Computational Statistics and Data Analysis 112 (2017) 145–153

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Analysis of binary longitudinal data with time-varying effects

Seonghyun Jeong[a,1], Minjae Park[b,1], Taeyoung Park[c,*]

[a] Department of Statistics, North Carolina State University, Raleigh, NC, USA
[b] College of Business Administration, Hongik University, Seoul, Republic of Korea
[c] Department of Applied Statistics, Yonsei University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 20 September 2016
Received in revised form 24 February 2017
Accepted 3 March 2017
Available online 22 March 2017

Keywords:
Longitudinal data
Probit mixed model
Nonparametric regression
Partial collapse
Repeated measures

ABSTRACT

This paper considers the analysis of longitudinal data where a binary response variable is observed repeatedly for each subject over time. In analyzing such data, regression coefficients are commonly assumed constant over time, which may not properly account for the time-varying effects of some subject characteristics on a sequence of binary outcomes. This paper proposes a Bayesian method for the analysis of binary longitudinal data with time-varying regression coefficients and random effects to account for nonlinear subject-specific effects over time as well as between-subject variation. The proposed method facilitates posterior computation via the method of partial collapse and accommodates spatially inhomogeneous smoothness of nonparametric functions without overfitting via a basis search technique. The proposed method is illustrated with a simulated study and the binary longitudinal data from the German socioeconomic panel study.

1. Introduction

A binary response variable is often measured repeatedly for each subject over time, along with each subject's time-dependent personal characteristics. In analyzing such binary longitudinal data, a model needs to account for nonlinear subject-specific effects on a sequence of binary responses and between-subject variability present in longitudinal data. Mixed models with random effects are thus generally adopted because they provide effective modeling of the longitudinal nature of data (Laird and Ware, 1982).

Among some link functions that provide binary transformation, the probit link is often preferred because it avoids the unrealistic assumption of independence of irrelevant alternatives of logistic models (Imai and van Dyk, 2005). Probit mixed models are, in this sense, widely used to analyze binary longitudinal data and, until recently, have been studied in various statistical settings (Varin and Czado, 2010; Kyung et al., 2010; Soyer and Sung, 2013).

A significant extension to probit mixed models can be developed by allowing the effect of predictors on a sequence of binary responses to vary with time. That is, the regression coefficients of the probit mixed models are not assumed constant but rather change with a time-dependent effect modifier, e.g., time itself. Regression models that have coefficients varying with other effect modifiers are referred to as varying-coefficient models (Hastie and Tibshirani, 1993), which are widely used to capture the time-varying effect of predictors on a response variable in a longitudinal study (Hoover et al., 1998; Wu et al., 1998; Huang et al., 2002).

* Corresponding author.
E-mail address: tpark@yonsei.ac.kr (T. Park).

1 Equally contributed.

http://dx.doi.org/10.1016/j.csda.2017.03.007
0167-9473/

Most references on varying-coefficient models focus on the estimation of the unknown functions of varying regression coefficients. In particular, the tuning of surface roughness parameters (e.g., bandwidths for kernel methods, smoothing penalty for smoothing splines, or the number of basis terms for regression splines) in estimating nonparametric functions depends on local variation of trajectories (Hoover et al., 1998; Fan and Zhang, 1999; Huang et al., 2002). The classical smoothing methods, however, suffer from the curse of dimensionality because multiple roughness parameters are required to be tuned on a high-dimensional parameter space when multiple varying coefficients are present. It is also well known that they cannot incorporate significant spatial inhomogeneity of a varying coefficient without major modifications (Ruppert et al., 2003). Under the Bayesian framework, however, there have been several attempts to estimate a spatially adaptive function (Smith and Kohn, 1996; Baladandayuthapani et al., 2005). In particular, Jeong and Park (2016) proposed a data-driven method to account for spatial inhomogeneity of varying coefficients in Gaussian linear mixed models via Bayesian basis selection, while avoiding the curse of dimensionality problems. Within the generalized linear model framework, there exist studies that consider regression models involving both random effects and smoothly varying regression coefficients (Zhang, 2004; Lu and Zhang, 2009). To the best of our knowledge, however, no method has been proposed to simultaneously address all issues mentioned above for varying-coefficient mixed models with non-Gaussian longitudinal responses.

The objective of this article is twofold. The first is to develop a Bayesian semiparametric method to estimate the time-varying and subject-specific effects of predictors on binary longitudinal responses in a probit mixed model framework. The second is to develop a data-driven Bayesian method to estimate the unknown varying-coefficient functions in probit mixed models. In this way, the proposed method reduces modeling bias in estimating varying-coefficient functions for probit varying-coefficient mixed models and adapts to spatially inhomogeneous smoothness of the unknown varying-coefficient functions without the possibility of overfitting. Thus it is not necessary to predetermine predictors with constant effects and the data-driven method allows all predictors to be assumed to have time-varying effects in the absence of prior knowledge.

The rest of this article is organized as follows. In Section 2, we introduce a probit mixed model with time-varying coefficients and develop a flexible representation of regression splines used to approximate the unknown varying coefficients. Section 3 specifies prior distributions and proposes the efficient Bayesian semiparametric analysis of binary longitudinal data with time-varying effects via the method of partial collapse. Section 4 conducts a simulation study to validate the proposed method. In Section 5, the proposed method is applied to the binary longitudinal data from the German socioeconomic panel study to examine a relationship between working status and subject characteristics. Finally, Section 6 concludes with discussion.

2. Model specification

Consider a binary response variable Y_{ij}, representing the jth observation within subject i. Given a linear predictor η_{ij}, a probit regression model is expressed as

$$P(Y_{ij} = 1|\eta_{ij}) = \Phi(\eta_{ij}), \quad i = 1, \ldots, N,\ j = 1, \ldots, n_i, \tag{1}$$

where Φ denotes the cumulative density function of a standard Gaussian random variable. To account for the time-varying effects of predictors and subject-specific effects in binary longitudinal data, we model the linear predictor η_{ij} as

$$\eta_{ij} = \mathbf{w}_{ij}^\top \boldsymbol{\alpha}(t_{ij}) + \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{z}_{ij}^\top \mathbf{b}_i, \tag{2}$$

where $\mathbf{w}_{ij} = (w_{ij1}, \ldots, w_{ijp})^\top$, $\mathbf{x}_{ij} = (x_{ij1}, \ldots, x_{ijq})^\top$, and $\mathbf{z}_{ij} = (z_{ij1}, \ldots, z_{ijr})^\top$ are $p \times 1$, $q \times 1$, and $r \times 1$ vectors of covariates, respectively, t_{ij} is a time-dependent effect modifier such as age or observation time, $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \ldots, \alpha_p(\cdot))^\top$ is a $p \times 1$ vector of smoothly varying coefficient functions to $\mathbf{w}_{ij}$, $\boldsymbol{\beta}$ is a $q \times 1$ vector of fixed effects, and $\mathbf{b}_i$ is an $r \times 1$ vector of random effects following an independent and identical multivariate Gaussian distribution, $\mathbf{b}_i \overset{\text{i.i.d.}}{\sim} N_r(\mathbf{0}, \boldsymbol{\Psi})$, with $\boldsymbol{\Psi}$ an $r \times r$ positive definite covariance matrix. From the Bayesian perspective, it is useful to express the model in (1) using an auxiliary latent random variable,

$$L_{ij} = \eta_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N_1(0, 1), \tag{3}$$

so that Y_{ij} can be represented as

$$Y_{ij} = \begin{cases} 1, & \text{if } L_{ij} > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Under some smoothness assumptions, the varying coefficient $\alpha_k(\cdot)$ is well approximated as a linear combination of spline basis terms. The number of knots for the basis terms controls the shape and smoothness of a fitted curve in regression splines. To reduce modeling bias and prevent overfitting, we develop a data-driven method that uses data to automatically adapt to the curvature of a varying-coefficient function by determining the unknown number and location of knots in a probit mixed model framework. For data-driven basis selection, we consider a set of potential basis terms with M_k knot candidates and let data choose significant knots that account for possible spatial inhomogeneity.

More specifically, we use the $(M_k+2)\times 1$ vector of potential radial basis functions defined by

$$B_k(t) = \left\{1, t, |t-\tau_{k1}|^3, \ldots, |t-\tau_{kM_k}|^3\right\}^\top, \tag{4}$$

where the abscissae, $\min(t_{ij}) < \tau_{k1} < \cdots < \tau_{kM_k} < \max(t_{ij})$, are candidates for knot locations, and we introduce the $(M_k+1)\times 1$ vector of latent knot indicator variables $\boldsymbol{\gamma}_k = (\gamma_{k0}, \gamma_{k1}, \ldots, \gamma_{kM_k})^\top$, where $\gamma_{km} = 1$ if the $(m+2)$th term in (4) is used as a basis term and 0 otherwise, for $m = 0, 1, \ldots, M_k$. For notational simplicity, we define $|\boldsymbol{\gamma}_k| = \sum_{m=0}^{M_k}\gamma_{km}$ and $|\boldsymbol{\gamma}| = \sum_{k=1}^{p}|\boldsymbol{\gamma}_k|$ with $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}_k\}_{k=1}^{p}$. Let $B_{\boldsymbol{\gamma}_k}(\cdot)$ denote a $(|\boldsymbol{\gamma}_k|+1)\times 1$ subvector of $B_k(\cdot)$, whose elements consist of the constant basis term and the other basis terms determined by $\boldsymbol{\gamma}_k$, and let $\boldsymbol{\phi}_{\boldsymbol{\gamma}_k}$ denote a $(|\boldsymbol{\gamma}_k|+1)\times 1$ vector of the corresponding coefficients. Each varying coefficient $\alpha_k(\cdot)$ can then be approximated as

$$\alpha_k(t) \approx B^\top_{\boldsymbol{\gamma}_k}(t)\boldsymbol{\phi}_{\boldsymbol{\gamma}_k}.$$

Given an $n_i \times (|\boldsymbol{\gamma}|+p)$ matrix defined by

$$\mathbf{W}^\star_{\boldsymbol{\gamma},i} = \begin{pmatrix} w_{i11}B^\top_{\boldsymbol{\gamma}_1}(t_{i1}) & \cdots & w_{i1p}B^\top_{\boldsymbol{\gamma}_p}(t_{i1}) \\ \vdots & \ddots & \vdots \\ w_{in_i1}B^\top_{\boldsymbol{\gamma}_1}(t_{in_i}) & \cdots & w_{in_ip}B^\top_{\boldsymbol{\gamma}_p}(t_{in_i}) \end{pmatrix}, \tag{5}$$

the model in (3) is rewritten in a matrix form as

$$\mathbf{L}_i = \mathbf{W}^\star_{\boldsymbol{\gamma},i}\boldsymbol{\phi}_{\boldsymbol{\gamma}} + \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N_{n_i}(\mathbf{0}, \mathbf{I}_{n_i}), \tag{6}$$

where $\mathbf{L}_i$ is an $n_i\times 1$ vector of latent variables for subject i, $\boldsymbol{\phi}_{\boldsymbol{\gamma}} = (\boldsymbol{\phi}^\top_{\boldsymbol{\gamma}_1}, \ldots, \boldsymbol{\phi}^\top_{\boldsymbol{\gamma}_p})^\top$ is a $(|\boldsymbol{\gamma}|+p)\times 1$ vector of the corresponding coefficients to $\mathbf{W}^\star_{\boldsymbol{\gamma},i}$, and $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})^\top$ and $\mathbf{Z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{in_i})^\top$ are $n_i \times q$ and $n_i \times r$ design matrices, respectively. Given the latent variables $\mathbf{L}_i$, the model in (6) is equivalent to linear mixed models with $(\boldsymbol{\phi}_{\boldsymbol{\gamma}}, \boldsymbol{\beta})$ treated as fixed effects, except that the dimension of $\boldsymbol{\phi}_{\boldsymbol{\gamma}}$ is varying. When $\boldsymbol{\gamma}$ is given, the method of partial collapse can be adopted to facilitate posterior computation within the linear mixed model framework.

Note that the matrix $\mathbf{W}^\star_{\boldsymbol{\gamma},i}$ contains all columns of $\mathbf{W}_i = (\mathbf{w}_{i1}, \ldots, \mathbf{w}_{in_i})^\top$ because the spline basis terms in (4) contain the constant basis term, and thus the sets of covariates $\mathbf{w}_{ij}$ and $\mathbf{x}_{ij}$ should be disjoint for identifiability. Predetermining which predictors have constant effects rather than time-varying effects is a practical issue and such prior knowledge is rarely available in practice. Because the proposed data-driven method automatically searches for significant basis terms, however, it is not necessary to predetermine predictors with constant effects in our method. Robustness against overfitting is illustrated in Section 5.

Another practical issue is the choice of candidates for knot location. The appropriate number of knot candidates should be incorporated to capture an unknown functional feature and avoid a high local variance (Ruppert et al., 2003; Jeong and Park, 2016). As a guideline, 20 to 30 knot candidates are recommended if an effect modifier t_{ij} is a continuous variable. If t_{ij} is ordinal with a few distinct values, knot candidates must be placed at appropriate locations and their number must be smaller than the number of distinct values. In our analyses, we use the sample quantiles of t_{ij}'s to avoid issues that may arise when the spacing of observations for an effect modifier is not constant.

3. Bayesian analysis

3.1. Prior specification

To complete a Bayesian specification of the proposed model, this section specifies prior distributions for unknown model components. First, we assign a beta-binomial prior distribution to $\boldsymbol{\gamma}_k$,

$$\pi(\boldsymbol{\gamma}_k) = \frac{\mathcal{B}(|\boldsymbol{\gamma}_k|+a, M_k+1-|\boldsymbol{\gamma}_k|+b)}{\mathcal{B}(a,b)}, \quad k = 1, \ldots, p, \tag{7}$$

where $\mathcal{B}(\cdot,\cdot)$ is a beta function. By setting $a = b = 1$, the prior in (7) becomes noninformative in that it assigns equal probabilities to the number of basis terms, i.e., $\pi(|\boldsymbol{\gamma}_k|) \propto 1$; see Scott and Berger (2010) for discussion. Second, given $(\boldsymbol{\gamma}, \kappa, \boldsymbol{\Psi})$, a multivariate Gaussian prior distribution is assigned to the fixed effects $\boldsymbol{\theta}_{\boldsymbol{\gamma}} = (\boldsymbol{\phi}^\top_{\boldsymbol{\gamma}}, \boldsymbol{\beta}^\top)^\top$,

$$\boldsymbol{\theta}_{\boldsymbol{\gamma}}|(\boldsymbol{\gamma}, \kappa, \boldsymbol{\Psi}) \sim N_{|\boldsymbol{\gamma}|+p+q}\left(\mathbf{0}, \kappa\left[\sum_{i=1}^{N}\mathbf{C}^\top_{\boldsymbol{\gamma},i}\left(\mathbf{I}_{n_i} + \mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i^\top\right)^{-1}\mathbf{C}_{\boldsymbol{\gamma},i}\right]^{-1}\right),$$

where $\mathbf{C}_{\boldsymbol{\gamma},i} = \left(\mathbf{W}^\star_{\boldsymbol{\gamma},i}, \mathbf{X}_i\right)$ is an $n_i \times (|\boldsymbol{\gamma}|+p+q)$ design matrix and κ is considered as a dispersion factor. This prior distribution can be viewed as the Zellner's g-prior (Zellner, 1986) for linear mixed models, which is convenient for efficient posterior computation with nice properties of invariance (Jeong and Park, 2016). Although the prior specification is straightforward, it is well known that the choice of a fixed value of κ is complicated by the paradoxes of g-priors (Liang et al., 2008). We thus treat κ as a random variable by assigning an inverse gamma prior distribution,

$$\kappa \sim IG\left(1/2, \sum_{i=1}^{N} n_i/2\right),$$

which becomes a Cauchy prior for $\boldsymbol{\theta}_{\boldsymbol{\gamma}}$ (Zellner and Siow, 1980). Lastly, we assign an inverse Wishart prior distribution to $\boldsymbol{\Psi}$,

$$\boldsymbol{\Psi} \sim IW(u, \mathbf{V}),$$

with fixed constants u and $\mathbf{V}$ that make the corresponding prior distribution vaguely informative.

3.2. Efficient posterior inference

Given the prior specification, our goal is to generate samples from the joint posterior distribution $\pi(\boldsymbol{\theta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \kappa, \mathbf{b}, \boldsymbol{\Psi}, \mathbf{L}|\mathbf{Y})$ where $\mathbf{b} = \{\mathbf{b}_i\}_{i=1}^N$ is a collection of random effects, $\mathbf{L} = \{\mathbf{L}_i\}_{i=1}^N$ is a collection of latent variables, and $\mathbf{Y} = \{\mathbf{Y}_i\}_{i=1}^N$ is a collection of binary responses. To facilitate posterior computation, we consider partially collapsing $(\mathbf{b}, \boldsymbol{\theta}_{\boldsymbol{\gamma}})$ out of the proposed model. More specifically, closed-form expressions are available for the following marginal posterior distributions,

$$\pi(\boldsymbol{\theta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \kappa, \boldsymbol{\Psi}, \mathbf{L}|\mathbf{Y}) = \int \pi(\boldsymbol{\theta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \kappa, \mathbf{b}, \boldsymbol{\Psi}, \mathbf{L}|\mathbf{Y}) d\mathbf{b},$$
$$\pi(\boldsymbol{\gamma}, \kappa, \boldsymbol{\Psi}, \mathbf{L}|\mathbf{Y}) = \int \pi(\boldsymbol{\theta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \kappa, \boldsymbol{\Psi}, \mathbf{L}|\mathbf{Y}) d\boldsymbol{\theta}_{\boldsymbol{\gamma}},$$

which allows to construct an efficient posterior sampling scheme via the method of partial collapse (van Dyk and Park, 2008; Park and van Dyk, 2009; van Dyk and Jiao, 2015). By qualitatively examining mixing and autocorrelation plots and quantitatively assessing convergence with the $\hat{R}^{1/2}$ statistic (Gelman and Rubin, 1992) for all model parameters of fixed dimension, we can confirm that subchains for the parameters mix very well and exhibit fast convergence characteristics; refer to Park and Min (2016) for general discussion about efficiency of the method of partial collapse for linear mixed models. The resulting Metropolis–Hastings within a partially collapsed Gibbs sampler is constructed as follows:

Step 1. Draw $\boldsymbol{\gamma}_k$ from the conditional distribution of $\boldsymbol{\gamma}_k$ given $(\boldsymbol{\gamma}_{-k}, \kappa, \boldsymbol{\Psi}, \mathbf{L}, \mathbf{Y})$, which is proportional to

$$\pi(\boldsymbol{\gamma}_k|\boldsymbol{\gamma}_{-k}, \kappa, \boldsymbol{\Psi}, \mathbf{L}, \mathbf{Y}) \propto (\kappa+1)^{-(|\boldsymbol{\gamma}|+p+q)/2} \mathcal{B}(|\boldsymbol{\gamma}_k| + a, M_k + 1 - |\boldsymbol{\gamma}_k| + b)$$
$$\times \exp\left(\frac{\kappa}{2(\kappa+1)} \boldsymbol{\xi}(\boldsymbol{\gamma}, \boldsymbol{\Psi}, \mathbf{L})^\top \boldsymbol{\Xi}(\boldsymbol{\gamma}, \boldsymbol{\Psi})^{-1} \boldsymbol{\xi}(\boldsymbol{\gamma}, \boldsymbol{\Psi}, \mathbf{L})\right), \quad k = 1, \ldots, p,$$

where $\boldsymbol{\gamma}_{-k}$ denotes all elements of $\boldsymbol{\gamma}$ except $\boldsymbol{\gamma}_k$, and we have

$$\boldsymbol{\Xi}(\boldsymbol{\gamma}, \boldsymbol{\Psi}) = \sum_{i=1}^{N} \mathbf{C}_{\boldsymbol{\gamma},i}^\top \left(\mathbf{I}_{n_i} + \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^\top\right)^{-1} \mathbf{C}_{\boldsymbol{\gamma},i},$$
$$\boldsymbol{\xi}(\boldsymbol{\gamma}, \boldsymbol{\Psi}, \mathbf{L}) = \sum_{i=1}^{N} \mathbf{C}_{\boldsymbol{\gamma},i}^\top \left(\mathbf{I}_{n_i} + \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^\top\right)^{-1} \mathbf{L}_i.$$

To efficiently draw $\boldsymbol{\gamma}_k$, we construct a consecutive sequence of blocks for the elements of $\boldsymbol{\gamma}_k$ in each iteration, where the first element γ_{k0} for the linear basis term consists of the first block of size 1 and the sizes of subsequent blocks are randomly chosen among 2, 3, and 4. For example, a possible sequence of blocks consists of $\{(\gamma_{k0}), (\gamma_{k1}, \gamma_{k2}, \gamma_{k3}), (\gamma_{k4}, \gamma_{k5}), \ldots, (\gamma_{k,M_k-3}, \gamma_{k,M_k-2}, \gamma_{k,M_k-1}, \gamma_{kM_k})\}$. Then we use the block sampling scheme proposed by Kohn et al. (2001), which is the Metropolis–Hastings algorithm with a conditional prior distribution for each block as a proposal distribution. When $\boldsymbol{\gamma}_{k,B}$ denotes a certain block to be updated, $\boldsymbol{\gamma}'_{k,B}$ is generated from the conditional prior distribution $\pi(\boldsymbol{\gamma}_{k,B}|\boldsymbol{\gamma}_{k,\backslash B})$ and the move from $\boldsymbol{\gamma}_{k,B}$ to $\boldsymbol{\gamma}'_{k,B}$ is accepted with probability

$$\min\left\{1, \frac{\pi(\boldsymbol{\gamma}'_{k,B}, \boldsymbol{\gamma}_{k,\backslash B}|\boldsymbol{\gamma}_{-k}, \kappa, \boldsymbol{\Psi}, \mathbf{L}, \mathbf{Y})\pi(\boldsymbol{\gamma}_{k,B}|\boldsymbol{\gamma}_{k,\backslash B})}{\pi(\boldsymbol{\gamma}_{k,B}, \boldsymbol{\gamma}_{k,\backslash B}|\boldsymbol{\gamma}_{-k}, \kappa, \boldsymbol{\Psi}, \mathbf{L}, \mathbf{Y})\pi(\boldsymbol{\gamma}'_{k,B}|\boldsymbol{\gamma}_{k,\backslash B})}\right\},$$

where $\boldsymbol{\gamma}_{k,\backslash B}$ represents all elements of $\boldsymbol{\gamma}_k$ except those in $\boldsymbol{\gamma}_{k,B}$. We repeat this sampling procedure for all blocks in $\boldsymbol{\gamma}_k$.

Step 2. Draw $\boldsymbol{\theta}_{\boldsymbol{\gamma}}$ from $\pi(\boldsymbol{\theta}_{\boldsymbol{\gamma}}|\boldsymbol{\gamma}, \kappa, \boldsymbol{\Psi}, \mathbf{L}, \mathbf{Y})$, which is multivariate Gaussian,

$$\boldsymbol{\theta}_{\boldsymbol{\gamma}}|(\boldsymbol{\gamma}, \kappa, \boldsymbol{\Psi}, \mathbf{L}, \mathbf{Y}) \sim N_{|\boldsymbol{\gamma}|+p+q}\left(\frac{\kappa}{\kappa+1} \boldsymbol{\Xi}(\boldsymbol{\gamma}, \boldsymbol{\Psi})^{-1} \boldsymbol{\xi}(\boldsymbol{\gamma}, \boldsymbol{\Psi}, \mathbf{L}), \frac{\kappa}{\kappa+1} \boldsymbol{\Xi}(\boldsymbol{\gamma}, \boldsymbol{\Psi})^{-1}\right).$$

Step 3. Draw $\mathbf{b}$ from $\pi(\mathbf{b}|\boldsymbol{\theta}_\gamma, \boldsymbol{\gamma}, \kappa, \boldsymbol{\Psi}, \mathbf{L}, \mathbf{Y})$, which is a product of N independent multivariate Gaussian distributions,

$$\mathbf{b}_i|(\boldsymbol{\theta}_\gamma, \boldsymbol{\gamma}, \kappa, \boldsymbol{\Psi}, \mathbf{L}, \mathbf{Y}) \overset{\text{ind}}{\sim} N_r\left(\hat{\mathbf{b}}_i(\boldsymbol{\theta}_\gamma, \boldsymbol{\gamma}, \boldsymbol{\Psi}, \mathbf{L}),\ \boldsymbol{\Sigma}(\boldsymbol{\Psi})\right), \quad i = 1, \dots, N,$$

where

$$\hat{\mathbf{b}}_i(\boldsymbol{\theta}_\gamma, \boldsymbol{\gamma}, \boldsymbol{\Psi}, \mathbf{L}) = \boldsymbol{\Psi}\mathbf{Z}_i^\top \left(\mathbf{I}_{n_i} + \mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i^\top\right)^{-1}\left(\mathbf{L}_i - \mathbf{C}_{\gamma,i}\boldsymbol{\theta}_\gamma\right),$$
$$\boldsymbol{\Sigma}(\boldsymbol{\Psi}) = \boldsymbol{\Psi} - \boldsymbol{\Psi}\mathbf{Z}_i^\top \left(\mathbf{I}_{n_i} + \mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i^\top\right)^{-1}\mathbf{Z}_i\boldsymbol{\Psi}.$$

Step 4. Draw κ from $\pi(\kappa|\boldsymbol{\theta}_\gamma, \boldsymbol{\gamma}, \mathbf{b}, \boldsymbol{\Psi}, \mathbf{L}, \mathbf{Y})$, which is inverse gamma,

$$\kappa|(\boldsymbol{\theta}_\gamma, \boldsymbol{\gamma}, \mathbf{b}, \boldsymbol{\Psi}, \mathbf{L}, \mathbf{Y}) \sim IG\left(\frac{1 + |\boldsymbol{\gamma}| + p + q}{2}, \frac{\sum_{i=1}^N n_i + \boldsymbol{\theta}_\gamma^\top \Xi(\boldsymbol{\gamma}, \boldsymbol{\Psi})\boldsymbol{\theta}_\gamma}{2}\right).$$

Step 5. Draw $\boldsymbol{\Psi}$ from $\pi(\boldsymbol{\Psi}|\boldsymbol{\theta}_\gamma, \boldsymbol{\gamma}, \kappa, \mathbf{b}, \mathbf{L}, \mathbf{Y})$, where the conditional distribution of $\boldsymbol{\Psi}$ given $(\boldsymbol{\theta}_\gamma, \boldsymbol{\gamma}, \kappa, \mathbf{b}, \mathbf{L}, \mathbf{Y})$ is proportional to

$$\pi(\boldsymbol{\Psi}|\boldsymbol{\theta}_\gamma, \boldsymbol{\gamma}, \kappa, \mathbf{b}, \mathbf{L}, \mathbf{Y}) \propto \mathcal{IW}\left(\boldsymbol{\Psi}; u + N, \left[\mathbf{V} + \sum_{i=1}^N \mathbf{b}_i\mathbf{b}_i^\top\right]\right) \times \mathcal{N}_{|\boldsymbol{\gamma}|+p+q}\left(\boldsymbol{\theta}_\gamma; \mathbf{0}, \kappa\ \Xi(\boldsymbol{\gamma}, \boldsymbol{\Psi})^{-1}\right), \tag{8}$$

where $\mathcal{N}_d(\cdot\,; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the density function of a d-dimensional multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and $\mathcal{IW}(\cdot\,; \nu, \boldsymbol{\Lambda})$ denotes the density function of an inverse Wishart distribution with ν degrees of freedom and scale matrix $\boldsymbol{\Lambda}$. To draw $\boldsymbol{\Psi}$, we use a Metropolized independent sampler with an inverse Wishart proposal distribution given by

$$\mathcal{IW}\left(\boldsymbol{\Psi}; u + N, \left[\mathbf{V} + \sum_{i=1}^N \mathbf{b}_i\mathbf{b}_i^\top\right]\right).$$

The move from $\boldsymbol{\Psi}$ to $\boldsymbol{\Psi}'$ is then accepted with probability

$$\min\left\{1, \frac{\mathcal{N}_{|\boldsymbol{\gamma}|+p+q}\left(\boldsymbol{\theta}_\gamma; \mathbf{0}, \kappa\ \Xi(\boldsymbol{\gamma}, \boldsymbol{\Psi}')^{-1}\right)}{\mathcal{N}_{|\boldsymbol{\gamma}|+p+q}\left(\boldsymbol{\theta}_\gamma; \mathbf{0}, \kappa\ \Xi(\boldsymbol{\gamma}, \boldsymbol{\Psi})^{-1}\right)}\right\}.$$

As $N \to \infty$ and under stationarity, the target distribution in (8) is dominated by the first term; refer to Jeong and Park (2016) for details.

Step 6. Draw $\mathbf{L}$ from $\pi(\mathbf{L}|\boldsymbol{\theta}_\gamma, \boldsymbol{\gamma}, \kappa, \mathbf{b}, \boldsymbol{\Psi}, \mathbf{L}, \mathbf{Y})$, which is a product of independent truncated Gaussian distributions,

$$L_{ij}|(\boldsymbol{\theta}_\gamma, \boldsymbol{\gamma}, \kappa, \mathbf{b}, \boldsymbol{\Psi}, \mathbf{L}, \mathbf{Y}) \sim \begin{cases} TN_{(-\infty,0)}(\{\mathbf{C}_{\gamma,i}\boldsymbol{\theta}_\gamma + \mathbf{Z}_i\mathbf{b}_i\}_j, 1), & \text{if } Y_{ij} = 0, \\ TN_{[0,\infty)}(\{\mathbf{C}_{\gamma,i}\boldsymbol{\theta}_\gamma + \mathbf{Z}_i\mathbf{b}_i\}_j, 1), & \text{if } Y_{ij} = 1, \end{cases} \quad i = 1, \dots, N,\ j = 1, \dots, n_i,$$

where $\{\mathbf{C}_{\gamma,i}\boldsymbol{\theta}_\gamma + \mathbf{Z}_i\mathbf{b}_i\}_j$ denotes the jth element of $\mathbf{C}_{\gamma,i}\boldsymbol{\theta}_\gamma + \mathbf{Z}_i\mathbf{b}_i$ and TN_A denotes a Gaussian distribution truncated to region A.

The above posterior sampling scheme consists of a functionally incompatible set of conditional distributions, and thus permuting the order of sampling steps may upset the target stationary distribution of the corresponding Markov transition kernel (van Dyk and Park, 2008; Park and van Dyk, 2009). Therefore, great care must be taken to maintain the target distribution in implementing the posterior sampling scheme.

4. Simulation study

In this section, we conduct a simulation study to assess the performance of the proposed method. Test data are generated with two time-varying coefficients given by

$$\alpha_1(t) = \sin(2\pi t),$$
$$\alpha_2(t) = 3\exp[-200(t - 0.2)^2] + \exp[-50(t - 0.6)^2],$$

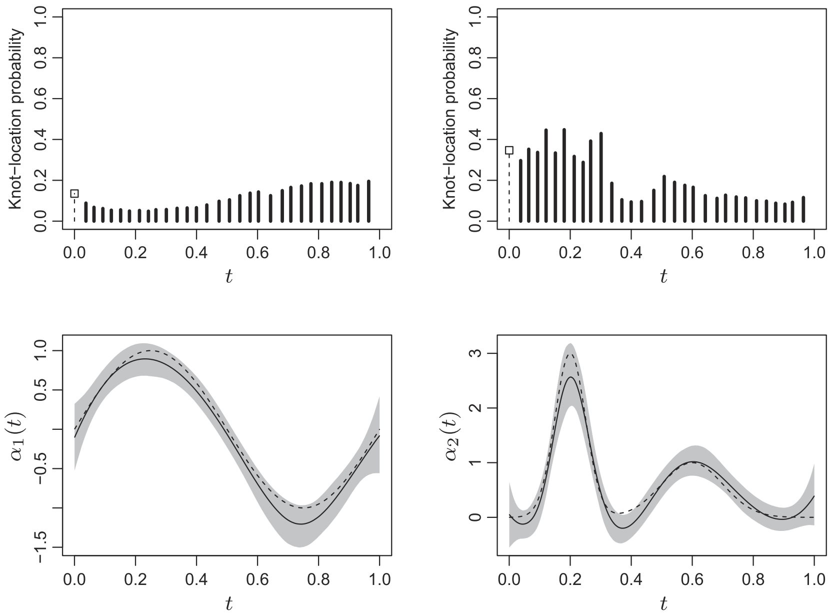

Fig. 1. Posterior summaries for two varying coefficients. The top two panels show the posterior inclusion probabilities of a linear basis term (open square with dashed lines) and knot-based basis terms (solid lines). The bottom two panels show the true varying-coefficient functions (dashed lines), point-wise posterior medians of the functions (solid lines), and point-wise 95% credible intervals (gray regions).

for $0 < t < 1$, three fixed effects given by $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^\top = (-0.5, 0.5, 1)^\top$, and a covariance matrix for two random effects given by

$$\boldsymbol{\Psi} = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{12} & \psi_{22} \end{pmatrix} = \begin{pmatrix} 0.5 & 0.2 \\ 0.2 & 1 \end{pmatrix}.$$

The number n_i of observations for subject i is randomly generated from Poisson$(10) + 1$, the number N of subjects is set to 200, the values $\{t_{ij} : i = 1, \ldots, N,\ j = 1, \ldots, n_i\}$ are randomly drawn from a uniform distribution between 0 and 1, and all values of known covariates $\{(\mathbf{w}_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}) :\ i = 1, \ldots, N,\ j = 1, \ldots, n_i\}$ are generated from independent standard Gaussian distributions. We use the 30 quantiles of observation times as possible candidates for knot location used to approximate varying-coefficient functions. With two over-dispersed starting values, we run the proposed sampling scheme for 10 000 iterations and convergence is assessed with the $\hat{R}^{1/2}$ statistic for all model parameters of fixed dimension. Posterior inference is then made with the second halves of the two chains.

Posterior summaries for the two varying coefficients are presented in Fig. 1. The top two panels of Fig. 1 show the estimated posterior inclusion probabilities of a linear basis term and knot-based basis terms for each of the two varying-coefficients. The estimated posterior inclusion probabilities of all the basis terms for the first varying coefficient $\alpha_1(t)$ tend to be uniform over the interval $(0, 1)$, while those for the second varying coefficient $\alpha_2(t)$ are non-uniform. Such results illustrate that the proposed data-driven method can automatically adjust the spatially inhomogeneous smoothness of the varying coefficients by using more knot-based basis terms in a high-curvature region especially for the second varying coefficient. The bottom two panels of Fig. 1 show the point-wise posterior medians for each of the two varying coefficients with point-wise 95% credible intervals. The true varying-coefficient functions are correctly estimated by the point-wise posterior medians and well covered by the corresponding point-wise 95% credible intervals. Fig. 2 shows the marginal posterior distributions of model parameters other than the varying-coefficient functions. In Fig. 2, the true values of the model parameters used to generate test data are well covered by the corresponding marginal posterior distributions, thereby illustrating the validity and performance of the proposed method.

5. Application to binary longitudinal data

5.1. Data description and modeling procedure

The proposed method is illustrated using the binary longitudinal data from the German socioeconomic panel study (Riphahn et al., 2003). The data set comes from an unbalanced panel study with 7293 subjects in Germany for the years 1984–1988, 1991, and 1994. The number of observations for each subject ranges from 1 to 7, and the data have 27,326 observations in total.

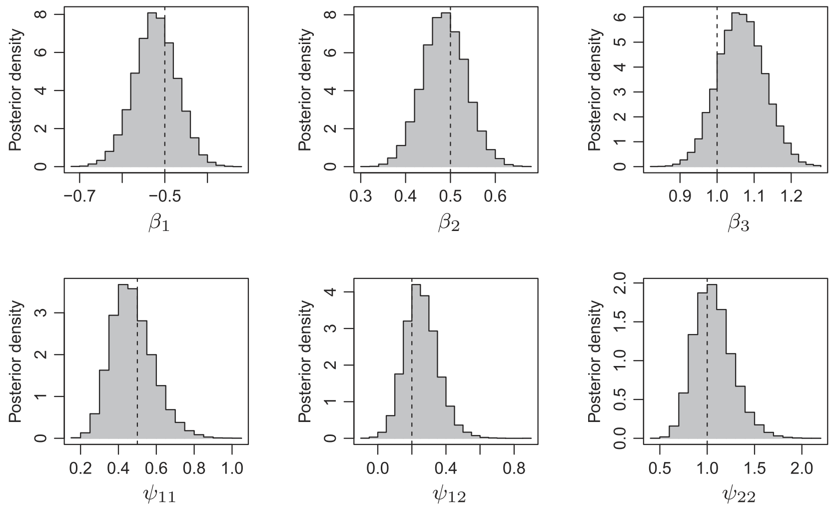

Fig. 2. Marginal posterior distributions for model parameters other than the varying-coefficient functions. The vertical dashed lines represent the true values of the parameters.

In our analysis, we are supposed to examine a relationship between working status ($1 = $ working, $0 = $ otherwise) and subject characteristics given by G_i (gender; $1 = $ male, $0 = $ female), E_{ij} (education; years of schooling), M_{ij} (marital status; $1 = $ married, $0 = $ otherwise), H_{ij} (degree of handicap; 0 to 100), K_{ij} (household with any kids under 16; $1 = $ yes, $0 = $ no), and S_{ij} (personal health satisfaction; 0–10). A hypothesis is that the relationship between working status and the subject characteristics may dynamically change with age of subject (denoted by A_{ij}), and thus the proposed method is used to test if such a functional relationship really exists.

It is reasonable to allow for the possibility of an interaction between gender and the other predictors on working status because working conditions for women are likely to be different from those for men. Thus, a flexible model involving gender-specific effects is constructed as

$$\eta_{ij} = \begin{cases} \alpha_0^m(A_{ij}) + E_{ij}\alpha_E^m(A_{ij}) + K_{ij}\alpha_K^m(A_{ij}) + S_{ij}\alpha_S^m(A_{ij}) + \beta_M^m M_{ij} + \beta_H^m H_{ij} + b_i, & \text{if male,} \\ \alpha_0^f(A_{ij}) + E_{ij}\alpha_E^f(A_{ij}) + K_{ij}\alpha_K^f(A_{ij}) + S_{ij}\alpha_S^f(A_{ij}) + \beta_M^f M_{ij} + \beta_H^f H_{ij} + b_i, & \text{if female,} \end{cases} \tag{9}$$

which can be expressed as the model in (2) by using the indicator variable G_i. In the model in (9), the effects of M_{ij} and H_{ij} on working status are assumed to be constant over age, which is based on our preliminary investigation.

5.2. Analysis and results

Posterior summaries for the varying coefficients in the model (9) are given in Fig. 3, where the left panels correspond to the posterior estimates of gender-specific varying-coefficient functions and the right panels correspond to a difference in varying coefficient between different gender groups. The first row of Fig. 3 shows the results of varying-intercept functions which decrease nonlinearly with age of males and females. A difference in varying intercept between the two gender groups is significantly positive for subjects under the age of 35, illustrating that young males are more likely to be employed than young females. The second row of Fig. 3 shows that varying-coefficient functions for education increase nonlinearly with age of subjects, where the varying-coefficient functions are significantly different between males and females under the age of 35. In particular, higher education levels negatively affect the probability of employment for males in their 20s and early 30s after which the estimated effect of education becomes positive and constant. The third row of Fig. 3 shows the results of varying-coefficient functions for whether or not there are any kids under 16 in household. By using data-driven basis terms in regression splines, the varying coefficient function for males is estimated as a constant of about zero, illustrating that the proposed method prevents overfitting. That is, for males, the probability of being employed is not affected by having any kids under 16 in household. By contrast, the estimated effect of having any kids under 16 in household on the probability of being employed varies with age of females and is more negative for younger females. The gender difference in the varying coefficient becomes smaller as subjects get older but remains significant until the age of 60. The last row of Fig. 3 shows the results of varying-coefficient functions for personal health satisfaction. The estimated effect of personal health satisfaction on the probability of employment increases with age of males but is not significantly different from zero for females, which also illustrates robustness against overfitting. The gender difference in the varying coefficient is not significant for subjects in their 20s after which the gender difference becomes significantly positive.

The summary statistics of the fixed effects and the variance of the random effect, $\mathrm{Var}(b_i) = \psi$, in the model (9) are given in Table 1. The fixed effect of marital status is estimated to be positive for males and negative for females, while

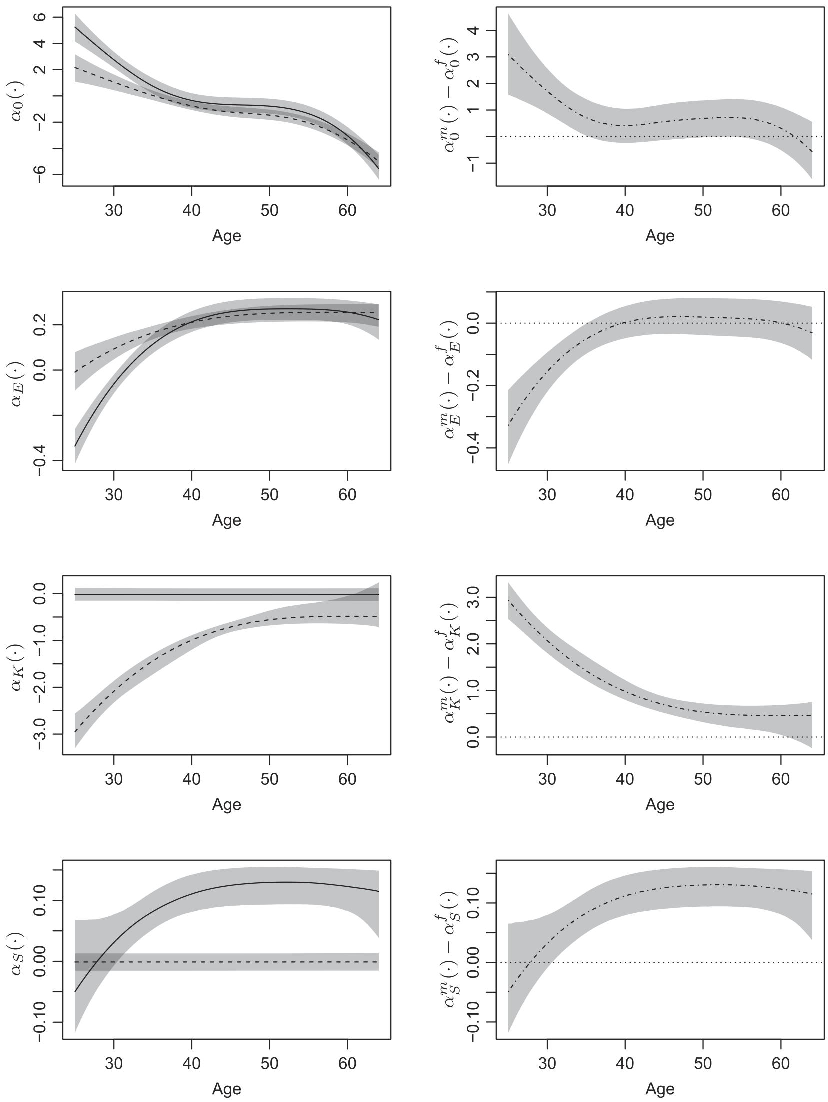

Fig. 3. Posterior summaries for varying coefficients in the analysis of the German socioeconomic panel study data. The left panels show the point-wise posterior medians of varying-coefficient functions for males (solid lines) and females (dashed lines) with point-wise 95% credible intervals (gray regions), while the right panels show the point-wise posterior medians of difference (dashed–dotted lines) in varying coefficient between the two gender groups with point-wise 95% credible intervals (gray regions). The dotted lines in the right panels represent a difference of zero.

Table 1
Posterior summary statistics of selected model parameters.

Parameter	Mean	Median	2.5 percentile	97.5 percentile
β_M^m	0.6662	0.6673	0.5264	0.8024
β_M^f	−0.6865	−0.6861	−0.7894	−0.5864
β_H^m	−0.0250	−0.0250	−0.0300	−0.0199
β_H^f	−0.0169	−0.0169	−0.0216	−0.0122
ψ	3.3685	3.3599	3.0719	3.7127

both are significant. That is, married males are more likely to be employed than unmarried males, while the opposite is true for females. As expected, a higher degree of handicap negatively affects the probability of being employed for both males and females. The bottom row of Table 1 illustrates that a large amount of between-subject variation present in binary longitudinal data is accounted for by using a random effect in a model.

6. Discussion

This article proposes a Bayesian semiparametric approach for inference on probit varying-coefficient mixed models constructed for binary longitudinal data. Our focus is in particular on the estimation of the time-varying and subject-specific effects of predictors on binary longitudinal responses, while using data-driven basis terms to flexibly estimate unknown varying-coefficient functions. To do so, probit models are extended to a longitudinal setting that allows the effect of predictors on a sequence of binary responses to vary over time and accounts for subject-specific effects by adding random effects. Parameters associated with the resulting probit varying-coefficient mixed models are efficiently estimated with the method of partial collapse that enables data-driven basis search for approximating varying-coefficient functions.

An interesting area for future research is the generalization to other classes of generalized linear mixed models in order to analyze longitudinal data following an exponential family distribution. This can be achieved by developing a variable selection method for generalized linear mixed models. In addition, the proposed method can be generalized to the cases where missing values are present in the response or predictor variables, which are common in longitudinal studies.

Acknowledgments

M. Park's research was supported by the National Research Foundation of Korea Grant funded by Korea Government (NRF-2014S1A5A8012594) and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2014R1A1A2053679). T. Park's research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2014R1A1A2056049).

References

Baladandayuthapani, V., Mallick, B.K., Carroll, R.J., 2005. Spatially adaptive Bayesian penalized regression splines (p-splines). J. Comput. Graph. Statist. 14 (2), 378–394.
Fan, J., Zhang, W., 1999. Statistical estimation in varying coefficient models. Ann. Statist. 27 (5), 1491–1518.
Gelman, A., Rubin, D.B., 1992. Inference from iterative simulations using multiple sequences (with discussion). Statist. Sci. 7, 457–472.
Hastie, T., Tibshirani, R., 1993. Varying-coefficient models. J. R. Stat. Soc. Ser. B Stat. Methodol. 55 (4), 757–796.
Hoover, D.R., Rice, J.A., Wu, C.O., Yang, L.-P., 1998. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika 85 (4), 809–822.
Huang, J.Z., Wu, C.O., Zhou, L., 2002. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. Biometrika 89 (1), 111–128.
Imai, K., van Dyk, D.A., 2005. A Bayesian analysis of the multinomial probit model using marginal data augmentation. J. Econometrics 124, 311–334.
Jeong, S., Park, T., 2016. Bayesian semiparametric inference on functional relationships in linear mixed models. Bayesian Anal. 11 (4), 1137–1163.
Kohn, R., Smith, M., Chan, D., 2001. Nonparametric regression using linear combinations of basis functions. Stat. Comput. 11 (4), 313–322.
Kyung, M., Gill, J., Casella, G., et al., 2010. Estimation in Dirichlet random effects models. Ann. Statist. 38 (2), 979–1009.
Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38 (4), 963–974.
Liang, F., Paulo, R., Molina, G., Clyde, M.A., Berger, J.O., 2008. Mixtures of g-priors for Bayesian variable selection. J. Amer. Statist. Assoc. 103 (481), 410–423.
Lu, Y., Zhang, R., 2009. Smoothing spline estimation of generalised varying-coefficient mixed model. J. Nonparametr. Stat. 21 (7), 815–825.
Park, T., Min, S., 2016. Partially collapsed Gibbs sampling for linear mixed-effects models. Comm. Statist. Simulation Comput. 45 (1), 165–180.
Park, T., van Dyk, D.A., 2009. Partially collapsed Gibbs samplers: Illustrations and applications. J. Comput. Graph. Statist. 18 (2), 283–305.
Riphahn, R.T., Wambach, A., Million, A., 2003. Incentive effects in the demand for health care: a bivariate panel count data estimation. J. Appl. Econometrics 18 (4), 387–405.
Ruppert, D., Wand, M.P., Carroll, R.J., 2003. Semiparametric Regression. Cambridge University Press.
Scott, J.G., Berger, J.O., 2010. Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. Ann. Statist. 38 (5), 2587–2619.
Smith, M., Kohn, R., 1996. Nonparametric regression using Bayesian variable selection. J. Econometrics 75 (2), 317–343.
Soyer, R., Sung, M., 2013. Bayesian dynamic probit models for the analysis of longitudinal data. Comput. Statist. Data Anal. 68, 388–398.
van Dyk, D.A., Jiao, X., 2015. Metropolis-Hastings within partially collapsed Gibbs samplers. J. Comput. Graph. Statist. 24 (2), 301–327.
van Dyk, D.A., Park, T., 2008. Partially collapsed Gibbs samplers: Theory and methods. J. Amer. Statist. Assoc. 193 (482), 790–796.
Varin, C., Czado, C., 2010. A mixed autoregressive probit model for ordinal longitudinal data. Biostatistics 11 (1), 127–138.
Wu, C.O., Chiang, C.-T., Hoover, D.R., 1998. Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. J. Amer. Statist. Assoc. 93 (444), 1388–1402.
Zellner, A., 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In: Bayesian Inference and Decision Techniques: Essays in Honor of Bruno De Finetti. Vol. 6. pp. 233–243.
Zellner, A., Siow, A., 1980. Posterior odds ratios for selected regression hypotheses. Trabajos Estadística Invest. Oper. 31 (1), 585–603.
Zhang, D., 2004. Generalized linear mixed models with varying coefficients for longitudinal data. Biometrics 60 (1), 8–15.