Semiparametric Bayesian inference on generalized linear measurement error models

Nian-Sheng Tang1 · De-Wang Li1 · An-Min Tang1

Received: 11 October 2014 / Revised: 9 December 2015 / Published online: 1 February 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

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-
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 covariate 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, \( x_i \) be a \( r \times 1 \) vector of the unobserved covariate variables, and \( v_i \) be a \( p \times 1 \) vector of the observed covariate variables for the \( i \)th 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 \( z_1, \ldots, z_n \), where \( z_i = (x_i^\top, v_i^\top)^\top \) for \( i = 1, \ldots, n \). Given \( z_i \), 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 \( \theta_i \) and a mean that is a function of \( z_i \). That is, for \( i = 1, \ldots, n \), the conditional probability density function of \( y_i \) given \( z_i \) is given by

\[
p(y_i|z_i, \phi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi) \right\}
\]

with \( \mu_i = \text{E}(y_i|z_i) = \hat{b}(\theta_i) \) and \( V_i = \text{var}(y_i|z_i) = \phi \tilde{b}(\theta_i) \), where \( \phi \) is a scale parameter, \( b(\cdot) \) and \( c(\cdot, \cdot) \) are specific differentiable functions, \( \hat{b}(\theta_i) = \partial b(\theta_i)/\partial \theta_i \) and \( \tilde{b}(\theta_i) = \partial^2 b(\theta_i)/\partial \theta_i^2 \). The conditional mean \( \mu_i \) is assumed to satisfy

\[
\eta_i = h(\mu_i) = x_i^\top \beta_x + v_i^\top \beta_v = z_i^\top \beta,
\]

where \( h(\cdot) \) is a monotonic differentiable link function, \( \beta = (\beta_x^\top, \beta_v^\top)^\top \) 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 $x_i$ is measured $m$ times for individual $i$, giving outcomes $w_{ij}$ for $j = 1, \ldots, m$, the structural ME model can be expressed as

$$w_{ij} = x_i + u_{ij}, \quad (2.3)$$

where the MEs $u_{ij}$'s are assumed to follow an unknown distribution, and are independent of $x_i$.

Following Lee et al. (2008), we use the DP mixture model to specify the distribution of $u_{ij}$. That is, $u_{ij} \sim \sum_{g=1}^{\infty} \pi_g N_r(\alpha_g, \Omega_g)$ with $(\alpha_g, \Omega_g) \sim \mathcal{P}$ and $\mathcal{P} \sim \text{DP}(\tau P_0)$, where $\pi_g$ is a random probability weight between 0 and 1 such that $0 \leq \pi_g \leq 1$ and $\sum_{g=1}^{\infty} \pi_g = 1$, $N_r(\alpha_g, \Omega_g)$ denotes the multivariate normal distribution with mean $\alpha_g$ and covariance matrix $\Omega_g$, $\mathcal{P}$ 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 $\tau$ is a weight assigning a priori to the base distribution and represents the certainty of $P_0$ as the distribution of $\mathcal{P}$. 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 $u_{ij}$ is zero (Carroll et al. 2006).

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

$$u_{ij} \sim \sum_{g=1}^{G} \pi_g N_r(\alpha_g, \Omega_g) \quad \text{with} \quad \alpha_g = \alpha_g^* - \sum_{l=1}^{G} \pi_l \alpha_l^* \quad \text{and} \quad (\alpha_g^*, \Omega_g) \sim P_0, \quad (2.4)$$

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

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

with $\nu_g \sim \text{Beta}(1, \tau)$ 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 $u_{ij}$ via the MCMC algorithm because of the complicated posterior distribution involved. To this end, we generate $u_{ij}$ from $N_r(\alpha_{L_{ij}}, \Omega_{L_{ij}})$ in terms of a latent variable $L_{ij} \in \{1, \ldots, G\}$, where $\alpha_{L_{ij}} = \alpha_{L_{ij}}^* - \sum_{g=1}^{G} \pi_g \alpha_g^*$ in which $\alpha_{L_{ij}}^*$ is sampled from the following reformulated model. Let $\pi = \{\pi_1, \ldots, \pi_G\}$, $\alpha^* = \{\alpha_1^*, \ldots, \alpha_G^*\}$ and $\Omega = \{\Omega_1, \ldots, \Omega_G\}$ in which $\Omega_g = \text{diag}(\omega_{g1}, \ldots, \omega_{gr})$ for $g = 1, \ldots, G$. It follows from Lee et al. (2008) that Equation (2.4) can be rewritten as

$$L_{ij} | \pi \sim \sum_{g=1}^{G} \pi_g \delta_g (\cdot) \quad \text{and} \quad (\pi, \alpha^*, \Omega) \sim f_1(\pi) f_2(\alpha^*) f_3(\Omega), \quad (2.6)$$
where $\delta_g(\cdot)$ is a discrete probability measure concentrated at $g$, $f_1(\pi)$ is specified by the stick-breaking prior as given in Eq. (2.5), the prior distribution of $\alpha_g^*$ involved in $f_2(\alpha^*) = \prod_{g=1}^{G} p(\alpha_g^*)$ is given by

$$
\alpha_g^* | \xi, \Psi \overset{\text{i.i.d.}}{\sim} \Gamma_r(\xi, \Psi), \quad \xi | \psi^0, \omega \sim \Gamma_r(\psi^0, \omega), \quad \psi_j^{-1} | c_1, c_2 \sim \Gamma(c_1, c_2) \quad \text{for } j = 1, \ldots, r,
$$

(2.7)

where $\Psi = \text{diag}(\psi_1, \ldots, \psi_r)$, $\psi^0, \omega, c_1$ 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 $\omega_{gj}$ involved in $f_3(\Omega) = \prod_{g=1}^{G} \prod_{j=1}^{r} p(\omega_{gj})$ is given by

$$
\omega_{gj}^{-1} | \phi_{gj}^0, \phi_j^b \sim \Gamma(\phi_{gj}^0, \phi_j^b) \quad \text{and} \quad \phi_{gj}^b | \phi_j^b, \phi_j^d \sim \Gamma(\phi_j^b, \phi_j^d),
$$

(2.8)

where $\phi_{gj}^0, \phi_j^b$ and $\phi_j^d$ 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

$$
x_{ki} = \gamma_{k0} + \gamma_{k1} v_1 + \epsilon_{ki}, \quad \epsilon_{ki} \overset{\text{i.i.d.}}{\sim} \text{N}(0, \sigma_x^2),
$$

(2.9)

where $\gamma_{k0}$ is an intercept, $\gamma_{k1} = (\gamma_{k1}, \ldots, \gamma_{kp})^T$ is a $p \times 1$ vector of unknown regression parameters, and $\epsilon_{ki}$'s are residuals and assumed to be independent of the covariates $v_i$ and MEs $u_{ij}$'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 $\sigma_x^2$ and $\gamma_{k0}$ to be some pre-specified values.

Let $y = \{y_1, \ldots, y_n\}$, $x = \{x_1, \ldots, x_n\}$, $v = \{v_1, \ldots, v_n\}$, $u = \{u_1, \ldots, u_n\}$ and $w = \{w_1, \ldots, w_n\}$ in which $x_i = (x_{i1}, \ldots, x_{ir})^T$, $u_i = \{u_{i1}, \ldots, u_{imi}\}$ and $w_i = \{w_{i1}, \ldots, w_{imi}\}$ for $i = 1, \ldots, n$. Denote $\theta_y = \{\beta, \phi\}$, $\theta_u = \{\tau, \pi, \alpha^*, \Omega\}$, $\theta_y = \{\gamma_{10}, \ldots, \gamma_{r0}, \gamma_{1y}, \ldots, \gamma_{ry}, \sigma_x^2\}$ and $\theta = \{\theta_y, \theta_u, \theta_y\}$. Under the above assumptions, the joint probability density function for $y, w, u, x$ is given by

$$
P(y, w, u, x | v, \theta) = \prod_{i=1}^{n} \left\{ p(y_i | x_i, v_i; \theta_y) p(w_i | x_i; \theta_u) p(x_i | v_i; \theta_y) \right\}.
$$

(2.10)

To make Bayesian inference on parameters in $\{\tau, \theta_y, \theta_y\}$, it is necessary to specify their corresponding priors. Similar to Lee and Tang (2006), we consider the following priors for parameters $\tau, \beta, \phi, \gamma_k = (\gamma_{k0}, \gamma_{k1}^T)$ for $k = 1, \ldots, r$, and $\sigma_x^2$:

$$
\begin{align*}
\tau | a_1, a_2 & \sim \Gamma(a_1, a_2), \quad \beta | \phi, \beta^0 \sim \text{N}_{r+p}(\beta^0, \phi^{-1} \Omega_{\beta}), \\
\phi^{-1} | a_3, a_4 & \sim \Gamma(a_3, a_4), \quad \gamma_k^0 | \gamma_k^0 \sim \text{N}_{p+1}(\gamma_k^0, \Omega_{\gamma_k}), \quad \sigma_x^{-2} | c_3, c_4 \sim \Gamma(c_3, c_4),
\end{align*}
$$
where \(a_1, a_2, a_3, a_4, b_0, H_0^{\beta}, y_k^0, H_0^{\gamma^k}, c_3\) and \(c_4\) are hyperparameters whose values are assumed to be given by the prior information. Thus, we specify standard conjugate priors for parameters \(\tau, \phi^{-1}, \gamma^k\) and \(\sigma_x^{-2}\). The associated hyperparameters can be determined in a relatively straightforward manner based on previous applications or data-dependent inputs (Raftery 1996; Richardson 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 \(\tau\) which controls the behavior of \(u_{ij}\). 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 \(\{\tau, \theta_y, \theta_{y'}\}\) 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 \(\{\tau, \theta_y, \theta_{y'}\}\) 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 \(p(\xi, \Psi, \tau, \phi^b, \pi, L, \alpha^*, \Omega, u, \beta, \phi, \gamma, \sigma_x^2, x|y, w, v)\), and Bayesian estimates of unknown parameters and covariates subject to MEs are obtained from the generated sequence of random observations, where \(\phi^b = \{\phi_1^b, \ldots, \phi_r^b\}\), \(\gamma = \{\gamma_1^*, \ldots, \gamma_r^*\}\) and \(L = \{L_{ij}: i = 1, \ldots, n, j = 1, \ldots, m\}\). In this algorithm, observations \(\{\xi, \Psi, \tau, \phi^b, \pi, L, \alpha^*, \Omega, u, \beta, \phi, \gamma, \sigma_x^2, x\}\) are iteratively drawn from the following conditional distributions: \(p(\xi|\alpha^*, \Psi)\), \(p(\Psi|\alpha^*, \xi)\), \(p(\tau|\pi)\), \(p(\phi^b|\Omega)\), \(p(\pi|L, \tau)\), \(p(L|\pi, \alpha, \Omega, u)\), \(p(\alpha^*|\xi, \Psi, \Omega, L, u)\), \(p(\beta|y, x, v, \phi)\), \(p(\Omega|\alpha, \phi^b, L, u)\), \(p(\phi|\beta, y, x, v)\), \(p(\gamma|x, v, \sigma_x^2)\), \(p(\sigma_x^2|x, v, \gamma)\), \(p(x|y, v, u, w, \alpha, \beta, \phi, \sigma_x^2, \gamma, L)\) and \(p(u|\alpha, \Omega, L, x, w, \theta_u)\). These conditional distributions are presented in the Appendix.

#### 3.1 Bayesian estimates

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

\[
\hat{\beta} = \frac{1}{L} \sum_{\ell=1}^{L} \beta^{(\ell)}, \quad \hat{\phi} = \frac{1}{L} \sum_{\ell=1}^{L} \phi^{(\ell)}, \quad \hat{y} = \frac{1}{L} \sum_{\ell=1}^{L} y^{(\ell)}, \quad \hat{x} = \frac{1}{L} \sum_{\ell=1}^{L} x^{(\ell)}, \quad \hat{\sigma}_x^2 = \frac{1}{L} \sum_{\ell=1}^{L} \sigma_x^{2(\ell)}.
\]

(3.1)

The consistent estimates of the posterior covariance matrices \(\text{var}(\beta|y, v, w)\), \(\text{var}(\phi|y, v, w)\), \(\text{var}(y|y, v, w)\) and \(\text{var}(\sigma_x^2|y, v, w)\) can be obtained from the sample covari-
ance matrices of the above generated observations. For example, $\hat{\text{var}}(\beta | y, v, w) = \frac{1}{\Sigma^{-1}} \sum_{i=1}^{n} (\beta_i^{(e)} - \hat{\beta}) (\beta_i^{(e)} - \hat{\beta})^T$. The standard errors for components of $\beta$ 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 $R = \{y, v, w\}$ be a full data set, and $R(i) = \{(y_j, v_j, w_j) : j = 1, \ldots, n, j \neq i\}$ be a subset of $R$ with the $i$th individual deleted. To assess the effect of the $i$th individual on the posterior distribution of parameter vector $\theta = \{\tau, \theta_y, \theta_w\}$, we use the following Kullback–Leibler (K–L) divergence

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

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

To measure the effect of the $i$th individual on the posterior mean of $\theta$, we use the following Cook’s distance

$$CD(i) = (\hat{\theta} - \hat{\theta} (i))^T W_\theta^{-1} (\hat{\theta} - \hat{\theta} (i))$$

as another Bayesian case influence measure, where $\hat{\theta} = \int \theta p(\theta | R)d\theta$ and $\hat{\theta} (i) = \int \theta p(\theta | R(i))d\theta$ are the posterior means of $\theta$ for the full data set $R$ and the deleted data set $R(i)$, respectively, and $W_\theta$ is selected to be the posterior covariance matrix of $\theta$. 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 $p(\theta | R)$ and $p(\theta | R(i))$. 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

$$p(\theta | R(i)) = \frac{\{p_i(\theta)\}^{-1} p(\theta | R)}{\int \{p_i(\theta)\}^{-1} p(\theta | R)d\theta},$$

where $p_i(\theta) = p(y_i, v_i, w_i | \theta)$. Substituting Eq. (3.2) into the definition of $KL(i)$ yields
\[ KL(i) = \log E_{\theta | R} \{ p_i(\hat{\theta}) \}^{-1} + E_{\theta | R} \{ \log p_i(\hat{\theta}) \}, \]

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

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

where \( p_i(\theta^{(\ell)}) = p(y_i, v_i, w_i | \theta^{(\ell)}) \).

On the other hand, to compute \( CD(i) \), we need evaluating \( \hat{\theta} \), \( \hat{\theta}^{(i)} \) and \( W_\theta \). It follows from Eq. (3.2) and the definitions of \( \hat{\theta} \), \( \hat{\theta}^{(i)} \) and \( W_\theta \) that \( \hat{\theta} = E_{\theta | R}(\theta) \), 
\[
\hat{\theta}^{(i)} = E_{\theta | R}(\theta | p_i(\theta))^{-1}|/E_{\theta | R}(p_i(\theta))^{-1} \text{ and } W_\theta = \var\theta | R(\theta). \]

Similarly, the MCMC approximations of \( \hat{\theta} \), \( \hat{\theta}^{(i)} \) and \( W_\theta \) are given by

\[
\hat{\theta} = \frac{1}{L} \sum_{\ell=1}^{L} \theta^{(\ell)}, \quad \hat{\theta}^{(i)} = \frac{1}{L} \sum_{\ell=1}^{L} \theta^{(\ell)}(p_i(\theta^{(\ell)})|^{-1}, \quad W_\theta = \frac{1}{L - 1} \sum_{\ell=1}^{L} (\hat{\theta}^{(i)} - \hat{\theta})(\theta^{(i)} - \hat{\theta})^T, \tag{3.4}\]

respectively.

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

\[
p_i(\theta) = \int p(y_i, v_i, w_i, u_i, x_i | \theta) du_i dx_i = E_{u_i, x_i} \{ p(y_i, v_i, w_i | \theta, x_i, u_i) \}, \tag{3.5}\]

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

- **Step 0** Specifying the initial value \( x_i^{(0)} \) of \( x_i \). Generally, we can take \( x_i^{(0)} \) to be the mean of \( w_{ij} \)’s.
- **Step 1** Sampling Gibbs sample \( u_i^{(t)} \) from the conditional distribution \( p(u_i | x_i^{(t-1)}, \theta_u^{(t)}) \).
- **Step 2** Drawing Gibbs sample \( x_i^{(t)} \) from the conditional distribution \( p(x_i | u_i^{(t)}, \theta_u^{(t)}) \).
- **Step 3** Repeating Steps 1 and 2 for \( T \) times.
- **Step 4** Getting the nested Gibbs samples \( \{(u_i^{(t)}, x_i^{(t)}) : t = 1, \ldots, T\} \) from \( p(u_i, x_i | \theta_u) \). Thus, the MCMC approximation of \( p_i(\theta^{(\ell)}) \) is given by

\[
p_i(\theta^{(\ell)}) \approx \frac{1}{T} \sum_{t=1}^{T} p(y_i, v_i, w_i | \theta^{(t)}, x_i^{(t)}, u_i^{(t)}). \tag{3.6}\]
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 $u_{ij}$ 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 $\{(y_i, v_i, w_i, x_i) : i = 1, \ldots, n\}$ 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 $\eta_i = \log(\mu_i) = x_i^T \beta_x + v_i^T \beta_v$, where $v_i$’s were generated from a multivariate normal distribution with mean zero and covariance matrix $0.25 I_3$ and components $x_{1i}$ and $x_{2i}$ in $x_i$ were generated via Eq. (2.9). In this case, $\phi = 1$ relating to Eq. (2.1) is a constant. The true values of $\beta_x = (\beta_0, \beta_1)^T$, $\beta_v = (\beta_2, \beta_3, \beta_4)^T$, $\gamma_k = (\gamma_{k0}, \gamma_{k1}, \gamma_{k2}, \gamma_{k3})^T$ and $\sigma_k^2$ were taken to be $\beta_x = (0.4, -0.3)^T$, $\beta_v = (0.4, 0.3, 0.4)^T$, $\gamma_k = (0.2, 0.2, 0.2, 0.5)^T$ for $k = 1$ and 2, and $\sigma_k^2 = 1$, respectively. To test the effectiveness of using the TCDP prior to approximate distributions of MEs $u_{ij} = (u_{ij1}, u_{ij2})^T$, we considered the following eight distributional assumptions for $u_{ijk}$.

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

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

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

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

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

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

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 \( \{u_{ijk}: i = 1, \ldots, n, j = 1, \ldots, m, k = 1, 2\} \) via Eq. (2.3).

The hyperparameter values of the prior distributions for \( \tau \) and \( \sigma_x^2 \) 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 \( \varphi_d^j = 3 \), \( \varphi_c^j = 150 \) and \( \varphi_d^j = 20 \) 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 \( \tau \), we set \( a_1 = 200 \) and \( a_2 = 10 \) to generate large values of \( \tau \) which lead to more unique covariate MEs. For the conjugate prior of \( \sigma_x^{-2} \), we set \( c_3 = 10 \) and randomly generated \( c_4 \) from a uniform distribution \( U(9, 10) \). For the hyperparameters \( \xi^0 \) and \( \Psi^0 \), we took \( \xi^0 = 0 \) and \( \Psi^0 = I_r \) 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 \( \beta \) and \( \gamma_k \).

- **Type A** The hyperparameters corresponding to the priors of \( \beta \) and \( \gamma_k \) were taken to be \( \beta^0 = (0.4, -0.3, 0.4, 0.3, 0.4)^\top \), \( H_{\beta}^0 = 0.25I_5 \), \( \gamma_k^0 = (0.2, 0.2, 0.2, 0.5)^\top \) and \( H_{ky}^0 = 0.25I_4 \). This can be regarded as a situation with good prior information.

- **Type B** The hyperparameters corresponding to the priors of \( \beta \) and \( \gamma_k \) were taken to be \( \beta^0 = 1.5 \times (0.4, -0.3, 0.4, 0.3, 0.4)^\top \), \( H_{\beta}^0 = 0.75I_5 \), \( \gamma_k^0 = 1.5 \times (0.2, 0.2, 0.2, 0.5)^\top \) and \( H_{ky}^0 = 0.75I_4 \). This can be regarded as a situation with inaccurate prior information.

- **Type C** The hyperparameters corresponding to the priors of \( \beta \) and \( \gamma_k \) were taken to be \( \beta^0 = (0, 0, 0, 0, 0)^\top \), \( H_{\beta}^0 = 10I_5 \), \( \gamma_k^0 = (0, 0, 0, 0)^\top \) and \( H_{ky}^0 = 10I_4 \). 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, \( \mathcal{L} = 5000 \) 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 \( \beta \) 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 \( \beta \) and \( \gamma_k \).
| Par. Type | Type A | Type B | Type C | Type A | Type B | Type C |
|----------|--------|--------|--------|--------|--------|--------|
|          | Bias   | RMS    | Bias   | Bias   | RMS    | Bias   |
| $\gamma_{10}$ | 0.000 | 0.079 | 0.001 | 0.085 | 0.003 | 0.087 |
| $\gamma_{11}$ | 0.000 | 0.144 | 0.004 | 0.126 | 0.017 | 0.167 |
| $\gamma_{12}$ | 0.013 | 0.154 | 0.043 | 0.160 | 0.021 | 0.162 |
| $\gamma_{13}$ | 0.025 | 0.145 | 0.002 | 0.159 | 0.001 | 0.164 |
| $\gamma_{20}$ | 0.003 | 0.089 | 0.011 | 0.085 | 0.003 | 0.081 |
| $\gamma_{21}$ | 0.011 | 0.142 | 0.010 | 0.165 | 0.018 | 0.164 |
| $\gamma_{22}$ | 0.015 | 0.154 | 0.021 | 0.175 | 0.008 | 0.154 |
| $\gamma_{23}$ | 0.001 | 0.135 | 0.015 | 0.156 | 0.010 | 0.170 |
| $\beta_0$ | 0.007 | 0.064 | 0.006 | 0.068 | 0.008 | 0.066 |
| $\beta_1$ | 0.012 | 0.080 | 0.011 | 0.077 | 0.005 | 0.081 |
| $\beta_2$ | 0.017 | 0.107 | 0.087 | 0.154 | 0.007 | 0.147 |
| $\beta_3$ | 0.003 | 0.118 | 0.023 | 0.134 | 0.007 | 0.142 |
| $\beta_4$ | 0.022 | 0.122 | 0.021 | 0.127 | 0.002 | 0.137 |
| $\sigma_x^2$ | 0.011 | 0.077 | 0.027 | 0.092 | 0.015 | 0.090 |
| $\gamma_{10}$ | 0.000 | 0.070 | 0.003 | 0.076 | 0.001 | 0.071 |
| $\gamma_{11}$ | 0.009 | 0.133 | 0.032 | 0.150 | 0.017 | 0.143 |
| $\gamma_{12}$ | 0.008 | 0.126 | 0.006 | 0.133 | 0.012 | 0.157 |
| $\gamma_{13}$ | 0.012 | 0.147 | 0.013 | 0.142 | 0.009 | 0.144 |
| $\gamma_{20}$ | 0.004 | 0.075 | 0.000 | 0.078 | 0.009 | 0.080 |
| $\gamma_{21}$ | 0.015 | 0.134 | 0.004 | 0.148 | 0.008 | 0.156 |
| $\gamma_{22}$ | 0.001 | 0.124 | 0.019 | 0.136 | 0.001 | 0.142 |
| $\gamma_{23}$ | 0.014 | 0.135 | 0.024 | 0.146 | 0.017 | 0.141 |
| $\beta_0$ | 0.005 | 0.056 | 0.001 | 0.054 | 0.001 | 0.055 |
| $\beta_1$ | 0.015 | 0.061 | 0.004 | 0.074 | 0.004 | 0.058 |
| $\beta_2$ | 0.017 | 0.114 | 0.085 | 0.154 | 0.008 | 0.128 |
| $\beta_3$ | 0.009 | 0.119 | 0.007 | 0.117 | 0.017 | 0.133 |
| $\beta_4$ | 0.009 | 0.129 | 0.010 | 0.140 | 0.010 | 0.141 |
| $\sigma_x^2$ | 0.004 | 0.068 | 0.008 | 0.075 | 0.013 | 0.067 |
| $\gamma_{10}$ | 0.011 | 0.071 | 0.006 | 0.081 | 0.005 | 0.074 |
| $\gamma_{11}$ | 0.002 | 0.137 | 0.014 | 0.135 | 0.002 | 0.144 |
| $\gamma_{12}$ | 0.014 | 0.126 | 0.005 | 0.157 | 0.013 | 0.130 |
| $\gamma_{13}$ | 0.004 | 0.136 | 0.003 | 0.131 | 0.007 | 0.138 |
| $\gamma_{20}$ | 0.001 | 0.069 | 0.002 | 0.080 | 0.008 | 0.073 |
| $\gamma_{21}$ | 0.018 | 0.138 | 0.008 | 0.158 | 0.010 | 0.135 |
| $\gamma_{22}$ | 0.013 | 0.148 | 0.009 | 0.125 | 0.017 | 0.134 |
| $\gamma_{23}$ | 0.009 | 0.148 | 0.015 | 0.148 | 0.003 | 0.149 |
Table 1 continued

| Par. | Type A | Type B | Type C | Type A | Type B | Type C |
|------|--------|--------|--------|--------|--------|--------|
|      | Bias   | RMS    | Bias   | RMS    | Bias   | RMS    |
| $\beta_0$ | 0.006  | 0.063  | 0.000  | 0.065  | 0.008  | 0.065  |
| $\beta_1$ | 0.000  | 0.064  | 0.000  | 0.067  | 0.004  | 0.064  |
| $\beta_2$ | 0.010  | 0.123  | 0.113  | 0.173  | 0.004  | 0.120  |
| $\beta_3$ | 0.006  | 0.107  | 0.013  | 0.137  | 0.001  | 0.135  |
| $\beta_4$ | 0.003  | 0.115  | 0.017  | 0.148  | 0.018  | 0.157  |
| $\gamma_{10}$ | 0.019  | 0.067  | 0.015  | 0.074  | 0.007  | 0.077  |
| $\gamma_{11}$ | 0.008  | 0.134  | 0.021  | 0.120  | 0.007  | 0.113  |
| $\gamma_{12}$ | 0.006  | 0.112  | 0.005  | 0.114  | 0.002  | 0.113  |
| $\gamma_{13}$ | 0.008  | 0.119  | 0.019  | 0.126  | 0.030  | 0.140  |
| $\gamma_{20}$ | 0.001  | 0.081  | 0.002  | 0.072  | 0.012  | 0.067  |
| $\gamma_{21}$ | 0.000  | 0.134  | 0.014  | 0.122  | 0.001  | 0.133  |
| $\gamma_{22}$ | 0.000  | 0.107  | 0.027  | 0.129  | 0.007  | 0.127  |
| $\gamma_{23}$ | 0.002  | 0.144  | 0.034  | 0.159  | 0.011  | 0.135  |
| $\sigma^2_x$ | 0.000  | 0.056  | 0.002  | 0.049  | 0.012  | 0.060  |

Assumption 7

| Par. | Type A | Type B | Type C | Type A | Type B | Type C |
|------|--------|--------|--------|--------|--------|--------|
|      | Bias   | RMS    | Bias   | RMS    | Bias   | RMS    |
| $\gamma_{10}$ | 0.000  | 0.062  | 0.009  | 0.082  | 0.015  | 0.067  |
| $\gamma_{11}$ | 0.008  | 0.134  | 0.021  | 0.120  | 0.007  | 0.142  |
| $\gamma_{12}$ | 0.006  | 0.112  | 0.005  | 0.114  | 0.002  | 0.113  |
| $\gamma_{13}$ | 0.008  | 0.119  | 0.019  | 0.126  | 0.030  | 0.140  |
| $\gamma_{20}$ | 0.001  | 0.081  | 0.002  | 0.072  | 0.012  | 0.067  |
| $\gamma_{21}$ | 0.000  | 0.134  | 0.014  | 0.122  | 0.001  | 0.133  |
| $\gamma_{22}$ | 0.000  | 0.107  | 0.027  | 0.129  | 0.007  | 0.127  |
| $\gamma_{23}$ | 0.002  | 0.144  | 0.034  | 0.159  | 0.011  | 0.135  |
| $\beta_0$ | 0.013  | 0.067  | 0.008  | 0.066  | 0.010  | 0.062  |
| $\beta_1$ | 0.011  | 0.128  | 0.122  | 0.158  | 0.022  | 0.118  |
| $\beta_2$ | 0.011  | 0.109  | 0.024  | 0.114  | 0.016  | 0.115  |
| $\beta_3$ | 0.000  | 0.112  | 0.003  | 0.135  | 0.015  | 0.124  |
| $\sigma^2_x$ | 0.002  | 0.075  | 0.008  | 0.069  | 0.008  | 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 | 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 |        |        |        |        |        |        |        |        |        |        |        |        |
| $\beta_0$ | 0.009  | 0.108  | 0.016  | 0.118  | 0.007  | 0.115  | 0.022  | 0.222  | 0.044  | 0.224  | 0.099  | 0.220  |
| $\beta_1$ | 0.014  | 0.128  | 0.010  | 0.135  | 0.016  | 0.136  | 0.009  | 0.232  | 0.075  | 0.240  | 0.018  | 0.245  |
| $\beta_2$ | 0.005  | 0.113  | 0.005  | 0.118  | 0.005  | 0.120  | 0.018  | 0.133  | 0.013  | 0.138  | 0.021  | 0.144  |
| $\beta_3$ | 0.012  | 0.112  | 0.017  | 0.119  | 0.014  | 0.121  | 0.008  | 0.125  | 0.011  | 0.130  | 0.009  | 0.135  |
| $\beta_4$ | 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 2 |        |        |        |        |        |        |        |        |        |        |        |        |
| $\beta_0$ | 0.013  | 0.200  | 0.066  | 0.283  | 0.015  | 0.348  | 0.018  | 0.156  | 0.005  | 0.224  | 0.052  | 0.285  |
| $\beta_1$ | 0.003  | 0.251  | 0.048  | 0.342  | 0.001  | 0.416  | 0.017  | 0.178  | 0.014  | 0.244  | 0.047  | 0.308  |
| $\beta_2$ | 0.003  | 0.134  | 0.008  | 0.140  | 0.002  | 0.144  | 0.004  | 0.128  | 0.003  | 0.133  | 0.008  | 0.138  |
| $\beta_3$ | 0.011  | 0.117  | 0.014  | 0.121  | 0.009  | 0.124  | 0.001  | 0.142  | 0.006  | 0.147  | 0.001  | 0.152  |
| $\beta_4$ | 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 3 |        |        |        |        |        |        |        |        |        |        |        |        |
| $\beta_0$ | 0.029  | 0.132  | 0.025  | 0.139  | 0.032  | 0.144  | 0.019  | 0.225  | 0.001  | 0.283  | 0.041  | 0.327  |
| $\beta_1$ | 0.008  | 0.148  | 0.013  | 0.158  | 0.008  | 0.163  | 0.004  | 0.234  | 0.028  | 0.297  | 0.001  | 0.341  |
| $\beta_2$ | 0.007  | 0.120  | 0.010  | 0.126  | 0.006  | 0.129  | 0.004  | 0.112  | 0.007  | 0.118  | 0.003  | 0.120  |
| $\beta_3$ | 0.003  | 0.119  | 0.007  | 0.125  | 0.002  | 0.128  | 0.001  | 0.127  | 0.001  | 0.133  | 0.001  | 0.137  |
| $\beta_4$ | 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 4 |        |        |        |        |        |        |        |        |        |        |        |        |
| $\beta_0$ | 0.013  | 0.125  | 0.019  | 0.133  | 0.011  | 0.138  | 0.014  | 0.211  | 0.075  | 0.238  | 0.096  | 0.260  |
| $\beta_1$ | 0.047  | 0.165  | 0.058  | 0.182  | 0.051  | 0.188  | 0.023  | 0.209  | 0.128  | 0.242  | 0.069  | 0.242  |
| $\beta_2$ | 0.020  | 0.133  | 0.015  | 0.139  | 0.022  | 0.143  | 0.006  | 0.145  | 0.001  | 0.152  | 0.008  | 0.155  |
| $\beta_3$ | 0.006  | 0.114  | 0.010  | 0.120  | 0.006  | 0.122  | 0.001  | 0.116  | 0.003  | 0.122  | 0.002  | 0.126  |
| $\beta_4$ | 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 \(\{(y_i, v_i, w_i, x_i) : i = 1, \ldots, 200\}\) 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 $u_{ijlk}$ and $\hat{u}_{ijlk}$ in the simulation studies

| Assump. | Mean | SD | Type A | | | Type B | | | Type C | | |
|---------|------|----|--------|---|---|--------|---|---|--------|---|---|
|         | $u_{ij1}$ | $\hat{u}_{ij1}$ | $u_{ij2}$ | $\hat{u}_{ij2}$ | $u_{ij3}$ | $\hat{u}_{ij3}$ | $u_{ij4}$ | $\hat{u}_{ij4}$ | $u_{ij5}$ | $\hat{u}_{ij5}$ |
| 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 |

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.

The 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):

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

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

To make Bayesian analysis for the above considered LOGMEM, we assumed the following prior distributions for $\boldsymbol{y} = (\gamma_0, \gamma_v)^T$ and $\boldsymbol{\beta} = (\beta_0, \beta_x, \beta_v)^T$: $p(\boldsymbol{y}) \sim N_2(\boldsymbol{y}^0, 0.25 \mathbf{I}_2)$ and $p(\boldsymbol{\beta}) \sim N_3(\boldsymbol{\beta}^0, 0.25 \mathbf{I}_3)$, where the hyperparameters $\boldsymbol{y}^0$ and $\boldsymbol{\beta}^0$.
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   |
| $\beta_0$ | 4.409  | 0.018 | 4.394 | 0.018 | -2.877 | 0.478 |
| $\beta_x$ | -0.047  | 0.021 | -0.034 | 0.020 | 0.038  | 0.115 |
| $\beta_v$ | 8.032  | 0.460 | 3.039 | 0.453 | 0.342  | 0.211 |
| $\gamma_0$ | -1.305 | 0.324 | -0.174 | 0.115 | –  | –  |
| $\gamma_v$ | -0.289 | 0.201 | -0.151 | 0.193 | –  | –  |
| $\sigma^2_x$ | 0.118  | 0.004 | 0.112 | 0.004 | –  | –  |

$\beta^0$ were taken to be their corresponding Bayesian estimates obtained from the non-informative priors on $\gamma$ and $\beta$ (e.g., $p(\gamma) \sim N_2(0, 20I_2)$ and $p(\beta) \sim N_3(0, 20I_3)$). 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 $\tau$ leading to more unique covariate MEs. For the hyperparameters $c_3$ and $c_4$, we randomly generated $c_3$ from a uniform distribution $U(1, 100)$ to yield relatively diffuse values of $\sigma^2_x$. For the hyperparameters $\xi^0$ and $\Psi^0$, we took $\xi^0 = 0$ and $\Psi^0 = I_2$ to satisfy the condition of the centered DP procedure. Similar to simulation studies, we set $\varphi^a_j = 3$, $\varphi^c_j = 100$ and $\varphi^d_j = 20$ 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. 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:

$$
\text{logit}\left[\Pr(y_i = 1|x_i, v_i)\right] = \beta_0 + x_i \beta_x + v_i \beta_v \text{ for } i = 1, \ldots, n\text{ under the above specified prior of } \beta = (\beta_0, \beta_x, \beta_v)\top, \text{ where } x_i = (w_{i1} + w_{i2})/2. $$

The corresponding results were presented in Table 4, which showed that our considered logistic ME model led to a smaller estimate and a smaller SE of parameter $\beta_x$ 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 \( \beta_x \) and \( \beta_v \).

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.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

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 \( p(\xi, \Psi, \varphi^b, \pi, \alpha^*, \Omega, L, \tau, \beta, \phi, \gamma, \sigma_\xi^2, u, x|y, w, v) \). The Gibbs sampler is implemented by iteratively drawing observations from the following conditional distributions: \( p(\xi|\alpha^*, \Psi) \), \( p(\Psi|\alpha^*, \xi) \), \( p(\tau|\pi) \), \( p(\varphi^b|\Omega) \), \( p(\pi|L, \tau) \), \( p(L|\pi, \alpha, \Omega, u) \), \( p(\alpha^*|\xi, \Psi, \Omega, L, u) \), \( p(\beta|y, x, v, \phi) \), \( p(\Omega|\alpha, \varphi^b, L, u) \), \( p(\phi|\beta, y, x, v) \), \( p(\pi|x, v, \sigma_\xi^2) \), \( p(\sigma_\xi^2|x, v, \gamma) \), \( p(x|y, v, u, w, \alpha, \beta, \phi, \sigma_\xi^2, y, L) \) and \( p(u|\alpha, \Omega, L, x, w, \theta_u) \). 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 \( u_{ij} \) in terms of the latent variable \( L_{ij} \) for \( i = 1, \ldots, n \) and \( j = 1, \ldots, m \), we first generate \( \alpha^* = (\alpha_1^*, \ldots, \alpha_G^*) \) and \( \Omega = (\Omega_1, \ldots, \Omega_G) \) 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 \( u_{ij} \) from the multivariate normal distribution 
\( N_r(\alpha_{L_{ij}}, \Omega_{L_{ij}}) \) with 
\( \alpha_{L_{ij}} = \alpha_{L_{ij}}^* - \sum_{g=1}^{G} \pi_g \alpha_g^* \). Since it is rather difficult to directly 
sample observations from the posterior distribution of \((\xi, \Psi, \varphi^b, \pi, \alpha^*, \Omega, L, \tau, u)\), 
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 \( \xi \) is 
\( p(\xi | \alpha^*, \Psi) \sim N_r(\alpha_{\xi}, \Sigma_{\xi}) \), where 
\( \Sigma_{\xi} = (G\Psi^{-1} + \Psi^{-1})^{-1} \) and \( \alpha_{\xi} = \Sigma_{\xi} (\Psi^{-1} \xi + \Psi^{-1} \sum_{g=1}^{G} \alpha_g^*) \).

**Step (b)** For \( j = 1, \ldots, r \), the \( j \)th diagonal element of \( \Psi \) given \((\alpha^*, \xi)\) is distributed as 
\( p(\psi_j^{-1} | \alpha^*, \xi) \sim \Gamma(c_1 + G/2, c_2 + \frac{1}{2} \sum_{g=1}^{G} (\alpha_g^* - \xi_j)^2) \), 
where \( \alpha_g^* \) is the \( j \)th element of \( \alpha_g^* \) and \( \xi_j \) is the \( j \)th element of \( \xi \).

**Step (c)** Following the same argument of Chow et al. (2011), the conditional 
distribution \( p(\tau | \pi) \) is given by 
\( p(\tau | \pi) \sim \Gamma(a_1 + G - 1, a_2 - \sum_{g=1}^{G} \log(1 - v_g)) \), 
where \( v_g \) 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 \( \varphi_j^b \) is given by 
\( p(\varphi_j^b | \Omega) \sim \Gamma(\varphi_j^c, \varphi_j^d + \sum_{g=1}^{G} \omega_{gj}^{-1}) \), where \( \omega_{gj}^{-1} \) is the \( j \)th diagonal element of \( \Omega_g \).

**Step (e)** It can be shown that the conditional distribution \( p(\pi|L, \tau) \) follows a 
generalized Dirichlet distribution, i.e., 
\( p(\pi | L, \tau) \sim \varphi(a_1^*, b_1^*, \ldots, a_{G-1}^*, b_{G-1}^*) \), 
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 \( p(\pi | L, \tau) \) can be conducted by (1) sampling 
\( v_1^* \) from a Beta \((a_1^*, b_1^*)\) distribution, (2) sampling \( \pi_1, \ldots, \pi_G \) with the following 
expressions: 
\( \pi_1 = v_1^*, \quad \pi_g = v_g^* \prod_{j=1}^{g-1}(1 - v_j^*) \) for \( g = 2, \ldots, G-1 \) and 
\( \pi_G = 1 - \sum_{j=1}^{G-1} \pi_j \).

**Step (f)** We consider the conditional distribution \( p(\alpha^*_g | \xi, \Psi, \Omega, L, u) \). 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 \( g = 1, \ldots, G \), \( \alpha_g^* \) is drawn from the following 
conditional distribution: 
\( p(\alpha_g^* | \xi, \Psi) \sim N_r(\xi, \Sigma)^{-1} \), and \( p(\alpha_g^* | \Psi, \Omega, L, u) \sim N_r(A_g, B_g) \) with 
\( B_g = (\Psi^{-1} + \sum_{(i,j):L_{ij}=g} \Omega_{L_{ij}}^{-1})^{-1} \) and 
\( A_g = B_g (\Psi^{-1} \xi + \sum_{(i,j):L_{ij}=g} \Omega_{L_{ij}}^{-1} u_{ij}) \) for \( l \in \{L_1^*, \ldots, L_d^*\} \). Given 
\( \alpha^*, \alpha_g = \alpha_g^* - \sum_{j=1}^{G} \pi_j \alpha_j^* \) for \( g = 1, \ldots, G \), 
\( \alpha^* = \{\alpha_1^*, \ldots, \alpha_G^*\} \) and 
\( \alpha = \{\alpha_1, \ldots, \alpha_G\} \).

**Step (g)** The conditional distribution of \( \Omega \) is similar to the step (f). For \( j = 1, \ldots, r \), 
the \( j \)th diagonal element of \( \Omega_g \) \( (g = 1, \ldots, G) \) is generated from the following 
conditional distribution: 
\( p(\omega_{gj}^{-1} | \alpha, \varphi^b, L, u) \sim \Gamma(\varphi_{gj}^c, \varphi_{gj}^d) \) for \( g \neq \{L_1^*, \ldots, L_d^*\} \), 
and \( p(\omega_{gj}^{-1} | \alpha, \varphi^b, L, u) \sim \Gamma(\varphi_{gj}^c + d_{gj}/2, \varphi_{gj}^d + \frac{1}{2} \sum_{(i,\ell):L_{i\ell}=g} (u_{i\ell \ell} - \alpha_{gj})^2) \) for \( g \in \{L_1^*, \ldots, L_d^*\} \), where \( u_{i\ell \ell} \) is the \( j \)th element of \( u_{i\ell \ell} \), \( \alpha_{gj} \) is the \( j \)th element of \( \alpha_g \) 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 $p(L_{ij} | \pi, \alpha, \Omega, u) \sim \text{Multinomial}(\pi^*_{ij1}, \ldots, \pi^*_{ijG})$, where $\pi^*_{ijg}$ is proportional to $\pi_g p(u_{ij} | \alpha_g, \Omega_g)$ and $p(u_{ij} | \alpha_g, \Omega_g) \sim \text{N}(\alpha_g, \Omega_g)$ for $g = 1, \ldots, G$.

Step (i) Consider the conditional distribution of $u_{ij}$. It is easily shown that the conditional distribution $p(u_{ij} | \alpha, \Omega, L_{ij}, \theta_u, w, x_i) \propto p(u_{ij} | \alpha_{L_{ij}}, \Omega_{L_{ij}})$ $p(w_{ij} | u_{ij}, x_i, \theta_u)$ is a non-standard distribution. Thus, we cannot directly sample $u_{ij}$ from its conditional distribution. Here, the Metropolis–Hastings algorithm is employed to sample observation $u_{ij}$ from its conditional distribution via the following steps. At the $t$th iteration with a current value $u_{ij}^{(t)}$, a new candidate $u_{ij}$ is generated from the normal distribution $N_m(u_{ij}^{(t)}, \sigma^2_u \mathbb{D}_{u_{ij}})$, where $\mathbb{D}_{u_{ij}} = (\Omega_{L_{ij}}^{-1} + \mathbb{C}_{ij})^{-1}$ with $\mathbb{C}_{ij} = -\partial^2 \log p(w_{ij} | u_{ij}, x_i, \theta_u) / \partial u_{ij} \partial u_{ij}^T | u_{ij} = u_{ij}^{(t)}$. Then, the new candidate $u_{ij}$ is accepted with probability

$$\min \left\{ 1, \frac{p(u_{ij} | \alpha_{L_{ij}}, \Omega_{L_{ij}}) p(w_{ij} | u_{ij}, x_i, \theta_u)}{p(u_{ij}^{(t)} | \alpha_{L_{ij}}, \Omega_{L_{ij}}) p(w_{ij} | u_{ij}^{(t)}, x_i, \theta_u)} \right\}.$$ 

The variance $\sigma^2_u$ can be chosen such that the average acceptance rate is approximately 0.25 or more.

Step (j) The conditional distribution $p(\phi^{-1} | y, x, v, \beta)$ is proportional to

$$\phi^{-(a_3 + p + r - 1)} \exp \left\{ -\frac{1}{\phi} \left( a_4 + \frac{1}{2} (\beta - \beta_0)^T (H_\beta^{-1}) (\beta - \beta_0) - \sum_{i=1}^n (y_i \theta_i - b(\theta_i)) + \sum_{i=1}^n c(y_i, \phi) \right) \right\},$$

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

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

The variance $\sigma^2_{\phi}$ 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 $p(\phi^{-1} | y, x, v, \beta) \sim \Gamma(r + p + a_3, a_4 + 0.5(\beta - \beta_0)^T (H_\beta^{-1}) (\beta - \beta_0) - \sum_{i=1}^n (y_i \theta_i - b(\theta_i) + c(y_i)))$. Also, if $c(y_i, \phi) = \xi c(y_i)$, we obtain $p(\phi^{-1} | y, x, v, \beta) \sim \Gamma(r + p + a_3, a_4 + 0.5(\beta - \beta_0)^T (H_\beta^{-1}) (\beta - \beta_0) - \sum_{i=1}^n (y_i \theta_i - b(\theta_i)))$, where $\xi$ is a constant that does not depend on $\phi$ and $y_i$.

Step (k) The conditional distribution $p(\beta | y, x, v, \phi)$ can be expressed as

$$p(\beta | y, x, v, \phi) \propto \exp \left\{ \frac{1}{\phi} \sum_{i=1}^n (y_i \theta_i - b(\theta_i)) - \frac{1}{2\phi} (\beta - \beta_0)^T (H_\beta^{-1}) (\beta - \beta_0) \right\}.$$
Similarly, the Metropolis–Hastings algorithm for simulating observations from the conditional distribution \( p(\beta | y, x, v, \phi) \) is implemented as follows. Given the current value \( \beta^{(t)} \), a new candidate \( \beta \) is generated from \( N_{p+r}(\beta^{(t)}, \sigma_\beta^2 \Sigma_{\beta}) \) and is accepted with probability \( \min(1, p(\beta | y, x, v, \phi) / p(\beta^{(t)} | y, x, v, \phi)) \), where \( \Sigma_{\beta} = \phi^{-1} \left( \frac{1}{2} \sum_{i=1}^{n} V_i C_i^\top C_i + (H_\beta^0)^{-1} \right)^{-1} \) with \( C_i = (x_i, v_i^\top) \) and \( V_i = \hat{h}^{-2}(\mu_i) \hat{b}^{-1}(\theta_i) \) in which \( \hat{h}(a) = d h(a)/da \) and \( \hat{b}(a) = d^2 b(a)/da^2 \).

*Step (l)* The conditional distribution \( p(\gamma^*_y | x, v, \sigma_\gamma^2) \) is given by \( p(\gamma^*_y | x, v, \sigma_\gamma^2) \sim N(\mu^*_y, \Omega^*_y) \), where \( \Omega^*_y = (\sum_{i=1}^{n} v_i^* v_i^*/\sigma_\gamma^2 + (H_\gamma^0)^{-1})^{-1} \) and \( \mu^*_y = \Omega^*_y (\sum_{i=1}^{n} v_i^* x_i)/\sigma_\gamma^2 \) with \( v_i^* = (1, v_i^\top)^\top \).

*Step (m)* It is easily shown that the conditional distribution \( p(\sigma_\chi^2 | x, v, y) \) is given by \( p(\sigma_\chi^2 | x, v, y) \sim \Gamma(c_3 + \frac{n_r}{2}, c_4 + \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{r} (x_{ki} - y_{k0} - y_{k1} v_i)^2) \).

*Step (n)* The conditional distribution \( p(x_i | y_i, v_i, y, w_i, \alpha, \Omega, \beta, \sigma_\alpha^2, \phi, L_i) \) is proportional to

\[
\exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{\phi} - \sum_{k=1}^{r} \frac{(x_{ki} - y_{k0} - y_{k1} v_i)^2}{2\sigma_\chi^2} - \frac{1}{2} \sum_{j=1}^{m} (w_{ij} - x_i - \alpha_{L_{ij}})^\top \Omega_{L_{ij}}^{-1} (w_{ij} - x_i - \alpha_{L_{ij}}) \right\},
\]

where \( L_i = \{L_{ij} : j = 1, \ldots, m\} \).

The Metropolis–Hastings algorithm for sampling observations from \( p(x_i | y_i, v_i, y, w_i, \alpha, \Omega, \beta, \sigma_\alpha^2, \phi, L_i) \) is implemented as follows. Given the current value \( x_i^{(t)} \), a new candidate \( x_i \) is generated from \( N_r(x_i^{(t)}, \sigma_\alpha^2 H_x) \) and is accepted with probability

\[
\min \left\{ 1, \frac{p(x_i | y_i, v_i, y, w_i, \alpha, \Omega, \beta, \sigma_\alpha^2, \phi, L_i)}{p(x_i^{(t)} | y_i, v_i, y, w_i, \alpha, \Omega, \beta, \sigma_\alpha^2, \phi, L_i)} \right\},
\]

where \( H_x = (V_i \beta x \beta^\top + \sum_{j=1}^{m} \Omega_{L_{ij}}^{-1} + \sigma_{\chi}^{-2} I_r)^{-1} \) with \( V_i = \hat{h}^{-2}(\mu_i) \) \( \hat{b}^{-1}(\theta_i) |_{x_i = x_i^{(t)}} \). The variance \( \sigma_\alpha^2 \) can be chosen such that the average acceptance rate is approximately 0.25 or more.

References

Aitkin M, Rocc 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
Richardson 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