	0.135	0.019	0.138

To illustrate our proposed Bayesian case-deletion influence measures, we conducted the second simulation study. In this simulation study, the data set $\{(y_i, \boldsymbol{v}_i, \boldsymbol{w}_i, \boldsymbol{x}_i) : i = 1, \ldots, 200\}$ was generated by using the same setup as specified in the first simulation study, but outliers were created by changing y_i as $y_i + 30$ for $i = 1, 100$ and 150. We calculated the corresponding values of diagnostics $KL(i)$ and $CD(i)$ for the above generated data set including outliers. Results were presented in Fig. 2. Examination of Fig. 2 indicated that cases 1, 100 and 150 were detected to be influential as expected.

4.2 An example

To illustrate our proposed methods, we considered a data set from Framingham heart study, which has been analyzed by Carroll et al. (2006, Section 9.10) and Muff et al. (2015) via a logistic ME model with the normality assumption of covariate ME. The

Table 3 Means and standard deviations (SD) of u_{ijk} and $\hat{u}_{ijk}$ in the simulation studies

Assump.	Mean	SD	Type A		Type B		Type C	
			EMean	ESD	EMean	ESD	EMean	ESD
	u_{ij1}		$\hat{u}_{ij1}$					
1	0.0	1.200	0.002	1.110	0.002	1.108	0.002	1.113
2	0.0	0.583	0.000	0.522	0.001	0.522	0.000	0.521
3	0.0	1.761	−0.001	1.756	−0.001	1.756	0.000	1.758
4	0.0	1.962	0.007	1.899	0.009	1.896	0.010	1.896
5	0.0	1.000	−0.001	0.970	0.000	0.970	−0.001	0.958
6	0.0	0.612	−0.008	0.636	−0.007	0.642	−0.007	0.636
7	0.0	0.882	0.001	0.844	0.001	0.845	0.002	0.846
8	0.0	0.167	0.000	0.150	0.000	0.150	0.000	0.150
	u_{ij2}		$\hat{u}_{ij2}$					
1	0.0	1.200	0.000	1.108	−0.001	1.112	−0.002	1.106
2	0.0	0.583	0.000	0.524	0.001	0.521	0.000	0.522
3	0.0	1.761	−0.001	1.756	−0.001	1.757	0.000	1.756
4	0.0	1.962	0.007	1.897	0.009	1.897	0.009	1.899
5	0.0	1.000	−0.001	0.972	−0.001	0.975	−0.001	0.975
6	0.0	0.612	−0.007	0.638	−0.007	0.639	−0.008	0.636
7	0.0	0.882	0.001	0.839	0.000	0.824	0.001	0.833
8	0.0	0.167	0.000	0.149	0.000	0.150	0.000	0.149

Mean and SD denote empirical mean and standard deviation of the generated random samples, respectively; EMean and ESD represent mean and standard deviation of the sampled posterior samples, respectively

data set consisted of a series of exams taken over two years. Here, we only analyzed the data set from exam 3 with $n = 1615$ men aged between 31 and 65. We took y_i to be the indicator for coronary heart disease, which was assumed to follow a Bernoulli distribution, v_i to be the indicator for smoking and x_i to be the transformed (unobserved) long-term blood pressure (i.e., $x_i = \log(\text{SBP}_i - 50)$), where SBP was an abbreviation of the systolic blood pressure. Since it is impossible to measure the long-term SBP, measurements at single clinical visits had to be used as a proxy (Muff et al. 2015). Also, due to daily variations or deviations in the measurement instrument, the single-visit measures might considerably differ from the long-term blood pressure (Carroll et al. 2006). Hence, to estimate the magnitude of the error, SBP had been measured twice at different examinations. The two proxy measures for x_i were denoted as w_{i1} and w_{i2}, respectively. Following Muff et al. (2015), we fitted the data set via the following logistic ME model (LOGMEM):

$$\text{logit}\{\Pr(y_i = 1|x_i, v_i)\} = \beta_0 + x_i\beta_x + v_i\beta_v, \quad w_{ij} = x_i + u_{ij}, \quad x_i = \gamma_0 + \gamma_v v_i + \varepsilon_i$$

for $i = 1, \ldots, n$ and $j = 1, 2$, where $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma_x^2)$.

To make Bayesian analysis for the above considered LOGMEM, we assumed the following prior distributions for $\boldsymbol{\gamma} = (\gamma_0, \gamma_v)^\top$ and $\boldsymbol{\beta} = (\beta_0, \beta_x, \beta_v)^\top$: $p(\boldsymbol{\gamma}) \sim N_2(\boldsymbol{\gamma}^0, 0.25\boldsymbol{I}_2)$ and $p(\boldsymbol{\beta}) \sim N_3(\boldsymbol{\beta}^0, 0.25\boldsymbol{I}_3)$, where the hyperparameters $\boldsymbol{\gamma}^0$ and

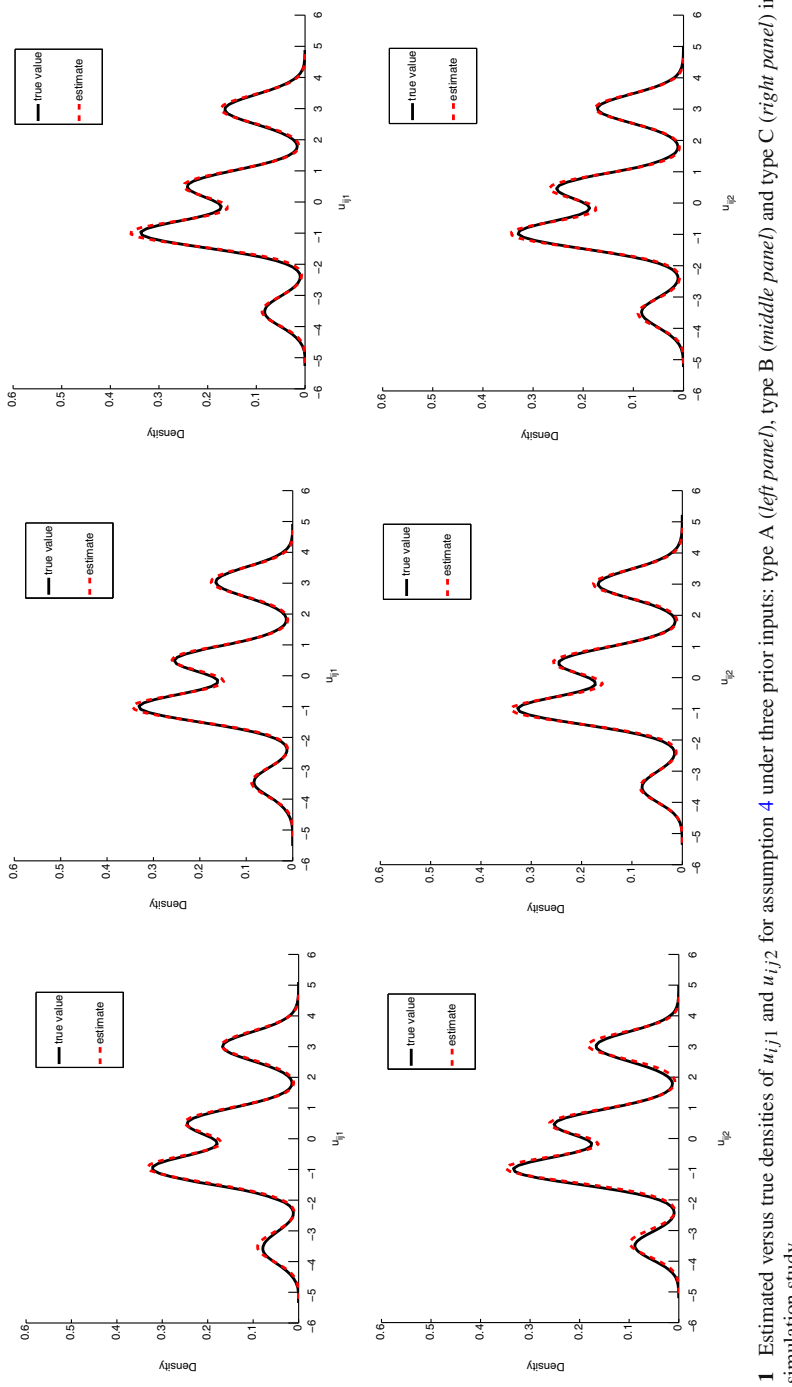

Fig. 1 Estimated versus true densities of u_{ij1} and u_{ij2} for assumption 4 under three prior inputs: type A (*left panel*), type B (*middle panel*) and type C (*right panel*) in the first simulation study

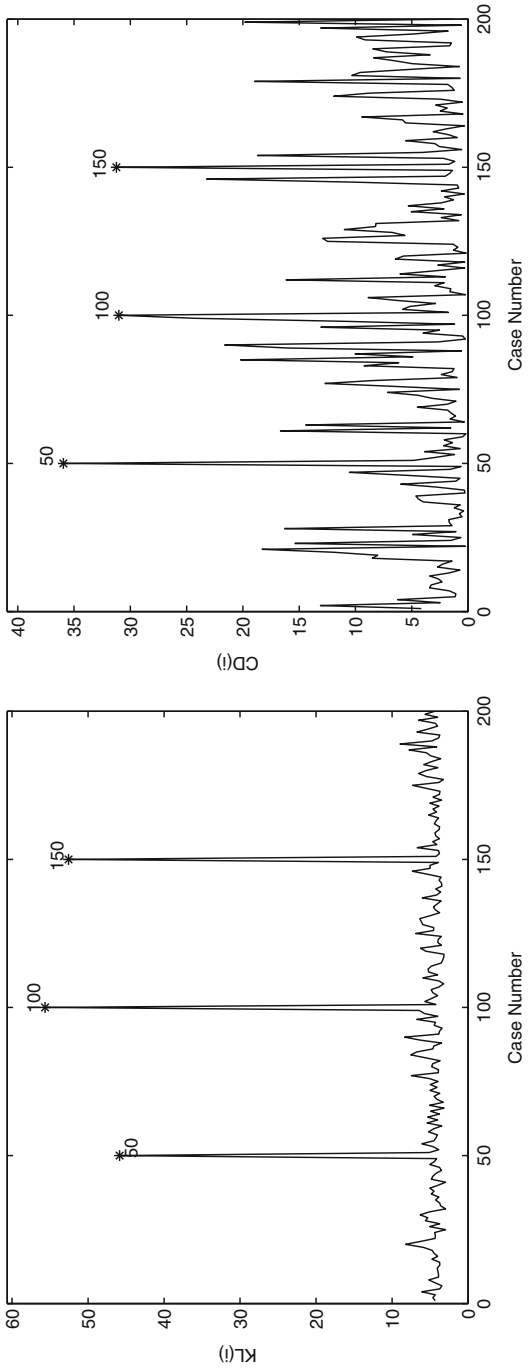

Fig. 2 Index plots of $KL(i)$ (*left panel*) and $CD(i)$ (*right panel*) in the simulation study

Table 4 Bayesian estimates (Est) and standard errors (SE) of parameters in the real example

Parameter	Full data set		Reduced data set		Without ME	
	Est	SE	Est	SE	Est	SE
β_0	4.409	0.018	4.394	0.018	−2.877	0.478
β_x	−0.047	0.021	−0.034	0.020	0.038	0.115
β_v	8.032	0.460	3.039	0.453	0.342	0.211
γ_0	−1.305	0.324	−0.174	0.115	–	–
γ_v	−0.289	0.201	−0.151	0.193	–	–
σ_x^2	0.118	0.004	0.112	0.004	–	–

$\boldsymbol{\beta}^0$ were taken to be their corresponding Bayesian estimates obtained from the non-informative priors on $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ (e.g., $p(\boldsymbol{\gamma}) \sim N_2(\mathbf{0}, 20\boldsymbol{I}_2)$ and $p(\boldsymbol{\beta}) \sim N_3(\mathbf{0}, 20\boldsymbol{I}_3)$). For the hyperparameters a_1 and a_2, we took $a_1 = 250$ and set a_2 to be a value generated randomly from a uniform distribution $U(25, 30)$ to yield large values of τ leading to more unique covariate MEs. For the hyperparameters c_3 and c_4, we randomly generated c_3 from a uniform distribution $U(1, 10)$ and randomly sampled c_4 from a uniform distribution $U(1, 100)$ to yield relatively diffuse values of σ_x^2. For the hyperparameters $\boldsymbol{\xi}^0$ and $\boldsymbol{\Psi}^0$, we took $\boldsymbol{\xi}^0 = \mathbf{0}$ and $\boldsymbol{\Psi}^0 = \boldsymbol{I}_2$ to satisfy the condition of the centered DP procedure. Similar to simulation studies, we set $\varphi_j^a = 3$, $\varphi_j^c = 100$ and $\varphi_j^d = 20$ for $j = 1, \ldots, r$, $c_1 = 5$ and allowed c_2 to be generated randomly from a uniform distribution $U(2, 6)$. The preceding presented MCMC algorithm with $G = 250$ was used to obtain Bayesian estimates of parameters and MEs u_{ij}'s. Similarly, the EPSR values of all unknown parameters were computed by using three parallel sequences of observations generated from three different starting values of unknown parameters. Their EPSR values were less than 1.2 after about 20,000 iterations. We collected 10,000 observations after 20,000 burn-in iterations to evaluate Bayesian estimates of parameters. Results were given in Table 4. Examination of Table 4 showed that the SBP has a positive effect on the coronary heart disease, whilst smoking has a slightly negative effect on the coronary heart disease and the SBP. Also, for comparison, we evaluated Bayesian estimates of parameters in the following logistic model: $\text{logit}\{\Pr(y_i = 1|x_i, v_i)\} = \beta_0 + x_i\beta_x + v_i\beta_v$ for $i = 1, \ldots, n$ under the above specified prior of $\boldsymbol{\beta} = (\beta_0, \beta_x, \beta_v)^\top$, where $x_i = (w_{i1} + w_{i2})/2$. The corresponding results were presented in Table 4, which showed that our considered logistic ME model leaded to a smaller estimate and a smaller SE of parameter β_x than the logistic model without ME.

To illustrate our proposed case deletion influence measures, we computed the values of diagnostics $KL(i)$ and $CD(i)$, which were presented in Fig. 3. Examination of Fig. 3 indicated that cases 10, 59, 207, 208, 222, 362, 367, 386, 391, 456, 501, 530, 533, 709, 976, 1093, 1096, 1162, 1187, 1336, 1430 and 1502 were detected to be influential by diagnostics $KL(i)$ and $CD(i)$. To investigate the effect of these influential observations on Bayesian estimates of unknown parameters, we also calculated Bayesian estimates of unknown parameters for our considered data set with these influential cases deleted. The corresponding results were given in Table 4. Examination of Table 4 indicated that

these influential individuals have a relatively large influence on Bayesian estimates of β_x and β_v.

5 Discussion

We discussed Bayesian estimates of unknown parameters and Bayesian case-deletion diagnostics for generalized linear mixed models with covariates subject to MEs. Under the unknown distribution assumptions of random MEs, we used the TCDP mixture model to approximate the distribution of random ME. We also obtained Bayesian estimates of unknown parameters and random MEs and their standard errors, and presented two Bayesian case-deletion influence diagnostics to detect influential observations. Simulation studies and a real example were used to illustrate our proposed methodologies. The empirical results showed that (i) the TCDP mixture model approximation can well capture characteristics of the distribution for random ME; and (ii) our proposed methods can be used to effectively detect influential observations.

This paper considered the balanced repeated measurement for the covariate subject to ME so that we can use the TCDP mixture model to approximate the distribution of random ME. When an unbalanced repeated measurement for the covariate subject to ME is considered, other methods may be employed to address the issue, which is our further work.

Acknowledgements The authors are grateful for the Editor, an Associate Editor and two referees for their valuable suggestions and comments that greatly improved the manuscript. The research was supported by grants from the National Science Fund for Distinguished Young Scholars of China (11225103), and the National Natural Science Foundation of China (11561074), and Research Fund for the Doctoral Program of Higher Education of China (20115301110004). We also thank Mahmoud Torabi for providing the Framingham Heart Study dataset.

Appendix: Conditional distributions

To obtain Bayesian estimates of unknown parameters and covariates subject to MEs in our considered GLMEMs, the Gibbs sampler is adopted to draw a sequence of random observations from the joint posterior distribution $p(\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\varphi}^b, \boldsymbol{\pi}, \boldsymbol{\alpha}^*, \boldsymbol{\Omega}, \boldsymbol{L}, \tau, \boldsymbol{\beta}, \phi, \boldsymbol{\gamma}, \sigma_x^2, \boldsymbol{u}, \boldsymbol{x}|\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{v})$. The Gibbs sampler is implemented by iteratively drawing observations from the following conditional distributions: $p(\boldsymbol{\xi}|\boldsymbol{\alpha}^*, \boldsymbol{\Psi})$, $p(\boldsymbol{\Psi}|\boldsymbol{\alpha}^*, \boldsymbol{\xi})$, $p(\tau|\boldsymbol{\pi})$, $p(\boldsymbol{\varphi}^b|\boldsymbol{\Omega})$, $p(\boldsymbol{\pi}|\boldsymbol{L}, \tau)$, $p(\boldsymbol{L}|\boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{u})$, $p(\boldsymbol{\alpha}^*|\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\Omega}, \boldsymbol{L}, \boldsymbol{u})$, $p(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \phi)$, $p(\boldsymbol{\Omega}|\boldsymbol{\alpha}, \boldsymbol{\varphi}^b, \boldsymbol{L}, \boldsymbol{u})$, $p(\phi|\boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v})$, $p(\boldsymbol{\gamma}|\boldsymbol{x}, \boldsymbol{v}, \sigma_x^2)$, $p(\sigma_x^2|\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{\gamma})$, $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{u}, \boldsymbol{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi, \sigma_x^2, \boldsymbol{\gamma}, \boldsymbol{L})$ and $p(\boldsymbol{u}|\boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{L}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{\theta}_u)$. The conditional distributions required in implementing the above Gibbs sampler are summarized as follows.

Steps (a)–(h) Conditional distributions related to the nonparametric components

To sample $\boldsymbol{u}_{ij}$ in terms of the latent variable L_{ij} for $i = 1, \ldots, n$ and $j = 1, \ldots, m$, we first generate $\boldsymbol{\alpha}^* = (\boldsymbol{\alpha}_1^*, \ldots, \boldsymbol{\alpha}_G^*)$ and $\boldsymbol{\Omega} = (\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_G)$ from their corresponding

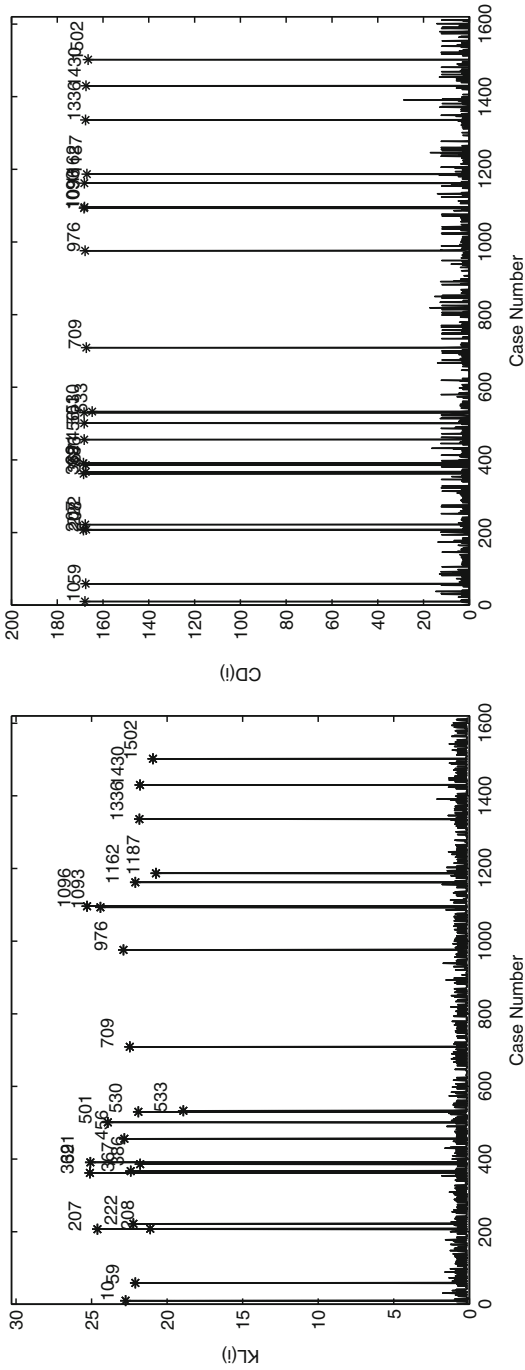

Fig. 3 Index plots of $KL(i)$ (*left panel*) and $CD(i)$ (*right panel*) in the real example

posterior distributions and then draw $\boldsymbol{u}_{ij}$ from the multivariate normal distribution $N_r(\boldsymbol{\alpha}_{L_{ij}}, \boldsymbol{\Omega}_{L_{ij}})$ with $\boldsymbol{\alpha}_{L_{ij}} = \boldsymbol{\alpha}^*_{L_{ij}} - \sum_{g=1}^{G} \pi_g \boldsymbol{\alpha}^*_g$. Since it is rather difficult to directly sample observations from the posterior distribution of $(\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\varphi}^b, \boldsymbol{\pi}, \boldsymbol{\alpha}^*, \boldsymbol{\Omega}, \boldsymbol{L}, \tau, \boldsymbol{u})$, the blocked Gibbs sampler is employed to solve the above difficulties. The conditional distributions relating to implement Gibbs sampling of the nonparametric components are given as follows.

Step (a) The conditional distribution for $\boldsymbol{\xi}$ is $p(\boldsymbol{\xi}|\boldsymbol{\alpha}^*, \boldsymbol{\Psi}) \sim N_r(\boldsymbol{\alpha}_\xi, \boldsymbol{\Sigma}_\xi)$, where $\boldsymbol{\Sigma}_\xi = (G\boldsymbol{\Psi}^{-1} + \boldsymbol{\Psi}^{0^{-1}})^{-1}$ and $\boldsymbol{\alpha}_\xi = \boldsymbol{\Sigma}_\xi(\boldsymbol{\Psi}^{0^{-1}}\boldsymbol{\xi}^0 + \boldsymbol{\Psi}^{-1}\sum_{g=1}^{G}\boldsymbol{\alpha}^*_g)$.
Step (b) For $j = 1, \ldots, r$, the jth diagonal element of $\boldsymbol{\Psi}$ given $(\boldsymbol{\alpha}^*, \boldsymbol{\xi})$ is distributed as $p(\psi_j^{-1}|\boldsymbol{\alpha}^*, \boldsymbol{\xi}) \sim \Gamma(c_1 + G/2, c_2 + \frac{1}{2}\sum_{g=1}^{G}(\alpha^*_{gj} - \xi_j)^2)$, where α^*_{gj} is the jth element of $\boldsymbol{\alpha}^*_g$ and ξ_j is the jth element of $\boldsymbol{\xi}$.
Step (c) Following the same argument of Chow et al. (2011), the conditional distribution $p(\tau|\boldsymbol{\pi})$ is given by $p(\tau|\boldsymbol{\pi}) \sim \Gamma\,(a_1 + G - 1, a_2 - \sum_{g=1}^{G}\log(1 - \nu^*_g))$, where ν^*_g is a random weight sampled from the beta distribution and is sampled in step (e).
Step (d) For $j = 1, \ldots, r$, the conditional distribution of φ^b_j is given by $p(\varphi^b_j|\boldsymbol{\Omega}) \sim \Gamma\,(\varphi^c_j, \varphi^d_j + \sum_{g=1}^{G}\omega^{-1}_{gj})$, where ω^{-1}_{gj} is the jth diagonal element of $\boldsymbol{\Omega}_g$.
Step (e) It can be shown that the conditional distribution $p(\boldsymbol{\pi}|\boldsymbol{L}, \tau)$ follows a generalized Dirichlet distribution, i.e., $p(\boldsymbol{\pi}|\boldsymbol{L}, \tau) \sim \wp(a^*_1, b^*_1, \cdots, a^*_{G-1}, b^*_{G-1})$, where $a^*_g = 1 + d_g, b^*_g = \tau + \sum_{j=g+1}^{G} d_j$ for $g = 1, \ldots, G-1$, and d_g is the number of L_{ij}'s (and thus individuals) whose value equals g. Sampling observations from the conditional distribution $p(\boldsymbol{\pi}|\boldsymbol{L}, \tau)$ can be conducted by (1) sampling ν^*_l from a Beta (a^*_l, b^*_l) distribution, (2) sampling $\pi_1, \ldots, \pi_G$ with the following expressions: $\pi_1 = \nu^*_1$, $\pi_g = \nu^*_g \prod_{j=1}^{g-1}(1 - \nu^*_j)$ for $g = 2, \ldots, G-1$ and $\pi_G = 1 - \sum_{l=1}^{G-1}\pi_l$.
Step (f) We consider the conditional distribution $p(\boldsymbol{\alpha}^*|\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\Omega}, \boldsymbol{L}, \boldsymbol{u})$. Let $L^*_1, \ldots, L^*_d$ be the d unique values of $\{L_{11}, \ldots, L_{1r}, \ldots, L_{n1}, \ldots L_{nr}\}$ (i.e., unique number of "clusters"). For $g = 1, \ldots, G$, $\boldsymbol{\alpha}^*_g$ is drawn from the following conditional distribution: $p(\boldsymbol{\alpha}^*_g|\boldsymbol{\xi}, \boldsymbol{\Psi}) \sim N_r(\boldsymbol{\xi}, \boldsymbol{\Psi})$ for $g \notin \{L^*_1, \ldots, L^*_d\}$, and $p(\boldsymbol{\alpha}^*_g|\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\Omega}, \boldsymbol{L}, \boldsymbol{u}) \sim N_r(\boldsymbol{A}_g, \boldsymbol{B}_g)$ with $\boldsymbol{B}_g = (\boldsymbol{\Psi}^{-1} + \sum_{\{(i,j):L_{ij}=g\}}\boldsymbol{\Omega}^{-1}_{L_{ij}})^{-1}$ and $\boldsymbol{A}_g = \boldsymbol{B}_g(\boldsymbol{\Psi}^{-1}\boldsymbol{\xi} + \sum_{\{(i,j):L_{ij}=g\}}\boldsymbol{\Omega}^{-1}_{L_{ij}}\boldsymbol{u}_{ij})$ for $l \in \{L^*_1, \ldots, L^*_d\}$. Given $\boldsymbol{\alpha}^*$, $\boldsymbol{\alpha}_g = \boldsymbol{\alpha}^*_g - \sum_{j=1}^{G}\pi_j\boldsymbol{\alpha}^*_j$ for $g = 1, \ldots, G$, $\boldsymbol{\alpha}^* = \{\boldsymbol{\alpha}^*_1, \ldots, \boldsymbol{\alpha}^*_G\}$ and $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_G\}$.
Step (g) The conditional distribution of $\boldsymbol{\Omega}$ is similar to the step (f). For $j = 1, \ldots, r$, the jth diagonal element of $\boldsymbol{\Omega}_g$ $(g = 1, \ldots, G)$ is generated from the following conditional distribution: $p(\omega^{-1}_{gj}|\boldsymbol{\alpha}, \boldsymbol{\varphi}^b, \boldsymbol{L}, \boldsymbol{u}) \sim \Gamma(\varphi^c_j, \varphi^b_j)$ for $g \notin \{L^*_1, \ldots, L^*_d\}$, and $p(\omega^{-1}_{gj}|\boldsymbol{\alpha}, \boldsymbol{\varphi}^b, \boldsymbol{L}, \boldsymbol{u}) \sim \Gamma(\varphi^c_j + d_g/2, \varphi^b_j + \frac{1}{2}\sum_{\{(i,\ell):L_{i\ell}=g\}}(u_{i\ell j} - \alpha_{gj})^2)$ for $g \in \{L^*_1, \ldots, L^*_d\}$, where $u_{i\ell j}$ is the jth element of $\boldsymbol{u}_{i\ell}$, α_{gj} is the jth element of $\boldsymbol{\alpha}_g$ and d_g is the number of $L_{i\ell}$'s whose value equals g for $i = 1, \ldots, n$ and $\ell = 1, \ldots, m$.

Step (h) The conditional distribution of L_{ij} can be shown to be $p(L_{ij}|\boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{u}) \overset{\text{i.i.d.}}{\sim} \text{Multinomial}(\pi^*_{ij1}, \ldots, \pi^*_{ijG})$, where π^*_{ijg} is proportional to $\pi_g\, p(\boldsymbol{u}_{ij}|\boldsymbol{\alpha}_g, \boldsymbol{\Omega}_g)$ and $p(\boldsymbol{u}_{ij}|\boldsymbol{\alpha}_g, \boldsymbol{\Omega}_g) \sim N_r(\boldsymbol{\alpha}_g, \boldsymbol{\Omega}_g)$ for $g = 1, \ldots, G$.
Step (i) Consider the conditional distribution of $\boldsymbol{u}_{ij}$. It is easily shown that the conditional distribution $p(\boldsymbol{u}_{ij}|\boldsymbol{\alpha}, \boldsymbol{\Omega}, L_{ij}, \boldsymbol{\theta}_u, \boldsymbol{w}, \boldsymbol{x}_i) \propto p(\boldsymbol{u}_{ij}|\boldsymbol{\alpha}_{L_{ij}}, \boldsymbol{\Omega}_{L_{ij}})\, p(\boldsymbol{w}_{ij}|\boldsymbol{u}_{ij}, \boldsymbol{x}_i, \boldsymbol{\theta}_u)$ is a non-standard distribution. Thus, we cannot directly sample $\boldsymbol{u}_{ij}$ from its conditional distribution. Here, the Metropolis–Hastings algorithm is employed to sample observation $\boldsymbol{u}_{ij}$ from its conditional distribution via the following steps. At the tth iteration with a current value $\boldsymbol{u}_i^{(t)}$, a new candidate $\boldsymbol{u}_{ij}$ is generated from the normal distribution $N_m(\boldsymbol{u}_{ij}^{(t)}, \sigma_u^2 \mathbb{D}_{u_{ij}})$, where $\mathbb{D}_{u_{ij}} = (\boldsymbol{\Omega}^{-1}_{L_{ij}} + \mathbb{C}_{ij})^{-1}$ with $\mathbb{C}_{ij} = -\partial^2 \log p(\boldsymbol{w}_{ij}|\boldsymbol{u}_{ij}, \boldsymbol{x}_i, \boldsymbol{\theta}_u)/\partial \boldsymbol{u}_{ij} \partial \boldsymbol{u}_{ij}^T|_{\boldsymbol{u}_{ij}=\boldsymbol{u}_{ij}^{(t)}}$. Then, the new candidate $\boldsymbol{u}_{ij}$ is accepted with probability

$$\min\left\{1, \frac{p(\boldsymbol{u}_{ij}|\boldsymbol{\alpha}_{L_{ij}}, \boldsymbol{\Omega}_{L_{ij}})\, p(\boldsymbol{w}_{ij}|\boldsymbol{u}_{ij}, \boldsymbol{x}_i, \boldsymbol{\theta}_u)}{p(\boldsymbol{u}_{ij}^{(m)}|\boldsymbol{\alpha}_{L_{ij}}, \boldsymbol{\Omega}_{L_{ij}})\, p(\boldsymbol{w}_{ij}|\boldsymbol{u}_{ij}^{(m)}, \boldsymbol{x}_i, \boldsymbol{\theta}_u)}\right\}.$$

The variance σ_u^2 can be chosen such that the average acceptance rate is approximately 0.25 or more.
Step (j) The conditional distribution $p(\phi^{-1}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \boldsymbol{\beta})$ is proportional to

$$\phi^{-(a_3+p+r-1)} \exp\left\{-\frac{1}{\phi}\left(a_4 + \frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}^0)^T (\boldsymbol{H}_\beta^0)^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}^0) - \sum_{i=1}^n (y_i\theta_i - b(\theta_i))\right) + \sum_{i=1}^n c(y_i, \phi)\right\},$$

which is generally a non-standard or familiar distribution. In this case, the Metropolized independence sampler algorithm (Liu 2001) can be employed to sample observations from the posterior $p(\phi|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \boldsymbol{\beta})$. At the tth iteration with a current value $\phi^{(t)}$, a new candidate ϕ is drawn from $h(\phi) \sim N(\phi^{(t)}, \sigma_\phi^2) I(0, \infty)$ and is accepted with probability

$$\min\left\{1, \frac{p(\phi|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \boldsymbol{\beta}) h(\phi^{(t)})}{p(\phi^{(t)}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \boldsymbol{\beta}) h(\phi)}\right\}.$$

The variance σ_ϕ^2 can be chosen such that the average acceptance rate is approximately 0.25 or more. Particularly, if $c(y_i, \phi) = c(y_i)/\phi$, we have $p(\phi^{-1}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \boldsymbol{\beta}) \sim \Gamma(r + p + a_3, a_4 + 0.5(\boldsymbol{\beta} - \boldsymbol{\beta}^0)^\top (\boldsymbol{H}_\beta^0)^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}^0) - \sum_{i=1}^n (y_i\theta_i - b(\theta_i) + c(y_i)))$. Also, if $c(y_i, \phi) = \zeta c(y_i)$, we obtain $p(\phi^{-1}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \boldsymbol{\beta}) \sim \Gamma(r + p + a_3, a_4 + 0.5(\boldsymbol{\beta} - \boldsymbol{\beta}^0)^\top (\boldsymbol{H}_\beta^0)^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}^0) - \sum_{i=1}^n (y_i\theta_i - b(\theta_i)))$, where ζ is a constant that does not depend on ϕ and y_i.
Step (k) The conditional distribution $p(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \phi)$ can be expressed as

$$p(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \phi) \propto \exp\left\{\frac{1}{\phi}\sum_{i=1}^n (y_i\theta_i - b(\theta_i)) - \frac{1}{2\phi}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top (\boldsymbol{H}_\beta^0)^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\}.$$

Similarly, the Metropolis–Hastings algorithm for simulating observations from the conditional distribution $p(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \phi)$ is implemented as follows. Given the current value $\boldsymbol{\beta}^{(t)}$, a new candidate $\boldsymbol{\beta}$ is generated from $N_{p+r}(\boldsymbol{\beta}^{(t)}, \sigma_\beta^2 \boldsymbol{\Sigma}_\beta)$ and is accepted with probability $\min\{1, p(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \phi)/p(\boldsymbol{\beta}^{(t)}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \phi)\}$, where $\boldsymbol{\Sigma}_\beta = \phi^{-1}\left(\frac{1}{2}\sum_{i=1}^n V_i \boldsymbol{C}_i^\top \boldsymbol{C}_i + (\boldsymbol{H}_\beta^0)^{-1}\right)^{-1}$ with $\boldsymbol{C}_i = (x_i, \boldsymbol{v}_i^\top)$ and $V_i = \dot{h}^{-2}(\mu_i)\ddot{b}^{-1}(\theta_i)$ in which $\dot{h}(a) = dh(a)/da$ and $\ddot{b}(a) = d^2b(a)/da^2$.

Step (l) The conditional distribution $p(\boldsymbol{\gamma}_k^*|\boldsymbol{x}, \boldsymbol{v}, \sigma_x^2)$ is given by $p(\boldsymbol{\gamma}_k^*|\boldsymbol{x}, \boldsymbol{v}, \sigma_x^2) \sim N(\boldsymbol{\mu}_{\gamma k}^*, \boldsymbol{\Omega}_{\gamma k}^*)$, where $\boldsymbol{\Omega}_{\gamma k}^* = (\sum_{i=1}^n \boldsymbol{v}_i^* \boldsymbol{v}_i^{*\top}/\sigma_x^2 + (\boldsymbol{H}_{\gamma k}^0)^{-1})^{-1}$ and $\boldsymbol{\mu}_{\gamma k}^* = \boldsymbol{\Omega}_{\gamma k}^*(\sum_{i=1}^n \boldsymbol{v}_i^* x_{ki}/\sigma_x^2 + (\boldsymbol{H}_{\gamma k}^0)^{-1}\boldsymbol{\gamma}_k^0)$ with $\boldsymbol{v}_i^* = (1, \boldsymbol{v}_i^\top)^\top$.

Step (m) It is easily shown that the conditional distribution $p(\sigma_x^2|\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{\gamma})$ is given by $p(\sigma_x^{-2}|\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{\gamma}) \sim \Gamma(c_3 + \frac{nr}{2}, c_4 + \frac{1}{2}\sum_{i=1}^n \sum_{k=1}^r (x_{ki} - \gamma_{k0} - \boldsymbol{\gamma}_{kv}^\top \boldsymbol{v}_i)^2)$.

Step (n) The conditional distribution $p(\boldsymbol{x}_i|y_i, \boldsymbol{v}_i, \boldsymbol{\gamma}, \boldsymbol{w}_i, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{\beta}, \sigma_x^2, \phi, \boldsymbol{L}_i)$ is proportional to

$$\exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi} - \sum_{k=1}^r \frac{(x_{ki} - \gamma_{k0} - \boldsymbol{\gamma}_{kv}^\top \boldsymbol{v}_i)^2}{2\sigma_x^2} - \frac{1}{2}\sum_{j=1}^m (\boldsymbol{w}_{ij} - \boldsymbol{x}_i - \boldsymbol{\alpha}_{L_{ij}})^\top \boldsymbol{\Omega}_{L_{ij}}^{-1} (\boldsymbol{w}_{ij} - \boldsymbol{x}_i - \boldsymbol{\alpha}_{L_{ij}})\right\},$$

where $\boldsymbol{L}_i = \{L_{ij} : j = 1, \ldots, m\}$.

The Metropolis–Hastings algorithm for sampling observations from $p(\boldsymbol{x}_i|y_i, \boldsymbol{v}_i, \boldsymbol{\gamma}, \boldsymbol{w}_i, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{\beta}, \sigma_x^2, \phi, \boldsymbol{L}_i)$ is implemented as follows. Given the current value $\boldsymbol{x}_i^{(t)}$, a new candidate $\boldsymbol{x}_i$ is generated from $N_r(\boldsymbol{x}_i^{(t)}, \sigma_a^2 \boldsymbol{H}_{x_i})$ and is accepted with probability

$$\min\left\{1, \frac{p(\boldsymbol{x}_i|y_i, \boldsymbol{v}_i, \boldsymbol{\gamma}, \boldsymbol{w}_i, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{\beta}, \sigma_x^2, \phi, \boldsymbol{L}_i)}{p(\boldsymbol{x}_i^{(t)}|y_i, \boldsymbol{v}_i, \boldsymbol{\gamma}, \boldsymbol{w}_i, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{\beta}, \sigma_x^2, \phi, \boldsymbol{L}_i)}\right\},$$

where $\boldsymbol{H}_{x_i} = (V_i \boldsymbol{\beta}_x \boldsymbol{\beta}_x^\top/\phi + \sum_{j=1}^m \Omega_{L_{ij}}^{-1} + \sigma_x^{-2}\boldsymbol{I}_r)^{-1}$ with $V_i = \dot{h}^{-2}(\mu_i)\ddot{b}^{-1}(\theta_i)|_{x_i = x_i^{(t)}}$. The variance σ_a^2 can be chosen such that the average acceptance rate is approximately 0.25 or more.

References

Aitkin M, Rocci R (2002) A general maximum likelihood analysis of measurement error in generalized linear models. Stat Comput 12:163–174

Battauz M (2011) Laplace approximation in measurement error models. Biom J 53:411–425

Battauz M, Bellio R (2011) Structural modeling of measurement error in generalized linear models with rasch measures as covariates. Psychometrika 76:40–56

Buzas JS, Stefanski LA (1996) Instrumental variable estimation in generalized linear measurement error models. J Am Stat Assoc 91:999–1006

Cancho VG, Lachos VH, Ortega EMM (2010) A nonlinear regression model with skew-normal errors. Stat Pap 51:547–558

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2006) Measurement error in nonlinear models: a modern perspective, 2nd edn. Chapman and Hall/CRC, Boca Raton

Carlin BP, Polson NG (1991) An expected utility approach to influence diagnostics. J Am Stat Assoc 86:1013–1021
Cook RD, Weisberg S (1982) Residuals and influence in regression. Chapman and Hall, New York
Chow SM, Tang NS, Yuan Y, Song XY, Zhu HT (2011) Bayesian estimation of semiparametric nonlinear dynamic factor analysis models using the Dirichlet process prior. Br J Math Stat Psychol 64:69–106
Dunson DB (2006) Bayesian dynamic modeling of latent trait distributions. Biostatistics 7:551–568
Ferguson TS (1973) A Bayesian analysis of some nonparametric problems. Ann Stat 1:209–230
Fong Y, Rue H, Wakefield J (2010) Bayesian influence for generalized linear mixed models. Biostatistics 11:397–412
Gelman A, Meng XL, Stern H (1996) Posterior predictive assessment of model fitness via realized discrepancies. Stat Sin 6:733–807
Geman S, Geman D (1984) Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell 6:721–741
Guha S (2008) Posterior simulation in the generalized linear mixed model with semiparametric random effects. J Comput Graph Stat 17:410–425
Gustafson P (2004) Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments. Chapman and Hall/CRC, Boca Raton
Jackson D, White IR, Carpenter J (2012) Identifying influential observations in Bayesian models by using Markov chain Monte Carlo. Stat Med 31:1238–1248
Kleinman KP, Ibrahim JG (1998) A semi-parametric Bayesian approach to generalized linear mixed models. Stat Med 17:2579–2596
Lachos VH, Angolini T, Abanto-Valle CA (2011) On estimation and local influence analysis for measurement errors models under heavy-tailed distributions. Stat Pap 52:567–590
Lachos VH, Cancho VG, Aoki R (2010) Bayesian analysis of skew-*t* multivariate null intercept measurement error model. Stat Pap 51:531–545
Lee SY, Lu B, Song XY (2008) Semiparametric Bayesian analysis of structural equation models with fixed covariates. Stat Med 15:2341–2360
Lee SY, Tang NS (2006) Bayesian analysis of nonlinear structural equation models with nonignorable missing data. Psychometrika 71:541–564
Liu JS (2001) Monte Carlo strategies in scientific computing. Springer, New York
Muff S, Riebler A, Held L, Rue H, Saner P (2015) Bayesian analysis of measurement error models using integrated nested laplace approximations. J R Stat Soc Ser C 64:231–252
Raftery AE (1996) Approximate Bayes factors and accounting for model uncertainty in generalised linear models. Biometrika 83:2510–266
Richarson S, Green DJ (1997) On Bayesian analysis of mixture with unknown numbers of components (with discussion). J R Stat Soc Ser B 59:731–792
Singh S, Jain K, Sharma S (2014) Replicated measurement error model under exact linear restrictions. Stat Pap 55:253–274
Stefanski LA, Carroll RJ (1985) Covariate measurement error in logistic regression. Ann Stat 13:1335–1351
Stefanski LA, Carroll RJ (1987) Conditional scores and optimal scores for generalized linear measurement-error models. Biometrika 74:703–716
Tang NS, Duan XD (2012) A semiparametric Bayesian approach to generalized partial linear mixed models for longitudinal data. Comput Stat Data Anal 77:4348–4365
Tang NS, Tang AM, Pan DD (2014) Semiparametric Bayesian joint models of multivariate longitudinal and survival data. Comput Stat Data Anal 77:113–129
Yang M, Dunson DB, Baird D (2010) Semiparametric Bayes hierarchical models with mean and variance constraints. Comput Stat Data Anal 54:2172–2186