Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Semiparametric Bayesian joint models of multivariate longitudinal and survival data

Q1 Nian-Sheng Tang *, An-Min Tang, Dong-Dong Pan

Department of Statistics, Yunnan University, Kunming 650091, PR China

ARTICLE INFO

Article history:
Received 26 December 2012
Received in revised form 25 November 2013
Accepted 22 February 2014
Available online xxxx

Keywords:
Bayesian case-deletion diagnostic
Centered Dirichlet process prior
Joint models
Longitudinal data
Semiparametric Bayesian analysis
Survival data

ABSTRACT

Joint models for longitudinal and survival data are often used to investigate the association between longitudinal data and survival data in many studies. A common assumption for joint models is that random effects are distributed as a fully parametric distribution such as multivariate normal distribution. The fully parametric distribution assumption of random effects is relaxed by specifying a centered Dirichlet Process Mixture Model (CDPMM) for a general distribution of random effects because of some good properties of CDPMM such as inducing zero mean and continuous probability distribution of random effects. A computationally feasible Bayesian case-deletion diagnostic based on the ϕ-divergence is proposed to identify the potential influential cases in the joint models. Several simulation studies and a real example are used to illustrate our proposed methodologies.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Joint models for longitudinal and survival data (JMLS) are often employed to investigate the association between longitudinal and survival data in many studies such as HIV/AIDS clinical trials, and have received a lot of attention from both the frequentist and Bayesian perspectives. For example, see De Gruttola and Tu (1994), Tsiatis et al. (1995), Faucett and Thomas (1996), Wulfsohn and Tsiatis (1997), Henderson et al. (2000), Wang and Taylor (2001), Xu and Zeger (2001), Law et al. (2002), Song et al. (2002), Chen et al. (2002, 2004), Brown and Ibrahim (2003), Tsiatis and Davidian (2004), Brown et al. (2005), Chi and Ibrahim (2006, 2007), Ding and Wang (2008), Song and Wang (2008), Ye et al. (2008), Hu et al. (2009), Rizopoulos et al. (2009), Albert and Shih (2010), Zhu et al. (2012), among others. Q2

All the above mentioned studies have assumed that the random effects in JMLS are distributed as a fully parametric distribution such as a multivariate normal distribution. However, in some applications, the parametric distribution assumption of random effects in JMLS may be questioned (Brown and Ibrahim, 2003; Rizopoulos and Ghosh, 2011) even though parameter estimators in JMLSs are rather robust to random effect misspecification (Rizopoulos et al., 2008). Therefore, relaxing the parametric distribution assumption of random effects in JMLS have been received a lot of attention in past years (Rizopoulos et al., 2008). For example, Song et al. (2002) developed a semiparametric likelihood approach to JMLS via the EM algorithm based on the assumption that the random effects have a smooth density; Brown and Ibrahim (2003) presented a semiparametric Bayesian approach to JMLS by relaxing the distributional assumptions for the longitudinal model via Dirichlet process (DP) (Ferguson, 1973) priors on the parameters defining the longitudinal model. Recently, Rizopoulos and Ghosh (2011) proposed a new semiparametric multivariate JMLS with the single survival outcome by using a natural cubic spline-based method to flexibly capture the possibly nonlinear shapes of the subject-specific evolutions and

* Corresponding author.
E-mail address: nstang@ynu.edu.cn (N.-S. Tang).

http://dx.doi.org/10.1016/j.csda.2014.02.015
0167-9473/© 2014 Elsevier B.V. All rights reserved.

UNCORRECTED PROOF

relax the distribution of the random effects via a DP prior. However, they did not consider (1) multivariate survival outcomes in JMLS; (2) detection of the potential influential observations via Bayesian case-deletion approach which is an important step in the analysis of a data set; (3) flexible continuous distribution for the random effects; and (4) zero means of the random effects using the DP prior. Hence, the main purpose of this paper is to develop a novel Bayesian approach to JMLS for fully addressing the above mentioned problems based on the centered Dirichlet Process Mixture Model (CDPMM) specification of random effect distribution because CDPMM allows for more flexibility in modeling the random effect distribution, which is useful for the efficient estimation and for proper model-based variability estimates of the regression coefficients.

The study of this paper is motivated by a data set from a clinical trial conducted by the International Breast Cancer Study Group (IBCSG) (Chi and Ibrahim, 2006). In the IBCSG trial, each premenopausal woman with node-positive breast cancer was randomly assigned in a 2×2 factorial design to receive either the adjuvant chemotherapy or the reintroduction of three single courses of delayed chemotherapy. Different therapeutic procedure may have a direct affect on disease-free-survival (DFS) and overall survival (OS), as well as age, estrogen receptor (ER) status (negative/positive) and the number of positive nodes of the tumor, and the toxicity of a therapeutic procedure may adversely affect a patient's qualify of life (QOL), which is specifically related to DFS and OS. Four indicators of health-related QOL, including physical well-being (lousy-good), mood (miserable–happy), appetite (none-good) and perceived coping ("How much effort does it cost you to cope with your illness?" (a great deal-none)), were assessed at baseline and at months 3 and 18 after randomization. Under the normality assumption of random effects, Chi and Ibrahim (2006) proposed a joint likelihood approach to jointly model multidimensional QOL and the bivariate failure time random variables DFS and OS based on the subset of the IBCSG data set, which were collected from $n = 832$ patients from Switzerland, Sweden and New Zealand/Australia; Zhu et al. (2012) also considered the Bayesian local influence assessment of JMLS for the data set. However, to the best of our knowledge, there is little work done on developing some novel Bayesian case-deletion diagnostic measures to detect the potential influential observations for the data set that was fitted by JMLS using the CDPMM to allow flexible continuous distribution for the random effects.

Bayesian case-deletion diagnostics for detecting the potential influential observations (or sets of observations) have been proposed for many models such as normal linear regression models, mixed-effects models, generalized linear mixed models and survival models based on the Kullback–Leibler (K–L) divergence and the conditional predictive ordinate (CPO). For example, see Carlin and Polson (1991), Zeger and Karim (1991), Bradlow and Zaslavsky (1997), Weiss and Cho (1998), Spiegelhalter et al. (2002), Cho et al. (2009), Fong et al. (2010), Jackson et al. (2012). But, extending these existing Bayesian case-deletion diagnostics to our considered JMLS has computational challenge because of the complexity of the considered models and the unknown distribution of the random effects. To overcome the above mentioned difficulties, Markov chain Monte Carlo (MCMC) algorithm is employed to develop a Bayesian case-deletion influence diagnostic to assess the effect of cases (or sets of observations) on estimations of parameters based on the ϕ-divergence in the paper. A computationally feasible formula for the proposed Bayesian case-deletion diagnostic is also presented because the closed-form for the ϕ-divergence is not available.

The rest of this article is organized as follows. In Section 2, we describe a general semiparametric JMLS by using a centered Dirichlet Process Mixture Model (CDPMM) to specify the distribution of the random effects. Section 3 develops a Bayesian MCMC algorithm to make Bayesian inference on the JMLS by using the stick-breaking prior presentation for the centered DP prior and the blocked Gibbs sampler, together with the Metropolis–Hastings algorithm. Also, a Bayesian case-deletion diagnostic measure is proposed to detect the potential influential observations based on the ϕ-divergence in Section 3. Four simulation studies and an example are used to illustrate our proposed methodologies in Section 4. Some concluding remarks are given in Section 5. Technical details are presented in the Appendix A.

2. Model and notation

2.1. Notation

Consider the data from n independent individuals. For each individual, we consider K longitudinal responses and M time-to-event outcomes. Suppose that y_{ijk} is the observation of the kth longitudinal response measured at time t_{ijk} for the ith individual and $\mathbf{Y}_{ik} = (y_{i1k}, \ldots, y_{in_{ik}k})^T$ is the observed longitudinal process for the kth response for $i = 1, \ldots, n$, $k = 1, \ldots, K$ and $j = 1, \ldots, n_{ik}$. For the ith individual, we observe the event time $T_{im} = \min(T^*_{im}, C_{im})$ and the event indicator $\delta_{im} = \mathbf{1}(T^*_{im} \leq C_{im})$ for the mth time-to-event outcome for $m = 1, \ldots, M$, where $\mathbf{1}(A)$ is an indicator function of an event A, and T^*_{im} and C_{im} are the true survival time and the censoring time, respectively.

2.2. Generalized linear mixed longitudinal models

Let $\boldsymbol{b}_i = (\boldsymbol{b}^T_{i1}, \ldots, \boldsymbol{b}^T_{iK})^T$ be time-independent random effects underlying both the longitudinal and survival processes for the ith individual. It is usually assumed that $\boldsymbol{b}_i$'s have zero mean (Zhu et al., 2012), which facilitates the interpretation of other fixed effects in the joint model, and all components of the longitudinal outcomes and the time-to-event outcomes given $\boldsymbol{b}_i$ are conditionally independent. Specifically, we suppose that $y_{i1k}, \ldots, y_{in_{ik}k}$ given $\boldsymbol{b}_{ik}$ are conditionally independent and each y_{ijk} given $\boldsymbol{b}_{ik}$ follows the following exponential family distribution

$$p(y_{ijk}|\boldsymbol{b}_{ik}, \phi_k) = \exp[\phi_k^{-1}\{y_{ijk}\vartheta_{ijk} - s(\vartheta_{ijk})\} + c(y_{ijk}, \phi_k)], \quad (1)$$

Please cite this article in press as: Tang, N.-S., et al., Semiparametric Bayesian joint models of multivariate longitudinal and survival data. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.02.015

UNCORRECTED PROOF

where ϕ_k is a scale parameter, $c(y, \phi)$ is a function only depending on y and ϕ, ϑ_{ijk} is the scalar canonical parameter. Then, we have $\mu_{ijk} = E(y_{ijk}|\boldsymbol{b}_{ik}) = \dot{s}(\vartheta_{ijk})$ and $v_{ijk} = \text{var}(y_{ijk}|\boldsymbol{b}_{ik}) = \phi_k \ddot{s}(\vartheta_{ijk})$, where $\dot{s}(\vartheta) = \partial s(\vartheta)/\partial\vartheta$ and $\ddot{s}(\vartheta) = \partial^2 s(\vartheta)/\partial\vartheta^2$. It is assumed that the conditional mean μ_{ijk} satisfies

$$g_k(\mu_{ijk}) = \eta_{ik}(t_{ijk}, \boldsymbol{b}_{ik}), \tag{2}$$

where $g_k(\mu)$ is a monotone differential link function of μ and $\eta_{ik}(t, \boldsymbol{b}_{ik})$ is a parametric or nonparametric function of t, which may depend on other covariates of interest such as gender or age. Throughout this paper, we assume that $\eta_{ik}(t, \boldsymbol{b}_{ik})$ has the following form

$$\eta_{ik}(t, \boldsymbol{b}_{ik}) = \boldsymbol{R}_{ik}(t)\boldsymbol{\beta}_k + \boldsymbol{W}_{ik}(t)\boldsymbol{b}_{ik}, \tag{3}$$

where $\boldsymbol{\beta}_k$ $(p \times 1)$ and $\boldsymbol{b}_{ik}$ $(s \times 1)$ are fixed effects and random effects, respectively, $\boldsymbol{R}_{ik}(t)$ and $\boldsymbol{W}_{ik}(t)$ are known design matrices at time t, respectively. Here, we assume that $\boldsymbol{b}_i$'s are independent of all covariates in $\boldsymbol{R}_{ik}(t)$ and $\boldsymbol{W}_{ik}(t)$. The model defined in Eqs. (1)–(3) is referred to as a generalized linear mixed longitudinal model, which includes mixed discrete, ordinal and continuous responses and most existing models for longitudinal data.

2.3. Multivariate survival models

We consider the following multivariate survival model for survival process. Let $S_m(t|\boldsymbol{b}_i)$ and $\lambda_m(t|\boldsymbol{b}_i)$ be the marginal survival function and the marginal hazard function of the mth time-to-event outcome for the ith individual, respectively. For $m = 1, \ldots, M$, we assume that $\lambda_m(t|\boldsymbol{b}_i)$ has the following form

$$\lambda_m(t|\boldsymbol{b}_i) = \lambda_{m0}(t)\exp\{\boldsymbol{\alpha}_m^T\boldsymbol{\eta}_i(t, \boldsymbol{b}_i) + \boldsymbol{x}_i^T\boldsymbol{\gamma}_m\}, \tag{4}$$

where $\boldsymbol{\alpha}_m = (\alpha_{m1}, \ldots, \alpha_{mK})^T$ and $\boldsymbol{\gamma}_m = (\gamma_{m1}, \ldots, \gamma_{md})^T$ are vectors of unknown parameters, $\boldsymbol{\eta}_i(t, \boldsymbol{b}_i) = (\eta_{i1}(t, \boldsymbol{b}_{i1}), \ldots, \eta_{iK}(t, \boldsymbol{b}_{iK}))^T$, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{id})^T$ is a vector of time-independent covariates for the ith individual, and $\lambda_{m0}(t)$ is an unknown baseline hazard function. Also, it is assumed that $\boldsymbol{b}_i$'s are independent of $\boldsymbol{x}_i$. Then, the marginal survival function is given by $S_m(t|\boldsymbol{b}_i) = \exp\{-\int_0^t \lambda_m(u|\boldsymbol{b}_i)du\}$. Generally, the piecewise constant hazard model, Gamma process model, Beta process model or a Dirichlet process model can be employed to specify the prior distribution of the baseline hazard $\lambda_{m0}(t)$. Here, we consider the piecewise constant hazard model for $\lambda_{m0}(t)$. Following Ibrahim et al. (2004), $\lambda_{m0}(t)$ can be written as $\lambda_{m0}(t) = \sum_{l=1}^L h_{ml}\mathbf{1}(t \in (c_{m,l-1}, c_{ml}])$, where $0 = c_{m0} < c_{m1} < c_{m2} < \cdots < c_{mL}$ with $c_{mL} > \max(T_{1m}, \ldots, T_{nm})$ and $(c_{m,l-1}, c_{ml}]$ is the lth interval of finite partition of the mth time-to-event outcome time axis for $l = 1, \ldots, L$ and $m = 1, \ldots, M$. Let $\boldsymbol{h}_m = \{h_{m1}, \ldots, h_{mL}\}$. Then, the hazard rate $\lambda_m(T_{im}|\boldsymbol{b}_i, \boldsymbol{h}_m)$ can be written as

$$\lambda_m(T_{im}|\boldsymbol{b}_i, \boldsymbol{h}_m) = \sum_{l=1}^L h_{ml}\exp\{\boldsymbol{\alpha}_m^T\boldsymbol{\eta}_i(T_{im}, \boldsymbol{b}_i) + \boldsymbol{x}_i^T\boldsymbol{\gamma}_m\}\mathbf{1}(T_{im} \in (c_{m,l-1}, c_{ml}]), \tag{5}$$

and survival probability $S_m(T_{im}|\boldsymbol{b}_i, \boldsymbol{h}_m)$ can be written as

$$S_m(T_{im}|\boldsymbol{b}_i, \boldsymbol{h}_m) = \exp\left\{-\sum_{l=1}^{L_{im}} h_{ml}\int_{V_{iml-1}}^{V_{iml}} \exp\{\boldsymbol{\alpha}_m^T\boldsymbol{\eta}_i(u, \boldsymbol{b}_i) + \boldsymbol{x}_i^T\boldsymbol{\gamma}_m\}du\right\}, \tag{6}$$

where L_{im} satisfies $c_{m,L_{im}-1} < T_{im} \le c_{m,L_{im}}$, and $V_{iml} = c_{ml}$ when $l < L_{im}$ and $V_{iml} = T_{im}$ when $l = L_{im}$. Let $B_{iml} = \int_{V_{im,l-1}}^{V_{iml}} \exp(\boldsymbol{\alpha}_m^T\boldsymbol{\eta}_i(u, \boldsymbol{b}_i) + \boldsymbol{x}_i^T\boldsymbol{\gamma}_m)du$ for $l \le L_{im}$, and $B_{iml} = 0$ otherwise, which can be obtained by using the numerical integration method such as the Gaussian quadrature method. Then, $S_m(T_{im}|\boldsymbol{b}_i, \boldsymbol{h}_m)$ can be rewritten as

$$S_m(T_{im}|\boldsymbol{b}_i, \boldsymbol{h}_m) = \exp\left\{-\sum_{l=1}^{L_{im}} h_{ml}B_{iml}\right\}. \tag{7}$$

Let $\boldsymbol{T}_i = (T_{i1}, \ldots, T_{iM})^T$, $\boldsymbol{\theta}_y = \{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \phi_1, \ldots, \phi_K\}$, $\boldsymbol{\theta}_T = \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_M\}$, $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{iM})^T$, $\boldsymbol{h} = \{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_M\}$ and $\boldsymbol{\theta} = \{\boldsymbol{\theta}_y, \boldsymbol{\theta}_T\}$. For the ith individual, the likelihood function of $(\boldsymbol{T}_i, \boldsymbol{\delta}_i)$ given $\boldsymbol{b}_i$ is given by Q3

$$P(\boldsymbol{T}_i, \boldsymbol{\delta}_i|\boldsymbol{b}_i, \boldsymbol{h}; \boldsymbol{\theta}) = \prod_{m=1}^M S_m(T_{im}|\boldsymbol{b}_i, \boldsymbol{h}_m)\{\lambda_m(T_{im}|\boldsymbol{b}_i, \boldsymbol{h}_m)\}^{\delta_{im}}. \tag{8}$$

2.4. Specification of random effect distribution via CDPMM

We consider the distribution of random effects $\boldsymbol{b}_i$ with zero mean, where $\boldsymbol{b}_i = (\boldsymbol{b}_{i1}^T, \ldots, \boldsymbol{b}_{iK}^T)^T$. For random effects models, the traditional assumption for $\boldsymbol{b}_i$ is that $\boldsymbol{b}_i$'s are independently and identically distributed (i.i.d.) as a multivariate normal distribution $N_q(\boldsymbol{0}, \boldsymbol{\Phi})$ with $q = sK$. However, the normality assumption for $\boldsymbol{b}_i$ may be violated in some applications (Ohlssen et al., 2007). Rather than the traditional normality assumption for $\boldsymbol{b}_i$, we specify the distribution of $\boldsymbol{b}_i$ by using a DP mixture

UNCORRECTED PROOF

of normals, that is, $\boldsymbol{b}_i \overset{\text{i.i.d.}}{\sim} \sum_{g=1}^{\infty} \pi_g N_q(\boldsymbol{\mu}_g, \boldsymbol{\Omega}_g)$ with $(\boldsymbol{\mu}_g, \boldsymbol{\Omega}_g) \sim \mathscr{P}$, where $\mathscr{P}$ is a random probability with an unknown form. Obviously, due to the unknown form of $\mathscr{P}$, it is quite difficult to make Bayesian inference on parameters in $\boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_T$ on the basis of the above specified models. To solve the difficulty, one common approach is to specify a DP prior to approximate $\mathscr{P}$, i.e., $\mathscr{P} \sim \text{DP}(\tau F_0)$ in which F_0 is a base distribution that serves as a starting-point for constructing the nonparametric distribution, and τ is a weight that indicates the researcher's certainty of F_0 as the distribution of $\mathscr{P}$.

A popular way for specifying the DP prior $\text{DP}(\tau F_0)$ is the stick-breaking prior representation (Sethuraman, 1994), which leads to non-zero mean of random effects (Yang et al., 2010; Li et al., 2011) and discrete probability distribution of random effects (Ohlssen et al., 2007), which indicates that the direct use of the DP for density estimation is inappropriate when the underlying density of random effects is continuous. Hence, motivated by Ohlssen et al. (2007) and Yang et al. (2010), we consider the following CDPMM for random effects. The CDPMM for specifying the prior distribution of $\boldsymbol{b}_i$ (Ohlssen et al., 2007; Yang et al., 2010) can be expressed as

$$\boldsymbol{b}_i \overset{\text{i.i.d.}}{\sim} \sum_{g=1}^{\infty} \pi_g N_q(\boldsymbol{\mu}_g, \boldsymbol{\Omega}_g) \quad \text{with } \boldsymbol{\mu}_g = \boldsymbol{\mu}_g^* - \sum_{g=1}^{\infty} \pi_g \boldsymbol{\mu}_g^* \quad \text{and} \quad (\boldsymbol{\mu}_g^*, \boldsymbol{\Omega}_g) \overset{\text{i.i.d.}}{\sim} F_0,$$

where π_g is a random probability weight chosen to be independent of $(\boldsymbol{\mu}_g^*, \boldsymbol{\Omega}_g)$ such that $0 \le \pi_g \le 1$ and $\sum_{g=1}^{\infty} \pi_g = 1$. For the sake of computational convenience, we consider the following mixture model of the truncated approximation DP for $\mathscr{P}$:

$$\boldsymbol{b}_i \overset{\text{i.i.d.}}{\sim} \sum_{g=1}^{G} \pi_g N_q(\boldsymbol{\mu}_g, \boldsymbol{\Omega}_g) \quad \text{with } \boldsymbol{\mu}_g = \boldsymbol{\mu}_g^* - \sum_{g=1}^{G} \pi_g \boldsymbol{\mu}_g^* \quad \text{and} \quad (\boldsymbol{\mu}_g^*, \boldsymbol{\Omega}_g) \overset{\text{i.i.d.}}{\sim} F_0, \tag{9}$$

where $1 \le G < \infty$, and π_g is defined by the following stick-breaking procedure:

$$\pi_1 = \nu_1 \quad \text{and} \quad \pi_g = \nu_g \prod_{\iota=1}^{g-1} (1 - \nu_\iota) \quad \text{for } g = 2, \ldots, G, \tag{10}$$

where $\nu_g \overset{\text{i.i.d.}}{\sim} \text{Beta}(1, \tau)$ for $g = 1, \ldots, G-1$, and $\nu_G = 1$ so that $\sum_{g=1}^{G} \pi_g = 1$. The prior distribution for the unknown parameter τ is given by $\tau \sim \Gamma(a_1, a_2)$ with prespecified hyperparameters a_1 and a_2, and the details for selecting G can refer to Ishwaran and Zarepour (2000) and Ohlssen et al. (2007).

Sampling observations from posterior distributions of $\boldsymbol{b}_i$ with the above specified DP prior is quite difficult and inefficient. An efficient and flexible method for solving the above question within a Markov chain Monte Carlo (MCMC) framework is to represent $\boldsymbol{b}_i$ in terms of a latent variable $L_i \in \{1, \ldots, G\}$, which records each $\boldsymbol{b}_i$'s cluster membership and conveys its parametric value to the distribution of $\boldsymbol{b}_i$. Let $\boldsymbol{L} = \{L_1, \ldots, L_n\}$, $\boldsymbol{\pi} = \{\pi_1, \ldots, \pi_G\}$, $\boldsymbol{\mu}^* = \{\boldsymbol{\mu}_1^*, \ldots, \boldsymbol{\mu}_G^*\}$ and $\boldsymbol{\Omega} = \{\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_G\}$ in which $\boldsymbol{\Omega}_g = \text{diag}(\omega_{g1}, \ldots, \omega_{gq})$ for $g = 1, \ldots, G$. Then, Eq. (9) can be reformulated by

$$L_i | \boldsymbol{\pi} \overset{\text{i.i.d}}{\sim} \sum_{g=1}^{G} \pi_g \delta_g(\cdot) \quad \text{and} \quad (\boldsymbol{\pi}, \boldsymbol{\mu}^*, \boldsymbol{\Omega}) \sim f_1(\boldsymbol{\pi}) f_2(\boldsymbol{\mu}^*) f_3(\boldsymbol{\Omega}), \tag{11}$$

where $\delta_g(\cdot)$ denotes a discrete probability measure concentrated at g, $f_1(\boldsymbol{\pi})$ is specified by the stick-breaking prior as given in Eq. (10), the hierarchical structure for $\boldsymbol{\mu}^*$ related to $f_2(\boldsymbol{\mu}^*) = \prod_{g=1}^{G} f_2(\boldsymbol{\mu}_g^*)$ is specified by

$$\boldsymbol{\mu}_g^* | \boldsymbol{\zeta}, \boldsymbol{\Psi} \overset{\text{i.i.d}}{\sim} N_q(\boldsymbol{\zeta}, \boldsymbol{\Psi}), \qquad \boldsymbol{\zeta} | \boldsymbol{\zeta}^0, \boldsymbol{\Psi}^0 \sim N_q(\boldsymbol{\zeta}^0, \boldsymbol{\Psi}^0), \qquad \psi_j^{-1} | c_1, c_2 \sim \Gamma(c_1, c_2) \quad \text{for } j = 1, \ldots, q, \tag{12}$$

and the hierarchical structure for $\boldsymbol{\Omega}$ related to $f_3(\boldsymbol{\Omega}) = \prod_{g=1}^{G} \prod_{j=1}^{q} f_3(\omega_{gj})$ is specified by

$$\omega_{gj}^{-1} | \omega_j^a, \varpi_j \sim \Gamma(\omega_j^a, \varpi_j) \quad \text{and} \quad \varpi_j | \varpi_j^a, \varpi_j^b \sim \Gamma(\varpi_j^a, \varpi_j^b), \tag{13}$$

where $\boldsymbol{\Psi} = \text{diag}(\psi_1, \ldots, \psi_q)$, $\boldsymbol{\zeta}^0$, $\boldsymbol{\Psi}^0$, c_1, c_2, ω_j^a, ϖ_j^a and ϖ_j^b are prespecified hyperparameters, and $\Gamma(c_1, c_2)$ denotes the Gamma distribution with parameters c_1 and c_2. Selection of these hyperparameters can refer to Ohlssen et al. (2007). Given the values of L_i, $\boldsymbol{\mu}^*$ and $\boldsymbol{\Omega}$, we can sample $\boldsymbol{b}_i$ from $N_q(\boldsymbol{\mu}_{L_i}, \boldsymbol{\Omega}_{L_i})$ with $\boldsymbol{\mu}_{L_i} = \boldsymbol{\mu}_{L_i}^* - \Sigma_{g=1}^{G} \boldsymbol{\mu}_g^*$.

2.5. The joint likelihood function

Let $\boldsymbol{Y} = \{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n\}$ in which $\boldsymbol{Y}_i = \{\boldsymbol{Y}_{i1}, \ldots, \boldsymbol{Y}_{iK}\}$ for $i = 1, \ldots, n$, $\boldsymbol{T} = \{\boldsymbol{T}_1, \ldots, \boldsymbol{T}_n\}$, $\boldsymbol{\Delta} = \{\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_n\}$ and $\boldsymbol{b} = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$. Suppose that $\boldsymbol{\theta}_b$ contains all hyperparameters associated with distribution of $\boldsymbol{b}$, whilst $\boldsymbol{\theta}_h$ contains all hyperparameters related to distribution of $\boldsymbol{h}$. Under the above assumptions, the joint likelihood function for $(\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{\Delta}, \boldsymbol{b}, \boldsymbol{h})$ is given by

$$P(\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{\Delta}, \boldsymbol{b}, \boldsymbol{h} | \boldsymbol{\theta}_y, \boldsymbol{\theta}_T, \boldsymbol{\theta}_b, \boldsymbol{\theta}_h) = \prod_{i=1}^{n} \left\{ p(\boldsymbol{Y}_i | \boldsymbol{b}_i; \boldsymbol{\theta}_y) p(\boldsymbol{T}_i, \delta_i | \boldsymbol{b}_i, \boldsymbol{h}; \boldsymbol{\theta}) p(\boldsymbol{b}_i | \boldsymbol{\theta}_b) p(\boldsymbol{h} | \boldsymbol{\theta}_h) \right\}, \tag{14}$$

UNCORRECTED PROOF

where $p(\boldsymbol{Y}_i|\boldsymbol{b}_i;\boldsymbol{\theta}_y)=\prod_{k=1}^{K}\prod_{j=1}^{n_{ik}}p(y_{ijk}|\boldsymbol{b}_{ik};\boldsymbol{\theta}_y)$. Based on the above presented joint likelihood function, a Bayesian approach can be developed to make statistical inference on parameters in $\boldsymbol{\theta}=\{\boldsymbol{\theta}_y,\boldsymbol{\theta}_T\}$ by using the Gibbs sampler (Geman and Geman, 1984) and the Metropolis–Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970) for our considered joint models for longitudinal and survival data.

3. Bayesian analysis of joint models

To derive conditional distributions of unknown parameters in $\{\boldsymbol{\beta}_k,\phi_k,\boldsymbol{\alpha}_m,\boldsymbol{\gamma}_m\}$, it is necessary to specify their corresponding priors. Similar to Chow et al. (2011), we consider the following prior distributions for $\boldsymbol{\beta}_k$, ϕ_k, h_{ml}, $\boldsymbol{\alpha}_m$ and $\boldsymbol{\gamma}_m$:

$$\begin{array}{l}\boldsymbol{\beta}_k|\boldsymbol{\beta}_k^0,\boldsymbol{H}_\beta^0\sim N_p(\boldsymbol{\beta}_k^0,\boldsymbol{H}_\beta^0),\qquad \phi_k^{-1}|\phi_k^a,\phi_k^b\sim\Gamma(\phi_k^a,\phi_k^b),\qquad h_{ml}|\kappa_{m10},\kappa_{m11}\sim\Gamma(\kappa_{ml0},\kappa_{ml1}),\\ \boldsymbol{\alpha}_m|\boldsymbol{\alpha}_m^0,\boldsymbol{H}_\alpha^0\sim N_K(\boldsymbol{\alpha}_m^0,\boldsymbol{H}_\alpha^0)\quad\text{and}\quad \boldsymbol{\gamma}_m|\boldsymbol{\gamma}_m^0,\boldsymbol{H}_\gamma^0\sim N_d(\boldsymbol{\gamma}_m^0,\boldsymbol{H}_\gamma^0),\end{array}\tag{15}$$

for $k=1,\ldots,K$ and $m=1,\ldots,M$, where $\boldsymbol{\beta}_k^0$, $\boldsymbol{H}_\beta^0$, ϕ_k^a, ϕ_k^b, κ_{ml0}, κ_{ml1}, $\boldsymbol{\alpha}_m^0$, $\boldsymbol{H}_\alpha^0$, $\boldsymbol{\gamma}_m^0$ and $\boldsymbol{H}_\gamma^0$ are hyperparameters whose values are specified by the prior information.

Let $\boldsymbol{R}=\{\boldsymbol{R}_{ik}(t_{ijk}):i=1,\ldots,n,k=1,\ldots,K,j=1,\ldots,n_{ik}\}$, $\boldsymbol{W}=\{\boldsymbol{W}_{ik}(t_{ijk}):i=1,\ldots,n,k=1,\ldots,K,j=1,\ldots,n_{ik}\}$ and $\boldsymbol{X}=\{\boldsymbol{x}_i:i=1,\ldots,n\}$. Based on the above specified priors, the joint posterior probability density for $\{\boldsymbol{b},\boldsymbol{h},\boldsymbol{\theta}\}$ given $(\boldsymbol{Y},\boldsymbol{R},\boldsymbol{W},\boldsymbol{X},\boldsymbol{T},\boldsymbol{\Delta})$ is proportional to

$$\begin{aligned}&\exp\Bigg[\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{j=1}^{n_{ik}}\{(y_{ijk}\vartheta_{ijk}-s(\vartheta_{ijk}))/\phi_k+c(y_{ijk},\phi_k)\}-\sum_{i=1}^{n}\sum_{m=1}^{M}\sum_{l=1}^{L_{im}}h_{ml}B_{mil}\\&+\sum_{i=1}^{n}\sum_{m=1}^{M}\delta_{im}\left(\boldsymbol{\alpha}_m^T\boldsymbol{\eta}_i(T_{im},\boldsymbol{b}_i)+\boldsymbol{x}_i^T\boldsymbol{\gamma}_m\right)+\sum_{m=1}^{M}\sum_{l=1}^{L}\{d_{ml}\log(h_{ml})-\kappa_{ml1}h_{ml}\}\\&-\frac{1}{2}\sum_{k=1}^{K}(\boldsymbol{\beta}_k-\boldsymbol{\beta}_k^0)^T(\boldsymbol{H}_\beta^0)^{-1}(\boldsymbol{\beta}_k-\boldsymbol{\beta}_k^0)+\sum_{k=1}^{K}\{(1-\phi_k^a)\log(\phi_k)-\phi_k^b/\phi_k\}\\&-\frac{1}{2}\sum_{m=1}^{M}(\boldsymbol{\alpha}_m-\boldsymbol{\alpha}_m^0)^T(\boldsymbol{H}_\alpha^0)^{-1}(\boldsymbol{\alpha}_m-\boldsymbol{\alpha}_m^0)-\frac{1}{2}\sum_{m=1}^{M}(\boldsymbol{\gamma}_m-\boldsymbol{\gamma}_m^0)^T(\boldsymbol{H}_\gamma^0)^{-1}(\boldsymbol{\gamma}_m-\boldsymbol{\gamma}_m^0)\\&+\sum_{m=1}^{M}\sum_{l=1}^{L}(\kappa_{ml0}-1)\log(h_{ml})\Bigg]\prod_{i=1}^{n}p(\boldsymbol{b}_i|\boldsymbol{\theta}_b),\end{aligned}\tag{16}$$

where $d_{ml}=\sum_{i=1}^{n}\delta_{im}\mathbf{1}(T_{im}\in(c_{m,l-1},c_{ml}])$, and $p(\boldsymbol{b}_i|\boldsymbol{\theta}_b)$ is given in Eqs. (9)–(13). Clearly, it is rather difficult to make statistical inference on $(\boldsymbol{b},\boldsymbol{h},\boldsymbol{\theta})$ via the above presented joint posterior distribution. Due to the recent development in statistical computing, the MH algorithm within the Gibbs sampler is employed to sample a sequence of random observations from the joint posterior distribution $p(\boldsymbol{\zeta},\boldsymbol{\Psi},\boldsymbol{\varpi},\boldsymbol{\pi},\boldsymbol{\mu}^*,\boldsymbol{\Omega},\boldsymbol{L},\tau,\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{b},\boldsymbol{h}|\boldsymbol{Y},\boldsymbol{R},\boldsymbol{W},\boldsymbol{X},\boldsymbol{T},\boldsymbol{\Delta})$, then Bayesian estimates are obtained from the generated sequence of random observations, where $\boldsymbol{\varpi}=(\varpi_1,\ldots,\varpi_q)^T$, $\boldsymbol{\beta}=\{\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_K\}$, $\boldsymbol{\phi}=\{\phi_1,\ldots,\phi_K\}$, $\boldsymbol{\alpha}=\{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_M\}$ and $\boldsymbol{\gamma}=\{\boldsymbol{\gamma}_1,\ldots,\boldsymbol{\gamma}_M\}$. In this algorithm, observations $\{\boldsymbol{\zeta},\boldsymbol{\Psi},\boldsymbol{\varpi},\boldsymbol{\pi},\boldsymbol{\mu}^*,\boldsymbol{\Omega},\boldsymbol{L},\tau,\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{b},\boldsymbol{h}\}$ are generated iteratively from the following conditional distributions: $p(\boldsymbol{\zeta}|\boldsymbol{\mu}^*,\boldsymbol{\Psi},\boldsymbol{b})$, $p(\boldsymbol{\Psi}|\boldsymbol{\mu}^*,\boldsymbol{\zeta})$, $p(\boldsymbol{\varpi}|\boldsymbol{\Omega})$, $p(\boldsymbol{\pi}|\boldsymbol{L},\tau)$, $p(\boldsymbol{\mu}^*|\boldsymbol{\zeta},\boldsymbol{\Psi},\boldsymbol{\Omega},\boldsymbol{L},\boldsymbol{b})$, $p(\boldsymbol{\Omega}|\boldsymbol{\mu},\boldsymbol{\varpi},\boldsymbol{L},\tau)$, $p(\boldsymbol{L}|\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{b})$, $p(\tau|\boldsymbol{\pi})$, $p(\boldsymbol{\beta}_k|\phi_k,\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{b},\boldsymbol{h},\boldsymbol{Y},\boldsymbol{R},\boldsymbol{W},\boldsymbol{X},\boldsymbol{T},\boldsymbol{\Delta})$, $p(\phi_k|\boldsymbol{\beta}_k,\boldsymbol{b},\boldsymbol{Y},\boldsymbol{R},\boldsymbol{W})$, $p(\boldsymbol{\alpha}_m|\boldsymbol{\beta},\boldsymbol{\gamma}_m,\boldsymbol{b},\boldsymbol{h}_m,\boldsymbol{R},\boldsymbol{W},\boldsymbol{X},\boldsymbol{T},\boldsymbol{\Delta})$, $p(\boldsymbol{\gamma}_m|\boldsymbol{\beta},\boldsymbol{\alpha}_m,\boldsymbol{b},\boldsymbol{h}_m,\boldsymbol{R},\boldsymbol{W},\boldsymbol{X},\boldsymbol{T},\boldsymbol{\Delta})$, $p(\boldsymbol{b}|\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{L},\boldsymbol{\theta}_y,\boldsymbol{\theta}_T,\boldsymbol{h},\boldsymbol{R},\boldsymbol{W},\boldsymbol{X},\boldsymbol{T},\boldsymbol{\Delta})$ and $p(h_{ml}|\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\alpha}_m,\boldsymbol{\gamma}_m,\boldsymbol{b},\boldsymbol{R},\boldsymbol{W},\boldsymbol{X},\boldsymbol{T},\boldsymbol{\Delta})$.

The above conditional distributions are presented in the Appendix. Clearly, the conditional distributions $p(\boldsymbol{\zeta}|\boldsymbol{\mu}^*,\boldsymbol{\Psi},\boldsymbol{b})$, $p(\boldsymbol{\Psi}|\boldsymbol{\mu}^*,\boldsymbol{\zeta})$, $p(\varpi|\boldsymbol{\Omega})$, $p(\tau|\boldsymbol{\pi})$ and $p(h_{ml}|\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\alpha}_m,\boldsymbol{\gamma}_m,\boldsymbol{b},\boldsymbol{R},\boldsymbol{W},\boldsymbol{X},\boldsymbol{T},\boldsymbol{\Delta})$ are some familiar distributions such as normal and Gamma distributions. Drawing observations from these conditional distributions is straightforward and fast. But conditional distributions $p(\boldsymbol{\mu}^*|\boldsymbol{\zeta},\boldsymbol{\Psi},\boldsymbol{\Omega},\boldsymbol{L},\boldsymbol{b})$, $p(\boldsymbol{\Omega}|\boldsymbol{\mu},\varpi,\boldsymbol{L},\tau)$, $p(\boldsymbol{\pi}|\boldsymbol{L},\tau)$, $p(\boldsymbol{L}|\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{b})$, $p(\boldsymbol{\beta}_k|\phi_k,\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{b},\boldsymbol{h},\boldsymbol{Y},\boldsymbol{R},\boldsymbol{W},\boldsymbol{X},\boldsymbol{T},\boldsymbol{\Delta})$, $p(\phi_k|\boldsymbol{\beta}_k,\boldsymbol{b},\boldsymbol{Y},\boldsymbol{R},\boldsymbol{W})$, $p(\boldsymbol{\alpha}_m|\boldsymbol{\beta},\boldsymbol{\gamma}_m,\boldsymbol{b},\boldsymbol{h}_m,\boldsymbol{R},\boldsymbol{W},\boldsymbol{X},\boldsymbol{T},\boldsymbol{\Delta})$, $p(\boldsymbol{\gamma}_m|\boldsymbol{\beta},\boldsymbol{\alpha}_m,\boldsymbol{b},\boldsymbol{h}_m,\boldsymbol{R},\boldsymbol{W},\boldsymbol{X},\boldsymbol{T},\boldsymbol{\Delta})$ and $p(\boldsymbol{b}|\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{L},\boldsymbol{\theta}_y,\boldsymbol{\theta}_T,\boldsymbol{h},\boldsymbol{R},\boldsymbol{W},\boldsymbol{X},\boldsymbol{T},\boldsymbol{\Delta})$ are some unfamiliar and complicated distributions, thus generating observations from these conditional distributions is quite difficult. Hence, the commonly used MH algorithm is employed to sample observations from these conditional distributions. The details for implementing MH algorithm are summarized in the Appendix.

3.1. Bayesian estimates

Observations generated by using the Gibbs sampler together with the MH algorithm are used to obtain the joint Bayesian estimates of parameters in $\{(\boldsymbol{\beta}_k,\phi_k,\boldsymbol{\alpha}_m,\boldsymbol{\gamma}_m):k=1,\ldots,K,m=1,\ldots,M\}$ and random effects $\boldsymbol{b}_i$ and piecewise constant h_{ml} in the baseline hazard function as well as their corresponding standard errors.

Please cite this article in press as: Tang, N.-S., et al., Semiparametric Bayesian joint models of multivariate longitudinal and survival data. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.02.015

UNCORRECTED PROOF

Let $\{(\boldsymbol{\beta}_k^{(\ell)}, \phi_k^{(\ell)}, \boldsymbol{\alpha}_m^{(\ell)}, \boldsymbol{\gamma}_m^{(\ell)}, \boldsymbol{b}_i^{(\ell)}, h_{ml}^{(\ell)}) : \ell = 1, \ldots, \mathfrak{L}\}$ be an observed sample of $\{\boldsymbol{\beta}_k, \phi_k, \boldsymbol{\alpha}_m, \boldsymbol{\gamma}_m, \boldsymbol{b}_i, h_{ml}\}$ generated from the joint conditional distribution $p(\boldsymbol{\beta}, \phi, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{b}, \boldsymbol{h}|\boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})$ via the preceding presented Gibbs sampler together with the MH algorithm. Then, Bayesian estimates of $\boldsymbol{\beta}_k$, ϕ_k, $\boldsymbol{\alpha}_m$, $\boldsymbol{\gamma}_m$, $\boldsymbol{b}_i$ and h_{ml} can be obtained by

$$\hat{\boldsymbol{\beta}}_k = \frac{1}{\mathfrak{L}}\sum_{\ell=1}^{\mathfrak{L}} \boldsymbol{\beta}_k^{(\ell)}, \qquad \hat{\phi}_k = \frac{1}{\mathfrak{L}}\sum_{\ell=1}^{\mathfrak{L}} \phi_k^{(\ell)}, \qquad \hat{\boldsymbol{\alpha}}_m = \frac{1}{\mathfrak{L}}\sum_{\ell=1}^{\mathfrak{L}} \boldsymbol{\alpha}_m^{(\ell)},$$

$$\hat{\boldsymbol{\gamma}}_m = \frac{1}{\mathfrak{L}}\sum_{\ell=1}^{\mathfrak{L}} \boldsymbol{\alpha}_m^{(\ell)}, \qquad \hat{\boldsymbol{b}}_i = \frac{1}{\mathfrak{L}}\sum_{\ell=1}^{\mathfrak{L}} \boldsymbol{b}_i^{(\ell)}, \qquad \hat{h}_{ml} = \frac{1}{\mathfrak{L}}\sum_{\ell=1}^{\mathfrak{L}} h_{ml}^{(\ell)}.$$

Also, the baseline hazard function $\lambda_{m0}(t)$ can be estimated by $\hat{\lambda}_{m0}(t) = \Sigma_{l=1}^{L}\hat{h}_{ml}\mathbf{1}(t \in (c_{m,l-1}, c_{ml}])$. Similarly, the consistent estimates of the posterior covariance matrices $\text{var}(\boldsymbol{\beta}_k|\boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})$, $\text{var}(\phi_k|\boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W})$, $\text{var}(\boldsymbol{\alpha}_m|\boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})$ and $\text{var}(\boldsymbol{\gamma}_m|\boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})$ can be obtained via the sample covariance matrices of the above generated observations. For example, $\widehat{\text{var}}(\boldsymbol{\beta}_k|\boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta}) = \frac{1}{\mathfrak{L}-1}\Sigma_{\ell=1}^{\mathfrak{L}}(\boldsymbol{\beta}_k^{(\ell)} - \hat{\boldsymbol{\beta}}_k)(\boldsymbol{\beta}_k^{(\ell)} - \hat{\boldsymbol{\beta}}_k)^T$. Thus, the standard errors for components of $\boldsymbol{\beta}_k$ can be obtained from the diagonal elements of the sample covariance matrix.

3.2. Bayesian case-deletion influence analysis

Let $\boldsymbol{D}$ be a full data set, and $\boldsymbol{D}_{-i}$ be a subset of $\boldsymbol{D}$ with the data corresponding to the ith individual deleted. Let $p(\boldsymbol{\theta}|\boldsymbol{D})$ and $p(\boldsymbol{\theta}|\boldsymbol{D}_{-i})$ be the posterior distributions of $\boldsymbol{\theta}$ for the full data set $\boldsymbol{D}$ and the reduced data set $\boldsymbol{D}_{-i}$, respectively. The Kullback–Leibler (K–L) divergence is a widely used measure for investigating the effect of the ith case on the joint posterior distribution of $\boldsymbol{\theta}$. Similar to the K–L divergence, we consider the following Bayesian diagnostic measure

$$D_\varphi(i) = \int \varphi(\widetilde{R}_{(i)}(\boldsymbol{\theta}))p(\boldsymbol{\theta}|\boldsymbol{D})d\boldsymbol{\theta} = E_{\theta|D}\{\varphi(\widetilde{R}_{(i)}(\boldsymbol{\theta}))\},$$

where $\widetilde{R}_{(i)}(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\boldsymbol{D}_{-i})/p(\boldsymbol{\theta}|\boldsymbol{D})$ and $\varphi(\cdot)$ is a convex function with $\varphi(1) = 0$, and $E_{\theta|D}$ is the expectation taken with respect to the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{D})$. Clearly, $D_\varphi(i)$ measures the discrepancy between two posterior distributions $p(\boldsymbol{\theta}|\boldsymbol{D}_{-i})$ and $p(\boldsymbol{\theta}|\boldsymbol{D})$ (Weiss and Cook, 1992; Weiss, 1996; Ibrahim, Sinha, Zhu, 2009) and it can be regarded as a Bayesian analogue of the likelihood displacement as discussed in Cook and Weisberg (1982), and the observation for the large value of $D_\varphi(i)$ is identified to be influential. Different selection of $\varphi(\cdot)$ corresponds to various divergences, for example, $\varphi(u) = u\log(u)$ and $-\log(u)$ correspond to the K–L divergence, $\varphi(u) = (u-1)\log(u)$ corresponds to the symmetric K–L divergence, and $\varphi(u) = 0.5|u-1|$ and $\varphi(u) = (u-1)^2$ correspond to the L_1-distance and the χ^2-divergence, respectively (Weiss, 1992).

To compute $D_\varphi(i)$, we need to calculate the marginal posterior density functions $p(\boldsymbol{\theta}|\boldsymbol{D})$ and $p(\boldsymbol{\theta}|\boldsymbol{D}_{-i})$. It is easily shown that $p(\boldsymbol{\theta}|\boldsymbol{D}_{-i}) = \{p_i(\boldsymbol{\theta})\}^{-1}p(\boldsymbol{\theta}|\boldsymbol{D})/\int\{p_i(\boldsymbol{\theta})\}^{-1}p(\boldsymbol{\theta}|\boldsymbol{D})d\boldsymbol{\theta}$, where $p_i(\boldsymbol{\theta}) = p(\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{\delta}|\boldsymbol{\theta}; \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X})/p(\boldsymbol{Y}_{-i}, \boldsymbol{T}_{-i}, \boldsymbol{\delta}_{-i}|\boldsymbol{\theta}; \boldsymbol{R}_{-i}, \boldsymbol{W}_{-i}, \boldsymbol{X}_{-i})$. For our considered models, $p_i(\boldsymbol{\theta}) = p(\boldsymbol{Y}_i, \boldsymbol{T}_i, \boldsymbol{\delta}_i|\boldsymbol{\theta}; \boldsymbol{R}_i, \boldsymbol{W}_i, \boldsymbol{x}_i)$ because of the independence of individuals. Thus, combining the above equations yields

$$D_\varphi(i) = E_{\theta|D}\left[\varphi\left(\frac{\{p_i(\boldsymbol{\theta})\}^{-1}}{E_{\theta|D}\{p_i(\boldsymbol{\theta})\}^{-1}}\right)\right]. \tag{17}$$

Specifically, if we take $\varphi(u)$ to be the K–L divergence (i.e., $\varphi(u) = -\log(u)$), $D_\varphi(i)$ has the following expression: $D_\varphi(i) = \log E_{\theta|D}\{p_i(\boldsymbol{\theta})\}^{-1} + E_{\theta|D}\{\log p_i(\boldsymbol{\theta})\}$.

It is easily seen from Eq. (17) that computing $D_\varphi(i)$ heavily depends on computation of $p_i(\boldsymbol{\theta})$, which is given by

$$p_i(\boldsymbol{\theta}) = \int p(\boldsymbol{Y}_i, \boldsymbol{T}_i, \boldsymbol{\delta}_i, \boldsymbol{b}_i, \boldsymbol{h}|\boldsymbol{\theta}; \boldsymbol{R}_i, \boldsymbol{W}_i, \boldsymbol{x}_i)d\boldsymbol{b}_i d\boldsymbol{h} = E_h[E_{b_i}\{p(\boldsymbol{Y}_i, \boldsymbol{T}_i, \boldsymbol{\delta}_i|\boldsymbol{\theta}; \boldsymbol{R}_i, \boldsymbol{W}_i, \boldsymbol{x}_i, \boldsymbol{b}_i, \boldsymbol{h})\}],$$

where E_{b_i} and E_h denote the expectation taken with respect to distributions of $\boldsymbol{b}_i$ (i.e., $p(\boldsymbol{b}_i|\boldsymbol{\theta}_b)$) and $\boldsymbol{h}$ (i.e., $p(\boldsymbol{h}|\boldsymbol{\theta}_h)$), respectively. Clearly, it is quite difficult and complicated to evaluate $p_i(\boldsymbol{\theta})$ because of high-dimensional integral involved and the complexity of $p(\boldsymbol{Y}_i, \boldsymbol{T}_i, \boldsymbol{\delta}_i, \boldsymbol{b}_i, \boldsymbol{h}|\boldsymbol{\theta}; \boldsymbol{R}_i, \boldsymbol{W}_i, \boldsymbol{x}_i)$. Particularly, if $\boldsymbol{b}_i$ and $\boldsymbol{h}$ follow some unfamiliar standard distribution, the Monte Carlo method is used to approximate $p_i(\boldsymbol{\theta})$, that is,

$$p_i(\boldsymbol{\theta}) \approx \frac{1}{FJ}\sum_{f=1}^{F}\sum_{j=1}^{J} p(\boldsymbol{Y}_i, \boldsymbol{T}_i, \boldsymbol{\delta}_i|\boldsymbol{\theta}; \boldsymbol{R}_i, \boldsymbol{W}_i, \boldsymbol{x}_i, \boldsymbol{b}_i^{(f)}, \boldsymbol{h}^{(j)}),$$

where observations $\{\boldsymbol{b}_i^{(f)} : f = 1, \ldots, F\}$ and $\{\boldsymbol{h}^{(j)} : j = 1, \ldots, J\}$ are generated from $p(\boldsymbol{b}_i|\boldsymbol{\theta}_b)$ and $p(\boldsymbol{h}|\boldsymbol{\theta}_h)$, respectively. For our conducted simulation studies and real example in Section 4, we take $F = 2000$ and $J = 2000$.

4. Simulation studies

To investigate the finite performance of the preceding proposed Bayesian procedures under known population assumption of random effects $\boldsymbol{b}_i$, we generated data using the JMLS described in Section 2 with different distributions

Please cite this article in press as: Tang, N.-S., et al., Semiparametric Bayesian joint models of multivariate longitudinal and survival data. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.02.015

UNCORRECTED PROOF

for $\boldsymbol{b}_i$ but the same sample size (i.e., $n = 200$). 50 replications were conducted to evaluate Bayesian results for each of the following presented simulation assumptions.

In the following simulation studies, we considered two longitudinal markers in the generalized linear mixed-effects longitudinal model and bivariate survival times in the multivariate survival model, that is, $K = 2$ and $M = 2$. For the ith individual, two longitudinal markers were monitored simultaneously at t_{ij} ($t_{ij1} = t_{ij2} = t_{ij}$), where $t_{ij} \in \{0.00, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50\}$ and $t_{ij} \le \max(T_{i1}, T_{i2})$, and were predictive of the bivariate survival times. The trajectory function was given by

$$\eta_{ik}(t_{ijk}, b_{ik}) = \beta_{k0} + t_{ijk}\beta_{k1} + r_i\beta_{k2} + b_{ik}, \tag{18}$$

where r_i was generated from a uniform distribution $U(0, 1)$ representing a baseline covariate in the longitudinal model. Conditional on $\boldsymbol{b}_i = (b_{i1}, b_{i2})^T$, the two event and censoring times were assumed to be independent and their marginal hazard functions were given by

$$\lambda_m(t|\boldsymbol{b}_i) = \lambda_{m0}(t)\exp\{\alpha_{m1}\eta_{i1}(t, b_{i1}) + \alpha_{m2}\eta_{i2}(t, b_{i2}) + \boldsymbol{x}_i^T\boldsymbol{\gamma}_m\}, \tag{19}$$

for $m = 1$ and 2, where $\boldsymbol{x}_i = (x_{i1}, x_{i2})^T$ was a vector of time-independent covariates, and x_{i1} was generated from $N(0, 0.25^2)$ and x_{i2} was generated from $N(0.50x_{i1}, 0.4^2)$. The censoring times C_{im}'s were independently generated via $C_{im} = 1.2\mathbf{1}(1.2 > 4.5u_{im}) + 4.5u_{im}\mathbf{1}(1.2 \le 4.5u_{im})$, where u_{im}'s were independently generated from a uniform distribution $U(0, 1)$ for $m = 1$ and $m = 2$, which gives a uncensored rate between 80% and 85%.

In MCMC algorithm, we took $L = 100$ subintervals with equal lengths in the piecewise constant hazard model and $G = 25$ in the truncated approximation CDPMM, and considered the following hyperparameters for h_{ml}, $\boldsymbol{\zeta}$, $\boldsymbol{\Psi}$, ψ_j^{-1}, $\boldsymbol{\varpi}$, ω_{gj}, $\boldsymbol{\beta}_k$, ϕ_k, $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$: $\kappa_{1l0} = 12$, $\kappa_{1l1} = 20$, $\kappa_{2l0} = 14$ and $\kappa_{2l1} = 20$ for $l = 1, \ldots, 100$, $\boldsymbol{\zeta}^0 = (0, 0)^T$, $\boldsymbol{\Psi}^0 = \boldsymbol{I}_2$, $c_1 = 11$, $c_2 = 2.5$, $\varpi_j^a = 80$, $\varpi_j^b = 20$, $\omega_j^a = 3$, $\boldsymbol{H}_\beta^0 = 0.1\boldsymbol{I}_3$, $\boldsymbol{H}_\alpha^0 = 0.1\boldsymbol{I}_2$, $\boldsymbol{H}_\gamma^0 = 0.1\boldsymbol{I}_2$, $\phi_1^a = 10$ and $\phi_1^b = 4$ which are not required in the second simulation because ϕ_1 is a constant, $\phi_2^a = 10$ and $\phi_2^b = 4$; whilst the hyperparameters $\boldsymbol{\beta}_k^0$ and $\boldsymbol{\alpha}_m^0$ and $\boldsymbol{\gamma}_m^0$ were taken to be their corresponding true values for $k = 1$ and 2 and $m = 1$ and 2, respectively. To generate large value of τ (Chow et al., 2011), which leads to more unique $\boldsymbol{b}_{ik}$ values, we took the hyperparameters a_1 and a_2 to be 25 and 5 in our conducted simulation studies, respectively.

4.1. Performance assessment of Bayesian estimates

4.1.1. Normal distributional assumptions for longitudinal data

In the first simulation study, the kth longitudinal response y_{ijk} was generated from the following model $y_{ijk} = \eta_{ik}(t_{ijk}, b_{ik}) + \varepsilon_{ijk}$ with the following three assumptions of $\boldsymbol{b}_i$, where $\eta_{ik}(t_{ijk}, \boldsymbol{b}_{ik})$ is specified in Eq. (18), ε_{ij1} and ε_{ij2} were independently generated from $N(0, \sigma_{11}^2)$ and $N(0, \sigma_{22}^2)$ with $\sigma_{11}^2 = \sigma_{22}^2 = 0.4$, respectively, and the true values of $\beta_{k\ell}$ for $k = 1$ and 2 and $\ell = 0, 1$ and 2 were taken to be $\beta_{10} = -0.9$, $\beta_{11} = 0.8$, $\beta_{12} = 1.0$, $\beta_{20} = -1.1$, $\beta_{21} = 0.75$ and $\beta_{22} = 0.8$. In this case, $\phi_1 = \sigma_{11}^2$ and $\phi_2 = \sigma_{22}^2$. The survival time T_{im}^* was independently generated from Eq. (19) with the following specifications: $\lambda_{10}(t) = 0.6$ and $\lambda_{20}(t) = 0.7$, $\alpha_{11} = 1.2$, $\alpha_{12} = 0.8$, $\alpha_{21} = 0.9$, $\alpha_{22} = 0.8$, $\boldsymbol{\gamma}_1 = (0.85, -0.85)^T$ and $\boldsymbol{\gamma}_2 = (-0.85, 0.85)^T$. Here, we considered a special generalized linear mixed-effects longitudinal model because the normal distribution is a special case of exponential family distribution.

To test the effectiveness of using CDPMM to approximate distribution of random effects $\boldsymbol{b}_i$, we considered the following four sets of distributional assumptions for $\boldsymbol{b}_i$.

Assumption 1. The true distribution of $\boldsymbol{b}_i$ was assumed to be a normal distribution: $b_{i1} \sim N(0, 0.5^2)$ and $b_{i2} \sim N(0, 0.5^2)$. This assumption was used to show that CDPMM can still capture characteristics of the normal distribution as a special case even when the normality assumption holds.

Assumption 2. We specified the true distribution of $\boldsymbol{b}_i$ as $b_{i1} = b_{i1}^* - 1$ and $b_{i2} = b_{i2}^* - 1.5$, where $b_{i1}^* \overset{\text{i.i.d}}{\sim} \Gamma(4, 4)$ and $b_{i2}^* \overset{\text{i.i.d}}{\sim} \Gamma(9, 6)$. In this case, the means of b_{i1} and b_{i2} were 0, respectively, which indicates that the posited model is identifiable.

Assumption 3. We specified the true distribution of $\boldsymbol{b}_i$ to be bimodal:

$$b_{i1} \overset{\text{i.i.d}}{\sim} 0.6N(-0.4, 0.2^2) + 0.4N(0.6, 0.2^2) \quad \text{and} \quad b_{i2} \overset{\text{i.i.d}}{\sim} 0.5N(-0.5, 0.2^2) + 0.5N(0.5, 0.2^2).$$

For each data set generated from the above presented procedure under each assumption of $\boldsymbol{b}_i$, the preceding proposed MCMC algorithm combining the stick-break prior of CDPMM and the Gibbs sampler together with the MH algorithm was used to obtain Bayesian estimates of parameters and random effects. Three parallel sequences of observations generated from three different starting values were used to compute the EPSR values (Geman and Geman, 1984) of parameters. Q4 From Fig. S1 in Appendix B, for all the test runs, we observed that the EPSR values of all parameters were less than 1.2 after about 10 000 iterations. Thus, $\mathfrak{L} = 5000$ observations collected after 10 000 iterations were used to compute the

Please cite this article in press as: Tang, N.-S., et al., Semiparametric Bayesian joint models of multivariate longitudinal and survival data. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.02.015

UNCORRECTED PROOF

Table 1a
Bayesian estimates of parameters for the multivariate JMLS with CDPMM assumption of random effects fitted to 50 replicated data sets under the true distribution of $\boldsymbol{b}_i$: normal distribution (Assumption 1), Gamma distribution (Assumption 2), bimodal distribution (Assumption 3) and normal or Gamma distribution (Assumption 4).

Par.	Assumption 1		Assumption 2		Assumption 3		Assumption 4	
	RB (%)	RMS	RB (%)	RMS	RB (%)	RMS	RB (%)	RMS
α_{11}	0.055	0.096	1.435	0.129	2.106	0.122	1.597	0.155
α_{12}	−0.655	0.133	1.812	0.123	−2.948	0.138	5.544	0.141
α_{21}	0.576	0.104	1.685	0.115	0.249	0.123	1.642	0.130
α_{22}	−0.882	0.132	3.205	0.125	1.248	0.129	3.478	0.128
γ_{11}	−1.791	0.134	1.544	0.146	−0.435	0.125	−0.563	0.150
γ_{12}	1.033	0.124	0.111	0.096	−1.551	0.108	1.296	0.107
γ_{21}	0.707	0.137	0.609	0.126	2.451	0.115	3.176	0.135
γ_{22}	−0.558	0.099	0.988	0.112	−0.318	0.121	3.446	0.130
β_{10}	0.293	0.066	−0.197	0.075	0.493	0.082	3.035	0.091
β_{11}	1.556	0.041	0.445	0.043	0.311	0.037	1.306	0.053
β_{12}	−0.858	0.116	−1.160	0.103	1.111	0.099	−1.861	0.138
β_{20}	−0.977	0.068	−0.366	0.070	0.698	0.065	0.606	0.066
β_{21}	−0.969	0.043	−1.380	0.045	1.497	0.043	−0.989	0.030
β_{22}	−1.757	0.112	1.482	0.115	−1.979	0.108	3.596	0.097
σ_{11}	−0.679	0.017	−0.757	0.018	−0.624	0.020	–	–
σ_{22}	−0.092	0.020	−1.037	0.019	−0.939	0.020	−2.003	0.013
ΣRMS		1.442		1.460		1.435		1.564

Table 1b
Means and standard deviations of b_{i1} and b_{i2} for the multivariate JMLS with CDPMM assumption of random effects fitted to 50 replicated data sets under the true distribution of $\boldsymbol{b}_i$: normal distribution (Assumption 1), Gamma distribution (Assumption 2), bimodal distribution (Assumption 3) and normal or Gamma distribution (Assumption 4).

Random effects	True mean	Est mean	SD	Est SD
	Assumption 1			
b_{i1}	0.000	0.005	0.500	0.437
b_{i2}	0.000	0.004	0.500	0.436
	Assumption 2			
b_{i1}	0.000	0.005	0.500	0.440
b_{i2}	0.000	0.005	0.500	0.434
	Assumption 3			
b_{i1}	0.000	−0.000	0.530	0.473
b_{i2}	0.000	0.000	0.539	0.489
	Assumption 4			
b_{i1}	0.000	0.021	0.500	0.409
b_{i2}	0.000	0.002	0.500	0.453

simulation results for the 50 replications. It roughly took 0.65 min in a Thinkpad SL400 serve to run 1000 iterations for our proposed semiparametric Bayesian approach, and 0.60 min for the parametric Bayesian approach with $\boldsymbol{b}_i$ following a normal distribution. Our programs were written in Matlab and R language, and are available upon request from the first author.

Results were reported in Table 1a, where 'RB' is the relative bias, i.e., RB $= \frac{1}{50}\sum_{i=1}^{50}(\hat{\gamma}^{(i)}/\gamma - 1)$ in which $\hat{\gamma}^{(i)}$ is Bayesian estimate of parameter γ in the ith replication, and 'RMS' is the root mean square between the estimates based on 50 replications and its true value. Table 1a showed that Bayesian estimates are reasonably accurate regardless of the distributional assumptions of $\boldsymbol{b}_i$ in the sense that their absolute values of RB are less than 10% and their RMS values are less than 0.2.

To investigate the accuracy of using CDPMM to approximate distribution of $\boldsymbol{b}_i$, we computed means and standard deviations of $\hat{\boldsymbol{b}}_i$ across individuals for a randomly selected replication. Results were given in Table 1b, where 'True mean' denotes the true mean of random effect, 'Est mean' denotes the estimated mean of random effect, 'SD' is the true standard deviation of random effect and 'Est SD' is the estimated standard deviation of random effect. Densities of $\hat{b}_{i1}$ and $\hat{b}_{i2}$ for a randomly selected replication against their corresponding true densities under Assumptions 1–3 were plotted in Figs. 1, S2(a) and (b) in the Appendix B. These results showed that (i) the mean and standard deviation of the true distribution of $\boldsymbol{b}_i$ can be estimated well by our proposed method under our considered three assumptions; (ii) the conditional distributions based on the truncated approximation CDPMM were flexible enough to capture the general shapes of the different distributions of $\boldsymbol{b}_i$ used in our considered three assumptions; (iii) the shapes of the estimated distributions in Assumption 1 were more accurately matched with the true distributions of $\boldsymbol{b}_i$ than those in the other two assumptions because the true densities (i.e., multivariate normal) of $\boldsymbol{b}_i$ were of the same form as the base distribution; (iv) in Assumptions 2–3, the estimated density was generally similar with the density of the true distribution of $\boldsymbol{b}_i$, but slight differences were observed close to the tail areas in Assumption 2 and close to the inflection points in Assumption 3; (v) the estimated baseline hazard functions $\hat{\lambda}_{m0}(t)$ slightly wave around true baseline hazard functions $\lambda_{m0}(t)$ under our considered three assumptions but the

Please cite this article in press as: Tang, N.-S., et al., Semiparametric Bayesian joint models of multivariate longitudinal and survival data. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.02.015

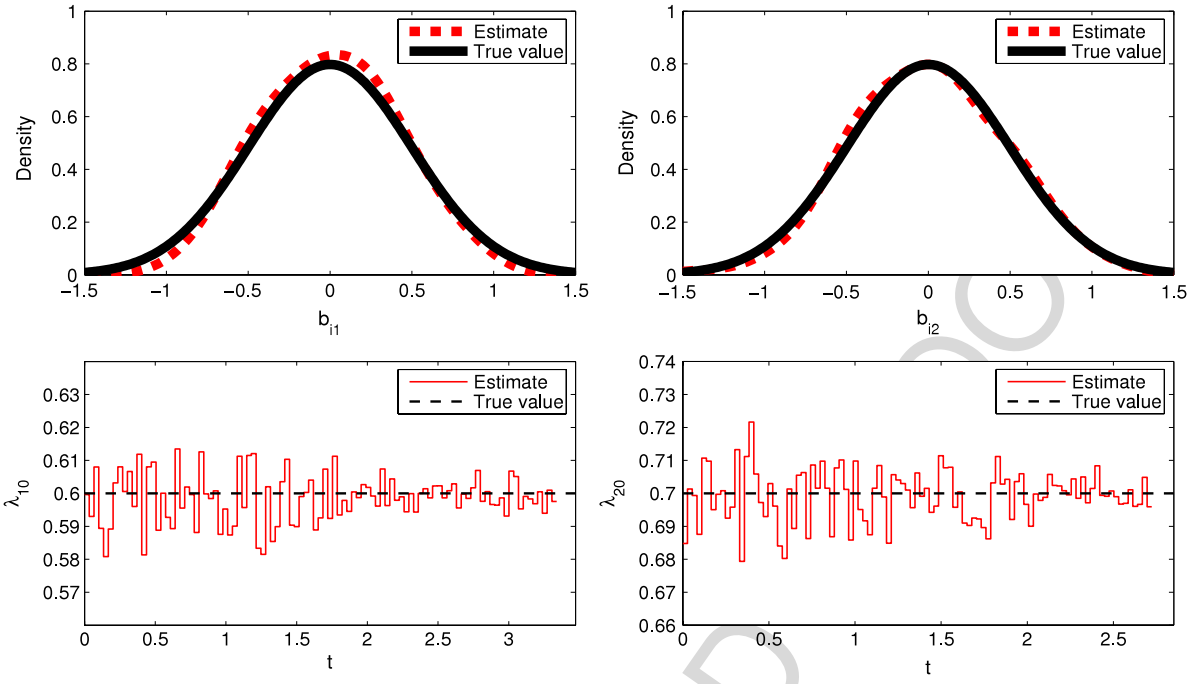


Fig. 1. Estimated densities versus true densities for random effects b_{i1} and b_{i2} and the estimated baseline hazard functions for λ_{10} and λ_{20} under Assumption 1 with a randomly selected replication in the first simulation study.

absolute difference at any time point is less than 0.03. All above findings indicated that the preceding proposed estimation procedures can well recover the true information under various considered distributional assumptions for $\boldsymbol{b}_i$.

4.1.2. Poisson distributional assumptions for longitudinal data

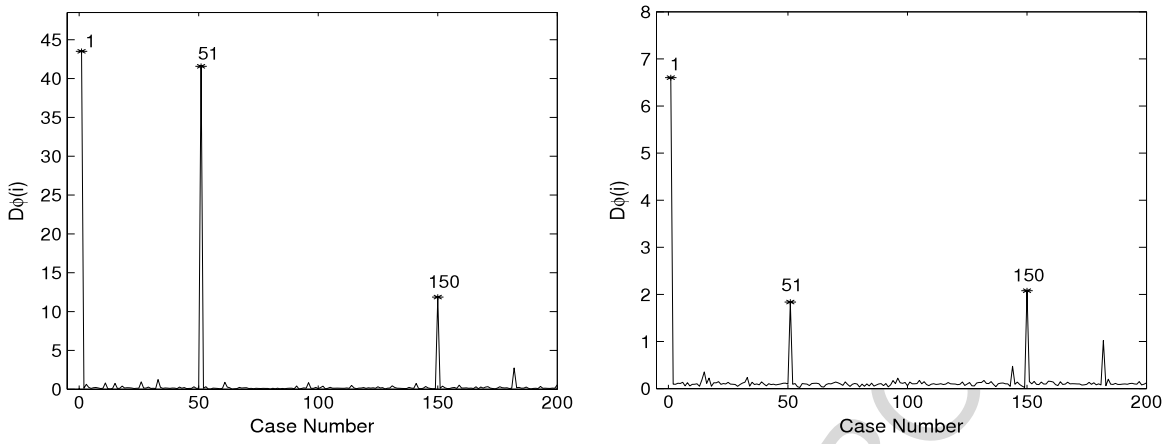
In the second simulation study, we assume that y_{ij1} was distributed as a Poisson distribution with mean parameter μ_{ij1} satisfying $\log(\mu_{ij1}) = \eta_{i1}(t_{ij1}, b_{i1}) = \beta_{10} + t_{ij1}\beta_{11} + r_i\beta_{12} + b_{i1}$, and y_{ij2} was distributed as a normal distribution $N(\mu_{ij2}, \sigma_{22}^2)$ with μ_{ij2} satisfying $\mu_{ij2} = \eta_{i2}(t_{ij2}, b_{i2}) = \beta_{20} + t_{ij2}\beta_{21} + r_i\beta_{22} + b_{i2}$. In this case, $\phi_1 = 1$ which is a fixed constant and $\phi_2 = \sigma_{22}^2$. Clearly, the above considered longitudinal model is a special case of the model defined in Eqs. (1)–(3) and includes mixed discrete and continuous responses. Also, to investigate the effectiveness of our proposed methods, we considered the following distributional assumption for $\boldsymbol{b}_i$.

Assumption 4. We specified the true distribution of $\boldsymbol{b}_i$ as $b_{i1} \overset{\text{i.i.d}}{\sim} N(0, 0.5^2)$ and $b_{i2} = b_{i2}^* - 1$ in which $b_{i2}^* \overset{\text{i.i.d}}{\sim} \Gamma(4, 4)$.

In this simulation study, data $\{(y_{ij1}, y_{ij2}) : i = 1, \ldots, 200, j = 1, \ldots, n_i\}$ were generated from the above given Poisson distribution and normal distribution with the above specified distribution of $\boldsymbol{b}_i$ (i.e., Assumption 4) and the same time points and values of r_i as presented in the first simulation study, and the following true values of parameters: $\beta_{10} = -0.5$, $\beta_{11} = 0.4$, $\beta_{12} = 0.3$, $\beta_{20} = 0.2$, $\beta_{21} = 0.3$ and $\beta_{22} = 0.4$. The survival times T_{i1}^* and T_{i2}^* were independently generated from the marginal hazard function given in Eq. (19) with the same value of $\boldsymbol{x}_i$ and the following true values of parameters: $\lambda_{10}(t) = 0.6$ and $\lambda_{20}(t) = 0.7$, $\alpha_{11} = 0.6$, $\alpha_{12} = 0.8$, $\alpha_{21} = 0.5$, $\alpha_{22} = 0.7$, $\boldsymbol{\gamma}_1 = (0.25, 0.35)^T$ and $\boldsymbol{\gamma}_2 = (0.20, 0.30)^T$. Again, for each data set generated from the above specified setting, the above proposed MCMC algorithm was used to obtain Bayesian estimates of parameters and random effects based on 50 replications and the same specification of the hyperparameters as given in the first simulation study. Results were presented in Tables 1a and 1b and Fig. S2(c). These results indicated that (1) Bayesian estimates were reasonably accurate as observed in the first simulation study; (2) the estimated means and SDs of $\hat{\boldsymbol{b}}_i$'s across individuals were close to the true means and SDs of $\boldsymbol{b}_i$'s and their bias values were less than 0.03; (3) the distributions of $\hat{b}_{i1}$ and $\hat{b}_{i2}$ were generally reasonable approximations to the shapes of the true distributions of b_{i1} and b_{i2}. These observations showed that our proposed estimation procedures can well recover the true information under the mixed distribution assumptions of longitudinal responses.

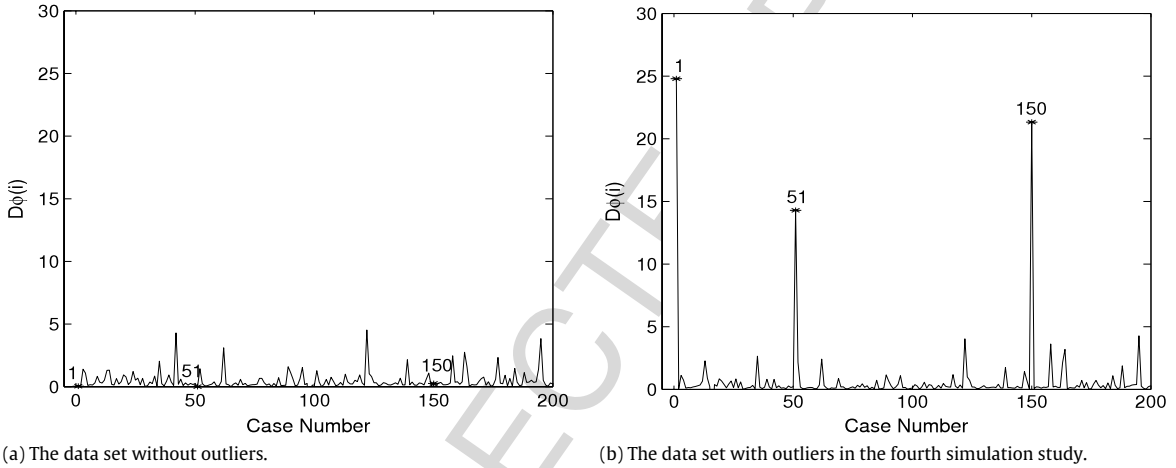
To compare the performance of our proposed MCMC method with that given in Zhu et al. (2012) under a probably misspecified distribution of $\boldsymbol{b}_i$, we computed Bayesian estimates of parameters for the data sets generated in the first and second simulation studies via the MCMC algorithm developed in Zhu et al. (2012) under multivariate normal assumption of $\boldsymbol{b}_i$: $\boldsymbol{b}_i \overset{\text{i.i.d.}}{\sim} N_2(\boldsymbol{\mu}_w, \boldsymbol{\Phi}_w)$, where $\boldsymbol{\mu}_w = (0, 0)^T$ and the prior distribution for $\boldsymbol{\Phi}_w$ was taken to be $\boldsymbol{\Phi}_w \sim IW_2(0.25\boldsymbol{I}_2, 10)$ in which IW denotes the inverted Wishart distribution. To save the space, we displayed these results in Table S1 in the

Please cite this article in press as: Tang, N.-S., et al., Semiparametric Bayesian joint models of multivariate longitudinal and survival data. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.02.015



(a) Perturbing only longitudinal data. (b) Perturbing only survival data.

Fig. 2. Index plots of $D_\phi(i)$ for (a) perturbing only longitudinal data and (b) perturbing only survival data in the third simulation study.



(a) The data set without outliers. (b) The data set with outliers in the fourth simulation study.

Fig. 3. Index plots of $D_\phi(i)$ for (a) the data set without outliers and (b) the data set with three outliers in the fourth simulation study.

Appendix B, which indicated that Zhu et al.'s method was sensitivity to distribution assumption of $\boldsymbol{b}_i$, whilst our method is robust to misspecified distribution of $\boldsymbol{b}_i$ in terms of their corresponding values of RMS.

4.2. Performance assessment of Bayesian diagnostics

In the subsection, we used the preceding proposed Bayesian case-deletion influence measures to detect the potential influential cases. In the third simulation study, the data set was generated by using the same setup as specified in the first simulation study together with distribution Assumption 1 of $\boldsymbol{b}_i$, but outliers were created by the following ways: (I) changing longitudinal data $\boldsymbol{Y}_1$ as $\boldsymbol{Y}_1 - 4$, and $\boldsymbol{Y}_i = \boldsymbol{Y}_i + 4$ for $i = 51$ and 150; (II) changing survival data T_{im} as $T_{im} + 8$ and δ_{im} as 1 for $i = 1$, 51 and 150, $m = 1, 2$. For each of the above created two sets of outliers, we computed their corresponding values of diagnostic $D_\varphi(i)$ by using the formula given in Eq. (17). Results were presented in Fig. 2. Examination of Fig. 2 showed that cases 1, 51 and 150 were detected to be influential as expected.

In the fourth simulation study, the data set was simulated by using the same setup as given in the second simulation study, but outliers were created by only changing $\boldsymbol{Y}_{i1}$ with Poisson distribution as $\boldsymbol{Y}_{i1} + 20$ for $i = 1, 51$ and 150. Similarly, diagnostic $D_\phi(i)$ was evaluated by using the formula given in Eq. (17). Fig. 3 presented index plots of $D_\phi(i)$ without and with outliers, respectively. These results indicated that cases 1, 51 and 150 were identified to be influential as expected, and there is not any case detected to be influential when there is not outliers in the considered data set.

5. Application to the IBCSG data

A data set taken from a clinical trial conducted by IBCSG was used to illustrate our proposed methodologies in this subsection. This data set has been analyzed by Zhu et al. (2012) and Chi and Ibrahim (2006) via JMLS under the normality

UNCORRECTED PROOF

Table 2a
Bayesian estimates (BE), standard deviations (SD) and 95% credible intervals (CI) for parameters in the survival model of the IBCSG data.

	DFS		OS	
	BE/SD	CI	BE/SD	CI
Appetite	0.302/0.156	(0.011, 0.607)	0.298/0.215	(−0.085, 0.700)
Perceived coping	−1.473/0.140	(−1.751, −1.189)	−2.169/0.211	(−2.620, −1.783)
Mood	−0.879/0.176	(−1.220, −0.549)	−1.192/0.251	(−1.652, −0.725)
Physical well-being	2.214/0.149	(1.903, 2.494)	3.159/0.224	(2.700, 3.579)
# Initial cycle	−0.339/0.160	(−0.644, −0.026)	−0.821/0.209	(−1.250, −0.411)
Reintroduction	−0.542/0.168	(−0.902, −0.228)	−1.341/0.215	(−1.809, −0.949)
# Initial cycle × reintroduction	0.153/0.215	(−0.296, 0.560)	1.168/0.261	(0.667, 1.683)
Age > 40	−1.770/0.181	(−2.153, −1.447)	−2.385/0.220	(−2.862, −1.996)
# Positive nodes > 4	1.342/0.131	(1.089, 1.605)	1.844/0.188	(1.499, 2.247)
ER(1 = positive)	−0.674/0.138	(−0.951, −0.399)	−1.442/0.195	(−1.816, −1.039)

assumption of random effects. Here, we reanalyzed the data set via JMLS under the unknown distribution assumption of random effects. Similar to Zhu et al. (2012), our main purpose of analyzing this data set was to use JMLS to jointly investigate the relationship between the multidimensional QOL (appetite, denoted as y_1; perceived coping, denoted as y_2; mood, denoted as y_3; and physical well-being, denoted as y_4) and the bivariate failure time variables, disease-free (DFS) with a censoring rate of 46.39% and overall survival (OS) with a censoring rate of 63.10%; and we transformed corresponding observed values of QOL to $\sqrt{100 - \text{QOL}}$. In this real example, the analyzed data set included 832 patients from Switzerland, Sweden and New Zealand/Australia with a total of 2154 QOL observations.

We fitted the IBCSG data set via the following JMLS:

$$\begin{aligned} &y_{ijk} = \eta_{ik}(t_{ijk}, b_{ik}) + \varepsilon_{ijk} = \boldsymbol{R}_{ik}(t_{ijk})\boldsymbol{\beta}_k + b_{ik} + \varepsilon_{ijk} \quad \text{and} \\ &\lambda_m(t|\boldsymbol{b}_i, \boldsymbol{x}_i) = \lambda_{m0}(t)\exp\{\alpha_{m1}\eta_{i1}(t, b_{i1}) + \cdots + \alpha_{m4}\eta_{i4}(t, b_{i4}) + \boldsymbol{x}_i^T\boldsymbol{\gamma}_m\}, \end{aligned} \tag{20}$$

for $i = 1, \ldots, 832$, $j = 1, \ldots, n_i$, $k = 1, \ldots, 4$ and $m = 1, 2$, where $y_{i1}(t_{ij1}), \ldots, y_{i4}(t_{ij4})$ were the observed values of the transformed QOLs for the ith patient at the jth time point, respectively, $\boldsymbol{R}_{ik}(t_{ijk}) = (1, z_{i1}, \ldots, z_{i6}, t_{ijk})$, $z_{i1}, \ldots, z_{i6}$ denoted the number of initial cycles, reintroduction, interaction of the number of initial cycles and reintroduction, age, residency for Switzerland and residency for Sweden, respectively, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{i6})^T$ included the number of initial cycles, reintroduction, interaction of the number of initial cycles and reintroduction, age, number of positive nodes, and ER status, $\boldsymbol{\beta}_k = (\beta_{k0}, \ldots, \beta_{k6}, \beta_{k7})^T$ and $\boldsymbol{\gamma}_m = (\gamma_{m1}, \ldots, \gamma_{m6})^T$. Further, we assumed that $\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{ij1}, \ldots, \varepsilon_{ij4})^T$ were independently and identically distributed as $N_4(\boldsymbol{0}, \boldsymbol{\Sigma})$ but $\boldsymbol{b}_i = (b_{i1}, \ldots, b_{i4})^T$ has an unknown distribution, which was specified by the preceding presented CDPMM in this data analysis. For the baseline hazard $\lambda_{m0}(t)$, we adopted the piecewise constant hazards model with $L = 250$ subintervals with equal lengths such that $\lambda_{m0}(t) = \Sigma_{l=1}^{L} h_{ml}\mathbf{1}(t \in (c_{m,l-1}, c_{ml}])$, where $c_{ml} = l(\max_i\{T_{mi}\} + 0.01)/L$ for $l = 1, \ldots, L$ and $c_{m0} = 0$.

In Bayesian analysis, we considered the following prior distributions: $\boldsymbol{\alpha}_m = (\alpha_{m1}, \ldots, \alpha_{m4})^T \sim N_2(\boldsymbol{\alpha}_m^0, 0.1\boldsymbol{I}_4)$, $\boldsymbol{\gamma}_m \sim N_6(\boldsymbol{\gamma}_m^0, 0.1\boldsymbol{I}_6)$, $\boldsymbol{\beta}_k \sim N_8(\boldsymbol{\beta}_k^0, 0.1\boldsymbol{I}_8)$, $\boldsymbol{\Sigma}^{-1} \sim \text{Wishart}_4(\boldsymbol{R}^0, \rho^0)$, and $\lambda_{ml} \sim \Gamma(\kappa_{ml0}, \kappa_{ml1})$ for $k = 1, \ldots, 4$, $m = 1, 2$ and $l = 1, \ldots, 250$, where $\boldsymbol{\alpha}_m^0, \boldsymbol{\gamma}_m^0, \boldsymbol{\beta}_k^0, \rho^0$ and $\boldsymbol{R}^0$ were the Bayesian estimates obtained from their corresponding non-informative prior distributions, and $\kappa_{ml0} = \kappa_{ml1} = 10$. For $\boldsymbol{b}_i$, we assumed that its prior distribution was specified by Eq. (9) with $G = 25$. The prior distributions for $\boldsymbol{\pi}, \boldsymbol{\zeta}, \boldsymbol{\Psi}, \varpi_j$ and $\boldsymbol{\Omega}$ given in Eqs. (10)–(13) were used to conduct Bayesian analysis with the following specified hyperparameters: $a_1 = 25$, $a_2 = 5$, $\boldsymbol{\zeta}_0 = \mathbf{1}_4$, $\boldsymbol{\Psi}_0 = \boldsymbol{I}_4$, $c_1 = 11$, $c_2 = 20$, $\omega_j^a = 6$, $\varpi_j^a = 200$ and $\varpi_j^b = 10$.

The preceding presented MCMC algorithm was used to obtain Bayesian estimates of parameters and random effects given in Eqs. (20). Similarly, EPSR values of all parameters were calculated by using three parallel sequences of observations generated from three different starting values, and their results indicated that the MCMC algorithm converged after 20 000 iterations because their EPSR values were less than 1.2 after about 20 000 iterations. Hence, $\mathfrak{L} = 10\,000$ observations collected after 20 000 iterations were used to calculate Bayesian estimates of parameters and random effects. To compare our proposed MCMC algorithm with the existing methods, for example, Zhu et al.'s (2012) approach with traditional normality assumption for random effects, we carried out our proposed methods to calculate the corresponding Bayesian estimates of parameters. Results were reported in Tables 2a and 2b and Figs. 4 and 5, which indicated that (i) Bayesian estimates obtained by using our method have smaller SDs than those from Zhu et al.'s (2012) method in most cases, and have shorter credible intervals than those from Chi and Ibrahim's (2006) method; (ii) traditional normality assumption for random effects is inappropriate because there are bimodal distributions for b_{i2}, b_{i3} and b_{i4}, and there is a skewed distribution for b_{i1}; (iii) the estimated baseline hazard functions $\hat{\lambda}_{10}(t)$ and $\hat{\lambda}_{20}(t)$ slightly wave around 1, so we have reason to believe that the true baseline hazard functions $\lambda_{m0}(t) \cong 1$ for $m = 1$ and 2.

Also, we calculated Bayesian case-deletion influence diagnostic $D_\varphi(i)$ based on Eq. (16) via our proposed MCMC algorithm. Results were presented in Fig. 6, which indicated that individuals 21, 56, 103, 130, 438 and 780 were detected to be influential, whilst individuals 103, 130 and 780 were also identified to be influential by Zhu et al. (2012). We deleted the six influential individuals and recalculated Bayesian estimates of parameters in JMLS for the IBCSG data set, which were reported in Tables S2(a) and (b) in the Appendix B. Examination of Tables S2(a) and (b) indicated that many estimates of parameters without the six influential individuals have smaller SD values and shorter CIs, hence the above detected six

UNCORRECTED PROOF

Table 2b
Bayesian estimates (BE) and standard deviations (SD) and 95% credible intervals (CI) for parameters in the longitudinal model of the IBCSG data.

Par.	Appetite	Perceived coping	Mood	Physical well-being
	BE/SD CI	BE/SD CI	BE/SD CI	BE/SD CI
Intercept	3.502/0.136 (3.220, 3.773)	5.733/0.119 (5.511, 5.962)	4.521/0.140 (4.237, 4.803)	3.735/0.121 (3.508, 3.953)
# Initial cycle	0.114/0.111 (−0.124, 0.335)	0.142/0.118 (−0.111, 0.369)	0.436/0.100 (0.232, 0.621)	0.418/0.094 (0.242, 0.597)
Reintroduction	−0.055/0.126 (−0.302, 0.195)	−0.068/0.123 (−0.328, 0.145)	0.095/0.136 (−0.163, 0.362)	0.112/0.118 (−0.098, 0.320)
# Initial cycle × reintroduction	−0.052/0.167 (−0.377, 0.261)	−0.063/0.161 (−0.378, 0.235)	−0.322/0.164 (−0.644,−0.011)	−0.279/0.157 (−0.613, 0.020)
Age > 40	0.395/0.111 (0.172, 0.612)	0.074/0.124 (−0.152, 0.326)	0.428/0.111 (0.216, 0.640)	0.662/0.098 (0.449, 0.843)
Residency: Swiss	0.118/0.118 (−0.120, 0.351)	0.208/0.100 (0.021, 0.389)	0.157/0.115 (−0.091, 0.397)	0.099/0.091 (−0.073, 0.261)
Residency: Sweden	0.203/0.119 (−0.020, 0.445)	0.492/0.113 (0.287, 0.728)	0.957/0.124 (0.696, 1.181)	0.664/0.093 (0.485, 0.843)
Time (in years)	−0.628/0.055 (−0.730,−0.518)	−0.779/0.045 (−0.857, −0.668)	−0.629/0.034 (−0.709, −0.557)	−0.621/0.033 (−0.675, −0.540)
Covariate matrix Σ	3.272/0.117 (3.051, 3.509)	0.663/0.096 (0.482, 0.857) 3.316/0.129 (3.066, 3.574)	1.426/0.094 (1.249, 1.618) 1.290/0.101 (1.095, 1.490) 3.078/0.114 (2.859, 3.307)	1.627/0.100 (1.439, 1.829) 0.805/0.091 (0.632, 0.988) 1.790/0.099 (1.597, 1.984) 3.603/0.126 (3.363, 3.853)

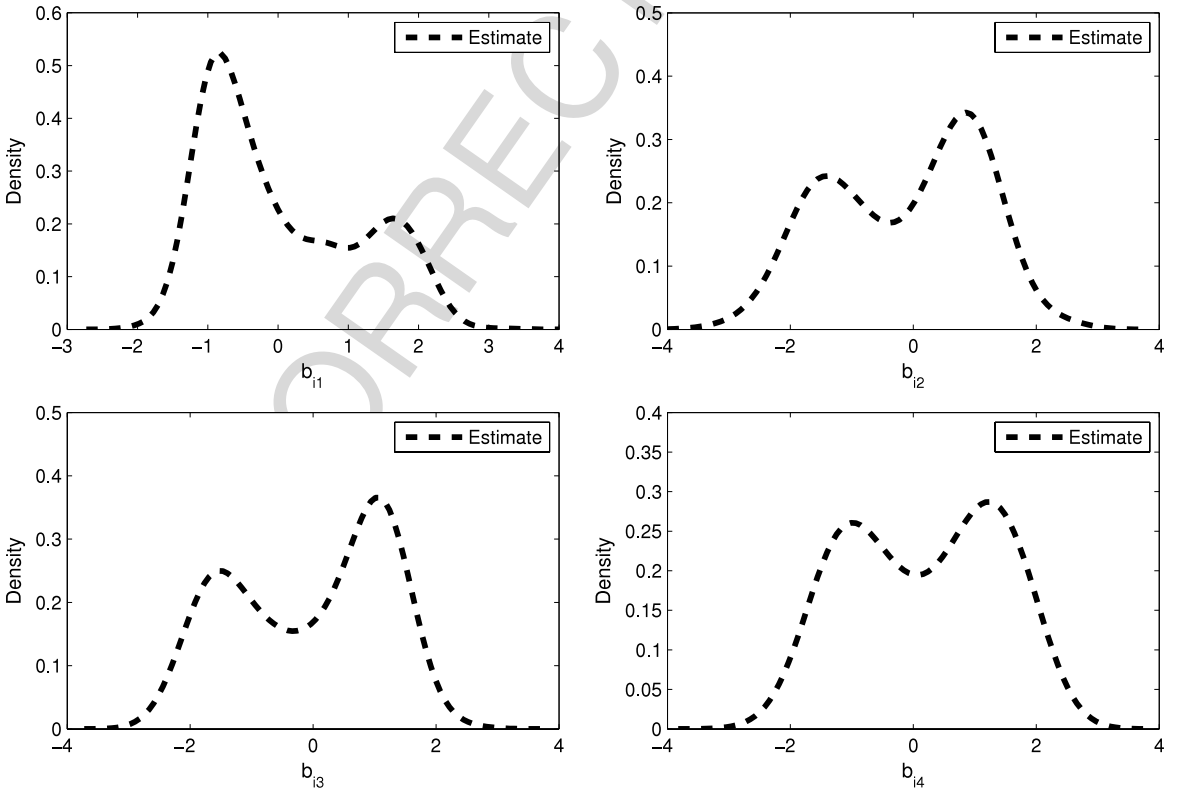


Fig. 4. Estimated densities of random effects $b_{i1}, \ldots, b_{i4}$ for the IBCSG data.

Please cite this article in press as: Tang, N.-S., et al., Semiparametric Bayesian joint models of multivariate longitudinal and survival data. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.02.015

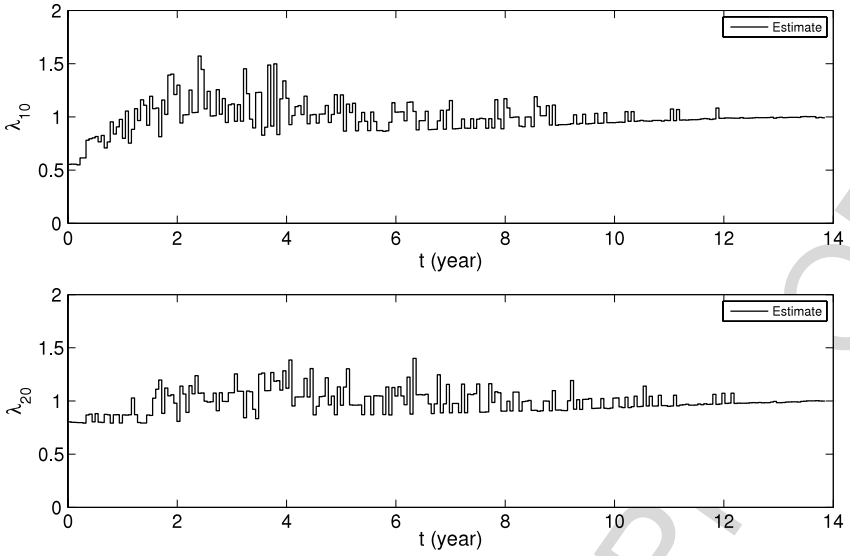


Fig. 5. Estimates of baseline hazard functions $\lambda_{10}(t)$ and $\lambda_{20}(t)$ with piecewise constant hazard model for the IBCSG data.

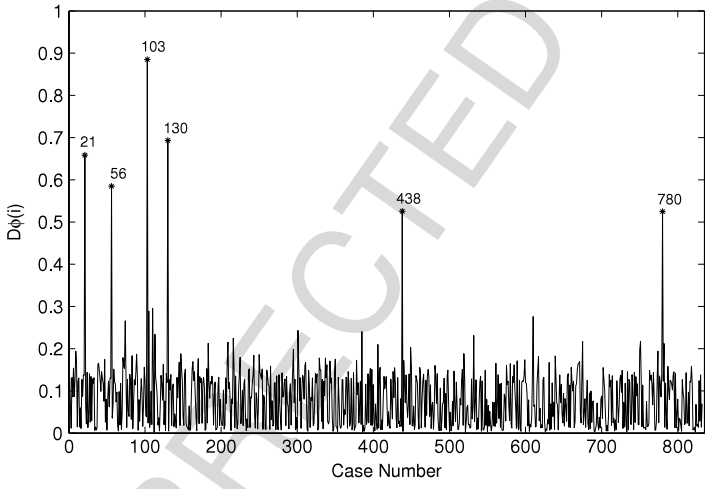


Fig. 6. $D_{\phi}(i)$ for the IBCSG data.

influential cases have a significant influence on parameter estimations. Results without individuals 21, 56, 103, 130, 438 and 780 and those without individuals 103, 130 and 780 were given in Tables S2(a) and (b), which showed that the former had smaller SD and shorter CI than the latter in most cases, thus it is reasonable to regard individuals 21, 56, and 438 as influential individuals.

6. Conclusion

This paper discusses Bayesian estimates of parameters and Bayesian case-deletion diagnostics for joint models for longitudinal and survival data under the unknown distribution assumption of random effects. To conduct Bayesian analysis for joint models for longitudinal and survival data, we used the truncated CDPMM to approximate distribution of random effects and used the piecewise constant baseline hazard model to approximate the baseline hazard function. A semiparametric Bayesian approach is developed to simultaneously obtain Bayesian estimates of parameters and random effects and their standard errors and other statistical results such as Bayesian case-deletion influence diagnostic by incorporating the MH algorithm within the Gibbs sampler framework. Four simulation studies and a real example from IBCSG are used to illustrate the proposed methodologies. Empirical results demonstrate that (1) the proposed semiparametric Bayesian method provides quite accurate estimates of parameters, although our proposed Bayesian method takes more computing time for running the same iterations than the parametric Bayesian method with the normality assumption for random effects; (2) the truncated CDPMM can well capture characteristics of distributions for random effects; and (3) the developed Bayesian case-deletion method can detect the potential influential observations well.

Please cite this article in press as: Tang, N.-S., et al., Semiparametric Bayesian joint models of multivariate longitudinal and survival data. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.02.015

UNCORRECTED PROOF

Uncited references

Q7

Kleinman and Ibrahim, 1998a, Kleinman and Ibrahim, 1998b, Lee et al., 2008 and McAuliffe et al., 2006.

Acknowledgments

The research was supported by grants from the National Science Fund for Distinguished Young Scholars of China (11225103), and the Research Fund for the Doctoral Program of Higher Education of China (20115301110004).

Appendix A. Conditional distributions used in the Gibbs sampling procedures

To obtain Bayesian estimates of parameters and random effects in our considered JMLSs, the Gibbs sampler is employed to draw a sequence of observations from the joint posterior distribution $p(\boldsymbol{\zeta}, \boldsymbol{\Psi}, \boldsymbol{\varpi}, \boldsymbol{\pi}, \boldsymbol{\mu}^*, \boldsymbol{\Omega}, \boldsymbol{L}, \tau, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{b}, \boldsymbol{h}|\boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})$. The Gibbs sampler is conducted by iteratively sampling observations from the following conditional distributions $p(\boldsymbol{\zeta}|\boldsymbol{\mu}^*, \boldsymbol{\Psi}, \boldsymbol{b})$, $p(\boldsymbol{\Psi}|\boldsymbol{\mu}^*, \boldsymbol{\zeta})$, $p(\boldsymbol{\mu}^*|\boldsymbol{\zeta}, \boldsymbol{\Psi}, \boldsymbol{\Omega}, \boldsymbol{L}, \boldsymbol{b})$, $p(\boldsymbol{\varpi}|\boldsymbol{\Omega})$, $p(\boldsymbol{\Omega}|\boldsymbol{\mu}, \boldsymbol{\varpi}, \boldsymbol{L}, \tau)$, $p(\tau|\boldsymbol{\pi})$, $p(\boldsymbol{\pi}|\boldsymbol{L}, \tau)$, $p(\boldsymbol{L}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{b})$, $p(\boldsymbol{b}|\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{L}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_T, \boldsymbol{h}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})$, $p(h_{ml}|\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha}_m, \boldsymbol{\gamma}_m, \boldsymbol{b}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})$, $p(\boldsymbol{\beta}_k|\boldsymbol{\phi}_k, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{b}, \boldsymbol{h}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})$, $p(\phi_k|\boldsymbol{\beta}_k, \boldsymbol{b}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W})$, $p(\boldsymbol{\alpha}_m|\boldsymbol{\beta}, \boldsymbol{\gamma}_m, \boldsymbol{b}, \boldsymbol{h}_m, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})$ and $p(\boldsymbol{\gamma}_m|\boldsymbol{\beta}, \boldsymbol{\alpha}_m, \boldsymbol{b}, \boldsymbol{h}_m, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})$. The conditional distributions required in conducting Gibbs sampler are presented as follows.

Steps (a)–(i): *conditional distributions related to sample random-effects*

To obtain the random observation of random effects $\boldsymbol{b}_i$ in terms of the latent variable L_i for $i = 1, \ldots, n$, we first simulate $\boldsymbol{\mu}^* = (\boldsymbol{\mu}_1^*, \ldots, \boldsymbol{\mu}_G^*)$ and $\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_G$ from their posterior distributions, then $\boldsymbol{b}_i \sim N_q(\boldsymbol{\mu}_{L_i}, \boldsymbol{\Omega}_{L_i})$ with $\boldsymbol{\mu}_g = \boldsymbol{\mu}_g^* - \Sigma_{g=1}^G \pi_g \boldsymbol{\mu}_g^*$. Because the posterior distribution $p(\boldsymbol{\zeta}, \boldsymbol{\Psi}, \varpi, \boldsymbol{\pi}, \boldsymbol{\mu}^*, \boldsymbol{\Omega}, \boldsymbol{L}, \tau, \boldsymbol{b}|\boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{h}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})$ is not a standard and complicated distribution, it is quite difficult to simulate observations from this posterior distribution. Hence, the blocked Gibbs sampler is adopted to sample observations from the following nine conditional distributions: $p(\boldsymbol{\zeta}|\boldsymbol{\mu}^*, \boldsymbol{\Psi}, \boldsymbol{b})$, $p(\boldsymbol{\Psi}|\boldsymbol{\mu}^*, \boldsymbol{\zeta})$, $p(\boldsymbol{\mu}^*|\boldsymbol{\zeta}, \boldsymbol{\Psi}, \boldsymbol{\Omega}, \boldsymbol{L}, \boldsymbol{b})$, $p(\varpi_j|\boldsymbol{\Omega})$, $p(\boldsymbol{\Omega}|\boldsymbol{\mu}, \boldsymbol{\varpi}, \boldsymbol{L}, \tau)$, $p(\tau|\boldsymbol{\pi})$, $p(\boldsymbol{\pi}|\boldsymbol{L}, \tau)$, $p(\boldsymbol{L}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{b})$ and $p(\boldsymbol{b}|\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{L}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_T, \boldsymbol{h}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})$. These conditional distributions are summarized as follows.

Step (a): conditional distribution of $\boldsymbol{\zeta}$ given $(\boldsymbol{\mu}^*, \boldsymbol{\Psi}, \boldsymbol{b})$ is given

$$\boldsymbol{\zeta}|\boldsymbol{\mu}^*, \boldsymbol{\Psi}, \boldsymbol{b} \sim N_q(\mathbb{A}, \mathbb{B}) \tag{A.1}$$

where $\mathbb{B} = (G\boldsymbol{\Psi}^{-1} + (\boldsymbol{\Psi}^0)^{-1})^{-1}$ and $\mathbb{A} = \mathbb{B}((\boldsymbol{\Psi}^0)^{-1}\boldsymbol{\zeta}^0 + \boldsymbol{\Psi}^{-1}\sum_{g=1}^G \boldsymbol{\mu}_g^*)$.

Step (b): for $j = 1, \ldots, q$, the diagonal elements of $\boldsymbol{\Psi}$ are conditionally distributed as

$$\psi_j^{-1}|\boldsymbol{\mu}^*, \boldsymbol{\zeta} \sim \Gamma\left(c_1 + \frac{G}{2}, c_2 + \frac{1}{2}\sum_{g=1}^G (\mu_{gj}^* - \zeta_j)^2\right), \tag{A.2}$$

where μ_{gj}^* is the jth element of $\boldsymbol{\mu}_g^*$ and ζ_j is the jth element of $\boldsymbol{\zeta}$.

Step (c): for $j = 1, \ldots, q$, $\varpi_j|\boldsymbol{\Omega}$ is conditionally distributed as

$$\varpi_j|\boldsymbol{\Omega} \sim \Gamma\left(\varpi_j^a, \varpi_j^b + \sum_{g=1}^G \omega_{gj}^{-1}\right), \tag{A.3}$$

where ω_{gj} is the jth diagonal element of $\boldsymbol{\Omega}_g$.

Step (d): following Ishwaran and Zarepour (2000), the conditional distribution of $\tau|\boldsymbol{\pi}$ can be expressed as

$$(\tau|\boldsymbol{\pi}) \sim \Gamma\left(a_1 + G - 1, a_2 - \sum_{g=1}^{G-1} \log(1 - \nu_g^*)\right), \tag{A.4}$$

where ν_g^* is a random weight sampled from the beta distribution and it is sampled with step (e).

Step (e): it is easily obtained that the conditional distribution of $\boldsymbol{\pi}|\boldsymbol{L}, \tau$ is distributed as the following generalized Dirichlet distribution:

$$\boldsymbol{\pi}|\boldsymbol{L}, \tau \sim \mathrm{Dir}(a_1^*, b_1^*, \ldots, a_{G-1}^*, b_{G-1}^*), \tag{A.5}$$

where $a_g^* = 1 + d_g$, $b_g^* = \tau + \sum_{\iota=g+1}^G d_\iota$ for $g = 1, \ldots, G-1$, and d_g is the number of L_is (and thus individuals) whose value equals to g. Simulating observation from the conditional distribution $\boldsymbol{\pi}|\boldsymbol{L}, \tau$ can be conducted as follows. First, ν_g^* is independently generated from a Beta distribution (a_g^*, b_g^*). Then, $\pi_1, \ldots, \pi_G$ are obtained from the following formulae:

$$\pi_1 = \nu_1^*, \qquad \pi_G = 1 - \sum_{g=1}^{G-1} \pi_g, \quad \text{and} \quad \pi_g = \prod_{\iota=1}^{g-1}(1 - \nu_\iota^*)\nu_g^*, \quad \text{for } g \neq 1 \text{ or } G. \tag{A.6}$$

Step (f): conditional distribution of $\boldsymbol{\mu}^*|\boldsymbol{\zeta}, \boldsymbol{\Psi}, \boldsymbol{\Omega}, \boldsymbol{L}, \boldsymbol{b}$.

UNCORRECTED PROOF

Let $L_1^*, \ldots, L_d^*$ be the d unique values of $\{L_1, \ldots, L_n\}$ (i.e., unique number of "clusters"), for $g = 1, \ldots, G$, $\boldsymbol{\mu}_g^*$ is conditionally distributed as follows:

$$\boldsymbol{\mu}_g^* | \boldsymbol{\zeta}, \boldsymbol{\Psi} \sim N_q(\boldsymbol{\zeta}, \boldsymbol{\Psi}) \quad \text{for } g \notin \{L_1^*, \ldots, L_d^*\},$$

$$\boldsymbol{\mu}_g^* | \boldsymbol{\zeta}, \boldsymbol{\Psi}, \boldsymbol{\Omega}, \boldsymbol{L}, \boldsymbol{b} \sim N_q(\mathbb{E}_g, \mathbb{F}_g) \quad \text{for } g \in \{L_1^*, \ldots, L_d^*\}, \tag{A.7}$$

where $\mathbb{F}_g = (\boldsymbol{\Psi}^{-1} + \Sigma_{\{i:L_i=g\}} \boldsymbol{\Omega}_i^{-1})^{-1}$ and $\mathbb{E}_g = \mathbb{F}_g(\boldsymbol{\Psi}^{-1}\boldsymbol{\zeta} + \Sigma_{\{i:L_i=g\}} \boldsymbol{\Omega}_i^{-1} \boldsymbol{b}_i)$ for $g \in \{L_1^*, \ldots, L_d^*\}$. Given $\boldsymbol{\mu}_g^*$, $\boldsymbol{\mu}_g = \boldsymbol{\mu}_g^* - \Sigma_{g=1}^G \pi_g \boldsymbol{\mu}_g^*$, $\boldsymbol{\mu}^* = \{\boldsymbol{\mu}_1^*, \ldots, \boldsymbol{\mu}_G^*\}$ and $\boldsymbol{\mu} = \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_G\}$.

Step (g): conditional distribution of $\boldsymbol{\Omega} | \boldsymbol{\mu}, \varpi, \boldsymbol{L}, \tau$.

Similar to the notation of step (f), given g, for $j = 1, \ldots, q$, the jth diagonal element of $\boldsymbol{\Omega}_g$ is conditionally distributed as

$$\omega_{g_j} \sim \Gamma(\omega_j^a, \varpi_j) \quad \text{for } g \notin \{L_1^*, \ldots, L_d^*\},$$

$$\omega_{g_j} \sim \Gamma\left(\frac{d_g}{2} + \omega_j^a, \varpi_j + \sum_{\{i:L_i=g\}} \frac{1}{2}(b_{i_j} - \mu_{g_j})^2\right) \quad \text{for } g \in \{L_1^*, \ldots, L_d^*\}, \tag{A.8}$$

where b_{i_j} is the jth element of $\boldsymbol{b}_i$ and μ_{g_j} is the jth element of $\boldsymbol{\mu}_g$. Given ω_{g_j}, $\boldsymbol{\Omega}_g = \text{diag}(\omega_{g_1}, \ldots, \omega_{g_q})$ and $\boldsymbol{\Omega} = \{\Omega_1, \ldots, \Omega_G\}$.

Step (h): the conditional distribution of $L_i | \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{b}$ is given by

$$L_i | \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{b} \overset{\text{i.i.d}}{\sim} \text{Multinomial}(\pi_{ig}^*, g = 1, \ldots, G), \tag{A.9}$$

where π_{ig}^* is proportional to $(\pi_g p(\boldsymbol{b}_i | \boldsymbol{\mu}_g, \boldsymbol{\Omega}_g))$ with $\boldsymbol{b}_i | \boldsymbol{\mu}_g, \boldsymbol{\Omega}_g \sim N_q(\boldsymbol{\mu}_g, \boldsymbol{\Omega}_g)$, and $\pi_g (g = 1, \ldots, G)$ are sampled from step (e).

Step (i): the conditional distribution for $\boldsymbol{b}$.

The conditional distribution of $p(\boldsymbol{b}_i | \boldsymbol{\mu}, \boldsymbol{\Omega}, L_i, \boldsymbol{\theta}_y, \boldsymbol{\theta}_T, \boldsymbol{h}, \boldsymbol{Y}_i, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}_i, \boldsymbol{\delta}_i)$ is non-standard and cannot be derived directly via Gibbs sampling for $i = 1, \ldots, n$. Specially,

$$p(\boldsymbol{b}_i | \boldsymbol{\mu}, \boldsymbol{\Omega}, L_i, \boldsymbol{\theta}_y, \boldsymbol{\theta}_T, \boldsymbol{h}, \boldsymbol{Y}_i, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}_i, \boldsymbol{\delta}_i) \propto p(\boldsymbol{b}_i | \boldsymbol{\mu}_{L_i}, \boldsymbol{\Omega}_{L_i}) p(\boldsymbol{Y}_i | \boldsymbol{b}_i, \boldsymbol{R}, \boldsymbol{W}; \boldsymbol{\theta}_y) p(\boldsymbol{T}_i, \boldsymbol{\delta}_i | \boldsymbol{b}_i, \boldsymbol{h}, \boldsymbol{X}; \boldsymbol{\theta}_T). \tag{A.10}$$

The Metropolis–Hastings algorithm used to sample observation $\boldsymbol{b}_i$ is implemented as follows. At the ℓth iteration with a current value $\boldsymbol{b}_i^{(\ell)}$, a new candidate $\boldsymbol{b}_i$ is drawn from the normal distribution $N_q(\boldsymbol{b}_i^{(\ell)}, \sigma_b^2 \boldsymbol{\Theta}_{b_i})$, where $\boldsymbol{\Theta}_{b_i} = (\boldsymbol{\Omega}_{L_i}^{-1} + \boldsymbol{\Xi}_i)^{-1}$ and $\boldsymbol{\Xi}_i = -\partial^2(\ln(p(\boldsymbol{Y}_i | \boldsymbol{b}_i; \boldsymbol{\theta}_y) p(\boldsymbol{T}_i, \boldsymbol{\delta}_i | \boldsymbol{b}_i, \boldsymbol{h}; \boldsymbol{\theta})))/\partial \boldsymbol{b}_i \partial \boldsymbol{b}_i^T |_{\boldsymbol{b}_i = \boldsymbol{b}_i^{(\ell)}}$. The new $\boldsymbol{b}_i$ is accepted with probability Q8

$$\min\left\{1, \frac{p(\boldsymbol{b}_i | \boldsymbol{\mu}_{L_i}, \boldsymbol{\Omega}_{L_i}) p(\boldsymbol{Y}_i | \boldsymbol{b}_i, \boldsymbol{R}, \boldsymbol{W}; \boldsymbol{\theta}_y) p(\boldsymbol{T}_i, \boldsymbol{\delta}_i | \boldsymbol{b}_i, \boldsymbol{h}, \boldsymbol{X}; \boldsymbol{\theta}_T)}{p(\boldsymbol{b}_i^{(\ell)} | \boldsymbol{\mu}_{L_i}, \boldsymbol{\Omega}_{L_i}) p(\boldsymbol{Y}_i | \boldsymbol{b}_i^{(\ell)}, \boldsymbol{R}, \boldsymbol{W}; \boldsymbol{\theta}_y) p(\boldsymbol{T}_i, \boldsymbol{\delta}_i | \boldsymbol{b}_i^{(\ell)}, \boldsymbol{h}, \boldsymbol{X}; \boldsymbol{\theta}_T)}\right\}.$$

The variance σ_b^2 can be chosen such that the average acceptance rate is approximately 0.25 or more.

Step (j): the conditional distribution of parameter ϕ_k $(k = 1, \ldots, K)$ or $\boldsymbol{\Sigma}$.

The conditional distribution $p(\phi_k | \boldsymbol{\beta}_k, \boldsymbol{b}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W})$ is proportional to

$$\phi_k^{\phi_k^a - 1} \exp\left\{-\phi_k^b \phi_k^{-1} + \phi_k^{-1} \sum_{i=1}^n \sum_{j=1}^{n_{ik}} (y_{ijk}\vartheta_{ijk} - s(\vartheta_{ijk})) + \sum_{i=1}^n \sum_{j=1}^{n_{ik}} c(y_{ijk}, \phi_k)\right\}. \tag{A.11}$$

Similar to Liu (2001), we adopt the following Metropolized independence sampler (MIS) algorithm to sample observations from the posterior distribution of $(\phi_k | \boldsymbol{\beta}_k, \boldsymbol{b}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W})$ given in (A.11). At the ℓth iteration with a current value $\phi_k^{(\ell)}$, a new candidate ϕ_k is generated from $h(\phi_k) \overset{\text{D}}{=} N(\phi_k^{(\ell)}, \sigma_{\phi_k}^2)\mathbf{1}(0, +\infty)$, and it is accepted with probability

$$\min\left\{1, \frac{p(\phi_k | \boldsymbol{\beta}_k, \boldsymbol{b}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}) h(\phi_k^{(\ell)})}{p(\phi_k^{(\ell)} | \boldsymbol{\beta}_k, \boldsymbol{b}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}) h(\phi_k)}\right\}.$$

The variance $\sigma_{\phi_k}^2$ can be chosen such that the average acceptance rate is approximately 0.25 or more. Specifically, in real example, when the prior distribution for $\boldsymbol{\Sigma}^{-1}$ is $\text{Wishart}_4(\rho^0, \boldsymbol{R}^0)$, we have

$$\boldsymbol{\Sigma}^{-1} | \boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W} \sim \text{Wishart}_4\left(\rho^0 + N, \left((\boldsymbol{R}^0)^{-1} + \sum_{i=1}^n \sum_{j=1}^{n_i} [\boldsymbol{y}_i(t_{ij}) - \eta_i(t_{ij}, \boldsymbol{b}_i)]^{\otimes 2}\right)^{-1}\right), \tag{A.12}$$

where $N = \sum_{i=1}^n n_i$.

Step (k): the Conditional distribution for h_{ml}.

It can be shown that

$$h_{ml} | \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta} \sim \Gamma\left(\kappa_{ml0} + d_{ml}, \kappa_{ml1} + \sum_{i=1}^n B_{mil}\right), \tag{A.13}$$

where $d_{ml} = \sum_{i=1}^n \delta_{im} \mathbf{1}\{T_{im} \in (c_{m,l-1}, c_{ml}]\}$, and B_{mil} is given in Eq. (7).

UNCORRECTED PROOF

Step (l): the Conditional distribution for $\boldsymbol{\alpha}_m$.
The conditional distribution $p(\boldsymbol{\alpha}_m|\boldsymbol{\beta}, \boldsymbol{\gamma}_m, \boldsymbol{b}, \boldsymbol{h}_m, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})$ is proportional to

$$\exp\left[\sum_{i=1}^{n}\left(\delta_{im}\boldsymbol{\alpha}_m^T\boldsymbol{\eta}_i(T_{im}, \boldsymbol{b}_i) - \sum_{l=1}^{L_{im}} h_{ml}\boldsymbol{B}_{mil}\right) - \frac{1}{2}(\boldsymbol{\alpha}_m - \boldsymbol{\alpha}_m^0)(\boldsymbol{H}_\alpha^0)^{-1}(\boldsymbol{\alpha}_m - \boldsymbol{\alpha}_m^0)^T\right], \tag{A.14}$$

where $\boldsymbol{\eta}_i(T_{im}, \boldsymbol{b}_i) = (\eta_{i1}(T_{im}, \boldsymbol{b}_{i1}), \ldots, \eta_{iK}(T_{im}, \boldsymbol{b}_{iK}))^T$. Similar to Liu (2001), we adopt the Metropolis–Hastings algorithm to sample observations from the posterior distribution of $\boldsymbol{\alpha}_m$ given in (A.14). At the ℓth iteration with a current value $\boldsymbol{\alpha}_m^{(\ell)}$, a new candidate $\boldsymbol{\alpha}_m$ is generated from the normal distribution $N_K(\boldsymbol{\alpha}_m^{(\ell)}, \sigma_\alpha^2\boldsymbol{\Omega}_\alpha)$, where $\boldsymbol{\Omega}_\alpha = ((\boldsymbol{H}_\alpha^0)^{-1} + \sum_{i=1}^n\sum_{l=1}^{L_{im}}\mathbb{B}_{mil})^{-1}$ and $\mathbb{B}_{mil} = \partial^2 B_{mil}/\partial\boldsymbol{\alpha}_m\partial\boldsymbol{\alpha}_m^T|_{\boldsymbol{\alpha}_m=\boldsymbol{\alpha}_m^{(\ell)}}$. The new candidate $\boldsymbol{\alpha}_m$ is accepted with probability

$$\min\left\{1, \frac{p(\boldsymbol{\alpha}_m|\boldsymbol{\beta}, \boldsymbol{\gamma}_m, \boldsymbol{b}, \boldsymbol{h}_m, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})}{p(\boldsymbol{\alpha}_m^{(\ell)}|\boldsymbol{\beta}, \boldsymbol{\gamma}_m, \boldsymbol{b}, \boldsymbol{h}_m, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})}\right\}.$$

Step (m): the Conditional distribution for $\boldsymbol{\gamma}_m$.
The conditional distribution $p(\boldsymbol{\gamma}_m|\boldsymbol{\beta}, \boldsymbol{\alpha}_m, \boldsymbol{b}, \boldsymbol{h}_m, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})$ is proportional to

$$\exp\left[\sum_{i=1}^{n}\left(\delta_{im}x_i^T\boldsymbol{\gamma}_m - \sum_{l=1}^{L_{im}} h_{ml}\boldsymbol{B}_{mil}\right) - \frac{1}{2}(\boldsymbol{\gamma}_m - \boldsymbol{\gamma}_m^0)(\boldsymbol{H}_\gamma^0)^{-1}(\boldsymbol{\gamma}_m - \boldsymbol{\gamma}_m^0)^T\right]. \tag{A.15}$$

Also, we adopt the Metropolis–Hastings algorithm to sample observations from the posterior distribution of $\boldsymbol{\gamma}_m$ given in (A.15). At the ℓth iteration with a current value $\boldsymbol{\gamma}_m^{(\ell)}$, a new candidate $\boldsymbol{\gamma}_m$ is generated from the normal distribution $N_q(\boldsymbol{\gamma}_m^{(\ell)}, \sigma_\gamma^2\boldsymbol{\Omega}_\gamma)$, where $\boldsymbol{\Omega}_\gamma = ((\boldsymbol{H}_\gamma^0)^{-1} + \sum_{i=1}^n\sum_{l=1}^{L_{im}}\mathbb{C}_{mil})^{-1}$ and $\mathbb{C}_{mil} = \partial^2 B_{mil}/\partial\boldsymbol{\gamma}_m\partial\boldsymbol{\gamma}_m^T|_{\boldsymbol{\gamma}_m=\boldsymbol{\gamma}_m^{(\ell)}}$. The new candidate $\boldsymbol{\gamma}_m$ is accepted with probability

$$\min\left\{1, \frac{p(\boldsymbol{\gamma}_m|\boldsymbol{\beta}, \boldsymbol{\alpha}_m, \boldsymbol{b}, \boldsymbol{h}_m, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})}{p(\boldsymbol{\gamma}_m^{(\ell)}|\boldsymbol{\beta}, \boldsymbol{\alpha}_m, \boldsymbol{b}, \boldsymbol{h}_m, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})}\right\}.$$

Step (n): the Conditional distribution of $\boldsymbol{\beta}_k$.
The conditional distribution $p(\boldsymbol{\beta}_k|\phi_k, \alpha, \boldsymbol{\gamma}, \boldsymbol{b}, \boldsymbol{h}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})$ is proportional to

$$\exp\left[\phi_k^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_{ik}}(y_{ijk}\vartheta_{ijk} - s(\vartheta_{ijk})) + \sum_{i=1}^{n}\sum_{j=1}^{n_{ik}}c(y_{ijk}, \phi_k) - \sum_{i=1}^{n}\sum_{m=1}^{M}\sum_{l=1}^{L_{im}}h_{ml}B_{mil}\right.$$
$$\left. + \sum_{i=1}^{n}\sum_{m=1}^{M}\delta_{im}\alpha_{mk}\eta_{ik}(T_{im}, b_{ik}) - \frac{1}{2}(\boldsymbol{\beta}_k - \boldsymbol{\beta}_k^0)(\boldsymbol{H}_\beta^0)^{-1}(\boldsymbol{\beta}_k - \boldsymbol{\beta}_k^0)^T\right], \tag{A.16}$$

where $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M\}$ and $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_M\}$. Similarly, the Metropolis–Hastings algorithm for simulating observations from the conditional distribution $p(\boldsymbol{\beta}_k|\phi_k, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{b}, \boldsymbol{h}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{T}, \boldsymbol{\Delta}, \boldsymbol{X})$ given in (A.16) is implemented as follows. Given the current value $\boldsymbol{\beta}_k^{(\ell)}$, a new candidate $\boldsymbol{\beta}_k$ is generated from $N_p(\boldsymbol{\beta}_k^{(\ell)}, \sigma_k^2\boldsymbol{\Omega}_k)$ and is accepted with probability

$$\min\left\{1, \frac{p(\boldsymbol{\beta}_k|\phi_k, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{b}, \boldsymbol{h}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})}{p(\boldsymbol{\beta}_k^{(\ell)}|\phi_k, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{b}, \boldsymbol{h}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{\Delta})}\right\},$$

where $\boldsymbol{\Omega}_k = ((\boldsymbol{H}_\beta^0)^{-1} + \Sigma_{i=1}^n\Sigma_{j=1}^{n_{ik}}\partial^2 p(y_{ijk}|\boldsymbol{b}_{ik}, \phi_k)/\partial\boldsymbol{\beta}_k\partial\boldsymbol{\beta}_k^T|_{\boldsymbol{\beta}_k=\boldsymbol{\beta}_k^{(\ell)}} + \Sigma_{i=1}^n\Sigma_{m=1}^M\Sigma_{l=1}^{L_{im}}h_{ml}\mathbb{D}_{mil})^{-1}$ with $\mathbb{D}_{mil} = \partial^2 B_{mil}/\partial\boldsymbol{\beta}_k\partial\boldsymbol{\beta}_k^T|_{\boldsymbol{\beta}_k=\boldsymbol{\beta}_k^{(\ell)}}$. Particularly, in real example, (A.16) becomes

$$\exp\left[\sum_{i=1}^{n}\sum_{j=1}^{n_i} -\frac{1}{2}\left(y_i(t_{ij}) - \eta_i(t_{ij}, \boldsymbol{b}_i)\right)^T\boldsymbol{\Sigma}^{-1}\left(y_i(t_{ij}) - \eta_i(t_{ij}, \boldsymbol{b}_i)\right) - \sum_{i=1}^{n}\sum_{m=1}^{M}\sum_{l=1}^{L_{im}}h_{ml}B_{mil}\right.$$
$$\left. + \sum_{i=1}^{n}\sum_{m=1}^{M}\delta_{im}\alpha_{mk}\eta_{ik}(T_{im}, b_{ik}) - \frac{1}{2}(\boldsymbol{\beta}_k - \boldsymbol{\beta}_k^0)(\boldsymbol{H}_\beta^0)^{-1}(\boldsymbol{\beta}_k - \boldsymbol{\beta}_k^0)^T\right],$$

and $\boldsymbol{\Omega}_k = ((\boldsymbol{H}_\beta^0)^{-1} + \Sigma_{i=1}^n\Sigma_{j=1}^{n_i}\sigma^{kk}\boldsymbol{q}_{ijk}\boldsymbol{q}_{ijk}^T + \Sigma_{i=1}^n\Sigma_{m=1}^M\Sigma_{l=1}^{L_{im}}h_{ml}\mathbb{D}_{mil})^{-1}$ with $\boldsymbol{q}_{ijk}^T = \boldsymbol{R}_{ik}(t_{ijk})$ and σ^{kk} is the (k, k)th component of $\boldsymbol{\Sigma}^{-1}$.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.csda.2014.02.015.

Please cite this article in press as: Tang, N.-S., et al., Semiparametric Bayesian joint models of multivariate longitudinal and survival data. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.02.015

UNCORRECTED PROOF

References

Albert, P.S., Shih, J.H., 2010. An approach for jointly modeling multivariate longitudinal measurements and discrete time-to-event data. Ann. Appl. Stat. 4, 1517–1532.
Bradlow, E.T., Zaslavsky, A.M., 1997. Case influence analysis in Bayesian inference. J. Comput. Graph. Statist. 6, 314–331.
Brown, E.R., Ibrahim, J.G., 2003. Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine. Biometrics 59, 686–693.
Brown, E.R., Ibrahim, J.G., DeGruttola, V., 2005. A flexible *B*-spline model for multiple longitudinal biomarkers and survival. Biometrics 61, 64–73.
Carlin, B.P., Polson, N.G., 1991. An expected utility approach to influence diagnostics. J. Amer. Statist. Assoc. 86, 1013–1021.
Chen, M.H., Ibrahim, J.G., Sinha, D., 2002. Bayesian inference for multivariate survival data with a surviving fraction. J. Multivariate Anal. 80, 101–126.
Chi, Y.Y., Ibrahim, J.G., 2006. Joint models for multivariate longitudinal and multivariate survival data. Biometrics 62, 432–445.
Chi, Y.Y., Ibrahim, J.G., 2007. Bayesian approaches to joint longitudinal and survival models accommodating both zero and nonzero cure fractions. Statist. Sinica 17, 445–462.
Cho, H., Ibrahim, J.G., Sinha, D., Zhu, H.T., 2009. Bayesian case influence diagnostics for survival models. Biometrics 65, 116–124.
Chow, S.M., Tang, N.S., Yuan, Y., Song, X.Y, Zhu, H.T., 2011. Bayesian estimation of semiparametric nonlinear dynamic factor analysis models using the Dirichlet process prior. British J. Math. Statist. Psych. 64, 69–106.
Cook, R.D., Weisberg, S., 1982. Residuals and Influence in Regression. Chapman and Hall, New York.
De Gruttola, V., Tu, X.M., 1994. Modeling progression of CD4-lymphocyte count and its relationship to survival time. Biometrics 50, 1003–1014.
Ding, J., Wang, J.L., 2008. Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. Biometrics 64, 546–556.
Faucett, C.L., Thomas, D.C., 1996. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. Stat. Med. 15, 1663–1685.
Ferguson, T.S., 1973. A Bayesian analysis of some nonparametric problems. Ann. Statist. 1, 209–230.
Fong, Y., Rue, H., Wakefield, J., 2010. Bayesian influence for generalized linear mixed models. Biostatistics 11, 397–412.
Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. 6, 721–741.
Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their application. Biometrika 57, 97–109.
Henderson, R., Diggle, P., Dobson, A., 2000. Joint modelling of longitudinal measurements and event time data. Biostatistics 1, 465–480.
Hu, W.H., Li, G., Li, N., 2009. A Bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data. Stat. Med. 29, 1601–1619.
Ibrahim, J.G., Chen, M.H., Sinha, D., 2004. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. Statist. Sinica 14, 847–867.
Ishwaran, H., Zarepour, M., 2000. Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. Biometrika 87, 371–390.
Jackson, D., White, I.R., Carpenter, J., 2012. Identifying influential observations in Bayesian models by using Markov chain Monte Carlo. Stat. Med. 31, 1238–1248.
Kleinman, K.P., Ibrahim, J.G., 1998a. A semi-parametric Bayesian approach to generalized linear mixed models. Stat. Med. 17, 2579–2596.
Kleinman, K.P., Ibrahim, J.G., 1998b. A semiparametric Bayesian approach to the random effects model. Biometrics 54, 921–938.
Law, N.J., Taylor, J.M.G., Sandler, H., 2002. The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. Biostatistics 3, 547–563.
Lee, S.Y., Lu, B., Song, X.Y., 2008. Semiparametric Bayesian analysis of structural equation models with fixed covariates. Stat. Med. 15, 2341–2360.
Li, Y.S, Müller, P., Lin, X.H., 2011. Center-adjusted inference for a nonparametric Bayesian random effect distribution. Statist. Sinica 21, 1201–1223.
Liu, J.S., 2001. Monte Carlo Strategies in Scientific Computing. Springer-Verlag, New York.
McAuliffe, J.D., Blei, D.M., Jordan, M.I., 2006. Nonparametric empirical Bayes for the Dirichlet process mixture model. Stat. Comput. 16, 5–14.
Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machine. J. Chem. Phys. 21, 1087–1091.
Ohlssen, D.I., Sharples, L.D., Spiegelhalter, D.J., 2007. Flexible random-effects models using Bayesian semi-parametric models: applications to institutional comparisons. Stat. Med. 26, 2088–2112.
Rizopoulos, D., Ghosh, P., 2011. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. Stat. Med. 30, 1366–1380.
Rizopoulos, D., Verbeke, G., Lesaffre, E., 2009. Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. J. R. Stat. Soc. 71, 637–654.
Rizopoulos, D., Verbeke, G., Molenberghs, G., 2008. Shared parameter models under random effects misspecification. Biometrika 95, 63–74.
Sethuraman, J., 1994. A constructive definition of Dirichlet priors. Statist. Sinica 4, 639–650.
Song, X., Davidian, M., Tsiatis, A.A., 2002. An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. Biostatistics 3, 511–528.
Song, X., Wang, C.Y, 2008. Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying. Biometrics 64, 557–566.
Spiegelhalter, D.J., Best, N.G, Carlin, B.P, Linde, A.V.D., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B 64, 583–639.
Tsiatis, A.A., Davidian, M., 2004. An overview of joint modeling of longitudinal and time-to-event data. Statist. Sinica 14, 793–818.
Tsiatis, A.A., DeGruttola, V., Wulfsohn, M.S., 1995. Modeling the relationship of the survival to longitudinal datameasure with error. Application to survival and CD4 counts inpatients with AIDS. J. Amer. Statist. Assoc. 90, 27–37.
Wang, Y., Taylor, J.M.G., 2001. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. J. Amer. Statist. Assoc. 96, 895–905.
Weiss, R.E., 1996. An approach to Bayesian sensitivity analysis. J. R. Stat. Soc. Ser. B Methodol. 58, 739–750.
Weiss, R.E., Cho, M., 1998. Bayesian marginal influence assessment. J. Statist. Plann. Inference 71, 163–177.
Weiss, R.E., Cook, R.D., 1992. A graphical case statistics for assessing posterior influence. Biometrics 79, 51–55.
Wulfsohn, M.S., Tsiatis, A.A., 1997. A joint model for survival and longitudinal data measured with error. Biometrics 53, 330–339.
Xu, J., Zeger, S.L., 2001. The evaluation of multiple surrogate endpoints. Biometrics 57, 81–87.
Yang, M., Dunson, D.B., Baird, D., 2010. Semiparametric Bayes hierarchical models with mean and variance constraints. Comput. Statist. Data Anal. 54, 2172–2186.
Ye, W., Lin, X., Taylor, J.M., 2008. Semiparametric modeling of longitudinal measurements and time-to-event data—a two-stage regression calibration approach. Biometrics 64, 1238–1246.
Zeger, S.L., Karim, M.R., 1991. Generalized linear models with random effects: a Gibbs sampling approach. J. Amer. Statist. Assoc. 86, 79–86.
Zhu, H., Ibrahim, J.G., Chi, Y.Y., Tang, N., 2012. Bayesian influence measures for joint models for longitudinal and survival data. Biometrics 68, 954–964.

UNCORRECTED PROOF

Please cite this article in press as: Tang, N.-S., et al., Semiparametric Bayesian joint models of multivariate longitudinal and survival data. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.02.015