BAYESIAN REGRESSION ANALYSIS WITH LONGITUDINAL MEASUREMENTS

A Dissertation

by

DUCHWAN RYU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2005

Major Subject: Statistics

BAYESIAN REGRESSION ANALYSIS WITH LONGITUDINAL MEASUREMENTS

A Dissertation

by

DUCHWAN RYU

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Bani K. Mallick
(Co-Chair of Committee)

Raymond J. Carroll
(Co-Chair of Committee)

Marina Vannucci
(Member)

Yalchin Efendiev
(Member)

Michael Longnecker
(Head of Department)

May 2005

Major Subject: Statistics

ABSTRACT

Bayesian Regression Analysis with Longitudinal Measurements. (May 2005)

Duchwan Ryu, B.S., Korea University;

M.S., Korea University

Co–Chairs of Advisory Committee: Dr. Bani K. Mallick
Dr. Raymond J. Carroll

Bayesian approaches to the regression analysis for longitudinal measurements are considered. The history of measurements from a subject may convey characteristics of the subject. Hence, in a regression analysis with longitudinal measurements, the characteristics of each subject can be served as covariates, in addition to possible other covariates. Also, the longitudinal measurements may lead to complicated covariance structures within each subject and they should be modeled properly.

When covariates are some unobservable characteristics of each subject, Bayesian parametric and nonparametric regressions have been considered. Although covariates are not observable directly, by virtue of longitudinal measurements, the covariates can be estimated. In this case, the measurement error problem is inevitable. Hence, a classical measurement error model is established. In the Bayesian framework, the regression function as well as all the unobservable covariates and nuisance parameters are estimated. As multiple covariates are involved, a generalized additive model is adopted, and the Bayesian backfitting algorithm is utilized for each component of the additive model. For the binary response, the logistic regression has been proposed, where the link function is estimated by the Bayesian parametric and nonparametric regressions. For the link function, introduction of latent variables make the computing fast.

In the next part, each subject is assumed to be observed not at the prespecified time-points. Furthermore, the time of next measurement from a subject is supposed to be dependent on the previous measurement history of the subject. For this outcome-dependent follow-up times, various modeling options and the associated analyses have been examined to investigate how outcome-dependent follow-up times affect the estimation, within the frameworks of Bayesian parametric and nonparametric regressions. Correlation structures of outcomes are based on different correlation coefficients for different subjects. First, by assuming a Poisson process for the follow-up times, regression models have been constructed. To interpret the subject-specific random effects, more flexible models are considered by introducing a latent variable for the subject-specific random effect and a survival distribution for the follow-up times. The performance of each model has been evaluated by utilizing Bayesian model assessments.

To Mikyung, Hannah and Mira

ACKNOWLEDGMENTS

I wish to express my appreciation to my advisors, Dr. Bani K. Mallick and Dr. Raymond J. Carroll, for their expert and patient direction of my Ph.D. program. Their experience and insight have been very helpful in the preparation of this dissertation. I shall ever be in debt to them for the influence that they have been in my life. I would also like to thank the other members of my committee, Dr. Marina Vannucci and Dr. Yalchin Efendiev for their help and advice in the preparation of this dissertation. I am also grateful to faculty, staff and colleagues at the Department of Statistics for their friendship and help during my time at Texas A&M University.

My sincere thanks go to Dr. Emmanuel Fernando and Dr. Cliff Spiegelman at the Texas Transportation Institute at Texas A&M University. Their support and encouragement are greatly appreciated.

For this work, I would like to thank M. Pepe and K. Seidel for the data on child growth, S. Berry for the Fortran source code of the Bayesian smoothing spline, and Dr. Debajyoti Sinha for valuable comments on the outcome-dependent follow-up times.

My most appreciative and loving thanks go to my parents. Without their constant love and unlimited support, this dissertation would have been impossible. I would also like to thank my parents-in-law for their encouragement and prayers. Last but not least, I am truly indebted to my wife, Mikyung Ha, for her endless love and support.

TABLE OF CONTENTS

LIST OF TABLES

TABLE Page

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

This dissertation has utilized Bayesian approaches to perform flexible modeling with longitudinal data. Longitudinal measurements often show interesting features. Sometimes they may contribute a measurement error to the primary model. As the characteristics of each individual can be estimated by the longitudinal measurements, a regression analysis with those characteristics can be considered. In this case, although the characteristics serve as covariates in the primary model, usually they are not observable. Hence, a regression with those characteristics involves a measurement error problem. The motivation of this measurement error problem stems from a study on the adulthood obesity. Whitaker et al. (1997) were interested in the extent to which childhood growth data can predict the likelihood of obesity in adulthood. From these longitudinal measurements of childhood growth, a simple linear regression of childhood BMI's with its monitoring ages provides the information of the initial childhood BMIs and the slope of the childhood BMIs. These random regression coefficients served as covariates to predict the likelihood of adulthood obesity of the corresponding individual. Although these regression coefficients could not be observed, they could be estimated them with some errors. Because the regression coefficient are unknown, the classical measurement error model is inevitable in this case.

For both continuous and binary responses, Bayesian parametric regression analyses have been performed. As the linear relationship is not appropriate, Bayesian nonparametric regressions have been considered. For a Bayesian nonparametric regression, a Bayesian natural cubic spline has been considered with a partially improper

---

The journal model is *Journal of the American Statistical Association.*

Gaussian prior. The multiple covariates have been included into the primary model by a generalized additive model, and dealt with the Bayesian backfitting algorithm. The measurement error problems in those regression analyses have been successfully handled in Bayesian framework.

Another interesting feature of longitudinal measurements based on an outcome-dependent follow-up times has been considered. In this case, individuals may not be observed at prespecified time-points. Furthermore, the time of next measurement for each individual may depend on the individual's history of previous measurements. For example, in the cardiotoxic effects of doxorubicin chemotherapy for the treatment of acute lymphoblastic leukemia in childhood (Lipsitz et al., 2002; Fitzmaurice et al., 2003), the design points are not pre-defined but determined by the preceding response. This outcome-dependent feature of measurements makes biased estimation of regression line. As noticed by Lipsitz et al. (2002); Fitzmaurice et al. (2003), even the least square estimates will be biased, which does not require the distributional assumption of response error.

For this problem, Bayesian parametric as well as nonparametric regressions have been applied by allowing different correlation coefficients for each individual. We introduce a novel models by utilizing a latent variable for the subject-specific random effect as well as relaxing the distribution of the follow-up times. For this flexible model, both Bayesian parametric and nonparametric regression have been explored. All these models have been assessed under Bayesian model choice criterion. For this model assessment, conditional predictive ordinate (CPO) has been customized and utilized. Each chapter can be outlined as follows.

1.1. Parametric regressions with measurement errors

The measurement error problem in parametric regression has been reviewed in Fuller (1987) for linear regression and in Carroll et al. (1995) for nonlinear regression. There are many and extensive studies involving the measurement error problem (Carroll et al., 1984; Pierce et al., 1992; Prentice, 1992; Rocke and Durbin, 2001; Black et al., 2003). Studies of childhood growth data also show interesting measurement error structure because the measurement error has not only additive structure but also multiplicative term.

For this problem, Bayesian parametric regressions have been considered, which can be classified as a structural method for the measurement error problem. The Bayesian parametric regression is extended to Bayesian nonparametric regressions in subsequent two chapters under a generalized linear model.

1.2. Bayesian nonparametric regression on continuous response with measurement errors

As a more flexible regression analysis, a nonparametric curve fitting to the childhood growth data can be considered. Traditional nonparametric regression has been reviewed in Eubank (1999), and the generalized additive model along with the Bayesian smoothing spline has been described in Hastie and Tibshirani (1990). There are also many other studies of the Bayesian nonparametric regression analysis including Wecker and Ansley (1983); Carter and Kohn (1994); Denison et al. (1997); Hastie and Tibshirani (1998).

Under the generalized additive model, by applying the Bayesian backfitting algorithm to the Bayesian nonparametric regression studied by Berry et al. (2002), their nonparametric regression has been extended to a two-dimensional covariates space, where covariates have measurement errors.

1.3. Bayesian nonparametric regression on binary response with measurement errors

Nonparametric regression on binary response has been studied by Diaconis and Freedman (1993); Albert and Chib (1993); Neal (1997); Wood and Kohn (1998); Shively et al. (1999); Qian et al. (2000). As the extension of the parametric logistic regression studied by Wang et al. (1999) and the nonparametric regression on the single covariate space, the nonparametric version of logistic regression on a two-dimensional covariate space with measurement error has been established.

For two covariates, a generalized additive model has been considered. In addition, latent variables for each component of the additive logit link function has been proposed with the idea of Holmes and Mallick (2003), to relieve the computational burden.

1.4. Bayesian nonparametric regression of outcome-dependent follow-up times

In distinction from the previous case, longitudinal measurements may have features of the covariate. For example, follow-up times may depend on the previous measurements. As Lipsitz et al. (2002); Fitzmaurice et al. (2003) have studied, the design points of longitudinal measurements may depend on the response of preceding measurement in each individual.

In this case, the correct specification of correlation structure is very important in the estimation of the marginal effect of covariates over individuals. By allowing different correlation coefficients for each subject in the covariance structure, Bayesian parametric as well as nonparametric regressions have been performed. This model is denoted Model-0.

1.5. Bayesian nonparametric regression under flexible model of outcome-dependent follow-up times

As a more general idea, say Model-1, in outcome-dependent follow-up, a more flexible model has been considered for the follow-up times. Although Model-0 in section 1.4 assumes distribution of follow-up times, there is no direct association between the follow-up times and the regression function. By utilizing a subject-specific random effect, the regression model is associated with the follow-up times, under Model-1.

1.6. Overview

There can be various types of longitudinal measurements. Among them, this dissertation has dealt with two cases. The first case occurs when the longitudinal measurements bring measurement errors into the primary model, and hence it leads to the measurement error problems. In the second case, the longitudinal measurements do not lead to measurement errors, but it makes the estimation of the primary model very dependent on the correlation structure of response.

CHAPTER II

PARAMETRIC REGRESSION ANALYSIS WHEN COVARIATES ARE SUBJECT-SPECIFIC PARAMETERS IN A RANDOM EFFECTS MODEL FOR LONGITUDINAL MEASUREMENTS

2.1. Introduction

The regression model often associates the response variable with longitudinal measurements of certain variables. A common way is to use the subject-specific long-term averages of longitudinal measurements as covariates. However in several applications we have to use rate (or other measurements) over time rather than the average as the major risk factor. In this problem, the covariates considered are not observable so the estimated covariate will lead to measurement error problems.

The motivation for this work is a study on an adult obesity (Whitaker et al., 1997). In the health study, child growth is monitored by recording heights and weights over time, among other measurements. The main interest in the study is the extent to which longitudinal growth data from childhood can predict the likelihood of obesity in adulthood. At a given age, the growth data of each child consists of body mass index (BMI) defined by weight/height^2 (kg/m^2). The usual way to examine the data is to obtain summary information from longitudinal BMI measurements over time and further use the information to investigate the association with adult obesity. Wang and Pepe (2000) used the long term average of BMI over a period of childhood. Wang et al. (1999) considered a more general approach to extract the summary information as regression coefficients based on longitudinal childhood growth data. Then they used multiple linear logistic regression on adulthood obesity with the previously obtained regression coefficients as covariates. They showed that "naive" implementation of this model by substituting subject-specific ordinary least squares estimates of the random effects in the primary generalized linear model yields biased inferences on its parameters. Thus viewing it as a measurement error problem, they considered regression calibration (Carroll et al., 1995), where the random effects are replaced in the primary model by estimated best linear unbiased predictors from the fit of the mixed model, which reduces but does not completely eliminate bias. Wang et al. (1999) proposed a pseudo-expected estimating equation (EEE) approach, which requires numerical integration to compute the conditional expectations and they developed an approximate-EEE to circumvent this problem.

All of these approaches have two components: the first one contains repeated observed measurements, which are assumed to follow a linear random effects model. The second component is the primary regression where the random coefficients of the random effects model are covariate variables. Because the random coefficients are not observable, the measurement error is inevitable. We consider Bayesian parametric logistic regression model as the second component. All of these approaches depend on the assumption that the relationship in the primary regression between the response and the covariates under transformation by a link function is linear.

In the parametric logistic regression, the measurement error brings an attenuation problem (Carroll et al., 1995). The added strengths of the Bayesian approach in this problem are (i) a unified hierarchical model to accommodate all the uncertainties and (ii) automatic adjustment of bias due to measurement error.

This chapter is the motivation to consider Bayesian nonparametric regressions. Section 2.2 describes a measurement error model and a primary model. Bayesian frameworks are explained in section 2.3. Section 2.4 presents a simple example of the application of the Bayesian parametric regression.

2.2. Model

Let Y_i be the outcome variable for the ith subject, $i = 1, \ldots, n$, and $\boldsymbol{W}_i = (W_{i1}, \ldots, W_{im_i})^T$ are the longitudinal measurements of a continuous variable at times $t_{i1}, \ldots, t_{im_i}$. In the first stage $\boldsymbol{W}_i$ follows a random effect model as

$$\boldsymbol{W}_i = \boldsymbol{D}_i \boldsymbol{X}_i + \boldsymbol{U}_i$$

where $\boldsymbol{D}_i$ is a full rank $(m_i \times q)$ design matrix; and $\boldsymbol{U}_i = (U_{i1}, \ldots, U_{im_i})^T$ are within-subject errors reflecting uncertainty in measuring $\boldsymbol{W}_i$, independently and identically with mean zero and variance σ_u^2, i.e., $\boldsymbol{U}_i \sim N(\boldsymbol{0}, \sigma_u^2 \boldsymbol{I}_{m_i})$, where $\boldsymbol{I}_l$ is the identity matrix of dimension $l \times l$, independent of $\boldsymbol{X}_i$. $\boldsymbol{X}_i$ are $(q \times 1)$ random effects representing unobserved subject-specific features of the longitudinal profiles. A typical example is that the jth row of $\boldsymbol{D}_i$ is $(1, t_{ij})$ with $q = 2$ and $\boldsymbol{X}_i = (X_{i1}, X_{i2})^T$ yields a linear random coefficient model for the longitudinal data representing the subject specific initial exposure level (intercept) and the rate of change (slope).

In the next stage, the primary regression model is a parametric regression. We consider two kinds of responses: (i) continuous and (ii) binary. Let regression function $\eta_i = \boldsymbol{X}_i^T \boldsymbol{\beta}$. For the continuous response, the data is assumed to be Gaussian, so the conditional distribution of Y_i given $\boldsymbol{X}_i$ follows normal distribution with mean η_i and variance σ_z^2. For the binary response, we consider logistic regression, so the conditional distribution of Y_i given $\boldsymbol{X}_i$ follows Bernoulli distribution with success probability p_i and $\text{logit}(p_i) = \eta_i$, where $\text{logit}(v) = \{1 + \exp(-v)\}^{-1}$. All variables are independent across i. It is further stipulated that $\boldsymbol{W}_i$ is a surrogate for $\boldsymbol{X}_i$ such that the distribution of $(Y_i | \boldsymbol{W}_i, \boldsymbol{X}_i)$ is that of $(Y_i | \boldsymbol{X}_i)$ independent of $\boldsymbol{W}_i$.

For non-Gaussian data, it is well known that conjugate priors do not exist for the regression coefficients. The computations are then potentially much harder particularly with measurement error. This difficulty is due to a possibly strong posterior correlation between the parameters. We explore the use of a random residual component with small variance σ_z^2 within the model as in Holmes and Mallick (2003). We extend the model by introducing latent variables as

$$\begin{aligned} Y_i &\overset{ind}{\sim} B(p_i), \qquad \text{logit}(p_i) = Z_i, \quad i = 1, \ldots, n, \\ Z_i &= \eta_i + \epsilon_i, \qquad \epsilon_i \overset{iid}{\sim} N(0, \sigma_z^2), \quad i = 1, \ldots, n, \end{aligned}$$

where $\eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$ and $\text{logit}(v) = \{1 + \exp(-v)\}^{-1}$. We assume X_{il} are independent of each other, and have means, μ_{x_l} and variances, $\sigma_{x_l}^2$, $l = 1, 2$.

2.3. Bayesian regression

A. Regression for continuous response

We consider multiple regression for the continuous response. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$, $\boldsymbol{X}_j = (X_{1j}, \ldots, X_{nj})^T$, $j = 1, 2$, and $\boldsymbol{X} = (\mathbf{1}, \boldsymbol{X}_1, \boldsymbol{X}_2)$. For nuisance parameters we assume conjugate priors such that $\sigma_z^2 \sim IG(A_z, B_z)$, $\mu_{x_l}^2 \sim N(A_{m_l}, B_{m_l})$, $\sigma_{x_l}^2 \sim IG(A_{x_l}, B_{x_l})$, $l = 1, 2$, and $\sigma_u^2 \sim IG(A_u, B_u)$. Assuming uniform prior for $\boldsymbol{\beta}$, the joint conditional of $(\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{X})$ are proportional to the following.

$$\begin{aligned} [\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{X}|\cdot] &\propto (\sigma_z^2)^n \exp\left\{-\frac{1}{2\sigma_z^2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\right\} \\ &\times \exp\left\{-\frac{1}{2\sigma_{x_1}^2}\sum_{i=1}^n (X_{i1} - \mu_{x_1})^2 - \frac{1}{2\sigma_{x_2}^2}\sum_{i=1}^n (X_{i2} - \mu_{x_2})^2\right\} \\ &\times \exp\left\{-\frac{1}{2\sigma_u^2}\sum_{i=1}^n \sum_{k=1}^{m_i} (W_{ik} - X_{i1} - X_{i2}t_{ik})^2\right\} \end{aligned}$$

We can generate $\boldsymbol{\beta}$ from the following full conditional.

$$[\boldsymbol{\beta}|\boldsymbol{Y}, \cdot] \sim N\left[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}, (\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma_z^2\right]$$

We can utilize Gibbs procedure (Geman and Geman, 1984) to generate $\boldsymbol{\beta}$ and other nuisance parameters. From the joint distribution, we can also derive the full conditional for $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ by the square completion such that

$$\begin{aligned}
[X_{i1}|X_{i2},\cdot] &\overset{ind}{\sim} N\left(\frac{A_{i1}}{B_{i1}},\frac{1}{B_{i1}}\right), \qquad i=1,\ldots,n,\\
[X_{i2}|X_{i1},\cdot] &\overset{ind}{\sim} N\left(\frac{A_{i2}}{B_{i2}},\frac{1}{B_{i2}}\right), \qquad i=1,\ldots,n,
\end{aligned}$$

where A_{ij} and B_{ij}, $i = 1,\ldots,n; j = 1,2$, can be summarized as follows. For $i = 1,\ldots,n$,

$$\begin{aligned}
A_{i1} &= \frac{\beta_1^2}{\sigma_z^2}+\frac{1}{\sigma_{x_1}^2}+\frac{m_i}{\sigma_u^2},\\
B_{i1} &= \frac{\beta_1}{\sigma_z^2}(Y_i-\beta_0-\beta_2X_{i2})+\frac{\mu_1}{\sigma_{x_1}^2}+\frac{1}{\sigma_u^2}\sum_{k=1}^{m_i}(W_{ik}-X_{i2}t_{ik}),\\
A_{i1} &= \frac{\beta_2^2}{\sigma_z^2}+\frac{1}{\sigma_{x_2}^2}+\frac{1}{\sigma_u^2}\sum_{k=1}^{m_i}t_{ik}^2,\\
B_{i2} &= \frac{\beta_2}{\sigma_z^2}(Y_i-\beta_0-\beta_1X_{i1})+\frac{\mu_2}{\sigma_{x_2}^2}+\frac{1}{\sigma_u^2}\sum_{k=1}^{m_i}t_{ik}(W_{ik}-X_{i1}).
\end{aligned}$$

B. Regression for binary response

For the ith response Y_i, $i = 1,\ldots,n$, we adopt Z_i as a latent variable. Let $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$, $\boldsymbol{X}_j = (X_{1j},\ldots, X_{nj})^T$, $j = 1,2$, $\boldsymbol{X} = (\mathbf{1}, \boldsymbol{X}_1, \boldsymbol{X}_2)$, and $\boldsymbol{Z} = (Z_1,\ldots, Z_n)^T$. We assume the same conjugate priors for nuisance parameters, σ_z^2, $\mu_{x_l}^2$, $\sigma_{x_l}^2$, $l = 1,2$, and σ_u^2, and uniform prior for $\boldsymbol{\beta}$, as continuous response. Then the joint conditional of $(\boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{X})$ given nuisance parameters are proportional to the following:

$$\begin{aligned}
[\boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{X}|\cdot] &\propto \prod_{i=1}^{n} \left\{p_i^{Y_i}(1-p_i)^{1-Y_i}\right\} \exp\left\{-\frac{1}{2\sigma_z^2}(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta})\right\} \\
&\times \exp\left\{-\frac{1}{2\sigma_{x_1}^2}\sum_{i=1}^{n}(X_{i1}-\mu_{x_1})^2 - \frac{1}{2\sigma_{x_2}^2}\sum_{i=1}^{n}(X_{i2}-\mu_{x_2})^2\right\} \\
&\times \exp\left\{-\frac{1}{2\sigma_u^2}\sum_{i=1}^{n}\sum_{k=1}^{m_i}(W_{ik}-X_{i1}-X_{i2}t_{ik})^2\right\}
\end{aligned}$$

By customizing the idea of Holmes and Held (2005) to the our parametric regression, we can generate $\boldsymbol{Z}$ and $\boldsymbol{\beta}$ jointly from their joint full conditional as with the continuous response. Let $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$. Then the conditional of $\boldsymbol{\beta}$ and the marginal conditional of $\boldsymbol{Z}$ integrated over $\boldsymbol{\beta}$ can be described as follows:

$$\begin{aligned}
[\boldsymbol{\beta}|\boldsymbol{Z}, \cdot] &\sim N\left[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Z}, (\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma_z^2\right] \\
[\boldsymbol{Z}|\cdot] &\propto \prod_{i=1}^{n}\left\{p_i^{Y_i}(1-p_i)^{1-Y_i}\right\} \exp\left\{-\frac{1}{2\sigma_z^2}\boldsymbol{Z}^T(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Z}\right\}
\end{aligned}$$

Because the full conditional of $\boldsymbol{Z}$ follows n-dimensional multivariate normal distribution, it is hard to generate random numbers of $\boldsymbol{Z}$. To cope with the high dimensional problem, we use the Gibbs sampling intensively. Let h_{ij} be the (i, j)th element of $\boldsymbol{I} - \boldsymbol{H}$. Then the ith element of $\boldsymbol{Z}$ can be generated by the Metropolis-Hasting's algorithm based on the distribution proportional to the following:

$$[Z_i|\boldsymbol{Z}_{-i}, \cdot] \propto p_i^{Y_i}(1-p_i)^{1-Y_i} N\left(Z_i - \frac{\sum_{j=1}^{n} h_{ij}Z_j}{h_{ii}}, \frac{\sigma_z^2}{h_{ii}}\right)$$

From the joint distribution, we can also derive the full conditional for $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ by the square completion such that

$$\begin{aligned}
[X_{i1}|X_{i2}, \cdot] &\overset{ind}{\sim} N\left(\frac{A_{i1}}{B_{i1}}, \frac{1}{B_{i1}}\right), \qquad i = 1, \ldots, n, \\
[X_{i2}|X_{i1}, \cdot] &\overset{ind}{\sim} N\left(\frac{A_{i2}}{B_{i2}}, \frac{1}{B_{i2}}\right), \qquad i = 1, \ldots, n,
\end{aligned}$$

where A_{ij} and B_{ij}, $i = 1, \ldots, n; j = 1, 2$, can be summarized as follows. For $i = 1, \ldots, n$,

$$
\begin{aligned}
A_{i1} &= \frac{\beta_1^2}{\sigma_z^2} + \frac{1}{\sigma_{x_1}^2} + \frac{m_i}{\sigma_u^2}, \\
B_{i1} &= \frac{\beta_1}{\sigma_z^2}(Z_i - \beta_0 - \beta_2 X_{i2}) + \frac{\mu_1}{\sigma_{x_1}^2} + \frac{1}{\sigma_u^2}\sum_{k=1}^{m_i}(W_{ik} - X_{i2}t_{ik}), \\
A_{i1} &= \frac{\beta_2^2}{\sigma_z^2} + \frac{1}{\sigma_{x_2}^2} + \frac{1}{\sigma_u^2}\sum_{k=1}^{m_i} t_{ik}^2, \\
B_{i2} &= \frac{\beta_2}{\sigma_z^2}(Z_i - \beta_0 - \beta_1 X_{i1}) + \frac{\mu_2}{\sigma_{x_2}^2} + \frac{1}{\sigma_u^2}\sum_{k=1}^{m_i} t_{ik}(W_{ik} - X_{i1}).
\end{aligned}
$$

We can utilize Gibbs procedure to generate $\boldsymbol{\beta}$, $\boldsymbol{Z}$, and other parameters similar to the continuous response case.

2.4. Parametric regression analysis on cardiotoxic data

We applied our method (when $q = 2$) to the childhood growth data, which is also used by Wang et al. (1999) and Whitaker et al. (1997). The data was collected from 330 subjects who had at least three measurements of BMI z-score (BMI-z) between ages 2.75 and 5.25. The adulthood BMI is calculated by taking average BMI over ages 21 through 29. We assess the extent to which the initial BMI-z value and the rate of change of the BMI-z value are predictive of adulthood BMI value (or obesity). The initial value and the rate of change are, respectively, the intercept and the slope of the simple linear regression on the childhood BMI-z with the monitoring age, which are not observable covariates of our linear model. We have the actual continuous data which is the observed adulthood BMI response for each subject. We also have the binary data where the adulthood BMI is dichotomized as obese or not using the critical values: 27.8 for male and 27.3 for female. Twenty samples of childhood BMIs are shown in Figure 1.

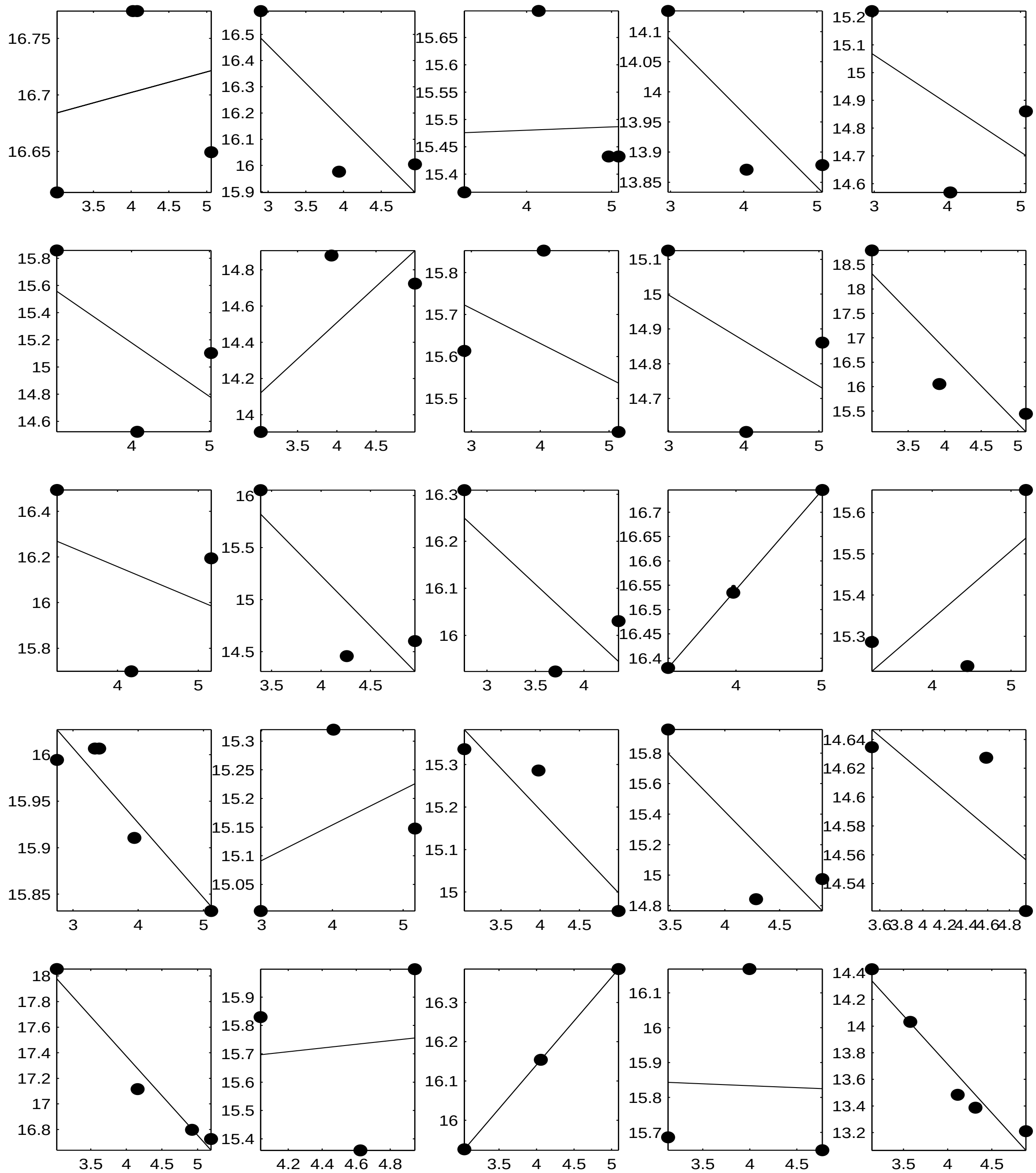

Fig. 1. Samples of childhood BMIs. *Each cell shows longitudinal measurement of BMIs from a child. The intercept and the slope of line are used as covariates of the corresponding individual, in the naive method.*

First for the continuous response, we considered regression on the adulthood BMI with the intercept and the slope of childhood BMI-z as covariates. We tried Bayesian linear regression. We also compared the results with the naive method using linear models. We performed traditional multiple linear regression using regression calibration method. In regression calibration, we imputed the conditional expectations of unknown covariates given response and the error-prone observations, $E(X_{il}|W_{i1}, \ldots, W_{im_i}, t_{i1}, \ldots, t_{im_i}, Y_i)$, in place of the unobserved covariate X_{il}, for $i = 1, \ldots, n$; $l = 1, 2$. Following Wang et al. (1999) the unknown parameters in the conditional distribution were estimated by the method of moment estimator and the regression parameters were estimated using the maximum likelihood approach. To check the performance, the mean residual sum of squares was considered, $\overline{\text{RSS}}=\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \widehat{f}_1(X_{i1}) - \widehat{f}_2(X_{i2})\}^2$. As shown in Table I, Bayesian linear regression outperformed all other methods.

Next we analyze the binary data using a logistic regression model and have observed very similar results. Table II shows Bayesian logistic regression performs better than all other competitors in terms of DIC. In summary for both the situations (continuous and binary), Bayesian parametric regression works better than others. In the subsequent two chapters, the above parametric regressions are extended to nonparametric regressions: Chapter III for continuous response and Chapter IV for non-continuous response.

Table I. Adulthood BMI: Performance of various parametric regression methods with continuous response.

Method	$\overline{\text{RSS}}$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
Regression Calibration	18.81	24.23	2.30	6.04
Naive Method	18.79	24.31	2.28	4.22
Bayes Method	17.36	24.22	2.22	5.95

NOTE: Bayes method outperformed the naive method and the regression calibration method $\overline{\text{RSS}}$.

Table II. Adulthood obesity: Performance of various regression methods with binary response.

Method	DIC	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
Regression Calibration	241.22	-2.04	1.05	4.49
Naive Method	248.20	-2.04	0.99	2.58
Bayes Method	240.33	-2.46	1.21	5.32

NOTE: The Bayesian nonparametric regression shows the best performance in DIC and the largest values of estimated coefficients.

CHAPTER III

NONPARAMETRIC REGRESSION ANALYSIS FOR CONTINUOUS RESPONSE WHEN COVARIATES ARE SUBJECT-SPECIFIC PARAMETERS IN A RANDOM EFFECTS MODEL FOR LONGITUDINAL MEASUREMENTS

3.1. Introduction

When the response is a continuous variable, we construct nonparametric regression model to associate the response variable with risk factors from longitudinal measurements. We consider an application to use rate (or other measurements) over time, as well as the average as the major risk factors. The multi-dimensional covariate space can make the nonparametric regression difficult. Further, the covariates considered are not observable so the estimated covariate will lead to measurement error problems.

In the health study of body mass index (BMI) defined by weight/height^2 (Whitaker et al., 1997), adulthood BMI is used to determine adulthood obesity. From BMI over a period of childhood, as Wang et al. (1999) did, we consider linear regression on BMI with age when it is monitored. Then, the regression coefficients will serve to extract the childhood summary information of each individual. We utilize those random coefficients as covariates to predict the adulthood BMI nonparametically. Because true regression coefficients are not observable, the measurement errors are involved in the regression. Hence, the primary model becomes nonparametric regression model with error-prone covariates. The presence of several covariates makes this nonparametric model more complex. Recently, advances in computer power have allowed statisticians to consider richer classes of models that were previously computationally prohibitive. We propose a Bayesian model based on smoothing spline (Eubank, 1999) to handle the nonlinearity. The two components of the modeling procedure, the measurement error model and the primary model, fit within a hierarchical Bayes model in a unified way. Under multi-covariate situation, the extension can be achieved by the backfitting algorithm which enables us to utilize the regressions in one dimension to yield the regression in the multidimensional covariate space. A Bayesian smoothing spline is used to estimate the unknown functions and explore MCMC algorithms to generate the fitted curves in multi-dimensional covariate space.

The nonparametric regression problem here is much more complicated than the usual additive model regression because the covariates under consideration are not directly observable. For example, in the childhood growth data, covariates are estimated by regression coefficients from the simple linear regression of childhood BMI with age. In the parametric logistic regression, as mentioned in Chapter II, measurement error creates an attenuation problem (Carroll et al., 1995). Berry et al. (2002) developed nonparametric regression using smoothing splines for a single covariate with measurement error. We propose a nonparametric regression for Gaussian response with multiple covariates measured with error. As mentioned earlier, A Bayesian smoothing spline approach has been examined to estimate the unknown functions. The advantage of the Bayesian approach in this problem are (i) a unified hierarchical model to accommodate all the uncertainties, (ii) an automatic adjustment of bias due to measurement error, and (iii) an automatic selection of the smoothing parameters in the additive model. Measurement error has large effects on both bias and variance and a smoothing parameter that is optimal for correctly measured covariate may be far from optimal in the presence of measurement error. An optimal choice of a smoothing parameter is hard in a measurement error problem and could be even harder in an additive model framework. The Bayesian approach automatically chooses the smoothing parameters for each covariate.

Subsequent section reminds a measurement error model and shows primary model of a Bayesian nonparametric regression (Section 3.2). The next sections explains a Bayesian nonparametric regression (Section 3.3). Last two sections demonstrate a Bayesian nonparametric regression with simulated data and the BMI data (Sections 3.4 and 3.5).

3.2. Model

The measurement error model is same as the parametric regression, but the primary regression model is based on natural cubic smoothing splines. For the ith subject, $i = 1, \ldots, n$, supposed longitudinal measurements $\boldsymbol{W}_i = (W_{i1}, \ldots, W_{im_i})^T$ are observed at $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{im_i})^T$. Denoting a full rank design matrix $\boldsymbol{D}_i = (\mathbf{1}_{m_i}, \boldsymbol{t}_i)$, where $\mathbf{1}_{m_i}$ is a $m_i \times 1$ vector of ones. Then, the measurement error model for covariates $\boldsymbol{X}_i = (X_{i1}, X_{i2})^T$ is described by

$$\boldsymbol{W}_i = \boldsymbol{D}_i \boldsymbol{X}_i + \boldsymbol{U}_i,$$

where $\boldsymbol{U}_i = (U_{i1}, \ldots, U_{im_i})^T$ are normal random errors with mean zero and variance $\sigma_u^2 \boldsymbol{I}_{m_i}$ and $\boldsymbol{I}_{m_i}$ is a $m_i \times m_i$ identity matrix. Note that, just for the simplicity, we assume independent identical variance of measurement errors over longitudinal measurements and over all subjects. Further, because we suppose non-differential measurement error, $\boldsymbol{X}_i$ and $\boldsymbol{U}_i$ are supposed to be independent of each other.

The primary model is a generalized additive model of two natural cubic splines (NCSs). Let Y_i be the ith response and f_l be the smoothing spline for the previous covariates X_{il}, $l = 1, 2$. Then the primary model will be expressed by

$$Y_i = f_1(X_{i1}) + f_2(X_{i2}) + \epsilon_i, \qquad i = 1, \ldots, n,$$

where ϵ_i are independent Gaussian errors with zero mean and a constant variance of σ_z^2.

3.3. Bayesian smoothing spline with measurement error

We consider the Bayesian natural cubic smoothing spline (NCS) to model the unknown functions (Hastie and Tibshirani, 1990; Berry et al., 2002). For the sake of simplicity, we first explain the NCS for a single f_l, the function corresponding to the lth covariate.

A. Smoothing spline

If the covariate ($X_{il},\ l = 1, 2;\ i = 1, \ldots, n$) is observable and response ($Y_i,\ i = 1, \ldots, n$) is continuous then NCS defines the spline basis functions with a knot at each distinctive value of the covariate X_l. The estimate of f_l minimizes the following penalized sum of squares over all possible NCS:

$$\sum_{i=1}^{n} \{Y_i - f_l(X_{il})\}^2 + \alpha_l \int_{\min(X_{il})}^{\max(X_{il})} \left\{ f_l^{''}(t) \right\}^2 dt,$$

where $f_l^{''}(\cdot)$ is the second derivative of $f_l(\cdot)$ and positive valued α_l is the smoothing parameter. Note that the NCS is a cubic smoothing spline with the boundary condition such that $f_l'(\cdot) = 0$ and $f_l''(\cdot) = 0$. Let $N_i(\cdot)$ be the i^{th} NCS basis function with knots $\{X_{1l}, \ldots, X_{nl}\}$, and $\boldsymbol{N} = \{N_j(X_{il})\}_{i,j=1,\ldots,n}$ be an $n \times n$ nonsingular natural splines basis matrix, and $\boldsymbol{\Omega} = \left[\int \left\{ N_i^{''}(t) N_j^{''}(t) \right\}^2 dt \right]_{i,j=1,\ldots,n}$. Since the NCS can be described by $f_l(X_{il}) = \sum_{j=1}^{n} c_j N_j(X_{il})$ with coefficients $c_j,\ j = 1, \ldots, n$, we can rewrite the above penalized sum of squares as

$$(\boldsymbol{Y} - \boldsymbol{N}\boldsymbol{c})^T (\boldsymbol{Y} - \boldsymbol{N}\boldsymbol{c}) + \alpha_l \boldsymbol{c}^T \boldsymbol{\Omega} \boldsymbol{c},$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ and $\boldsymbol{c} = (c_1, \ldots, c_n)^T$. Hence, if the smoothing parameter is given, the NCS is similar to the ridge regression with the hat matrix $\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{N} + \alpha_l\boldsymbol{\Omega})^{-1}\boldsymbol{N}^T$. The choice of a smoothing parameter is critical to determining the roughness of the estimated curve and can be achieved by the generalized cross validation or predictive risk estimator. Detailed procedures for the NCS can be found in Eubank (1999); Hastie and Tibshirani (1990). In our Bayesian hierarchical model we assume the smoothing parameter as an unknown and treat the uncertainty of the model through a prior distribution on the smoothing parameter. Model uncertainty relates to the fact that many different NCS models may offer nearly equally plausible representation of the data. Rather than using a single plug in estimation of the smoothing parameter, we will perform model mixing with respect to the smoothing parameter.

In the Bayesian approach, the function f_l is also treated as a random variable and assigned a prior density proportional to the partially improper Gaussian process (Raghavan and Cox, 1998; Hastie and Tibshirani, 2000) which is proportional to the following:

$$\tau_l^{\frac{n-2}{2}} \exp\left\{-\frac{\tau_l}{2}\boldsymbol{f}_l^T\boldsymbol{K}_l\boldsymbol{f}_l\right\},$$

where $\tau_l = \frac{\alpha_l}{\sigma_z^2}$ and σ_z^2 is the variance of responses. K_l is defined as satisfying $f_l^T\boldsymbol{K}_l f_l = \int \left\{f_l''(t)\right\}^2 dt$. Eubank (1999, p. 244) explained a method to construct the matrix $\boldsymbol{K}_l$ for the NCS. Another covariance structure for the Bayesian nonparametric curve, such as the state space model, can be found at Carter and Kohn (1994) and Wecker and Ansley (1983).

In a generalized additive model, we consider q NCSs associated with q covariates. For each NCS, a partially improper prior with a corresponding τ_l and $\boldsymbol{K}_l$, $l = 1, \ldots, q$ is assigned independently. Under the linear model representation of each f_l, we can calculate the full conditional distributions of f_ls in the Gibbs sampling framework. This is equivalent to a backfitting algorithm and known as a Bayesian back-fitting procedure (Hastie and Tibshirani, 2000). For additive models, problem arises from the identifiability of the mean levels of the unknown functions. To ensure identifiability, the functions f_l are constrained to have zero means, i.e. $\{\text{range}(x_l)\}^{-1}\int f_l(x_l)dx_l = 0$. This can be incorporated into estimation via MCMC by centering the function f_l about the mean of f_j in every iteration of the sampler. To ensure the posterior not being changed, the subtracted means are added to the intercept. Next we will develop a unified Bayesian hierarchical model combining the NCS model with measurement errors.

B. Bayesian hierarchical model

To develop the Bayesian hierarchical model, we need to assign prior distributions for all the unknowns. For a convenience of notation, let A and B with subscripts be known constants. There is no conjugate prior for $\boldsymbol{X}_l = (X_{1l}, \dots, X_{nl})^T$, $l = 1, \dots, q$, which could ease the computational burden. As mentioned earlier, we assume $\boldsymbol{X}_l$ follows a normal distribution with mean μ_{x_l} and variance $\sigma^2_{x_l}$, $l = 1, \dots, q$ and are independent of each other. Further we assume conjugate priors for μ_{x_l} and $\sigma^2_{x_l}$ as $\mu_{x_l} \sim N(A_{m_l}, B_{m_l})$, $\sigma^2_{x_l} \sim IG(A_{x_l}, B_{x_l})$, $l = 1, \dots, q$. We may assign more complicated distributions like the mixture of normals (Carroll et al., 1999) but as $\boldsymbol{X}_l$ continually changes throughout the MCMC algorithm, updating this mixture at every iteration can make the algorithm very slow. We also assume a conjugate prior for the variance of the Gaussian response $\boldsymbol{Y}{=}(Y_1, \dots, Y_n)^T$ as $\sigma^2_z \sim IG(A_z, B_z)$, and for the variance of measurement errors $\sigma^2_u \sim IG(A_u, B_u)$. The prior for τ_l is a Gamma distribution $\tau_l \sim G(A_{t_l}, B_{t_l})$ where τ_l is defined by $\frac{\alpha_l}{\sigma^2_z}$. The relationship between parameters is described in the DAG (directed acyclic graph) in Figure 2 for $q = 2$.

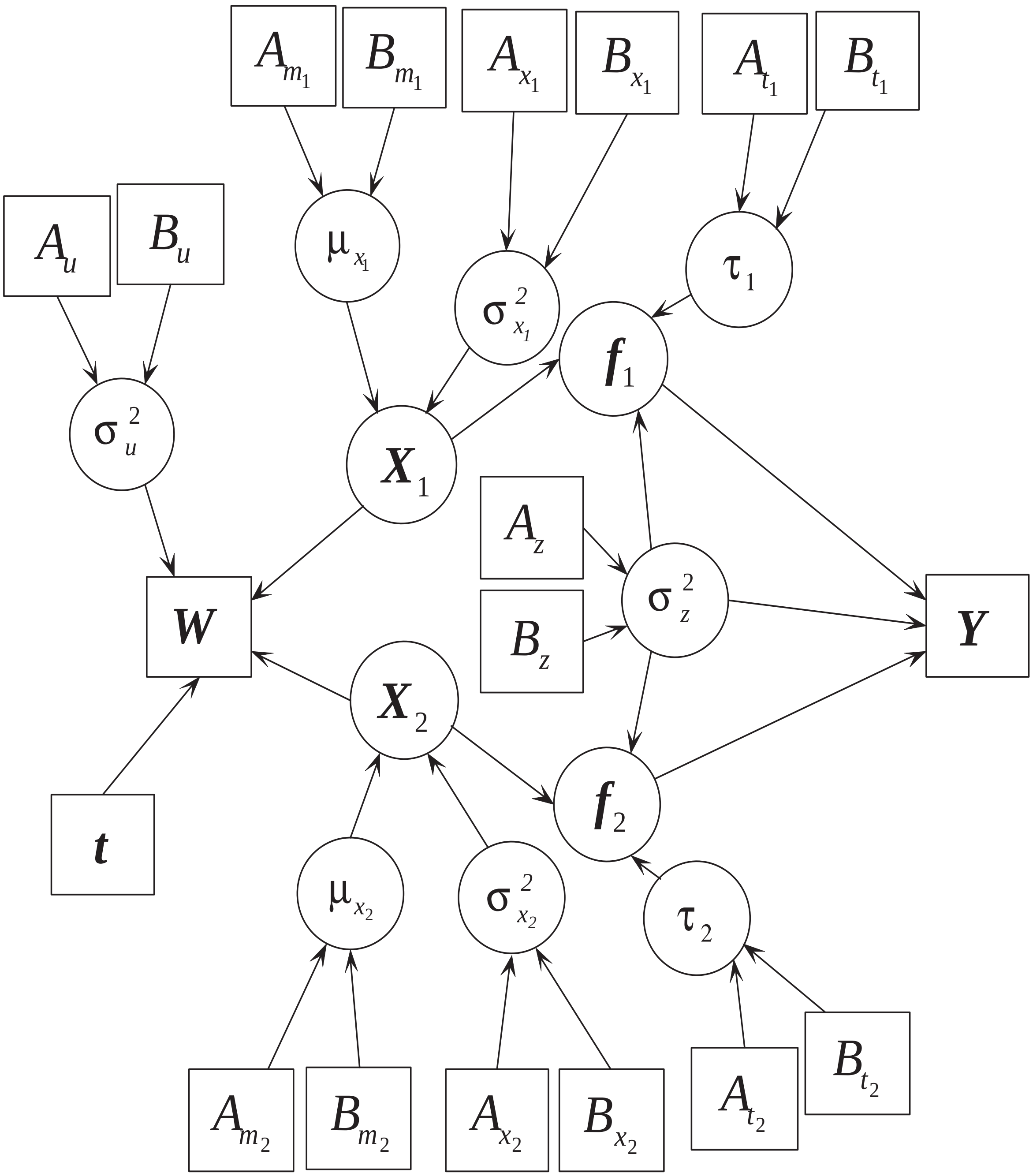

Fig. 2. Directed acyclic graph for hierarchical Bayesian model with continuous response. *Variables in the rectangles are observable or given, but variables in the circles are not observable. The information of* $\boldsymbol{X}_1$ *and* $\boldsymbol{X}_2$ *is given by the combination of* $\boldsymbol{W}$ *and* $\boldsymbol{t}$*. Each covariate is associated with separate nonparametric curve* $\boldsymbol{f}_j$, $j = 1, 2$.

Using the model and prior distributions, we can obtain the joint posterior distribution of the unknowns, which is proportional to

$$
\begin{aligned}
\propto \quad & \left(\sigma_z^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma_z^2}\sum_{i=1}^{n}\left\{Y_i - \sum_{l=1}^{q} f_l(X_{il})\right\}^2\right] \\
\times \quad & \prod_{l=1}^{q}\left[\tau_l^{\frac{n-2}{2}} \exp\left\{-\frac{\tau_l}{2}\boldsymbol{f}_l^T\boldsymbol{K}_l\boldsymbol{f}_l\right\}\tau_l^{A_{t_l}-1}\exp\left\{-\frac{\tau_l}{B_{t_l}}\right\}\right. \\
& \quad \times \left(\sigma_{x_l}^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_{x_l}^2}\sum_{i=1}^{n}(X_{il}-\mu_{x_l})^2\right\} \\
& \quad \times \left(\sigma_{x_l}^2\right)^{-(A_{x_l}+1)} \exp\left\{-\frac{1}{\sigma_{x_l}^2 B_{x_l}}\right\} \\
& \quad \times B_{m_l}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2B_{m_l}}(\mu_{x_l}-A_{m_l})^2\right\} \\
& \quad \left.\times \left(\sigma_z^2\right)^{-(A_z+1)} \exp\left\{-\frac{1}{\sigma_z^2 B_z}\right\}\right] \\
\times \quad & \left(\sigma_u^2\right)^{-\frac{1}{2}\sum_{i=1}^{n} m_i} \exp\left\{-\frac{1}{2\sigma_u^2}\sum_{i=1}^{n}(\boldsymbol{W}_i-\boldsymbol{D}_i\boldsymbol{X}_i)^T(\boldsymbol{W}_i-\boldsymbol{D}_i\boldsymbol{X}_i)\right\} \\
\times \quad & \left(\sigma_u^2\right)^{-(A_u+1)} \exp\left\{-\frac{1}{\sigma_u^2 B_u}\right\},
\end{aligned}
$$

where $\boldsymbol{f}_l = [f_l(X_{1l}), \ldots, f_l(X_{nl})]^T$.

Let $\boldsymbol{A}_l(\alpha_l) = (\boldsymbol{I} + \alpha_l\boldsymbol{K}_l)^{-1}$, for $l = 1, \ldots, q$, and $\boldsymbol{R}_{f_l}$ denotes the residual of additive model by excluding function $\boldsymbol{f}_l$. Further let residual sums of squares for each variable be $RSS_y = \sum_{i=1}^{n}\{Y_i - \sum_{l=1}^{q} f_l(X_{il})\}^2$, $RSS_{x_l} = \sum_{i=1}^{n}(X_{il} - \mu_{x_l})^2$, $l = 1, \ldots, q$, and residual sum of squares for each element be such that $RSS_{w_i} = (\boldsymbol{W}_i - \boldsymbol{D}_i\boldsymbol{X}_i)^T(\boldsymbol{W}_i - \boldsymbol{D}_i\boldsymbol{X}_i)$, $RSS_{y_i} = \{Y_i - \sum_{l=1}^{q} f_l(X_{il})\}^2$, and $RSS_{x_{il}} = (X_{il} - \mu_{x_{il}})^2$, $l = 1, \ldots, q;\ i = 1, \ldots, n$.

To execute Gibbs sampling, we need all the full conditional distributions which are given below:

$$
\begin{aligned}
\boldsymbol{f}_l|\cdot &\overset{ind}{\sim} N\left[\boldsymbol{A}_l(\alpha_l)\boldsymbol{R}_{f_l}, \boldsymbol{A}_l(\alpha_l)\sigma_z^2\right], \\
\tau_l|\cdot &\overset{ind}{\sim} G\left[\frac{n-2}{2}+A_{t_l}, \left(\frac{\boldsymbol{f}_l^T\boldsymbol{K}_l\boldsymbol{f}_l}{2}+\frac{1}{B_{t_l}}\right)^{-1}\right], \\
\sigma_z^2|\cdot &\overset{ind}{\sim} IG\left[\frac{n}{2}+A_z, \left(\frac{RSS_y}{2}+\frac{1}{B_z}\right)^{-1}\right], \\
\sigma_{x_l}^2|\cdot &\overset{ind}{\sim} IG\left[\frac{n}{2}+A_{x_l}, \left(\frac{RSS_{x_l}}{2}+\frac{1}{B_{x_l}}\right)^{-1}\right], \\
\mu_{x_l}|\cdot &\overset{ind}{\sim} N\left[\left(\frac{\sum_{i=1}^n X_{il}}{\sigma_{x_l}^2}+\frac{A_{m_l}}{B_{m_l}}\right)\left(\frac{n}{\sigma_{x_l}^2}+\frac{1}{B_{m_l}}\right)^{-1}, \left(\frac{n}{\sigma_{x_l}^2}+\frac{1}{B_{m_l}}\right)^{-1}\right], \\
\sigma_u^2|\cdot &\sim IG\left[\frac{\sum_{i=1}^n m_i}{2}+A_u, \left(\frac{\sum_{i=1}^n RSS_{w_i}}{2}+\frac{1}{B_u}\right)^{-1}\right],
\end{aligned}
$$

for $l = 1, \ldots, q$, where $G[A, B]$ denotes gamma distribution with the mean AB, and $IG[A, B]$ indicates inverse gamma distribution with mean $\{(A-1)B\}^{-1}$. The full conditionals of X_{il}, $i = 1, \ldots, n$; $l = 1, \ldots, q$, are not of standard forms and their densities are proportional to the following:

$$
[X_{il}|\cdot] \overset{ind}{\propto} \exp\left\{-\frac{RSS_{z_{il}}}{2\sigma_z^2}-\frac{RSS_{x_{il}}}{2\sigma_{x_l^2}}-\frac{RSS_{w_i}}{2\sigma_u^2}\right\}, \qquad i = 1, \ldots, n;\ l = 1, \ldots, q.
$$

From the full conditionals, we can see that the covariates are involved almost everywhere. Hence, the measurement error will seriously affect the estimation of link functions, as well as other parameters. In the naive approach, we replace the unobserved $\boldsymbol{X}$ by the least squares estimator for each subject. That will create a biased estimate for the unknown functions, and we will compare the results of a full Bayes approach to this naive approach.

Based on the full conditional distributions, Gibbs sampler (Gelfand and Smith, 1990) can generate each parameter from the joint posterior distribution. As the conditional distributions of the covariates, X_{il}, $i = 1, \ldots, n$, $l = 1, \ldots, q$ are not of standard forms so they are generated via Metropolis-Hasting's algorithm within the Gibbs sampling. The full conditions of $\boldsymbol{f}_l$ are from multivariate normal distributions; hence the regular generation method requires inverses of big covariance matrices. For an efficient computation, we utilize Cholesky's decomposition with the backward and the forward substitutions. Details are provided in Appendix B.

We draw samples of functions from their joint posterior distribution and use the pointwise mean curves as the natural estimate of the regression functions. A sampling based method provides us the flexibility to calculate pointwise median, credible intervals, or any other functionals of these regression functions. In our method, we allow a separate smoothness parameter for each different regression function in the additive model setup, so we have the flexibility to estimate curves with different degrees of smoothness (complexity) with all the uncertainty measures. Though we are mainly concerned with estimation of the functions, inferences (posterior mean and credible intervals) can be done for mismeasured Xs, easily using the corresponding MCMC samples.

3.4. Simulation study

The comparison between the Bayesian method with the naive estimator was performed by a series of simulations with a continuous response. The gold standard is the estimated Bayesian nonparametric curve with true covariate values (mentioned as the "no error" case in the tables). For simulations, we tried 200 cases and 400 cases ($n = 200$ and $n = 400$) of longitudinal data along with continuous responses. In each case, covariates and measurement error structure followed the simulation scheme of Wang et al. (1999) with slightly more correlations between covariates. We first generated unobservable covariates from normal distributions such that $\boldsymbol{X}_i =$

$\binom{X_{i1}}{X_{i2}} \overset{iid}{\sim} N\left[\mathbf{0}, \begin{pmatrix} 1 & -.1 \\ -.1 & .25 \end{pmatrix}\right]$, $i = 1, \ldots, n$. Each case was assumed to have four replicates ($m_i = 4$) of W_{ij}, which were simulated from the model $W_{ij} = X_{i1} + X_{i2}t_{ij} + U_{ij}$, where $t_{ij} \overset{ind}{\sim} N(j-1, 0.1^2)$, $j = 1, \ldots, 4; i = 1, \ldots, 200$. We considered the bivariate additive model. As a complicated function, $f_1(x)$ was taken from the sine family, which is a slight modification of Berry et al. (2002), and as a simple function, $f_2(x)$ was taken from the quadratic family such that

$$\begin{aligned} f_1(x) &= \frac{5\sin(\pi x/2)}{1 + 2x^2\{\text{sign}(x) + 1\}} + 2, \\ f_2(x) &= -2x^2 + 1. \end{aligned}$$

Finally, the responses were generated using additive errors generated from the normal distribution such that $\epsilon_i \overset{iid}{\sim} N(0, 0.3^2)$, $i = 1, \ldots, n$.

For the Bayesian model, we assigned flexible hyper priors such as $\sigma_u^2 \sim IG(1, 1)$, $\tau \sim G(3, 1/100)$, $\mu_{x_1} = \mu_{x_2} \sim N(0, 10^2)$, and $\sigma_{x_1}^2 = \sigma_{x_2}^2 \sim IG(1, 1)$. For the variance of the response we also assigned flexible prior as $\sigma_z^2 \sim IG(1, 1)$.

To evaluate the performance of each estimator, we calculated mean squared error (MSE) from the evaluated values of the estimated functions, evaluated at 101 grid points in the interval [-2,2] for f_1 and [-1,1] for f_2.

We generated twenty simulated data sets with different measurement error variances. In each simulation, we collected 10000 Markov Chain Monte Carlo (MCMC) samples after 50000 burning iterations. We examined the effect of increased variance of measurement error (σ_u^2) and the increased sample size (n) on the performance of the estimator. Usually for fixed sample size, larger σ^2 (more measurement error) worsens the performance of the estimator (increase the MSE). The results are presented in Table III, which shows that the Bayesian method performed distinctly better than the naive method and close to the regression with true covariate values for all the

Table III. Table for simulated continuous data: Average MSE from 20 simulated data for each situation.

Situation	Link	Criterion	NoErr	Naive	Bayes
$n = 200$, $\sigma_u^2 = 0.49$	f_1	MSE	0.0090	1.4253	0.1818
		Bias^2	0.0007	1.3371	0.0393
		Var	0.0083	0.0883	0.1424
	f_2	MSE	0.0054	0.1716	0.0391
		Bias^2	0.0001	0.1083	0.0067
		Var	0.0053	0.0633	0.0325
$n = 200$, $\sigma_u^2 = 1.0$	f_1	MSE	0.0090	2.3403	0.6226
		Bias^2	0.0007	2.1988	0.1334
		Var	0.0083	0.1414	0.4891
	f_2	MSE	0.0054	0.3424	0.0904
		Bias^2	0.0001	0.2761	0.0231
		Var	0.0053	0.0663	0.0673
$n = 400$, $\sigma_u^2 = 1.0$	f_1	MSE	0.0056	2.2024	0.8297
		Bias^2	0.0004	2.1295	0.1325
		Var	0.0052	0.0729	0.6972
	f_2	MSE	0.0033	0.3570	0.0671
		Bias^2	0.0001	0.3094	0.0243
		Var	0.0032	0.0475	0.0428

NOTE: The Bayesian method shows much better performance than the naive method in terms of MSE for all the situations. Its performance is almost as good as knowing the true covariate values.

cases.

For the example with continuous response, we plotted the true curve, the posterior mean curve, and 95% pointwise credible intervals obtained from the MCMC samples in Figure 3. For both of the functions, the Bayes estimate is pretty close to the true one. We also overlay the naive estimate on the Bayes estimate. The Bayes estimate is distinctly better than the naive estimate.

3.5. Childhood growth data analysis for adulthood BMI

We applied nonparametric regression (when $q = 2$) to the childhood growth data, which is used in Chapter II. As with the continuous response, we considered regression on adulthood BMI with the intercept and slope of childhood BMI-z as unobservable covariates. We tried Bayesian nonparametric regression in a two dimensional covariate space. We also compared the results with the naive method using nonparametric models. In regression calibration, we substituted the conditional expectations of unknown covariates given response and the error-prone observations, $E(X_{il}|W_{i1}, \ldots, W_{im_i}, t_{i1}, \ldots, t_{im_i}, Y_i)$, in place of the unobserved covariate X_{il}, for $i = 1, \ldots, n$; $l = 1, 2$. Following Wang et al. (1999) the unknown parameters in the conditional distribution were estimated by the method of moment estimator and the regression parameters were estimated using the maximum likelihood approach. To check the performance, the mean residual sum of squares was considered, $\overline{\text{RSS}}=\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \widehat{f}_1(X_{i1}) - \widehat{f}_2(X_{i2})\}^2$. In Bayesian NCS, the Bayes method outperformed the naive estimator in $\overline{\text{RSS}}$ (17.20 vs. 1.56), although both of methods have less $\overline{\text{RSS}}$ than the best of the parametric regression methods (Bayes method, 17.36). Accordingly, in Figure 4, it is clear that both of the functions are not at all linear. Hence, in terms of exploring the nonlinear curves as well as to improve the fitting significantly, the nonparametric method is useful. In summary, the initial BMI-z value and the rate of change have significant nonlinear effects on adulthood BMI.

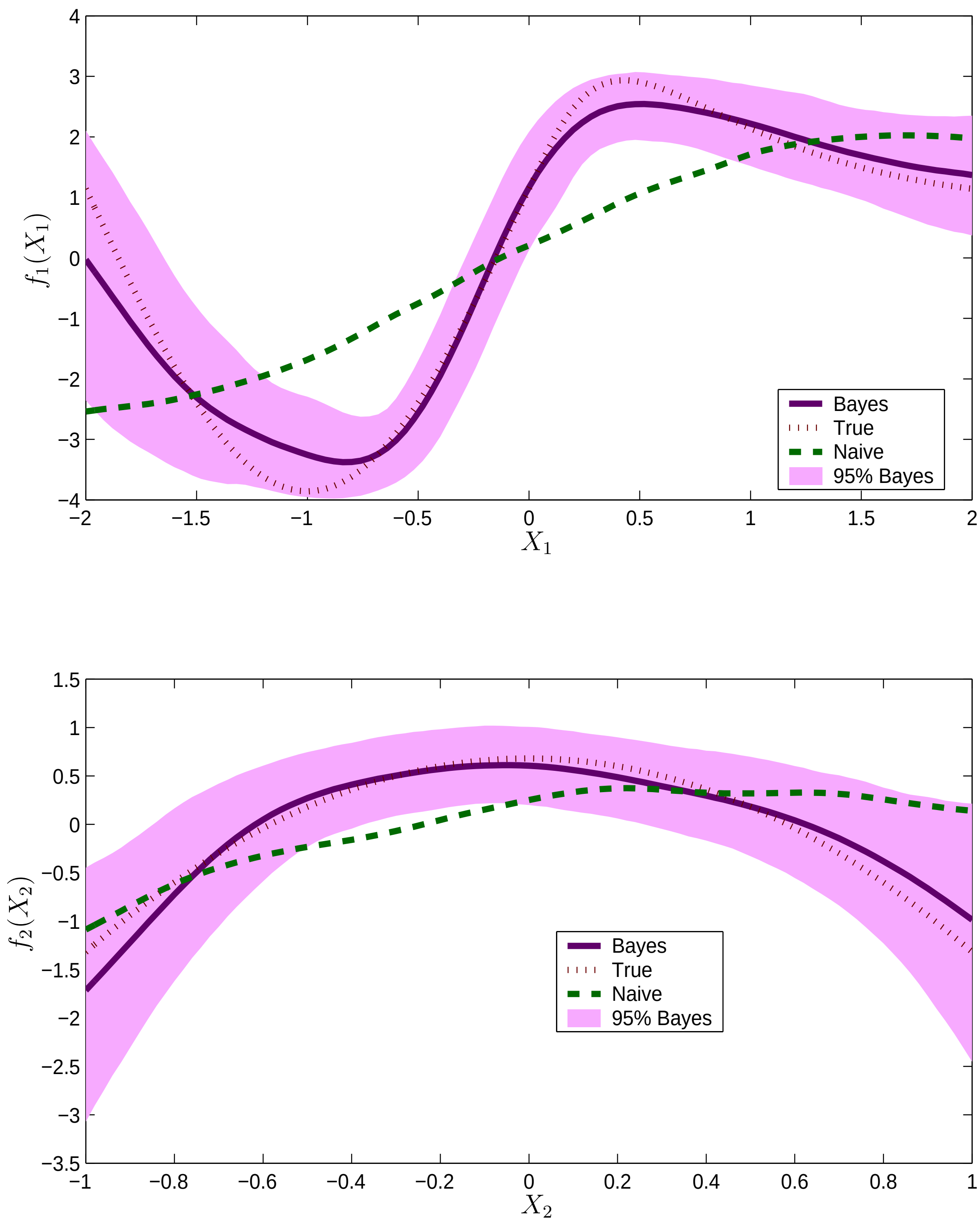

Fig. 3. Average fitted values (95% credible intervals) for 20 simulations when $\sigma_u^2 = 1.0$ and $n = 200$ with continuous response. *The Bayesian method almost perfectly detects the true curves, but the naive method fails to detect the true curves, and shows an almost linear pattern.*

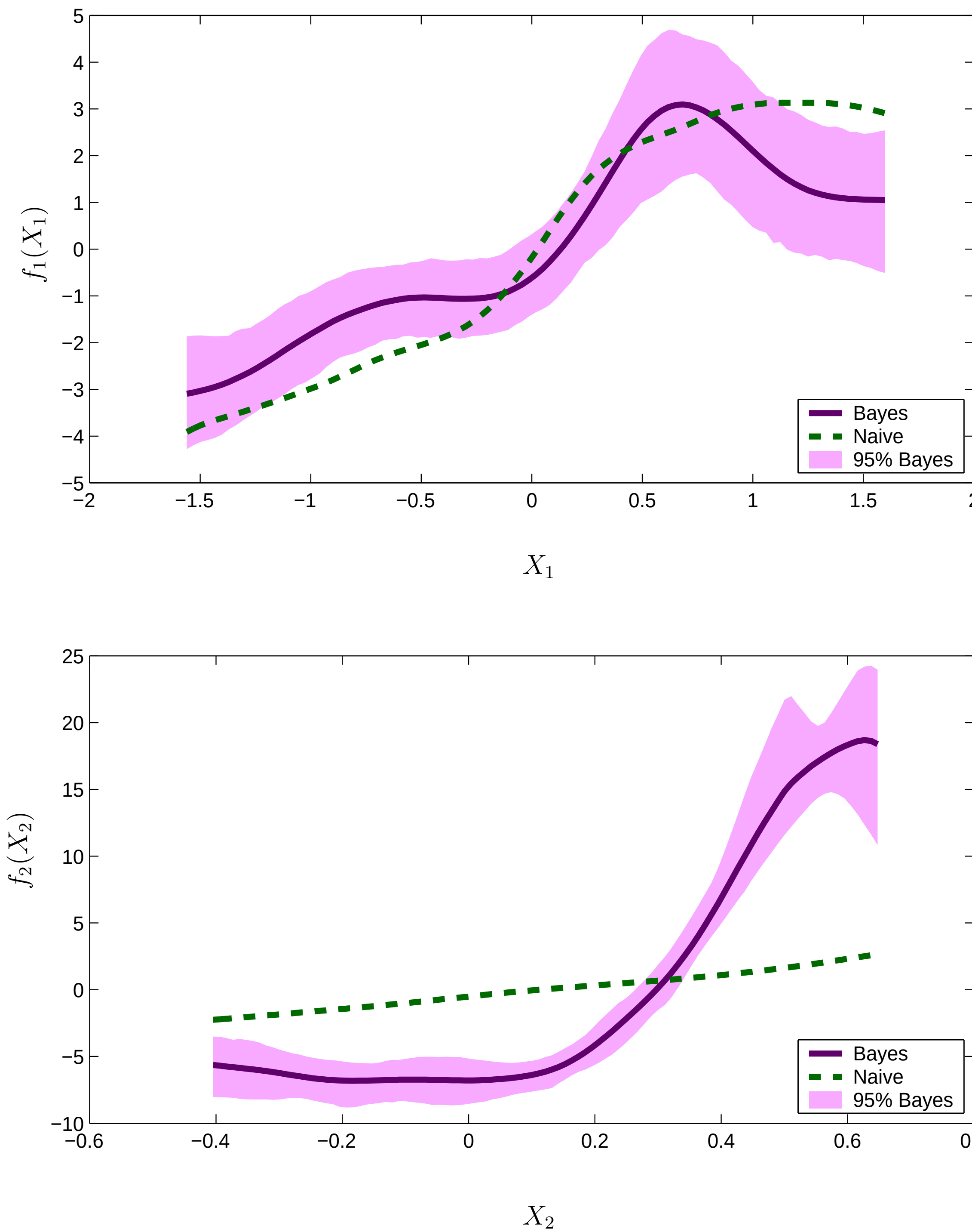

Fig. 4. Continuous regression for adulthood BMI. *The estimated nonparametric curves (95% credible intervals) for the average adulthood BMI show big difference in* f_2*, where the curve from the naive method slopes gently, but the curve from the Bayes method has a steep slope.*

CHAPTER IV

NONPARAMETRIC REGRESSION ANALYSIS FOR BINARY RESPONSE WHEN COVARIATES ARE SUBJECT-SPECIFIC PARAMETERS IN A RANDOM EFFECTS MODEL FOR LONGITUDINAL MEASUREMENTS

4.1. Introduction

Because the linear link function in logistic regression is often unrealistic, nonparametric logistic regression is considered with regard to the binary response, while the measurement error model is assumed to be same as in Chapter III.

With childhood growth data, Wang et al. (1999) tried multiple linear logistic regression on adulthood obesity with the random regression coefficients accomplished by the simple linear regression of childhood BMI, as mentioned before. They proposed a pseudo-EEE and developed approximate-EEE to circumvent measurement error problem. For the binary response, we apply a Bayesian framework to the measurement error problem, and extend the parametric primary regression model to the nonparametric primary regression model. The binary response, in the presence of several covariates and the involved measurement errors, make the nonparametric regression hard. However, the recent advanced computer power makes it possible.

As mentioned in Chapter III, the nonparametric regression problem here is much more complicated than usual additive model regression as the covariates under consideration are not directly observable. For non-Gaussian data, it is well known that the conjugate priors do not exist for the regression coefficients. The computations are then much harder and with the presence of measurement error it could be worse. By virtue of latent variables, we add a random residual component to the model in the spirit of Holmes and Mallick (2003).

By combining ideas from Wang et al. (1999), Berry et al. (2002), and Holmes and Mallick (2003), we propose nonparametric logistic regression with multiple covariates measured with error, through a Bayesian smoothing spline approach to estimate the unknown functions. The Bayesian approach in this problem enables us to deal with non-continuous response, in addition to the advantages of Chapter III.

For a binary response, Section 4.2 has quick summary of measurement error model and primary model. Section 4.3 explains Bayesian nonparametric regression for binary response. Sections 4.4 and 4.5 examine the Bayesian nonparametric regression with simulated data and the BMI data.

4.2. Model

We assume the same measurement error model and same notations as previous chapters such that

$$\boldsymbol{W}_i \;=\; \boldsymbol{D}_i\boldsymbol{X}_i + \boldsymbol{U}_i, \qquad i = 1, \ldots, n,$$

where $\boldsymbol{U}_i \overset{ind}{\sim} N(\boldsymbol{0}, \sigma_u^2 \boldsymbol{I}_{m_i})$, $\boldsymbol{X}_i$ is a vector of subject-specific random effects, $\boldsymbol{D}_i$ is a full rank design matrix, and $\boldsymbol{W}_i$ is a vector of error-prone longitudinal measurements.

Whereas, the primary regression model is supposed to be a generalized linear model (GLM), specifically a logistic regression model, so the conditional distribution of Y_i given $\boldsymbol{X}_i$ is a general exponential family of distributions such as

$$p(Y_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \phi) = \exp\left\{\frac{Y_i\eta_i - b(\eta_i)}{a(\phi)} + c(Y_i, \phi)\right\} = \mathcal{E}xp(\eta_i),$$

where η_i is a canonical parameter (a function of $\boldsymbol{X}_i$), ϕ is a dispersion parameter, and $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are known functions. In logistic regression, $\eta_i = \mathrm{logit}(p_i)$, $a(\phi) = 1$, $b(\eta_i) = -\log(1 - p_i)$, and $c(Y_i, \phi) = 0$. All variables considered are independent across i. Further $\boldsymbol{W}_i$ is supposed to be a surrogate for $\boldsymbol{X}_i$, that is, the distribution of $(Y_i|\boldsymbol{W}_i, \boldsymbol{X}_i)$ is that of $(Y_i|\boldsymbol{X}_i)$ which is independent of $\boldsymbol{W}_i$.

For non-Gaussian data, it is well known that conjugate priors do not exist for the regression coefficients. The computations are then potentially much harder, particularly with measurement error. This is due to possibly strong posterior correlation between the parameters. Although a Metropolis-Hastings algorithm has been commonly used in GLM, the construction of good proposals for the GLM is not trivial. A random residual component is utilized within the model as in Holmes and Mallick (2003). By introducing latent variables Z_{il} the model can be extended such as

$$\begin{aligned} Y_i &\overset{ind}{\sim} \mathcal{E}xp(\eta_i), & \eta_i = \sum_{l=1}^{q} Z_{il}, \quad i = 1, \dots, n, \\ Z_{il} &= f_l(X_{il}) + \epsilon_{il}, & \epsilon_{il} \overset{iid}{\sim} N(0, \sigma_{z_l}^2), \quad i = 1, \dots, n, \ l = 1, \dots, q. \end{aligned}$$

Suppose X_{il} to be independently from $N(\mu_{x_l}, \sigma_{x_l}^2)$, $l = 1, \dots, q$, and identically distributed across i. Further, assume that X_{il} is independent of Z_{il}s, then η_i has mean of $\sum_{l=1}^{q} f_l(X_{il})$ and variance of $\sum_{l=1}^{q} \sigma_{z_l}^2$.

4.3. Bayesian smoothing spline with measurement error

We consider additive q Bayesian natural cubic smoothing splines (NCSs) as the link function in the primary model. As in Chapter III, a partially improper Gaussian process (singular normal) is assigned as the prior such that

$$\tau_l^{\frac{n-2}{2}} \exp\left\{-\frac{\tau_l}{2} \boldsymbol{f}_l^T \boldsymbol{K}_l \boldsymbol{f}_l\right\}, \qquad l = 1, \dots, q,$$

where $\tau_l = \frac{\alpha_l}{\sigma_{z_l}^2}$, for a smoothing parameter α and $\mathrm{var}(Z_{il}) = \sigma_{zl}^2$, and $\boldsymbol{K}_l$ is a matrix satisfying $f_l^T \boldsymbol{K}_l f_l = \int \left\{f_l''(t)\right\}^2 dt$.

To construct the Bayes hierarchical model, we assign same priors for covariates

$\boldsymbol{X}_l = (X_{1l}, \ldots, X_{nl})^T$ and other necessary parameters such that for $l = 1, \ldots, q$, $\boldsymbol{X}_l \overset{ind}{\sim} N(\mu_{x_l}, \sigma^2_{x_l})$, $\mu_{x_l} \sim N(A_{m_l}, B_{m_l})$, $\sigma^2_{x_l} \sim IG(A_{x_l}, B_{x_l})$, and $\sigma^2_u \sim IG(A_u, B_u)$. By introducing q latent variables $\boldsymbol{Z}_l = (Z_{1l}, \ldots, Z_{nl})^T$, we also assume a conjugate prior for the variance of each latent variable as $\sigma^2_{z_l} \sim IG(A_{z_l}, B_{z_l})$ and τ_l as $\tau_l \sim G(A_{t_l}, B_{t_l})$, $l = 1, \ldots, q$. Note that each nonparametric curve induced by one covariate has its own latent variable. The relationship between parameters is described in the DAG (directed acyclic graph) in Figure 5 for $q = 2$.

From the model and prior distributions, the joint posterior distribution of the unknowns is established to be proportional to

$$
\begin{aligned}
\propto\ & \prod_{i=1}^{n} p(Y_i|\eta_i) \times \prod_{l=1}^{q} \Bigg[\left(\sigma^2_{z_l}\right)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2_{z_l}} \sum_{i=1}^{n} \left(Z_{il} - f_l(X_{il})\right)^2 \right\} \\
& \quad \times \tau_l^{\frac{n-2}{2}} \exp\left\{ -\frac{\tau_l}{2} \boldsymbol{f}_l^T \boldsymbol{K}_l \boldsymbol{f}_l \right\} \tau_l^{A_{t_l}-1} \exp\left\{ -\frac{\tau_l}{B_{t_l}} \right\} \\
& \quad \times \left(\sigma^2_{x_l}\right)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2_{x_l}} \sum_{i=1}^{n} (X_{il} - \mu_{x_l})^2 \right\} \\
& \quad \times \left(\sigma^2_{x_l}\right)^{-(A_{x_l}+1)} \exp\left\{ -\frac{1}{\sigma^2_{x_l} B_{x_l}} \right\} \ \times B_{m_l}^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2B_{m_l}} (\mu_{x_l} - A_{m_l})^2 \right\} \\
& \quad \times \left(\sigma^2_{z_l}\right)^{-(A_{z_l}+1)} \exp\left\{ -\frac{1}{\sigma^2_{z_l} B_{z_l}} \right\} \Bigg] \\
\times\ & \left(\sigma^2_u\right)^{-\frac{1}{2}\sum_{i=1}^{n} m_i} \exp\left\{ -\frac{1}{2\sigma^2_u} \sum_{i=1}^{n} (\boldsymbol{W}_i - \boldsymbol{D}_i \boldsymbol{X}_i)^T (\boldsymbol{W}_i - \boldsymbol{D}_i \boldsymbol{X}_i) \right\} \\
\times\ & \left(\sigma^2_u\right)^{-(A_u+1)} \exp\left\{ -\frac{1}{\sigma^2_u B_u} \right\},
\end{aligned}
$$

where $p(\cdot)$ is a density of the general exponential family, particularly Bernoulli density with a success probability p_i.

Let's use the same notations with Chapter III for $\boldsymbol{A}_l(\alpha_l)$, $\boldsymbol{f}_l$, and $\boldsymbol{Y}$, in addition to the latent variable $\boldsymbol{Z}_l = [Z_{1l}, \ldots, Z_{nl}]^T$. For the residual sums of squares, let RSS_{x_l}, RSS_{w_i} and $RSS_{x_{il}}$ stand for the same things in Chapter III. The residual sums of

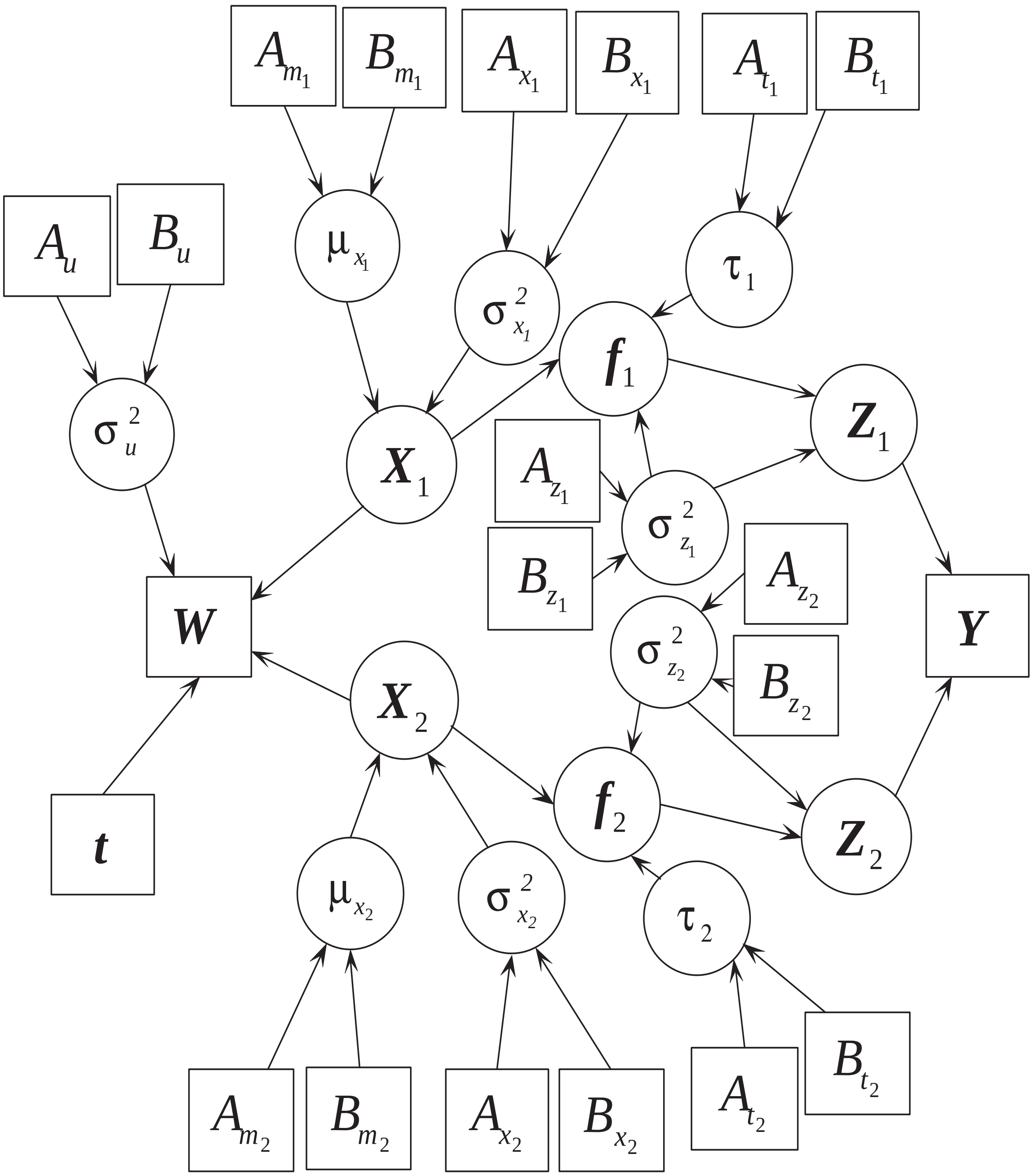

Fig. 5. Directed acyclic graph for hierarchical Bayesian model with binary response. *Variables in the rectangles are observable or given, but variables in the circles are not observable. The information of* $\boldsymbol{X}_1$ *and* $\boldsymbol{X}_2$ *is given by the combination of* $\boldsymbol{W}$ *and* $\boldsymbol{t}$*. Each covariate is associated with the separate nonparametric curve* $\boldsymbol{f}_j$, $j = 1, 2$*. Each curve has its own latent variable.*

squares induced by a latent variable $\boldsymbol{Z}_l$ are defined by $RSS_{z_l} = \sum_{i=1}^{n}\{Z_{il} - f_l(X_{il})\}^2$ and $RSS_{z_{il}} = \{Z_{il} - f_l(X_{il})\}^2$, $l = 1, \ldots, q;\ i = 1, \ldots, n$.

After performing simple algebra, all the full conditional distributions are achieved by:

$$
\begin{aligned}
\boldsymbol{f}_l|\cdot &\overset{ind}{\sim} N\left[\boldsymbol{A}_l(\alpha_l)\boldsymbol{Z}_l, \boldsymbol{A}_l(\alpha_l)\sigma_{z_l}^2\right],\\
\tau_l|\cdot &\overset{ind}{\sim} G\left[\frac{n-2}{2} + A_{t_l}, \left(\frac{\boldsymbol{f}_l^T\boldsymbol{K}_l\boldsymbol{f}_l}{2} + \frac{1}{B_{t_l}}\right)^{-1}\right],\\
\sigma_{z_l}^2|\cdot &\overset{ind}{\sim} IG\left[\frac{n}{2} + A_{z_l}, \left(\frac{RSS_{z_l}}{2} + \frac{1}{B_{z_l}}\right)^{-1}\right],\\
\sigma_{x_l}^2|\cdot &\overset{ind}{\sim} IG\left[\frac{n}{2} + A_{x_l}, \left(\frac{RSS_{x_l}}{2} + \frac{1}{B_{x_l}}\right)^{-1}\right],\\
\mu_{x_l}|\cdot &\overset{ind}{\sim} N\left[\left(\frac{\sum_{i=1}^{n} X_{il}}{\sigma_{x_l}^2} + \frac{A_{m_l}}{B_{m_l}}\right)\left(\frac{n}{\sigma_{x_l}^2} + \frac{1}{B_{m_l}}\right)^{-1}, \left(\frac{n}{\sigma_{x_l}^2} + \frac{1}{B_{m_l}}\right)^{-1}\right],\\
\sigma_u^2|\cdot &\sim IG\left[\frac{\sum_{i=1}^{n} m_i}{2} + A_u, \left(\frac{\sum_{i=1}^{n} RSS_{w_i}}{2} + \frac{1}{B_u}\right)^{-1}\right],
\end{aligned}
$$

for $l = 1, \ldots, q$, where $G[a, b]$ stands for a gamma distribution with mean ab, and $IG[a, b]$ indicates an inverse gamma distribution with mean $\{(a-1)b\}^{-1}$. In addition to the full conditionals of X_{il}, the full conditionals of Z_{il}, $i = 1, \ldots, n$; $l = 1, \ldots, q$ are not of standard forms but proportional to the following:

$$
\begin{aligned}
[X_{il}|\cdot] &\overset{ind}{\propto} \exp\left\{-\frac{RSS_{z_{il}}}{2\sigma_{z_l}^2} - \frac{RSS_{x_{il}}}{2\sigma_{x_l^2}} - \frac{RSS_{w_i}}{2\sigma_u^2}\right\},\\
[Z_{il}|\cdot] &\overset{ind}{\propto} p(Y_i|\eta_i)N\left[f_l(X_{il}), \sigma_{z_l}^2\right], \qquad i = 1, \ldots, n;\ l = 1, \ldots, q.
\end{aligned}
$$

From this conditional distribution, it is clear that by adopting Gaussian residual effects $\boldsymbol{Z}_l$, $l = 1, \ldots, q$, many of the conditional distributions for the model parameters are now of standard form, which greatly aids in the computations. To be specific, conditioning on $\boldsymbol{Z}_l$, the model for $\boldsymbol{f}_l$ is independent of $\boldsymbol{Y}$ and can be written as a standard Bayes linear regression of $\boldsymbol{Z}_l$ on the basis space defined by NCS. Hence, an efficient sampling of $p(\boldsymbol{f}_l|\boldsymbol{Z}_l, \boldsymbol{Y})$ is possible.

Even with the binary response, we can find that the covariates have roles at almost every full conditional distribution except for the variances of the latent variables. Accordingly, the measurement error will affect the estimation of link functions, as well as other parameters. In the naive approach, we replace the unobserved $\boldsymbol{X}$ by the least squares estimator for each subject, as in Chapter III. That will create a biased estimate for the unknown functions. The results from the naive approach are going to be compared with those from full Bayes approach.

To generate each parameter from the joint posterior distribution, we use Gibbs sampler (Gelfand and Smith, 1990). Within the Gibbs procedure, the covariates X_{il} and the latent variables Z_{il}, $i = 1, \ldots, n$, $l = 1, \ldots, q$ can be generated through Metropolis-Hasting's algorithm, because they do not follow standard forms. Customizing the idea of Holmes and Held (2005) used in Chapter II, we can generate Z_{il}, $i = 1, \ldots, n$; $l = 1, \ldots, q$, from the conditional distribution marginalized over $\boldsymbol{f}_l$, $l = 1, \ldots, q$. Detailed calculations are provided in Appendix C. As explained in Chapter III, the fitted curves $\boldsymbol{f}_l$ from multivariate normal distributions are generated by Cholesky's decomposition with the backward and the forward substitutions.

We draw samples of functions from their joint posterior distribution and use the pointwise mean curves as the natural estimate of the regression functions. The sampling based method provides us with the flexibility to calculate pointwise median, credible intervals, or any other functional values of these regression functions. In our method, we allow a separate smoothness parameter for different regression functions in the additive model setup, as in Chapter III. Hence, the estimated curves have flexibility with different degrees of smoothness (complexity) with all the uncertainty measures. In addition to the estimation of the functions, inferences (posterior mean and credible intervals) can be done for mismeasured Xs easily using the corresponding MCMC samples even with a binary response.

4.4. Simulation study

For this simulation study, the same data set with Chapter III was used, except for the response. The binary responses were generated from a Bernoulli distribution with success probability $p_i = [1 + \exp\{-f_1(X_{i1}) - f_2(X_{i2})\}]^{-1}$.

The Bayesian hierarchical model was established by assigning same flexible hyper priors for σ_u^2, τ, μ_{x_1}, μ_{x_2}, and $\sigma_{x_1}^2$. Because different variances are assumed for each latent variable, the priors for the variances of two latent variables are given by $\sigma_{z_1}^2 = \sigma_{z_2}^2 \sim IG(3, 3)$.

The performance of each estimator was evaluated by MSE at 101 grid points in the interval [-2,2] for f_1 and [-1,1] for f_2. In addition, we also examined the deviance information criterion (DIC) explained at Appendix A. Detail theory and procedure for DIC calculation can be found in Spiegelhalter et al. (2002); Mallick et al. (2002).

We examined twenty simulated data sets with different measurement error variance by collecting 10000 Markov Chain Monte Carlo (MCMC) samples after 50000 burning iterations, to check the effect of increased variance of measurement error (σ_u^2) and the increased sample size (n) on the performance of the estimator. For fixed sample size, in general, larger σ^2 (more measurement error) worsens the performance of the estimator (increase the MSE). The simulation results are reported in Table IV. The Bayes method outperformed the naive method in each situation. Although the overall performance of the Bayes method has been deteriorated compared to the continuous case, the Bayes method still has a better performance than the naive method. This bad performance is from the binary response, not from the estimation

Table IV. Table for simulated binary data: Average MSE and average DIC from 20 simulated data for each situation.

Situation	Link	Criterion	NoErr	Naive	Bayes
$n = 200$, $\sigma_u^2 = 0.49$	f_1	MSE	0.4825	1.5835	0.9806
		Bias^2	0.1442	1.4030	0.7180
		Var	0.3383	0.1804	0.2626
	f_2	MSE	0.1695	0.2710	0.2649
		Bias^2	0.0346	0.1737	0.1061
		Var	0.1350	0.0973	0.1588
		$\overline{\text{DIC}}$	115.17	181.61	175.63
$n = 200$, $\sigma_u^2 = 1.0$	f_1	MSE	0.4825	2.2095	1.4062
		Bias^2	0.1442	2.0737	0.9474
		Var	0.3383	0.1359	0.4588
	f_2	MSE	0.1695	0.4651	0.3395
		Bias^2	0.0346	0.3545	0.1373
		Var	0.1350	0.1106	0.2022
		$\overline{\text{DIC}}$	115.17	201.26	189.16
$n = 400$, $\sigma_u^2 = 1.0$	f_1	MSE	0.2633	2.5814	1.2636
		Bias^2	0.0965	2.4919	1.1279
		Var	0.1668	0.0895	0.1356
	f_2	MSE	0.1450	0.3501	0.2968
		Bias^2	0.0541	0.3050	0.1531
		Var	0.0909	0.0451	0.1436
		$\overline{\text{DIC}}$	248.89	418.20	411.66

NOTE: Bayes method performed better than the naive method in all the situations in terms of MSE as well as DIC.

of unknown covariates, because the estimated curve performs the worst even with true covariate values.

In Figure 6, we plotted the true curve, the posterior mean curve and 95% pointwise credible intervals obtained from the MCMC samples for the nonparametric regression with binary data. Though the Bayes method did not work as well as the continuous example, still it outperformed the naive method completely. In summary, even in the simulation with the binary response the Bayes method showed a superior performance to the naive estimates.

4.5. Childhood growth data analysis for adulthood obesity

We applied nonparametric logistic regression (when $q = 2$) to the childhood growth data which is used in Chapter II. Adulthood obesity has been considered as a binary response in two dimensional covariate space, where covariates are not observable intercepts and slopes of childhood BMIs. The analysis results of adulthood obesity were very similar to adulthood BMI. The Bayesian nonparametric logistic regression performs better than the naive method in terms of DIC (237.56 vs. 233.09), which is also better than the results of parametric logistic regression. Again, from Figure 7, we can realize that the non-linear effects of the initial BMI-z at age three and the rate of change of BMI-z on the response. Hence, to explore them, nonparametric regression is inevitable here. In summary for both of the situations (continuous and binary), the initial BMI-z value and the rate of change have significant nonlinear effects on adulthood BMI (or obesity).

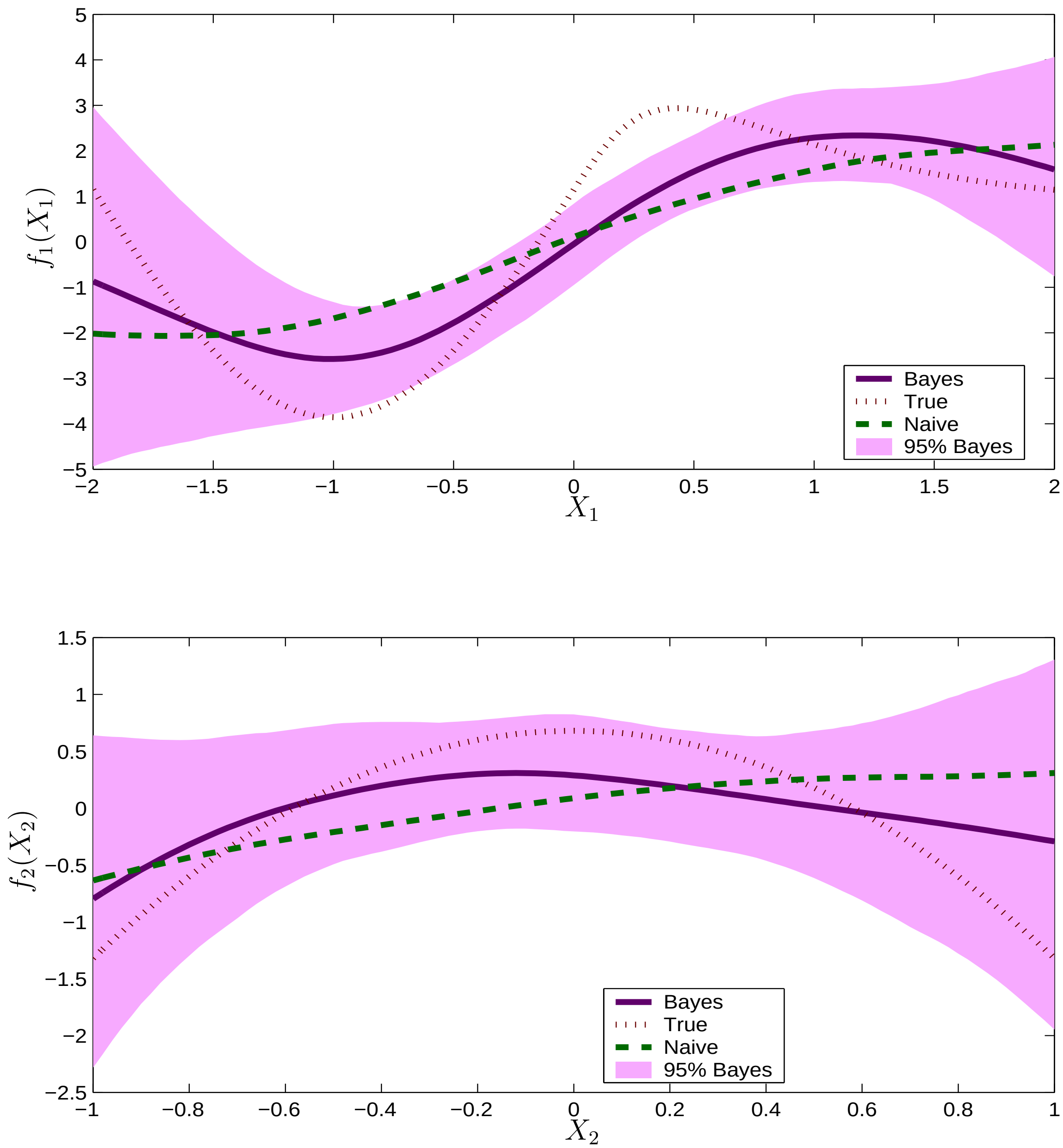

Fig. 6. Average fitted values (95% credible intervals) for twenty simulations when $\sigma_u^2 = 1.0$ and $n = 200$ with binary response. *The naive method fails to detect the true curve and shows an almost linear pattern. However, the Bayesian method detects the pattern of the true curve. In* $\boldsymbol{f}_1$, *the Bayesian method captures the valley and the mountain of the true curve, and in* $\boldsymbol{f}_2$, *it has a concave shape.*

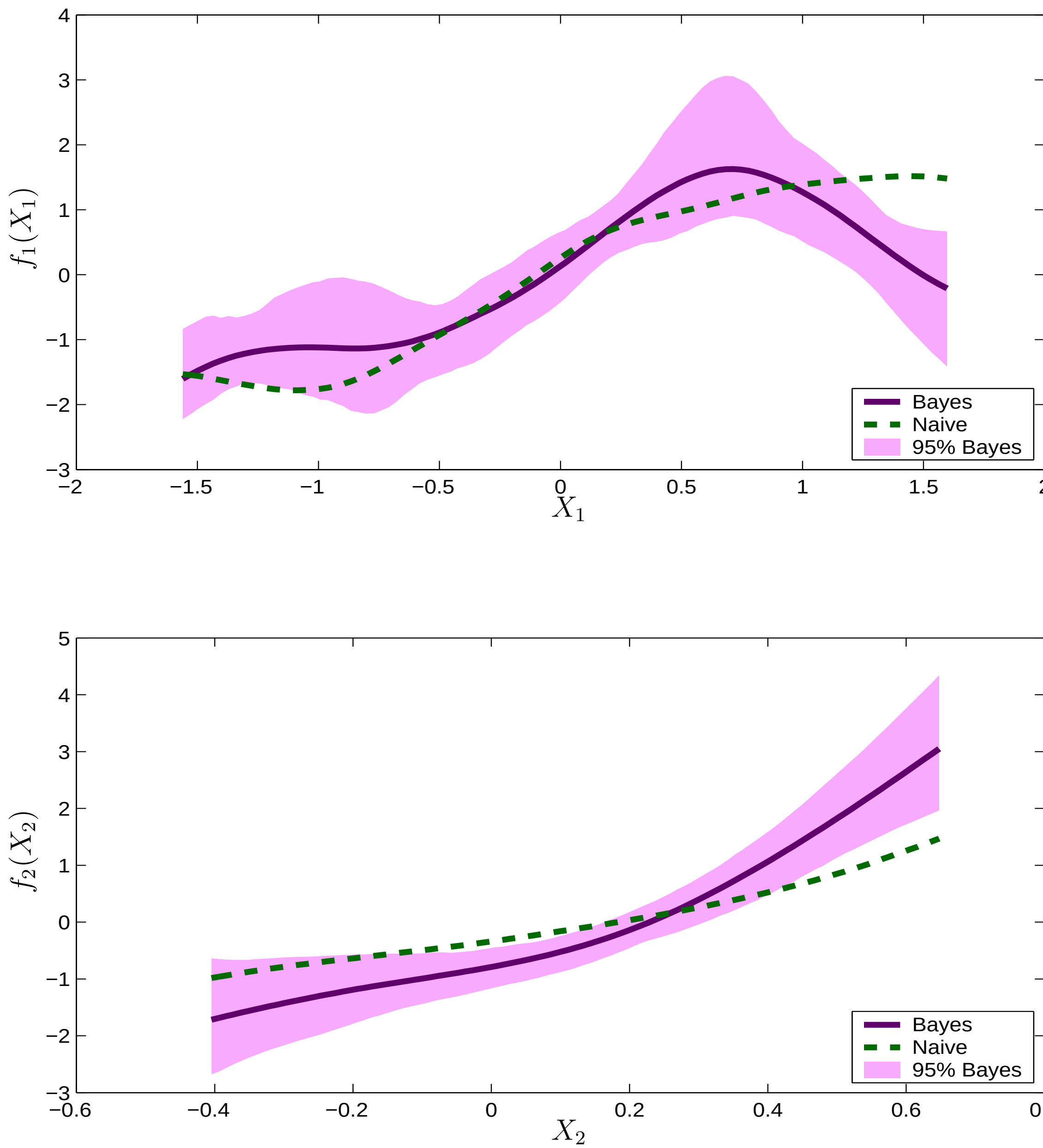

Fig. 7. Binary regression for the real data. *Estimated nonlinear functions (95% credible intervals) under the nonparametric model with BMI obesity data.*

CHAPTER V

BAYESIAN NONPARAMETRIC REGRESSION OF OUTCOME-DEPENDENT FOLLOW-UP TIMES

5.1. Introduction

In many observational longitudinal studies, individuals are not measured at pre-specified regular intervals. We consider studies where the time-points of measurements are unequally spaced and time of a follow-up measurement can depend on the history of past clinic visits and of previous outcomes of that individual, often called 'outcome-dependent follow-up.' This situation can arise when an individual with a poor disease history requires more care and hence more frequent visits to a doctor. For example, Lipsitz et al. (2002) investigated the cardiotoxic effects of doxorubicin chemotherapy for the treatment of acute lymphoblastic leukemia in childhood. Although doxorubicin has proven to be successful in curing leukemia, it can cause progressive abnormalities of the heart in long-term survivors. The primary longitudinal outcome variable of the study was the patient's heart-wall thickness, which was measured via echocardiogram during every clinic visit of the child. The time of the next follow-up visit was based on the physician's judgement about the child's history of disease and condition of health at the current clinic visit. When previous echocardiograms had shown a history of abnormalities of the heart, the next echocardiogram was expected to be scheduled sooner than what is typical for a normal patient. Consequently, the times of observation were unequally spaced and varied from child to child. Our interest lies in estimating the regression parameters for the longitudinal outcome, as well as the prediction of future outcomes for an individual, using Bayesian techniques. Further, we are also interested in applying Bayesian techniques when the interval between two echocardiograms might also depend on other (unobserved) health factors beyond what was captured directly via the past history of echocardiogram.

There has been a recent surge of interest in techniques to analyze studies with outcome-dependent follow-up. Parametric regression methods have been proposed in Liang and Zeger (1986); Lipsitz et al. (2002); Fitzmaurice et al. (2003); Chen (2003), and nonparametric regression methods have been proposed by Wang (1998); Opsomer et al. (2001). In a parametric likelihood framework, as Lipsitz et al. (2002) pointed out, misspecification of the correlation structure among longitudinal measurements result in biased estimation of regression parameters. In nonparametric regression, outcome-dependent follow-up leads to undersmoothing behavior of the cross-validation fit (Opsomer et al., 2001). However, Bayesian methods to analyze such data have not been proposed.

In this chapter and the next, Bayesian approaches with both parametric and nonparametric regression models has been considered. For the parametric regression, noninformative priors for regression parameters are applied. For the nonparametric regression, Bayesian natural cubic smoothing splines with partially improper Gaussian prior (Berry et al., 2002) are considered. Lipsitz et al. (2002) assumed a common variance and correlation structures based on a common correlation coefficient over all individuals, and constructed a parametric regression model. Hence, when the exchangeable (compound symmetry) correlation structure is assumed, their model will have a common subject-specific covariance over all individuals. In addition, they did not provide a direct association between the outcome process and the follow-up time process, so that their model is only affected by a type of correlation structure not by a specific follow-up time process. In this chapter, Bayesian approaches using a model very similar to Lipsitz et al. (2002) have been explored and then, in the next chapter, it has been extended by introducing a subject-specific latent variable. Furthermore, the correlation structure of errors of the regression model has also been extended by allowing different correlations for different individuals. These models considered in this chapter and the next are more general than the model of Lipsitz et al. (2002) in that they allow more a much bigger class of models of association among the responses at different time-points and also more general modelling of the intervals between measurement times.

In section 5.2, we describe parametric and nonparametric regression models, and in section 5.3, we present priors and induced posteriors for parametric and nonparametric regression. To explore the effect of misspecification of the type of correlation structure, section 5.4 customizes a Bayesian model diagnostic and section 5.5 provides a simple simulation for some types of correlation structure. Finally, in section 5.6 we illustrate the proposed methods using data from the longitudinal study of the cardiotoxic effects of doxorubicin chemotherapy discussed earlier.

5.2. Model

For the individual $i = 1, \cdots, n$, the unequally spaced observed follow-up times are denoted by $\{t_{i1} < \ldots < t_{im_i}\}$. Let $Y_{ij} = Y_i(t_{ij})$ be the response from the individual i at the follow-up time t_{ij}. Following Lipsitz et al. (2002); Fitzmaurice et al. (2003), the first follow-up time t_{i1} is considered fixed by design and is treated as a part of the covariates. In cardiotoxicity data (Lipsitz et al., 2002), t_{i1} is the time of first clinic visit since the end of chemotherapy of individual i and cardiologists consider t_{i1} as a key measure in predicting future course of heart function. Given the first follow-up time t_{i1}, the subsequent follow-up times can be modeled by a process of follow-up times $N_i(t)$. We model the time elapsed between follow-up times

$$U_{ik} = t_{i,k} - t_{i,k-1}, \qquad k = 2, \ldots, m_i;\ i = 1, \ldots, n.$$

The time interval U_{ik} of individual i is expected to be dependent on the history of previous measurements $\mathscr{Y}_{ik} = (Y_{i1}, \cdots, Y_{i,k-1})$ and history of follow-up times $\mathscr{U}_{ik} = (t_{i1}, \cdots, U_{ik,k-1})$. We begin with Model-0, studied by Lipsitz et al. (2002); Fitzmaurice et al. (2003), where follow-up times depend only on the observed history $\mathscr{Y}_{ik}$ of longitudinal measurements. This modeling assumption is given by

$$[U_{ik}|\mathscr{Y}_{ik}, \mathscr{U}_{ik}] = [U_{ik}|\mathscr{Y}_{ik}] \quad \text{and} \quad [Y_{ik}|U_{ik}, \mathscr{Y}_{ik}, \mathscr{U}_{ik}] = [Y_{ik}|U_{ik}, \mathscr{Y}_{ik}]\ .$$

Under this assumption of Model-0, the conditional density of U_{ik} can remain unspecified, because follow-up time process $(U_{i2}, \ldots, U_{im_i})$ do not make any direct contribution to the estimation of the parameters associated with the longitudinal response variable Y, where $[Y_{ik}|U_{ik}, \mathscr{Y}_{ik}, \mathscr{U}_{ik}]$ and $[U_{ik}|Y_{ik}]$ do not share any common parameter. For details of the direct impact of the follow-up times to the regression in this model, we refer to Lipsitz et al. (2002); Fitzmaurice et al. (2003).

Under Model-0, the outcome process is given by

$$Y(t) \;=\; \mu(t) + \epsilon_t,$$

where $\mu(t)$ is a regression function of covariates at time t and ϵ_t is a Gaussian process with zero mean. We use $\epsilon_{ij} = \epsilon_{t_{ij}}$, to simplify the notation. The error processes ϵ_{it} are assumed to be independent among individuals, but dependent within each individual. Assume common variance $Var(\epsilon_{it}) = \sigma^2$ and covariance $Cov(\epsilon_{it}, \epsilon_{is}) = \sigma^2\rho_{its}$ with a subject-specific correlation function $-1 < \rho_{its} < 1$. Hence the covariance of errors at observed follow-up times, $\text{cov}\{\epsilon_{11}, \ldots, \epsilon_{nm_n}\}$, can be described a blocked diagonal matrix

$$\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{\Sigma}_e = \sigma^2 \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_n \end{pmatrix},$$

where $\boldsymbol{\Sigma}_i$, $i = 1, \ldots, n$, is the $m_i \times m_i$ correlation matrix of outcome measures from individual i. We may consider a correlation structure which depends on the time elapsed between follow-up times in each individual. For example, we can consider the correlation function $\rho_{its} = \rho_i^{|t-s|}$, which implies that the $(k, j)^{th}$ element of $\boldsymbol{\Sigma}_i$, $i = 1, \ldots, n$, is given by

$$\mathrm{corr}(Y_{ik}, Y_{ij}) = \begin{cases} \rho_i^{|t_{ik} - t_{ij}|}, & \text{if } k \neq j, \\ 1, & \text{if } k = j. \end{cases}$$

For the regression function $\mu(t)$, we consider an additive model with follow-up times and other covariates. The regression function can be described as

$$\mu_i(t) = f(t) + \boldsymbol{X}_i(t)\boldsymbol{\beta}, \qquad i = 1, \ldots, n,$$

where $f(t)$ is a function for follow-up time t, $\boldsymbol{X}_i(t)$ is a $1 \times (q + 1)$ row vector of a constant 1 and other q covariates at t, and $\boldsymbol{\beta}$ is a $q + 1$ dimensional column vector of regression coefficients. For the parametric regression model, $f(t)$ can be assumed to be, say, a known order polynomial function of t with unknown coefficients. As an alternative, we can use a natural cubic smoothing spline for the nonparametric regression model of $f(t)$. To avoid the identifiability problem, we set $\sum_{ij} f(t_{ij}) = 0$ for nonparametric regression model and set the intercept term in $f(t)$ to be zero for parametric regression model.

5.3. Bayesian regressions

Bayesian hierarchical models have been considered for analysis using Model-0. Unknown parameters of models considered in section 5.2 are supposed to follow flexible priors, and posterior quantities of interest are achieved by utilizing the Gibbs sampling (Geman and Geman, 1984). Arguments begin with Bayesian parametric regression (B-P), and then proceed to Bayesian nonparametric regression (B-NP).

A. Parametric regression

For the function $f(t)$, suppose a polynomial function of known order p:

$$f(t) = \theta_1 t + \cdots + \theta_p t^p \ ,$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$ is the unknown parameter vector. Assume uniform priors for $\boldsymbol{\theta}$ and the regression parameters of covariates so that their posterior means are comparable to the maximum likelihood estimates. In addition, consider uniform priors for correlations such that $\rho_i \sim U(-1, 1)$, $i = 1, \ldots, n$, and $\sigma^2 \sim IG(A_s, B_s)$ with known A_s and B_s for the common variance of ϵ_{it}.

Let $N = \sum_{i=1}^{n} m_i$. Then, for individual i, we have $m_i \times 1$ vector of response $\boldsymbol{Y}_i$, $m_i \times 1$ vector of follow-up times $\boldsymbol{t}_i$, and $m_i \times q$ matrix of other covariates $\boldsymbol{X}_i$. Let $\boldsymbol{Z}_i = (\boldsymbol{t}_i, \mathbf{1}_{m_i}, \boldsymbol{X}_i)$, where $\mathbf{1}_{m_i}$ is $m_i \times 1$ vector of ones, and $\boldsymbol{R}_{y_i} = (\boldsymbol{Y}_i - \boldsymbol{Z}_i \boldsymbol{\Theta})$, where $\boldsymbol{\Theta} = (\boldsymbol{\theta}^T, \boldsymbol{\beta}^T)^T$ is a $(p + q + 1) \times 1$ vector of regression coefficients $\boldsymbol{\theta}$ from follow-up times and $\boldsymbol{\beta}$ from constant term and other covariates, $i = 1, \ldots, n$. Further denote $\boldsymbol{Y}$, $\boldsymbol{t}$, $\boldsymbol{X}$, $\boldsymbol{Z}$ and $\boldsymbol{R}$ as the stacked versions of the corresponding vectors or matrices over all individuals. Then, the joint density given follow-up times is proportional to the following:

$$
\begin{aligned}
\propto & \quad \left(\sigma^2\right)^{-\frac{N}{2}} \prod_{i=1}^{n} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{R}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{R}_i\right\} \\
\times & \quad \left(\sigma^2\right)^{-(A_s+1)} \exp\left\{-\frac{1}{\sigma^2 B_s}\right\},
\end{aligned}
$$

where $\boldsymbol{\Sigma}_i$ are i^{th} diagonal element of the error correlation matrix $\boldsymbol{\Sigma}_e$. The prior densities of nuisance parameters do not need to be considered in the calculation of the joint p.d.f. because nuisance parameters are only involved in the follow-up time process and there is no direct association between the follow-up time process and the outcome process.

From the joint density, full conditionals for $\boldsymbol{\Theta}$ and other parameters σ^2 and ρ_i, $i = 1, \ldots, n$, are driven by

$$
\begin{aligned}
\boldsymbol{\Theta}|\cdot & \sim N\left[\left(\boldsymbol{Z}^T\boldsymbol{\Sigma}_e^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^T\boldsymbol{\Sigma}_e^{-1}\boldsymbol{Y}, \left(\boldsymbol{Z}^T\boldsymbol{\Sigma}_e^{-1}\boldsymbol{Z}\right)^{-1}\sigma^2\right] \\
\sigma^2|\cdot & \sim IG\left[\frac{N}{2} + A_s, \left(\frac{1}{2}\boldsymbol{R}_y^T\boldsymbol{\Sigma}_e^{-1}\boldsymbol{R}_y + \frac{1}{B_s}\right)^{-1}\right], \\
\rho_i|\cdot & \overset{ind}{\propto} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{R}_{y_i}^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{R}_{y_i}\right\}, \qquad i = 1, \ldots, n.
\end{aligned}
$$

Since the posterior of ρ_i does not have a standard distribution, Metropolis-Hasting's algorithm needs to be utilized to generate samples from the density. The de-constraint transformation (Chen et al., 2000) has been applied to sample ρ_i such that $\rho_i = \frac{\exp(\xi_i)}{1+\exp(\xi_i)}$, and generate ρ_i via ξ_i which has a density such that

$$
\pi(\xi_i|\cdot) = \pi(\rho_i|\cdot)\frac{\exp(\xi_i)}{\{1 + \exp(\xi_i)\}^2}, \qquad i = 1, \ldots, n,
$$

where $\pi(\cdot)$ indicates the density function of the corresponding arguments. For the Metropolis-Hastings algorithm, a normal proposal density $N(\widehat{\xi_i}, \widehat{\sigma^2_{\xi_i}})$ has been considered, where $\widehat{\xi_i}$ can be achieved by Nelder-Mead algorithm and $\widehat{\sigma^2_{\xi_i}}$ can be achieved by the inverse of the numerically approximated information number of ξ_i, evaluated at $\widehat{\xi}_i$. Details of the procedure can be found at Chen et al. (2000, pg. 25).

B. Nonparametric regression

For a Bayesian nonparametric regression (B-NP), the $f(t)$ is estimated by Bayesian natural cubic smoothing spline (NCS), and the effects of other covariates are modeled by linear function. Let $\boldsymbol{f}_i = [f(t_{i1}), \ldots, f(t_{im_i})]^T$, $i = 1, \ldots, n$, for the individual i, and let $\boldsymbol{f} = [\boldsymbol{f}_1^T, \ldots, \boldsymbol{f}_n^T]^T$ and $N = \sum_{i=1}^n m_i$. For functional values $\boldsymbol{f}$ of all individuals, suppose a partially improper prior which is often used whenever response errors are assumed independent (Berry et al., 2002; Ryu and Mallick, 2004) such that

$$\boldsymbol{f} \sim Singular\ Normal[\mathbf{0}, \alpha^{-1}\sigma^2\boldsymbol{K}^-],$$

where α is a smoothing parameter for the NCS, and $\boldsymbol{K}$ is $N \times N$ matrix with rank $N-2$ satisfying $\boldsymbol{f}^T\boldsymbol{K}\boldsymbol{f} = \int\{f''(t)\}^2 dt$. The matrix $\boldsymbol{K}$ can be achieved by the method described at Eubank (1999, p. 244). Note that $\boldsymbol{K}$ in Eubank (1999) is based on the sorted $\boldsymbol{t}$. Without loss of generality, a conjugate prior is assigned to $\tau = \frac{\alpha}{\sigma^2}$ instead of α such that $\tau \sim G(A_t, B_t)$. For the nuisance parameters σ^2 and ρ_i, $i = 1, \ldots, n$, same priors are assigned as in the Bayesian parametric regression (B-P). In Model-0, the distribution of follow-up times needs not to be explicit. The directed acyclic graph of the Bayesian nonparametric regression under Model-0 is shown in Figure 8. Not being confused with Bayesian parametric regression, let $\boldsymbol{R}_y = (\boldsymbol{Y} - \boldsymbol{f} - \boldsymbol{X}\boldsymbol{\beta})$ and $\boldsymbol{R}_{y_i}$ be the component of $\boldsymbol{R}_y$ corresponding to the individual i. Utilizing the same notations with parametric regression, the joint density given follow-up times is proportional to

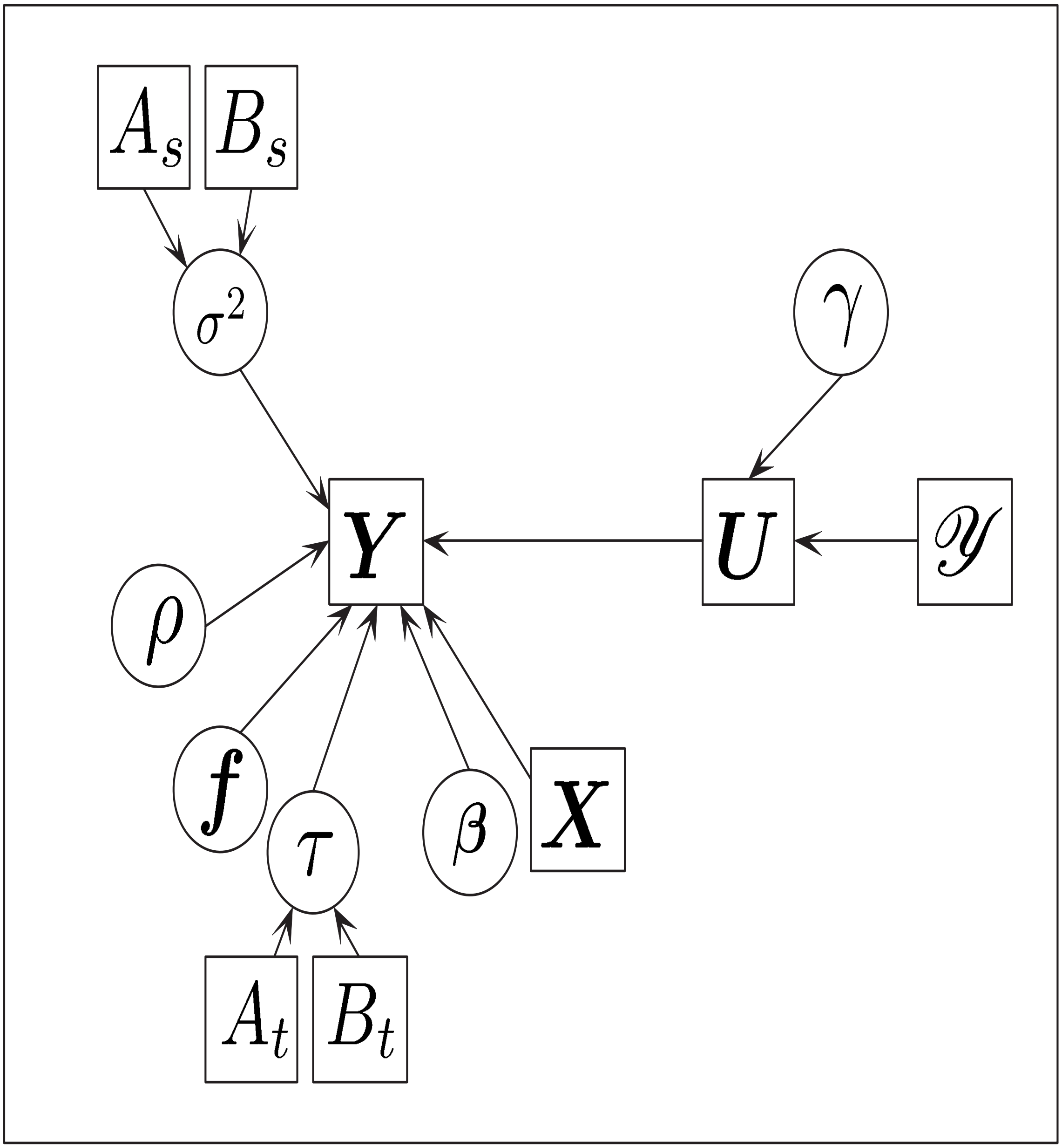

Fig. 8. Directed acyclic graph for Bayesian nonparametric regression under Model-0.

the following:

$$
\begin{aligned}
\propto & \quad (\sigma^2)^{-\frac{N}{2}} \prod_{i=1}^{n} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{R}_{y_i}^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{R}_{y_i}\right\} \\
\times & \quad \tau^{\frac{N-2}{2}} \exp\left\{-\frac{\tau}{2}\boldsymbol{f}^T \boldsymbol{K} \boldsymbol{f}\right\} \\
\times & \quad \tau^{A_t - 1} \exp\left\{-\frac{\tau}{B_t}\right\} \\
\times & \quad (\sigma^2)^{-(A_s+1)} \exp\left\{-\frac{1}{\sigma^2 B_s}\right\},
\end{aligned}
$$

where we restrict $\sum_{ij} f(t_{ij}) = 0$ to prevent the identifiability problem in the additive model. From the joint density, full conditionals for $\boldsymbol{f}$, τ, and other nuisance parameters are achieved. Let $\boldsymbol{A}(\alpha) = (\boldsymbol{\Sigma}_e^{-1} + \alpha \boldsymbol{K})^{-1}$. Then the full conditionals can be summaries as the following:

$$
\begin{aligned}
\boldsymbol{f}|\cdot & \sim N\left[\boldsymbol{A}(\alpha)\{\boldsymbol{\Sigma}_e^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\}, \boldsymbol{A}(\alpha)\sigma^2\right] \\
\tau|\cdot & \sim G\left[\frac{N-2}{2} + A_t, \left(\frac{\boldsymbol{f}^T \boldsymbol{K} \boldsymbol{f}}{2} + \frac{1}{B_t}\right)^{-1}\right], \\
\boldsymbol{\beta}|\cdot & \sim N\left[\left(\boldsymbol{X}^T \boldsymbol{\Sigma}_e^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \{\boldsymbol{\Sigma}_e^{-1}(\boldsymbol{Y} - \boldsymbol{f})\}, \left(\boldsymbol{X}^T \boldsymbol{\Sigma}_e^{-1} \boldsymbol{X}\right)^{-1} \sigma^2\right] \\
\sigma^2|\cdot & \sim IG\left[\frac{N}{2} + A_s, \left(\frac{1}{2}\boldsymbol{R}_y^T \boldsymbol{\Sigma}_e^{-1} \boldsymbol{R}_y + \frac{1}{B_s}\right)^{-1}\right], \\
\rho_i|\cdot & \propto |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{R}_{y_i}^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{R}_{y_i}\right\}, \qquad i = 1, \ldots, n,
\end{aligned}
$$

where $\boldsymbol{\Sigma}_i$ is the i^{th} diagonal element of the blocked diagonal matrix $\boldsymbol{\Sigma}_e$. As in the parametric regression, ρ_i, $i = 1, \ldots, n$, is generated via Nelder-Mead algorithm. For samplings from non-standard full conditionals including ρ_i, Gibbs sampling is utilized.

5.4. Model diagnostics

The conditional predictive ordinate (CPO) statistics introduced by Gelfand et al. (1992) is a useful model assessment tool using the marginal posterior predictive density of each response given data from rest of the observations. Let $\boldsymbol{\xi}$ be all parameters in the model, $\boldsymbol{D}$ be the data from all subjects, and $\boldsymbol{D}_{(i)}$ be the data not from the subject i. Further let $\boldsymbol{Y}_i$ be the response from subject i. Then, the CPO statistic for the subject i is defined as

$$\begin{aligned}
\text{CPO}_i &= f(\boldsymbol{Y}_i|\boldsymbol{D}_{(i)}) = \int_{\boldsymbol{\xi}} f(\boldsymbol{Y}_i|\boldsymbol{\xi})\pi(\boldsymbol{\xi}|\boldsymbol{D}_{(i)})d\boldsymbol{\xi} \\
&= \left\{E_{\boldsymbol{\xi}|\boldsymbol{D}}\left(\frac{1}{f(\boldsymbol{Y}_i|\boldsymbol{\xi})}\right)\right\}^{-1}.
\end{aligned}$$

First consider Model-0 with follow-up times $\boldsymbol{U}_i = (t_{i1}, U_{i2}, \ldots, U_{im_i})^T$ of the subject i. For Model-0, the distribution of $\boldsymbol{U}_i$ is ignorable for the predictive distribution of $\boldsymbol{Y}_i$, hence for given $\boldsymbol{U}_i$, the CPO for $\boldsymbol{Y}_i$ can be computed as

$$\text{CPO}_{i0} = \left[E_{\boldsymbol{\xi}|\boldsymbol{D}}\left\{\frac{1}{f(\boldsymbol{Y}_i|\boldsymbol{U}_i, \boldsymbol{\xi})}\right\}\right]^{-1},$$

where $\boldsymbol{\xi} = (\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$. Although the integration for the calculations of CPO is not trivial, we can utilize the MCMC samples of parameters and subject-specific random effects. Let the superscripts (q), $q = 1, \ldots, Q$, be the qth MCMC samples for corresponding parameters or subject-specific random effect. Then, under Model-0, the CPO of the subject i can be estimated by

$$\widehat{\text{CPO}}_{i0} = \left[\frac{1}{Q}\sum_{q=1}^{Q}\left\{\boldsymbol{\Psi}\left(\boldsymbol{\mu}^{(q)}, \sigma^{2(q)}\boldsymbol{\Sigma}_i^{(q)}\right)\right\}^{-1}\right]^{-1},$$

where $\boldsymbol{\Psi}(a, b)$ is a normal density with mean a and variance b. As a summary statistic of CPO over all subjects, the logarithm of the pseudomarginal likelihood (LPML) has been considered. The LPML (Ibrahim et al., 2001) is defined by

$$\text{LPML} = \frac{1}{n}\sum_{i=1}^{n}\log(\text{CPO}_i).$$

Hence, larger value of LPML indicates better fit of the regression model.

5.5. Simulation study

Leaving extra-covariates out of the consideration, only follow-up times were considered as covariates in our simulations. Bayesian parametric regression (B-P) and Bayesian nonparametric regression (B-NP) were compared to each other, under Model-0, by assuming three types of correlation structures of the outcome process:

$$\text{AR1} \;:\; \text{corr}(Y_{ik}, Y_{ij}) = \begin{cases} \rho_i^{|t_{ik}-t_{ij}|}, & \text{if } k \neq j \\ 1, & \text{if } k = j \end{cases},$$

$$\text{IND} \;:\; \text{corr}(Y_{ik}, Y_{ij}) = \begin{cases} 0, & \text{if } k \neq j \\ 1, & \text{if } k = j \end{cases},$$

$$\text{EXCH} \;:\; \text{corr}(Y_{ik}, Y_{ij}) = \begin{cases} \rho_i, & \text{if } k \neq j \\ 1, & \text{if } k = j \end{cases},$$

where ρ_i indicates the correlation coefficient of the subject i. The simulation results of the maximum likelihood estimation (MLE) from SAS PROC MIXED which Lipsitz et al. (2002) used were also compared to B-P and B-NP.

By assuming AR1 correlation structure, five simulation data sets were generated for 50 subjects ($n = 50$). For given follow-up time t, responses of the subject i were generated from:

$$\begin{aligned} Y_i(t) &= \mu(t) + \eta \log W_i + \epsilon_i(t), \\ \mu(t) &= \sin\left(\frac{(t+2)/3}{t^3/1000+1}\right) - \frac{1}{10}, \end{aligned}$$

where the response errors $\epsilon_i(t)$ were from $(\epsilon_{11}, \ldots, \epsilon_{nm_n})^T \sim N(0, \sigma^2 \mathbf{\Sigma}_e)$, $\sigma^2 = 0.1$, and W_i is a subject-specific random effect with the contribution coefficient η. In the follow-up times of the subject i, the entering time t_{i1} were generated from $Exp(2)$, where

$Exp(a)$ is the exponential distribution with mean a. The subsequent follow-up times were generated by the intervals between consecutive follow-ups $\boldsymbol{U}_i = (U_{i2}, \ldots, U_{im_i})$. Denoting the history of outcomes from the subject i before the k^{th} follow-up time as $\mathscr{Y}_{ik}$, U_{ik} were generated from an extreme value distribution:

$$f(U_{ik}|\mathscr{Y}_{ik}, W_i) = h_0(U_{ik})W_i \exp\{\gamma Y_{i,k-1} - H_0(U_{ik})W_i e^{\gamma Y_{i,k-1}}\}, \quad k = 2, \ldots, m_i,$$

where $h_0(t)$ and $H_0(t)$ are the baseline hazard and the corresponding cumulative hazard, respectively. True values of parameters and latent variable in the model of outcome process were assumed such that $\eta = -1$ and $\log W_i \sim N(0, \kappa)$, where $\kappa = 0.1$. The correlation coefficients ρ_i of the subject i in the correlation structures were generated from $U(0.4, 0.6)$. For the follow-up times, true parameter values were also assigned by $\lambda = 0.1$ with $\gamma = -1$ or 1. In addition, when $\gamma = -1$, ρ_i were also generated from $U(0.01, 0.99)$, and named the case as 'wide ρ_i'. Furthermore, the censoring time for the subject i was considered as the 75^{th} percentile of the distribution of $\boldsymbol{U}_i$ to keep it same with the simulation in Lipsitz et al. (2002). At most 20 follow-ups were allowed while their values are less than 20 ($m_i < 20$, $t_{ij} < 20$, where $j = 1, \ldots, m_i$ and $i = 1, \ldots, 50$). This simulation data set will be re-visited in section 6.4 under more complex model (Model-1).

To explore the effects of different types of assumed correlation structures and the performances of Bayesian parametric and nonparametric regression under Model-0, MSE and CPO have been evaluated for each case. In B-P, conjugate vague priors have been assigned for σ^2 and κ such that $IG(1, 1)$, where $IG(a, b)$ indicates inverse-gamma with mean $\{(a-1)b\}^{-1}$. Another conjugate prior were assigned for λ such that $G(1, 1)$, where $G(a, b)$ is a gamma distribution with mean ab. For other parameters, non-informative uniform priors have been assumed. In addition, for the nonparametric regression, the prior of smoothing parameter was set to be $\tau \sim G(3, 10)$. After 500 burning time, we took posterior mean of 1000 iterations as the estimates of parameters and regression function.

As shown in Table V, the correct specification of correlation structure (AR1) is indispensable for good fit. Among regression method, B-NP has shown the best performance in MSE. Figure 9 shows similar estimated curves by different types of correlation structures under Model-0, while AR1 performs slightly better. Utilizing Bayesian model assessment discussed in section 5.4, for the Bayesian parametric and nonparametric regressions, Table VI confirms the results of MSE with respect to CPO. In Table VI, larger value of summary statistic LPML=$\sum_{i=1}^{n} \log \mathrm{CPO}_i$ indicates a better fit of the model. For every value of γ, the Bayesian parametric regression (B-P) yields a smaller CPO statistic relative to the Bayesian nonparametric regression (B-NP), indicating a better fit for the Bayesian nonparametric regression. In LPML, the difference by three types of correlation structures and regression method are more distinguishable. Figure 9 also shows better fitting by assuming AR1 under Model-0.

5.6. Longitudinal study of cardiotoxicity

The effective treatment of chemotherapy doxorubicin of acute lymphoblastic leukemia in children has late cardiotoxic effects. To study cardiotoxic effects, the wall thickness of the heart was measured by examining echocardiograms. Data used in analysis were collected from 111 patients who had been completed chemotherapy. The considered covariates are cumulative doses of doxorubicin/m^2 of body surface (Dose, dichotomized at 350mg; 1 = 350 mg or more, 0 = less than 350 mg), age when the last treatment is taken (Age, ranging from 1.4 to 20.1 years), gender (Sex, 1 = female, 0 = male), and follow-up times since the end of chemotherapy (t, ranging from 0.1 to 13.86 years). Among the patients, 91 were measured more than once. For

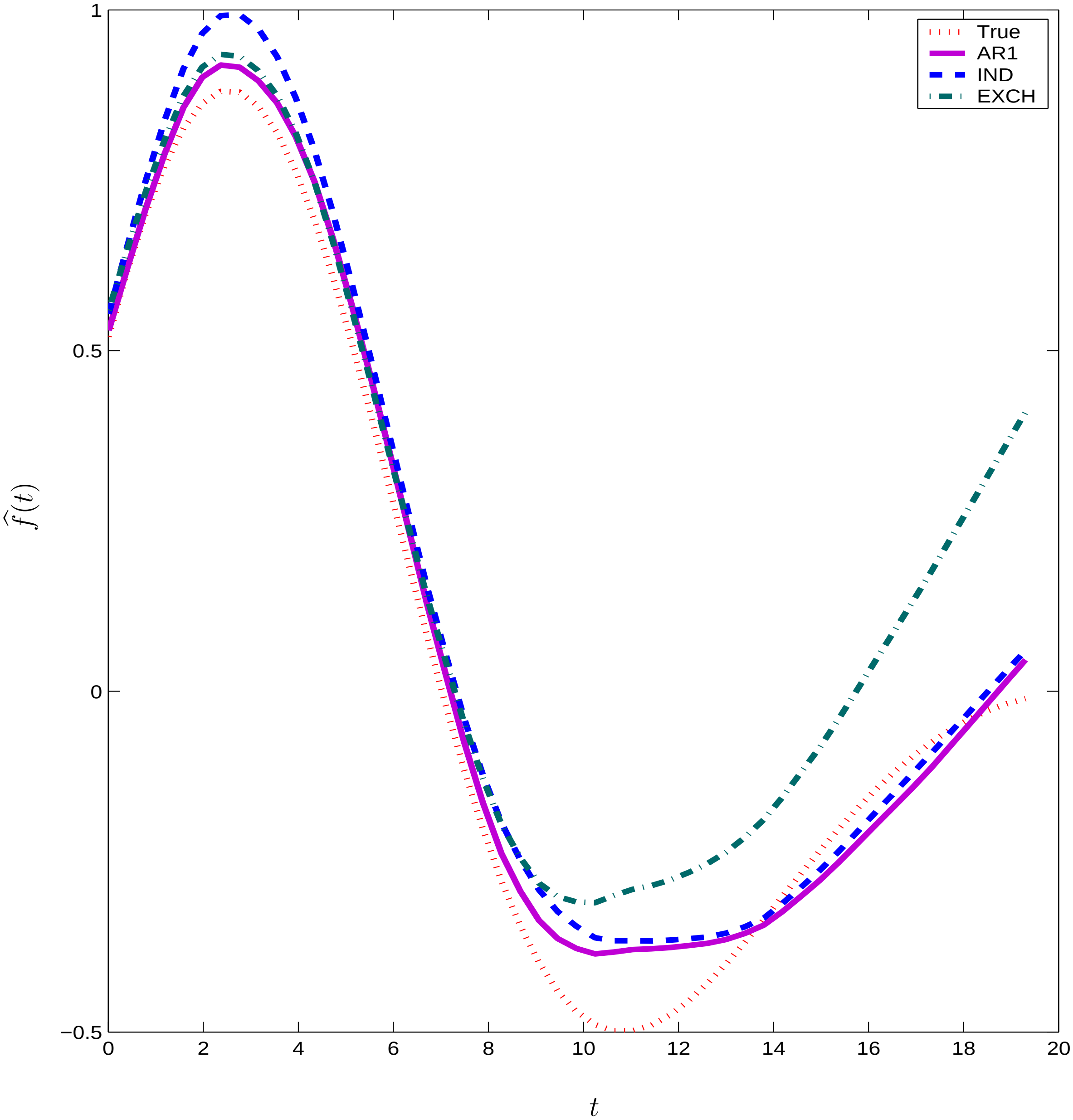

Fig. 9. Bayesian nonparametric fit under Model-0 when $\gamma = 1$. *At the beginning IND has a big bias and at the end EXCH has a big bias, while AR1 (true correlation structure) constantly shows a good estimation.*

those patients, the time to perform echocardiogram is not pre-specified but depends on previous outcomes. From all patients including twenty single visit patients, 329 measurement outputs were available. Figure 10 shows some samples of longitudinal measurements of heart wall thickness from ten patients.

According to Lipsitz et al. (2002), preliminary analysis of the data suggested the following regression function for the mean wall thickness of the heart at the measurement time t_{ij} from the individual i such that

$$\mu(t_{ij}) = \gamma_0 + \gamma_1 t_{ij} + \gamma_2 t_{ij}^2 + \beta_1 \text{age}_i + \beta_2 \text{gender}_i + \beta_3 \text{dose}_i + \beta_4 (\text{gender}_i \times \text{dose}_i).$$

As in section 5.5, with three correlation structures (AR1, IND, and EXCH), three regression methods (MLE, B-P, and B-NP) were explored, under Model-0.

As shown in Table VII, in most of cases, the estimated parameters are similar to each other, under the assumption of correlation structure as EXCH, the estimated regression coefficients for follow-up times (t_{ij}), Age, and Dose are slightly higher than under other correlation structures (IND and EXCH). The residual sum of squares (RSS) is given by

$$\text{RSS} = \frac{1}{N} \sum_{i}^{n} \sum_{j=1}^{m_i} \left\{ Y_{ij} - \widehat{\mu}(t_{ij}) \right\}^2,$$

where $N = \sum_{i=1}^{n}$. The RSS are not distinguishable under Model-0. Furthermore, outcomes from an individual may not be independent. Hence, the RSS may not assess model properly (Neter et al., 1990, pg. 484).

Since RSS only considers the fitted values based on entire data, alternatively, condition probability ordinate (CPO) has been considered as a better assessment tool to evaluate the effects of each model on the prediction at each data point. Table VIII describes summary statistic LPML of CPO. According to the table, AR1 and EXCH

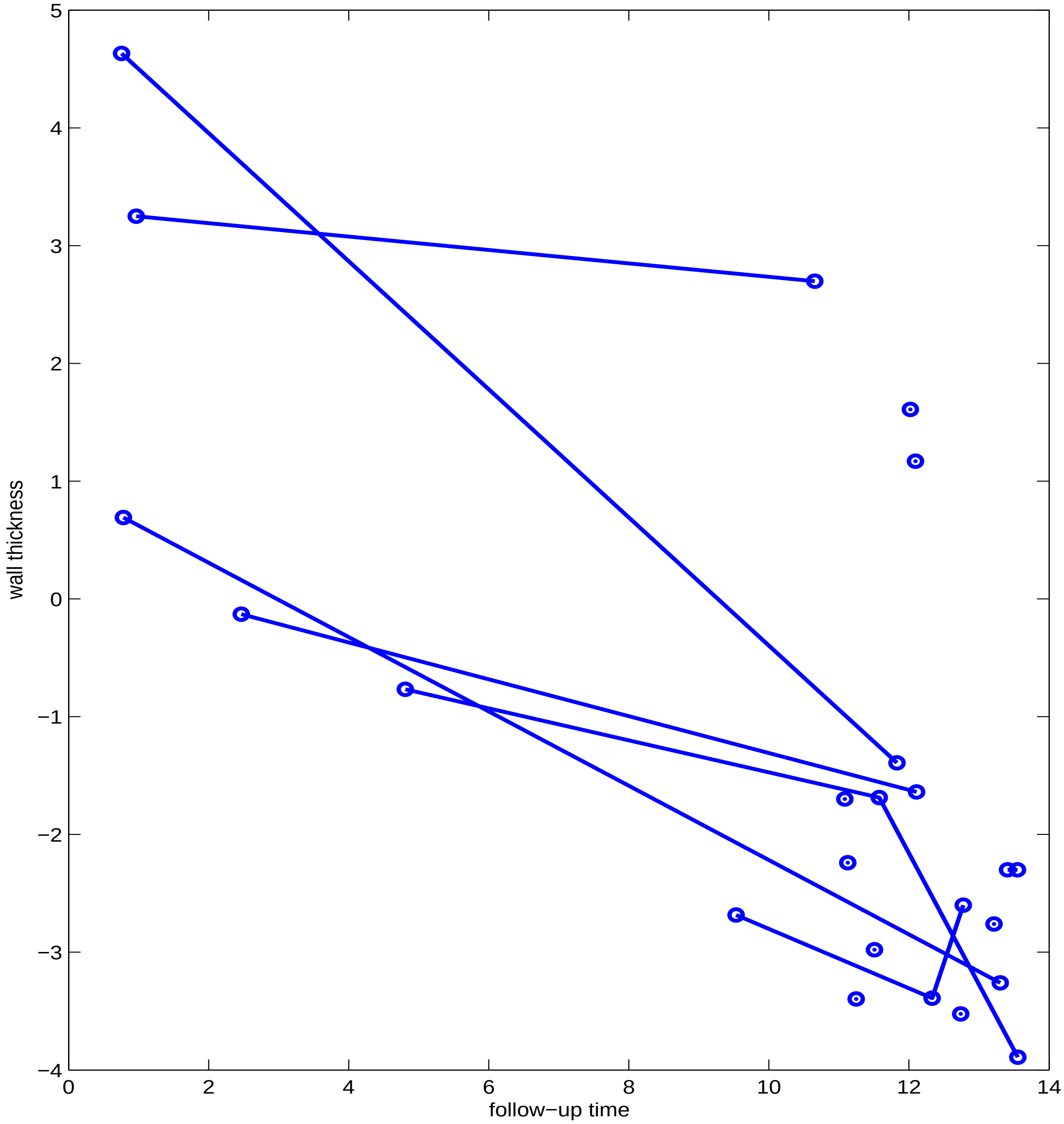

Fig. 10. Samples of longitudinal measurements of heart wall thicknesses. *Each line shows longitudinal measurements from one patient.*

bring better performance than IND, while EXCH possesses slightly more preferable results than AR1. On the other hand, the improvement of the performance by B-NP are negligible.

Figure 11 indicates that B-NP with EXCH correlation type leads to the most slope of the fitted curve, which is evaluated as the best fit with respect to LPML. Under Model-0, B-NP and B-P show very similar trend to each other (Figure 12), especially at the beginning and the ending of the curve. Hence, assuming Model-0, above results indicate that the type of correlation structure works more critically to the better fit than the regression method.

Table V. Mean Square Error (MSE) of Model-0: Five simulation data sets of 50 subjects.

Condition	Method	Correlation Structure		
		AR1	IND	EXCH
$\gamma = 1$, $\rho_i \sim U(0.4, 0.6)$	MLE	0.0634	0.0707	0.0692
	B-P	0.0654	0.0705	0.0697
	B-NP	0.0162	0.0255	0.0214
$\gamma = -1$, $\rho_i \sim U(0.4, 0.6)$	MLE	0.1104	0.1311	0.1090
	B-P	0.1086	0.1298	0.1083
	B-NP	0.0301	0.0502	0.0303
$\gamma = -1$, $\rho_i \sim U(0.01, 0.99)$	MLE	0.1082	0.1344	0.1086
	B-P	0.1067	0.1342	0.1079
	B-NP	0.0251	0.0521	0.0245

Note: In most of all cases, AR1 shows smaller MSE. In some other cases, AR1 and EXCH have similar MSEs which is less than that from IND. Comparing the regression method, while B-P and MLE have similar performance to each other, B-NP has much better performance.

Table VI. Logarithm of the Pseudomarginal Likelihood (LPML) of Model-0 for simulated data.

Condition	Method	Correlation Structure		
		AR1	IND	EXCH
$\gamma = 1$	B-P	-77.11	-108.85	-85.95
$\rho_i \sim U(0.4, 0.6)$	B-NP	-65.22	-90.92	-70.85
$\gamma = -1$	B-P	-95.85	-112.13	-104.06
$\rho_i \sim U(0.4, 0.6)$	B-NP	-75.51	-88.41	-80.07
$\gamma = -1$	B-P	-70.65	-99.88	-82.43
$\rho_i \sim U(0.01, 0.99)$	B-NP	-56.53	-73.36	-58.61

Table VII. Estimated regression parameters in cardiotoxic data under Model-0.

Method	CORR	t_{ij}	t_{ij}^2	Age	Sex	Dose	S*D	RSS
MLE	AR1	-0.1877	0.0062	0.0092	0.6148	0.5069	-1.0453	3.0604
	IND	-0.1845	0.0059	-0.0105	0.6212	0.5177	-1.0567	3.0379
	EXCH	-0.1464	0.0028	0.0117	0.7010	0.7089	-1.4141	3.0774
B-P	AR1	-0.1712	0.0054	0.0103	0.5892	0.5501	-1.2444	3.0709
	IND	-0.1835	0.0058	-0.0107	0.6158	0.5107	-1.0392	3.0379
	EXCH	-0.1517	0.0030	0.0128	0.7365	0.7258	-1.4405	3.0792
B-NP	AR1			0.0099	0.6275	0.6063	-1.2998	3.0639
	IND			-0.0116	0.6156	0.5433	-1.0574	3.0286
	EXCH			0.0101	0.6945	0.7405	-1.4162	3.0726

Table VIII. LPML under Model-0: Cardiotoxic data with 111 subjects.

Method	Correlation Structure		
	AR1	IND	EXCH
B-P	-639.43	-663.44	-633.47
B-NP	-636.50	-662.51	-633.88

Note: Summary statistic LPML=$\sum_{i=1}^{n} \log \mathrm{CPO}_i$ indicates better fit with larger value. In types of correlation structures, IND obtains the worst fit with all of the regressions (B-P and B-NP). While EXCH takes slight improvement of fit, AR1 and EXCH lead to similar results. Although B-NP has slightly better performance than B-P, the improvement is negligible.

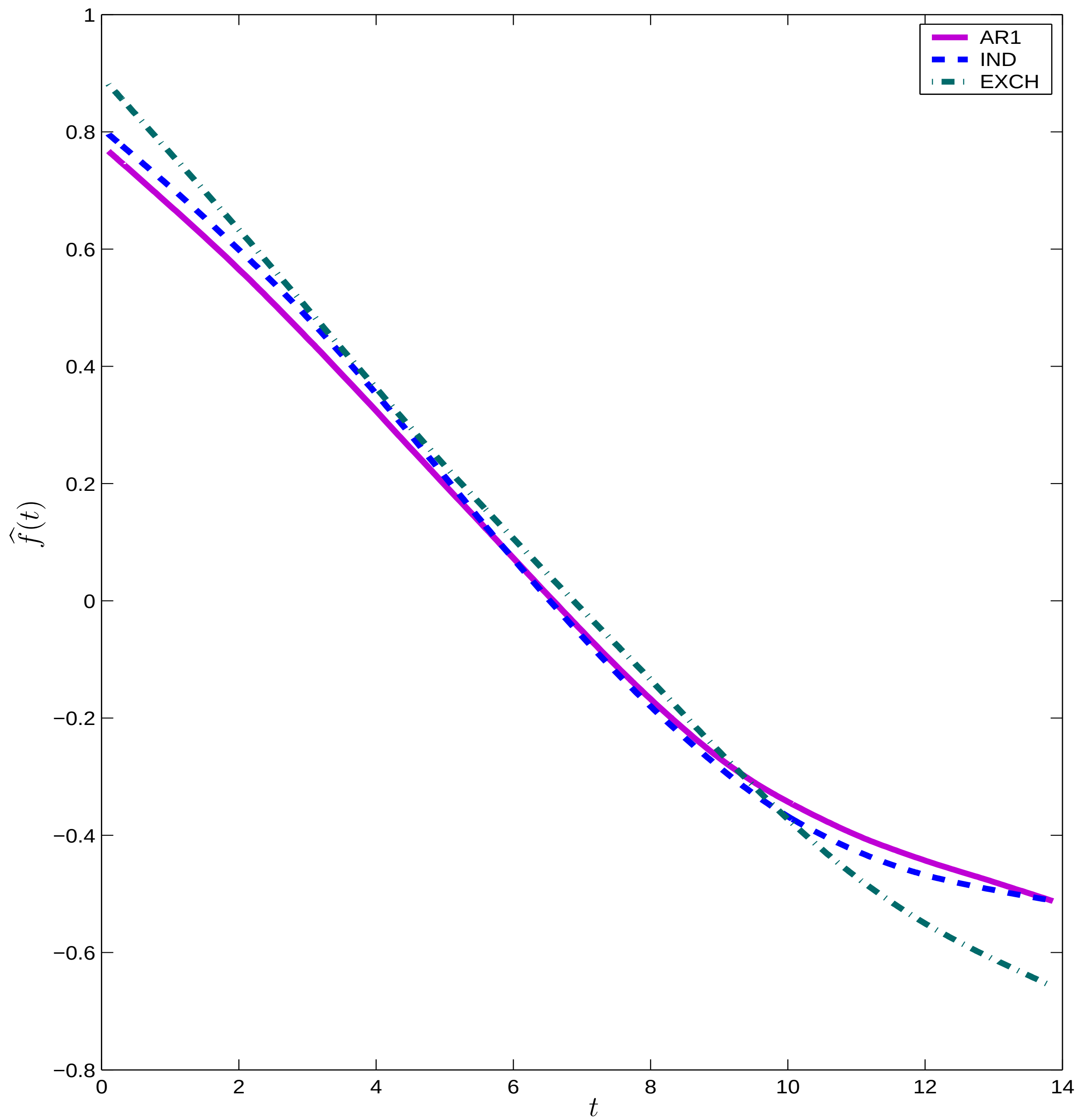

Fig. 11. Bayesian nonparametric fit for cardiotoxic data under Model-0. *While all three structures show similar trends, fitted curved from IND ends more slowly than others.*

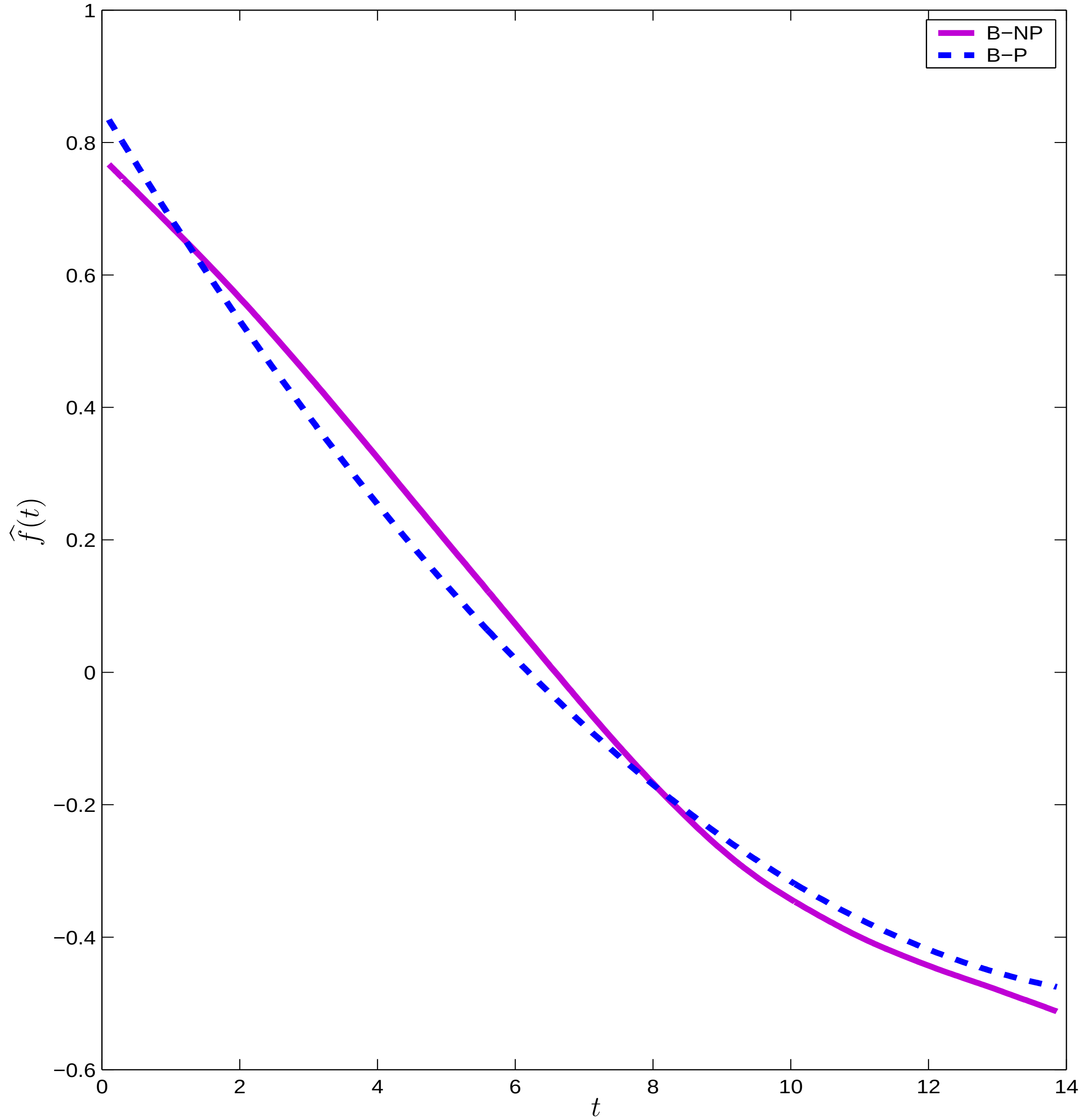

Fig. 12. Bayesian parametric and nonparametric fit for cardiotoxic data under Model-0. *Curves are estimated by the correlation type EXCH, under Model-0. B-P and B-NP produce very similar curves to each other.*

CHAPTER VI

BAYESIAN NONPARAMETRIC REGRESSION UNDER FLEXIBLE MODEL OF OUTCOME-DEPENDENT FOLLOW-UP TIMES

The same longitudinal measurement study with Chapter V has been considered, where the association among individual measurement times is assumed to be completely explained by observed history of longitudinal measurements. In this chapter, unobservable subject-specific random effects are considered in the outcome process for more flexible modeling.

Same notations of Chapter V have been utilized in this chapter. In section 6.1, parametric and nonparametric regression models are described, and in section 6.2, priors and induced posteriors for parametric and nonparametric regression are developed. To evaluate the performance of regression models and relaxed model assumption, section 6.3 customizes Bayesian method of model assessment and section 6.4 provides a simple simulation. Section 6.5 summaries the longitudinal study of the cardiotoxic effects of doxorubicin chemotherapy under more flexible modeling.

6.1. Model

As an alternative to Model-0 in Chapter V, a more complex model, called Model-1, has been considered. This new Model-1, addresses the modeling of the joint distribution of follow-up times $\boldsymbol{U}_i = (U_{i2}, \cdots, U_{im_i})$ for individual i via a latent frailty variable W_i, reflecting the subject-specific random effect. As in Model-0, let $\mu_i(t)$ be the regression function, and ϵ_{it} be the Gaussian error process with zero mean and covariance process

$\sigma^2\rho_{its}$. We describe outcome process as following model.

$$Y_i(t) = \mu_i(t) + \eta \log W_i + \epsilon_{it},$$
$$\log(W_i) \sim N(0, \kappa), \quad j = 1, \ldots, m_i, \; i = 1, \ldots, n,$$

where κ is a variance of the random subject-specific frailty effect W_i and η is the coefficient of the random frailty effect to the response $Y_i(t)$. In this model, follow-up time process contributes to the estimation of longitudinal response process $Y_i(t)$ through the shared frailty W_i. When a subject has only one observation ($m_i = 1$), we do not have follow-up process. However, when a subject has more than one observation ($m_i \geq 2$), we assume U_{ik}, $k = 2, \cdots, m_i$, given the frailty W_i and the observed history $(\mathscr{U}_{ik}, \mathscr{Y}_{ik})$ follow the survival distribution with the hazard rate

$$h_{ik}(u|\mathscr{U}_{ik}, \mathscr{Y}_{ik}, W_i) = h_0(u) W_i \exp(\gamma Y_{i,k-1}), \quad j = 2, \ldots, m_i,$$

where $h_0(u)$ is a baseline hazard rate function and γ is a coefficient of autoregressive model. We also assume that W_i are independent of the response errors ϵ_{it}. Please note that when $\eta = 0$, the model reduces essentially to Model-0 and distribution of observed follow-up times does not contribute to the estimation of parameters of the longitudinal response process $Y(t)$. This follow-up times model can also include a term involving a regression coefficient for the covariate vector $\boldsymbol{X}_i(t)$, however, for the data examples considered here, we omit this term from Model-1. Please note that model for follow-up times considered in Model-1 is similar to the log-normal frailty model for multivariate survival data considered by Hougaard (2000).

Let $H_0(u) = \int_0^u h_0(v)dv$ be the cumulative baseline hazard function. Then, equivalently, we can write the density function of U_{ik}, for $k = 2, \ldots, m_i$; $i = 1, \ldots, n$, as

$$f_{ik}(u|\mathscr{Y}_{ik}, \mathscr{U}_{ik}, W_i) = h_0(u) W_i \exp\left\{\gamma Y_{i,k-1} - H_0(u) W_i e^{\gamma Y_{i,k-1}}\right\}, \quad k = 2, \ldots, m_i.$$

For the sake of the simplicity, we assume $h_0(t) = \lambda$. More information about the estimation of baseline hazard function can be found at Fan et al. (1997); Horowitz and Lee (2004); Klein and Moeschberger (1997).

6.2. Bayesian regressions

As in chapter V, Bayesian hierarchical models have been considered for Bayesian parametric and nonparametric regressions under Model-1. Modeling starts with Bayesian parametric regression (B-P) and then proceed to Bayesian nonparametric regression (B-NP).

A. Parametric regression

The follow-up time $\boldsymbol{U}_i$ is modeled by utilizing a survival distribution with baseline hazard λ and the subject-specific random effect W_i is supposed to be from log-normal distribution with mean zero and variance κ as mentioned in section 6.1. Consider the same priors with Model-0 for the regression parameters and nuisance parameters introduced in Model-0. Additionally, suppose priors for parameters related to follow-up times such that $\lambda \sim G(A_l, B_l)$ and $\kappa \sim IG(A_0, B_0)$. For other parameters, suppose non-informative uniform priors. The vectors $\boldsymbol{Y}$, $\boldsymbol{t}$, $\boldsymbol{X}$, $\boldsymbol{Z}$, and $\boldsymbol{\Theta}$ denote same values as before. Let $\boldsymbol{R}^*_{y_i} = (\boldsymbol{Y}_i - \boldsymbol{Z}_i \boldsymbol{\Theta} - \eta \log W_i \boldsymbol{1})$, $i = 1, \ldots, n$, and $\boldsymbol{R}^*_y = (\boldsymbol{R}^T_{y_1}, \ldots, \boldsymbol{R}^T_{y_n})^T$.

Then the joint p.d.f can be summarized by

$$
\begin{aligned}
&\propto \quad (\sigma^2)^{-\frac{N}{2}} \prod_{i=1}^{n} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \boldsymbol{R}_{y_i}^{*T} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{R}_{y_i}^{*} \right\} \\
&\times \quad \prod_{i=1}^{n} \left(\kappa^{\frac{1}{2}} W_i \right)^{-1} \exp\left\{ -\frac{1}{\kappa} (\log W_i)^2 \right\} \\
&\times \quad \prod_{i=1}^{n} \prod_{j=2}^{m_i} \lambda W_i \exp\left(\gamma Y_{i,j-1} - \lambda U_{ij} W_i e^{\gamma Y_{i,j-1}} \right) \\
&\times \quad (\sigma^2)^{-(A_s+1)} \exp\left\{ -\frac{1}{\sigma^2 B_s} \right\} \lambda^{A_l - 1} \exp\left(-\frac{\lambda}{B_l} \right) \kappa^{-(A_k+1)} \exp\left(-\frac{1}{B_k \kappa} \right),
\end{aligned}
$$

where the conditional density for U_{ik} is exponential with hazard $\lambda W_i \exp(\gamma Y_{i,k-1})$. From the joint density, full conditionals for parameters and latent variable are calculated as the following:

$$
\begin{aligned}
\boldsymbol{\Theta}|\cdot \quad &\sim \quad N\left[\left(\boldsymbol{Z}^T \boldsymbol{\Sigma}_e^{-1} \boldsymbol{Z} \right)^{-1} \boldsymbol{Z}^T \boldsymbol{\Sigma}_e^{-1} (\boldsymbol{Y} - \eta \log \boldsymbol{W}), \left(\boldsymbol{Z}^T \boldsymbol{\Sigma}_e^{-1} \boldsymbol{Z} \right)^{-1} \sigma^2 \right] \\
\sigma^2|\cdot \quad &\sim \quad IG\left[\frac{N}{2} + A_s, \left(\frac{1}{2} \boldsymbol{R}_y^{*T} \boldsymbol{\Sigma}_e^{-1} \boldsymbol{R}_y^{*} + \frac{1}{B_s} \right)^{-1} \right], \\
\rho_i|\cdot \quad &\overset{ind}{\propto} \quad |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \boldsymbol{R}_{y_i}^{*T} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{R}_{y_i}^{*} \right\}, \qquad i = 1, \ldots, n, \\
W_i|\cdot \quad &\overset{ind}{\propto} \quad G\left[m_i, \left(\lambda \sum_{j=2}^{m_i} U_{ij} e^{\gamma Y_{i,j-1}} \right)^{-1} \right] \mathit{lognormal}\left[\frac{D_{w_i}}{C_{w_i}}, \frac{1}{C_{w_i}} \right], \quad i = 1, \ldots, n, \\
\eta|\cdot \quad &\sim \quad N\left(\frac{D_\eta}{C_\eta}, \frac{\sigma^2}{C_\eta} \right), \\
\lambda|\cdot \quad &\sim \quad G\left[\sum_{i=1}^{n} (m_i - 1) + A_l, \left\{ \sum_{i=1}^{n} \sum_{j=2}^{m_i} U_{ij} W_i e^{\gamma Y_{i,j-1}} + \frac{1}{B_l} \right\}^{-1} \right], \\
\gamma|\cdot \quad &\propto \quad \exp\left\{ \gamma \sum_{i=1}^{n} \sum_{j=2}^{m_i} Y_{i,j-1} - \lambda \sum_{i=1}^{n} \sum_{j=2}^{m_i} W_i U_{ij} e^{\gamma Y_{i,j-1}} \right\} \\
\kappa|\cdot \quad &\sim \quad IG\left[\frac{n}{2} + A_k, \left\{ \frac{1}{2} \sum_{i=1}^{n} (\log W_i)^2 + \frac{1}{B_k} \right\}^{-1} \right],
\end{aligned}
$$

where $\boldsymbol{W} = (W_1 \mathbf{1}_{m_1}^T, \ldots, W_n \mathbf{1}_{m_n}^T)^T$ and $\mathbf{1}_s$ is a $s \times 1$ vector of ones. Note that, for any subject with $m_i = 1$, the conditional of W_i is log-normal and conditional density of ρ_i is not required. The normalizing constants in the full conditional of W_i are $C_{w_i} = \frac{\eta^2}{\sigma^2} \mathbf{1}^T \boldsymbol{\Sigma}_i^{-1} \mathbf{1} + \frac{1}{\kappa}$, $D_{w_i} = \frac{\eta}{\sigma^2} \mathbf{1}^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{Z}_i \boldsymbol{\Theta})$, $i = 1, \ldots, n$, respectively, and the constants in the full conditional of η are $C_\eta = \sum_{i=1}^n (\log W_i)^2 (\mathbf{1}_{m_i}^T \boldsymbol{\Sigma}_i^{-1} \mathbf{1}_{m_i})$, and $D_\eta = \sum_{i=1}^n \log W_i \{\mathbf{1}_{m_i}^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{Z}_i \boldsymbol{\Theta})\}$. The full conditional of W_i does not follow any standard density when $m_i \geq 2$, but it has a form similar to a product of extreme value densities.

By utilizing the Gibbs sampling, samples from the posterior distributions of the parameters establish the samples from the joint p.d.f. Although ρ_i, W_i, and γ do not follow any known standard density, Metropolis-Hastings algorithm enables to generate them. As in chapter V, Nelder-Mead algorithm is applied to generate ρ_i.

B. Nonparametric regression

In Model-1, the priors for the follow-up times need to be specified explicitly. As in the parametric regression, let $\boldsymbol{U}_i = (U_{i2}, \ldots, U_{im_i})$ be the elapsed times and W_i be the subject-specific random effect from the individual i, $i = 1, \ldots, n$. Same distributions of $\boldsymbol{U}_i$ and W_i as in the parametric regression are assumed. In addition, the same priors for λ, η, κ, and γ as in the parametric regression are also assumed. Without ambiguity with parametric regression, we re-define the vector of residual such that $\boldsymbol{R}_{y_i}^* = (\boldsymbol{Y}_i - \boldsymbol{f}_i - \boldsymbol{X}_i \boldsymbol{\beta} - \eta \log W_i \mathbf{1}_{m_i})$, $i = 1, \ldots, n$, and $\boldsymbol{R}_y^*$ be the stacked version over all individuals. Then, the modified joint density for the nonparametric regression can be described by the following:

$$
\begin{aligned}
\propto\;& (\sigma^2)^{-\frac{N}{2}} \prod_{i=1}^{n} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{R}_{y_i}^{*T}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{R}_{y_i}^{*}\right\} \\
\times\;& \prod_{i=1}^{n}\left(\kappa^{\frac{1}{2}}W_i\right)^{-1}\exp\left\{-\frac{1}{2\kappa}(\log W_i)^2\right\} \\
\times\;& \prod_{i=1}^{n}\prod_{j=2}^{m_i}\lambda W_i \exp\left(\gamma Y_{i,j-1}-\lambda U_{ij}W_i e^{\gamma Y_{i,j-1}}\right) \\
\times\;& \tau^{\frac{N-2}{2}}\exp\left\{-\frac{\tau}{2}\boldsymbol{f}^T\boldsymbol{K}\boldsymbol{f}\right\}\tau^{A_t-1}\exp\left\{-\frac{\tau}{B_t}\right\} \\
\times\;& (\sigma^2)^{-(A_s+1)}\exp\left\{-\frac{1}{\sigma^2 B_s}\right\}, \\
\times\;& \lambda^{A_l-1}\exp\left(-\frac{\lambda}{B_l}\right)\kappa^{-(A_k+1)}\exp\left(-\frac{1}{B_k\kappa}\right),
\end{aligned}
$$

where $\boldsymbol{f}$ is restricted to be $\sum_{ij} f(t_{ij}) = 0$ to prevent the identifiability problem in the additive model. The directed acyclic graph for Bayesian nonparametric regression under Model-1 can be described by Figure 13. After all, the joint distribution is a combination of the Bayesian parametric regression under Model-1 and the Bayesian nonparametric regression under Model-0. Followings are full conditionals which are different from Model-0 in nonparametric regression:

$$
\begin{aligned}
\boldsymbol{f}|\cdot \;&\sim\; N\left[\boldsymbol{A}(\alpha)\boldsymbol{\Sigma}_e^{-1}(\boldsymbol{Y}-\boldsymbol{X\beta}-\eta\log\boldsymbol{W}),\boldsymbol{A}(\alpha)\sigma^2\right] \\
\boldsymbol{\beta}|\cdot \;&\sim\; N\left[\left(\boldsymbol{X}^T\boldsymbol{\Sigma}_e^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma}_e^{-1}\left(\boldsymbol{Y}-\boldsymbol{f}-\eta\log\boldsymbol{W}\right),\left(\boldsymbol{X}^T\boldsymbol{\Sigma}_e^{-1}\boldsymbol{X}\right)^{-1}\sigma^2\right] \\
\sigma^2|\cdot \;&\sim\; IG\left[\frac{N}{2}+A_s,\left(\frac{1}{2}\boldsymbol{R}_y^{*T}\boldsymbol{\Sigma}_e^{-1}\boldsymbol{R}_y^{*}+\frac{1}{B_s}\right)^{-1}\right], \\
\rho_i|\cdot \;&\propto\; |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{R}_{y_i}^{*T}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{R}_{y_i}^{*}\right\}, \qquad i=1,\ldots,n, \\
W_i|\cdot \;&\overset{ind}{\propto}\; G\left[m_i,\left(\lambda\sum_{j=2}^{m_i}U_{ij}e^{\gamma Y_{i,j-1}}\right)^{-1}\right] lognormal\left[\frac{D_{w_i}^*}{C_{w_i}},\frac{1}{C_{w_i}}\right], \quad i=1,\ldots,n, \\
\eta|\cdot \;&\sim\; N\left(\frac{D_\eta^*}{C_\eta},\frac{\sigma^2}{C_\eta}\right),
\end{aligned}
$$

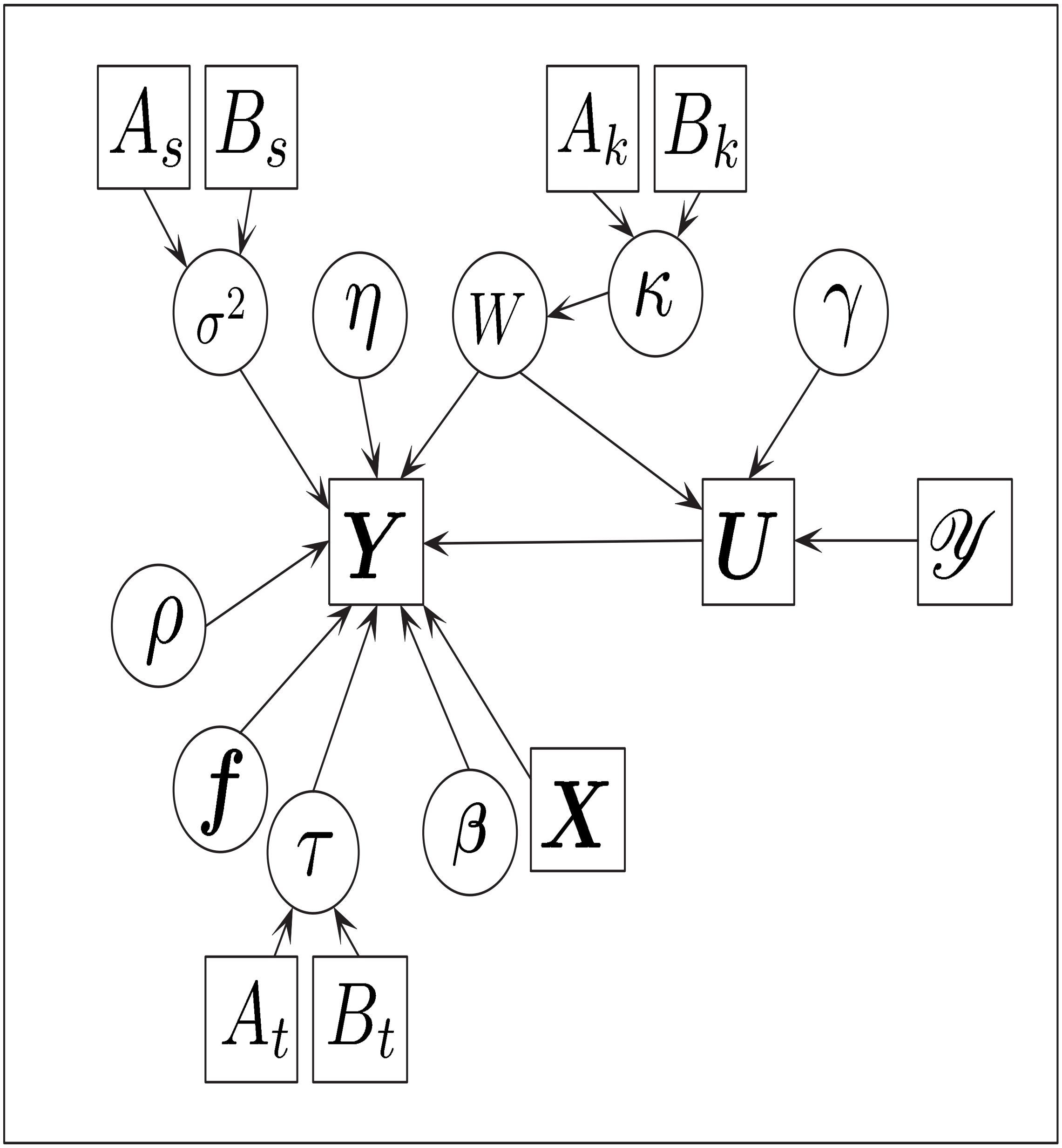

Fig. 13. Directed acyclic graph for Bayesian nonparametric regression under Model-1.

where $\mathbf{\Sigma}_i$, $\mathbf{\Sigma}_e$, and $\boldsymbol{W}$ are defined same as before. The constants in denominators of the full conditionals of W_i and η remain same as in the parametric regression, but the constants in the numerators need minor modifications such that $D^*_{w_i} = \frac{\eta}{\sigma^2}\mathbf{1}^T\mathbf{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{f}_i - \boldsymbol{X}_i\boldsymbol{\beta})$, and $D^*_\eta = \sum_{i=1}^n \log W_i\{\mathbf{1}^T_{m_i}\mathbf{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{f}_i - \boldsymbol{X}_i\boldsymbol{\beta})\}$. Other parameters have same full conditionals as in the parametric regression under Model-1; $[\lambda|\cdot]$ follows Gamma, $[\gamma|\cdot]$ is proportional to the product of generalized extreme value distributions, and $[\kappa|\cdot]$ follows inverse Gamma. The full conditional of the smoothing parameter is also same as in the nonparametric regression under Model-0 such that $[\tau|\cdot]$ follows Gamma.

Through the Gibbs sampling, samples for all parameters and latent variables are generated from the joint p.d.f. For non-standard full conditionals, Metropolis-Hasting's algorithm is utilized. Especially for ρ_i, Nelder-Mead algorithm is also applied, as in the parametric regression under Model-1.

6.3. Model diagnostics

For Model-1, $\boldsymbol{W} = (W_1, \ldots, W_n)^T$ is the vector of subject-specific random effects of all subjects. The distribution of $\boldsymbol{U}_i$ is associated with distribution of $\boldsymbol{Y}_i$ via the frailty W_i. The commonly used CPO defined by Gelfand et al. (1992) leading up to $\text{CPO}_{i1} = \int f(\boldsymbol{Y}_i, \boldsymbol{U}_i|\boldsymbol{\xi})\pi(\boldsymbol{\xi}|\boldsymbol{D}_{(i)})d\boldsymbol{W}d\boldsymbol{\xi}$ is not comparable to CPO_{i0} of Model-0, which is based only on the predictive distribution of $\boldsymbol{Y}_i|\boldsymbol{D}_{(i)}$. To remedy this problem of comparing CPO_{i0} based on $\boldsymbol{Y}_i|\boldsymbol{D}_{(i)}$ versus CPO_{i1} based on $\boldsymbol{Y}_i, \boldsymbol{U}_i|\boldsymbol{D}_{(i)}$, we introduce a novel idea of Conditional CPO (CCPO) for Model-1 based on $\boldsymbol{Y}_i, \boldsymbol{U}_i|\boldsymbol{D}_{(i)}$. For

Model-1, given $\boldsymbol{U}_i$, CCPO of the subject i is

$$\begin{aligned}\text{CCPO}_i &= \int f(\boldsymbol{Y}_i|\boldsymbol{U}_i, \boldsymbol{W}, \boldsymbol{\xi}) f(\boldsymbol{W}, \boldsymbol{\xi}|\boldsymbol{D}_{(i)}) d\boldsymbol{W} d\boldsymbol{\xi} \\ &= \frac{E_{W_i,\boldsymbol{\xi}|\boldsymbol{D}}\left(\frac{1}{f(\boldsymbol{U}_i|W_i,\boldsymbol{\xi})}\right)}{E_{W_i,\boldsymbol{\xi}|\boldsymbol{D}}\left(\frac{1}{f(\boldsymbol{Y}_i,\boldsymbol{U}_i|W_i,\boldsymbol{\xi})}\right)}\end{aligned}$$

The above equation shows that CCPO can be computed using samples from full posterior $\pi(\boldsymbol{\xi}|\boldsymbol{D})$. Details for the calculation of CCPO can be found at the Appendix.

The integration for the calculations of CCPO is not trivial. However, the MCMC samples of parameters and subject-specific random effects can be utilized to obtain CCPO. Under Model-1, to calculate CCPO, we need values of the density of $\boldsymbol{U}_i$ and the joint density of $(\boldsymbol{Y}_i, \boldsymbol{U}_i)$ given $(W_i, \boldsymbol{\xi})$. From the joint density in section 6.2, we can drive the following density functions:

$$\begin{aligned}f(\boldsymbol{Y}_i, \boldsymbol{U}_i|W_i, \boldsymbol{\xi}) &\propto \boldsymbol{\Psi}[\boldsymbol{\mu}(\boldsymbol{t}_i) + \eta \log W_i, \sigma^2 \boldsymbol{\Sigma}_i] \\ &\times \prod_{j=2}^{m_i} \lambda W_i \exp\left(\gamma Y_{i,j-1} - \lambda U_{ij} W_i e^{\gamma Y_{i,j-1}}\right), \\ f(\boldsymbol{U}_i|W_i, \boldsymbol{\xi}) &= \frac{f(\boldsymbol{Y}_i, \boldsymbol{U}_i|W_i, \boldsymbol{\xi})}{f(\boldsymbol{Y}_i|\boldsymbol{U}_i, W_i, \boldsymbol{\xi})} \\ &\propto \prod_{j=2}^{m_i} \left\{\lambda W_i \exp\left(\gamma Y_{i,j-1} - \lambda U_{ij} W_i e^{\gamma Y_{i,j-1}}\right)\right\},\end{aligned}$$

where $\boldsymbol{\Psi}(a, b)$ is a normal density with mean a and variance b. Hence, with MCMC samples, CCPO can be achieved by

$$\widehat{\text{CCPO}}_{i1} = \frac{\frac{1}{Q}\sum_{q=1}^{Q}\left(\frac{1}{f(\boldsymbol{U}_i|W_i^{(q)},\boldsymbol{\xi}^{(q)})}\right)}{\frac{1}{Q}\sum_{q=1}^{Q}\left(\frac{1}{f(\boldsymbol{Y}_i,\boldsymbol{U}_i|W_i^{(q)},\boldsymbol{\xi}^{(q)})}\right)}$$

As a summary statistic of CCPO, the LPML in section 5.4 is utilized.

$$\text{LPML} = \frac{1}{n}\sum_{i=1}^{n} \log(\text{CCPO}_i).$$

6.4. Simulation study

The five simulated data sets in section 5.5 have been re-used. Hence, simulated data includes three cases: $\gamma = 1$, $\gamma = -1$, and wide ρ. Under Model-1, three types of correlation structures (AR1, IND, and EXCH) defined in section 5.5 have been assessed with Bayesian parametric regression (B-P) and Bayesian nonparametric regression (B-NP). In addition to the priors in section 5.5, a non-informative uniform prior were assigned to the contribution coefficient η, and flexible conjugate priors were assigned to the variances of the outcome process and the subject-specific random effect such that $\lambda \sim G(1,1)$ and $\kappa \sim IG(1,1)$.

For the model assessment, as in section 5.5, MSE and customized CPO statistic have been utilized. Tables IX and X show that B-NP outperforms B-P. In Table X, larger value of summary statistic LPML=$\sum_{i=1}^{n} \log \text{CCPO}_i$ indicates a better fit of the model. For every value of γ, the Bayesian parametric regression (B-P) yields a smaller CPO statistic relative to the Bayesian nonparametric regression (B-NP), indicating a better fit for the Bayesian nonparametric regression. For the comparison of two models by MSE (Table V versus Table IX) and by LPML (Table VI versus Table X), Model-1 gives better fits than Model-0 for all conditions and correlation structures (AR1, IND, EXCH). In addition, Model-1 brings less variation in CPO statistics than those of Model-0 when we vary the assumed correlation structure. This indicates the robust behavior of Model-1 relative to Model-0 with respect to any assumed correlation structure. Apparently, the correct specification of the type of correlation structure also bring a better fit than other, as Lipsitz et al. (2002)

Table IX. Mean Square Error (MSE) of Model-1: Five simulation data sets of 50 subjects.

Condition	Method	Correlation Structure		
		AR1	IND	EXCH
$\gamma = 1$	B-P	0.0676	0.0704	0.0695
$\rho_i \sim U(0.4, 0.6)$	B-NP	0.0159	0.0225	0.0220
$\gamma = -1$	B-P	0.1078	0.1089	0.1076
$\rho_i \sim U(0.4, 0.6)$	B-NP	0.0288	0.0291	0.0306
$\gamma = -1$	B-P	0.1069	0.1061	0.1081
$\rho_i \sim U(0.01, 0.99)$	B-NP	0.0224	0.0261	0.0239

Note: In every case, B-NP has much smaller MSE than B-P. Comparing to Table V, MSE of Model-1 has been less affected by the type of the correlation structure.

Table X. Logarithm of the Pseudomarginal Likelihood (LPML) of Model-1 for simulated data.

Condition	Method	Correlation Structure		
		AR1	IND	EXCH
$\gamma = 1$	B-P	-74.54	-85.15	-85.58
$\rho_i \sim U(0.4, 0.6)$	B-NP	-53.00	-58.01	-59.53
$\gamma = -1$	B-P	-77.25	-91.79	-90.76
$\rho_i \sim U(0.4, 0.6)$	B-NP	-55.61	-57.34	-68.21
$\gamma = -1$	B-P	-67.82	-80.71	-81.67
$\rho_i \sim U(0.01, 0.99)$	B-NP	-49.98	-53.50	-66.28

have studied. Hence, the best performance is achieved by B-NP and AR1 under Model-1. Figure 14 shows fitted curves by B-NP with different types of correlation structure. Figures 15 and 16 compare the performance of B-NP, which shows the best performance under each model. Even in this comparison, Model-1 produces better performance. In addition, Figure 17 indicates that the true value of ρ_i is not estimated perfectly but estimated well by assuming different ρ_i on each subject.

6.5. Longitudinal study of cardiotoxicity

Under Model-1, the same cardiotoxic data in section 5.6 have been analyzed. With the same regression function $\mu(t)$, three types of correlation structures (AR1, IND, and EXCH) have been applied to Bayesian parametric and nonparametric regressions (B-P and B-NP), which have been performed under Model-0 in section 5.6. However, under Model-1, the subject-specific random effect (W_i, $i = 1, \ldots, n$) and the parameters in the distribution of follow-up times are required to be estimated.

Table XI summarizes the estimated parameters and residual sum of squares. The estimated parameters, under Model-1, takes similar values to section 5.6. Estimating parameters ($\mu(t_{ij})$, η, W_i) with its corresponding posterior means of MCMC iterations ($\widehat{\mu}(t_{ij})$, $\widehat{\eta}$, $\widehat{W}_i$), for the subject i, the residual sum of squares (RSS) is given by

$$\text{RSS} = \frac{1}{N} \sum_{i}^{n} \sum_{j=1}^{m_i} \left\{ Y_{ij} - \widehat{\mu}(t_{ij}) - \widehat{\eta} \log \widehat{W}_i \right\}^2 ,$$

where $N = \sum_{i=1}^{n}$. Although RSS showed similar performance in B-NP and B-P, for given correlation structure, with the same reason of section 5.6, LPML is preferred to RSS. In LPML, B-NP with EXCH has produced the best performance (Table XII). The estimated curves in Figures 18 and 19 also show similar patterns by different types of correlation structures and different regression methods. In addition, as shown in

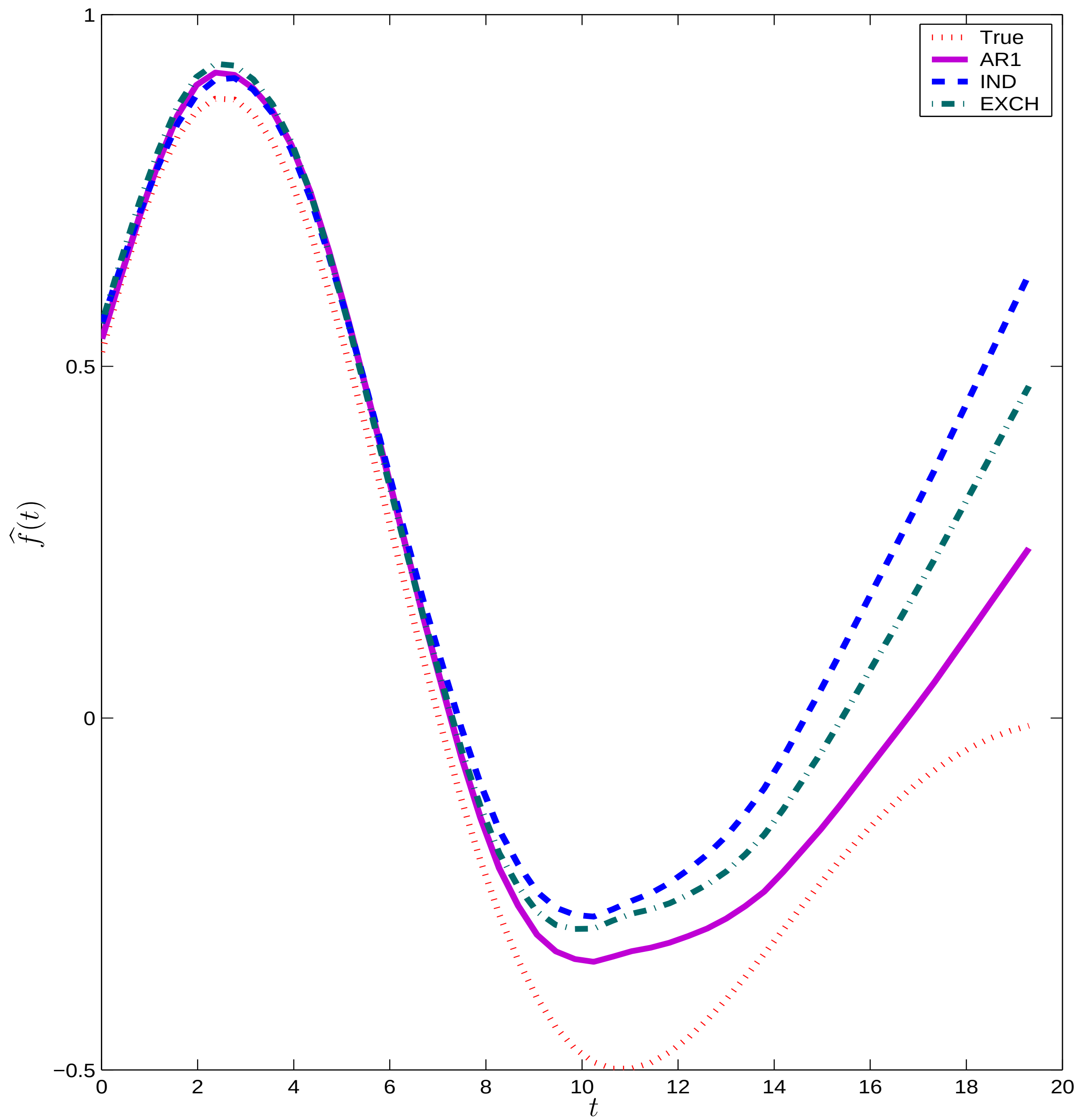

Fig. 14. Bayesian nonparametric fit under Model-1 when $\gamma = 1$. *At the beginning three correlation structures show a similar result, but at the end AR1 leads to the best estimation. The correct specification of correlation structure is important to estimate the true curve under Model-1.*

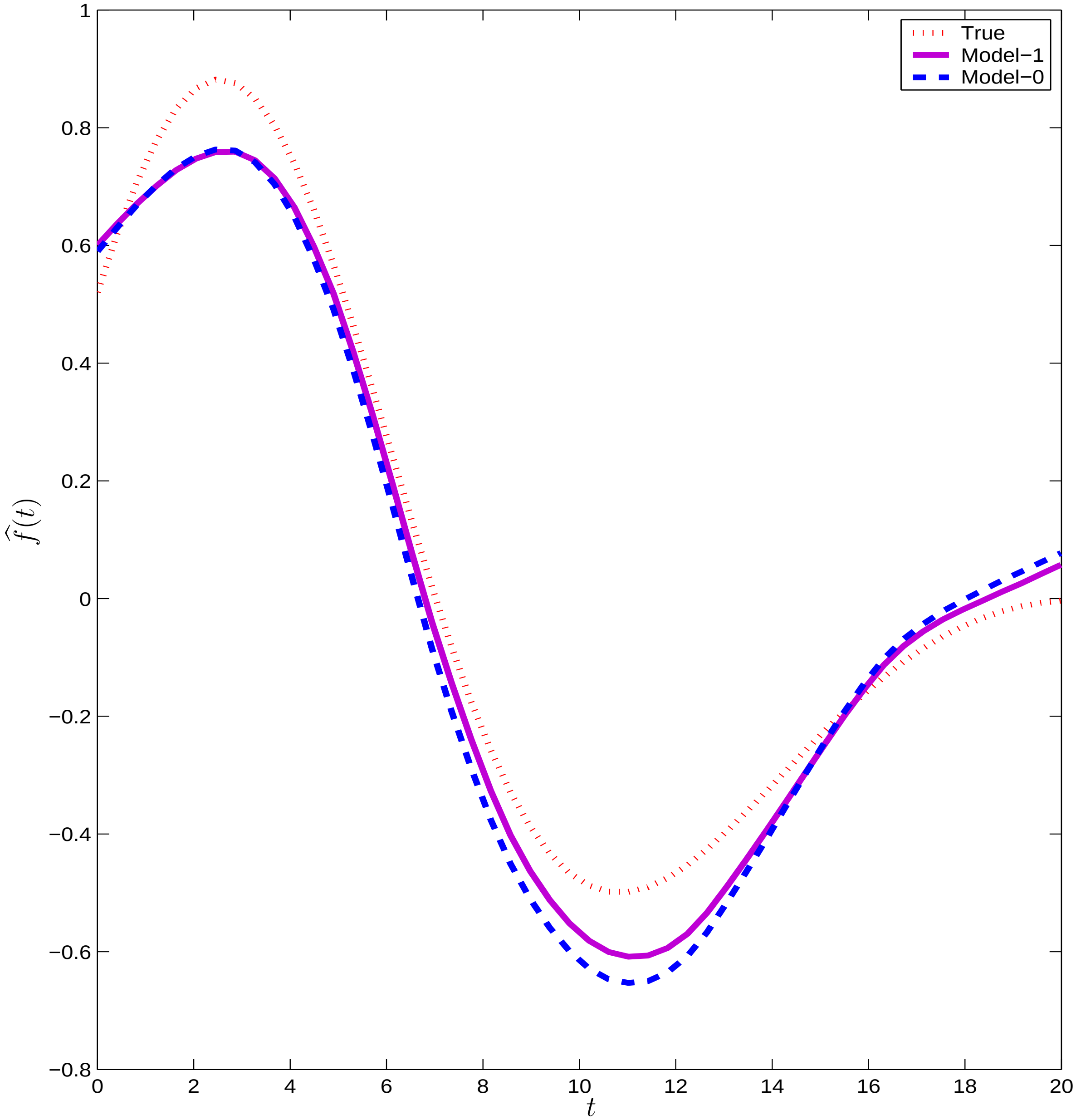

Fig. 15. Bayesian nonparametric fit under Model-0 and Model-1. *When the coefficient of autoregression* $\gamma = -1$*, we assumed true AR1 correlation structure in the fitting, under Model-0 and Model-1. Two models produce similar lines, but Model-1 shows slightly better performance. Because Model-1 reflects the effect of outcome-dependency, it improves the performance of nonparametric fitting.*

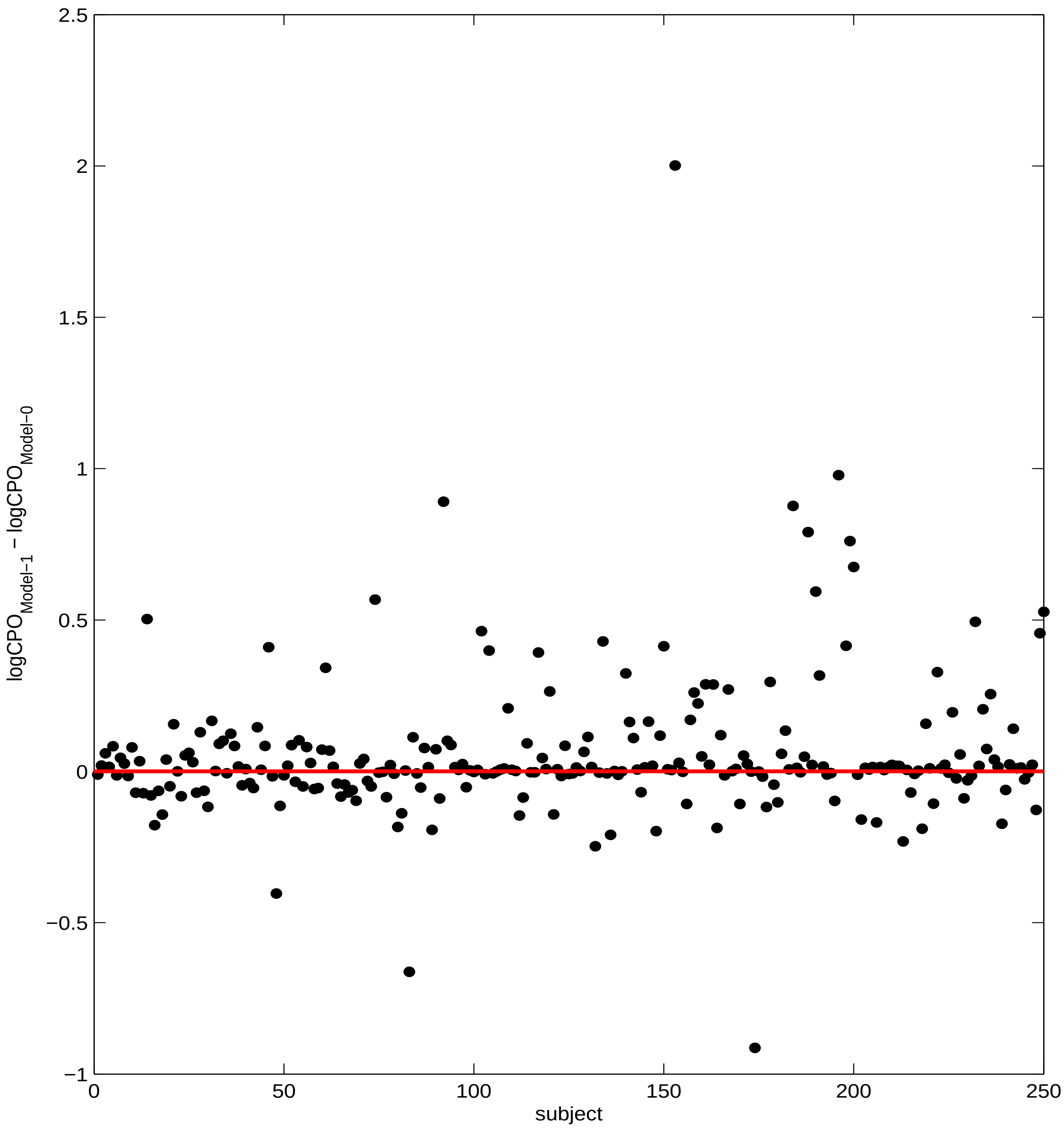

Fig. 16. Comparison CPOs from Model-1 to Model-0 for simulated data. *While Model-1 and Model-0 produce similar CPO, 60% of log CPO ratios for Model-1 versus Model-0 are positive. That is, Model-1 explains data better than Model-0.*

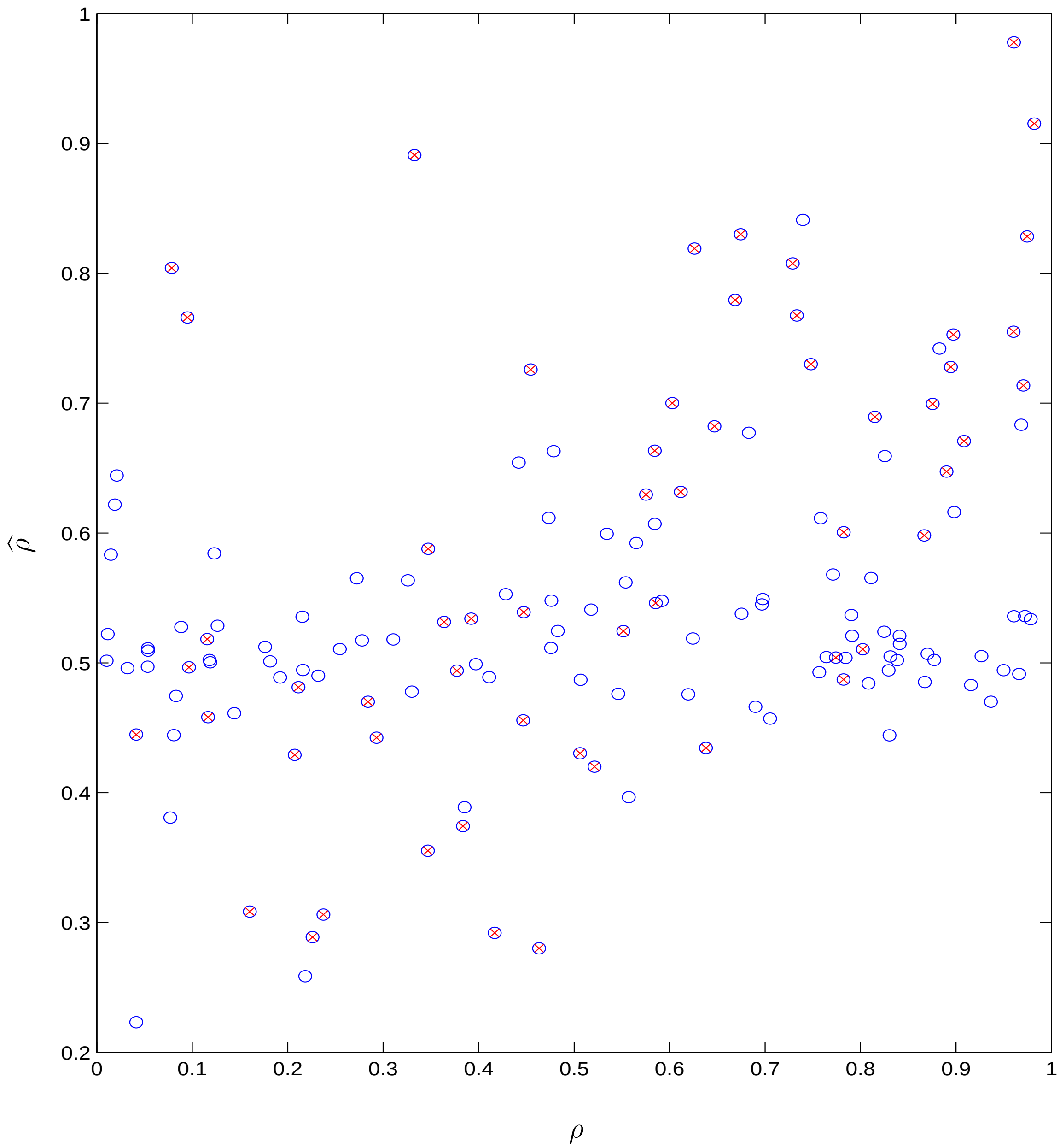

Fig. 17. Estimated correlation coefficient of each subject. *For the case of wide* ρ*, Model-1 has been performed under AR1. Because* ρ_i *is defined only when* $m_i > 1$*, plot indicates* $\widehat{\rho}$ *for the corresponding individuals. Especially, when* $m_i > 3$*,* $\widehat{\rho}_i$ *is denoted by* $\otimes$ *to indicate individuals with more information of follow-up times. In most cases, the true* ρ_i *is relatively well estimated. In addition, more information of follow-ups results in better estimation of* ρ_i*.*

Figure 20, each individual shows a different subject-specific random effect. Hence, Model-1 is expected to bring more reliable results.

Comparing the results from Model-0 and Model-1 by RSS (Table VII versus Table XI), Model-1 has slightly less RSS in each type of correlation structure. With respect to LPML (Table VIII versus Table XII), Model-1 also shows slight improvement of the performance.

In summary, the correct specification of the type of correlation structure is critical to get better performance of fit. In addition, the application of B-NP and Model-1 instead of B-P and Model-0 results in better fit of the regression model with the outcome-dependent follow-up data.

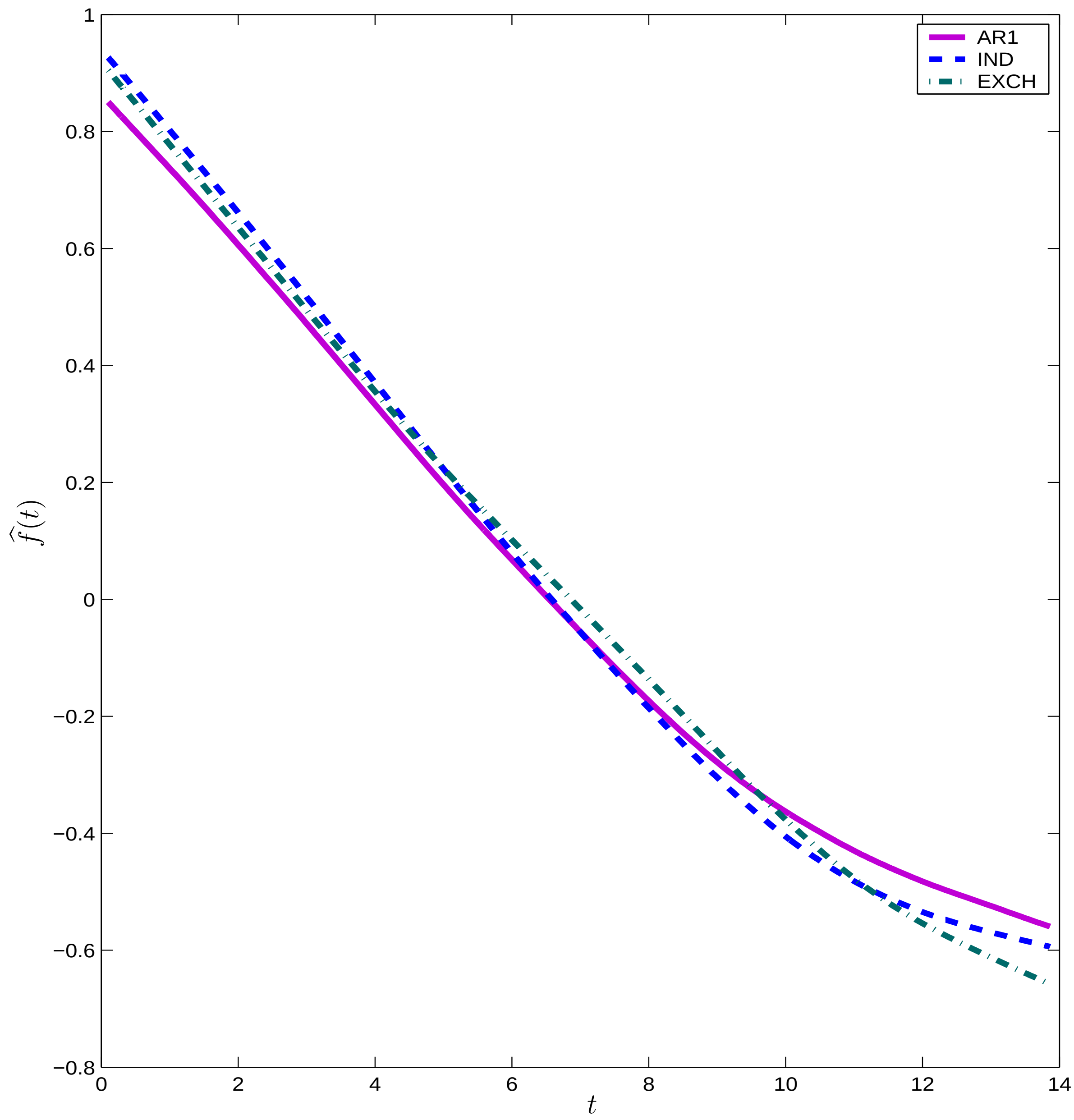

Fig. 18. Bayesian nonparametric fit for cardiotoxic data under Model-0. *While all three structures show similar trends, fitted curved from IND ends more slowly than others.*

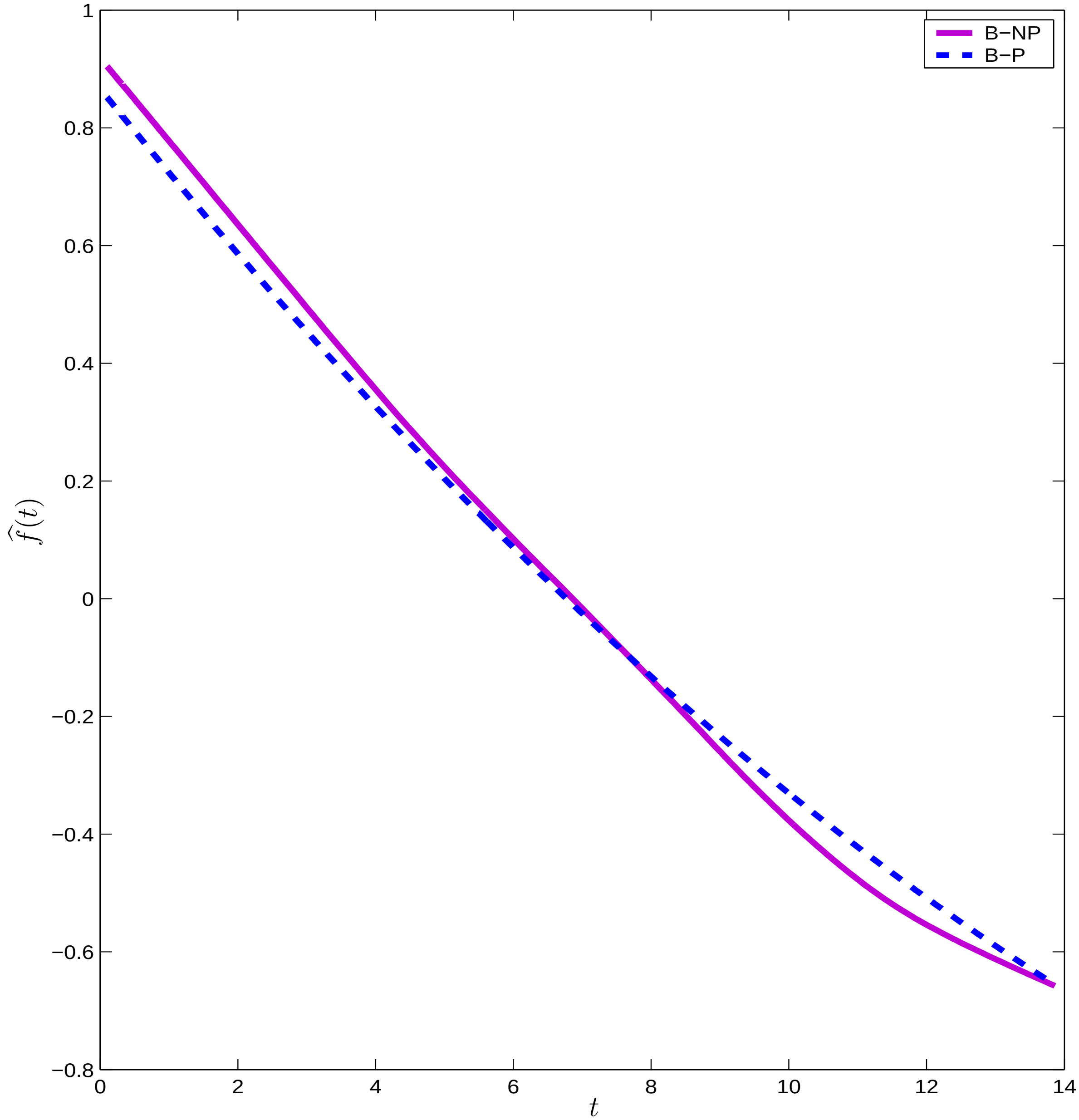

Fig. 19. Bayesian parametric and nonparametric fit for cardiotoxic data under Model-0. *Curves are estimated by the correlation type EXCH, under Model-0. B-P and B-NP produce very similar curves to each other.*

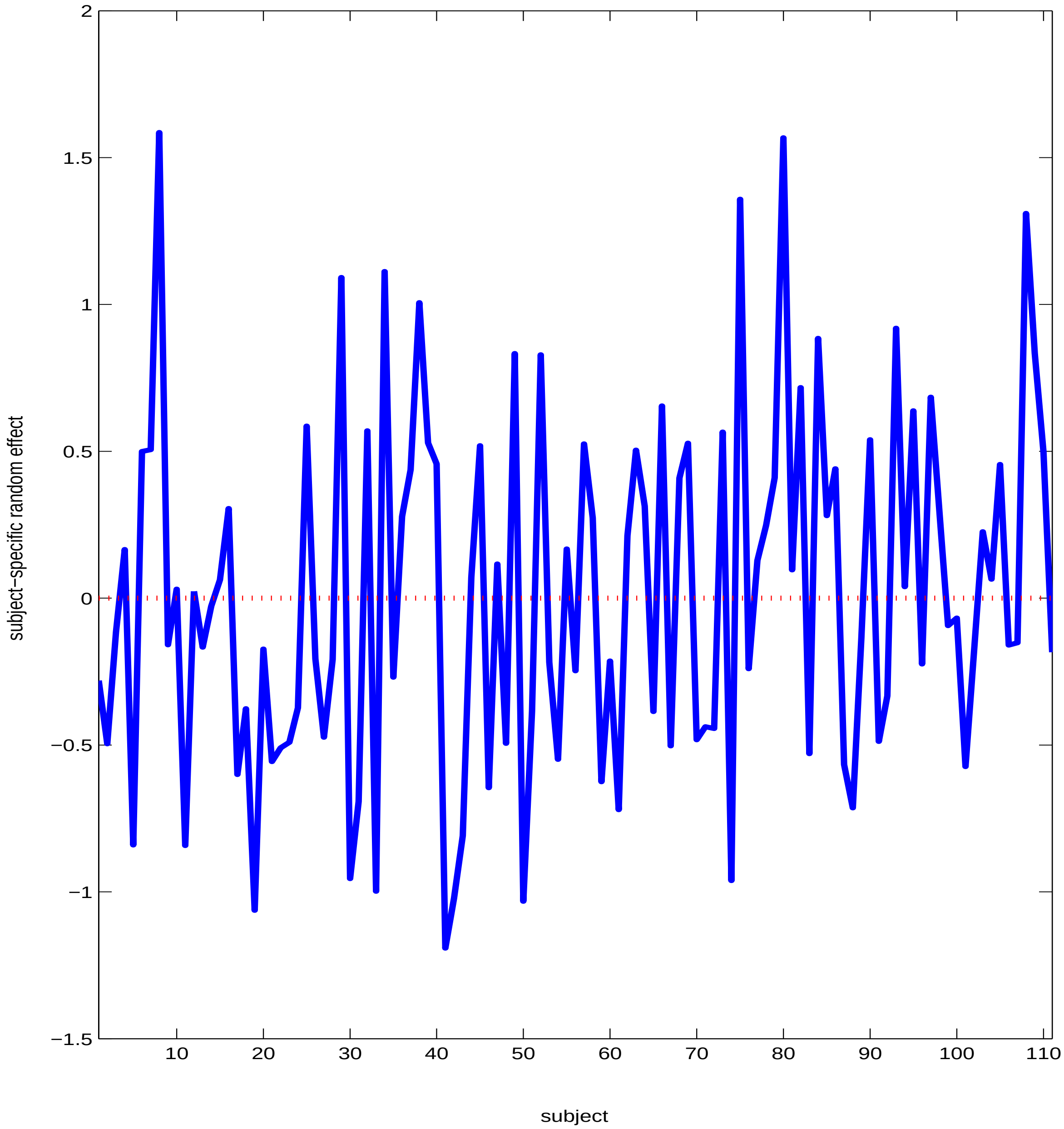

Fig. 20. Estimated subject-specific random effects for cardiotoxic data. *From Bayesian nonparametric regression with EXCH correlation type, subject-specific random effects show different values for each subject, where of* $\widehat{\eta}$ *is estimated by* 0.3276.

Table XI. Estimated regression parameters in cardiotoxic data under Model-1.

Method	CORR	t_{ij}	t_{ij}^2	Age	Sex	Dose	S*D	RSS
B-P	AR1	-0.1882	0.0061	0.0104	0.5610	0.6364	-1.1945	2.5310
	IND	-0.2025	0.0062	0.0004	0.5475	0.6246	-0.9721	2.0384
	EXCH	-0.1453	0.0026	0.0091	0.6358	0.7354	-1.3170	2.9402
B-NP	AR1			0.0120	0.5812	0.6325	-1.1745	2.5612
	IND			-0.0013	0.5409	0.6308	-0.9368	2.0099
	EXCH			0.0099	0.6912	0.8019	-1.3904	2.9489

Table XII. LPML under Model-1: Cardiotoxic data with 111 subjects.

Method	Correlation Structure		
	AR1	IND	EXCH
B-P	-638.72	-663.43	-631.98
B-NP	-633.97	-660.54	-629.79

Note: Summary statistic LPML=$\sum_{i=1}^{n} \log \mathrm{CPO}_i$ indicates better fit with larger value. As in Table VIII, AR1 and EXCH accomplish better fit and B-NP appears to be slightly advantageous, while all values are slightly better than that.

REFERENCES

Albert, J. H. and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association*, 88:669–679.

Berry, S. M., Carroll, R. J., and Ruppert, D. (2002), "Bayesian Smoothing and Regression Splines for Measurement Error Problems." *Journal of the American Statistical Association*, 97:160–169.

Black, D., Sanders, S., and Taylor, L. (2003), "Measurement of Higher Education in the Census and Current Population Survey." *Journal of the American Statistical Association*, 98:545–554.

Carroll, R. J., Roeder, K., and Wasserman, L. (1999), "Flexible Parametric Measurement Error Models." *Biometrics*, 55:44–54.

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995), *Measurement Error in Nonlinear Models*. New York: Chapman and Hall.

Carroll, R. J., Spiegelman, C. H., Lan, K. K. G., Bailey, K. T., and Abbott, R. D. (1984), "On Errors-in-Variables for Binary Regression Models." *Biometrika*, 71:19–25.

Carter, C. K. and Kohn, R. (1994), "On Gibbs Sampling for State Space Models." *Biometrika*, 81:541–553.

Chen, H. Y. (2003), "A Note on the Prospective Analysis of Outcome-Dependent Samples." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65:575–584.

Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000), *Monte Carlo Methods in Bayesian Computation*. New York: Springer.

Denison, D., Mallick, B., and Smith, A. (1997), "Bayesian MARS." URL citeseer.ist.psu.edu/denison97bayesian.html.

Diaconis, P. and Freedman, D. A. (1993), "Nonparametric Binary Regression: A Bayesian Approach." *The Annals of Statistics*, 21:2108–2137.

Eubank, R. L. (1999), *Nonparametric Regression and Spline Smoothing* (Second Edition). New York: Marcel Dekker, Inc.

Fan, J., Gijbels, I., and King, M. (1997), "Local Likelihood and Locan Partial Likelihood in Hazard Regression." *The Annals of Statistics*, 25(4):1661–1690.

Fitzmaurice, G. M., Lipsitz, S. R., Ibrahim, J. G., Gelber, R., and Lipshultz, S. (2003), "Parameter Estimation for Binary Regression Models in Longitudinal Studies with Outcome-Dependent Follow-Up." Technical Report, Medical University of South Carolina, Charleston, South Carolina.

Fuller, W. (1987), *Measurement Error Models*. New York: John Wiley.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992), "Model Determination Using Predictive Distributions with Implementation via Sampling-Based Methods (with Discussion)." In *Bayesian Statistics* **4**, J. M. Bernardo, J. O. Berger, A.P. Dawid, and A. F. M. Smith (eds), 147-167. New York: Oxford University Press.

Gelfand, A. E. and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association*, 85:398–409.

Geman, S. and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Hastie, T. and Tibshirani, R. (1990), *Generalized Additive Models*. New York: Chapman and Hall.

— (1998), "Bayesian Backfitting." Technical Report, Stanford University.

— (2000), "Bayesian Backfitting (with Comments and a Rejoinder by the Authors)." *Statistical Science*, 15(3):196–223.

Holmes, C. and Mallick, B. K. (2003), "Generalized Nonlinear Modeling with Multivariate Smoothing Splines." *Journal of the American Statistical Association*, 98:352–368.

Holmes, C. C. and Held, L. (2005), "Bayesian Auxiliary Variable Models for Binary and Polychotomous Regression." *Bayesian Analysis* (forthcoming).

Horowitz, J. L. and Lee, S. (2004), "Semiparametric Estimation of a Panel Data Proportional Hazards Model with Fixed Effects." *Journal of Econometrics*, 119(1):155–198.

Hougaard, P. (2000), *Analysis of Multivariate Survival Data*. New York: Springer-Verlag.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001), "Criterion-based Methods for Bayesian Model Assessment." *Statistica Sinica*, 11:419–443.

Klein, J. P. and Moeschberger, M. L. (1997), *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Marcel Dekker, Inc.

Liang, K.-Y. and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika*, 73:13–22.

Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J. G., Gelber, R., and Lipshultz, S. (2002), "Parameter Estimation in Longitudinal Studies with Outcome-Dependent Follow-Up." *Biometrics*, 58:621–630.

Mallick, B. K., Hoffman, F. O., and Carroll, R. J. (2002), "Semiparametric Regression Modeling with Mixture of Berkson and Classical Error, with Fallout from the Nevada Site." *Biometrics*, 58:13–20.

Neal, R. M. (1997), "Monte Carlo Implementation of Gaussian Process Models for Bayesian Regression and Classification." Technical Report No. 9702, Department of Statistics, University of Toronto.

Neter, J., Wasserman, W., and Kutner, M. H. (1990), *Applied Linear Statistical Models* (Third Edition). Chicago: Irwin.

Opsomer, J., Wang, Y., and Yang, Y. (2001), "Nonparametric Regression with Correlated Errors." *Statistical Science*, 16(2):134–153.

Pierce, D. A., Stram, D. O., Vaeth, M., and Schafer, D. W. (1992), "The Errors-in-Variables Problem: Considerations Provided by Radiation Dose-Response Analyses of the A-Bomb Survivor Data." *Journal of the American Statistical Association*, 87:351–359.

Prentice, R. L. (1992), "Measurement Error and Results from Analytic Epidemiology: Dietary Fat and Breast Cancer." *Journal of the National Cancer Institute*, 88:1738–1747.

Qian, S. S., Lavine, M., and Stow, C. A. (2000), "Univariate Bayesian Nonparametric Binary Regression with Application in Environmental Management." *Environmental and Ecological Statistics*, 7:77–91.

Raghavan, N. and Cox, D. D. (1998), "Analysis of the posterior spline estiators in logistic regression." *Journal of Statistical Planning and Inference*, 71:117–136.

Rocke, D. and Durbin, B. (2001), "A Model for Measurement Error for Gene Expression Arrays." *Journal of Computational Biology*, 8(6):557–569.

Ryu, D. and Mallick, B. (2004), "Bayesian Nonparametric Regression Analysis when Covariates Are Subject-specific Parameters in a random effects Model for Longitudinal Measurements." Submitted to Biometrics.

Shively, T. S., Kohn, R., and Wood, S. (1999), "Variable Selection and Function Estimation in Additive Nonparametric Regression Using a Data-Based Prior." *Journal of the American Statistical Association*, 94:777–806.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64:583–639.

Wang, C. Y. and Pepe, M. S. (2000), "Expected estimating equations to accommodate covariate measurement error." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62:509–524.

Wang, C. Y., Wang, N., and Wang, S. (1999), "Regression Analysis When Covariates Are Regression Parameters of a Random Effects Model for Observed Longitudinal Measurements." *Biometrics*, 56:487–495.

Wang, Y. (1998), "Smoothing Spline Models with Correlated Random Errors." *Journal of the American Statistical Association*, 93(441):341–348.

Wecker, W. E. and Ansley, C. F. (1983), "The Signal Extraction Approach to Nonlinear Regression and Spline Smoothing." *Journal of the American Statistical Association*, 78:81–89.

Whitaker, R. C., Wright, J. A., Pepe, M. S., Seidel, D., K., and Dietz, W. H. (1997), "Predicting Obesity in Young Adulthood from Childhood and Parental Obesity." *The New England Journal of Medicine*, 337:869–873.

Wood, S. and Kohn, R. (1998), "A Bayesian Approach to Robust Binary Nonparametric Regression." *Journal of the American Statistical Association*, 93:203–213.

APPENDIX A

DEVIANCE INFORMATION CRITERION (DIC)

In the logistic regression, we can use the deviance information criterion (DIC) to evaluate the performance of the model. Let $Y_i \overset{ind}{\sim} B(1, p_i)$ and $\text{logit}(p_i) = f(X_i)$ for $i = 1, \ldots, n$. Suppose $\widehat{p}_i$ is a estimate of p_i then, the deviance of the model defined as follows.

$$D = -2 \sum_{i=1}^{n} \left\{ Y_i \log \widehat{p}_i + (1 - Y_i) \log(1 - \widehat{p}_i) \right\}.$$

In MCMC iterations, let $\widehat{f}_i^{(j)}$ be the value of the generated link function from the j^{th} iteration evaluated at X_i. Then the corresponding estimated $\widehat{p}_i^{(j)} = \{1 + \exp(-\widehat{f}_i^{(j)})\}^{-1}$. Similarly, the estimated average probability is $\widehat{p}_i^{mean}$ is the Monte Carlo averages of these $\widehat{p}_i^{(j)}$. Let D_j be the deviance from the j^{th} iteration and D_{mean} be the deviance from the posterior mean. Then the DIC can be defined as follow:

$$\begin{aligned} \text{DIC} &= \overline{D} + p_D, \\ \overline{D} &= \frac{1}{L} \sum_{j=1}^{L} D_j, \\ p_D &= \overline{D} - D_{mean}, \end{aligned}$$

where L is the number of MCMC samples used.

APPENDIX B

COMPUTATIONAL NOTES

1. Duplicated values of covariate

   We assumed distinctive values of covariates for different subjects. However, in the real world with childhood growth data, we may have the same value of covariates because of the truncation errors in the recording. These duplicated values of the covariates made almost no problem in the generations of parameters and latent variables, except for the NCS during Gibbs procedure. Hence, for the NCS, we rounded off the average value of the latent variable when the corresponding values of the covariate are the same.

2. Calculation of covariance matrix of the full conditional distribution of Bayesian NCS

   Under Model I in section 3.3, to generate the Bayesian NCS, $\boldsymbol{f}_1$, from its full conditional distribution, $N\left[\boldsymbol{A}_1(\alpha_1)\boldsymbol{Z}_1, \boldsymbol{A}_1(\alpha_1)\sigma^2_{z_1}\right]$, we could use Cholesky's decomposition of $\boldsymbol{A}_1(\alpha_1)$, because $\boldsymbol{A}_1(\alpha_1)$ is symmetric positive definite. Assuming an upper triangular matrix $\boldsymbol{L}$ and Cholesky's decomposition of $\boldsymbol{L}^\top\boldsymbol{L} = \boldsymbol{A}_1(\alpha_1)$, we could generate $\boldsymbol{f}_1$ by generating $\boldsymbol{V}$ from $N(0, \sigma^2_{z_1}\boldsymbol{I})$ and taking $\boldsymbol{A}_1(\alpha_1)\boldsymbol{Z}_1 + \boldsymbol{L}^\top\boldsymbol{V}$. However, this method requires the matrix inversion of $\boldsymbol{I} + \alpha_1\boldsymbol{K}_1$, which takes more computing time and leads to truncation errors.

   Without the matrix inversion of $\boldsymbol{I}+\alpha_1\boldsymbol{K}_1$, we utilized Cholesky's decomposition of $\boldsymbol{L}^{*\top}\boldsymbol{L}^* = \boldsymbol{I} + \alpha_1\boldsymbol{K}_1$, where $\boldsymbol{L}^*$ is an upper triangular matrix. We first solved $\boldsymbol{L}^{*\top}\boldsymbol{v} = \boldsymbol{Z}_1$ for $\boldsymbol{v}$ by forward substitution, and then solved $\boldsymbol{L}^*\boldsymbol{u} = \boldsymbol{v}$ for $\boldsymbol{u}$ by backward substitution instead of the direct calculation of $\boldsymbol{A}_1(\alpha_1)\boldsymbol{Z}_1$. With $\boldsymbol{V}$ from $N(0, \sigma_{z_1}^2 \boldsymbol{I})$, we solved $\boldsymbol{L}^* \boldsymbol{w} = \boldsymbol{V}$ for $\boldsymbol{w}$ by backward substitution rather than $\boldsymbol{L}^\top \boldsymbol{V}$.

When $\boldsymbol{I} + \alpha_1 \boldsymbol{K}_1$ is ill conditioned, that is, badly scaled and close to non-negative definite, Cholesky's decomposition is not possible. Because our estimated or generated covariate X_{i1} in Model I is unequally spaced, it can produce badly scaled $\boldsymbol{I} + \alpha_1 \boldsymbol{K}_1$. Sometimes a small value for the generated smoothing parameter α can lead to a badly scaled matrix. When the matrix inversion of $\boldsymbol{I} + \alpha_1 \boldsymbol{K}_1$ is possible with high condition number, we used eigenvalue decomposition of $\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top = \boldsymbol{A}_1(\alpha_1)$, where $\boldsymbol{U}$ is an orthogonal matrix and $\boldsymbol{D}$ is a diagonal matrix, and then constructed diagonal matrix $\boldsymbol{D}^*$ with non-negative elements of $\boldsymbol{D}$ and zeros for the negative elements of $\boldsymbol{D}$. Then, with the generated normal random vector $\boldsymbol{V}$, we produced $\boldsymbol{f}_1$ such that $\boldsymbol{f}_1 = \boldsymbol{A}_1(\alpha_1)\boldsymbol{Z}_1 + \boldsymbol{U}\sqrt{\boldsymbol{D}^*}\boldsymbol{V}$, where $\sqrt{\boldsymbol{D}^*}$ consists of the square roots of elements of $\boldsymbol{D}^*$.

APPENDIX C

GENERATION LATENT VARIABLE FROM JOINT POSTERIOR

In Gibbs sampling scheme of Section 4.3, rather than drawing $\boldsymbol{Z}_l$ and $\boldsymbol{f}_l$ from the full conditional distributions, we decided to draw from the joint distribution as $[\boldsymbol{f}_l, \boldsymbol{Z}_l|\boldsymbol{Y}, \alpha_l, \sigma^2_{z_l}] = [\boldsymbol{Z}|\boldsymbol{Y}, \alpha_l, \sigma^2_{z_l}][\boldsymbol{f}_l|\boldsymbol{Z}_l, \alpha_l, \sigma^2_{z_l}]$. In the first expression, $\boldsymbol{Z}_l$ has been drawn from the conditional distribution, marginalized over $\boldsymbol{f}_l$, which will reduce the autocorrelation and improve mixing in the Markov Chain. In the second expression, the distribution of $\boldsymbol{f}_l$ is conditionally independent of $\boldsymbol{Y}$. We can obtain these distributions as

$$\begin{aligned}
[\boldsymbol{f}_l, \boldsymbol{Z}_l|\boldsymbol{Y}, \alpha_l, \sigma^2_{z_l}] &\propto \left\{\prod_{i=1}^{n} p(Y_i|\eta_i)\right\} \\
&\times \exp\left\{-\frac{1}{2\sigma^2_{z_l}}(\boldsymbol{Z}_l - \boldsymbol{f}_l)^T(\boldsymbol{Z}_l - \boldsymbol{f}_l) - \frac{\alpha_l}{2\sigma^2_{z_l}}\boldsymbol{f}_l^T\boldsymbol{K}_l\boldsymbol{f}_l\right\} \\
&\propto \left\{\prod_{i=1}^{n} p(Y_i|\eta_i)\right\} \exp\left[-\frac{1}{2\sigma^2_{z_l}}\boldsymbol{Z}^T\left\{\boldsymbol{I} - \boldsymbol{A}_l(\alpha_l)\right\}\boldsymbol{Z}_l\right] \\
&\times \exp\left[-\frac{1}{2\sigma^2_{z_l}}\left\{\boldsymbol{f}_l - \boldsymbol{A}_l(\alpha_l)\boldsymbol{Z}_l\right\}^T \boldsymbol{A}_l(\alpha_l)^{-1}\left\{\boldsymbol{f}_l - \boldsymbol{A}_l(\alpha_l)\boldsymbol{Z}_l\right\}\right] \\
&= [\boldsymbol{Z}_l|\boldsymbol{Y}, \alpha_l, \sigma^2_{z_l}][\boldsymbol{f}_l|\boldsymbol{Z}_l, \alpha_l, \sigma^2_{z_l}],
\end{aligned}$$

where $p(\cdot)$ is a density of the general exponential family with the response Y_i, $\eta_i = \sum_{l=1}^{q} Z_{il}$, $i = 1, \ldots, n$, and $\boldsymbol{A}_l(\alpha_l) = (\boldsymbol{I} + \alpha_l\boldsymbol{K}_l)^{-1}$. Let $\boldsymbol{B}_l(\alpha_l) = \boldsymbol{I} - \boldsymbol{A}_l(\alpha_l)$. By the matrix notation $\boldsymbol{Q}_l$ and $\boldsymbol{R}_l$ from Eubank (1999) such that $\boldsymbol{K}_l = \boldsymbol{Q}_l\boldsymbol{R}_l^{-1}\boldsymbol{Q}_l^T$, $\boldsymbol{B}_l(\alpha_l) = \boldsymbol{Q}_l(\frac{1}{\alpha_l}\boldsymbol{R}_l + \boldsymbol{Q}_l^T\boldsymbol{Q}_l)^{-1}\boldsymbol{Q}_l^T$, which is a $n \times n$ matrix with rank $n-2$. Then, in the above factorization, the full conditional of $\boldsymbol{f}_l$ is same as before, while the full conditional of $\boldsymbol{Z}_l$ is a marginal integrated over $\boldsymbol{f}_l$ such that

$$[\boldsymbol{Z}_l|\boldsymbol{Y}, \cdot] \;\propto\; \left\{\prod_{i=1}^{n} p(Y_i|\eta_i)\right\} N\left[\boldsymbol{0}, \boldsymbol{B}_l(\alpha_l)^{-}\right],$$

where $\boldsymbol{B}_l(\alpha_l)^-$ indicates the generalized inverse of $\boldsymbol{B}_l(\alpha_l)$. To generate $\boldsymbol{Z}_l$ from its full conditional distribution, we can use the Gibbs samples for each element of $\boldsymbol{Z}_l$. Let Z_{il} be the i^{th} element of $\boldsymbol{Z}_l$, $\boldsymbol{Z}_{-il}$ be $\boldsymbol{Z}_l$ without i^{th} element, and b_{ij} be the element of i^{th} row and j^{th} column of $\boldsymbol{B}_l(\alpha_l)$, then full conditional of Z_{il} can be expressed as the following:

$$
\begin{aligned}
[Z_{il}|\boldsymbol{Z}_{-il}, \boldsymbol{Y}, \cdot] &\propto p(Y_i|\eta_i) \exp\left[-\frac{1}{2\sigma_{z_l}^2}\left\{b_{ii}Z_{il}^2 + 2Z_{il}\left(\sum_{j=1}^{n} b_{ij}Z_{jl} - b_{ii}Z_{il}\right)\right\}\right] \\
&\propto p(Y_i|\eta_i) N\left[\frac{b_{ii}Z_{il} - \sum_{j=1}^{n} b_{ij}Z_{jl}}{b_{ii}}, \frac{\sigma_{z_l}^2}{b_{ii}}\right].
\end{aligned}
$$

APPENDIX D

COMPUTATION OF CPO

For the data from the ith subject, let $Y_{ij} = Y_i(t_{ij})$, $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im_i})^T$, $\boldsymbol{U}_i = (t_{i1}, U_{i2}, \ldots, U_{im_i})^T$, $\boldsymbol{D}_i = (\boldsymbol{Y}_i^T, \boldsymbol{U}_i^T)^T$, $\boldsymbol{D} = (\boldsymbol{D}_1^T, \ldots, \boldsymbol{D}_n^T)^T$, and let subscript (i) indicate vector without ith case. As a usual case, suppose $\boldsymbol{U}_i$ is a covariate which does not depend on response. Then, the CPO statistic is defined as

$$\begin{aligned}
\text{CPO}_i &= f(\boldsymbol{Y}_i|\boldsymbol{D}_{(i)}) = \int_{\boldsymbol{\xi}} f(\boldsymbol{Y}_i|\boldsymbol{\xi})\pi(\boldsymbol{\xi}|\boldsymbol{D}_{(i)})d\boldsymbol{\xi} \\
&= \int_{\boldsymbol{\xi}} f(\boldsymbol{Y}_i|\boldsymbol{\xi}) \frac{f(\boldsymbol{D}_{(i)}|\boldsymbol{\xi})\pi(\boldsymbol{\xi})}{\int_{\boldsymbol{\xi}} f(\boldsymbol{D}_{(i)}|\boldsymbol{\xi})\pi(\boldsymbol{\xi})d\boldsymbol{\xi}} d\boldsymbol{\xi} \\
&= \frac{f(\boldsymbol{D})}{\int_{\boldsymbol{\xi}} \frac{1}{f(\boldsymbol{Y}_i|\boldsymbol{\xi})} f(\boldsymbol{D}|\boldsymbol{\xi})\pi(\boldsymbol{\xi})d\boldsymbol{\xi}} \\
&= \frac{1}{\int_{\boldsymbol{\xi}} \frac{1}{f(\boldsymbol{Y}_i|\boldsymbol{\xi})} \pi(\boldsymbol{\xi}|\boldsymbol{D})d\boldsymbol{\xi}} \\
&= \left\{ E_{\boldsymbol{\xi}|\boldsymbol{D}} \left(\frac{1}{f(\boldsymbol{Y}_i|\boldsymbol{\xi})} \right) \right\}^{-1}.
\end{aligned}$$

Under Model-0, let $\boldsymbol{\xi}$ be all parameters in the model except for γ which is related only on follow-up times $\boldsymbol{U}$. Then, for given $\boldsymbol{U}_i$, CPO_{i0} is defined by

$$\text{CPO}_{i0} = P(\boldsymbol{Y}_i|\boldsymbol{D}_{(i)}) = \int f(\boldsymbol{Y}_i|\boldsymbol{U}_i, \gamma, \boldsymbol{\xi})\pi(\gamma, \boldsymbol{\xi}|\boldsymbol{D}_{(i)})d\gamma d\boldsymbol{\xi}.$$

Note that $\boldsymbol{Y}_i$ does not depend on γ, $\boldsymbol{U}_i$ does not depend on $\boldsymbol{\xi}$, and $\pi(\gamma, \boldsymbol{\xi}) = \pi(\gamma)\pi(\boldsymbol{\xi})$, that is γ and $\boldsymbol{\xi}$ are separable. Hence, for given $\boldsymbol{U}_i$, CPO$_{i0}$ can be re-written by

$$\begin{aligned}
\text{CPO}_{i0} &= \int f(\boldsymbol{Y}_i|\boldsymbol{U}_i,\boldsymbol{\xi})\pi(\boldsymbol{\xi}|\boldsymbol{D}_{(i)})d\boldsymbol{\xi} \\
&= \int \frac{f(\boldsymbol{Y}_i|\boldsymbol{U}_i,\boldsymbol{\xi})f(\boldsymbol{Y}_{(i)}|\boldsymbol{U}_{(i)},\boldsymbol{\xi})f(\boldsymbol{U}_{(i)},\boldsymbol{\xi})}{\int f(\boldsymbol{D}_{(i)}|\gamma,\boldsymbol{\xi})\pi(\gamma,\boldsymbol{\xi})d\gamma d\boldsymbol{\xi}}d\boldsymbol{\xi} \\
&= \frac{\int \frac{1}{f(\boldsymbol{U}_i)}f(\boldsymbol{Y},\boldsymbol{U},\boldsymbol{\xi})d\boldsymbol{\xi}}{\int f(\boldsymbol{Y}_{(i)}|\boldsymbol{U}_{(i)},\boldsymbol{\xi})f(\boldsymbol{U}_{(i)}|\gamma)\pi(\gamma)\pi(\boldsymbol{\xi})d\boldsymbol{\xi}d\gamma} \\
&= \frac{\frac{1}{f(\boldsymbol{U}_i)}f(\boldsymbol{Y},\boldsymbol{U})}{f(\boldsymbol{U}_{(i)})\int f(\boldsymbol{Y}_{(i)}|\boldsymbol{U}_{(i)},\boldsymbol{\xi})\pi(\boldsymbol{\xi})d\boldsymbol{\xi}} \\
&= \left\{f(\boldsymbol{U})\int \frac{1}{f(\boldsymbol{Y}_i|\boldsymbol{U}_i,\boldsymbol{\xi})}\frac{f(\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{\xi})\pi(\boldsymbol{\xi})}{f(\boldsymbol{Y},\boldsymbol{U})}d\boldsymbol{\xi}\right\}^{-1} \\
&= \left\{\int \frac{1}{f(\boldsymbol{Y}_i|\boldsymbol{U}_i,\boldsymbol{\xi})}\pi(\boldsymbol{\xi}|\boldsymbol{Y},\boldsymbol{U})d\boldsymbol{\xi}\right\}^{-1} \\
&= \left[E_{\boldsymbol{\xi}|\boldsymbol{D}}\left\{\frac{1}{f(\boldsymbol{Y}_i|\boldsymbol{U}_i,\boldsymbol{\xi})}\right\}\right]^{-1}
\end{aligned}$$

Under Model-1, let $\boldsymbol{W} = (W_1,\ldots,W_n)^T$, $\boldsymbol{\xi}$ be a vector of all parameters, and other notations be the same as before. Then, for given $\boldsymbol{U}_i$, CCPO$_{i1}$ is defined by

$$\begin{aligned}
\text{CCPO}_{i1} &= \int f(\boldsymbol{Y}_i|\boldsymbol{U}_i,\boldsymbol{W},\boldsymbol{\xi})f(\boldsymbol{W},\boldsymbol{\xi}|\boldsymbol{D}_{(i)})d\boldsymbol{W}d\boldsymbol{\xi} \\
&= \int \frac{f(\boldsymbol{Y}_i|\boldsymbol{U}_i,\boldsymbol{W},\boldsymbol{\xi})f(\boldsymbol{Y}_{(i)}|\boldsymbol{U}_{(i)},\boldsymbol{W},\boldsymbol{\xi})f(\boldsymbol{U}_{(i)},\boldsymbol{W},\boldsymbol{\xi})}{\int f(\boldsymbol{D}_{(i)}|\boldsymbol{W},\boldsymbol{\xi})f(\boldsymbol{W},\boldsymbol{\xi})d\boldsymbol{W}d\boldsymbol{\xi}}d\boldsymbol{W}d\boldsymbol{\xi} \\
&= \frac{\int \frac{1}{f(\boldsymbol{U}_i|W_i,\boldsymbol{\xi})}f(\boldsymbol{Y},\boldsymbol{U},W_i,\boldsymbol{\xi})dW_i d\boldsymbol{\xi}}{\int \frac{1}{f(\boldsymbol{Y}_i,\boldsymbol{U}_i|W_i,\boldsymbol{\xi})}f(\boldsymbol{Y},\boldsymbol{U},W_i,\boldsymbol{\xi})dW_i d\boldsymbol{\xi}} \\
&= \frac{\int \frac{1}{f(\boldsymbol{U}_i|W_i,\boldsymbol{\xi})}f(W_i,\boldsymbol{\xi}|\boldsymbol{D})f(\boldsymbol{D})dW_i d\boldsymbol{\xi}}{\int \frac{1}{f(\boldsymbol{Y}_i,\boldsymbol{U}_i|W_i,\boldsymbol{\xi})}f(W_i,\boldsymbol{\xi}|\boldsymbol{D})f(\boldsymbol{D})dW_i d\boldsymbol{\xi}} \\
&= \frac{E_{W_i,\boldsymbol{\xi}|\boldsymbol{D}}\left(\frac{1}{f(\boldsymbol{U}_i|W_i,\boldsymbol{\xi})}\right)}{E_{W_i,\boldsymbol{\xi}|\boldsymbol{D}}\left(\frac{1}{f(\boldsymbol{Y}_i,\boldsymbol{U}_i|W_i,\boldsymbol{\xi})}\right)}
\end{aligned}$$

VITA

Duchwan Ryu was born in Kwanhwa, Korea on June 26, 1966 to Hoan Ryu and Kyungja Yoon. He received his B.S. degree in statistics from the Department of Statistics of Korea University, Seoul, Korea in February, 1989. He received his M.S. degree in statistics from Korea University, 1991 under the direction of Dr. Myoung-Shik Jhun. He continued his studies at Texas A&M University in statistics under the direction of Dr. Bani K. Mallick and Dr. Raymond J. Carroll, and received his Ph.D. degree in May 2005. His permanent address is 482-18 DongA Villa B-302, Yokbuk-dong, Yongin-si, Kyounggi-do, South Korea 449-932.

The typist for this dissertation was Duchwan Ryu.