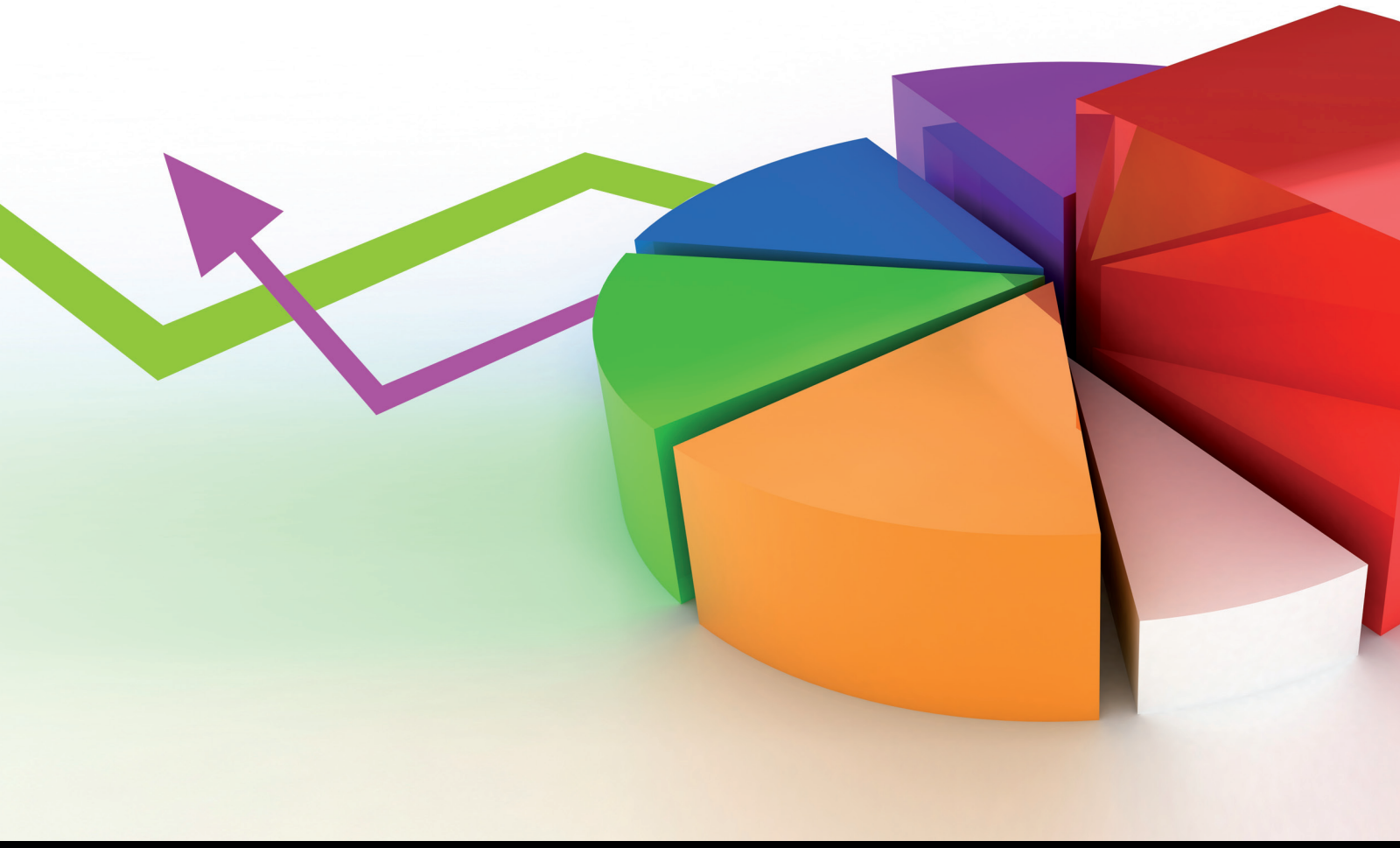
Journal of Probability and Statistics

Joint Models and Their Applications

Guest Editors: Yangxin Huang, Lang Wu, Grace Y. Yi, and Wenbin Lu



Joint Models and Their Applications

Journal of Probability and Statistics

Joint Models and Their Applications

Guest Editors: Yangxin Huang, Lang Wu, Grace Y. Yi, and Wenbin Lu

Editorial Board

Contents

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2012, Article ID 463506, 2 pages
doi:10.1155/2012/463506

Editorial

Joint Models and Their Applications

Yangxin Huang,[1] Lang Wu,[2] Grace Y. Yi,[3] and Wenbin Lu[4]

[1] *Department of Epidemiology & Biostatistics, College of Public Health, University of South Florida, Tampa, FL 33612, USA*
[2] *Department of Statistics, University of British Columbia, Vancouver, BC, Canada V6T 1Z2*
[3] *Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada N2L 3G1*
[4] *Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA*

Correspondence should be addressed to Yangxin Huang, yhuang@health.usf.edu

Received 27 December 2011; Accepted 27 December 2011

A common objective in longitudinal studies is to characterize the relationship between a longitudinal response process and a time to event. Considerable interest has been focused on the so-called joint models, where models for the event time distribution and longitudinal data are often specified through a common set of latent random effects. Joint models of longitudinal data and/or survival data have received great attention in the literature over the past two decades and are becoming increasingly active area of statistics research. The importance of these models is well recognized, partly due to the fact that longitudinal data and survival data arise frequently in practice. Despite the extensive literature on this topic, these models continue to be a main research stream since they offer many advantages over separate analysis of longitudinal data and/or survival data. To accelerate the development of advanced tools and knowledge of joint models, a number of important issues remain to be addressed such as computational issues, model diagnostics and selections, joint models with skew distributions, and various choices of longitudinal models and survival models. To stimulate the continuing efforts to understand various joint model development and associated statistical inference methods with their applications in biomedical, biological, engineering, and other studies, in this special issue, we have invited a few papers that address some of those issues.

The paper by L. Wu et al., in this special issue, provides a brief overview of various formulations of joint models for longitudinal and survival data. Commonly used methods including the likelihood method and two-stage methods are discussed in detail, and other joint modeling methods such as Bayesian approach are also briefly presented. Computational issues are investigated. A real data analysis and a simulation study are provided to compare the performance of various methods. The paper by M. Liu and W. Lu delivers a semiparametric marginal inference approach for longitudinal outcomes in the presence

of informative dropouts. The dependence between longitudinal outcome and informative dropout time is characterized by a conditional mean model, and the longitudinal regression coefficients are estimated by a class of conditional generalized estimating equations. The proposed method is robust against a class of latent-variable models for longitudinal data with informative dropouts and is computationally easy to implement. G. A. Dagne and Y. Huang's paper presents joint Tobit models for a left-censored response variable with skewness and covariate variables with substantial measurement errors. The proposed joint models are a skew-t nonlinear mixed-effects Tobit model for the response process and a skew-t nonparametric mixed-effects model for covariate process under a Bayesian framework. A real data example is used to illustrate the proposed methods. The paper by M. Murawska et al. proposes a two-stage approach that summarizes the longitudinal information with nonlinear mixed-effects model at the first stage, and includes empirical Bayes estimates of the subject-specific parameters as predictors in the Cox model for time-to-event at the second stage. To take into account the uncertainty of the estimated subject-specific parameters, the authors use a Monte Carlo approach and sample from the posterior distribution of the random effects given the observed data. The paper by S. Gurmu and G. A. Dagne, in this special issue, develops zero-inflated bivariate-ordered probit model and carries out estimation using Markov Chain Monte Carlo techniques in the context of Bayesian analysis of a joint model of ordered outcomes. The authors analyze the socioeconomic determinants of individual problem of smoking and chewing tobacco using household tobacco survey data with substantial proportion of zeros. The example shows that the use of a model that ignores zero-inflation masks differential effects of covariates on nonusers and users.

Yangxin Huang
Lang Wu
Grace Y. Yi
Wenbin Lu

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2012, Article ID 194194, 18 pages
doi:10.1155/2012/194194

Research Article

A Two-Stage Joint Model for Nonlinear Longitudinal Response and a Time-to-Event with Application in Transplantation Studies

Magdalena Murawska,[1] Dimitris Rizopoulos,[1] and Emmanuel Lesaffre[1,2]

[1] *Department of Biostatistics, Erasmus University Medical Center, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands*
[2] *I-Biostat, Catholic University of Leuven, B-3000 Leuven, Belgium*

Correspondence should be addressed to Magdalena Murawska, m.murawska@erasmusmc.nl

Received 7 July 2011; Revised 27 October 2011; Accepted 6 November 2011

Academic Editor: Grace Y. Yi

In transplantation studies, often longitudinal measurements are collected for important markers prior to the actual transplantation. Using only the last available measurement as a baseline covariate in a survival model for the time to graft failure discards the whole longitudinal evolution. We propose a two-stage approach to handle this type of data sets using all available information. At the first stage, we summarize the longitudinal information with nonlinear mixed-effects model, and at the second stage, we include the Empirical Bayes estimates of the subject-specific parameters as predictors in the Cox model for the time to allograft failure. To take into account that the estimated subject-specific parameters are included in the model, we use a Monte Carlo approach and sample from the posterior distribution of the random effects given the observed data. Our proposal is exemplified on a study of the impact of renal resistance evolution on the graft survival.

1. Introduction

Many medical studies involve analyzing responses together with event history data collected for each patient. A well-known and broadly studied example can be found in AIDS research, where CD4 cell counts taken at different time points are related to the time to death. These data need to be analyzed using a joint modeling approach in order to properly take into account the association between the longitudinal data and the event times. The requirement for a joint modeling approach is twofold. Namely, when focus is on the longitudinal outcome, events cause nonrandom dropout that needs to be accounted for in order to obtain valid inferences. When focus is on the event times, the longitudinal responses cannot be simply

included in a relative risk model because they represent the output of an internal time-dependent covariate [1].

In this paper, we focus on a setting that shares some similarities with the standard joint modeling framework described above, but also has important differences. In particular, we are interested in the association between longitudinal responses taken before the actual followup for the time-to-event has been initiated. This setting is frequently encountered in transplantation studies, where patients in the waiting list provide a series of longitudinal outcomes that may be related to events occurring after transplantation. A standard analysis in transplantation studies is to ignore the longitudinal information and use only the last available measurement as a baseline covariate in a model for the allograft survival. It is, however, evident that such an approach discards valuable information. An alternative straightforward approach is to put all longitudinal measurements as covariates in the survival model. Nevertheless, there are several disadvantages with this approach. First, it would require spending many additional degrees of freedom, one for each of the longitudinal measurements. Second, patients with at least one missing longitudinal response need to be discarded, resulting in a great loss of efficiency. Finally, we may encounter multicollinearity problems due to the possibly high correlation between the longitudinal measurements at different time points.

Nowadays, when it comes to measuring the association between a longitudinal marker and an event-time outcome, a standard approach is to use the joint model postulated by Faucett and Thomas [2] and Wulfsohn and Tsiatis [3]. In this setting, the longitudinal responses are considered realizations of an endogenous time-dependent covariate (Kabfleish and Prentice [1]), which is measured with error and for which we do not have the complete history of past values available. To account for these features we estimate in the joint modeling framework the joint distribution of the survival and longitudinal processes. Unlike in the multivariate approach, where we have to interpret the estimates for each longitudinal measurement separately, the joint modeling approach allows to get more insight in the nature of the relation between the two analyzed processes since it assumes some underlying process for the longitudinal measures.

However, in contrast with the standard joint modeling setting, in our case (i.e., transplantation studies) the longitudinal responses do not constitute an endogenous time-dependent variable measured at the same period as the time to event. In particular, since the longitudinal measurements are collected prior to transplantation, occurrence of an event (i.e., graft failure after transplantation) does not cause nonrandom dropout in the longitudinal outcome. Nevertheless, the problem of measurement error still remains. Ignoring the measurement error affects not only the standard errors of the estimates of interest, but also it can cause attenuation of the coefficients towards zero [4]. To overcome this, we propose a two-stage modeling approach that appropriately summarizes the longitudinal information before the start of followup by means of a mixed effects model and then uses this information to model the time to event by including the Empirical Bayes estimates of the subject-specific parameters as predictors in the Cox model. To account for the fact that we include the estimates and not the true values of the parameters, we use a Monte Carlo approach and sample from the posterior distribution of the random effects. The proposed method does not require joint maximization neither fitting the random effects model for each event time as in the two-stage approach of Tsiatis et al. [5]. We compare this approach with the "naive" one when the uncertainty about the estimates from the first step is not taken into account, as well as with the full Bayesian approach. Our approach shares similarities with the two-stage approach of Albert and Shih [6]. They considered a model, in which a discrete event

time distribution is modeled as a linear function of the random slope of the longitudinal process estimated from the linear-mixed model. The bias from informative dropout was reduced by using the conditional distribution of the longitudinal process given the dropout time to construct the complete data set. To account for the measurement error in the mean of the posterior distribution of the random effects, the variance, that incorporates the error in estimating the fixed effects in the longitudinal model, was used. However, we use sampling not to impute missing values and correct for nonrandom dropout but in order to account for the variability in the predicted longitudinal covariates that are then used in survival model. A method of adjusting for measurement error in covariates, which was used by Albert and Shih, does not apply in our case since it requires the discrete time-to-event and linear model for longitudinal data. The time-to-event could be discretized but still we have a nonlinear model for longitudinal data.

Our research is motivated by data from an international prospective trial on kidney-transplant patients. The study has two arms, where in the first arm donors' kidneys were administered to cold storage and in the second arm they were administered to machine perfusion (MP). The advantage of machine perfusion is the possibility of measuring different kidney's parameters reflecting the state of the organ. One of the parameters of interest is renal resistance level (RR), which has been measured at 10 minutes, 30 minutes, 1 hour, 2 hours, 4 hours, and just before transplantation. Our aim here is to study the association of the renal resistance evolution profile with the risk of graft failure. The time of last measurement was different for different patients and often unknown exactly. However, based on the medical consult and visual inspection of the individual profiles, the last measurement was chosen to be taken at 6 hours for each patient.

The rest of the paper is organized as follows. Section 2 provides the general modeling framework with the definition of the two submodels for the longitudinal and survival data, respectively. Section 3 describes the estimation methods for the full likelihood and the proposed two-stage approach. In Section 4, we apply the two-stage approach to the renal data. Section 5 contains the setup and the results for the simulation study. Finally, in Section 6 we discuss the proposed methodology.

2. Joint Modeling Framework

Let $Y_i(u)$ denote the longitudinal profiles for individual i, $i = 1, 2, \ldots, N$. We assume that $Y_i(u)$ are collected for the ith individual prior to the specified time t_i, $u \in (0, t_i)$. Let $t = 0$ denote the time of the first longitudinal measurement and t_i the time of the last collected measurement. t_i can be different for different individuals, and we denote by m_i the number of longitudinal measurements for subject i collected until time t_i and by u_{ij} the time of jth measurement. Denote by $T_i^* \geq t_i$ the true survival time for individual i. Since the survival time is right censored, we observe only $T_i = \min(T_i^*, C_i)$, where $C_i \geq t_i$ is the censoring time with the failure indicator Δ_i, which equals to 1 if the failure is observed and 0 otherwise, that is, $\Delta_i = I(T_i \leq C_i)$ with $I(\cdot)$ denoting the indicator function. We will assume that censoring is independent of all other survival and covariate information. In addition, we assume that the observed longitudinal responses $Y_i(u)$ are measured with error (i.e., biological variation) around the true longitudinal profile $W_i(u)$, that is,

$$\begin{aligned} Y_i(u) &= W_i(u) + \varepsilon_i(u), \quad \text{with } \varepsilon_i(u) \sim N\left(0, \sigma^2\right), \\ &\operatorname{cov}\left(\varepsilon_i(u), \varepsilon_i\left(u'\right)\right) = 0, \quad u' \neq u. \end{aligned} \tag{2.1}$$

We will consider the longitudinal response that exhibits nonlinear profiles in time. Therefore, we approximate $W_i(u)$ by means of a nonlinear mixed effects model:

$$W_i(u) = f\left(u; \phi_i\right), \quad \text{with } \phi_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\alpha}_i, \tag{2.2}$$

where $f(\cdot)$ is a nonlinear function, parameterized by the vector ϕ_i. The parameters ϕ_i control the shape of the nonlinear function, and subscript i denotes that each subject may have its own nonlinear evolution in time in the family $f(\cdot;\phi)$. For the subject-specific parameter ϕ_i, we assume a standard mixed model structure with $\mathbf{X}_i$ denoting the fixed effects design matrix with corresponding regression coefficients $\boldsymbol{\beta}$, $\mathbf{Z}_i$ the random effects design matrix, and $\boldsymbol{\alpha}_i$ the random effects. The random effects $\boldsymbol{\alpha}_i$ are assumed to be independent and normally distributed with mean zero and variance-covariance matrix $\mathbf{D}$.

For the event process, we postulate the standard relative risk model of the form:

$$\lambda_i(t) = \lambda_0(t)\exp\left(\boldsymbol{\gamma}^T\phi_i\right), \tag{2.3}$$

where $\lambda_i(t)$ is the hazard function and $\lambda_0(t)$ is the baseline hazard, which can be modeled parametrically or left completely unspecified. The subject-specific parameters ϕ_i summarize the longitudinal evolutions of the response for each subject, and therefore coefficients $\boldsymbol{\gamma}$ measure the strength of the association between the different characteristics of the underlying subject-specific nonlinear evolution of the longitudinal profiles and the risk for an event. Within the formulation of the two submodels (2.2) and (2.3), the same random effects now account for both the association between the longitudinal and event outcomes, and the correlation between the repeated measurements in the longitudinal process.

In the particular transplantation setting that will be analyzed in the following study, $Y_i(u)$ are the renal resistance level measurements collected for the ith donor prior to the transplantation time t_i and the same index i is used to denote the allograft transplanted to the ith patient. Time $t = 0$ represents the time that the kidney is removed from the donor and put in cold storage or in a perfusion machine.

3. Estimation

3.1. Full Likelihood Framework: Bayesian Approach

In the standard joint modeling framework, the estimation is typically based on maximum likelihood or Bayesian methods (MCMC). This proceeds under the following set of conditional independence assumptions:

$$\begin{aligned} p(T_i, \Delta_i, \mathbf{Y}_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}) &= p(T_i, \Delta_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_t)p\left(\mathbf{Y}_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y\right), \\ p\left(\mathbf{Y}_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y\right) &= \prod_{j=1}^{m_i} p\left(Y_i(u_{ij}) \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y\right). \end{aligned} \tag{3.1}$$

In particular, we assume that given the random effects the longitudinal process is independent from the event times, and moreover, the longitudinal measurements are independent from each other.

Maximum likelihood methods use the joint likelihood and maximize the log-likelihood function $l_i(\boldsymbol{\theta}) = \sum_i \log p(T_i, \Delta_i, \mathbf{Y_i}; \boldsymbol{\theta})$. This requires numerical integration and optimization, which makes the fit difficult, especially in high-dimensional random effects settings. Standard options for numerical integration are Gaussian quadrature, Laplace approximation, or Monte Carlo sampling [7, 8]. Maximization of the approximated log-likelihood is based on an EM algorithm [3, 5, 9–11]. Several authors proposed a Bayesian approach (MCMC) [2, 12, 13]. Bayesian estimation, that generalizes a joint model for the case with multivariate longitudinal data, has been discussed by Ibrahim et al. [14]. Brown and Ibrahim [15] considered semiparametric model relaxing the distributional assumption for the random effects. In most papers, the longitudinal submodel is a linear mixed-effects model. Joint models with nonlinear mixed-effects submodels have been less studied in the literature [16]. Nonlinear mixed models are more common in pharmacokinetics and pharmacodynamics, where they are jointly modeled with nonrandom dropout using NONMEM software. Several authors considered a Bayesian approach with a nonlinear mixed model and informative missingness [17, 18].

Here we will proceed under the Bayesian paradigm to estimate the model parameter. Under the conditional independence assumption (3.1), the posterior distribution of the parameters and the latent terms, conditional on the observed data, are derived as

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}_i \mid T_i; \Delta_i; \mathbf{Y}_i) \propto \prod_{i=1}^{N} \prod_{j=1}^{m_i} \{p(Y_i(u_{ij}) \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y)\} p(T_i, \Delta_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_t) p(\boldsymbol{\alpha}_i; \boldsymbol{\theta}_\alpha) p(\boldsymbol{\theta}_y, \boldsymbol{\theta}_t, \boldsymbol{\theta}_\alpha), \quad (3.2)$$

where $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_y^T, \boldsymbol{\theta}_t^T, \boldsymbol{\theta}_\alpha^T)$ is a vector of parameters from the longitudinal and survival models and the vector of the random effects, respectively, and $p(\cdot)$ denotes the appropriate probability density function. The likelihood contribution for the ith subject conditionally on the random terms is given by

$$\begin{aligned} p(\mathbf{Y_i}, T_i, \Delta_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}) &= p(\mathbf{Y_i} \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y) p(T_i, \Delta_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_t) \\ &= \left[\lambda_0(T_i) \exp\{\boldsymbol{\gamma}^T \phi_i(\boldsymbol{\alpha}_i)\}\right]^{\Delta_i} \exp\left[-\int_0^{T_i} \lambda_0(t) \exp\left\{\boldsymbol{\gamma}^T \phi_i(\boldsymbol{\alpha}_i)\right\} dt\right] \\ &\quad \times \frac{1}{(2\pi\sigma^2)^{m_i/2}} \exp\left[-\sum_{j=1}^{m_i} \frac{\{W_i(u_{ij}, \boldsymbol{\alpha}_i) - Y_i(u_{ij})\}^2}{2\sigma^2}\right]. \end{aligned} \quad (3.3)$$

The baseline hazard can be assumed of a specific parametric form, for example, the Weibull hazard. For the priors of the model parameters, we make standard assumptions following Ibrahim et al. [14]. In particular, for the regression coefficients $\boldsymbol{\beta}$ of the longitudinal submodel and for the coefficients $\boldsymbol{\gamma}$ of survival submodel, we used multivariate normal priors. For variance-covariance matrices, we assumed an inverse Wishart distribution and for the variance-covariance parameters we took as a prior an inverse gamma. For all parameters, the vague priors have been chosen.

The implementation of the Cox and piecewise constant hazard models is typically based on the counting process notation introduced by Andersen and Gill [19] and formulated

by Clayton [20]. In particular, we treat the counting process increments $dN_i(t)$ in the time interval $[t, t + \Delta t]$ as independent Poisson random variables with means $\Lambda_i(t)dt$:

$$\Lambda_i(t)dt = \omega_i(t)\exp\left(\gamma^T \phi_i\right) d\Lambda_0(t), \tag{3.4}$$

where $\omega_i(t)$ is an observed process taking the value 1 if subject i is observed at time t and 0 otherwise and $d\Lambda_0(t)$ is the increment in the integrated baseline hazard function occurring during the time interval $[t, t + \Delta t]$. Since the conjugate prior for the Poisson mean is the gamma distribution, we assume the conjugate-independent increments prior suggested by Kalbfleisch [21], namely,

$$d\Lambda_0(t) \sim \text{Gamma}\left(c * d\Lambda_0^*(t), c\right), \tag{3.5}$$

where $d\Lambda_0^*(t)$ is a prior mean hazard with c being a scaling parameter representing the "strength" of our prior beliefs. The gamma prior was also chosen for the baseline risk parameter of the Weibull distribution in parametric survival model. Alternatively to implement the Cox model in a fully Bayesian approach, one may use the "multinomial-Poisson trick" described in the OpenBUGS manual that is equivalent to assuming independent increments in the cumulative hazard function. The increments are treated as failure times, and noninformative priors are given for their logarithms. Analogically to the Cox model, a piecewise constant hazard model was implemented. We have modeled baseline hazard using a step function with 3 quantiles t_1, t_2, and t_3 as changing points assuring the same number of events in between. Let t_0 denote the start of the followup, t_4 the maximum censoring time, and $d\Lambda_{0k}(t)$ the increment in the integrated baseline hazard function occurring during the time interval $[t_k, t_{k+1}]$, $k = 0, 1, 2, 3$. Then for different intervals, we specify a separate prior hazard mean $d\Lambda_0^*(t)$ and

$$d\Lambda_{0k}(t) \sim \text{Gamma}\left(c * d\Lambda_{0k}^*(t), c\right). \tag{3.6}$$

Similarly as for the Cox model, the results were not sensitive with respect to the choice of the hyperparameters as long as the priors were sufficiently diffuse. The above nonparametric approach can be criticized as having the independent priors for the hazard distribution. However, as suggested by Kalbfleisch [21] a mixture of gamma priors can be considered as an alternative. Moreover, in a piecewise constant model one can also include change points as unknown parameters in the model as proposed in a Bayesian context by Patra and Dey [22] and applied by Casellas [23].

In order to assess convergence for the full Bayesian model, standard MCMC diagnostic plots were used. The burn-in size was set to 10000 iterations, which was chosen based on the visual inspection of the trace plots and confirmed by the the Raftery and Lewis diagnostics. The same number of iterations were used for constructing the summary statistics. Based on the autocorrelation plots, we have chosen every 30th iteration. Therefore, in total to obtain 10000 iterations for the final inferenc 300000 iterations were required after the burn-in part. Additionally, we run a second parallel chain and used Gelman and Rubin diagnostic plots to assess the convergence.

3.2. Two-Stage Approach

As mentioned in Section 1, the longitudinal measurements in our setting do not constitute an internal time-dependent covariate, since the events took place after the last longitudinal measurement was collected. In particular, since events do not cause nonrandom dropout, the event process does not carry any information for the longitudinal outcome. Mathematically, this means that information for the random effects $\boldsymbol{\alpha}_\mathbf{i}$ is actually only coming from the longitudinal responses, that is,

$$p\left(\boldsymbol{\alpha}_i \mid Y_i(u_{ij}); T_i; \Delta_i; \boldsymbol{\theta}_y\right) = p\left(\boldsymbol{\alpha}_i \mid Y_i(u_{ij}); \boldsymbol{\theta}_y\right). \tag{3.7}$$

Thus, we can avoid the computational complexity of the full likelihood framework presented in Section 3.1 by employing a two-stage approach. More specifically, at Stage I, we obtain $\widehat{\boldsymbol{\theta}}_y$ by maximizing the log-likelihood:

$$l_y\left(\boldsymbol{\theta}_y\right) = \sum_{i=1}^{N} \int p\left(\mathbf{Y}_\mathbf{i} \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y\right) p\left(\boldsymbol{\alpha}_i; \boldsymbol{\theta}_y\right) d\boldsymbol{\alpha}_i. \tag{3.8}$$

This requires numerical integration, and we use a Gaussian quadrature for that purpose. Then we obtain the corresponding empirical Bayes estimates:

$$\widehat{\boldsymbol{\alpha}}_i = \arg\max_{\alpha} \left[\log p\left(Y_i \mid \boldsymbol{\alpha}; \widehat{\boldsymbol{\theta}}_y\right) + \log p\left(\boldsymbol{\alpha}; \widehat{\boldsymbol{\theta}}_y\right)\right] \tag{3.9}$$

and compute the predictions:

$$\widehat{\phi}_i = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{Z}_i\widehat{\boldsymbol{\alpha}}_i. \tag{3.10}$$

At Stage II, we fit the relative risk model:

$$\lambda_i(t) = \lambda_0(t) \exp\left(\boldsymbol{\gamma}^T \widehat{\phi}_i\right). \tag{3.11}$$

However, a potential problem in the above is that $\widehat{\phi}_i$ is not the true subject-specific parameters but rather an estimate with a standard error. If we ignore this measurement error, the regression coefficients $\boldsymbol{\gamma}_i$ will be possibly attenuated. To overcome this problem, we propose here a sampling approach to account for the variability in $\widehat{\phi}_i$, very close in spirit to the Bayesian approach of Section 3.1. In particular, we use the following sampling scheme.

Step 1. simulate $\boldsymbol{\theta}_y^{(m)} \sim N(\widehat{\boldsymbol{\theta}}_y,\ \widehat{\mathrm{var}}(\widehat{\boldsymbol{\theta}}_y))$.

Step 2. simulate $\boldsymbol{\alpha}_i^{(m)} \sim [\boldsymbol{\alpha}_i \mid \mathbf{Y}_\mathbf{i}, \boldsymbol{\theta}_y^{(m)}]$.

Step 3. calculate $\phi_i^{(m)} = \mathbf{X}\boldsymbol{\beta}^{(m)} + \mathbf{Z}_i\boldsymbol{\alpha}_i^{(m)}$ and fit the relative risk model $\lambda_i(t) = \lambda_0(t) \exp\{\boldsymbol{\gamma}^T \phi_i^{(m)}\}$ from which $\widehat{\boldsymbol{\theta}}_t^{(m)} = \widehat{\boldsymbol{\gamma}}^{(m)}$ and $\widehat{\mathrm{var}}(\widehat{\boldsymbol{\theta}}_t^{(m)})$ are kept.

Steps 1–3 are repeated $m = 1, \ldots, M$ times.

Step 1 takes into account the variability of the MLEs, and Step 2 the variability of $\boldsymbol{\alpha}_\mathbf{i}$. Moreover, because the distribution in Step 2 is not of a standard form, we use

a independent Metropolis-Hastings algorithm to sample from it with multivariate t-proposal density centered at an Empirical Bayes estimates $\widehat{\boldsymbol{\alpha}}_i$, covariance matrix $\widehat{\text{var}}(\widehat{\boldsymbol{\alpha}}_i)$, and $df = 4$. The low number of degrees of freedom was chosen to ensure that the proposal density has heavy tails to provide sufficient coverage of the target density $[\boldsymbol{\alpha}_i \mid \mathbf{Y}_i, \boldsymbol{\theta}_y]$. The variance-covariance matrix estimated from the nonlinear mixed model was additionally scaled by some parameter *Scale*. The tuning parameter allows to control the acceptance rate through the range of the proposed distribution. If the range is too narrow, the proposed values will be close to the current ones leading to low rejection rate. On the contrary, if the range is too large, the proposed values are far away from the current ones leading to high rejection rate. We chose the acceptance rate to be 0.5 following Carlin and Louis [24] that suggests a desirable acceptance rates of Metropolis-Hastings algorithms to be around 1/4 for the dependence (random walk) M-H version and 1/2 for the independent M-H. Roberts et al. [25] recommended to use the acceptance rate close to 1/4 for high dimensions and 1/2 for the models with dimensions 1 or 2. They discussed this issue in the context of the random walk proposal density. The authors showed that if the target and proposal densities are normal, then the scale of the latter should be tuned so that the acceptance rate is approximately 0.45 in one-dimensional problems and approximately 0.23 as the number of dimensions approaches infinity, with the optimal acceptance rate being around 0.25 in as low as six dimensions. In our case, an independent version of Metropolis-Hastings algorithm is applied. The proposal density in the algorithm does not depend on the current point as in the random-walk Metropolis algorithm. Therefore, for this sampler to work well, we want to have a proposal density that mimics the target distribution and have the acceptance rate be as high as possible. In order to obtain a desirable acceptance rate one needs to run a sampling algorithm for a number of iterations for a given data set and compute an acceptance rate and then repeat the procedure changing the tuning parameter until the chosen acceptance rate, is obtained. Usually a small number of iterations (100–500) is sufficient for the purpose of calibration. More details about the Metropolis-Hastings acceptance-rejection procedure can be found in the supplementary material (available online at doi:10.1155/2012/194194). A final estimate of $\boldsymbol{\theta}_t$ is obtained using the mean of the estimates from all M iterations:

$$\overline{\widehat{\boldsymbol{\theta}}}_t = \frac{\sum_{m=1}^{M} \widehat{\boldsymbol{\theta}}_t^m}{M}. \tag{3.12}$$

To obtain the SE of $\overline{\widehat{\boldsymbol{\theta}}}_t$, we use the variance-covariance matrix $\mathbf{V}$:

$$\widehat{\mathbf{V}} = \widehat{\mathbf{W}} + \frac{(M+1)\widehat{\mathbf{B}}}{M}, \tag{3.13}$$

where $\widehat{\mathbf{W}}$ is the average within-iteration variance and $\widehat{\mathbf{B}}$ is the between-iteration variance, that is,

$$\begin{aligned} \widehat{\mathbf{W}} &= \frac{\sum_{m=1}^{M} \widehat{\mathbf{U}}^m}{M}, \\ \widehat{\mathbf{B}} &= \frac{1}{M-1} \sum_{m=1}^{M} \left(\widehat{\boldsymbol{\theta}}_t^m - \overline{\widehat{\boldsymbol{\theta}}}_t\right)\left(\widehat{\boldsymbol{\theta}}_t^m - \overline{\widehat{\boldsymbol{\theta}}}_t\right)^T. \end{aligned} \tag{3.14}$$

$\widehat{\mathbf{U}}^m$ represents a variance-covariance matrix estimated in iteration m for $\widehat{\boldsymbol{\gamma}}^m$.

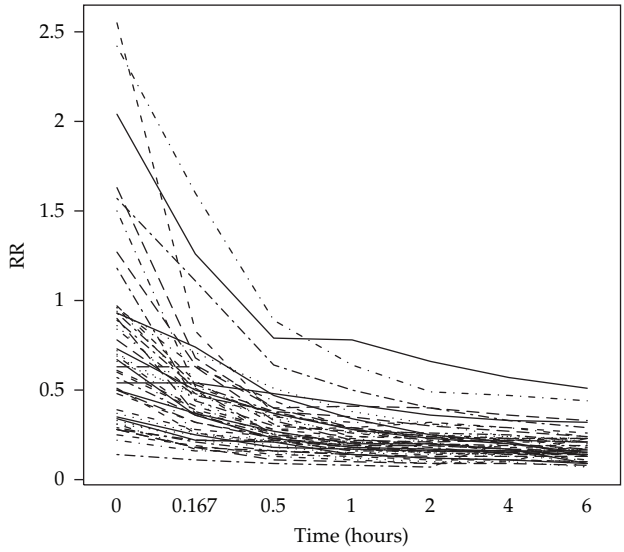


Figure 1: Individual profiles of renal resistance level for 50 sampled donors.

4. Analysis of the RR Data

4.1. Models' Specification

We apply the proposed two-stage procedure and a fully Bayesian approach to the transplantation study introduced in Section 1. The data was taken from an international prospective trial on 337 kidney pairs, which aimed to compare two different types of storage, namely, cold storage and machine perfusion (MP). Here we focus on the second arm. Our main outcome of interest is graft survival time censored after 1 year. At the end of the study, only 26 graft failures were observed. The renal resistance level (RR) was expected to be an important risk factor for graft failure. It was measured using the perfusion machine at the moment of taking the organ out from a donor ($t = 0$), and thereafter at 10 minutes, 30 minutes, 1 hour, 2 hours, 4 hours, and just before transplantation. As mentioned in the Section 1, the time of last measurement was different for different patients and sometimes unknown. However, there was a clear asymptote visible from the individual profiles that was reached after about 5 hours by each patient. Based on that behavior and after the medical consult, we chose the last measurement to be taken at 6 hours for each patient. Other variables of interest include the age of the donor, donor's region (3 countries considered), and donor's type (heart beating or non-heart-beating).

The individual profiles of 50 randomly selected kidney donors are presented in Figure 1. This plot confirms the biological expectation that allografts exhibit their highest renal resistance levels just after being extracted from the donor. Thereafter, they show a smooth decrease in RR until they reach an asymptote above zero indication that there is no "perfect flow" through the kidney. Furthermore, we observe that the initial RR level, the rate of decrease, and the final RR level differ from donor to donor. Additional descriptive plots for our data are presented in the supplementary material.

In the first step of our analysis, we aim to describe the evolution of the renal resistance level in time. Motivated by both biological expectation and Figure 1, we postulate the following nonlinear function:

$$f(t) = \phi_1 + \phi_2 e^{-\phi_3 t}, \tag{4.1}$$

where ϕ_1 is a lower asymptote, $\phi_1 + \phi_2$ is an initial value for $t = 0$, and ϕ_3 is the rate of decrease from ϕ_2 to ϕ_1 (see also Figure 2 in supplementary material).

To accommodate for the shapes of RR evolutions observed in Figure 1, we need to constraint ϕ_1, ϕ_2, and ϕ_3 to be positive. Moreover, in order to allow for individual donor effects, we use the following formulation:

$$Y_i(t) = W_i(t) + \varepsilon(t), \text{ with } W_i(t) = f_i(t) = \exp(\phi_{1i}) + \exp(\phi_{2i}) e^{-\exp(\phi_{3i})t}, \tag{4.2}$$

where

$$\begin{aligned}
\phi_1 &= \beta_{10} + \beta_{11}\ \text{Donor Age} + \beta_{12}\ \text{Donor Type} + \beta_{13}\ \text{Donor Reg1} + \beta_{14}\ \text{Donor Reg2} + \alpha_1,\\
\phi_2 &= \beta_{20} + \beta_{21}\ \text{Donor Age} + \beta_{22}\ \text{Donor Type} + \beta_{23}\ \text{Donor Reg1} + \beta_{24}\ \text{Donor Reg2} + \alpha_2,\\
\phi_3 &= \beta_{30} + \beta_{31}\ \text{Donor Age} + \beta_{32}\ \text{Donor Type} + \beta_{33}\ \text{Donor Reg1} + \beta_{34}\ \text{Donor Reg2} + \alpha_3,
\end{aligned} \tag{4.3}$$

and $\boldsymbol{\alpha}_\mathbf{i} \sim N(0, D)$, $\varepsilon(t) \sim N(0, \sigma^2)$ with $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ and $\operatorname{cov}(\alpha_i, \varepsilon(t)) = 0$. In the second step, the predicted parameters (ϕ_1, ϕ_2, ϕ_3) summarizing the RR evolution of the nonlinear mixed model are included in the graft survival model. The final model for graft survival was of the form:

$$\lambda_i(u) = \lambda_0(u) \exp\left(\gamma_1 \widehat{\phi}_{1i} + \gamma_2 \widehat{\phi}_{2i} + \gamma_3 \widehat{\phi}_{3i}\right). \tag{4.4}$$

To investigate the impact of ignoring that the covariate $\widehat{\phi}_i$ is measured with error, we compared the naive approach in which we ignored this measurement error and our proposal that accounts for the uncertainty in $\widehat{\phi}_\mathbf{i}$ via Monte Carlo sampling. We used Metropolis-Hastings algorithm with independent t-proposal and acceptance rate around 50% for the reason given in Section 3.2. We simulated $M = 10000$ samples with additional initial step of the scaling parameter calibration in order to achieve the desirable acceptance rate. The final estimates (and SE) of the parameters associated with RR covariates were calculated using the formulas described in Section 3.2. We compared the results from the two-stage procedure with the estimates obtained from the fully Bayesian joint model fitted for the data using the priors specified in Section 3.1.

The analysis was performed using *R* Statistical Software. Packages *survival* and *nlme* were used for the two submodels fit, and a separate code was written by the first author for the sampling part. The fully Bayesian model was fitted using OpenBUGS software with the priors specified in Section 3.1. In particular, for the $p \times p$ variance-covariance matrices of multivariate normal priors, we used inverse Wishart distribution with p degrees of freedom. For the variance-covariance parameter of the normal longitudinal response, we took as a prior an inverse-Gamma $(10^{-3}, 10^{-3})$. For the baseline risk parameter of the Weibull distribution in

survival submodel, a Gamma$(10^{-3}, 10^{-3})$ prior was used. To analyze the data using the fully Bayesian Cox model described in Section 3.1, we chose the scaling parameter c in a gamma prior for the independent increments to be equal 0.001 and a prior mean $d\Lambda_0^*(t) = 0.1$. We did not observe any substantial difference for the different values of parameter c as long as c was small enough to keep the prior noninformative. We do not recommend too small values of the scaling parameter c as they can lead to the computation problems. Analogically we have chosen gamma priors for the piecewise constant hazard model. The code for the Bayesian full joint model, as well the R codes for the sampling two-stage procedure, is available from the authors on request.

4.2. Results

The results for the nonlinear-mixed model are presented in Table 1, for the two-stage approach and in supplementary material, for the full Bayesian approach with Weibull survival model. The results for the longitudinal part for the full Bayesian approach with Cox and piecewise constant hazard models were similar (not presented). It can be observed, based on two-stage model results, that only Donor Age had a significant impact on the RR asymptote. Donor Type and Region had a significant impact on the steepness parameter. However, we keep all the covariates in the model for the purpose of prediction for the second stage. The mean RR profiles for Heart-Beating and Non-Heart-Beating donors from different regions together with fitted values based on the obtained nonlinear mixed model are given in the supplementary material.

In the second step of the analysis, we applied at first the naive approach and used the estimates $\widehat{\phi}_1, \widehat{\phi}_2$, and $\widehat{\phi}_3$ from the nonlinear mixed model as fixed covariates in the final Cox models for graft survival. Table 2 presents the results for the survival submodel for the all approaches, namely, the plug-in method, two-stage approach, and the fully Bayesian model. For the fully Bayesian approach, the results for the parametric Weibull model together with Cox and piecewise constant hazard models are listed. The results from Table 2 indicate that only the RR asymptote could have a significant impact on graft survival.

We observe that the estimates are close or almost identical as in plug-in model. SE of the Cox regression coefficients for the model with sampling are greater than SE from the plug-in model in Table 2(a), especially for the parameter ϕ_3. The increase in SE is somewhat the expected and is caused by the additional variability in covariates captured by the sampling approach. The fully Bayesian model produces similar results to our semi-Bayesian sampling model with somewhat lower SE. We do not observe substantial difference between fully parametric and nonparametric models in a fully Bayesian approach. Since in the analyzed real data the number of events is small, the fully Bayesian Cox and piecewise constant hazard Bayesian models were expected to produce similar results. We did not observe any substantial difference for the different values of hyperparameters.

Even though it is hard to compare exactly the computational time for the two approaches, the rough estimation of the total computational time needed to estimate and assess the convergence (2 chains) of the full Bayesian model was about 21.6 hours and depended on the implemented survival model. A similar computational time was needed for the full Bayesian model with the Cox survival model and piecewise constant hazard model with a slightly more time needed for the parametric Weibull model. For the two-stage approach, the total computational time was about 10 hours using the Intel(R) Core(TM)2 Duo T9300 2.5 GHz and 3.5 GB RAM.

Table 1: Parameter estimates, standard errors, and 95% confidence intervals from the nonlinear mixed model for RR.

Effect	Parameter	Estimate	SE	(95% CI)
Fixed effects				
ϕ_1				
Constant	β_{10}	2.838	0.094	(2.654; 3.022)
Donor Age	β_{11}	0.005	0.002	(0.001; 0.009)
Donor Type (HB versus NHB)	β_{12}	−0.102	0.068	(−0.235; 0.031)
Donor Region 1 versus 3	β_{13}	−0.078	0.065	(−0.205; 0.049)
Donor Region 2 versus 3	β_{14}	−0.072	0.072	(−0.213; 0.069)
ϕ_2				
Constant	β_{20}	3.510	0.211	(3.096; 3.924)
Donor Age	β_{21}	0.004	0.004	(−0.004; 0.012)
Donor Type (HB versus NHB)	β_{22}	−0.064	0.154	(−0.365; 0.238)
Donor Region 1 versus 3	β_{23}	−0.107	0.147	(−0.395; 0.181)
Donor Region 2 versus 3	β_{24}	0.033	0.163	(−0.286; 0.352)
ϕ_3				
Constant	β_{30}	1.010	0.186	(0.645; 1.375)
Donor Age	β_{31}	0.003	0.003	(−0.003; 0.009)
Donor Type (HB versus NHB)	β_{32}	0.402	0.130	(0.147; 0.657)
Donor Region 1 versus 3	β_{33}	−0.284	0.125	(−0.529; −0.039)
Donor Region 2 versus 3	β_{34}	−0.032	0.138	(−0.302; 0.238)
Random effects				
se(α_1)	d_{11}	0.396		
se(α_2)	d_{22}	0.955		
se(α_3)	d_{33}	0.572		
cov(α_1, α_2)	d_{12}	0.257		
cov(α_1, α_3)	d_{13}	−0.053		
cov(α_2, α_3)	d_{23}	0.023		
se(ε_{ij})	σ	7.507		

5. Simulations

5.1. Design

We have conducted a number of simulations to investigate the performance of our proposed two-stage method. In particular, we compared the plug-in method that uses the Empirical Bayes estimates $\widehat{\phi}_i$ from the nonlinear mixed model and ignores the measurement error, the two-stage Monte Carlo sampling approach that accounts for the variability in $\widehat{\phi}_i$, and the fully Bayesian approach. In the fully Bayesian approach, only the parametric Weibull model was considered.

For the longitudinal part, the data were simulated for 500 patients from model (5.1) with $\phi_{1i} = \beta_{10} + \alpha_{1i}$, $\phi_{2i} = \beta_{20} + \alpha_{2i}$ and $\phi_{3i} = \beta_{30} + \alpha_{3i}$, $\boldsymbol{\alpha}_i \sim N(0, \mathbf{D})$, $Y \sim N(f(t), \sigma^2)$. The variance-covariance matrix $\mathbf{D}$ of the random effects was chosen to be $\mathbf{D} = \text{vech}(0.6,\ 0.01, -0.01,\ 0.6,\ 0.01,\ 0.3)$. We kept 7 measurement points as in the original analyzed data set as well as the nonequal distances between them. The variance of the

Table 2: Parameter estimates, SE, and 95% confidence/credibility intervals from proportional hazards Cox model for graft survival for plug-in method (a), sampled covariates (b), and fully Bayesian approach (c, d, e).

(a) Graft survival, plug-in

Effect	Parameter	log(HR)	SE	(95% CI)
$\exp(\phi_1)$	γ_1	0.052	0.022	(0.009; 0.095)
$\exp(\phi_2)$	γ_2	−0.005	0.005	(−0.015; 0.005)
$\exp(\phi_3)$	γ_3	0.053	0.158	(−0.257; 0.363)

(b) Graft survival, sampling two-stage

Effect	Parameter	log(HR)	SE	(95% CI)
$\exp(\phi_1)$	γ_1	0.053	0.024	(0.006; 0.100)
$\exp(\phi_2)$	γ_2	−0.006	0.008	(−0.022; 0.010)
$\exp(\phi_3)$	γ_3	0.055	0.185	(−0.308; 0.418)

(c) Graft survival, fully Bayesian, Weibull

Effect	Parameter	log(HR)	SE	(95% HPD)
$\exp(\phi_1)$	γ_1	0.058	0.023	(0.013; 0.103)
$\exp(\phi_2)$	γ_2	−0.005	0.008	(−0.020; 0.011)
$\exp(\phi_3)$	γ_3	0.056	0.180	(−0.299; 0.411)

(d) Graft survival, fully Bayesian, Cox

Effect	Parameter	log(HR)	SE	(95% HPD)
$\exp(\phi_1)$	γ_1	0.056	0.023	(0.010; 0.101)
$\exp(\phi_2)$	γ_2	−0.006	0.008	(−0.022; 0.010)
$\exp(\phi_3)$	γ_3	0.055	0.171	(−0.280; 0.390)

(e) Graft survival, fully Bayesian, piecewise constant hazard

Effect	Parameter	log(HR)	SE	(95% HPD)
$\exp(\phi_1)$	γ_1	0.054	0.024	(0.007; 0.102)
$\exp(\phi_2)$	γ_2	−0.005	0.009	(−0.022; 0.012)
$\exp(\phi_3)$	γ_3	0.054	0.179	(−0.297; 0.405)

measurement error σ^2 was chosen to be 0.25, 1, and 25. Survival times were simulated for each patient using the exponential model with the rate parameter equal $\exp(\lambda)$, where

$$\lambda = \gamma_1 \exp(\phi_1) + \gamma_2 \exp(\phi_2) + \gamma_3 \exp(\phi_3). \tag{5.1}$$

We considered scenarios with fixed coefficients $\gamma_1 = 0.5$, $\gamma_2 = 0.5$, and $\gamma_3 = -0.2$. The censoring mechanism was simulated independently using an exponential distribution $\text{Exp}(\lambda_C)$. Parameter λ_C was changed in order to control proportion of censored observations. Different scenarios with 40% and 20% of censoring were examined. For each simulated data set we fitted four survival models, namely, the gold standard model that uses the true simulated values ϕ_i, the plug-in model, the sampling model, and fully Bayesian joint model. Neither nonparametric Cox nor piecewise constant hazard model were considered in a fully Bayesian approach since we have simulated the data from the parametric exponential model and wanted to compare the proposed two-stage approach with the "best" fully Bayesian

model. All the prior settings, size of burn-in, number of iterations, and so forth, for the fully Bayesian model were the same as for the real data analysis.

5.2. Results

In Table 3, we present the average results for 200 simulations of different scenarios are presented. The results from our sampling model were very close to the results obtained for the fully Bayesian model with slightly smaller bias and RMSE for the fully Bayesian approach. That was due to the better estimation of random effects variability in fully Bayesian approach. The plug-in method produced the biggest bias that sometimes with combination with the small variability of the estimates around the biased mean resulted in larger RMSE than in sampling approach. As the measurement error or the percentage of censored observations increased, the estimates of survival submodel were more biased with larger RMSE for all approaches. The increase in bias was more severe when the measurement error variance was increased rather than when the percentage of to censoring was higher. This bias was, however, decreased when the number of repeated measures per individual was increased. This has to do with the amount of information that is available in the data for the estimation of $\widehat{\phi}_i$. As it is known from the standard mixed models literature [26], the degree of shrinkage in the subject-specific predicted values is proportional to σ and inversely proportional to n_i and σ_α. To compare the relation between variance of the random effects and variance of the measurement error, one can use intraclass correlation (ICC) defined as the proportion of the total variability that is explained by the clustering with a given random effect. ICC was similar for the simulated and the real data for the biggest σ and increased in a simulation data as σ decreased.

Since the calculations for the simulation study were highly computationally intensive, we have used the cluster with about 20 nodes with AMD Quad-Core Opteron 835X, 4 × 2 GHz, and 16 GB RAM per node. The analysis for the the 200 simulated data sets for a single scenario took about 65.5 hours using the Bayesian approach and 31.2 hours using the two-stage approach.

6. Discussion

We have proposed a two-stage method that can be used in a joint analysis of longitudinal and time to event data when the longitudinal data are collected before the start of followup for survival, and the interest is in estimation of the impact of longitudinal profiles on survival. The modeling strategy is based on specification of two separate submodels for the longitudinal and time to event data. First, the longitudinal outcome is modeled using a random effects model. Then the survival outcome is modeled using the Empirical Bayes estimates of the subject-specific effects from the first stage. The variability of the estimates from the first stage is properly taken into account using a Monte Carlo approach by sampling from the posterior distribution of the random effects given the data.

As it was demonstrated, ignoring the additional variability of the subject-specific estimates when modeling survival leads to some bias, and in particular, attenuates the regression coefficients towards zero [4]. That was also confirmed by our simulation study. In comparison with the fully Bayesian approach, the proposed partially Bayesian method produced similar results with substantially less number of iterations required. This is due to the fact that sampling was conducted already around the EB estimates, and there is no

Table 3: Bias and residual mean squared error (RMSE) for the method with true ϕ_i (GS), Empirical Bayes estimates $\widehat{\phi}_i$ (Plug-in), sampled ϕ_i, and fully Bayesian approach.

7 time points						
% censoring	20			40		
σ	0.5					
	γ_1	γ_2	γ_3	γ_1	γ_2	γ_3
GS	0.00 (0.04)	−0.02 (0.03)	0.01 (0.03)	−0.01 (0.04)	0.02 (0.04)	−0.02 (0.04)
plug-in	−0.05 (0.06)	−0.04 (0.05)	0.06 (0.07)	−0.08 (0.09)	−0.04 (0.05)	0.12 (0.12)
sampling	−0.04 (0.05)	0.03 (0.08)	0.02 (0.07)	−0.05 (0.11)	−0.02 (0.06)	0.03 (0.10)
Bayesian	−0.03 (0.04)	−0.02 (0.04)	0.01 (0.02)	−0.01 (0.04)	−0.02 (0.04)	0.02 (0.07)
σ	1					
GS	0.04 (0.05)	0.04 (0.07)	−0.03 (0.07)	−0.05 (0.09)	−0.04 (0.06)	−0.03 (0.05)
plug-in	−0.07 (0.08)	−0.08 (0.09)	0.07 (0.09)	−0.10 (0.12)	−0.08 (0.09)	0.08 (0.11)
sampling	−0.07 (0.09)	−0.06 (0.10)	−0.02 (0.11)	−0.05 (0.12)	0.05 (0.11)	−0.03 (0.12)
Bayesian	0.01 (0.03)	0.05 (0.06)	−0.03 (0.07)	0.05 (0.06)	0.04 (0.06)	−0.04 (0.07)
σ	5					
GS	0.04 (0.06)	0.05 (0.06)	0.04 (0.08)	0.05 (0.10)	0.01 (0.05)	−0.02 (0.06)
plug-in	−0.09 (0.10)	−0.10 (0.11)	0.08 (0.11)	−0.20 (0.22)	−0.21 (0.22)	0.14 (0.18)
sampling	0.08 (0.13)	0.06 (0.12)	−0.05 (0.12)	0.07 (0.14)	−0.05 (0.13)	−0.11 (0.18)
Bayesian	0.09 (0.10)	0.05 (0.09)	−0.09 (0.10)	−0.09 (0.10)	0.08 (0.12)	−0.12 (0.18)
14 time points						
% censoring	20			40		
σ	0.5					
	γ_1	γ_2	γ_3	γ_1	γ_2	γ_3
GS	−0.03 (0.03)	0.00 (0.02)	−0.02 (0.03)	0.02 (0.03)	−0.03 (0.04)	0.02 (0.04)
plug-in	−0.02 (0.03)	−0.03 (0.04)	0.05 (0.07)	−0.02 (0.04)	−0.03 (0.04)	0.05 (0.06)
sampling	0.03 (0.04)	0.02 (0.06)	0.02 (0.07)	0.02 (0.04)	0.04 (0.05)	0.02 (0.08)
Bayesian	−0.03 (0.04)	−0.02 (0.04)	−0.02 (0.04)	0.02 (0.04)	0.03 (0.04)	−0.05 (0.06)
σ	1					
GS	−0.03 (0.04)	−0.03 (0.04)	−0.01 (0.03)	0.00 (0.03)	−0.02 (0.04)	0.05 (0.06)
plug-in	−0.09 (0.06)	−0.05 (0.06)	0.06 (0.07)	−0.02 (0.04)	−0.04 (0.05)	0.11 (0.11)
sampling	0.04 (0.08)	0.02 (0.08)	−0.02 (0.07)	−0.02 (0.04)	−0.02 (0.08)	0.04 (0.09)
Bayesian	−0.03 (0.04)	0.04 (0.05)	−0.03 (0.05)	0.02 (0.04)	0.03 (0.05)	0.06 (0.07)
σ	5					
GS	−0.03 (0.04)	−0.03 (0.04)	0.01 (0.04)	−0.01 (0.04)	−0.02 (0.04)	0.05 (0.06)
plug-in	−0.05 (0.06)	−0.10 (0.11)	0.07 (0.09)	−0.10 (0.11)	−0.09 (0.10)	0.11 (0.12)
sampling	0.04 (0.09)	0.04 (0.11)	−0.05 (0.11)	0.07 (0.12)	0.05 (0.11)	−0.06 (0.16)
Bayesian	0.03 (0.05)	0.03 (0.08)	−0.05 (0.10)	0.02 (0.04)	0.06 (0.10)	−0.09 (0.14)

needed for a burn-in part as in a standard fully Bayesian approach. We used 10000 iterations per subject, which was about the size of burn-in needed in the fully Bayesian models. No thinning was used in our approach, based on the visual inspection of the trace plots. Though it is hard to compare the fully Bayesian approach and the two-stage approach with respect to the computational time precisely, the rough approximation of the total computational time required for the two-stage approach was about half in comparison with the fully Bayesian

approach. The fully Bayesian approach provided similar results with the two-stage approach for the special setting we have considered here. However, fitting a fully Bayesian model was a bit of "overdone" in the sense that by design the longitudinal data could not be affected by the survival. Since in many transplantation studies, the longitudinal data are collected before the start of followup for survival; therefore, using our method in that cases seems to be more appropriate than using a fully Bayesian approach. We recommend the proposed approach not only for the particular transplantation studies but in any setting that shares the similarity of the separated followup periods for the two analyzed endpoints. That is, for example, when the event process does not carry any information for the longitudinal outcome and the condition (3.7) from Section 3.2 holds. The simulation results indicate that even if the data come from the real joint setting in which (3.7) may not hold, the proposed two-stage procedure can be comparable to the fully Bayesian approach.

Since the sampling in the proposed method relies strongly on the results of the first part, the accurate estimation of all parameters of nonlinear mixed model is a key feature and should be performed carefully. This can be problematic when the deviation from normality of the random effects, is suspected. However, it was shown that even for the nonnormal random effects one can still use a standard software such as *nlmixed* in SAS with just a small change in a standard program code. In such cases, the probability integral transformation (PIT) proposed by Nelson et al. [27] can be used or the reformulation of the likelihood proposed by Liu and Yu [28]. An alternative is fitting a Bayesian model only to estimate the longitudinal submodel in the first stage, instead of the likelihood methods. Fitting nonlinear mixed models using Bayesian standard software can be, however, problematic due to the high nonlinearity in random effects that is caused both by the nonlinear function of the longitudinal profiles and by the possible restrictions on parameters [29].

In comparison with the two-stage approach proposed by Tsiatis et al. [5], our method is less computationally intensive since it does not require fitting as many mixed models as there are event times in the data. An alternative, that is somewhat simpler to implement and does not require any assumption about the distribution on the underlying random effects, is the conditional score approach proposed by Tsiatis and Davidian [11]. However, this method is less efficient than the methods based on likelihood. The focus in the discussed approaches is on the association between the longitudinal and event time processes. However, in transplantation studies when the two followup periods for longitudinal and survival outcomes are often separated, the interest is rather in making an inference on the marginal event-time distribution. This is similar to the Bayesian approach proposed by Xu and Zeger [12], that uses the longitudinal data as auxiliary information or surrogate for time-to-event data. Our approach is particularly useful in this setting. Since each of the two submodels is fitted separately, a standard software can be used to implement our method with just a small part of additional programming for Monte Carlo sampling.

Another advantage of the proposed two-stage method is that it can be easily generalized from survival to other types of models as it was applied for the binary Delayed Graft Failure (DGF) indicator (results not shown) in the analysis of the renal data. For that purpose in the second step of the two-stage procedure, the survival model was replaced by the logistic regression model for the DGF indicator. The first stage of the proposed approach could be also modified allowing for other types of longitudinal response and other types of mixed models. Therefore, instead of using a nonlinear mixed model a linear mixed model or generalized linear mixed model (GLMMs) can be considered depending on the type and the shape of the longitudinal response. In the presented real data example, we have chosen the three parameters that described the evolution of the longitudinal response. However, for the

particular question of interest, one can easily choose the most convenient parametrization for the longitudinal model and use the selected parameters to analyze the nonlongitudinal response in the second stage.

Acknowledgment

The authors thank J. M. Smits from Eurotransplant International Foundation in Leiden for providing the data set analyzed in the paper and for the medical consult regarding the application of the proposed method.

References

[1] J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ, USA, Second edition, 2002.

[2] C. L. Faucett and D. C. Thomas, "Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach," *Statistics in Medicine*, vol. 15, no. 15, pp. 1663–1685, 1996.

[3] M. S. Wulfsohn and A. A. Tsiatis, "A joint model for survival and longitudinal data measured with error," *Biometrics. Journal of the International Biometric Society*, vol. 53, no. 1, pp. 330–339, 1997.

[4] R. L. Prentice, "Covariate measurement errors and parameter estimation in a failure time regression model," *Biometrika*, vol. 69, no. 2, pp. 331–342, 1982.

[5] A. Tsiatis, V. DeGruttola, and M. Wulfsohn, "Modeling the relationship of survival to longitudinal data measured with error: applications to survival and CD4 counts in patients with AIDS," *Journal of the American Statistical Association*, vol. 90, pp. 27–37, 1995.

[6] P. S. Albert and J. H. Shih, "On estimating the relationship between longitudinal measurments and time-to-event data using a simple two-stage procedure," *Biometrics*, vol. 66, pp. 983–991, 2009.

[7] D. Rizopoulos, G. Verbeke, and E. Lesaffre, "Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data," *Journal of the Royal Statistical Society B*, vol. 71, no. 3, pp. 637–654, 2009.

[8] R. Henderson, P. Diggle, and A. Dobson, "Joint modelling of longitudinal measurements and event time data," *Biostatistics*, vol. 4, pp. 465–480, 2000.

[9] V. DeGruttola and X. Tu, "Modeling progression of CD4 lymphocyte count and its relationship to survival time," *Biometrics*, vol. 50, pp. 1003–1014, 1994.

[10] S. Self and Y. Pawitan, "Modeling a marker of disease progression and onset of disease," in *AIDS Epidemiology: Methodological Issues*, N. P. Jewell, K. Dietz, and V.T. Farewell, Eds., Birkhauser, Boston, Mass, USA, 1992.

[11] A. A. Tsiatis and M. Davidian, "A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error," *Biometrika*, vol. 88, no. 2, pp. 447–458, 2001.

[12] J. Xu and S. L. Zeger, "Joint analysis of longitudinal data comprising repeated measures and times to events," *Journal of the Royal Statistical Society C*, vol. 50, no. 3, pp. 375–387, 2001.

[13] Y. Wang and J. M. G. Taylor, "Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome," *Journal of the American Statistical Association*, vol. 96, no. 455, pp. 895–905, 2001.

[14] J. G. Ibrahim, M.-H. Chen, and D. Sinha, "Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials," *Statistica Sinica*, vol. 14, no. 3, pp. 863–883, 2004.

[15] E. R. Brown and J. G. Ibrahim, "A Bayesian semiparametric joint hierarchical model for longitudinal and survival data," *Biometrics*, vol. 59, no. 2, pp. 221–228, 2003.

[16] L. Wu, X. J. Hu, and H. Wu, "Joint inference for nonlinear mixed-effects models and time to event at the presence of missing data," *Biostatistics*, vol. 9, pp. 308–320, 2008.

[17] C. Hu and M. E. Sale, "A joint model for nonlinear longitudinal data with informative dropout," *Journal of Pharmacokinetics and Pharmacodynamics*, vol. 30, no. 1, pp. 83–103, 2003.

[18] N. A. Kaciroti, T. E. Raghunathan, M. A. Schork, and N. M. Clark, "A Bayesian model for longitudinal count data with non-ignorable dropout," *Journal of the Royal Statistical Society C*, vol. 57, no. 5, pp. 521–534, 2008.

[19] P. K. Andersen and R. D. Gill, "Cox's regression model for counting processes: a large sample study," *The Annals of Statistics*, vol. 10, no. 4, pp. 1100–1120, 1982.

[20] D. Clayton, "Bayesian analysis of frailty models," Tech. Rep., Medical Research Council Biostatistics Unit., Cambridge, UK, 1994.

[21] J. D. Kalbfleisch, "Non-parametric Bayesian analysis of survival time data," *Journal of the Royal Statistical Society B*, vol. 40, no. 2, pp. 214–221, 1978.

[22] K. Patra and D. K. Dey, "A general class of change point and change curve modeling for life time data," *Annals of the Institute of Statistical Mathematics*, vol. 54, no. 3, pp. 517–530, 2002.

[23] J. Casellas, "Bayesian inference in a piecewise Weibull proportional hazards model with unknown change points," *Journal of Animal Breeding and Genetics*, vol. 124, no. 4, pp. 176–184, 2007.

[24] B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall, New York, NY, USA, 2000.

[25] G. O. Roberts, A. Gelman, and W. R. Gilks, "Weak convergence and optimal scaling of random walk Metropolis algorithms," *The Annals of Applied Probability*, vol. 7, no. 1, pp. 110–120, 1997.

[26] G. Verbeke and G. Molenberghs, *Linear mixed models for longitudinal data*, Springer Series in Statistics, Springer, New York, NY, USA, 2000.

[27] K. P. Nelson, S. R. Lipsitz, G. M. Fitzmaurice, J. Ibrahim, M. Parzen, and R. Strawderman, "Using the probability integral transformation for non-normal random effects in non-linear mixed models," *Journal of Computational and Graphical Statistics*, vol. 15: 3957, 2004.

[28] L. Liu and Z. Yu, "A likelihood reformulation method in non-normal random effects models," *Statistics in Medicine*, vol. 27, no. 16, pp. 3105–3124, 2008.

[29] M. Davidian and D. M. Giltinan, *Nonlinear Models for Repeated Measurment Data*, Chapman and Hall, New York, NY, USA, 1998.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2012, Article ID 617678, 26 pages
doi:10.1155/2012/617678

Research Article

Bayesian Approach to Zero-Inflated Bivariate Ordered Probit Regression Model, with an Application to Tobacco Use

Shiferaw Gurmu[1] and Getachew A. Dagne[2]

[1] *Department of Economics, Andrew Young School of Policy Studies, Georgia State University, P.O. Box 3992, Atlanta, GA 30302, USA*
[2] *Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, Tampa, FL 33612, USA*

Correspondence should be addressed to Shiferaw Gurmu, sgurmu@gsu.edu

Received 13 July 2011; Revised 18 September 2011; Accepted 2 October 2011

Academic Editor: Wenbin Lu

This paper presents a Bayesian analysis of bivariate ordered probit regression model with excess of zeros. Specifically, in the context of joint modeling of two ordered outcomes, we develop zero-inflated bivariate ordered probit model and carry out estimation using Markov Chain Monte Carlo techniques. Using household tobacco survey data with substantial proportion of zeros, we analyze the socioeconomic determinants of individual problem of smoking and chewing tobacco. In our illustration, we find strong evidence that accounting for excess zeros provides good fit to the data. The example shows that the use of a model that ignores zero-inflation masks differential effects of covariates on nonusers and users.

1. Introduction

This paper is concerned with joint modeling of two ordered data outcomes allowing for excess zeros. Economic, biological, and social science studies often yield data on two ordered categorical variables that are jointly dependent. Examples include the relationship between desired and excess fertility [1, 2], helmet use and motorcycle injuries [3], ownership of dogs and televisions [4], severity of diabetic retinopathy of the left and right eyes [5], and self-assessed health status and wealth [6]. The underlying response variables could be measured on an ordinal scale. It is also common in the literature to generate a categorical or grouped variable from an underlying quantitative variable and then use ordinal response regression model (e.g., [4, 5, 7]). The ensuing model is usually analyzed using the bivariate ordered probit model.

Many ordered discrete data sets are characterized by excess of zeros, both in terms of the proportion of nonusers and relative to the basic ordered probit or logit model. The zeros may be attributed to either corner solution to consumer optimization problem or errors in recording. In the case of individual smoking behavior, for example, the zeros may be recorded for individuals who never smoke cigarettes or for those who either used tobacco in the past or are potential smokers. In the context of individual patents applied for by scientists during a period of five years, zero patents may be recorded for scientists who either never made patent applications or for those who do but not during the reporting period [8]. Ignoring the two types of zeros for nonusers or nonparticipants leads to model misspecification.

The univariate as well as bivariate zero-inflated count data models are well established in the literature for example, Lambert [9], Gurmu and Trivedi [10], Mullahy [11], and Gurmu and Elder [12]. The recent literature presents a Bayesian treatment of zero-inflated Poisson models in both cross-sectional and panel data settings (see [13, 14], and references there in). By contrast, little attention has been given to the problem of excess zeros in the ordered discrete choice models. Recently, an important paper by Harris and Zhao [15] developed a zero-inflated univariate ordered probit model. However, the problem of excess zeros in ordered probit models has not been analyzed in the Bayesian framework. Despite recent applications and advances in estimation of bivariate ordered probit models [1–6], we know of no studies that model excess zeros in bivariate ordered probit models.

This paper presents a Bayesian analysis of bivariate ordered probit model with excess of zeros. Specifically, we develop a zero-inflated ordered probit model and carry out the analysis using the Bayesian approach. The Bayesian analysis is carried out using Markov Chain Monte Carlo (MCMC) techniques to approximate the posterior distribution of the parameters. Bayesian analysis of the univariate zero-inflated ordered probit will be treated as a special case of the zero-inflated bivariate order probit model. The proposed models are illustrated by analyzing the socioeconomic determinants of individual choice problem of bivariate ordered outcomes on smoking and chewing tobacco. We use household tobacco prevalence survey data from Bangladesh. The observed proportion of zeros (those identifying themselves as nonusers of tobacco) is about 76% for smoking and 87% for chewing tobacco.

The proposed approach is useful for the analysis of ordinal data with natural zeros. The empirical analysis clearly shows the importance of accounting for excess zeros in ordinal qualitative response models. Accounting for excess zeros provides good fit to the data. In terms of both the signs and magnitudes of marginal effects, various covariates have differential impacts on the probabilities associated with the two types of zeros, nonparticipants and zero-consumption. The usual analysis that ignores excess of zeros masks these differential effects, by just focusing on observed zeros. The empirical results also show the importance of taking into account the uncertainty in the parameter estimates. Another advantage of the Bayesian approach to modeling excess zeros is the flexibility, particularly computational, of generalizing to multivariate ordered response models.

The rest of the paper is organized as follows. Section 2 describes the proposed zero-inflated bivariate probit model. Section 3 presents the MCMC algorithm and model selection procedure for the model. An illustrative application using household tobacco consumption data is given in Section 4. Section 5 concludes the paper.

2. Zero-Inflated Bivariate Ordered Probit Model

2.1. The Basic Model

We consider the basic Bayesian approach to a bivariate latent variable regression model with excess of zeros. To develop notation, let $\widetilde{y}_{1i}^{*}$ and $\widetilde{y}_{2i}^{*}$ denote the bivariate latent variables. We

consider two observed ordered response variables $\tilde{y}_{1i}$ and $\tilde{y}_{2i}$ taking on values $0, 1, \ldots, J_r$, for $r = 1, 2$. Define two sets of cut-off parameters $\alpha^r = (\alpha_{r2}, \alpha_{r3}, \ldots, \alpha_{rJr})$, $r = 1, 2$, where the restrictions $\alpha_{r0} = -\infty$, $\alpha_{rJ_r+1} = \infty$, and $\alpha_{r1} = 0$ have been imposed. We assume that $(\tilde{y}^*_{1i}, \tilde{y}^*_{2i})' \equiv \tilde{\mathbf{y}}^*_i$ follows a bivariate regression model

$$\tilde{y}^*_{ri} = \mathbf{x}'_{ri}\boldsymbol{\beta}_r + \varepsilon_{ri}, \quad r = 1, 2, \tag{2.1}$$

where $\mathbf{x}_{ri}$ is a K_r-variate of regressors for the ith individual ($i = 1, \ldots, N$) and ε_{ri} are the error terms. For subsequent analysis, let $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2)'$, $\boldsymbol{\epsilon}_i = (\epsilon_{1i}, \epsilon_{2i})'$, and

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{x}'_{1i} & 0' \\ 0' & \mathbf{x}'_{2i} \end{pmatrix}. \tag{2.2}$$

Analogous to the univariate case, the observed bivariate-dependent variables are defined as

$$\tilde{y}_{ri} = \begin{cases} 0 & \text{if } \tilde{y}^*_{ri} \le 0, \\ 1 & \text{if } 0 < \tilde{y}^*_{ri} \le \alpha_{r2}, \\ j & \text{if } \alpha_{rj} < \tilde{y}^*_{ri} \le \alpha_{rj+1}, \ j = 2, 3, \ldots, J_r - 1, \\ J_r & \text{if } \tilde{y}^*_i \le \alpha_{rJ_r}, \end{cases} \tag{2.3}$$

where $r = 1, 2$. Let $\tilde{\mathbf{y}}_i = (\tilde{y}_{1i}, \tilde{y}_{2i})'$.

We introduce inflation at the point $(\tilde{y}_{1i} = 0, \ \tilde{y}_{2i} = 0)$, called the zero-zero state. As in the univariate case, define the participation model:

$$\begin{aligned} s^*_i &= \mathbf{z}'_i\gamma + \mu_i, \\ s_i &= I\left(s^*_i > 0\right). \end{aligned} \tag{2.4}$$

In the context of the zero-inflation model, the observed response random vector $\mathbf{y}_i = (y_{1i}, y_{2i})'$ takes the form

$$\mathbf{y}_i = s_i\tilde{\mathbf{y}}_i. \tag{2.5}$$

We observe $\mathbf{y}_i = \mathbf{0}$ when either the individual is a non-participant ($s_i = 0$) or the individual is a zero-consumption participant ($s_i = 1$ and $\tilde{\mathbf{y}}_i = \mathbf{0}$). Likewise, we observe positive outcome (consumption) when the individual is a positive consumption participant for at least one good ($s_i = 1$ and $\tilde{\mathbf{y}}_i \neq \mathbf{0}$).

Let $\Phi(a)$ and $\phi(a)$ denote the respective cumulative distribution and probability density functions of standardized normal evaluated at a. Assuming normality and that μ_i is

uncorrelated with $(\varepsilon_{1i}, \varepsilon_{2i})$, but $\operatorname{corr}(\varepsilon_{1i}, \varepsilon_{2i}) = \rho_{12} \neq 0$, and each component with unit variance, the zero-inflated bivariate ordered probit (ZIBOP) distribution is

$$f_b(\mathbf{y}_i^*, \mathbf{y}_i, s_i^*, s_i \mid \mathbf{X}_i, \mathbf{z}_i, \Psi) = \begin{cases} \Pr(s_i = 0) + (1 - \Pr(s_i = 0))\Pr(\tilde{y}_{1i} = 0, \tilde{y}_{2i} = 0), & \text{for } (\tilde{y}_{1i}, \tilde{y}_{2i}) = (0,0) \\ (1 - \Pr(s_i = 0))\Pr(\tilde{y}_{1i} = j, \tilde{y}_{2i} = l), & \text{for } (\tilde{y}_{1i}, \tilde{y}_{2i}) \neq (0,0), \end{cases} \tag{2.6}$$

where $j = 0, 1, \ldots, J_1$, $l = 0, 1, \ldots, J_2$, $\Pr(s_i = 0) = \Phi(-\mathbf{z}_i'\gamma)$, $\Pr(s_i = 1) = \Phi(-\mathbf{z}_i'\gamma)$. Further, for $(\tilde{y}_{1i}, \tilde{y}_{2i}) = (0,0)$ in (2.6), we have $\alpha_{r0} = -\infty, \alpha_{r1} = 0$ for $r = 1, 2$ so that

$$\Pr(\tilde{y}_{1i} = 0, \tilde{y}_{2i} = 0) = \Phi_2(-\mathbf{x}_{1i}'\beta_1, -\mathbf{x}_{2i}'\beta_2, \rho_{12}), \tag{2.7}$$

where $\Phi_2(\cdot)$ is the cdf for the standardized bivariate normal. Likewise, $\Pr(\tilde{y}_{1i} = j, \tilde{y}_{2i} = l)$ in (2.6) are given by

$$\begin{aligned} \Pr(\tilde{y}_{1i} = j, \tilde{y}_{2i} = l) &= \Phi_2(\alpha_{1j+1} - \mathbf{x}_{1i}'\beta_1, \alpha_{2l+1} - \mathbf{x}_{2i}'\beta_2; \rho_{12}) \\ &\quad - \Phi_2(\alpha_{1j} - \mathbf{x}_{1i}'\beta_1, \alpha_{2l} - \mathbf{x}_{2i}'\beta_2, \rho_{12}) \quad \text{for } j = 1, \ldots, J_1 - 1;\ l = 1, \ldots, J_2 - 1; \\ \Pr(\tilde{y}_{1i} = J_1, \tilde{y}_{2i} = J_2) &= 1 - \Phi_2(\alpha_{1J_1} - \mathbf{x}_{1i}'\beta_1, \alpha_{2J_2} - \mathbf{x}_{2i}'\beta_2, \rho_{12}). \end{aligned} \tag{2.8}$$

The ensuing likelihood contribution for N-independent observations is

$$\begin{aligned} \mathcal{L}_b(\mathbf{y}^*, \mathbf{y}, s^*, s \mid \mathbf{X}, \mathbf{z}, \Psi_b) &= \prod_{i=1}^{N} \prod_{(j,l)=(0,0)} \left[\Pr(s_i = 0) + (1 - \Pr(s_i = 0))\Pr(\tilde{y}_{1i} = 0, \tilde{y}_{2i} = 0)\right]^{d_{ijl}} \\ &\quad \times \prod_{i=1}^{N} \prod_{(j,l)\neq(0,0)} \left[(1 - \Pr(s_i = 0))\Pr(\tilde{y}_{1i} = j, \tilde{y}_{2i} = l)\right]^{d_{ijl}}, \end{aligned} \tag{2.9}$$

where $d_{ijl} = 1$ if $\tilde{y}_{1i} = j$ and $\tilde{y}_{2i} = l$, and $d_{ijl} = 0$ otherwise. Here, the vector Ψ_b consists of $\beta, \gamma, \alpha^1, \alpha^2$, and the parameters associated with the trivariate distribution of (ϵ, μ).

Regarding identification of the parameters in the model defined by (2.1) through (2.5) with normality assumption, we note that the mean parameter (joint choice probability associate with the observed response vector $\mathbf{y}_i$) depends nonlinearly on the probability of zero inflation ($\Phi(-\mathbf{z}_i'\gamma)$) and choice probability ($\Pr(\tilde{y}_{1i} = j, \tilde{y}_{2i} = l)$) coming from the BOP submodel. Since the likelihood function for ZIBOP depends separately on the two regression components, the parameters of ZIBOP model with covariates are identified as long as the model is estimated by full maximum likelihood method. The same or different sets of covariates can affect the two components via $\mathbf{z}_i$ and $\mathbf{x}_{ri}$. When using quasi-likelihood estimation or generalized estimating equations methods rather than full ML, the class of identifiable zero-inflated count and ordered data models is generally more restricted; see, for example, Hall and Shen [16] and references there in. Although the parameters in the ZIBOP

model above are identified through a nonlinear functional form estimated by ML, for more robust identification we can use traditional exclusion restrictions by including instrumental variables in the inflation equation, but excluding them from the ordered choice submodel. We follow this strategy in the empirical section.

About 2/3 of the observations in our tobacco application below have a double-zero-state, $(y_1 = 0, y_2 = 0)$. Consequently, we focused on a mixture constructed from a point mass at $(0,0)$ and a bivariate ordered probit. In addition to allowing for inflation in the double-zero-state, our approach can be extended to allow for zero-inflation in each component.

2.2. Marginal Effects

It is common to use marginal or partial effects to interpret covariate effects in nonlinear models; see, for example, Liu et al. [17]. Due to the nonlinearity in zero-inflated ordered response models and in addition to estimation of regression parameters, it is essential to obtain the marginal effects of changes in covariates on various probabilities of interest. These include the effects of covariates on probability of nonparticipation (zero-inflation), probability of participation, and joint and/or marginal probabilities of choice associated with different levels of consumption.

From a practical point of view, we are less interested in the marginal effects of explanatory variables on the joint probabilities of choice from ZIBOP. Instead, we focus on the marginal effects associated with the marginal distributions of y_{ri} for $r = 1, 2$. Define a generic (scalar) covariate w_i that can be a binary or approximately continuous variable. We obtain the marginal effects of a generic covariate w_i on various probabilities assuming that the regression results are based on ZIBOP. If w_i is a binary regressor, then the marginal effect of w_i on probability, say P, is the difference in the probability evaluated at 1 and 0, conditional on observable values of covariates: $P(w_i = 1) - P(w_i = 0)$. For continuous explanatory variables, the marginal effect is given by the partial derivative of the probability of interest with respect to w_i, $\partial P(\cdot)/\partial w_i$.

Regressor w_i can be a common covariate in vectors of regressors $\mathbf{x}_{ri}$ and $\mathbf{z}_i$ or appears in either $\mathbf{x}_{ri}$ or $\mathbf{z}_i$. Focusing on the continuous regressor case, the marginal effects of w_i in each of the three cases are presented below. First, consider the case of common covariate in participation and main parts of the model, that is, w_i in both $\mathbf{x}_{ri}$ and $\mathbf{z}_i$. The marginal effect on the probability of participation is given by

$$M_i(s_i = 1) = \frac{\partial \Pr(s_i = 1)}{\partial w_i} = \phi\left(\mathbf{z}_i'\boldsymbol{\gamma}\right)\gamma_{w_i}, \tag{2.10}$$

where again $\phi(\cdot)$ is the probability density function (pdf) of the standard normal distribution and γ_{w_i} is the coefficient in the inflation part associated with variable w_i. In terms of the zeros category, the effect on the probability of nonparticipation (zero inflation) is

$$M_i(s_i = 0) = \frac{\partial \Pr(s_i = 0)}{\partial w_i} = -\phi\left(-\mathbf{z}_i'\boldsymbol{\gamma}\right)\gamma_{w_i}, \tag{2.11}$$

while

$$M_i(s=1,\tilde{y}_{ri}=0) = \frac{\partial \Pr(s_i=1)\Pr(\tilde{y}_{ri}=0)}{\partial w_i} = \Phi(-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)\phi(\mathbf{z}_i'\boldsymbol{\gamma})\gamma_{w_i} - \Phi(\mathbf{z}_i'\boldsymbol{\gamma})\phi(-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)\beta_{rw_i}, \quad r=1,2, \tag{2.12}$$

represents the marginal effect on the probability of zero-consumption. Here the scalar β_{rw_i} is the coefficient in the main part of the model associated with w_i.

Continuing with the case of common covariate, the marginal effects of w_i on the probabilities of choice are given as follows. First, the total marginal effect on the probability of observing zero-consumption is obtained as a sum of the marginal effects in (2.11) and (2.12); that is,

$$M_i(y_{ri}=0) = \left[\Phi(-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)-1\right]\phi(\mathbf{z}_i'\boldsymbol{\gamma})\gamma_{w_i} - \Phi(\mathbf{z}_i'\boldsymbol{\gamma})\phi(-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)\beta_{rw_i}. \tag{2.13}$$

The effects for the remaining choices for outcomes $r=1,2$ are as follows:

$$\begin{aligned}
M_i(y_{ri}=1) &= \left[\Phi(\alpha_{r2}-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)-\Phi(-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)\right]\phi(\mathbf{z}_i'\boldsymbol{\gamma})\gamma_{w_i} \\
&\quad - \Phi(\mathbf{z}_i'\boldsymbol{\gamma})\left[\phi(\alpha_{r2}-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)-\phi(-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)\right]\beta_{rw_i}; \\
M_i(y_{ri}=j) &= \left[\Phi(\alpha_{r,j+1}-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)-\Phi(\alpha_{rj}-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)\right]\phi(\mathbf{z}_i'\boldsymbol{\gamma})\gamma_{w_i} \\
&\quad - \Phi(\mathbf{z}_i'\boldsymbol{\gamma})\left[\phi(\alpha_{r,j+1}-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)-\phi(\alpha_{rj}-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)\right]\beta_{rw_i}, \quad \text{for } j=2,\ldots,J_r-1; \\
M_i(y_{ri}=J_r) &= \left[1-\Phi(\alpha_{r,J_r}-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)\right]\phi(\mathbf{z}_i'\boldsymbol{\gamma})\gamma_{w_i} + \Phi(\mathbf{z}_i'\boldsymbol{\gamma})\phi(\alpha_{r,J_r}-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)\beta_{rw_i}.
\end{aligned} \tag{2.14}$$

Now consider case 2, where a generic independent variable w_i is included only in $\mathbf{x}_{ri}$, the main part of the model. In this case, covariate w_i has obviously no direct effect on the inflation part. The marginal effects of w_i on various choice probabilities can be presented as follows:

$$M_i(y_{ri}=j) = \frac{\partial \Pr(y_{ri}=j)}{\partial w_i} = -\Phi(\mathbf{z}_i'\boldsymbol{\gamma})\left[\phi(\alpha_{r,j+1}-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)-\phi(\alpha_{rj}-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)\right]\beta_{rw_i}, \quad \text{for } j=0,1,\ldots,J_r, \tag{2.15}$$

with $\alpha_{r0}=-\infty$, $\alpha_{r1}=0$, and $\alpha_{r,J_r+1}=\infty$. The marginal effects in (2.15) can be obtained by simply setting $\gamma_{w_i}=0$ in (2.13) and (2.14).

For case 3, where w_i appears only in $\mathbf{z}_i$, its marginal effects on participation components given in (2.10) and (2.11) will not change. Since $\beta_{rw_i}=0$ in case 3, the partial effects of w_i on various choice probabilities take the form:

$$M_i(y_{ri}=j) = \left[\Phi(\alpha_{r,j+1}-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)-\Phi(\alpha_{rj}-\mathbf{x}_{ri}'\boldsymbol{\beta}_r)\right]\phi(\mathbf{z}_i'\boldsymbol{\gamma})\gamma_{w_i} \quad \text{for } j=0,1,\ldots,J_r. \tag{2.16}$$

Again, we impose the restrictions $\alpha_{r0}=-\infty$, $\alpha_{r1}=0$ and $\alpha_{r,J_r+1}=\infty$.

As noted by a referee, it is important to understand the sources of covariate effects and the relationship between the marginal effects and the coefficient estimates. Since

$$\Pr(y_{ri} = j) = \left[\Pr(s_i = 1)\Pr(\widetilde{y}_{ri} = j)\right] \tag{2.17}$$

for $j = 0, 1, \ldots, J_r$, the total effect of a generic covariate w_i on probability of consumption at level j comes from two (weighted) sources: the participation part ($\Pr(s_i = 1)$) and the main ordered probit part ($\Pr(\widetilde{y}_{ri} = j)$) such that

$$\frac{\partial \Pr(s_i = 1)}{\partial w_i} = \phi(\mathbf{z}_i'\boldsymbol{\gamma})\gamma_{w_i}; \tag{2.18}$$

$$\frac{\partial \Pr(\widetilde{y}_{ri} = j)}{\partial w_i} = -\left[\phi(\alpha_{r,j+1} - \mathbf{x}_{ri}'\boldsymbol{\beta}_r) - \phi(\alpha_{rj} - \mathbf{x}_{ri}'\boldsymbol{\beta}_r)\right]\beta_{rw_i} \tag{2.19}$$

with $\alpha_{r0} = -\infty$, $\alpha_{r1} = 0$, s and $\alpha_{r,J_r+1} = \infty$. This shows that $\text{sign}(\gamma_{w_i})$ is the same as $\text{sign}(\partial \Pr(s_i = 1)/\partial w_i)$—the participation effect in (2.18)—but $\text{sign}(\beta_{rw_i})$ is not necessarily the same as the sign of $(\partial \Pr(\widetilde{y}_{ri} = j)/\partial w_i)$. The latter is particularly true in the left tail of the distribution, where the coefficient (β_{rw_i}) and the main (unweighted) effect in (2.19) have opposite signs because

$$\left\{-\left[\phi(\alpha_{r,j+1} - \mathbf{x}_{ri}'\boldsymbol{\beta}_r) - \phi(\alpha_{rj} - \mathbf{x}_{ri}'\boldsymbol{\beta}_r)\right]\right\} \equiv \varpi \tag{2.20}$$

is negative. In this case, a positive effect coming from the main part requires β_{rw_i} to be negative. By contrast, ϖ is positive in the right tail, but can be positive or negative when the terms $(\alpha_{r,j} - \mathbf{x}_{ri}'\boldsymbol{\beta}_r)$ and $(\alpha_{r,j+1} - \mathbf{x}_{ri}'\boldsymbol{\beta}_r)$ are on the opposite sides of the mode of the distribution. This shows that a given covariate can have opposite effects in the participation and main models. Since the total effect of an explanatory variable on probability of choice is a weighted average of (2.18) and (2.19), interpretation of results should focus on marginal effects of covariates rather than the signs of estimated coefficients. This is the strategy adopted in the empirical analysis below.

2.3. *A Special Case*

Since the zero-inflated univariate ordered probit (ZIOP) model has not been analyzed previously in the Bayesian framework, we provide a brief sketch of the basic framework for ZIOP. The univariate ordered probit model with excess of zeros can be obtained as a special case of the ZIBOP model presented previously. To achieve this, let $\rho_{12} = 0$ in the ZIBOP model and focus on the first ordered outcome with $r = 1$. In the standard ordered response approach, the model for the latent variable $\widetilde{y}_{1i}^*$ is given by (2.1) with $r = 1$. The observed ordered variable $\widetilde{y}_{1i}$ can be presented compactly as

$$\widetilde{y}_{1i} = \sum_{j=0}^{J} jI\left(\alpha_{1j} < \widetilde{y}_{1i}^* \leq \alpha_{1j+1}\right), \tag{2.21}$$

where $I(w \in A)$ is the indicator function equal to 1 or 0 according to whether $w \in A$ or not. Again $\alpha_{10}, \alpha_{11}, \ldots, \alpha_{1J_1}$ are unknown threshold parameters, where we set $\alpha_{10} = -\infty$, $\alpha_{11} = 0$, and $\alpha_{1J_1+1} = \infty$.

Zero-inflation is now introduced at point $\tilde{y}_{1i} = 0$. Using the latent variable model (2.4) for the zero inflation, the observed binary variable is given by $s_i = I(s_i^* > 0)$, where $I(s_i^* > 0) = 1$ if $s_i^* > 0$, and 0 otherwise. In regime 1, $s_i = 1$ or $s_i^* > 0$ for participants (e.g., smokers), while, in regime 0, $s_i = 0$ or $s_i^* \leq 0$ for nonparticipants. In the context of the zero-inflation model, the observed response variable takes the form $y_{1i} = s_i \tilde{y}_{1i}$. We observe $y_{1i} = 0$ when either the individual is a non-participant ($s_i = 0$) or the individual is a zero-consumption participant ($s_i = 1$ and $\tilde{y}_{1i} = 0$). Likewise, we observe positive outcome (consumption) when the individual is a positive consumption participant ($s_i = 1$ and $\tilde{y}_{1i}^* > 0$).

Assume that ϵ_1 and μ are independently distributed. Harris and Zhao [15] also consider the case where ϵ_1 and μ are correlated. In the context of our application, the correlated model did not provide improvements over the uncorrelated ZIOP in terms of deviance information criterion. The zero-inflated ordered multinomial distribution, say $\Pr(y_{1i})$, arises as a mixture of a degenerate distribution at zero and the assumed distribution of the response variable $\tilde{y}_{1i}$ as follows:

$$f_1\left(y_{1i}^*, y_{1i}, s_i^*, s_i \mid \mathbf{x}_{1i}, \mathbf{z}_i, \Psi_1\right) = \begin{cases} \Pr(s_i = 0) + \Pr(s_i = 1)\Pr\left(\tilde{y}_{1i} = 0\right), & \text{for } j = 0 \\ \Pr(s_i = 1)\Pr\left(\tilde{y}_{1i} = j\right), & \text{for } j = 1, 2, \ldots, J_1, \end{cases} \tag{2.22}$$

where, for any parameter vector Ω_{10} associated with the distribution of (ϵ_1, μ), $\Psi_1 = (\boldsymbol{\beta}_1, \boldsymbol{\gamma}, \boldsymbol{\alpha}^1, \Omega_{10})$ with $\boldsymbol{\alpha}^1 = (\alpha_{12}, \ldots, \alpha_{1J_1})$. For simplicity, dependence on latent variables, covariates, and parameters has been suppressed on the right-hand side of (2.22). The likelihood based on N-independent observations takes the form

$$\begin{aligned} \mathcal{L}_1\left(y_1^*, y_1, s^*, s \mid \mathbf{x}_1, \mathbf{z}, \Psi_1\right) &= \prod_{i=1}^{N} \prod_{j=0}^{J_1} \left[\Pr\left(y_{1i} = j \mid \mathbf{x}_{1i}, \mathbf{z}_i, \Psi_1\right)\right]^{d_{ij}} \\ &= \prod_{i=1}^{N} \prod_{j=0} \left[\Pr(s_i = 0) + \Pr(s_i = 1)\Pr\left(\tilde{y}_{1i} = j\right)\right]^{d_{ij}} \\ &\quad \times \prod_{i=1}^{N} \prod_{j>0} \left[\Pr(s_i = 1)\Pr\left(\tilde{y}_{1i} = j\right)\right]^{d_{ij}}, \end{aligned} \tag{2.23}$$

where, for example, $y_1^* = (y_1^*, \ldots, y_N^*)'$, and $d_{ij} = 1$ if individual i chooses outcome j, or $d_{ij} = 0$ otherwise.

Different choices of the specification of the joint distribution of (ϵ_{1i}, μ_i) give rise to various zero-inflated ordered response models. For example, if the disturbance terms in the latent variable equations are normally distributed, we get the zero-inflated ordered probit model of Harris and Zhao [15]. The zero-inflated ordered logit model can be obtained by assuming that ϵ_{1i} and μ_i are independent, each of the random variables following the logistic distribution with cumulative distribution function defined as $\Lambda(a) = e^a/(1 + e^a)$. Unlike the ordered probit framework, the ordered logit cannot lend itself easily to allow for correlation

between bivariate discrete response outcomes. Henceforth, we focus on the ordered probit paradigm in both univariate and bivariate settings.

Assuming that ϵ_{1i} and μ_i are independently normally distributed, each with mean 0 and variance 1, the required components in (2.22) and consequently (2.23) are given by:

$$\begin{aligned}
\Pr(s_i = 0) &= \Phi\left(-\mathbf{z}_i'\boldsymbol{\gamma}\right),\\
\Pr\left(\widetilde{y}_{1i} = 0\right) &= \Phi\left(-x_{1i}'\boldsymbol{\beta}_1\right),\\
\Pr\left(\widetilde{y}_{1i} = j\right) &= \Phi\left(\alpha_{1j+1} - \mathbf{x}_{1i}'\boldsymbol{\beta}_1\right) - \Phi\left(\alpha_{1j} - \mathbf{x}_{1i}'\boldsymbol{\beta}_1\right), \quad \text{for } j = 1,\ldots,J_1 - 1 \text{ with } \alpha_{10} = 0,\\
\Pr\left(\widetilde{y}_{1i} = J_1\right) &= 1 - \Phi\left(\alpha_{1J_1} - \mathbf{x}_{1i}'\boldsymbol{\beta}_1\right).
\end{aligned} \tag{2.24}$$

The marginal effects for the univariate ZIOP are given by Harris and Zhao [15]. Bayesian analysis of the univariate ZIOP will be treated as a special case of the zero-inflated bivariate order probit model in the next section.

3. Bayesian Analysis

3.1. Prior Distributions

The Bayesian hierarchical model requires prior distributions for each parameter in the model. For this purpose, we can use noninformative conjugate priors. There are two reasons for adopting noninformative conjugate priors. First, we prefer to let the data dictate the inference about the parameters with little or no influence from prior distributions. Secondly, the noninformative priors facilitate resampling using Markov Chain Monte Carlo algorithm (MCMC) and have nice convergence properties. We assume noninformative (vague or diffuse) normal priors for regression coefficients β, with mean β^* and variance Ω_β which are chosen to make the distribution proper but diffuse with large variances. Similarly, $\gamma \sim N(\gamma^*, \Omega_\gamma)$.

In choosing prior distributions for the threshold parameters, α's, caution is needed because of the order restriction on them. One way to avoid the order restriction is to reparameterize them. Following Chib and Hamilton [18] treatment in the univariate ordered probit case, we reparameterize the ordered threshold parameters

$$\tau_{r2} = \log(\alpha_{r2}); \quad \tau_{rj} = \log\left(\alpha_{rj} - \alpha_{rj-1}\right), \quad j = 3,\ldots,J_r;\ r = 1,2 \tag{3.1}$$

with the inverse map

$$\alpha_{rj} = \sum_{m=2}^{j} \exp(\tau_{rm}), \quad j = 2,\ldots,J_r;\ r = 1,2. \tag{3.2}$$

For $r = 1,2$, let $\boldsymbol{\tau}^r = (\tau_{r2}, \tau_{r3}, \ldots, \tau_{rJ})'$ so that $\boldsymbol{\tau} = (\boldsymbol{\tau}^1, \boldsymbol{\tau}^2)$. We choose normal prior $\boldsymbol{\tau} \sim N(\boldsymbol{\tau}^*, \boldsymbol{\Omega}_\tau)$ without order restrictions for τ_r's.

The only unknown parameter associate with the distribution of $(\boldsymbol{\epsilon}, \mu)$ in (2.1) and (2.4) is ρ_{12}, the correlation between ϵ_1 and ϵ_2. The values of ρ_{12} by definition are restricted to be in

the −1 to 1 interval. Therefore, the choice for prior distribution for ρ_{12} can be *uniform* $(-1,1)$ or a proper distribution based on reparameterization. Let ν denote the hyperbolic arc-tangent transformation of ρ_{12}, that is,

$$\nu = a\tanh(\rho_{12}), \tag{3.3}$$

and taking hyperbolic tangent transformation of ν gives us back $\rho_{12} = \tanh(\nu)$. Then parameter ν is asymptotically normal distributed with stabilized variance, $1/(N-3)$, where N is the sample size. We may also assume that $\nu \sim N(\nu^*, \sigma_\nu^2)$.

3.2. Bayesian Analysis via MCMC

For carrying out a Bayesian inference, the joint posterior distribution of the parameters of the ZIBOP model in (2.6) conditional on the data is obtained by combining the likelihood function given in (2.9) and the above-specified prior distributions via Bayes' theorem, as:

$$\begin{aligned} f(\Psi_b \mid \mathbf{x},\mathbf{z}) \propto & \prod_{i=1}^{N} \prod_{(j,l)=(0,0)} \left[\Phi(-\mathbf{z}_i'\boldsymbol{\gamma}) + \Phi(\mathbf{z}_i'\boldsymbol{\gamma})\Phi_2(-\mathbf{x}_{1i}'\beta_1, -\mathbf{x}_{2i}'\beta_2, \rho_{12})\right]^{d_{ijl}} \\ & \times \prod_{i=1}^{N} \prod_{(j,l)\neq(0,0)} \begin{bmatrix} \Phi(\mathbf{z}_i'\boldsymbol{\gamma}) \left[\Phi_2(\alpha_{1j+1} - \mathbf{x}_{1i}'\beta_1, \alpha_{2l+1} - \mathbf{x}_{2i}'\beta_2; \rho_{12}) \right. \\ \left. -\Phi_2(\alpha_{1j} - \mathbf{x}_{1i}'\beta_1, \alpha_{2l} - \mathbf{x}_{2i}'\beta_2, \rho_{12})\right] \end{bmatrix}^{d_{ijl}} \\ & \times f(\boldsymbol{\Psi}_b), \end{aligned} \tag{3.4}$$

where $f(\boldsymbol{\Psi}_b) \propto f(\boldsymbol{\beta})f(\boldsymbol{\gamma})f(\boldsymbol{\tau})f(\nu)$ and the parameter vector Ψ_b now consists of $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$, $\boldsymbol{\gamma}$, $\boldsymbol{\tau} = (\boldsymbol{\tau}^1, \boldsymbol{\tau}^2)$, s and $\nu = a\tanh(\rho_{12})$. Here $f(\boldsymbol{\beta}) \propto |\Omega_\beta|^{-1/2}\exp\{-1/2(\boldsymbol{\beta} - \boldsymbol{\beta}^*)'\Omega_\beta^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}^*)\}$; $f(\boldsymbol{\gamma}) \propto |\Omega_\gamma|^{-1/2}\exp\{-1/2(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*)'\Omega_\gamma^{-1}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*)\}$; $f(\boldsymbol{\tau}) \propto |\Omega_\tau|^{-1/2}\exp\{-1/2(\boldsymbol{\tau} - \boldsymbol{\tau}^*)'\Omega_\tau^{-1}(\boldsymbol{\tau} - \boldsymbol{\tau}^*)\}$; τ_{rj} are defined in (3.1), and α_{rj} are given via the inverse map (3.2).

Full conditional posterior distributions are required to implement the MCMC algorithm [19–22], and they are given as follows:

(1) fixed effects:

(a) zero state:

$$f(\boldsymbol{\gamma} \mid \mathbf{x},\mathbf{z},\Psi_{-\gamma}) \propto \left|\Omega_\gamma\right|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*)'\Omega_\gamma^{-1}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*)\right\} \times f(\Psi_b \mid \mathbf{x},\mathbf{z}); \tag{3.5}$$

(b) nonzero state:

$$f(\boldsymbol{\beta} \mid \mathbf{x},\mathbf{z},\Psi_{-\beta}) \propto \left|\Omega_\beta\right|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}^*)'\Omega_\beta^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}^*)\right\} \times f(\Psi_b \mid \mathbf{x},\mathbf{z}). \tag{3.6}$$

(2) thresholds:

$$f(\boldsymbol{\tau} \mid \mathbf{x}, \mathbf{z}, \Psi_{-\tau}) \propto |\Omega_\tau|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\tau}-\boldsymbol{\tau}^*)'\Omega_\tau^{-1}(\boldsymbol{\tau}-\boldsymbol{\tau}^*)\right\} \times \prod_{i=1}^{N} \prod_{(j,l)\neq(0,0)} \begin{bmatrix} \Phi(\mathbf{z}_i'\boldsymbol{\gamma}) \left[\Phi_2\left(\alpha_{1j+1} - \mathbf{x}_{1i}'\beta_1, \alpha_{2l+1} - \mathbf{x}_{2i}'\beta_2; \rho_{12}\right)\right. \\ \left. -\Phi_2\left(\alpha_{1j} - \mathbf{x}_{1i}'\beta_1, \alpha_{2l} - \mathbf{x}_{2i}'\beta_2, \rho_{12}\right)\right] \end{bmatrix}^{d_{ijl}}. \tag{3.7}$$

(3) bivariate correlation:

$$f(\nu \mid \mathbf{x}, \mathbf{z}, \Psi_{-\nu}) \propto \sigma_\nu^{-1} \exp\left\{-\frac{(\nu-\nu^*)^2}{2\sigma_\nu^2}\right\} \times f(\Psi_b \mid \mathbf{x}, \mathbf{z}). \tag{3.8}$$

The MCMC algorithm simulates direct draws from the above full conditionals iteratively until convergence is achieved. A single long chain [23, 24] is used for the proposed model. Geyer [23] argues that using a single longer chain is better than using a number of smaller chains with different initial values. We follow this strategy in our empirical analysis.

The Bayesian analysis of the univariate ZIOP follows as a special case of that of the ZIBOP presented above. In particular, the joint posterior distribution of the parameters of the ZIOP model in (2.22) conditional on the data is obtained by combining the likelihood function given in (2.23) and the above-specified prior distributions (with modified notations) via Bayes' theorem, as follows:

$$\begin{aligned} f(\Psi \mid \mathbf{x}, \mathbf{z},) \propto & \prod_{i=1}^{N} \prod_{j=0} \left[\Phi(-\mathbf{z}_i'\boldsymbol{\gamma}) + \Phi(\mathbf{z}_i'\boldsymbol{\gamma})\Phi(-\mathbf{x}_i'\boldsymbol{\beta})\right]^{d_{ij}} \\ & \times \prod_{i=1}^{N} \prod_{j>0} \left[\Phi(\mathbf{z}_i'\boldsymbol{\gamma})\left\{\Phi(\alpha_{j+1} - \mathbf{x}_i'\boldsymbol{\beta}) - \Phi(\alpha_j - \mathbf{x}_i'\boldsymbol{\beta})\right\}\right]^{d_{ij}} \\ & \times f(\boldsymbol{\beta}) f(\boldsymbol{\gamma}) f(\boldsymbol{\tau}), \end{aligned} \tag{3.9}$$

where, using notation of Section 2.3 for $\boldsymbol{\beta}$ and the other parameter vectors, $f(\boldsymbol{\beta}) \propto |\Omega_\beta|^{-1/2} \exp\{-1/2(\boldsymbol{\beta}-\boldsymbol{\beta}^*)'\Omega_\beta^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\}$; $f(\boldsymbol{\gamma}) \propto |\Omega_\gamma|^{-1/2} \exp\{-1/2(\boldsymbol{\gamma}-\boldsymbol{\gamma}^*)'\Omega_\gamma^{-1}(\boldsymbol{\gamma}-\boldsymbol{\gamma}^*)\}$; $f(\boldsymbol{\tau}) \propto |\Omega_\tau|^{-1/2} \exp\{-1/2(\boldsymbol{\tau}-\boldsymbol{\tau}^*)'\Omega_\tau^{-1}(\boldsymbol{\tau}-\boldsymbol{\tau}^*)\}$, $\tau_2 = \log(\alpha_2)$ and $\tau_j = \log(\alpha_j - \alpha_{j-1})$, $j = 3, \ldots, J$. Apart from dropping the bivariate correlation, we basically replace the bivariate normal cumulative distribution $\Phi_2(\cdot, \cdot; \rho_{12})$ by the univariate counterpart $\Phi(\cdot)$. Details are available upon request from the authors.

Apart from Bayesian estimation of the regression parameters, the posterior distributions of other quantities of interest can be obtained. These include posteriors for marginal effects and probabilities for nonparticipation, zero-consumption, and joint outcomes of interest. These will be considered in the application section. Next, we summarize model selection procedure.

The commonly used criteria for model selection like BIC and AIC are not appropriate for the multilevel models (in the presence of random effects), which complicates the counting

of the true number of free parameters. To overcome such a hurdle, Spiegelhalter et al. [25] proposed a Bayesian model comparison criterion, called Deviance Information Criterion (DIC). It is given as

$$\text{DIC} = \text{goodness-of-fit} + \text{penalty for complexity}, \tag{3.10}$$

where the "goodness-of-fit" is measured by the deviance for $\theta = (\beta, \gamma, \alpha)$

$$D(\theta) = -2 \log \mathcal{L}(\text{data} \mid \theta) \tag{3.11}$$

and complexity is measured by the "effective number of parameters":

$$\begin{aligned} pD &= E_{\theta|y}[D(\theta)] - D\left(E_{\theta|y}[\theta]\right) \\ &= \overline{D} - D\left(\overline{\theta}\right); \end{aligned} \tag{3.12}$$

that is, posterior mean deviance minus deviance evaluated at the posterior mean of the parameters. The DIC is then defined analogously to AIC as

$$\begin{aligned} \text{DIC} &= D\left(\overline{\theta}\right) + 2pD \\ &= \overline{D} + pD. \end{aligned} \tag{3.13}$$

The idea here is that models with smaller DIC should be preferred to models with larger DIC. Models are penalized both by the value of $\overline{D}$, which favors a good fit, but also (similar to AIC and BIC) by the effective number of parameters pD. The advantage of DIC over other criteria, for Bayesian model selection, is that the DIC is easily calculated from the MCMC samples. In contrast, AIC and BIC require calculating the likelihood at its maximum values, which are not easily available from the MCMC simulation.

4. Application

4.1. Data

We consider an application to tobacco consumption behavior of individuals based on the 2001 household Tobacco Prevalence survey data from Bangladesh. The Survey was conducted in two administrative districts of paramount interest for tobacco production and consumption in the country. Data on daily consumption of smoking and chewing tobacco along with other socioeconomic and demographic characteristics and parental tobacco consumption habits were collected from respondents of 10 years of age and above. The data set has been used previously by Gurmu and Yunus [26] in the context of binary response models. Here we focus on a sample consisting of 6000 individual respondents aged between 10 and 101 years.

The ordinal outcomes $y_r = 0, 1, 2, 3$ used in this paper correspond roughly to zero, low, moderate, and high levels of tobacco consumption in the form of smoking (y_1) or chewing tobacco (y_2), respectively. The first dependent variable y_1 for an individual's daily

Table 1: Bivariate frequency distribution for intensity of tobacco use.

Smoke group	Chew group			Total (N)
	0	1	2	
0	3931	302	324	4557
1	265	12	6	283
2	526	35	37	598
3	498	29	35	562
Total (N)	5220	378	402	6000

cigarette smoking intensities assumes the following 4 choices: $y_1 = 0$ if nonsmoker, $y_1 = 1$ if smoking up to 7 cigarettes per day, $y = 2$ if smoking between 8 and 12 cigarettes daily, and $y_1 = 3$ if smoking more than 12 cigarettes daily; likewise, for the intensity of chewing tobacco, $y_2 = 0$ if reported not chewing tobacco, $y_2 = 1$ if uses up to 7 chewing tobacco, and $y_2 = 2$ if consuming 7 or more chewing tobacco. The frequency distribution of cigarette smoking and tobacco chewing choices in Table 1 shows that nearly 66% of the respondents identify themselves as nonusers of tobacco. Our modeling strategy recognizes that these self-identified current nonusers of tobacco may include either individuals who never smoke or chew tobacco (genuine nonusers) or those who do, but not during the reporting period (potential users of tobacco). For example, potential tobacco users may include those who wrongly claim to be nonusers, previous tobacco users that are currently nonusers, and those most likely to use tobacco in the future due to changes in, say, prices and income. Table 1 also shows that 76% of the respondents are non-smokers and nearly 87% identify themselves as nonusers of tobacco for chewing. Given the extremely high proportion of observed zeros coupled with sparse cells on the right tail, we employ the zero-inflated bivariate ordered probit framework.

Table 2 gives definition of the explanatory variables as well as their means and standard deviations. The respondents are more likely to be Muslim, married, in early thirties, live in rural area, and have about 7 years of formal schooling. Although the country is mostly agrarian, only around 11% of the respondents were related to agricultural occupation in either doing agricultural operations on their own farms or working as agricultural wage laborers. About 12% of the respondents belong to the service occupation. The benchmark occupational group consists of business and other occupations. More than one-half of the fathers and slightly less than two-thirds of the mothers of the respondents currently use or have used tobacco products in the past.

Among the variables given in Table 2, the two indicators of parental use of tobacco products are included in $\mathbf{z}$ as part of the participation equation (2.4). The rest of the variables are included in $\mathbf{x}_r$ and $\mathbf{z}$ of (2.1) and (2.4). To allow for nonlinear effects, age and education enter all three equations using a quadratic form. Due to lack of data on prices, our analysis is limited to the study of other economic and demographic determinants of participation, smoking, and chewing tobacco.

4.2. Results

We estimate the standard bivariate ordered probit (BOP) and zero-inflated bivariate ordered probit regression models for smoking and chewing tobacco and report estimation results for parameters, marginal effects, and choice probabilities, along measures of model selection.

Table 2: Definition and summary statistics for independent variables.

Name	Definition	Mean[b]	St. Dev.
Age[a]	Age in years	30.35	(14.9)
Education[a]	Number of years of formal schooling	6.83	(4.7)
Income	Monthly family income in 1000s of Taka	7.57	(10.3)
Male	= 1 if male	54.6	
Married	= 1 if married	57.2	
Muslim	= 1 if religion is Islam	78.4	
Father use	= 1 if father uses tobacco	54.0	
Mother use	= 1 if mother uses tobacco	65.1	
Region	= 1 if Rangpur resident, = 0 if Chittagong resident	49.7	
Urban	= 1 if urban resident	38.0	
Agriservice	= 1 if agriculture labor or service occupation	23.2	
Self-employed	= 1 if self-employed or household chores	30.7	
Student	= 1 if student	26.8	
Other	= 1 if business or other occupations (control)	19.3	

[a] In implementation, we also include age squared and education squared.
[b]The means for binary variables are in percentage.

Table 3: Goodness-of-fit statistics via DIC.

Model	Dbar	Dhat	pD	DIC
Bivariate ordered probit (BOP)	11417.1	11386.9	30.1	11447.2
Zero-inflated BOP	11301.1	11270.3	29.8	11329.9

Dbar: Posterior mean of deviance, Dhat: Deviance evaluated at the posterior mean of the parameters, pD: Dbar-Dhat, the effective number of parameters, and DIC: Deviance information criterion.

An earlier version of this paper reports results from the standard ordered probit model as well as the uncorrelated and correlated versions of the univariate zero-inflated ordered probit model for smoking tobacco. Convergence of the generated samples is assessed using standard tools (such as trace plots and ACF plots) within WinBUGS software. After initial 10,000 burn-in iterations, every 10th MCMC sample thereafter was retained from the next 100,000 iterations, obtaining 10,000 samples for subsequent posterior inference of the unknown parameters. The slowest convergence is observed for some parameters in the inflation submodel. By contrast, the autocorrelations functions for most of the marginal effects die out quickly relative to those for the associated parameters.

Table 3 reports the goodness-of-fit statistics for the standard bivariate ordered probit model and its zero-inflated version, ZIBOP. The ZIBOP regression model clearly dominates BOP in terms of DIC and its components; compare the DIC of 11330 for the former and 11447 for the latter model. Table 4 gives posterior means, standard deviations, medians, and the 95 percent credible intervals (in terms of the 2.5 and 97.5 percentiles) of the parameters and choice probabilities from ZIBOP model. For comparison, the corresponding results from BOP

Table 4: Posterior mean, standard deviation, and 95% credible intervals of parameters from zibop for smoking and chewing tobacco.

Variable	Mean	St. dev.	2.50%	Median	97.50%
Main (β_1,α_1): smoking (y_1):					
Age/10	0.672	0.119	0.444	0.685	0.894
Age square/100	−0.070	0.012	−0.093	−0.071	−0.046
Education	−0.071	0.014	−0.097	−0.071	−0.042
Education square	0.001	0.001	−0.002	0.001	0.003
Income	0.000	0.002	−0.005	0.000	0.005
Male	2.092	0.086	1.925	2.091	2.269
Married	0.213	0.070	0.074	0.213	0.353
Muslim	−0.053	0.052	−0.157	−0.053	0.049
Region	−0.007	0.048	−0.102	−0.007	0.086
Urban	−0.096	0.051	−0.198	−0.097	0.004
Agriservice	−0.234	0.056	−0.345	−0.233	−0.125
Self-employed	−0.246	0.087	−0.414	−0.247	−0.069
student	−0.476	0.137	−0.742	−0.478	−0.204
α_{12}	0.284	0.017	0.252	0.283	0.318
α_{13}	0.987	0.030	0.928	0.987	1.048
Main (β_2,α_2): chewing (y_2)					
Age/10	0.649	0.133	0.382	0.658	0.893
Age square/100	−0.046	0.013	−0.071	−0.046	−0.019
Education	−0.020	0.016	−0.052	−0.020	0.012
Education square	−0.002	0.001	−0.005	−0.002	0.000
Income	0.001	0.003	−0.004	0.002	0.007
Male	−0.479	0.081	−0.641	−0.479	−0.320
Married	−0.025	0.075	−0.171	−0.025	0.122
Muslim	−0.072	0.056	−0.181	−0.072	0.039
Region	0.417	0.051	0.317	0.418	0.517
Urban	−0.080	0.058	−0.194	−0.079	0.035
Agriservice	0.052	0.074	−0.096	0.052	0.194
Self-employed	0.127	0.092	−0.058	0.126	0.309
Student	−0.450	0.221	−0.887	−0.448	−0.023
α_{22}	0.484	0.023	0.439	0.484	0.531
Inflation (γ):					
Age/10	−0.012	2.044	−4.755	0.253	2.861
Age square/100	0.509	0.552	−0.197	0.398	1.812
Education	−0.218	0.115	−0.476	−0.204	−0.024
Education square	0.028	0.011	0.010	0.026	0.053
Income	0.006	0.022	−0.027	0.003	0.059
Male	0.239	0.827	−1.582	0.417	1.379
Married	2.306	4.478	−0.416	0.500	16.900
Muslim	−0.528	0.356	−1.331	−0.494	0.068
Mother	−0.170	0.267	−0.716	−0.164	0.345
Father	−0.119	0.330	−0.664	−0.160	0.605
Region	0.630	0.291	0.061	0.625	1.222

Table 4: Continued.

Variable	Mean	St. dev.	2.50%	Median	97.50%
Urban	0.040	0.357	−0.737	0.071	0.675
Agrservice	5.312	5.416	1.017	2.674	20.470
Self-employed	3.783	5.025	0.124	1.275	17.990
Sstudent	−0.344	0.411	−1.154	−0.339	0.466
ρ_{12}	−0.185	0.033	−0.249	−0.186	−0.119
Select probabilities:					
$P\ (y_1 = 0)$	0.760	0.004	0.752	0.760	0.768
$P\ (y_2 = 0)$	0.871	0.004	0.864	0.871	0.879
$P\ (y_1 = 0, y_2 = 0)$	0.662	0.005	0.652	0.662	0.671
P (zero-inflation)	0.242	0.048	0.151	0.243	0.323

Results for the constant terms in the main and inflation parts have been suppressed for brevity.

are shown in Table 6 of the appendix. Both models predict significant negative correlation between the likelihood of smoking and chewing tobacco. The posterior estimates of the cut-off points are qualitatively similar across models. In what follows, we focus on discussion of results from the preferred ZIBOP model. The 95% credible interval for the correlation parameter ρ_{12} from the zero-inflated model is in the range −0.25 to −0.12, indicating that smoking and chewing tobacco are generally substitutes. Results of selected predicted choice probabilities (bottom of Table 4) show that the ZIBOP regression model provides very good fit to the data. The posterior mean for the probability of (zero, zero)-inflation is about 24% while the 95% credible interval is [0.15, 0.32], indicating that a substantial proportion of zeros may be attributed to nonparticipants. These results underscore the importance of modeling excess zeros in bivariate ordered probit models.

To facilitate interpretation of results, we report in Tables 5 and 7 the same set of posterior estimates for the marginal effects from ZIBOP and BOP models, respectively. Since age and education enter the three equations non-linearly, we report the total marginal effects coming from the linear and quadratic parts. We examine closely the marginal effects on the unconditional marginal probabilities at all levels of smoking and chewing tobacco ($y_1 = 0, 1, 2, 3$; $y_2 = 0, 1, 2$). The marginal effects reported in Table 5 show that the results for covariates are generally plausible. Age has a negative impact on probabilities of moderate and heavy use of tobacco. For heavy smokers, education has a significant negative impact on the probability of smoking cigarettes. An additional year of schooling on average decreases probability of smoking by about 6.9% for heavy smokers. Among participants, being male or married has positive impact on probability of smoking, while the effects for being Muslim, urban resident, and student are largely negative. Male respondents are more likely to smoke cigarettes while women respondents are more likely to use chewing tobacco with heavy intensity, a result which is in line with custom of the country [26].

Using (2.13), we decompose the marginal effect on probability of observing zero-consumption into two components: the effect on nonparticipation (zero inflation) and zero-consumption. For each explanatory variable, this decomposition is shown in Table 5 in the first three rows for smoking and in rows 1, 7, and 8 for chewing tobacco. For most variables, the effects on probabilities of nonparticipation and zero-consumption are on average opposite in sign, but this difference seems to diminish at the upper tail of the distribution. For example, looking at the posterior mean for age under smoking, getting older by one more year

Table 5: Posterior mean, standard deviation, and 95% credible intervals of marginal effects of covariates on probability of smoking and chewing tobacco (ZIBOP model).

Variable	Probability	Mean	St. dev.	2.50%	Median	97.50%
Age	Nonparticipation	−0.0259	0.0129	−0.0556	−0.0236	−0.0078
	Zero-consumption, y_1	0.0463	0.0102	0.0294	0.0453	0.0687
	All zeros, $y_1 = 0$	0.0204	0.0059	0.0078	0.0213	0.0304
	$y_1 = 1$	0.0058	0.0035	0.0009	0.0053	0.0138
	$y_1 = 2$	−0.0014	0.0029	−0.0057	−0.0019	0.0055
	$y_1 = 3$	−0.0690	0.0235	−0.1223	−0.0658	−0.0344
	Zero-consumption, y_2	0.0403	0.0116	0.0195	0.0386	0.0675
	All zeros, $y_2 = 0$	0.0145	0.0064	0.0018	0.0149	0.0264
	$y_2 = 1$	−0.0034	0.0021	−0.0071	−0.0035	0.0008
	$y_2 = 2$	−0.0019	0.0014	−0.0043	−0.0020	0.0011
Education	Nonparticipation	−0.2823	0.0768	−0.4260	−0.2837	−0.1252
	Zero-consumption, y_1	0.2447	0.0749	0.0917	0.2459	0.3851
	All zeros, $y_1 = 0$	−0.0377	0.0241	−0.0853	−0.0374	0.0094
	$y_1 = 1$	0.0498	0.0141	0.0231	0.0494	0.0789
	$y_1 = 2$	0.0241	0.0102	0.0045	0.0239	0.0444
	$y_1 = 3$	−0.5557	0.1536	−0.8415	−0.5588	−0.2417
	Zero-consumption, y_2	0.3136	0.0772	0.1561	0.3159	0.4546
	All zeros, $y_2 = 0$	0.0313	0.0161	−0.0009	0.0315	0.0618
	$y_2 = 1$	−0.0134	0.0080	−0.0288	−0.0135	0.0027
	$y_2 = 2$	−0.0222	0.0119	−0.0455	−0.0221	0.0009
Income	Nonparticipation	−0.0004	0.0015	−0.0038	−0.0002	0.0022
	Zero-consumption, y_1	0.0003	0.0014	−0.0022	0.0002	0.0035
	All zeros, $y_1 = 0$	−0.0001	0.0004	−0.0009	−0.0001	0.0008
	$y_1 = 1$	0.0001	0.0003	−0.0005	0.0000	0.0008
	$y_1 = 2$	0.0000	0.0002	−0.0003	0.0000	0.0004
	$y_1 = 3$	−0.0007	0.0030	−0.0075	−0.0004	0.0044
	Zero-consumption, y_2	0.0001	0.0016	−0.0025	0.0000	0.0036
	All zeros, $y_2 = 0$	−0.0002	0.0005	−0.0011	−0.0002	0.0007
	$y_2 = 1$	0.0001	0.0002	−0.0003	0.0001	0.0004
	$y_2 = 2$	0.0001	0.0001	−0.0002	0.0001	0.0003
Male	Nonparticipation	−0.0254	0.0599	−0.1268	−0.0305	0.1012
	Zero-consumption, y_1	−0.3595	0.0611	−0.4900	−0.3540	−0.2565
	All zeros, $y_1 = 0$	−0.3849	0.0116	−0.4078	−0.3849	−0.3618
	$y_1 = 1$	0.0630	0.0040	0.0555	0.0630	0.0711
	$y_1 = 2$	0.1560	0.0065	0.1435	0.1559	0.1689
	$y_1 = 3$	0.1659	0.0083	0.1503	0.1657	0.1829
	Zero-consumption, y_2	0.1012	0.0623	−0.0309	0.1064	0.2075
	All zeros, $y_2 = 0$	0.0758	0.0126	0.0511	0.0759	0.1004
	$y_2 = 1$	0.0501	0.0033	0.0438	0.0500	0.0567
	$y_2 = 2$	−0.1258	0.0112	−0.1478	−0.1258	−0.1040
Married	Nonparticipation	−0.0680	0.0777	−0.2274	−0.0433	0.0346
	Zero-consumption, y_1	0.0200	0.0705	−0.0778	−0.0001	0.1692
	All zeros, $y_1 = 0$	−0.0480	0.0149	−0.0796	−0.0472	−0.0207

Table 5: Continued.

Variable	Probability	Mean	St. dev.	2.50%	Median	97.50%
	$y_1 = 1$	0.0056	0.0035	0.0006	0.0047	0.0132
	$y_1 = 2$	0.0161	0.0061	0.0060	0.0154	0.0296
	$y_1 = 3$	0.0263	0.0073	0.0116	0.0264	0.0406
	Zero-consumption, y_2	0.0709	0.0791	−0.0371	0.0474	0.2349
	All zeros, $y_2 = 0$	0.0028	0.0119	−0.0200	0.0026	0.0269
	$y_2 = 1$	0.0628	0.0032	0.0566	0.0627	0.0693
	$y_2 = 2$	−0.0656	0.0115	−0.0888	−0.0654	−0.0434
Muslim	Nonparticipation	0.0393	0.0243	−0.0050	0.0384	0.0900
	Zero-consumption, y_1	−0.0239	0.0247	−0.0752	−0.0231	0.0216
	All zeros, $y_1 = 0$	0.0154	0.0090	−0.0016	0.0153	0.0334
	$y_1 = 1$	−0.0023	0.0011	−0.0044	−0.0022	−0.0002
	$y_1 = 2$	−0.0053	0.0027	−0.0106	−0.0053	−0.0001
	$y_1 = 3$	−0.0078	0.0060	−0.0200	−0.0077	0.0036
	Zero-consumption, y_2	−0.0260	0.0258	−0.0797	−0.0253	0.0222
	All zeros, $y_2 = 0$	0.0133	0.0092	−0.0046	0.0133	0.0315
	$y_2 = 1$	0.0613	0.0030	0.0554	0.0613	0.0674
	$y_2 = 2$	−0.0746	0.0091	−0.0926	−0.0746	−0.0569
Father use	Nonparticipation	0.0122	0.0187	−0.0251	0.0124	0.0487
	Zero-consumption, y_1	−0.0102	0.0158	−0.0411	−0.0104	0.0214
	All zeros, $y_1 = 0$	0.0020	0.0030	−0.0040	0.0019	0.0082
	$y_1 = 1$	−0.0005	0.0008	−0.0022	−0.0005	0.0011
	$y_1 = 2$	−0.0009	0.0014	−0.0037	−0.0009	0.0018
	$y_1 = 3$	−0.0005	0.0008	−0.0023	−0.0005	0.0011
	Zero-consumption, y_2	−0.0116	0.0179	−0.0464	−0.0118	0.0240
	All zeros, $y_2 = 0$	0.0006	0.0011	−0.0012	0.0003	0.0033
	$y_2 = 1$	−0.0003	0.0006	−0.0019	−0.0002	0.0007
	$y_2 = 2$	−0.0002	0.0005	−0.0014	−0.0001	0.0005
Mother use	Nonparticipation	0.0129	0.0257	−0.0343	0.0123	0.0634
	Zero-consumption, y_1	−0.0106	0.0215	−0.0527	−0.0103	0.0298
	All zeros, $y_1 = 0$	0.0024	0.0043	−0.0047	0.0020	0.0115
	$y_1 = 1$	−0.0006	0.0012	−0.0031	−0.0006	0.0014
	$y_1 = 2$	−0.0011	0.0019	−0.0051	−0.0009	0.0022
	$y_1 = 3$	−0.0007	0.0012	−0.0033	−0.0005	0.0012
	Zero-consumption, y_2	−0.0119	0.0242	−0.0587	−0.0118	0.0338
	All zeros, $y_2 = 0$	0.0010	0.0016	−0.0007	0.0004	0.0053
	$y_2 = 1$	−0.0006	0.0009	−0.0030	−0.0002	0.0005
	$y_2 = 2$	−0.0004	0.0007	−0.0023	−0.0001	0.0002
Region	Nonparticipation	−0.0480	0.0240	−0.0963	−0.0470	−0.0040
	Zero-consumption, y_1	0.0412	0.0237	−0.0039	0.0406	0.0889
	All zeros, $y_1 = 0$	−0.0068	0.0079	−0.0222	−0.0068	0.0086
	$y_1 = 1$	0.0021	0.0011	0.0001	0.0021	0.0046
	$y_1 = 2$	0.0033	0.0025	−0.0016	0.0033	0.0083
	$y_1 = 3$	0.0013	0.0052	−0.0087	0.0014	0.0114

Table 5: Continued.

Variable	Probability	Mean	St. dev.	2.50%	Median	97.50%
	Zero-consumption, y_2	−0.0206	0.0252	−0.0672	−0.0217	0.0301
	All zeros, $y_2 = 0$	−0.0686	0.0078	−0.0840	−0.0686	−0.0533
	$y_2 = 1$	0.0756	0.0038	0.0682	0.0755	0.0832
	$y_2 = 2$	−0.0070	0.0070	−0.0207	−0.0070	0.0072
Urban	Nonparticipation	−0.0062	0.0261	−0.0595	−0.0054	0.0428
	Zero-consumption, y_1	0.0217	0.0258	−0.0271	0.0211	0.0733
	All zeros, $y_1 = 0$	0.0155	0.0088	−0.0018	0.0155	0.0324
	$y_1 = 1$	−0.0007	0.0012	−0.0029	−0.0008	0.0017
	$y_1 = 2$	−0.0042	0.0028	−0.0096	−0.0042	0.0014
	$y_1 = 3$	−0.0106	0.0056	−0.0215	−0.0106	0.0006
	Zero-consumption, y_2	0.0181	0.0275	−0.0337	0.0178	0.0739
	All zeros, $y_2 = 0$	0.0119	0.0090	−0.0062	0.0120	0.0295
	$y_2 = 1$	0.0597	0.0036	0.0528	0.0597	0.0668
	$y_2 = 2$	−0.0716	0.0075	−0.0864	−0.0717	−0.0566
Agriservice	Nonparticipation	−0.1989	0.0521	−0.3092	−0.1960	−0.1062
	Zero-consumption, y_1	0.2102	0.0506	0.1202	0.2075	0.3161
	All zeros, $y_1 = 0$	0.0113	0.0098	−0.0084	0.0115	0.0297
	$y_1 = 1$	0.0058	0.0018	0.0026	0.0057	0.0097
	$y_1 = 2$	0.0023	0.0033	−0.0039	0.0021	0.0092
	$y_1 = 3$	−0.0194	0.0060	−0.0311	−0.0194	−0.0077
	Zero-consumption, y_2	0.1838	0.0530	0.0871	0.1811	0.2940
	All zeros, $y_2 = 0$	−0.0151	0.0126	−0.0400	−0.0150	0.0091
	$y_2 = 1$	0.0680	0.0049	0.0588	0.0678	0.0782
	$y_2 = 2$	−0.0529	0.0096	−0.0716	−0.0530	−0.0338
Self-employed	Nonparticipation	−0.1287	0.0693	−0.2542	−0.1191	−0.0122
	Zero-consumption, y_1	0.1590	0.0686	0.0431	0.1508	0.2845
	All zeros, $y_1 = 0$	0.0303	0.0166	−0.0034	0.0305	0.0627
	$y_1 = 1$	0.0005	0.0025	−0.0042	0.0005	0.0058
	$y_1 = 2$	−0.0075	0.0060	−0.0192	−0.0075	0.0043
	$y_1 = 3$	−0.0233	0.0089	−0.0398	−0.0237	−0.0046
	Zero-consumption, y_2	0.1034	0.0704	−0.0179	0.0941	0.2327
	All zeros, $y_2=0$	−0.0254	0.0147	−0.0546	−0.0251	0.0035
	$y_2 = 1$	0.0684	0.0047	0.0594	0.0681	0.0781
	$y_2 = 2$	−0.0430	0.0118	−0.0660	−0.0431	−0.0195
Student	Nonparticipation	0.0305	0.0357	−0.0312	0.0270	0.1076
	Zero-consumption, y_1	0.0548	0.0434	−0.0353	0.0564	0.1354
	All zeros, $y_1 = 0$	0.0852	0.0206	0.0437	0.0855	0.1247
	$y_1 = 1$	−0.0090	0.0027	−0.0149	−0.0089	−0.0041
	$y_1 = 2$	−0.0295	0.0079	−0.0455	−0.0294	−0.0143
	$y_1 = 3$	−0.0468	0.0106	−0.0657	−0.0475	−0.0244
	Zero-consumption, y_2	0.0284	0.0448	−0.0686	0.0313	0.1073
	All zeros, $y_2 = 0$	0.0588	0.0239	0.0065	0.0610	0.0995
	$y_2 = 1$	0.0390	0.0102	0.0207	0.0383	0.0604
	$y_2 = 2$	−0.0979	0.0142	−0.1211	−0.0994	−0.0659

Table 6: Posterior mean, standard deviation and 95% credible intervals of parameters from BOP for smoking and chewing tobacco.

Variable	Mean	St. Dev.	2.50%	Median	97.50%
Smoking (y_1) equation, (β_1,α_1)					
Age/10	1.029	0.095	0.828	1.030	1.199
Age square/100	−0.104	0.010	−0.123	−0.105	−0.082
Education	−0.078	0.014	−0.105	−0.078	−0.050
Education square	0.002	0.001	0.000	0.002	0.004
Income	0.000	0.002	−0.004	0.000	0.005
Male	2.066	0.091	1.888	2.067	2.245
Married	0.221	0.064	0.093	0.220	0.349
Muslim	−0.083	0.049	−0.177	−0.083	0.015
Region	0.041	0.043	−0.044	0.041	0.125
Urban	−0.091	0.048	−0.186	−0.091	0.002
Agriservice	−0.121	0.050	−0.219	−0.122	−0.023
Self-employed	−0.149	0.087	−0.318	−0.150	0.021
Sstudent	−0.720	0.093	−0.905	−0.719	−0.538
α_{12}	0.270	0.015	0.241	0.270	0.300
α_{13}	0.956	0.028	0.901	0.956	1.012
Chewing (y_2) equation, (β_2,α_2)					
Age/10	0.797	0.091	0.609	0.801	0.977
Age square/100	−0.059	0.010	−0.079	−0.059	−0.039
Education	−0.023	0.016	−0.055	−0.023	0.008
Education square	−0.002	0.001	−0.005	−0.002	0.001
Income	0.002	0.003	−0.004	0.002	0.007
Male	−0.441	0.074	−0.586	−0.441	−0.295
Married	−0.010	0.073	−0.153	−0.011	0.134
Muslim	−0.077	0.056	−0.187	−0.077	0.033
Region	0.430	0.049	0.334	0.430	0.528
Urban	−0.082	0.056	−0.193	−0.081	0.026
Agriservice	0.078	0.073	−0.067	0.078	0.222
Self employed	0.177	0.087	0.010	0.176	0.351
Student	−0.715	0.177	−1.070	−0.710	−0.378
α_{22}	0.480	0.023	0.436	0.480	0.525
ρ_{12}	−0.178	0.034	−0.244	−0.179	−0.111

Each equation includes father use and mother use variables as well as a constant term.

decreases probability of nonparticipation by about 2.6% but increases probability of zero-consumption by 4.6%, implying a net increase of 2.0% in predicted probability of observing zero. The effect of age in the case of chewing tobacco is qualitatively similar, negative effect on genuine nonusers and positive effect on potential tobacco users, with the latter dominating in the overall effect.

Income has opposite effects on probability of nonparticipation and zero-consumption, predicting on average that tobacco is an inferior good for nonparticipants and a normal good for participants. However, the 95% credible interval contains zero, suggesting that the

Table 7: Posterior mean, standard deviation, and 95% credible intervals of marginal effects of covariates on probability of smoking and chewing tobacco (BOP model).

Variable	Probability	Mean	St. dev.	2.50%	Median	97.50%
Age	All zeros, $y_1 = 0$	0.0368	0.0038	0.0288	0.0369	0.0438
	$y_1 = 1$	−0.0004	0.0003	−0.0009	−0.0004	0.0002
	$y_1 = 2$	−0.0073	0.0010	−0.0093	−0.0073	−0.0054
	$y_1 = 3$	−0.0292	0.0029	−0.0345	−0.0292	−0.0230
	All zeros, $y_2 = 0$	0.0213	0.0043	0.0125	0.0215	0.0298
	$y_2 = 1$	−0.0056	0.0013	−0.0082	−0.0057	−0.0031
	$y_2 = 2$	−0.0030	0.0008	−0.0047	−0.0030	−0.0014
Education	All zeros, $y_1 = 0$	−0.0342	0.0236	−0.0803	−0.0340	0.0126
	$y_1 = 1$	0.0038	0.0025	−0.0011	0.0039	0.0086
	$y_1 = 2$	0.0130	0.0084	−0.0039	0.0129	0.0293
	$y_1 = 3$	0.0174	0.0128	−0.0076	0.0172	0.0428
	All zeros, $y_2 = 0$	0.0322	0.0156	0.0002	0.0326	0.0616
	$y_2 = 1$	−0.0150	0.0077	−0.0296	−0.0151	0.0009
	$y_2 = 2$	−0.0201	0.0113	−0.0418	−0.0203	0.0024
Income	All zeros, $y_1 = 0$	−0.0001	0.0004	−0.0009	−0.0001	0.0007
	$y_1 = 1$	0.0000	0.0000	−0.0001	0.0000	0.0001
	$y_1 = 2$	0.0000	0.0001	−0.0002	0.0000	0.0002
	$y_1 = 3$	0.0000	0.0002	−0.0004	0.0000	0.0005
	All zeros, $y_2 = 0$	−0.0003	0.0005	−0.0012	−0.0003	0.0007
	$y_2 = 1$	0.0001	0.0002	−0.0002	0.0001	0.0004
	$y_2 = 2$	0.0001	0.0002	−0.0002	0.0001	0.0004
Male	All zeros, $y_1 = 0$	−0.3824	0.0121	−0.4064	−0.3826	−0.3586
	$y_1 = 1$	0.0641	0.0040	0.0567	0.0640	0.0722
	$y_1 = 2$	0.1540	0.0065	0.1416	0.1540	0.1667
	$y_1 = 3$	0.1643	0.0083	0.1487	0.1641	0.1807
	All zeros, $y_2 = 0$	0.0721	0.0123	0.0481	0.0721	0.0962
	$y_2 = 1$	0.0500	0.0032	0.0438	0.0500	0.0565
	$y_2 = 2$	−0.1222	0.0108	−0.1430	−0.1220	−0.1014
Married	All zeros, $y_1 = 0$	−0.0416	0.0124	−0.0666	−0.0415	−0.0174
	$y_1 = 1$	0.0039	0.0013	0.0015	0.0038	0.0067
	$y_1 = 2$	0.0131	0.0042	0.0053	0.0130	0.0218
	$y_1 = 3$	0.0246	0.0070	0.0106	0.0247	0.0385
	All zeros, $y_2 = 0$	0.0018	0.0118	−0.0207	0.0018	0.0254
	$y_2 = 1$	0.0622	0.0031	0.0563	0.0622	0.0685
	$y_2 = 2$	−0.0640	0.0114	−0.0873	−0.0640	−0.0420
Muslim	All zeros, $y_1 = 0$	0.0154	0.0092	−0.0029	0.0154	0.0331
	$y_1 = 1$	−0.0013	0.0008	−0.0028	−0.0013	0.0002
	$y_1 = 2$	−0.0044	0.0026	−0.0093	−0.0044	0.0008
	$y_1 = 3$	−0.0097	0.0058	−0.0211	−0.0097	0.0018
	All zeros, $y_2 = 0$	0.0126	0.0093	−0.0051	0.0125	0.0313
	$y_2 = 1$	0.0613	0.0031	0.0555	0.0613	0.0675
	$y_2 = 2$	−0.0739	0.0091	−0.0922	−0.0739	−0.0562
Father use	All zeros, $y_1 = 0$	0.7604	0.0042	0.7521	0.7604	0.7684

Table 7: Continued.

Variable	Probability	Mean	St. dev.	2.50%	Median	97.50%
	$y_1 = 1$	0.0477	0.0027	0.0426	0.0477	0.0531
	$y_1 = 2$	0.0982	0.0035	0.0915	0.0982	0.1051
	$y_1 = 3$	0.0937	0.0032	0.0874	0.0936	0.1000
	All zeros, $y_2 = 0$	0.8713	0.0039	0.8635	0.8713	0.8789
	$y_2 = 1$	0.0623	0.0030	0.0566	0.0623	0.0684
	$y_2 = 2$	0.0664	0.0030	0.0607	0.0664	0.0724
Mother use	All zeros, $y_1 = 0$	0.7604	0.0042	0.7521	0.7604	0.7684
	$y_1 = 1$	0.0477	0.0027	0.0426	0.0477	0.0531
	$y_1 = 2$	0.0982	0.0035	0.0915	0.0982	0.1051
	$y_1 = 3$	0.0937	0.0032	0.0874	0.0936	0.1000
	All zeros, $y_2 = 0$	0.8713	0.0039	0.8635	0.8713	0.8789
	$y_2 = 1$	0.0623	0.0030	0.0566	0.0623	0.0684
	$y_2 = 2$	0.0664	0.0030	0.0607	0.0664	0.0724
Region	All zeros, $y_1 = 0$	−0.0075	0.0079	−0.0229	−0.0075	0.0080
	$y_1 = 1$	0.0006	0.0007	−0.0007	0.0006	0.0020
	$y_1 = 2$	0.0022	0.0023	−0.0023	0.0022	0.0067
	$y_1 = 3$	0.0047	0.0049	−0.0050	0.0047	0.0144
	All zeros, $y_2 = 0$	−0.0691	0.0078	−0.0846	−0.0691	−0.0539
	$y_2 = 1$	0.0756	0.0038	0.0684	0.0755	0.0832
	$y_2 = 2$	−0.0065	0.0070	−0.0200	−0.0065	0.0072
Urban	All zeros, $y_1 = 0$	0.0167	0.0087	−0.0003	0.0167	0.0339
	$y_1 = 1$	−0.0014	0.0008	−0.0030	−0.0014	0.0000
	$y_1 = 2$	−0.0049	0.0026	−0.0100	−0.0049	0.0001
	$y_1 = 3$	−0.0104	0.0054	−0.0210	−0.0104	0.0002
	All zeros, $y_2 = 0$	0.0130	0.0088	−0.0041	0.0129	0.0303
	$y_2 = 1$	0.0592	0.0036	0.0524	0.0591	0.0664
	$y_2 = 2$	−0.0721	0.0074	−0.0866	−0.0721	−0.0576
Agriservice	All zeros, $y_1 = 0$	0.0218	0.0088	0.0043	0.0219	0.0390
	$y_1 = 1$	−0.0018	0.0007	−0.0032	−0.0018	−0.0004
	$y_1 = 2$	−0.0062	0.0025	−0.0110	−0.0062	−0.0013
	$y_1 = 3$	−0.0138	0.0057	−0.0250	−0.0139	−0.0027
	All zeros, $y_2 = 0$	−0.0127	0.0119	−0.0366	−0.0127	0.0106
	$y_2 = 1$	0.0656	0.0043	0.0572	0.0655	0.0742
	$y_2 = 2$	−0.0528	0.0094	−0.0711	−0.0529	−0.0338
Self employed	All zeros, $y_1 = 0$	0.0277	0.0162	−0.0039	0.0277	0.0592
	$y_1 = 1$	−0.0028	0.0018	−0.0065	−0.0027	0.0003
	$y_1 = 2$	−0.0087	0.0053	−0.0194	−0.0086	0.0012
	$y_1 = 3$	−0.0163	0.0093	−0.0335	−0.0165	0.0025
	All zeros, $y_2 = 0$	−0.0290	0.0144	−0.0578	−0.0286	−0.0017
	$y_2 = 1$	0.0686	0.0046	0.0600	0.0685	0.0779
	$y_2 = 2$	−0.0396	0.0116	−0.0617	−0.0398	−0.0162
Student	All zeros, $y_1 = 0$	0.1287	0.0155	0.0980	0.1286	0.1588
	$y_1 = 1$	−0.0173	0.0030	−0.0235	−0.0171	−0.0118

Table 7: Continued.

Variable	Probability	Mean	St. dev.	2.50%	Median	97.50%
	$y_1 = 2$	−0.0475	0.0069	−0.0614	−0.0475	−0.0343
	$y_1 = 3$	−0.0639	0.0063	−0.0758	−0.0640	−0.0510
	All zeros, $y_2 = 0$	0.0855	0.0151	0.0531	0.0866	0.1117
	$y_2 = 1$	0.0278	0.0071	0.0154	0.0274	0.0428
	$y_2 = 2$	−0.1133	0.0088	−0.1288	−0.1140	−0.0944

effect of income is weak. Generally, the opposing effects on probabilities of nonparticipation and zeroconsumption would have repercussions on both the magnitude and the statistical significance of the full effect of observing zero-consumption. Similar considerations apply to positive levels of consumption since the marginal effect on probability of observing consumption level j ($j = 1, 2, \ldots$) can be decomposed into the marginal effects on (i) participation $P(s_i = 1)$ and (ii) levels of consumption conditional on participation, $P(y_{ri} = j \mid s_i = 1)$. These results show that policy recommendations that ignore excess zeros may lead to misleading conclusions.

5. Conclusion

In this paper we analyze the zero-inflated bivariate ordered probit model in a Bayesian framework. The underlying model arises as a mixture of a point mass distribution at $(0, 0)$ for nonparticipants and the bivariate ordered probit distribution for participants. The Bayesian analysis is carried out using MCMC techniques to approximate the posterior distribution of the parameters. Using household tobacco survey data with substantial proportion of zeros, we analyze the socioeconomic determinants of individual problem of smoking and chewing tobacco. In our illustration, we find evidence that accounting for excess zeros provides very good fit to the data. The use of a model that ignores zero-inflation masks differential effects of covariates on nonusers and users at various levels of consumption, including zeros. The Bayesian approach to modeling excess zeros provides computational flexibility of generalizing to multivariate ordered response models as well as ordinal panel data models.

The proposed zero-inflated bivariate model is particularly useful when most of the bivariate ordered outcomes are zero ($y_1 = 0, y_2 = 0$). In addition to allowing for inflation in the double-zero state, our approach can be extended to allow for zero inflation in each component. If needed, other states in an ordered regression model may be inflated as well. These extensions need to be justified empirically on a case-by-case basis and are beyond the scope of this paper.

Appendices

A.

For more details see Tables 6 and 7.

B.

WinBUGS Code for Fitting the Proposed Models (see Algorithm 1).

```
#Variable names in the tobacco data are given in y[,1:21]
 model {
    for(h in 1:N) {
             ## participation model ###
    cov2[h]<- gama[1]+gama[2]*y[h,6]+gama[3]*y[h,7]+gama[4]*y[h,8]+gama[5]*y[h,9]
    cov3[h]<- gama[6]*y[h,10]+gama[7]*y[h,11]+gama[8]*y[h,12]+gama[9]*y[h,13]
    cov4[h]<- gama[10]*y[h,14] +gama[11]*y[h,15]+gama[12]*y[h,16]
    cov5[h]<- gama[13]*y[h,17]+gama[14]*y[h,18]+gama[15]*y[h,19] +gama[16]*y[h,20]
    cov[h] <- cov2[h]+cov3[h]+cov4[h] +cov5[h]
    pi[h]    <- phi(-cov[h])
    ph.5a[h] <- phi(cov[h])
            ### consumption model #####
                  #Smoking #
    covar2[h]<-beta[1]+beta[2]*y[h,6]+beta[3]*y[h,7]+beta[4]*y[h,8]+beta[5]*y[h,9]
    covar3[h]<-beta[6]*y[h,10]+beta[7]*y[h,11]+beta[8]*y[h,12]+beta[9]*y[h,13]
                    +beta2[1]*y[h,16]
    covar4[h]<-beta2[2]*y[h,17]+beta2[3]*y[h,18]+beta2[4]*y[h,19]+beta2[5]*y[h,20]
    covar[h]  <- covar2[h]+covar3[h]+covar4[h]
                    #Chewing #
    covar2.chew[h]<-beta.chew[1]+beta.chew[2]*y[h,6]+beta.chew[3]*y[h,7]
                         +beta.chew[4]*y[h,8
    covar3.chew[h] <- beta.chew[5]*y[h,9]+beta.chew[6]*y[h,10]+beta.chew[7]*y[h,11]
    covar4.chew[h] <- beta.chew[8]*y[h,12]+beta.chew[9]*y[h,13]
    covar5.chew[h] <- beta2.chew[1]*y[h,16]+beta2.chew[2]*y[h,17]
                          +beta2.chew[3]*y[h,18]
    covar6.chew[h] <- beta2.chew[4]*y[h,19]+ beta2.chew[5]*y[h,20]
    covar.chew2[h] <-covar2.chew[h]+covar3.chew[h]+covar4.chew[h]
    covar.chew3[h] <-covar5.chew[h]+covar6.chew[h]+covar7.chew[h]
    covar.chew[h]  <- covar.chew2[h]+covar.chew3[h]
        # Cumulative probability of <  j
    ph.2[h] <- (1/sqrt(2*3.14159))*exp(-0.5*covar[h]*covar[h])
    ph.3[h] <- (1/sqrt(2*3.14159))*exp(-0.5*(alpha[1]-covar[h])*(alpha[1]-covar[h]))
    ph.4[h] <- (1/sqrt(2*3.14159))*exp(-0.5*(alpha[2]-covar[h])*(alpha[2]-covar[h]))
    ph.5b[h]<- phi(-covar[h])
        #joint CDF probability for ((y1,y2)=(0,0))
    nu.0[h] <- -rho12*ph.2[h]/phi(-covar[h])
    s2.0[h] <-1+rho12*(-covar[h])*nu.0[h]-nu.0[h]*nu.0[h]
    Q.00[h] <-ph.5b[h]*phi((-covar.chew[h]-nu.0[h])/sqrt(s2.0[h]))

        #joint CDF probability for ((y1,y2)=(0,1))
    Q.01[h] <-ph.5b[h]*phi((alpha.chew-covar.chew[h]-nu.0[h])/sqrt(s2.0[h]))
         ......
        #joint CDF probability for ((y1,y2)=(3,2))
    Q.32[h] <-1

              mu[h,1] <- pi[h] + ph.5a[h]*Q.00[h]                    #p[0,0]
              mu[h,2] <- ph.5a[h]*(Q.01[h]-Q.00[h])                  #p[0,1]
              mu[h,3] <- ph.5a[h]*(Q.02[h]-Q.01[h])                  #p[0,2]
              mu[h,4] <- ph.5a[h]*(Q.10[h]-Q.00[h])                  #p[1,0]
              mu[h,5] <-  ph.5a[h]*(Q.11[h]-Q.10[h]-Q.01[h]+Q.00[h]) #p[1,1]
              mu[h,6] <- ph.5a[h]*(Q.12[h]-Q.11[h]-Q.02[h]+Q.01[h])  #p[1,2]
              mu[h,7] <- ph.5a[h]*(Q.20[h]-Q.10[h])                  #p[2,0]
              mu[h,8] <- ph.5a[h]*(Q.21[h]-Q.20[h]-Q.11[h]+Q.10[h])  #p[2,1]
              mu[h,9] <- ph.5a[h]*(Q.22[h]-Q.21[h]-Q.12[h]+Q.11[h])  #p[2,2]
              mu[h,10] <- ph.5a[h]*(Q.30[h]-Q.20[h])                 #p[3,0]
              mu[h,11] <- ph.5a[h]*(Q.31[h]-Q.30[h]-Q.21[h]+Q.20[h]) #p[3,1]
              mu[h,12] <- ph.5a[h]*(Q.32[h]-Q.31[h]-Q.22[h]+Q.21[h]) #p[3,2]
              y[h,21] ~dcat(mu[h,1:12])}}
```

Algorithm 1

Acknowledgments

The authors thank Alfonso Flores-Lagunes, the editor, two anonymous referees and seminar participants at the Conference on Bayesian Inference in Econometrics and Statistics, the Joint Statistical Meetings, the Southern Economics Association Conference, and Syracuse University for useful comments. Mohammad Yunus graciously provided the data used in this paper.

References

[1] C. Calhoun, "Estimating the distribution of desired family size and excess fertility," *The Journal of Human Resources*, vol. 24, pp. 709–724, 1989.
[2] C. Calhoun, "Desired and excess fertility in Europe and the United States: indirect estimates from World Fertility Survey data," *European Journal of Population*, vol. 7, no. 1, pp. 29–57, 1991.
[3] A. A. Weiss, "A bivariate ordered probit model with truncation: helmet use and motor cycle injuries," *Applied Statistics*, vol. 42, pp. 487–499, 1993.
[4] J. S. Butler and P. Chatterjee, "Tests of the specification of univariate and bivariate ordered probit," *The Review of Economics and Statistics*, vol. 79, pp. 343–347, 1997.
[5] A. Biswas and K. Das, "A bayesian analysis of bivariate ordinal data. Wisconsin epidemiologic study of diabetic retinopathy revisited," *Statistics in Medicine*, vol. 21, no. 4, pp. 549–559, 2002.
[6] Z. Sajaia, "Maximum likelihood estimation of a bivariate ordered probit model: implementation and Monte Carlo simulations," Tech. Rep., The World Bank, Working Paper, 2008.
[7] M. K. Munkin and P. K. Trivedi, "Bayesian analysis of the ordered probit model with endogenous selection," *Journal of Econometrics*, vol. 143, no. 2, pp. 334–348, 2008.
[8] P. E. Stephan, S. Gurmu, A. J. Sumell, and G. C. Black, "Who's patenting in the university? Evidence from the Survey of Doctorate Recipients," *Economics of Innovation and New Technology*, vol. 16, pp. 71–99, 2007.
[9] D. Lambert, "Zero-inflated poisson regression, with an application to defects in manufacturing," *Technometrics*, vol. 31, no. 1, pp. 1–14, 1992.
[10] S. Gurmu and P. K. Trivedi, "Excess zeros in count models for recreational trips," *Journal of Business and Economic Statistics*, vol. 14, no. 4, pp. 469–477, 1996.
[11] J. Mullahy, "Heterogeneity, excess zeros, and the structure of count data models," *Journal of Applied Econometrics*, vol. 12, no. 3, pp. 337–350, 1997.
[12] S. Gurmu and J. Elder, "A bivariate zero-inflated count data regression model with unrestricted correlation," *Economics Letters*, vol. 100, pp. 245–248, 2008.
[13] D. B. Hall, "Zero-inflated Poisson and binomial regression with random effects: a case study," *Biometrics*, vol. 56, no. 4, pp. 1030–1039, 2000.
[14] G. A. Dagne, "Hierarchical Bayesian analysis of correlated zero-inflated count data," *Biometrical Journal*, vol. 46, no. 6, pp. 653–663, 2004.
[15] M. N. Harris and X. Zhao, "A zero-inflated ordered probit model, with an application to modelling tobacco consumption," *Journal of Econometrics*, vol. 141, no. 2, pp. 1073–1099, 2007.
[16] D. B. Hall and J. Shen, "Robust estimation for zero-inflated Poisson regression," *Scandinavian Journal of Statistics*, vol. 37, no. 2, pp. 237–252, 2010.
[17] L. Liu, R. L. Strawderman, M. E. Cowen, and T. S. Ya-C. T. Shih, "A flexible two-part random effects model for correlated medical costs," *Journal of Health Economics*, vol. 29, no. 1, pp. 110–123, 2010.
[18] S. Chib and B. H. Hamilton, "Bayesian analysis of cross-section and clustered data treatment models," *Journal of Econometrics*, vol. 97, no. 1, pp. 25–50, 2000.
[19] D. Gamerman, *Markov Chain Monte Carlo*, Chapman & Hall, London, UK, 1997.
[20] A. Gelman, J. Carlin, H. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman & Hall, London, UK, 1995.
[21] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London, UK, 1996.
[22] L. Tierney, "Markov chains for exploring posterior distributions (with discussion)," *The Annals of Statistics*, vol. 22, no. 4, pp. 1701–1762, 1994.
[23] C. J. Geyer, "Practical markov chain Monte Carlo (with discussion)," *Statistical Science*, vol. 7, pp. 473–511, 1992.

[24] A. E. Raftery and S. Lewis, "Comment: one long run with diagnostics: implementation strategies for Markov Chain Monte Carlo," *Statistical Science*, vol. 7, pp. 493–549, 1992.
[25] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde, "Bayesian measures of model complexity and fit (with discussion)," *Journal of the Royal Statistical Society: Series B*, vol. 64, no. 4, pp. 583–639, 2002.
[26] S. Gurmu and M. Yunus, "Tobacco chewing, smoking and health knowledge: evidence from Bangladesh," *Economics Bulletin*, vol. 9, no. 12, pp. 1–9, 2008.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2012, Article ID 640153, 17 pages
doi:10.1155/2012/640153

Review Article

Analysis of Longitudinal and Survival Data: Joint Modeling, Inference Methods, and Issues

Lang Wu,[1] Wei Liu,[2] Grace Y. Yi,[3] and Yangxin Huang[4]

[1] *Department of Statistics, The University of British Columbia, Vancouver, BC, Canada V6T 1Z2*
[2] *Department of Mathematics and Statistics, York University, Toronto, ON, Canada M3J 1P3*
[3] *Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada N2L 3G1*
[4] *Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, Tampa, FL 33612, USA*

Correspondence should be addressed to Lang Wu, lang@stat.ubc.ca

Received 25 August 2011; Accepted 10 October 2011

Academic Editor: Wenbin Lu

In the past two decades, joint models of longitudinal and survival data have received much attention in the literature. These models are often desirable in the following situations: (i) survival models with measurement errors or missing data in time-dependent covariates, (ii) longitudinal models with informative dropouts, and (iii) a survival process and a longitudinal process are associated via latent variables. In these cases, separate inferences based on the longitudinal model and the survival model may lead to biased or inefficient results. In this paper, we provide a brief overview of joint models for longitudinal and survival data and commonly used methods, including the likelihood method and two-stage methods.

1. Introduction

Longitudinal data and survival data frequently arise together in practice. For example, in many medical studies, we often collect patients' information (e.g., blood pressures) repeatedly over time and we are also interested in the time to recovery or recurrence of a disease. Longitudinal data and survival data are often *associated* in some ways. The time to event may be associated with the longitudinal trajectories. *Separate* analyses of longitudinal data and survival data may lead to inefficient or biased results. Joint models of longitudinal and survival data, on the other hand, incorporate all information simultaneously and provide valid and efficient inferences.

Figure 1 shows a longitudinal dataset in which CD4 cell counts are measured repeatedly over time in an AIDS study. Here, the time to event could be time to viral rebound,

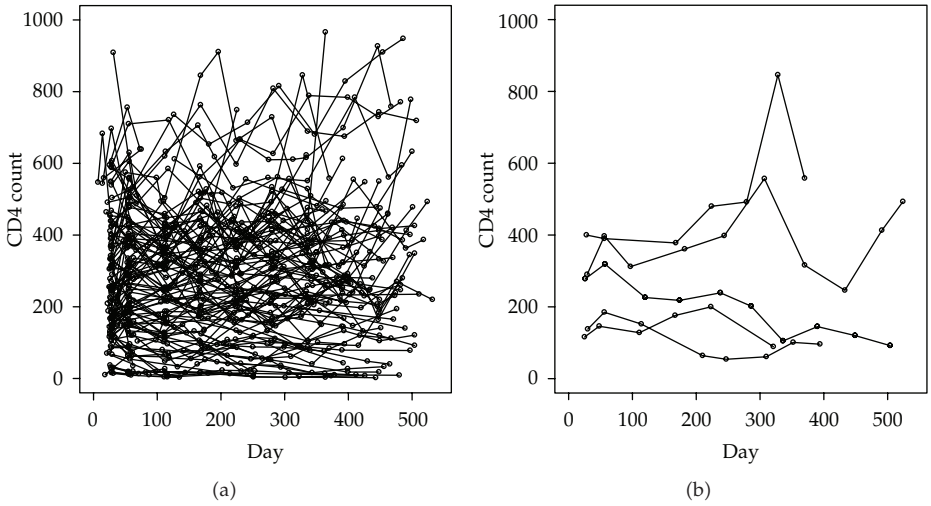


Figure 1: CD4 measurements over time. (a) All subjects. (b) Five randomly selected subjects.

time to dropout, or time to death, depending on the research objectives. Data analysis can mainly focus on either the longitudinal data or the survival data or both. When the analysis focuses on longitudinal data, we often need to address informative dropouts since dropouts are very common in longitudinal studies. When the analysis focuses on survival data, we often need to incorporate time-dependent covariates such as CD4 since the times to event may be associated with the covariate trajectories. Sometimes, the main interest may lie in the association between the longitudinal process and survival process. In any of these cases, joint models are required to feature correlated longitudinal and survival data.

Typically, joint models for longitudinal and survival data are required in the following situations:

(i) survival models with measurement errors in time-dependent covariates;

(ii) longitudinal models with informative dropouts;

(iii) longitudinal and survival processes are governed by a common latent process;

(iv) the use of external information for more efficient inference.

Joint models of longitudinal and survival data have attracted increasing attention over the last two decades. Tsiatis and Davidian [1] provided a nice overview of early work on joint models, including De Gruttola and Tu [2], Wulfsohn and Tsiatis [3], Henderson et al. [4], and Wang and Taylor [5], among others. More recent work includes Ding and Wang [6], Nathoo and Dean [7], Ye et al. [8], Albert and Shih [9], Jacqmin-Gadda et al. [10], Rizopoulos et al. [11], Wu et al. [12], Huang et al. [13], and Pan and Yi [14], among others. A typical model setting is to assume a mixed-effects model for the longitudinal data and a Cox model or an accelerated failure time (AFT) model for the survival data, with the two models sharing some random effects or variables. The likelihood method is often used, implemented by EM algorithms. Another common approach is based on two-stage methods, which are computationally simpler. Henderson et al. [4] allow different random effects in the longitudinal and survival models, but assume that the random effects are correlated. Bayesian methods have also been proposed [13, 15, 16].

Since the literature on joint models is quite extensive, it is difficult to review all references here. In this paper, we provide a brief review of the joint model literature. In Section 2, we describe a standard formulation of joint models. In Section 3, we review a commonly used two-stage method. In Section 4, we describe the standard likelihood method. In Section 5, we discuss some extensions of standard joint models. A real data example and a simulation study are presented in Section 6 to illustrate and evaluate the methods. We conclude the paper in Section 7 with discussion.

2. A Standard Formulation of a Joint Model

In this section, we consider a standard formulation of a joint model. In the literature, a typical setup is a survival model with measurement errors in time-dependent covariates, in which a linear mixed-effects (LME) model is often used to model time-dependent covariates to address covariate measurement errors and a Cox proportional hazards (PH) model is used for modelling the survival data. We focus on this setup to illustrate the basic ideas.

Consider a longitudinal study with n individuals in the sample. The objective is to model the time to an event of interest or survival time. Time-varying covariates are used in the survival model to partially explain the variation in the event times. Let s_i be the survival time for individual i, $i = 1, 2, \ldots, n$. Some individuals may not experience any events by the end of the study so their event times may be *right censored*. We assume that the censoring is random or *noninformative*. For individual i, let c_i be the censoring time, and let $\delta_i = I(s_i \leq c_i)$ be the censoring indicator such that $\delta_i = 0$ if the survival time for individual i is right censored and $\delta_i = 1$ otherwise. The observed survival data are $\{(t_i, \delta_i),\ i = 1, 2, \ldots, n\}$, where $t_i = \min(s_i, c_i)$.

In survival models, some time-dependent covariates may be measured with errors. For simplicity, we consider a single time-dependent covariate. Let z_{ij} be the *observed* covariate value for individual i at time u_{ij}, subject to measurement errors, $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m_i$. Let the corresponding unobserved *true* covariate value be z^*_{ij}. Denote $\mathbf{z}_i = (z_{i1}, \ldots, z_{im_i})^T$, and $\mathbf{z}^*_i = (z^*_{i1}, \ldots, z^*_{im_i})^T$. In many cases, the longitudinal covariate measurements are terminated at the event or censoring time t_i. For example, this is the case when the events are dropouts. In this case, we have $u_{im_i} \leq t_i$, and the covariate values after the event time t_i are all missing. Let $\mathbf{x}_i$ be covariates without measurement errors.

We consider the following Cox model for the survival data:

$$\lambda_i(t) = \lambda_0(t) \exp\left(z^*_i(t)\beta_1 + \mathbf{x}_i^T \boldsymbol{\beta}_2\right), \quad i = 1, \ldots, n, \tag{2.1}$$

where $\lambda_i(t)$ is the hazard function, $\lambda_0(t)$ is the unspecified baseline hazard function, and $\boldsymbol{\beta} = (\beta_1, \boldsymbol{\beta}_2^T)^T$ are unknown parameters. Survival model (2.1) assumes that the hazard function $\lambda_i(t)$ depends on the *unobserved true* covariate values $z^*_i(t)$, rather than the observed but mismeasured covariate value $z_i(t)$.

In Cox model (2.1), the time-dependent covariate value $z_i(t)$ should be available at any event time t. In practice, however, covariates are usually measured *intermittently* at times (say) $\{u_{ij},\ j = 1, 2, \ldots, m_i\}$ for individual i, with the measurement times possibly varying across individuals, leading to possible "missing covariates." Moreover, covariates may be measured with errors. Therefore, it is common to have both *missing data* and *measurement for errors* in time-dependent covariates, which must be addressed when conducting inference based on the Cox model (2.1). For simplicity, we assume that the missing covariate data are missing at random [17].

To address either missing data or measurement error or both, a standard approach is to model the time-dependent covariates. A common choice is the following LME model:

$$\mathbf{z}_i = U_i\boldsymbol{\alpha} + V_i\mathbf{a}_i + \boldsymbol{\epsilon}_i \equiv \mathbf{z}_i^* + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n, \tag{2.2}$$

where U_i and V_i are known design matrices, $\boldsymbol{\alpha}$ is a vector of fixed-effects parameters, $\mathbf{a}_i$ is a vector of random effects, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{im_i})^T$ is a vector of measurement errors, and the unobserved true covariates are $\mathbf{z}_i^* = U_i\boldsymbol{\alpha} + V_i\mathbf{a}_i$. Often, we assume that

$$\mathbf{a}_i \sim N(0, A), \qquad \boldsymbol{\epsilon}_i \sim N\left(0, \sigma^2 I\right), \tag{2.3}$$

and $\mathbf{a}_i$ and $\boldsymbol{\epsilon}_i$ are independent, where A is a covariance matrix, σ^2 is a parameter, and I is an identity matrix. Here, we focus on the case that the observations of the covariate process is truncated by the event time; that is, no covariate data are available after the event occurs (such as death or dropout).

Note that the survival model (2.1) and the longitudinal model (2.2) are *linked through the shared random effects* $\mathbf{a}_i$. In some applications, not necessarily in the measurement error context, the shared random effects may be viewed as a latent process that governs both the longitudinal process and the survival process. The shared random effects induce the *dependence* between the longitudinal and survival processes, and this dependence suggests the need of joint modelling.

There are two commonly used approaches for inference of joint models:

(i) two-stage methods,

(ii) likelihood methods.

In the following sections, we describe these two approaches in detail. Other approaches for joint modes have also been proposed, such as those based on estimating equations, but we omit them here for space consideration.

3. Two-Stage Methods

In the joint modelling literature, various two-stage methods have been proposed. A simple (naive) two-stage method is as follows.

Stage 1. Fit a LME model to the longitudinal covariate data, and estimate the missing or mismeasured covariates based on the fitted model.

Stage 2. Fit the survival model *separately*, with the missing or unobserved true covariate values substituted by their estimates from the first stage as if they were observed values and then proceed with the usual survival analysis.

Main advantages of the two-stage methods, including the modified two-stage methods as described below, are the simplicity and that they can be implemented with existing software. The limitation of those methods is that they may lead to biased inference for several reasons. First, in the estimation of the longitudinal covariate model parameters, the truncations resulted from the events are not incorporated. That is, the longitudinal covariate trajectories of subjects who experience an event may be different from those who do not

experience that event, so estimation of the parameters associated with the longitudinal covariate model in the first stage, based only on observed covariate data, may be biased. Second, the *uncertainty* of the estimation in the first stage is not incorporated in the second stage of the survival model estimation. Thus, standard errors of the parameter estimates of the survival model may be underestimated. Third, all information in the longitudinal process and the survival process is not fully combined in each model fitting to produce the most efficient estimates.

The bias in the estimation of the longitudinal model parameters caused by ignoring the informative truncations from the events may depend on the *strength of the association* between the longitudinal process and the survival process. The bias resulted from ignoring the estimation uncertainty in Stage 1 may depend on the *magnitude of measurement errors* in covariates. To address these issues, various modified two-stage methods have been proposed, leading to better two-stage methods.

Self and Pawitan [18] considered a two-stage method in which the least-square method was used to fit individual longitudinal covariate trajectories; the resulting estimates were used to impute the "true" covariate values in the survival models and inference was then based on the usual partial likelihood. Tsiatis et al. [19] considered an approximation to the hazard function and then using the approximation to construct the partial likelihood. They replaced the true covariate $z_i^*(t)$ by an empirical Bayes estimate of the conditional expectation $E(z_i^*(t) \mid z_i^H(t),\ t \le s_i)$, where $z_i^H(t) = \{z_i(u),\ u \le t\}$ is the covariate history up to time t. They obtained the empirical Bayes estimate from a standard fit of the LME model to the covariate data up to time t for all subjects still at risk at time t. Similar two-stage methods were also proposed in Bycott and Taylor [20] and Dafni and Tsiatis [21].

More recently, other two-stage methods have been developed in the literature. In the sequel, we review some of these recent methods. Following Prentice [22], we rewrite the survival model (2.1) as

$$\lambda_i(t; z_i(t), \mathbf{x}_i) = \lambda_0(t) E\left[\exp\left(z_i^*(t)\beta_1 + \mathbf{x}_i^T \boldsymbol{\beta}_2\right) \mid z_i(t), \mathbf{x}_i,\ t_i > t\right], \tag{3.1}$$

which involves an intractable conditional expectation. Following Dafni and Tsiatis [21] and Ye et al. [8], we approximate the above conditional expectation by

$$\begin{aligned} &E\left[\exp\left(z_i^*(t)\beta_1 + \mathbf{x}_i^T \boldsymbol{\beta}_2\right) \mid z_i(t), \mathbf{x}_i,\ t_i > t\right] \\ &\quad \approx \exp\left[E\left(z_i^*(t)\beta_1 + \mathbf{x}_i^T \boldsymbol{\beta}_2 \mid z_i(t), \mathbf{x}_i,\ t_i > t\right)\right]. \end{aligned} \tag{3.2}$$

A two-stage method may then proceed as follows. In the first step, we estimate the conditional expectation $E(z_i^*(t)\beta_1 + \mathbf{x}_i^T \boldsymbol{\beta}_2 \mid z_i(t), \mathbf{x}_i,\ t_i > t)$ by fitting the covariate model (2.2) to the observed longitudinal data and survival data. In the second step, we then substitute the conditional expectation (3.2) in (3.1) by its estimate from the first step and then we proceed with standard inference for the Cox model. Ye et al. [8] proposed two approaches for the first step, called *risk set regression calibration (RRC)* method and *ordinary regression calibration (ORC)* method, respectively. The idea is to fit the LME covariate model (2.2) to either the observed covariate data in the risk set or all observed covariate data.

Note that the bias resulted from the naive two-stage method is caused by the fact that the covariate trajectory is related to the length of followup. For example, subjects who drop

out early or die early may have different trajectories than those who stay in the study. Thus, much of the bias may be removed if we can recapture these missing covariate measurements due to truncation by incorporating the event time information. Albert and Shih [9] proposed to recapture the missing measurements by generating data from the conditional distribution of the covariate given the event time:

$$f(\mathbf{z}_i \mid s_i; \boldsymbol{\theta}) = \int f(\mathbf{z}_i \mid \mathbf{a}_i; \boldsymbol{\theta}) f(\mathbf{a}_i \mid s_i; \boldsymbol{\theta}) d\mathbf{a}_i, \tag{3.3}$$

where the covariate $\mathbf{z}_i$ and event time s_i are assumed to be conditionally independent given the random effects $\mathbf{a}_i$, and $\boldsymbol{\theta}$ contains all unknown parameters. They approximate the conditional density $f(\mathbf{z}_i \mid s_i; \boldsymbol{\theta})$ using a LME model, and then use standard software to simulate missing data from $f(\mathbf{z}_i \mid s_i; \boldsymbol{\theta})$. Once the missing measurements are simulated, the covariate model is then fitted to the "complete data," which are used in the second step. The procedure is iterated several times to incorporate the missing data uncertainty. Thus, the idea is similar to a multiple imputation method with nonignorable missing data. Such an approach may reduce the bias resulted from truncations.

To incorporate the estimation uncertainty in the first step, we may consider a *parametric bootstrap method* as follows.

Step 1. Generate covariate values based on the assumed covariate model, with the unknown parameters substituted by their estimates.

Step 2. Generate survival times from the fitted survival model.

Step 3. For each generated bootstrap dataset from Steps 1 and 2, fit the models using the two-stage method and obtain new parameter estimates.

Repeating the procedure B times (say, $B = 500$), we can obtain the standard errors for the fixed parameters from the sample covariance matrix across the B bootstrap datasets. This Bootstrap method may produce more reliable standard errors than the naive two-stage method if the assumed models are correct.

Two-stage methods have bearing with the *regression calibration* method in measurement error literature. Many of these two-stage methods may not completely remove biases. Moreover, they rely on certain assumptions and approximations. The validity of these assumptions and the accuracy of these approximations need to be further investigated.

4. Likelihood Methods

The likelihood method is perhaps the most widely used approach in the joint model literature. It provides a unified approach for inference, and it produces valid and the most efficient inference if the assumed models are correct. The likelihood method is based on the likelihood for both longitudinal data and survival data. However, since the likelihood function can be complicated, a main challenge for the likelihood method is computation.

4.1. The Likelihood

All the observed data are $\{(t_i, \delta_i, \mathbf{z}_i, \mathbf{x}_i),\ i = 1, 2, \ldots, n\}$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma, A, \boldsymbol{\lambda}_0)$ denote the collection of all unknown parameters in the models, where $\boldsymbol{\lambda}_0 = \{\lambda_0(t_i),\ i = 1, 2, \ldots, n\}$.

We assume that the censoring of survival data and the assessment process of longitudinal measurements are noninformative. The (overall) likelihood for all the observed data is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \int f(t_i, \delta_i \mid \mathbf{z}_i^*, \boldsymbol{\lambda}_0, \boldsymbol{\beta}) f\left(\mathbf{z}_i \mid \mathbf{a}_i, \boldsymbol{\alpha}, \sigma^2\right) f(\mathbf{a}_i \mid A) d\mathbf{a}_i, \tag{4.1}$$

where

$$\begin{aligned} f(t_i, \delta_i \mid \mathbf{z}_i^*, \boldsymbol{\lambda}_0, \boldsymbol{\beta}) &= \left[\lambda_0(t_i) \exp\left(z_i^*(t_i)\beta_1 + \mathbf{x}_i^T \beta_2\right)\right]^{\delta_i} \\ &\quad \times \exp\left[-\int_0^{t_i} \lambda_0(x) \exp\left(z_i^*(t_i)\beta_1 + \mathbf{x}_i^T \beta_2\right) dx\right], \\ f\left(\mathbf{z}_i \mid \mathbf{a}_i, \boldsymbol{\alpha}, \sigma^2\right) &= \left(2\pi\sigma^2\right)^{-m_i/2} \exp\left[-\frac{\left(\mathbf{z}_i - \mathbf{z}_i^*\right)^T \left(\mathbf{z}_i - \mathbf{z}_i^*\right)}{2\sigma^2}\right], \\ f(\mathbf{a}_i \mid A) &= (2\pi|A|)^{-1/2} \exp\left[-\frac{\left(\mathbf{a}_i^T A^{-1} \mathbf{a}_i\right)}{2}\right]. \end{aligned} \tag{4.2}$$

Parameter estimation can then be based on the observed-data likelihood $L(\boldsymbol{\theta})$ via the maximum likelihood method. Note that the baseline hazard $\lambda_0(t)$ in the Cox model is unspecified. It can be estimated using the nonparametric maximum likelihood method by assuming that $\lambda_0(t)$ takes discrete mass at each failure time t_i. Thus, the dimension of the parameter vector $\boldsymbol{\lambda}_0$ is equal to the number of unique failure times. This converts the semiparametric Cox model to a parametric model, but it introduces a major challenge since standard asymptotic theory for the maximum likelihood estimators (MLEs) may not apply due to the infinitely dimensional nature of $\boldsymbol{\lambda}_0$.

MLEs of the model parameters can either be obtained by a direct maximization of the observed data log likelihood or by using an EM algorithm. Since the observed data log likelihood involves an intractable integral, a direct maximization is often based on numerical integration techniques such as the Gaussian Hermite quadrature or Monte Carlo methods. These methods, however, can be quite computationally intensive if the dimension of the unobservable random effects $\mathbf{a}_i$ is not low. The EM algorithm is known for its stability and generality, so it is widely used for likelihood inference of joint models [1, 3, 11, 23]. Since the E-step of an EM algorithm still involves an intractable integral, Monte Carlo methods or Laplacian approximations are often used to approximate the conditional expectation in the E-step. In the M-step, the Newton-Raphson method is often used.

Hsieh et al. [24] noted that standard errors for estimators of the parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma)$ based on the Fisher information matrix may be problematic, because of the semiparametric nature of the joint model. They recommended a bootstrap method to obtain standard errors.

4.2. Computational Issues

A main challenge in the likelihood inference for joint models is the computational complexity, since numerical methods or Monte Carlo methods can be very computationally intensive

when the dimension of the random effects $\mathbf{a}_i$ is not small. Moreover, convergence of the EM algorithms can sometimes be an issue. Tsiatis and Davidian [1] and Tsiatis and Davidian [25] proposed an alternative approach, a so-called conditional score method, which makes no distributional assumption for the random effects $\mathbf{a}_i$. Their method treats $\mathbf{a}_i$ as "nuisance parameters" and conditions on an appropriate "sufficient statistic" for $\mathbf{a}_i$. Conditioning on this sufficient statistic would remove the dependence of the conditional distribution on the random effects $\mathbf{a}_i$. This approach leads to a set of unbiased estimating equations, similar to the usual partial likelihood score equations or generalized estimating equation (GEE). The resulting estimates are, under certain regularity conditions, consistent and asymptotically normal, although they may not be the most efficient. Moreover, their method is relatively simple to implement. Song et al. [26] considered an alternative approach to relax the normality assumption of $\mathbf{a}_i$. They assume that the random effects follow distributions in a class of smooth density family, including the standard normal density as a special case. They use the likelihood method for inference via an EM algorithm, but the computation is considerably more intensive than the conditional score method. Song and Wang [27] also proposed a local corrected score estimator and a local conditional score estimator, for which no distributional assumptions are needed for the underlying true covariates.

Approximate but computationally more efficient methods for joint models have also appeared in the literature, such as those based on Laplace approximations (e.g., [11, 12, 28]). When the dimension of the integration in the likelihood for joint models is high, the Laplace approximations offer considerably computational advantages over numerical or Monte Carlo methods. Note that the order of the Laplace approximation error is $O(m_i^{-1})$, which cannot be made arbitrarily accurate for a given dataset, where m_i is the number of within-individual measurements for individual i. Therefore, the Laplace approximation works well if the number of within-individual measurements is large. Approximate methods based on Taylor series approximations are similar to the Laplace approximation; that is, their performance improves as m_i increases.

Rizopoulos et al. [11] proposed to use the full exponential Laplace approximation in the E-step of the EM algorithm. Compared to the standard (first-order) Laplace approximation, the full exponential Laplace approximate method has approximation error of order $O(m_i^{-2})$ and requires a much smaller number of within-individual longitudinal measurements to produce reliable results. Lee et al. [28] suggested second-order Laplace approximations. However, these Laplace approximation methods cannot control the magnitude of the approximation errors, unlike Gaussian quadrature or Monte Carlo integration techniques.

5. Bayesian Methods

Bayesian joint models have also been studied by various authors, including Faucett and Thomas [29], Xu and Zeger [15], Wang and Taylor [5], Law et al. [30], Ibrahim et al. [16], and Huang et al. [13]. Joint models may contain many unknown parameters, which may lead to potential problems in inference. A main advantage of Bayesian methods is that they can borrow additional information from similar studies or from experts and incorporate this information in the current analysis, in the forms of prior distributions for the current model parameters. Thus, Bayesian methods can be very useful for inference of joint models.

For Bayesian joint models, the model parameters are assumed to follow some prior distributions, and inference is then based on the posterior distribution given the observed data. Let $\boldsymbol{\theta}$ denote the collection of unknown parameters in the joint model, and let $f(\boldsymbol{\theta})$

denote the prior distribution. Let $\mathfrak{D} = \{(t_i, \delta_i, \mathbf{z}_i, \mathbf{x}_i),\ i = 1, 2, \ldots, n\}$ denote all observed data. The joint *posterior distribution* for all unknown parameters $\boldsymbol{\theta}$ and random effects $\mathbf{a} = \{\mathbf{a}_i,\ i = 1, \ldots, n\}$ is then given by

$$f(\boldsymbol{\theta}, \mathbf{a} \mid \mathfrak{D}) \propto \prod_{i=1}^{n} \left[f(t_i, \delta_i \mid \mathbf{z}_i^*, \mathbf{x}_i, \boldsymbol{\theta}) f(\mathbf{z}_i \mid \mathbf{a}_i, \boldsymbol{\theta}) f(\mathbf{a}_i \mid A) \right] f(\boldsymbol{\theta} \mid \boldsymbol{\theta}_0), \tag{5.1}$$

where $\boldsymbol{\theta}_0$ are known hyperparameters. Bayesian inference is then based on Monte Carlo samples drawn from the posterior distribution $f(\boldsymbol{\theta}, \mathbf{a} \mid \mathfrak{D})$ using an MCMC algorithm such as the Gibbs sampler. For example, the posterior means and variances of the parameters can be estimated based on these Monte Carlo samples, and Bayesian inference can then be based on these estimated posterior means and variances. This Monte Carlo sampling can be done using the publically available WinBUGS software [31], which is quite general, flexible, and easy to use.

Like other Bayesian methods, it is desirable to check if the final results are sensitive to the choices of prior distributions. Sometimes, in the absence of prior information, noninformative priors or flat priors may be desirable.

6. Other Joint Models

In the previous sections, we have focused on joint models based on a Cox model for right-censored survival data and a LME model for longitudinal data. Other models for survival data and longitudinal data can also be considered in joint models. For example, for survival data, we may consider accelerated failure time (AFT) models and models for interval censored data and models for recurrent events. For longitudinal data, nonlinear, generalized linear mixed models or semiparametric/nonparametric mixed models can be utilized. Although the different survival models and longitudinal models can be employed, basic ideas and approaches for inference remain essentially the same. In the following, we briefly review some of these joint models.

6.1. Joint Models Based on an LME Model and an AFT Model

In joint modelling of longitudinal and survival data, we can use the AFT model to feature survival data. Here, we focus on an AFT model with measurement errors in time-dependent covariates. For longitudinal data, we again consider LME models for simplicity. The description below is based on Tseng et al. [23]. A semiparametric AFT model can be written in a form similar to the Cox model:

$$h_i(t) = h_0\left[\int_0^t \exp\{-z_i^*(u)\beta\}du\right] \exp\{-z_i^*(t)\beta\}, \tag{6.1}$$

where $h_i(t)$ is the hazard function of the ith individual at time t, $h_0(t)$ is the baseline hazard function, and $z_i^*(t)$ is the unobserved true covariate value at time t. For the observed measurements $z_i(t)$, we again consider the LME model (2.2).

Tseng et al. [23] proposed a likelihood method using an EM algorithm. The likelihood for all observed data is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \int f(t_i, \delta_i \mid \mathbf{z}_i^*, h_0, \beta) f\left(\mathbf{z}_i \mid \mathbf{a}_i, \boldsymbol{\alpha}, \sigma^2\right) f(\mathbf{a}_i \mid A) d\mathbf{a}_i, \tag{6.2}$$

where $f(\mathbf{z}_i \mid \mathbf{a}_i, \boldsymbol{\alpha}, \sigma^2)$ and $f(\mathbf{a}_i \mid A)$ are the same as those for (4.1) and

$$f(t_i, \delta_i \mid \mathbf{z}_i^*, h_0, \beta) = \left[h_0\{\phi(t_i; \boldsymbol{\theta}, \mathbf{a}_i)\} \frac{\partial \phi(t_i; \overline{\mathbf{z}}_i^*, \beta)}{\partial t_i} \right]^{\delta_i} \exp\left\{ -\int_0^{\phi(t_i; \overline{\mathbf{z}}_i^*, \beta)} h_0(u)du \right\}, \tag{6.3}$$

where $\overline{\mathbf{z}}_i^*$ denotes the covariate history and ϕ is a known function.

Handling the AFT structure in the joint modelling setting is more difficult than for the Cox model, since $f(t_i, \delta_i \mid \mathbf{z}_i^*, h_0, \boldsymbol{\beta})$ is more complicated and the baseline function $h_0\{\phi(t_i; \overline{\mathbf{z}}_i^*, \boldsymbol{\beta})\}$ involves unknown quantities $(\boldsymbol{\beta}, \boldsymbol{\alpha}, A, \mathbf{a}_i)$, while this is not the case in the Cox model. One cannot use the point mass function with masses assigned to all uncensored survival times t_i for the baseline hazard function h_0. In other words, in Cox models, the baseline hazard h_0 can be represented by a collection of parameters which are point masses, but this approach is not feasible for the AFT model because of its dependence on covariates via function $\phi(t_i; \overline{\mathbf{z}}_i^*, \boldsymbol{\beta})$. To circumvent this, Tseng et al. [23] assumed the baseline hazard function h_0 to be a step function, taking constant values between two consecutive failure times.

Tseng et al. [23] used an Monte Carlo EM algorithm to obtain the MLEs. The framework is similar to that in the previous section. They used a Monte Carlo method to approximate the conditional expectations in the E-step. The M-step involves more complicated computations due to the complicated baseline hazard h_0. To obtain the standard errors of the MLEs, the usual asymptotic formula based on Fisher information matrix may be questionable, so they used a bootstrap method.

6.2. Joint Models with Interval Censored Survival Data

In the previous sections, we have focused on right censored survival data and assume that either the *exact* survival times or censoring times are observed. In practice, however, we often cannot observe the exact survival nor censoring times, but we only know that events have occurred over certain time intervals. Such survival data are called *interval censored*. For simplicity, we assume that all individuals are assessed at the same times. Again, let S_i be the time to an event (survival time) for individual i, with observed value s_i. Let $\mathbf{r}_i = (r_{i1}, \ldots, r_{im})^T$ be the vector of event indicators such that $r_{ij} = 1$ if subject i has an event occurred from time t_{j-1} to time t_j, and let $r_{ij} = 0$ otherwise, $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$. We assume that $r_{i1} = 0$ for all i. Let $p_{ij} = P(t_{j-1} \le S_i < t_j)$, and let

$$\pi_{ij} = P(t_{j-1} \le S_i < t_j \mid S_i \ge t_{j-1}) = 1 - P(S_i \ge t_j \mid S_i \ge t_{j-1}). \tag{6.4}$$

Then, we have $p_{ij} = (1 - \pi_{i1})(1 - \pi_{i2}) \cdots (1 - \pi_{i,j-1})\pi_{ij}$. The probability function for the event indicator vector $\mathbf{r}_i$ can be written as

$$f(\mathbf{r}_i) = \prod_{j=1}^{m} p_{ij}^{r_{ij}} = \prod_{j=1}^{m} \pi_{ij}^{r_{ij}} (1 - \pi_{ij})^{1 - r_{ij}}, \tag{6.5}$$

which is the probability function for a Bernoulli distribution. We can introduce observed error-prone covariate value $\mathbf{z}_i$, with true value $\mathbf{z}_i^*$, and assume

$$\log\{-\log(1 - \pi_{ij})\} = \boldsymbol{\beta}^T \mathbf{z}_i^* + \gamma_j, \tag{6.6}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)^T$ are unknown parameters. Then, we can write the probability function of $\mathbf{r}_i$ as $f(\mathbf{r}_i \mid \mathbf{z}_i^*, \boldsymbol{\beta}, \boldsymbol{\gamma})$. Alternatively, we can assume that $\mathbf{r}_i$ depends on $\mathbf{z}_i^*$ only through the random effects $\mathbf{a}_i$ and writing the probability function of $\mathbf{r}_i$ as $f(\mathbf{r}_i \mid \mathbf{a}_i, \boldsymbol{\beta}, \boldsymbol{\gamma})$.

Let $f(\mathbf{z}_i \mid \mathbf{a}_i, \boldsymbol{\alpha}, \sigma)$ be the conditional probability density function, given the random effects $\mathbf{a}_i$, and $f(\mathbf{a}_i \mid A)$ be the marginal probability density function for $\mathbf{a}_i$ with covariance matrix A. Let $\boldsymbol{\theta}$ denote the collection of all parameters in all models. Then, the *likelihood* for all the observed data can be written as

$$L_o(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left[\int f(\mathbf{z}_i \mid \mathbf{a}_i, \boldsymbol{\alpha}, \sigma) f(\mathbf{r}_i \mid \mathbf{a}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}) f(\mathbf{a}_i \mid A) d\mathbf{a}_i \right]. \tag{6.7}$$

MLE of parameters $\boldsymbol{\theta}$ can be obtained by maximizing the observed data likelihood $L_o(\boldsymbol{\theta})$. Because the observed-data likelihood $L_o(\boldsymbol{\theta})$ can be difficult to evaluate due to its involvement of an intractable and possibly high-dimensional integral, one may proceed with Monte Carlo EM algorithms or other computationally more efficient approximate methods.

6.3. GLMM and NLME Models for Longitudinal Data

We have focused on LME models for modelling the longitudinal data. Other models for longitudinal data can also be considered. For example, one may consider nonlinear mixed-effects (NLME) models for modelling the longitudinal data in joint models [12, 32]. NLME models are often mechanistic models in the sense that they are typically based on the underlying mechanisms which generate the data. On the other hand, LME models are typically empirical models; that is, they are usually used to approximately describe the observed data without considering possible data-generation mechanisms. Thus, NLME models may be scientifically more desirable if such models exist. Similarly, for nonnormal longitudinal data, generalized linear mixed models (GLMMs) can be considered, which are special nonlinear models but are essentially empirical models as well.

When the longitudinal models are nonlinear, the general ideas of the two-stage methods and likelihood methods for joint models can still be applied. The complication is that computation becomes more demanding, because of the nonlinearity of the longitudinal models.

6.4. Joint Models with Missing Data

For longitudinal data, missing values are very common. When missing data are nonignorable in the sense that the missingness probability may be related to the missing values or the random effects, the missing data process is often needed to be incorporated in inferential procedures in order to obtain valid results. For likelihood methods, it is straightforward to incorporate missing data mechanisms in joint model inference. However, the computation becomes even more challenging. Wu et al. [12, 32] considered the missing data problems for joint models, using Monte Carlo EM algorithms and Laplace approximations.

7. Example and Simulation

7.1. An Illustrating Example

As an illustration, we consider an AIDS dataset which includes 46 HIV infected patients receiving an anti-HIV treatment. Viral load (i.e., plasma HIV RNA) and CD4 cell count were

repeatedly measured during the treatment. The number of viral load measurements for each individual varies from 4 to 10. It is known that CD4 is measured with substantial errors. About 11% viral load measurements are below the detection limit after the initial period of viral decay. We call the viral load below the detection limit as "viral suppression." We wish to check if the initial CD4 trajectories are predictive for the time to viral suppression.

Let s_i be the time to viral suppression, that is, the time from the start of the treatment to the first scheduled time when the viral load drops below the detection limit. The viral suppression times for patients whose viral loads never dropped below detection limit may be regarded as right censored, with the study end time as the censoring time. We employ two models, the Cox model (2.1) or the AFT model (6.1), to feature the time to viral suppression, where $z_i^*(t)$ is the unobserved true CD4 cell count at time t for individual i. We do not consider additional covariates here.

To address the measurement error in the time-dependent covariate CD4 cell count, we use the LME model to model the CD4 trajectories:

$$\text{CD4}_{ij} = (\alpha_0 + a_{i1}) + \left(\alpha_1 + a_{i2}u_{ij}\right) + \epsilon_{ij}, \tag{7.1}$$

where the parameters, random effects and random errors, and the assumed distributions are the same as those described in the previous sections. The fixed parameters (α_0, α_1) are population intercept and slope of the CD4 process and (a_{i1}, a_{i2}) are individual variations from the population averages. To avoid very small (large) estimates, which may be unstable, we standardize the CD4 cell counts and rescale the original time t (in days) so that the new time scale is between 0 and 1. We estimate the model parameters using the joint model method and the two-stage method with/out bootstrap standard error correction. The number $B = 500$ of the bootstrap samples is taken. For the joint model method, we consider the Cox model (2.1) and the AFT model (6.1) with $h_0(\cdot)$ being the Weibull baseline risk function. On the other hand, only the Cox model (2.1) is employed for the two-stage method for comparison. These analyses may be implemented by using the functions *coxph*(), *lme*(), and *jointModel*() in R software.

Table 1 presents the resulting estimates of main parameters of interest and their standard errors. We can see from Table 1 that under either the Cox or the AFT survival models, both the two-stage and the joint model methods produce similar estimates for the covariate (CD4) longitudinal model. However, the two-stage method may underestimate the standard errors of the parameter estimates since it does not incorporate the survival information in the estimation procedure. Note that the parameters in the Cox model and the AFT model have different interpretations, due to different model formulations, so they are not directly comparable.

The parameter β_1 in the survival models measures the effect of the true time-dependent covariate CD4 values on event times, so its estimate and the associated P value can be used to check if the true CD4 values are predictive for the times to viral suppression. Since the covariate CD4 is measured with errors, addressing the measurement error is the main focus of joint modelling in this application. Thus, the estimation of β_1 is of primary interest. For the joint model method, under either the Cox or the AFT models, there is some evidence that covariate CD4 is associated with the time to viral suppression, after measurement error has been addressed. It is seen that evidence of significance of the covariate effect is the strongest under the Cox model. On the other hand, the two-stage method may severely underestimate the covariate CD4 effect (the small value of $\widehat{\beta}_1$). Moreover, the naive two-stage method underestimates the standard error of $\widehat{\beta}_1$, due to failing to incorporate the estimating

Table 1: Analyses of the AIDS data under different models.

Model	Method		β_0	β_1	α_0	α_1	σ
Cox model	Two-stage	Estimate	—	0.315	−0.233	1.345	0.605
		SE	—	0.208	0.113	0.154	—
		P value	—	0.129	0.040	<0.001	—
		BSE	—	0.237	—	—	—
Cox model	Joint model	Estimate	—	0.648	−0.201	1.342	0.603
		SE	—	0.234	0.135	0.162	—
		P value	—	0.006	0.137	<0.001	—
AFT model	Joint model	Estimate	0.168	−0.487	−0.237	1.341	0.604
		SE	0.260	0.289	0.125	0.156	—
		P value	0.517	0.091	0.059	<0.001	—

SE: standard error; BSE: bootstrap standard error. The P values are from the Wald tests for testing if the corresponding parameters are zero or not.

uncertainty from the first step. This underestimation of standard error is somewhat corrected by the bootstrap method.

7.2. A Simulation Study

In this section, we conduct a simulation study to compare the joint model method and the two-stage method with/out bootstrap standard error correction. We generate 500 datasets from the time-dependent covariate CD4 process (7.1) in the example of the previous section and the Cox model (2.1) with constant baseline hazard function $\lambda_0(t) \equiv 1$ with emphasis on the effect of the time-dependent covariate. The measurement time points used in the simulation are the same as those in the example of the previous section. The true values of model parameters, given in Table 2, are similar to those in the example, and the variance-covariance matrix A of the random effect $\mathbf{a}_i$ is set to be diagonal. Again, we take $B = 500$ bootstrap samples for the bootstrap standard error in the two-stage method.

In Table 2, we report the simulation results of averages parameter estimates (Est), empirical standard errors (ESE), averages of asymptotic standard errors (ASE), average of bootstrap standard errors (BSE), empirical biases (Bias), mean square errors (MSE), and coverage rates (CR) for the 95% confidence intervals. We can see from Table 2 that the two-stage and the joint model methods produce similar parameter estimates (close to true parameters) except the one for the covariate effect β_1. In particular, the estimate $\widehat{\beta}_1$ based on the joint model method is very close to its true value, while the estimate based on the two-stage method is about one-third of its true value, which indicates that the two-stage method may underestimate the time-dependent covariate effect severely. The joint model method provides smaller mean square errors and more reasonable coverage rates for the 95% confidence intervals than the two-stage method. Moreover, the two-stage method may underestimate the standard deviation of $\widehat{\beta}_1$ and the bootstrap correction on this standard error seems plausible.

From the above results, we see that the joint likelihood method produces less biased estimates and more reliable standard errors than the two-stage method. These results have important implications. For example, if one uses Wald-type tests for model selection, the likelihood method would give more reliable results. However, two-stage methods are generally simpler and computationally quicker to output estimates than likelihood methods.

Table 2: Comparison of the two-stage Method and the joint likelihood method via a simulation study.

Method	Parameter	β_1	α_0	α_1	σ	A_{11}	A_{22}
	True value	0.6	−0.2	1.3	0.6	0.5	0.3
Two-stage	Est	0.183	−0.183	1.303	0.598	0.501	0.353
	ESE	0.216	0.111	0.164	0.025	0.114	0.209
	ASE	0.201	0.111	0.164	—	—	—
	BSE	0.250	—	—	—	—	—
	Bias	−0.417	0.017	0.003	−0.002	−0.001	0.053
	MSE	0.221	0.013	0.027	0.0007	0.013	0.046
	CR	42.8	95.6	94.4	—	—	—
Joint model	Est	0.6004	−0.175	1.296	0.598	0.492	0.321
	ESE	0.256	0.103	0.161	0.020	0.092	0.156
	ASE	0.249	0.099	0.163	—	—	—
	Bias	0.0004	0.025	−0.004	−0.002	−0.008	0.021
	MSE	0.066	0.011	0.026	0.0004	0.008	0.025
	CR	95.6	95.8	95.2	—	—	—

We can also compare the two methods with Bayesian methods. Note that, however, Bayesian methods are equivalent to the likelihood method when noninformative priors are used. We expect that Bayesian methods have similar performance to likelihood methods.

8. Discussion

We have provided a brief review of common joint models and methods for inference. In practice, when we need to consider a longitudinal process and an event process and suspect that the two processes may be associated, such as survival models with time-dependent covariates or longitudinal models with informative dropouts, it is important to use joint model methods for inference in order to avoid biased results. The literature on model selection for joint models is quite limited. In practice, the best longitudinal model can be selected based on the observed longitudinal data, and the best survival model can be selected based on the survival data, using standard model selection procedures for these models. Then, we specify reasonable link between the two models, such as shared random effects. To choose methods for inference, the joint likelihood method generally produces most reliable results *if* the assumed models and distributions are correct. On the other hand, the two-stage methods may be computationally simpler, and many existing models and methods for longitudinal data and survival data can be easily adapted. However, two-stage methods may not completely eliminate the biases in parameter estimates in some cases.

When the longitudinal covariate process terminates at event times, that is, when the longitudinal values are unavailable at and after the event times such as deaths or dropouts, the covariates are sometimes called *internal time-dependent covariates*. Sometimes, however, longitudinal covariate information is available at and after the event times. For example, CD4 measurements may be still available after patients have been diagnosed with AIDS. Such covariates are sometimes called *external covariates*. In joint models, it is important to distinguish internal and external time-dependent covariates. In particular, for internal time-dependent covariates, joint models are more desirable since separate analysis in this case may lead to more severe bias.

Survival models with measurement errors in time-dependent covariates have received much attention in the joint models literature. Another common situation is longitudinal models with informative dropouts, in which survival models can be used to model the dropout process. Both situations focus on characterizing the association between the longitudinal and survival processes. Some authors have also considered joint models in which the focus is on more efficient inference of the survival model, using longitudinal data as auxiliary information [15, 33, 34] or assume that the longitudinal process and the survival process are governed by a common latent process [4]. Nathoo and Dean [7] considered an interesting joint model in which an NLME model is used to model tree growth, with spatial correlation incorporated.

Joint models can also be extended to *multivariate cases*, in which more than one longitudinal processes and more than one event processes can be modelled simultaneously. Extensions are often conceptually straightforward, but computation and implementation can be more tedious than univariate cases. See Henderson et al. [4], Xu and Zeger [35], and Song et al. [26].

Zeng and Cai [36] derived some asymptotic results for maximum likelihood estimators in joint analysis of longitudinal and survival data. They showed the consistency of the maximum likelihood estimators, derived their asymptotic distributions, and showed that the maximum likelihood estimators in joint analysis are semiparametrically efficient.

Although there has been extensive research in joint models in the last two decades and the importance of joint models has been increasingly recognized, joint models are still not widely used in practice. A main reason is perhaps lack of software. Recently, Dimitris Rizopoulos has developed an R package called *JM* that can be used to fit joint models with normal longitudinal responses and event times under a maximum likelihood approach. Various options for the survival model and optimization/integration algorithms are provided, such as Cox models and AFT models for survival data and the Gauss-Hermite integration methods and Laplace approximations.

Acknowledgments

The authors are grateful to the editor and two referees for helpful and constructive comments. The research was partially supported by the Canada Natural Sciences and Engineering Research Council (NSERC) discovery grants to L. Wu, W. Liu, and G. Y. Yi and by NIAID/NIH Grant AI080338 and MSP/NSA Grant H98230-09-1-0053 to Y. Huang.

References

[1] A. A. Tsiatis and M. Davidian, "An overview of joint modeling of longitudinal and time-to-event data," *Statistica Sinica*, vol. 14, pp. 793–818, 2004.

[2] V. De Gruttola and X. M. Tu, "Modelling progression of CD4-lymphocyte count and its relationship to survival time," *Biometrics*, vol. 50, no. 4, pp. 1003–1014, 1994.

[3] M. S. Wulfsohn and A. A. Tsiatis, "A joint model for survival and longitudinal data measured with error," *Biometrics*, vol. 53, no. 1, pp. 330–339, 1997.

[4] R. Henderson, P. Diggle, and A. Dobson, "Joint modelling of longitudinal measurements and event time data," *Biostatistics*, vol. 1, pp. 465–480, 2000.

[5] Y. Wang and J. M. G. Taylor, "Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome," *Journal of the American Statistical Association*, vol. 96, no. 455, pp. 895–905, 2001.

[6] J. Ding and J.-L. Wang, "Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data," *Biometrics*, vol. 64, no. 2, pp. 546–556, 2008.

[7] F. S. Nathoo and C. B. Dean, "Spatial multistate transitional models for longitudinal event data," *Biometrics*, vol. 64, no. 1, p. 271–279, 326, 2008.
[8] W. Ye, X. Lin, and J. M. G. Taylor, "Semiparametric modeling of longitudinal measurements and time-to-event data—a two-stage regression calibration approach," *Biometrics*, vol. 64, no. 4, pp. 1238–1246, 2008.
[9] P. S. Albert and J. H. Shih, "On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure," *Biometrics*, vol. 66, no. 3, pp. 983–987, 2010.
[10] H. Jacqmin-Gadda, C. Proust-Lima, J. M. G. Taylor, and D. Commenges, "Score test for conditional independence between longitudinal outcome and time to event given the classes in the joint latent class model," *Biometrics*, vol. 66, no. 1, pp. 11–19, 2010.
[11] D. Rizopoulos, G. Verbeke, and E. Lesaffre, "Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data," *Journal of the Royal Statistical Society Series B*, vol. 71, no. 3, pp. 637–654, 2009.
[12] L. Wu, W. Liu, and X. J. Hu, "Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors," *Biometrics*, vol. 66, no. 2, pp. 327–335, 2010.
[13] Y. Huang, G. Dagne, and L. Wu, "Bayesian inference on joint models of HIV dynamics for time-to-event and longitudinal data with skewness and covariate measurement errors," *Statistics in Medicine*, vol. 30, no. 24, pp. 2930–2946, 2011.
[14] Q. Pan and G. Y. Yi, "An estimation method of marginal treatment effects on correlated longitudinal and survival outcomes," *Statistics and Its Inference*. In press.
[15] J. Xu and S. L. Zeger, "Joint analysis of longitudinal data comprising repeated measures and times to events," *Journal of the Royal Statistical Society Series C*, vol. 50, no. 3, pp. 375–387, 2001.
[16] J. G. Ibrahim, M. Chen, and D. Sinha, "Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials," *Statistica Sinica*, vol. 14, no. 3, pp. 863–883, 2004.
[17] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, Wiley Series in Probability and Statistics, Wiley-Interscience, New York, NY, USA, 2nd edition, 2002.
[18] S. Self and Y. Pawitan, "Modeling a marker of disease progression and onset of disease," in *AIDS Epidemiology: Methodological Issues*, N. P. Jewell, K. Dietz, and V. T. Farewell, Eds., Birkhäauser, Boston, Mass, USA, 1992.
[19] A. A. Tsiatis, V. DeGruttola, and M. S. Wulfsohn, "Modeling the relationship of survival to longitudinal data measured with error: applications to survival and CD4 counts in patients with AIDS," *Journal of the American Statistical Association*, vol. 90, pp. 27–37, 1995.
[20] P. Bycott and J. Taylor, "A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model," *Statistics in Medicine*, vol. 17, no. 18, pp. 2061–2077, 1998.
[21] U. G. Dafni and A. A. Tsiatis, "Evaluating surrogate markers of clinical outcome when measured with error," *Biometrics*, vol. 54, no. 4, pp. 1445–1462, 1998.
[22] R. L. Prentice, "Covariate measurement errors and parameter estimation in a failure time regression model," *Biometrika*, vol. 69, no. 2, pp. 331–342, 1982.
[23] Y. K. Tseng, F. Hsieh, and J. L. Wang, "Joint modelling of accelerated failure time and longitudinal data," *Biometrika*, vol. 92, no. 3, pp. 587–603, 2005.
[24] F. Hsieh, Y.-K. Tseng, and J.-L. Wang, "Joint modeling of survival and longitudinal data: likelihood approach revisited," *Biometrics*, vol. 62, no. 4, pp. 1037–1043, 2006.
[25] A. A. Tsiatis and M. Davidian, "A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error," *Biometrika*, vol. 88, no. 2, pp. 447–458, 2001.
[26] X. Song, M. Davidian, and A. A. Tsiatis, "A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data," *Biometrics*, vol. 58, no. 4, pp. 742–753, 2002.
[27] X. Song and C. Y. Wang, "Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients," *Biometrics*, vol. 64, no. 2, pp. 557–566, 2008.
[28] Y. Lee, J. A. Nelder, and Y. Pawitan, *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood*, vol. 106 of *Monographs on Statistics and Applied Probability*, Chapman & Hall/CRC, London, UK, 2006.
[29] C. L. Faucett and D. C. Thomas, "Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach," *Statistics in Medicine*, vol. 15, no. 15, pp. 1663–1685, 1996.
[30] N. J. Law, J. M. G. Taylor, and H. M. Sandler, "The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure," *Biostatistics*, vol. 3, pp. 547–563, 2002.

[31] D. J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter, "WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility," *Statistics and Computing*, vol. 10, no. 4, pp. 325–337, 2000.
[32] L. Wu, X. J. Hu, and H. Wu, "Joint inference for nonlinear mixed-effects models and time to event at the presence of missing data," *Biostatistics*, vol. 9, no. 2, pp. 308–320, 2008.
[33] C. L. Faucett, N. Schenker, and J. M. G. Taylor, "Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data," *Biometrics*, vol. 58, no. 1, pp. 37–47, 2002.
[34] J. W. Hogan and N. M. Laird, "Model-based approaches to analysing incomplete longitudinal and failure time data," *Statistics in Medicine*, vol. 16, no. 1–3, pp. 259–272, 1997.
[35] J. Xu and S. L. Zeger, "The evaluation of multiple surrogate endpoints," *Biometrics*, vol. 57, no. 1, pp. 81–87, 2001.
[36] D. Zeng and J. Cai, "Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time," *Annals of Statistics*, vol. 33, no. 5, pp. 2132–2163, 2005.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2012, Article ID 614102, 19 pages
doi:10.1155/2012/614102

Research Article

Mixed-Effects Tobit Joint Models for Longitudinal Data with Skewness, Detection Limits, and Measurement Errors

Getachew A. Dagne and Yangxin Huang

Department of Epidemiology & Biostatistics, College of Public Health, University of South Florida, Tampa, FL 33612, USA

Correspondence should be addressed to Getachew A. Dagne, gdagne@health.usf.edu

Received 29 May 2011; Accepted 13 August 2011

Academic Editor: Lang Wu

Complex longitudinal data are commonly analyzed using nonlinear mixed-effects (NLME) models with a normal distribution. However, a departure from normality may lead to invalid inference and unreasonable parameter estimates. Some covariates may be measured with substantial errors, and the response observations may also be subjected to left-censoring due to a detection limit. Inferential procedures can be complicated dramatically when such data with asymmetric characteristics, left censoring, and measurement errors are analyzed. There is relatively little work concerning all of the three features simultaneously. In this paper, we jointly investigate a skew-*t* NLME Tobit model for response (with left censoring) process and a skew-*t* nonparametric mixed-effects model for covariate (with measurement errors) process under a Bayesian framework. A real data example is used to illustrate the proposed methods.

1. Introduction

Modeling of longitudinal data is an active area of biostatistics and statistics research that has received a lot of attention in the recent years. Various statistical modeling and analysis methods have been suggested in the literature for analyzing such data with complex features (Higgins et al. [1], Liu and Wu [2], Wulfsohn and Tsiatis [3], and Wu [4]). However, there is a relatively little work done on simultaneously accounting for skewness, left censoring due to a detection limit (for example, a threshold below which viral loads are not quantifiable) and covariate measurement errors, which are inherent features of longitudinal data. This paper proposes a joint skew-*t* NLME Tobit model for a response and measurement errors in covariate by simultaneously accounting for left-censoring and skewness. Thus, the proposed model addresses three important features of longitudinal data such as viral load in an AIDS study.

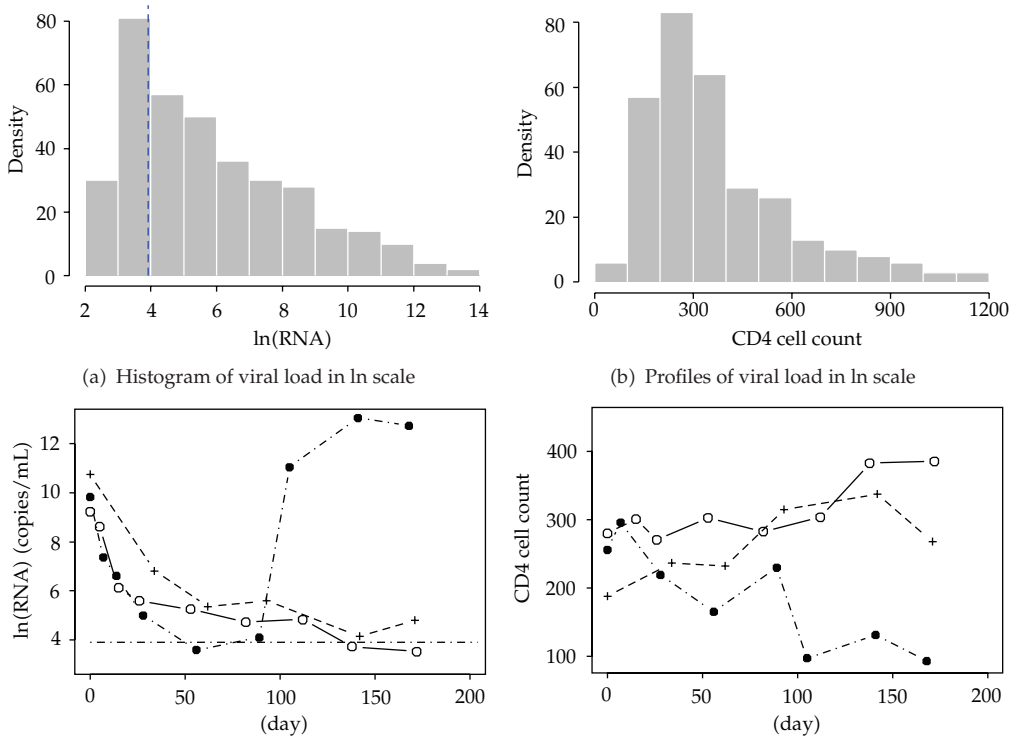
Firstly, our model relaxes the normality assumption for random errors and random-effects by using flexible skew-normal and skew-t distributions. It has been documented in the literature that the normality assumption lacks robustness against extreme values, obscures important features of between- and within-subject variations, and leads to biased or misleading results (Huang and Dagne [5], Verbeke and Lesaffre [6], and Sahu et al. [7]). Specially, nonnormal characteristics such as skewness with heavy tails appear very often in virologic responses. For example, Figures 1(a) and 1(b) displays the histograms of repeated viral load (in ln scale) and CD4 cell count measurements for 44 subjects enrolled in an AIDS clinical study (Acosta et al. [8]). For this data set, which is analyzed in this paper, both viral load (even after ln-transformation) and CD4 cell count are highly skewed, and thus a normality assumption may be violated.

Secondly, an outcome of a longitudinal study may be subject to a detection limit because of low sensitivity of current standard assays (Perelson et al. [9]). For example, for a longitudinal AIDS study, designed to collect data on every individual at each assessment, the response (viral load) measurements may be subject to left censoring due to a detection limit of quantification. Figures 1(c) and 1(d) shows the measurements of viral load and CD4 cell count for three randomly selected patients in the study. We can see that for some patients their viral loads are below detection limit (BDL), which is 50 (in copies/mL). When observations fall below the BDL, a common practice is to impute the censored values by either the detection limit or half of the detection limit (Wu [4], Ding and Wu [10], and Davidian and Giltinan [11]). Such *ad hoc* methods may produce biased results (Hughes [12]). In this paper, instead of arbitrarily imputing the observations below detection limit, we impute them using fully Bayesian predictive distributions based on a Tobit model (Tobin [13]), which is discussed in Section 2.

Thirdly, another feature of a longitudinal data set is the existence of time-varying covariates which suffer from random measurement errors. This is usually the case in a longitudinal AIDS study where CD4 cell counts are often measured with substantial measurement errors. Thus, any statistical inference without considering measurement errors in covariates may result in biased results (Liu and Wu [2], Wu [4], and Huang and Dagne [5]). In this paper, we jointly model measurement errors in covariate process along with the response process. The distributional assumption for the covariate model is a skew-t distribution which is relatively robust against potential extreme values and heavy tails.

Our research was motivated by the AIDS clinical trial considered by Acosta et al. [8]. In this study, 44 HIV-1-infected patients were treated with a potent artiretroviral regimen. RNA viral load was measured in copies/mL at study days 0, 7, 14, 28, 56, 84, 112, 140, and 168 of followup. Covariates such as CD4 cell counts were also measured throughout the study on similar scheme. In this study, the viral load detectable limit is 50 copies/mL, and there are 107 out of 357 (30 percent) of all viral load measurements that are below the detection limit. Previous studies show that change in viral load may be associated with change in CD4 cell counts. It is important to study the patterns of virological response to treatment in order to make clinical decisions and provide individualized treatments. Since viral load measurements appear to be skewed and censored, and in addition CD4 cell counts are typically measured with substantial errors and skewness, statistical analyses must take all these factors into account.

For longitudinal data, it is not clear how asymmetric nature, left censoring due to BDL, and covariate measurement error may interact and simultaneously influence inferential procedures. It is the objective of this paper to investigate the effects on inference when all of the three typical features exist in the longitudinal data. To achieve our objective,



(a) Histogram of viral load in ln scale

(b) Profiles of viral load in ln scale

(c) Profiles of CD4 cell count

(d) Histogram of CD4 cell count

Figure 1: The histograms (a,b) of viral load measured from RNA levels (in natural ln scale) and standardized CD4 cell count in plasma for 44 patients in an AIDS clinical trial study. Profiles (c,d) of viral load (response) in ln scale and CD4 cell count (covariate) for three randomly selected patients. The vertical and horizontal lines in (a) and (c) are below the detectable level of viral load (3.91 = ln(50)).

we employ a fairly general framework to accommodate a large class of problems with various features. Accordingly, we explore a flexible class of skew-elliptical (SE) distributions (see the Appendix for details) which include skew-normal (SN) and skew-t (ST) distributions as special cases for accounting skewness and heavy tails of longitudinal data, extend the Tobit model (Tobin [13]) to treat all left-censored observations as missing values, and investigate nonparametric mixed effects model for covariate measured with error under the framework of joint models. Because the SN distribution is a special case of the ST distribution when the degrees of freedom approach infinity, for the completeness and convenient presentation, we chose ST distributions to develop NLME Tobit joint models (i.e., the ST distribution is assumed for within-subject random errors and between-subject random effects). The skewness in both within-subject random errors and random-effects distributions may jointly contribute to the skewness of response and covariate variables in a longitudinal study, which makes the assumption of normality unrealistic.

The remaining of the paper is structured as follows. In Section 2, we present the joint models with ST distribution and associated Bayesian modeling approach in general forms so that they can be applicable to other scientific fields. In Section 3, we discuss specific joint models for HIV response process with left censoring and CD4 covariate process with measurement error that are used to illustrate the proposed methods using the data set

described above and report the analysis results. Finally, the paper concludes with some discussions in Section 4.

2. Joint Models and Bayesian Inferential Methods

2.1. Skew-t Mixed-Effects Tobit Joint Models

In this section, we present the models and methods in general forms so that our methods may be applicable to other areas of research. An approach we present in this paper treats censored values as realizations of a latent (unobserved) continuous variable that has been left-censored. This idea was popularized by Tobin ([13]) and the resulting model is commonly referred to as the Tobit model. Denote the number of subjects by n and the number of measurements on the ith subject by n_i. Let $y_{ij} = y_i(t_{ij})$ and $z_{ij} = z_i(t_{ij})$ be observed response and covariate for individual i at time t_{ij} $(i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, n_i)$ and q_{ij} denote the latent response variable that would be measured if the assay did not have a lower detectible limit ρ. In our case the Tobit model can be formulated as

$$y_{ij} = \begin{cases} q_{ij} & \text{if } q_{ij} > \rho, \\ \text{missing} & \text{if } q_{ij} \le \rho, \end{cases} \tag{2.1}$$

where ρ is a nonstochastic BDL, which in our example below is equivalent to ln(50). Note that the value of y_{ij} is missing when it is less than or equal to ρ.

For the response process with left-censoring, we consider the following NLME model with an ST distribution which incorporates possibly mismeasured time-varying covariates

$$\begin{aligned} y_{ij} &= g\left(t_{ij}, \mathbf{x}_{ij}, \boldsymbol{\beta}_{ij}\right) + e_{ij}, \qquad \mathbf{e}_i \ iid \sim ST_{n_i,\nu_e}\left(\mathbf{0}, \sigma^2\mathbf{I}_{n_i}, \boldsymbol{\Delta}(\boldsymbol{\delta}_{e_i})\right), \\ \boldsymbol{\beta}_{ij} &= \mathbf{d}\left(z^*_{ij}, \boldsymbol{\beta}, \mathbf{b}_i\right), \qquad \mathbf{b}_i \ iid \sim ST_{s_3,\nu_b}(\mathbf{0},\ \Sigma_{\mathrm{b}}, \boldsymbol{\Delta}(\delta_b)), \end{aligned} \tag{2.2}$$

where $\mathbf{x}_{ij}$ is an $s_1 \times 1$ design vector, $g(\cdot)$ is a linear or nonlinear known function, $\mathbf{d}(\cdot)$ is an s_1-dimensional vector-valued linear function, $\boldsymbol{\beta}_{ij}$ is an $s_1 \times 1$ individual-specific time-dependent parameter vector, $\boldsymbol{\beta}$ is an $s_2 \times 1$ population parameter vector $(s_2 \ge s_1)$; in the model (2.2), we assume that the individual-specific parameters $\boldsymbol{\beta}_{ij}$ depend on the true (but unobservable) covariate z^*_{ij} rather than the observed covariate z_{ij}, which may be measured with errors, and we discuss a covariate model (2.3) below.

It is noticed that we assume that an $s_3 \times 1$ vector of random effects $\mathbf{b}_i = (b_{i1}, \ldots, b_{is_3})^T$ $(s_3 \le s_1)$ follows a multivariate ST distribution with the unrestricted covariance matrix $\boldsymbol{\Sigma}_b$, the $s_3 \times s_3$ skewness diagonal matrix $\boldsymbol{\Delta}(\boldsymbol{\delta}_b) = \mathrm{diag}(\delta^b_1, \ldots, \delta^b_{s_3})$, and the degree of freedom ν_b; the model random error $\mathbf{e}_i = (e_{i1}, \ldots, e_{in_i})^T$ follows a multivariate ST distribution with the unknown scale parameter σ^2, the degree of freedom ν_e, and the $n_i \times n_i$ skewness diagonal matrix $\boldsymbol{\Delta}(\boldsymbol{\delta}_{e_i}) = \mathrm{diag}(\delta_{e_{i1}}, \ldots, \delta_{e_{in_i}})$, where the $n_i \times 1$ skewness parameter vector $\boldsymbol{\delta}_{e_i} = (\delta_{e_{i1}}, \ldots, \delta_{e_{in_i}})^T$. In particular, if $\delta_{e_{i1}} = \cdots = \delta_{e_{in_i}} \widehat{=} \delta_e$, then $\boldsymbol{\Delta}(\boldsymbol{\delta}_{e_i}) = \delta_e\mathbf{I}_{n_i}$ and $\boldsymbol{\delta}_{e_i} = \delta_e\mathbf{1}_{n_i}$ with $\mathbf{1}_{n_i} = (1, \ldots, 1)^T$; this indicates that we are interested in skewness of overall data set and is the case to be used in real data analysis in Section 3.

Covariate models have been investigated extensively in the literature (Higgins et al. [1], Liu and Wu [2], Wu [4], and Carroll et al. [14]). However, those models used the

normality assumption for random measurement errors. As we pointed out earlier, this assumption lacks robustness against departures from normality and may also lead to misleading results. In this paper, we extend the covariate models by assuming an ST distribution for the random errors. We adopt a flexible empirical nonparametric mixed-effects model with an ST to quantify the covariate process as follows:

$$z_{ij} = w(t_{ij}) + h_i(t_{ij}) + \epsilon_{ij}\left(\equiv z_{ij}^* + \epsilon_{ij}\right) \quad \boldsymbol{\epsilon}_i \ iid \sim ST_{n_i,\nu_\epsilon}\left(\mathbf{0}, \tau^2\mathbf{I}_{n_i}, \boldsymbol{\Delta}(\boldsymbol{\delta}_{\epsilon_i})\right), \tag{2.3}$$

where $w(t_{ij})$ and $h_i(t_{ij})$ are unknown nonparametric smooth fixed-effects and random effects functions, respectively, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^T$ follows a multivariate ST distribution with degrees of freedom ν_ϵ, the unknown scale parameter τ^2, and the $n_i \times n_i$ skewness diagonal matrix $\boldsymbol{\Delta}(\boldsymbol{\delta}_{\epsilon_i}) = \text{diag}(\delta_{\epsilon_{i1}}, \ldots, \delta_{\epsilon_{in_i}})$ with $n_i \times 1$ skewness parameter vector $\boldsymbol{\delta}_{\epsilon_i} = (\delta_{\epsilon_{i1}}, \ldots, \delta_{\epsilon_{in_i}})^T$. In particular, if $\delta_{\epsilon_{i1}} = \cdots = \delta_{\epsilon_{in_i}} \widehat{=} \delta_\epsilon$, then $\boldsymbol{\Delta}(\boldsymbol{\delta}_{\epsilon_i}) = \delta_\epsilon \mathbf{I}_{n_i}$ and $\boldsymbol{\delta}_{\epsilon_i} = \delta_\epsilon \mathbf{1}_{n_i}$. $z_{ij}^* = w(t_{ij}) + h_i(t_{ij})$ are the true but unobservable covariate values at time t_{ij}. The fixed smooth function $w(t)$ represents population average of the covariate process, while the random smooth function $h_i(t)$ is introduced to incorporate the large interindividual variation in the covariate process. We assume that $h_i(t)$ is the realization of a zero-mean stochastic process.

Nonparametric mixed-effects model (2.3) is more flexible than parametric mixed-effects models. To fit model (2.3), we apply a regression spline method to $w(t)$ and $h_i(t)$. The working principle is briefly described as follows and more details can be found in the literature (Davidian and Giltinan [11] and Wu and Zhang [15]). The main idea of regression spline is to approximate $w(t)$ and $h_i(t)$ by using a linear combination of spline basis functions. For instance, $w(t)$ and $h_i(t)$ can be approximated by a linear combination of basis functions $\boldsymbol{\Psi}_p(t) = \{\psi_0(t), \psi_1(t), ..., \psi_{p-1}(t)\}^T$ and $\boldsymbol{\Phi}_q(t) = \{\phi_0(t), \phi_1(t), ..., \phi_{q-1}(t)\}^T$, respectively. That is,

$$w(t) \approx w_p(t) = \sum_{l=0}^{p-1} \alpha_l \psi_l(t) = \boldsymbol{\Psi}_p(t)^T \boldsymbol{\alpha}, \qquad h_i(t) \approx h_{iq}(t) = \sum_{l=0}^{q-1} a_{il} \phi_l(t) = \boldsymbol{\Phi}_q(t)^T \mathbf{a}_i, \tag{2.4}$$

where $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_{p-1})^T$ is a $p \times 1$ vector of fixed-effects and $\mathbf{a}_i = (a_{i0}, \ldots, a_{i,q-1})^T$ $(q \leq p)$ is a $q \times 1$ vector of random-effects with $\mathbf{a}_i \ iid \sim ST_{q,\nu_a}(\mathbf{0}, \boldsymbol{\Sigma}_a, \boldsymbol{\Delta}(\boldsymbol{\delta}_a))$ with the unrestricted covariance matrix $\boldsymbol{\Sigma}_a$, the skewness diagonal matrix $\boldsymbol{\Delta}(\boldsymbol{\delta}_a) = \text{diag}(\delta_1^a, \ldots, \delta_q^a)$, and the degrees of freedom ν_a. Based on the assumption of $h_i(t)$, we can regard $\mathbf{a}_i$ as iid realizations of a zero-mean random vector. For our model, we consider natural cubic spline bases with the percentile-based knots. To select an optimal degree of regression spline and numbers of knots, that is, optimal sizes of p and q, the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) is often applied (Davidian and Giltinan [11] and Wu and Zhang [15]). Replacing $w(t)$ and $h_i(t)$ by their approximations $w_p(t)$ and $h_{iq}(t)$, we can approximate model (2.3) by the following linear mixed-effects (LME) model:

$$z_{ij} \approx \boldsymbol{\Psi}_p\left(t_{ij}\right)^T \boldsymbol{\alpha} + \boldsymbol{\Phi}_q\left(t_{ij}\right)^T \mathbf{a}_i + \epsilon_{ij} \approx z_{ij}^* + \epsilon_{ij}, \quad \boldsymbol{\epsilon}_i \ iid \sim ST_{n_i,\nu_\epsilon}\left(\mathbf{0}, \tau^2\mathbf{I}_{n_i}, \boldsymbol{\Delta}(\boldsymbol{\delta}_{\epsilon_i})\right). \tag{2.5}$$

2.2. Simultaneous Bayesian Inference

In a longitudinal study, such as the AIDS study described previously, the longitudinal response and covariate processes are usually connected physically or biologically. Statistical

inference based on the commonly used two-step method may be undesirable since it fails to take the covariate estimation into account (Higgins et al. [1]). Although a simultaneous inference method based on a joint likelihood for the covariate and response data may be favorable, the computation associated with the joint likelihood inference in joint models of longitudinal data can be extremely intensive and may lead to convergence problems and in some cases it can even be computationally infeasible (Liu and Wu [2] and Wu [4]). Here we propose a simultaneous Bayesian inference method based on MCMC procedure for longitudinal data of response with left censoring and covariate with measurement error. The Bayesian joint modeling approach may pave a way to alleviate the computational burdens and to overcome convergence problems.

We assume that $\mathbf{a}_i$, $\mathbf{b}_i$, $\boldsymbol{\epsilon}_i$, and $\mathbf{e}_i$ are mutually independent of each other. Following Sahu et al. [7] and properties of ST distribution, in order to specify the models (2.5) and (2.2) for MCMC computation, it can be shown that by introducing four random variable vectors $\mathbf{w}_{e_i} = (w_{e_{i1}}, \ldots, w_{e_{in_i}})^T$, $\mathbf{w}_{\epsilon_i} = (w_{\epsilon_{i1}}, \ldots, w_{e_{in_i}})^T$, $\mathbf{w}_{b_i} = (w_{b_{i1}}, \ldots, w_{b_{is_3}})^T$ and $\mathbf{w}_{a_i} = (w_{a_{i1}}, \ldots, w_{a_{iq}})^T$ and four random variables ξ_{e_i}, ξ_{ϵ_i}, ξ_{b_i}, and ξ_{a_i} $(i = 1, \ldots, n)$ based on the stochastic representation for the ST distribution (see the Appendix for details), z_{ij} and y_{ij} can be hierarchically formulated as

$$
\begin{aligned}
y_{ij} \mid \mathbf{b}_i, w_{e_{ij}}, \xi_{e_i};\ \boldsymbol{\beta}, \sigma^2, \delta_{e_{ij}} &\sim N\Big(g\Big(t_{ij},\ \mathbf{x}_{ij}, \mathbf{d}\Big(z^*_{ij}, \boldsymbol{\beta}, \mathbf{b}_i\Big)\Big) + \delta_{e_{ij}} w_{e_{ij}}, \xi^{-1}_{e_i}\sigma^2\Big),\\
w_{e_{ij}} &\sim N(0,1) I\Big(w_{e_{ij}} > 0\Big), \qquad \xi_{e_i} \mid \nu_e \sim G\Big(\frac{\nu_e}{2}, \frac{\nu_e}{2}\Big),\\
\mathbf{b}_i \mid \mathbf{w}_{b_i}, \xi_{b_i}; \boldsymbol{\Sigma}_b, \boldsymbol{\delta}_b &\sim N_{s_3}\Big(\boldsymbol{\Delta}(\boldsymbol{\delta}_b)\mathbf{w}_{b_i}, \xi^{-1}_{b_i}\boldsymbol{\Sigma}_b\Big),\\
\mathbf{w}_{b_i} &\sim N_{s_3}(\mathbf{0}, \mathbf{I}_{s_3}) I(\mathbf{w}_{b_i} > \mathbf{0}), \qquad \xi_{b_i} \mid \nu_b \sim G\Big(\frac{\nu_b}{2}, \frac{\nu_b}{2}\Big),\\
z_{ij} \mid \mathbf{a}_i, w_{\epsilon_{ij}}, \xi_{\epsilon_i}; \boldsymbol{\alpha}, \tau^2, \delta_{\epsilon_{ij}} &\sim N\Big(z^*_{ij} + \delta_{\epsilon_{ij}} w_{\epsilon_{ij}}, \xi^{-1}_{\epsilon_i}\tau^2\Big),\\
w_{e_{ij}} &\sim N(0,1) I\Big(w_{\epsilon_{ij}} > 0\Big), \qquad \xi_{\epsilon_i} \mid \nu_\epsilon \sim G\Big(\frac{\nu_\epsilon}{2}, \frac{\nu_\epsilon}{2}\Big),\\
\mathbf{a}_i \mid \mathbf{w}_{a_i}, \xi_{a_i}; \boldsymbol{\Sigma}_a, \boldsymbol{\delta}_a &\sim N_q\Big(\boldsymbol{\Delta}(\boldsymbol{\delta}_a)\mathbf{w}_{a_i}, \xi^{-1}_{a_i}\boldsymbol{\Sigma}_a\Big),\\
\mathbf{w}_{a_i} &\sim N_q(\mathbf{0}, \mathbf{I}_q) I(\mathbf{w}_{a_i} > \mathbf{0}), \qquad \xi_{a_i} \mid \nu_a \sim G\Big(\frac{\nu_a}{2}, \frac{\nu_a}{2}\Big),
\end{aligned}
\tag{2.6}
$$

where $G(\cdot)$ is a gamma distribution, $I(w_{e_{ij}} > 0)$ is an indicator function, and $w_{e_{ij}} \sim N(0,1)$ truncated in the space $w_{e_{ij}} > 0$ (standard half-normal distribution); $w_{\epsilon_{ij}}$, $\mathbf{w}_{a_i}$, and $\mathbf{w}_{b_i}$ can be defined similarly. z^*_{ij} is viewed as the true but unobservable covariate values at time t_{ij}. It is noted that, as discussed in the Appendix, the hierarchical model with the ST distribution (2.6) can be reduced to the following three special cases: (i) a model with skew-normal (SN) distribution as $\nu_e, \nu_\epsilon, \nu_b, \nu_a \to \infty$ and $\xi_{e_i}, \xi_{\epsilon_i}, \xi_{b_i}$ and $\xi_{a_i} \to 1$ with probability 1 (i.e., the four corresponding distributional specifications are omitted in (2.6)); (ii) a model with standard t-distribution as $\delta_{\epsilon_{ij}} = \delta_{e_{ij}} = 0$, $\boldsymbol{\delta}_b = \boldsymbol{\delta}_a = \mathbf{0}$, and thus the four distributional specifications of $w_{\epsilon_{ij}}$, $w_{e_{ij}}$, $\mathbf{w}_{a_i}$, and $\mathbf{w}_{b_i}$ are omitted in (2.6); (iii) a model with standard normal distribution as $\nu_e, \nu_\epsilon, \nu_a, \nu_b \to \infty$ and $\delta_{\epsilon_{ij}} = \delta_{e_{ij}} = 0$ and $\boldsymbol{\delta}_b = \boldsymbol{\delta}_a = \mathbf{0}$; in this case, the eight corresponding distributional specifications are omitted in (2.6).

Let $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau^2, \sigma^2, \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b, \nu_\epsilon, \nu_e, \nu_a, \nu_b, \boldsymbol{\delta}_a, \boldsymbol{\delta}_b, \boldsymbol{\delta}_{\epsilon_i}, \boldsymbol{\delta}_{e_i}; i = 1, \ldots, n\}$ be the collection of unknown parameters in models (2.2) and (2.5). To complete the Bayesian formulation, we need to specify prior distributions for unknown parameters in the models (2.2) and (2.5) as follows:

$$\begin{gathered}
\boldsymbol{\alpha} \sim N_p(\boldsymbol{\alpha}_0, \boldsymbol{\Lambda}_1), \quad \tau^2 \sim \mathrm{IG}(\omega_1, \omega_2), \quad \boldsymbol{\Sigma}_a \sim \mathrm{IW}(\boldsymbol{\Omega}_1, \rho_1), \quad \boldsymbol{\delta}_{\epsilon_i} \sim N_{n_i}(\mathbf{0}, \boldsymbol{\Gamma}_1), \\
\boldsymbol{\beta} \sim N_{s_2}(\boldsymbol{\beta}_0, \boldsymbol{\Lambda}_2), \quad \sigma^2 \sim \mathrm{IG}(\omega_3, \omega_4), \quad \boldsymbol{\Sigma}_b \sim \mathrm{IW}(\boldsymbol{\Omega}_2, \rho_2), \quad \boldsymbol{\delta}_{e_i} \sim N_{n_i}(\mathbf{0}, \boldsymbol{\Gamma}_2), \\
\nu_\epsilon \sim G(\nu_{\epsilon 0}, \nu_{\epsilon 1}) I(\nu_\epsilon > 3), \quad \nu_e \sim G(\nu_{e0}, \nu_{e1}) I(\nu_e > 3), \quad \nu_a \sim G(\nu_{a0}, \nu_{a1}) I(\nu_a > 3), \\
\nu_b \sim G(\nu_{b0}, \nu_{b1}) I(\nu_b > 3), \quad \boldsymbol{\delta}_a \sim N_q(\mathbf{0}, \boldsymbol{\Gamma}_3), \quad \boldsymbol{\delta}_b \sim N_{s_3}(\mathbf{0}, \boldsymbol{\Gamma}_4),
\end{gathered} \tag{2.7}$$

where the mutually independent Inverse Gamma (IG), Normal (N), Gamma (G), and Inverse Wishart (IW) prior distributions are chosen to facilitate computations (Pinheiro and Bates [16]). The hyperparameter matrices $\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \boldsymbol{\Gamma}_3$, and $\boldsymbol{\Gamma}_4$ can be assumed to be diagonal for convenient implementation.

Let $f(\cdot \mid \cdot), F(\cdot \mid \cdot)$ and $\pi(\cdot)$ denote a probability density function (pdf), cumulative density function (cdf), and prior density function, respectively. Conditional on the random variables and some unknown parameters, a detectable measurement y_{ij} contributes $f(y_{ij} \mid \mathbf{b}_i, w_{e_{ij}}, u_{e_i})$, whereas a nondetectable measurement contributes $F(\rho \mid \mathbf{b}_i, w_{e_{ij}}, u_{e_i}) \equiv P(y_{ij} < \rho \mid \mathbf{b}_i, w_{e_{ij}}, u_{e_i})$ in the likelihood. We assume that $\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau^2, \sigma^2, \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b, \nu_\epsilon, \nu_e, \boldsymbol{\delta}_{\epsilon_i}, \boldsymbol{\delta}_{e_i} (i = 1, \ldots, n)$ are independent of each other, that is, $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta})\pi(\tau^2)\pi(\sigma^2)\pi(\boldsymbol{\Sigma}_a)\pi(\boldsymbol{\Sigma}_b)\pi(\nu_\epsilon)\pi(\nu_e)\pi(\nu_a)\pi(\nu_b)\pi(\boldsymbol{\delta}_a)\pi(\boldsymbol{\delta}_b)\prod_i \pi(\boldsymbol{\delta}_{\epsilon_i})\pi(\boldsymbol{\delta}_{e_i})$. After we specify the models for the observed data and the prior distributions for the unknown model parameters, we can make statistical inference for the parameters based on their posterior distributions under the Bayesian framework. Letting $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^T$ and $\mathbf{z}_i = (z_{i1}, \ldots, z_{in_i})^T$, the joint posterior density of $\boldsymbol{\theta}$ based on the observed data can be given by

$$f(\boldsymbol{\theta} \mid \text{data}) \propto \left\{ \prod_{i=1}^{n} \int \int L_{\mathbf{y}_i} L_{\mathbf{z}_i} L_{\mathbf{a}_i} L_{\mathbf{b}_i} d\mathbf{a}_i d\mathbf{b}_i \right\} \pi(\theta), \tag{2.8}$$

where $L_{\mathbf{y}_i} = \prod_{j=1}^{n_i} f(y_{ij} \mid \mathbf{b}_i, w_{e_{ij}}, \xi_{e_i})^{1-c_{ij}} F(\rho \mid \mathbf{b}_i, w_{e_{ij}}, \xi_{e_i})^{c_{ij}} f(w_{e_{ij}} \mid w_{e_{ij}} > 0) f(\xi_{e_i})$ is the likelihood for the observed response data, c_{ij} is the censoring indicator such that y_{ij} is observed if $c_{ij} = 0$, and y_{ij} is left-censored if $c_{ij} = 1$, that is, $y_{ij} = q_{ij}$ if $c_{ij} = 0$, and y_{ij} is treated as missing if $c_{ij} = 1$, and $L_{\mathbf{z}_i} = \prod_{j=1}^{n_i} f(z_{ij} \mid \mathbf{a}_i, w_{e_{ij}}, \xi_{e_i}) f(w_{e_{ij}} \mid w_{e_{ij}} > 0) f(\xi_{\epsilon_i})$ is the likelihood for the observed covariate data $\{\mathbf{z}_i, i = 1, \ldots, n\}$, $L_{\mathbf{b}_i} = f(\mathbf{b}_i \mid \mathbf{w}_{b_i}, \xi_{b_i}) f(\mathbf{w}_{b_i} \mid \mathbf{w}_{b_i} > \mathbf{0}) f(\xi_{b_i})$, and $L_{\mathbf{a}_i} = f(\mathbf{a}_i \mid \mathbf{w}_{a_i}, \xi_{a_i}) f(\mathbf{w}_{a_i} \mid \mathbf{w}_{a_i} > \mathbf{0}) f(\xi_{a_i})$.

In general, the integrals in (2.8) are of high dimension and do not have closed form solutions. Therefore, it is prohibitive to directly calculate the posterior distribution of $\boldsymbol{\theta}$ based on the observed data. As an alternative, MCMC procedures can be used to sample based on (2.8) using the Gibbs sampler along with the Metropolis-Hasting (M-H) algorithm. An important advantage of the above representations based on the hierarchical models (2.6) and (2.7) is that they can be very easily implemented using the freely available WinBUGS software (Lunn et al. [17]) and that the computational effort is equivalent to the one necessary to fit the normal version of the model. Note that when using WinBUGS to implement our modeling approach, it is not necessary to explicitly specify the full conditional distributions. Thus we omit those here to save space.

3. Data Analysis

3.1. Specification of Models

We now analyze the data set described in Section 1 based on the proposed method. Among the 44 eligible patients, the number of viral load measurements for each patient varies from 4 to 9 measurements. As is evident from Figures 1(c) and 1(d), the interpatient variations in viral load appear to be large and these variations appear to change over time. Previous studies suggest that the interpatient variation in viral load may be partially explained by time-varying CD4 cell count (Wu [4] and Huang et al. [18]).

Models for covariate processes are needed in order to incorporate measurement errors in covariates. CD4 cell counts often have nonnegligible measurement errors, and ignoring these errors can lead to severely misleading results in a statistical inference (Carroll et al. [14]). In A5055 study, roughly 10 per cent of the CD4 measurement times are inconsistent with the viral load measurement times. Consequently, CD4 measurements may be missed at viral load measurement times mainly due to a different CD4 measurement scheme as designed in the study (e.g., CD4 measurements were missed at day 7 as displayed in Figures 1(c) and 1(d)). There seem to be no particular patterns for the missingness. Thus we assume that the missing data in CD4 are missing at random (MAR) in the sense of Rubin [19], so that the missing data mechanism can be ignored in the analysis. With CD4 measures collected over time from the AIDS study, we may model the CD4 process to partially address the measurement errors (Wu [4]). However, the CD4 trajectories are often complicated, and there is no well-established model for the CD4 process. We, thus, model the CD4 process empirically using a nonparametric mixed-effects model, which is flexible and works well for complex longitudinal data. We use linear combinations of natural cubic splines with percentile-based knots to approximate $w(t)$ and $h_i(t)$. Following the study in (Liu and Wu [2]), we set $\psi_0(t) = \phi_0(t) = 1$ and take the same natural cubic splines in the approximations (2.4) with $q \le p$ (in order to limit the dimension of random-effects). The values of p and q are determined based on the AIC/BIC criteria. The AIC/BIC values are evaluated for various models with $(p,q) = \{(1,1),(2,1),(2,2),(3,1),(3,2),(3,3)\}$ which was found that the model with $(p,q) = (3,3)$ has the smallest AIC/BIC values being 703.6/744.4. We thus adopted the following ST nonparametric mixed-effects CD4 covariate model:

$$z_{ij} = (\alpha_0 + a_{i0}) + (\alpha_1 + a_{i1})\psi_1(t_{ij}) + (\alpha_2 + a_{i2})\psi_2(t_{ij}) + \epsilon_{ij}\left(\equiv z_{ij}^* + \epsilon_{ij}\right), \tag{3.1}$$

where z_{ij} is the observed CD4 value at time t_{ij}, $\psi_1(\cdot)$ and $\psi_2(\cdot)$ are two basis functions given in Section 2.1 and taking the same natural cubic splines for $\phi(\cdot)$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)^T$ is a vector of population parameters (fixed-effects), $\mathbf{a}_i = (a_{i0}, a_{i1}, a_{i2})^T$ is a vector of random-effects, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^T \sim ST_{n_i,\nu_\epsilon}(\mathbf{0}, \tau^2\mathbf{I}_{n_i}, \delta_\epsilon\mathbf{I}_{n_i})$. In addition, in order to avoid too small or large estimates which may be unstable, we standardize the time-varying covariate CD4 cell counts (each CD4 value is subtracted by mean 375.46 and divided by standard deviation 228.57) and rescale the original time (in days) so that the time scale is between 0 and 1.

For the initial stage of viral decay after treatment, a biologically reasonable viral load model can be formulated by the uniexponential form (Ho et al. [20]), $V(t) = V(0)\exp(-\lambda t)$, where $V(t)$ is the total virus at time t and λ is the rate of change in viral load. To model the complete viral load trajectory, one possible extension of the model given above is to allow λ to vary over time. A simple determinant for time-varying λ is the linear function $\lambda(t) = a + bt$.

For HIV viral dynamic models, it is typical to take ln-transformation of the viral load in order to stabilize the variance and to speed up estimation algorithm (Ding and Wu [10]). After ln-transformation of $V(t)$, substituting λ by the linear function $\lambda(t) = a + bt$, we obtain the following quadratic linear mixed-effects model:

$$y_{ij} = \beta_{i0} + \beta_{ij1}t_{ij} + \beta_{ij2}t_{ij}^2 + e_{ij}, \tag{3.2}$$

where $y_{ij} = \ln(V_i(t_{ij}))$, parameter β_{i0} represents the initial viral load in ln scale, and parameters β_{ij1} and β_{ij2} incorporate change in viral decay rate over time, with $\lambda_{ij} \equiv -(\beta_{ij1} + \beta_{ij2}t_{ij})$ being the time-varying exponential decay rate. $\mathbf{e}_i = (e_{i1}, \ldots, e_{in_i})^T \sim \mathrm{ST}_{n_i,\nu_e}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i}, \delta_e\mathbf{I}_{n_i})$; $\boldsymbol{\beta}_{ij} = (\beta_{ij0}, \beta_{ij1}, \beta_{ij2})^T$ is a vector of individual parameters for the ith subject at time t_{ij}.

Since CD4 cell counts are measured with errors, we assume that the individual-specific and time-varying parameters $\boldsymbol{\beta}_{ij}$ are related to the summary of true CD4 values z_{ij}^* which may be interpreted as the "regularized" CD4 covariate value. As discussed by Wu [21], to determine whether CD4 values influence the dynamic parameters $\boldsymbol{\beta}_{ij}$, AIC/BIC criteria are used again as guidance (Pinheiro and Bates [16]) to find the following model

$$\beta_{i0} = \beta_1 + b_{i1}, \qquad \beta_{ij1} = \beta_2 + \beta_3 z_{ij}^* + b_{i2}, \qquad \beta_{ij2} = \beta_4 + \beta_5 z_{ij}^* + b_{i3}, \tag{3.3}$$

where $\mathbf{b}_i = (b_{i1}, \ldots, b_{i3})^T$ is individual random-effect, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_5)^T$ is a vector of population parameters. The model (3.3) indicates that the current (regularized) CD4 values z_{ij}^* rather than the past (observed) CD4 values z_{ij} are most predictive of the change in viral load at time t_{ij}. One possible explanation is that, since CD4 measurements for each individual are often sparse, the current CD4 value may be the best summary of immediate past CD4 values, while the early CD4 values may not be very predictive of the current change in viral load.

3.2. Model Implementation

In this section, we analyze the AIDS data set described in Section 1 to illustrate the proposed joint modeling method (denoted by JM) based on the joint models (3.2) in conjunction with the covariate model (3.1) and the corresponding specifications of prior distributions. As shown in Figures 1(a) and 1(b), the histograms of viral load in ln scale and CD4 cell count clearly indicate their asymmetric nature and it seems logical to fit the joint model with a skew distribution to the data set. Along with this consideration, the following statistical models with different distributions of both model errors and random-effects for both the response model (3.2) and the covariate model (3.1) are employed to compare their performance.

(i) **SN Model:** $\mathbf{e}_i$, $\boldsymbol{\epsilon}_i$, $\mathbf{b}_i$, and $\mathbf{a}_i$ follow an SN-distribution.

(ii) **ST Model:** $\mathbf{e}_i$, $\boldsymbol{\epsilon}_i$, $\mathbf{b}_i$, and $\mathbf{a}_i$ follow an ST-distribution.

(iii) **N Model:** $\mathbf{e}_i$, $\boldsymbol{\epsilon}_i$, $\mathbf{b}_i$, and $\mathbf{a}_i$ follow a normal (N) distribution.

We investigate the following *three scenarios*. First, since a normal distribution is a special case of an SN distribution when skewness parameter is zero, while the ST distribution reduces to the SN distribution when the degree of freedom approaches infinity, we investigate how an asymmetric (SN or ST) distribution contributes to modeling results and parameter estimation in comparison with a symmetric (normal) distribution. Second, we estimate the

model parameters by using the "naive" method (denoted by NV), which ignores measurement errors in CD4, and missing responses are imputed by the half (i.e., ln(25)) of the BDL. That is, the "naive" method only uses the observed CD4 values z_{ij} rather than true (unobservable) CD4 values z_{ij}^* in the response model (3.2) and the missing data in the Tobit model (2.1) is imputed by ln(25). We use it as a comparison to the JM proposed in Section 2. This comparison attempts to investigate how the measurement errors in CD4 and missing data in viral load together contribute to modeling results. Third, when covariates are measured with errors, a common approach is the so-called two-step (TS) method (Higgins et al. [1]): the first step estimates the "true" covariate values based on the covariate model (3.1); at the second step the covariate in the response model (2.6) is substituted by the estimate from the first step. Thus we use the two-step (TS) method to assess the performance of the JM method.

The progress in the Bayesian posterior computation due to MCMC procedures has made it possible to fit increasingly complex statistical models (Lunn et al. [17] and Huang et al. [18]). To choose the best model among candidate models, it has become more important to develop efficient model selection criteria. A recent publication by Spiegelhalter et al. [22] suggested a generalization of AIC called deviance information criterion (DIC). Since the structure of DIC allows for an automatic computation in WinBUGS, we use DIC to compare the models in this paper. As with other model selection criteria, we caution that DIC is not intended for identification of the "correct" model, but rather merely as a method of comparing a collection of alternative formulations. In our models with different distribution specifications for model errors, DIC can be used to find out how assumption of a skew-normal distribution contributes to virologic response in comparison with that of a normal distribution and how the proposed joint modeling approach influences parameter estimation compared with the "naive" method and imputation method.

To carry out the Bayesian inference, we need to specify the values of the hyperparameters in the prior distributions. In the Bayesian approach, we only need to specify the priors at the population level. The values of the hyperparameters were mostly chosen from previous studies in the literature (Liu and Wu [2], Huang and Dagne [5], Sahu et al. [7], Wu [21], and among others). We take weakly informative prior distribution for the parameters in the models. In particular, (i) fixed-effects were taken to be independent normal distribution $N(0, 100)$ for each component of the population parameter vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. (ii) For the scale parameters τ^2 and σ^2, we assume a limiting noninformative inverse gamma prior distribution, IG$(0.01, 0.01)$ so that the distribution has mean 1 and variance 100. (iii) The priors for the variance-covariance matrices of the random-effects $\boldsymbol{\Sigma}_a$ and $\boldsymbol{\Sigma}_b$ are taken to be inverse Wishart distributions IW$(\boldsymbol{\Omega}_1, \rho_1)$ and IW$(\boldsymbol{\Omega}_2, \rho_2)$ with covariance matrices $\boldsymbol{\Omega}_1 = \boldsymbol{\Omega}_2 = \text{diag}(0.01, 0.01, 0.01)$ and degrees of freedom $\rho_1 = \rho_2 = 5$, respectively. (iv) The degrees of freedom parameters ν_ϵ, ν_e, ν_a, and ν_b follow a truncated gamma distribution with two hyperparameter values being 1 and 0.1, respectively. (v) For each of the skewness parameters δ_e, δ_ϵ, δ_k^a, and δ_k^b $(k = 1, 2, 3)$, we choose independent normal distribution $N(0, 100)$, where we assume that $\boldsymbol{\delta}_{e_i} = \delta_e \mathbf{1}_{n_i}$ and $\boldsymbol{\delta}_{\epsilon_i} = \delta_\epsilon \mathbf{1}_{n_i}$ to indicate that we are interested in skewness of overall viral load data and overall CD4 cell count data. The MCMC sampler was implemented using WinBUGS software, and the program codes are available from authors on request. The convergence of MCMC implementation was assessed using standard tools (such as trace plots which are not shown here to save space) within WinBUGS, and convergence was achieved after initial 50,000 burn-in iterations. After convergence diagnostics was done, one long chain of 200,000 iterations, retaining every 20th, was run to obtain 10,000 samples for Bayesian inference. Next, we report analysis results of the three scenarios proposed above.

3.3. Comparison of Joint Modeling Results

The population posterior mean (PM), the corresponding standard deviation (SD), and 95% credible interval for fixed-effects parameters based on the three models (SN, ST, and N) for JM method are presented in the upper part of Table 1. The significant findings are presented as follows. (i) For the response model (3.2), where the most substantively interesting parameters are $(\beta_2, \beta_3, \beta_4, \beta_5)$, the estimates of β_2 and β_4, the linear coefficient and quadratic coefficient of time, respectively, under the three models, are significant since the 95% credible intervals do not contain zero. Among the coefficients of the true CD4 covariate (β_3, β_5) in model (3.3), the posterior means of β_5 are significantly different from zero for all the three models under JM method. Moreover, the posterior mean values for β_5 are quite different between models SN (−4.76), ST (−6.31), and N (−6.26), implying that the posterior means may be substantially biased if model distribution ignores skewness. We will see later that SN gives better fit than either ST or N. In addition, for the scale parameter σ^2, the posterior mean value (2.63) in N model is much larger than that of any other corresponding posterior means in SN and ST models. (ii) For parameter estimates of the CD4 covariate model (3.1), the posterior means of intercept α_0 and coefficient α_1 based on SN and ST models are significant, while the posterior mean of α_2 turns out to be nonsignificant under all the three models. For the scale parameter τ^2 of the covariate model, the posterior mean value (0.13) is the largest under N model. This is expected since the model based on ordinary normal distribution does not account for skewness and heaviness in tails for the type of data analyzed here.

To assess the goodness-of-fit of the proposed JM method, the diagnosis plots for the SN, ST, and N models comparing the residuals and the fitted values (Figures 2(a)–2(c)) and the observed values versus the fitted values (Figures 2(d)–2(f)). The distribution of the residuals for SN model looks tighter than those for either ST model or N model, showing a better fit. Similar results are observed by looking at the plots in Figures 2(d)–2(f). The plot for SN model has most of the points close the line showing a strong agreement between the observed and the fitted values. Clearly, it can be seen from the plots that N model, which ignores skewness, does not fit the data very well as compared to either SN model or ST model. Note that the horizontal line designates the below detection limit (BDL), which is at ln(50). The recorded observations less than BDL are not accurate and, therefore, have not been used in the analysis, but instead they were treated as missing and predicted values are obtained. These predicted values are plotted against the recorded observations below detection limit as shown in the lower-row plots. In general, from the model fitting results, both SN and ST models provide a reasonably good fit to the observed data even though SN model is slightly better than ST model.

In order to further investigate whether SN model under JM method can provide better fit to the data than ST model, the DIC values are obtained and found to be 863.0 for SN model and 985.6 for ST model. The DIC value for SN model is smaller than that of ST model, confirming that SN model is better than ST model in fitting the proposed joint model. As mentioned before, it is hard sometimes to tell which model is "correct" but which one fits data better. The model which fits the data better may be more appealing in order to describe the mechanism of HIV infection and CD4 changing process. Thus, based on the DIC criterion, the results indicate that SN model is relatively better than either ST model or N model. These findings are consistent with those displayed in the goodness-of-fit in Figure 2 indicating that SN model outperforms both ST model and N model. In summary, our results suggest that it is very important to assume an SN distribution for the response Tobit model and the CD4 covariate model in order to achieve reliable results, in particular if the data exhibit skewness,

Table 1: A summary of the estimated posterior mean (PM) of population (fixed-effects) and scale parameters, the corresponding standard deviation (SD) and lower limit (L_{CI}) and upper limit (U_{CI}) of 95% equal-tail credible interval (CI) as well as DIC values based on the joint modeling (JM), the naive (NV), and the two-step (TS) methods.

Method	Model		α_0	α_1	α_2	β_1	β_2	β_3	β_4	β_5	τ^2	σ^2	DIC
JM	SN	PM	−0.95	0.15	−0.23	5.62	−14.6	−2.34	11.7	−4.76	0.07	0.14	863.0
		L_{CI}	−1.58	0.06	−15.2	4.17	−22.1	−5.14	4.52	−9.92	0.04	0.01	
		U_{CI}	−0.01	0.90	14.8	7.59	−8.14	1.44	21.7	−0.62	0.11	0.64	
		SD	0.47	0.37	7.63	0.96	3.98	1.65	5.25	2.34	0.02	0.18	
	ST	PM	−0.94	0.34	−0.31	5.84	−12.0	−1.20	8.12	−6.31	0.04	0.21	985.6
		L_{CI}	−1.41	0.18	−14.1	4.15	−16.5	−5.72	2.20	−12.6	0.02	0.01	
		U_{CI}	−0.06	0.88	13.4	8.02	−7.72	2.72	19.2	−1.41	0.05	0.86	
		SD	0.35	0.26	7.09	1.10	2.22	2.22	4.14	2.77	0.01	0.26	
	N	PM	−0.21	0.45	−2.87	7.74	−15.4	−0.80	13.6	−6.26	0.13	2.63	1242.3
		L_{CI}	−0.46	0.22	−15.9	7.20	−18.3	−4.16	9.97	−11.7	0.11	2.06	
		U_{CI}	0.04	0.68	9.90	8.29	−12.6	2.53	17.2	−1.43	0.16	3.35	
		SD	0.13	0.12	6.54	0.28	1.48	1.73	1.85	2.61	0.01	0.33	
NV	SN	PM	—	—	—	5.03	−11.1	0.58	6.83	−2.10	—	0.10	1083.5
		L_{CI}	—	—	—	3.82	−13.6	−0.94	4.52	−4.18	—	0.01	
		U_{CI}	—	—	—	6.59	−8.73	2.08	9.18	0.07	—	0.35	
		SD	—	—	—	0.75	1.23	0.77	1.19	1.04	—	0.09	
TS	SN	PM	−0.99	0.19	2.71	5.91	−14.4	−1.24	8.47	−5.90	0.09	0.14	1023.8
		L_{CI}	−1.58	−0.43	−12.1	4.12	−22.1	−5.01	1.83	−10.6	0.05	0.01	
		U_{CI}	0.07	0.90	17.1	7.72	−8.50	2.16	21.2	−0.80	0.14	0.65	
		SD	0.42	0.36	7.54	1.05	3.88	1.79	5.14	2.52	0.02	0.18	

but not heaviness in the tails. Along with these observations, next we provide detailed fitting results and interpretations based on the SN Model.

3.4. Estimation Results Based on SN Model

For comparison, we used the "naive" (NV) method to estimate the model parameters presented in the lower part of Table 1 where the raw (observed) CD4 values z_{ij} rather than the true (unobserved) CD4 values z_{ij}^* are substituted in the response model (3.3). It can be seen that there are important differences in the posterior means for the parameters β_3 and β_5, which are coefficients of CD4 covariate. These posterior means are $\widehat{\beta}_3 = 0.58$ and $\widehat{\beta}_5 = -2.10$ for the NV method, and $\widehat{\beta}_3 = -2.34$ and $\widehat{\beta}_5 = -4.76$ for the JM method. The NV method may produce biased posterior means and may substantially overestimate the covariate CD4 effect. The estimated standard deviations (SD) for the CD4 effect (β_3 and β_5) using the JM method are 1.65 and 2.34, which are approximately twice as large as those (0.77 and 1.04) using the NV method, respectively, probably because the JM method incorporates the variation from fitting the CD4 process. The differences of the NV estimates and the JM estimates suggest that the estimated parameters may be substantially biased if measurement errors in CD4 covariate are ignored. We also obtained DIC value of 1083.5 for the NV method, while the DIC value for the JM method is 863.0. We can see from the estimated DIC values that the JM approach provides

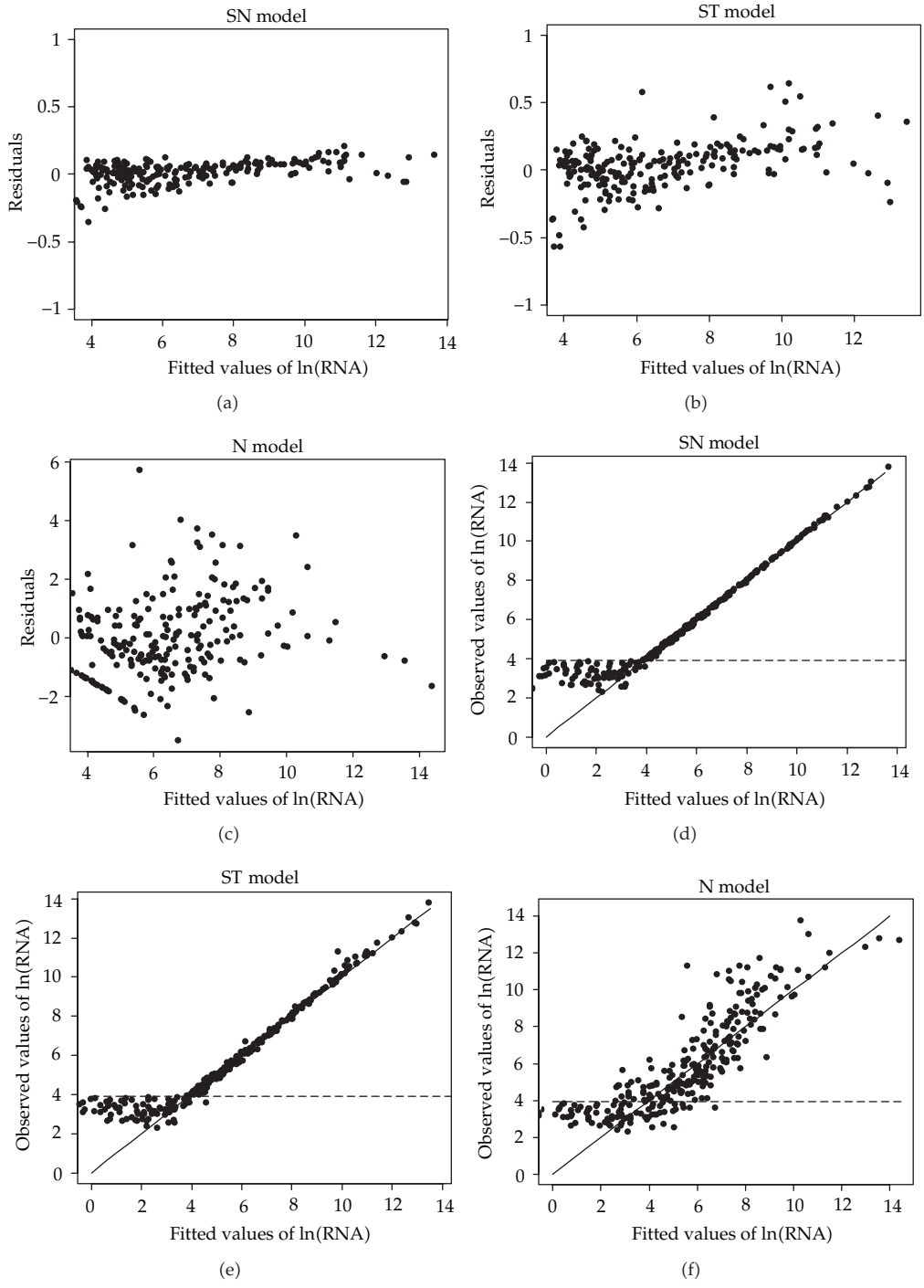


Figure 2: The goodness-of-fit. (a–c): Residuals versus fitted values of ln(RNA) under skew-normal (SN), skew-t (ST), and normal (N) models based on the JM method; the values below detection limit (ln(50)) are not included in the plots since there are no corresponding residuals but only predicted values. (d–f): Observed values versus fitted values of ln(RNA) under SN, ST, and N models, where the horizontal line at ln(50) represents the detection limit.

Table 2: A summary of the estimated posterior mean (PM) of skewness and degree of freedom parameters, the corresponding standard deviation (SD), and lower limit (L_{CI}) and upper limit (U_{CI}) of 95% equal-tail credible interval (CI) based on the joint modeling (JM), the naive (NV), and the two-step (TS) methods.

Method	Model		δ_e	δ_e	δ_1^a	δ_2^a	δ_3^a	δ_1^b	δ_2^b	δ_3^b	ν_e	ν_e	ν_a	ν_b
JM	SN	PM	0.41	2.34	0.58	0.34	0.26	0.52	−1.81	2.58	—	—	—	—
		L_{CI}	0.25	1.93	−0.62	−0.58	−16.1	−1.90	−10.1	−10.7	—	—	—	—
		U_{CI}	0.54	2.73	1.41	1.09	17.0	2.31	7.12	11.0	—	—	—	—
		SD	0.07	0.20	0.60	0.47	8.57	1.18	5.30	6.93	—	—	—	—
	ST	PM	0.05	2.26	0.87	0.04	−0.56	0.32	−4.70	6.45	3.32	10.2	14.0	14.4
		L_{CI}	−0.14	1.59	−0.35	−0.60	−16.5	−2.23	−10.1	−8.23	3.01	3.07	3.52	3.52
		U_{CI}	0.25	2.70	1.54	0.65	14.7	2.37	1.91	12.3	4.18	35.2	41.1	41.9
		SD	0.11	0.33	0.50	0.32	8.41	1.31	2.93	4.88	0.32	8.98	10.3	10.3
NV	SN	PM	—	2.24	—	—	—	0.80	0.15	5.53	—	—	—	—
		L_{CI}	—	1.95	—	—	—	−1.05	−1.71	3.62	—	—	—	—
		U_{CI}	—	2.55	—	—	—	2.25	2.30	7.74	—	—	—	—
		SD	—	0.15	—	—	—	0.92	1.00	1.06	—	—	—	—
TS	SN	PM	0.16	2.44	0.89	0.28	3.06	0.04	−0.94	5.18	—	—	—	—
		L_{CI}	−0.39	2.07	−0.48	−0.58	−11.7	−2.20	−8.53	−12.4	—	—	—	—
		U_{CI}	0.51	2.79	1.55	1.04	21.0	2.23	7.49	12.2	—	—	—	—
		SD	0.29	0.18	0.50	0.45	8.30	1.31	4.75	6.67	—	—	—	—

a better fit to the data in comparison with the NV method. Thus it is important to take the measurement errors into account when covariates are measured with errors.

Comparing the JM method against the two-step (TS) method from the lower part of Table 1, we can see that the TS estimates and the JM estimates are somewhat different. In particular, there are important differences in the posterior means for the parameters β_4 and β_5 which is directly associated CD4 covariate. For the parameter β_5, the posterior means are -4.76 (95% CI $= (-9.92, -0.62)$) and -5.90 (95% CI $= (-10.60, -0.80)$) for the JM and TS methods, respectively. The TS method slightly underestimates the effect of CD4 covariate.

The estimated results based on the JM method for SN model in Table 2 presents the estimated skewness parameters, and the only significant skewness parameters are those for the response model errors and CD4 covariate model errors, but not random-effects. These are $\widehat{\delta}_e = 2.34$ (95% CI $= (1.93, 2.73)$) and $\widehat{\delta}_e = 0.41$ (95% CI $= (0.25, 0.54)$) for viral load and CD4 cell count, respectively. They are significantly positive confirming the right-skewed viral load and CD4 cell count as was depicted in Figure 1. Thus, the results suggest that accounting for significant skewness, when the data exhibit skewness, provides a better model fit to the data and gives more accurate estimates to the parameters.

In summary, the results indicate that the SN model under the JM method is a better suited model for viral loads and CD4 covariate with measurement errors. Looking now at the estimated population initial stage of viral decay after treatment bases on the JM method, we get $\widehat{\lambda}(t) = -(-14.6 - 2.34z^*(t) + 11.7t - 4.76z^*(t)t)$, where $z^*(t)$ is the standardized true CD4 value at time t which may be interpreted as the "regularized" covariate value. Thus, the population viral load process may be approximated by $\widehat{V}(t) = \exp[5.62 - \widehat{\lambda}(t)t]$. Since the viral decay rate ($\lambda(t)$) is significantly associated with the true CD4 values (due to statistically significant estimate of β_5), this suggests that the viral load change $V(t)$ may be significantly associated with the true CD4 process. Note that, although the true association described

above may be complicated, the simple approximation considered here may provide a reasonable guidance and suggest a further research.

4. Discussion

Attempts to jointly fit the viral load data and CD4 cell counts with measurement errors are compromised by left censoring in viral load response due to detection limits. We addressed this problem using Bayesian nonlinear mixed-effects Tobit models with skew distributions. The models were fitted based on the assumption that the viral dynamic model (2.2) continues to hold for those unobserved left-censored viral loads. This assumption may be reasonable since the dynamic model considered here is a natural extension of a biologically justified model (Ding and Wu [10]). Even though left censoring effects are the focus of this paper, right-censoring (ceiling) effects can also be dealt with in very similar ways. It is therefore important for researchers to pay attention to censoring effects in a longitudinal data analysis, and Bayesian Tobit models with skew distributions make best use of both censored and uncensored data information.

Our results suggest that both ST (skew-t) and SN (skew-normal) models show superiority to the N (normal) model. Our results also indicate that the JM method outperformed the NV and TS methods in the sense that it produces more accurate parameter estimates. The JM method is quite general and so can be applied to other application areas, allowing accurate inferences of parameters while adjusting for skewness, left-censoring, and measurement errors. In short, skew distributions show potentials to gain efficiency and accuracy in estimating certain parameters when the normality assumption does not hold in the data.

The proposed NLME Tobit joint model with skew distributions can be easily fitted using MCMC procedure by using the WinBUGS package that is available publicly and has a computational cost similar to the normal version of the model due to the features of its hierarchically stochastic representations. Implementation via MCMC makes it straightforward to compare the proposed models and methods with various scenarios for real data analysis in comparison with symmetric distributions and asymmetric distributions for model errors. This makes our approach quite powerful and also accessible to practicing statisticians in the fields. In order to examine the sensitivity of parameter estimates to the prior distributions and initial values, we also conducted a limited sensitivity analysis using different values of hyperparameters of prior distributions and different initial values (data not shown). The results of the sensitivity analysis showed that the estimated dynamic parameters were not sensitive to changes of both priors and initial values. Thus, the final results are reasonable and robust, and the conclusions of our analysis remain unchanged (see Huang et al. [18] for more details).

The methods of this paper may be extended to accommodate various subpopulations of patients whose viral decay trajectories after treatment may differ. In addition, the purpose of this paper is to demonstrate the proposed models and methods with various scenarios for real data analysis for comparing asymmetric distributions for model errors to a symmetric distribution, although a limited simulation study might have been conducted to evaluate our results from different model specifications and the corresponding methods. However, since this paper investigated many different scenarios-based models and methods with real data analysis, the complex natures considered, especially skew distributions involved, will pose some challenges for such a simulation study which requires additional efforts, and it is beyond the purpose of this paper. We are currently investigating these related problems and will report the findings in the near future.

Appendix

A. Multivariate Skew-t and Skew Normal Distributions

Different versions of the multivariate skew-elliptical (SE) distributions have been proposed and used in the literature (Sahu et al. [7], Azzalini and Capitanio [23], Jara et al. [24], Arellano-Valle et al. [25], and among others). We adopt a class of multivariate SE distributions proposed by Sahu et al. [7], which is obtained by using transformation and conditioning and contains multivariate skew-t (ST) and skew-normal (SN) distributions as special cases. A k-dimensional random vector $\mathbf{Y}$ follows a k-variate SE distribution if its probability density function (pdf) is given by

$$f\left(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}(\boldsymbol{\delta}); m_{\nu}^{(k)}\right) = 2^k f\left(\mathbf{y} \mid \boldsymbol{\mu}, \mathbf{A}; m_{\nu}^{(k)}\right) P(\mathbf{V} > \mathbf{0}), \tag{A.1}$$

where $\mathbf{A} = \boldsymbol{\Sigma} + \boldsymbol{\Delta}^2(\boldsymbol{\delta})$, $\boldsymbol{\mu}$ is a location parameter vector, $\boldsymbol{\Sigma}$ is a covariance matrix, $\boldsymbol{\Delta}(\boldsymbol{\delta})$ is a skewness diagonal matrix with the skewness parameter vector $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_k)^T$; $\mathbf{V}$ follows the elliptical distribution $El(\boldsymbol{\Delta}(\boldsymbol{\delta})A^{-1}(\mathbf{y}-\boldsymbol{\mu}), \mathbf{I}_k - \boldsymbol{\Delta}(\boldsymbol{\delta})A^{-1}\boldsymbol{\Delta}(\boldsymbol{\delta}); m_{\nu}^{(k)})$ and the density generator function $m_{\nu}^{(k)}(u) = (\Gamma(k/2)/\pi^{k/2})(m_{\nu}(u)/\int_0^{\infty} r^{k/2-1} m_{\nu}(u)dr)$, with $m_{\nu}(u)$ being a function such that $\int_0^{\infty} r^{k/2-1} m_{\nu}(u)dr$ exists. The function $m_{\nu}(u)$ provides the kernel of the original elliptical density and may depend on the parameter ν. We denote this SE distribution by $\mathrm{SE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}(\boldsymbol{\delta}); m^{(k)})$. Two examples of $m_{\nu}(u)$, leading to important special cases used throughout the paper, are $m_{\nu}(u) = \exp(-u/2)$ and $m_{\nu}(u) = (u/\nu)^{-(\nu+k)/2}$, where $\nu > 0$. These two expressions lead to the multivariate ST and SN distributions, respectively. In the latter case, ν corresponds to the degree of freedom parameter.

Since the SN distribution is a special case of the ST distribution when the degree of freedom approaches infinity, for completeness, this section is started by discussing the multivariate ST distribution that will be used in defining the ST joint models considered in this paper. For detailed discussions on properties and differences among various versions of ST and SN distributions, see the references above. We consider a multivariate ST distribution introduced by Sahu et al. [7], which is suitable for a Bayesian inference since it is built using conditional method and is defined below.

An k-dimensional random vector $\mathbf{Y}$ follows an k-variate ST distribution if its pdf is given by

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}(\boldsymbol{\delta}), \nu) = 2^k t_{k,\nu}(\mathbf{y} \mid \boldsymbol{\mu}, \mathbf{A}) P(\mathbf{V} > \mathbf{0}). \tag{A.2}$$

We denote the k-variate t distribution with parameters $\boldsymbol{\mu}, \mathbf{A}$ and degrees of freedom ν by $t_{k,\nu}(\boldsymbol{\mu}, \mathbf{A})$ and the corresponding pdf by $t_{k,\nu}(\mathbf{y} \mid \boldsymbol{\mu}, \mathbf{A})$ henceforth, $\mathbf{V}$ follows the t distribution $t_{k,\nu+k}$. We denote this distribution by $\mathrm{ST}_{k,\nu}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}(\boldsymbol{\delta}))$. In particular, when $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_k$ and $\boldsymbol{\Delta}(\boldsymbol{\delta}) = \delta \mathbf{I}_k$, (A.2) simplifies to

$$\begin{aligned} f\left(\mathbf{y} \mid \boldsymbol{\mu}, \sigma^2, \delta, \nu\right) &= 2^k \left(\sigma^2 + \delta^2\right)^{-k/2} \frac{\Gamma((\nu+m)/2)}{\Gamma(\nu/2)(\nu\pi)^{k/2}} \left\{1 + \frac{(\mathbf{y}-\boldsymbol{\mu})^T(\mathbf{y}-\boldsymbol{\mu})}{\nu(\sigma^2+\delta^2)}\right\}^{-(\nu+k)/2} \\ &\quad \times T_{k,\nu+k}\left[\left\{\frac{\nu + \left(\sigma^2+\delta^2\right)^{-1}(\mathbf{y}-\boldsymbol{\mu})^T(\mathbf{y}-\boldsymbol{\mu})}{\nu+k}\right\}^{-1/2} \frac{\delta(\mathbf{y}-\boldsymbol{\mu})}{\sigma\sqrt{\sigma^2+\delta^2}}\right], \end{aligned} \tag{A.3}$$

where $T_{k,\nu+k}(\cdot)$ denotes the cumulative distribution function (cdf) of $t_{k,\nu+k}(\mathbf{0}, \mathbf{I}_k)$. However, unlike in the SN distribution to be discussed below, the ST density cannot be written as the product of univariate ST densities. Here $\mathbf{Y}$ are dependent but uncorrelated.

The mean and covariance matrix of the ST distribution $ST_{k,\nu}(\boldsymbol{\mu}, \sigma^2\mathbf{I}_k, \boldsymbol{\Delta}(\boldsymbol{\delta}))$ are given by

$$\begin{aligned} E(\mathbf{Y}) &= \boldsymbol{\mu} + \left(\frac{\nu}{\pi}\right)^{1/2} \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \boldsymbol{\delta}, \\ \operatorname{cov}(\mathbf{Y}) &= \left[\sigma^2\mathbf{I}_k + \boldsymbol{\Delta}^2(\boldsymbol{\delta})\right] \frac{\nu}{\nu-2} - \frac{\nu}{\pi}\left[\frac{\Gamma\{(\nu-1)/2\}}{\Gamma(\nu/2)}\right]^2 \boldsymbol{\Delta}^2(\delta). \end{aligned} \tag{A.4}$$

According to Lemma 1 of Azzalini and Capitanio [23], if $\mathbf{Y}$ follows $\mathrm{ST}_{k,\nu}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}(\boldsymbol{\delta}))$, it can be represented by

$$\mathbf{Y} = \boldsymbol{\mu} + \xi^{-1/2}\mathbf{X}, \tag{A.5}$$

where ξ follows a Gamma distribution $\Gamma(\nu/2, \nu/2)$, which is independent of $\mathbf{X}$, and $\mathbf{X}$ follows a k-dimensional skew-normal (SN) distribution, denoted by $\mathrm{SN}_k(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}(\boldsymbol{\delta}))$. It follows from (A.5) that $\mathbf{Y} \mid \xi \sim \mathrm{SN}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}/\xi, \boldsymbol{\Delta}(\boldsymbol{\delta}))$. By Proposition 1 of Arellano-Valle et al. [25], the SN distribution of $\mathbf{Y}$ conditional on ξ has a convenient stochastic representation as follows:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Delta}(\boldsymbol{\delta})|X_0| + \xi^{-1/2}\boldsymbol{\Sigma}^{1/2}X_1, \tag{A.6}$$

where X_0 and X_1 are two independent $N_k(\mathbf{0}, \mathbf{I}_k)$ random vectors. Note that the expression (A.6) provides a convenience device for random number generation and for implementation purpose. Let $\mathbf{w} = |X_0|$; then $\mathbf{w}$ follows a k-dimensional standard normal distribution $N_k(\mathbf{0}, \mathbf{I}_m)$ truncated in the space $\mathbf{w} > \mathbf{0}$ (i.e., the standard half-normal distribution). Thus, following Sahu et al. [7], a hierarchical representation of (A.6) is given by

$$\mathbf{Y} \mid \mathbf{w}, \xi \sim N_k\left(\boldsymbol{\mu} + \boldsymbol{\Delta}(\boldsymbol{\delta})\mathbf{w}, \xi^{-1}\boldsymbol{\Sigma}\right), \qquad \mathbf{w} \sim N_k(\mathbf{0}, \mathbf{I}_k)\mathbf{I}(\mathbf{w} > \mathbf{0}), \qquad \xi \sim G\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \tag{A.7}$$

where $G(\cdot)$ is a gamma distribution. Note that the ST distribution presented in (A.7) can be reduced to the following three special cases: (i) as $\nu \to \infty$ and $\xi \to 1$ with probability 1 (i.e., the last distributional specification is omitted), then the hierarchical expression (A.7) becomes an SN distribution $\mathrm{SN}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}(\boldsymbol{\delta}))$; (ii) as $\boldsymbol{\Delta}(\boldsymbol{\delta}) = \mathbf{0}$, then the hierarchical expression (A.7) is a standard multivariate t-distribution; (iii) as $\nu \to \infty, \xi \to 1$ with probability 1, and $\boldsymbol{\Delta}(\boldsymbol{\delta}) = \mathbf{0}$, then the hierarchical expression (A.7) is a standard multivariate normal distribution.

Specifically, if a k-dimensional random vector $\mathbf{Y}$ follows a k-variate SN distribution, then (A.2)–(A.4) revert to the following forms, respectively:

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}(\boldsymbol{\delta})) = 2^k|\mathbf{A}|^{-1/2}\phi_k\left\{\mathbf{A}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})\right\}P(\mathbf{V} > \mathbf{0}), \tag{A.8}$$

where $\mathbf{V} \sim N_k\{\boldsymbol{\Delta}(\boldsymbol{\delta})\mathbf{A}^{-1}(\mathbf{y} - \boldsymbol{\mu}),\ \mathbf{I}_k - \boldsymbol{\Delta}(\boldsymbol{\delta})\mathbf{A}^{-1}\boldsymbol{\Delta}(\boldsymbol{\delta})\}$, and $\phi_k(\cdot)$ is the pdf of $N_k(\mathbf{0}, \mathbf{I}_k)$. We denote the above distribution by $\mathrm{SN}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}(\boldsymbol{\delta}))$. An appealing feature of (A.8) is that it

gives independent marginal when $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)$. The pdf (A.8) thus simplifies to

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}(\boldsymbol{\delta})) = \prod_{i=1}^{k} \left[\frac{2}{\sqrt{\sigma_i^2 + \delta_i^2}} \phi \left\{ \frac{y_i - \mu_i}{\sqrt{\sigma_i^2 + \delta_i^2}} \right\} \Phi \left\{ \frac{\delta_i}{\sigma_i} \frac{y_i - \mu_i}{\sqrt{\sigma_i^2 + \delta_i^2}} \right\} \right], \tag{A.9}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively. The mean and covariance matrix are given by $E(\mathbf{Y}) = \boldsymbol{\mu} + \sqrt{2/\pi}\boldsymbol{\delta}$, $\operatorname{cov}(\mathbf{Y}) = \boldsymbol{\Sigma} + (1 - 2/\pi)\boldsymbol{\Delta}(\boldsymbol{\delta})^2$. It is noted that when $\boldsymbol{\delta} = \mathbf{0}$, the SN distribution reduces to usual normal distribution.

Acknowledgments

The authors are grateful to the Guest Editor and three reviewers for their insightful comments and suggestions that led to a marked improvement of the paper. They gratefully acknowledge A5055 study investigators for allowing them to use the clinical data from their study. This research was partially supported by NIAID/NIH Grant R03 AI080338 and MSP/NSA Grant H98230-09-1-0053 to Y. Huang.

References

[1] M. Higgins, M. Davidian, and D. M. Gilitinan, "A two-step approach to measurement error in time-dependent covariates in nonlinear mixed-effects models, with application to IGF-I pharmacokinetics," *Journal of the American Statistical Association*, vol. 92, pp. 436–448, 1997.

[2] W. Liu and L. Wu, "Simultaneous inference for semiparametric nonlinear mixed-effects models with covariate measurement errors and missing responses," *Biometrics*, vol. 63, no. 2, pp. 342–350, 2007.

[3] M. S. Wulfsohn and A. A. Tsiatis, "A joint model for survival and longitudinal data measured with error," *Biometrics*, vol. 53, no. 1, pp. 330–339, 1997.

[4] L. Wu, "A joint model for nonlinear mixed-effects models with censoring and covariants measured with error, with application to AIDS studies," *Journal of the American Statistical Association*, vol. 97, no. 460, pp. 955–964, 2002.

[5] Y. Huang and G. Dagne, "A Bayesian approach to joint mixed-effects models with a skew-normal distribution and measurement errors in covariates," *Biometrics*, vol. 67, no. 1, pp. 260–269, 2011.

[6] G. Verbeke and E. Lesaffre, "A linear mixed-effects model with heterogeneity in the random-effects population," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 217–221, 1996.

[7] S. K. Sahu, D. K. Dey, and M. D. Branco, "A new class of multivariate skew distributions with applications to Bayesian regression models," *The Canadian Journal of Statistics*, vol. 31, no. 2, pp. 129–150, 2003.

[8] E. P. Acosta, H. Wu, S. M. Hammer et al., "Comparison of two indinavir/ritonavir regimens in the treatment of HIV-infected individuals," *Journal of Acquired Immune Deficiency Syndromes*, vol. 37, no. 3, pp. 1358–1366, 2004.

[9] A. S. Perelson, P. Essunger, Y. Cao et al., "Decay characteristics of HIV-1-infected compartments during combination therapy," *Nature*, vol. 387, no. 6629, pp. 188–191, 1997.

[10] A. A. Ding and H. Wu, "Relationships between antiviral treatment effects and biphasic viral decay rates in modeling HIV dynamics," *Mathematical Biosciences*, vol. 160, no. 1, pp. 63–82, 1999.

[11] M. Davidian and D. M. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Chapman & Hall, London, UK, 1995.

[12] J. P. Hughes, "Mixed effects models with censored data with application to HIV RNA levels," *Biometrics*, vol. 55, no. 2, pp. 625–629, 1999.

[13] J. Tobin, "Estimation of relationships for limited dependent variables," *Econometrica*, vol. 26, pp. 24–36, 1958.

[14] R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu, *Measurement Error in Nonlinear Models: A Modern Perspective*, Chapman & Hall, Boca Raton, Fla, USA, 2nd edition, 2006.

[15] H. Wu and J. T. Zhang, "The study of long-term HIV dynamics using semi-parametric non-linear mixed-effects models," *Statistics in Medicine*, vol. 21, no. 23, pp. 3655–3675, 2002.
[16] J. C. Pinheiro and M. D. Bates, *Mixed-Effects Models in S and S-PLUS*, Springer, New York, NY, USA, 2000.
[17] D. J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter, "WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility," *Statistics and Computing*, vol. 10, no. 4, pp. 325–337, 2000.
[18] Y. Huang, D. Liu, and H. Wu, "Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system," *Biometrics*, vol. 62, no. 2, pp. 413–423, 2006.
[19] D. B. Rubin, "Inference and missing data," *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976.
[20] D. D. Ho, A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard, and M. Markowitz, "Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection," *Nature*, vol. 373, no. 6510, pp. 123–126, 1995.
[21] L. Wu, "Simultaneous inference for longitudinal data with detection limits and covariates measured with errors, with application to AIDS studies," *Statistics in Medicine*, vol. 23, no. 11, pp. 1715–1731, 2004.
[22] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van der Linde, "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society. Series B*, vol. 64, no. 4, pp. 583–639, 2002.
[23] A. Azzalini and A. Capitanio, "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t*-distribution," *Journal of the Royal Statistical Society. Series B*, vol. 65, no. 2, pp. 367–389, 2003.
[24] A. Jara, F. Quintana, and E. San Martín, "Linear mixed models with skew-elliptical distributions: a Bayesian approach," *Computational Statistics & Data Analysis*, vol. 52, no. 11, pp. 5033–5045, 2008.
[25] R. B. Arellano-Valle, H. Bolfarine, and V. H. Lachos, "Bayesian inference for skew-normal linear mixed models," *Journal of Applied Statistics*, vol. 34, no. 5-6, pp. 663–682, 2007.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2012, Article ID 734341, 15 pages
doi:10.1155/2012/734341

Research Article

A Semiparametric Marginalized Model for Longitudinal Data with Informative Dropout

Mengling Liu[1] and Wenbin Lu[2]

[1] *Division of Biostatistics, Department of Environmental Medicine, New York University School of Medicine, New York, NY 10016, USA*
[2] *Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA*

Correspondence should be addressed to Mengling Liu, mengling.liu@nyu.edu

Received 1 April 2011; Revised 11 July 2011; Accepted 27 July 2011

Academic Editor: Yangxin Huang

We propose a marginalized joint-modeling approach for marginal inference on the association between longitudinal responses and covariates when longitudinal measurements are subject to informative dropouts. The proposed model is motivated by the idea of linking longitudinal responses and dropout times by latent variables while focusing on marginal inferences. We develop a simple inference procedure based on a series of estimating equations, and the resulting estimators are consistent and asymptotically normal with a sandwich-type covariance matrix ready to be estimated by the usual plug-in rule. The performance of our approach is evaluated through simulations and illustrated with a renal disease data application.

1. Introduction

Longitudinal studies often encounter data attrition because subjects drop out before the designated study end. Both statistical analysis and practical interpretation of longitudinal data can be complicated by dropouts. For example, in the Modification of Diet in Renal Disease (MDRD) study [1, 2], one main interest was to investigate the efficacy of interventions of blood pressure control and diet modification on patients with impaired renal functions. The primary outcome was glomerular filtration rate (GFR), which measured filtering capacity of kidneys, and was repeatedly measured over the study period. However, some patients could leave the study prematurely for kidney transplant or dialysis, which precluded further GFR measurements. This resulted in a dropout mechanism that could relate to patients' kidney function and correlate with their GFR values. Other patients were followed to the end of the study or dropped out due to independent reasons. Thus, statistical analysis of longitudinal

GFR needs to take into consideration the presence of mixed types of informative and independent dropouts.

Many statistical models and inference approaches have been proposed to accommodate the nonignorable missingness into modeling longitudinal data (see reviews [3–8]). According to the target of inference and the interpretation of model parameters, existing methods can be classified into three categories: subject-specific inference, event-conditioning inference, and marginal inference. First, a widely used modeling strategy for longitudinal data with informative dropouts is to specify their joint distribution via shared or correlated latent variables. Under such model assumptions, the longitudinal parameters have a conditional, subject-specific interpretation (e.g., [9–11]). But the interpretation of longitudinal parameters usually changes with the number and characteristics of latent variables assumed, for example, a single random intercept versus a random intercept plus a random slope.

Second, event-conditioning approaches have also been widely used when the target of inference is within subgroups of patients with particular dropout patterns or when the dropout can potentially change the material characteristic of the longitudinal process (e.g., death). The inference is usually conducted conditioning on the dropout pattern or on the occurrence of the dropout event. Thus, model parameters have an event-conditioning subpopulation-averaged interpretation, for example, pattern-mixture models for the group expectation of each dropout pattern [3, 12]; treatment effects among survivors [13]; gender and age effects in mortal cohort [14]. Because the interpretation of such models is made by conditioning on a future event, event-conditioning approaches may be natural in a retrospective setting but may not be directly useful for the evaluation of treatment efficacy prospectively.

Lastly, when the research objective is to study covariate effects at population level in a dropout-free situation, marginal models address this concern directly. When data are without missing or missing completely at random (using Rubin's definition on missingness [15]), the estimation of model parameters can be carried out by the generalized estimating equation (GEE) approach assuming a "working" correlation matrix [16]. When dropouts are missing at random, the inverse probability-weighted GEE methods are commonly used [17, 18]. In the presence of informative dropouts, the class of selection models that were originally proposed to adjust selection bias in econometrics [19] have been widely used for the marginal analysis of longitudinal data [20–22]. Recently, the marginalized transition model [23] and marginalized pattern-mixture model [24] were proposed for binary longitudinal data with finite nonignorable nonresponse patterns. These marginalized approaches provide a powerful tool for studying the marginal association between longitudinal outcomes and covariates while incorporating nonignorable nonresponses.

In this paper, we shall adopt the idea of shared latent variables to account for the dependence between longitudinal responses and informative dropouts while focusing on marginal inference for the longitudinal responses. Here dropouts can occur on a continuous time scale. We develop an effective estimation procedure built on a series of asymptotically unbiased estimating equations with light computational burden. The resulting estimators for longitudinal parameters are shown to be consistent and asymptotically normal, with a sandwich-type variance-covariance matrix that can be estimated by the usual plug-in rule.

The remainder of the paper is organized as follows. In Section 2, we introduce the notation and the proposed semiparametric marginalized model. In Section 3, a simple estimating equation-based procedure is first proposed for the situation with pure informative dropouts and is extended to a more general situation where there is a mixture of random dropouts and informative dropouts. Asymptotic properties of resulting estimators are also studied.

Simulation studies and an application to a renal disease data set are given in Section 4. Some remarks are discussed in Section 5. All the technical details are provided in the appendix.

2. Notation and Model Specification

2.1. Data Notation

Consider that a longitudinal study follows n subjects over time period $[0, \tau]$. For the ith subject $i = 1, \ldots, n$, the *complete-data* consist of $\{Y_{ij}, X_{ij}, t_{ij},\ j = 1, \ldots, n_i\}$, where Y_{ij} is the value of response at the jth observation time t_{ij} and X_{ij} is a $p \times 1$ vector of covariates associated with response Y_{ij}. Note that X_{ij} includes baseline covariates that are separately denoted by Z_i and potential time-dependent covariates. Let T_i denote the informative dropout time and C_i denote the random censoring time that is independent of (Y_{ij}, T_i) given the covariates. In practice, we observe (T_i^*, δ_i), where $T_i^* = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$ taking the value of 1 if the informative dropout time is observed and 0 otherwise. Throughout the paper, let $I(\cdot)$ denote the indicator function. Due to the dropout, longitudinal responses and covariates can only be observed at $t_{ij} \leq T_i^*$. Hence, the *observed data* are $\{Y_{ij}, X_{ij}, t_{ij}, T_i^*, \delta_i,\ i = 1, \ldots, n,\ j = 1, \ldots, m_i\}$, where $m_i = \sum_{j=1}^{n_i} I(t_{ij} \leq T_i^*)$.

2.2. Semiparametric Marginalized Latent Variable Model

We first introduce the composition of our proposed model and then discuss the model motivation and interpretation. The first component is a marginal generalized linear model for longitudinal responses Y_{ij}'s:

$$g\{\mathrm{E}(Y_{ij} \mid \mathrm{X}_{ij})\} \triangleq g\{\mu_{ij}\} = \beta_M' \mathrm{X}_{ij}, \tag{2.1}$$

where $g(\cdot)$ is a known link function and μ_{ij} denotes the marginal expectation. The second component is a linear transformation model for the informative dropout time T_i:

$$H(T_i) = -\theta' Z_i + \eta_i, \tag{2.2}$$

where $H(\cdot)$ is an unspecified monotone transformation function and η_i is assumed to follow a known continuous distribution $F(\cdot)$ that is independent of Z_i. The last component is a conditional mean model characterizing the dependence between longitudinal responses and informative dropouts:

$$g\{\mathrm{E}(Y_{ij} \mid \mathrm{X}_{ij}, \eta_i)\} \triangleq g\{\mu_{ij}(\eta_i)\} = \Delta_{ij} + \alpha' b_{ij}(\eta_i), \tag{2.3}$$

where the latent random effects $\mathbf{b}_i(\eta_i) = \{b_{i1}'(\eta_i), \ldots, b_{in_i}'(\eta_i)\}$ are investigator-specified functions of η_i and covariates, and Δ_{ij} is an implicit parameter whose value is determined by the integral equation matching the conditional mean model (2.3) with the corresponding marginal model (2.1), that is,

$$g^{-1}(\beta_M' X_{ij}) = \mathrm{E}(Y_{ij} \mid X_{ij}) = \mathrm{E}\{\mathrm{E}(Y_{ij} \mid X_{ij}, \eta_i)\} = \int g^{-1}\{\Delta_{ij} + \alpha' b_{ij}(\eta)\} dF(\eta). \tag{2.4}$$

The marginal mean model (2.1) directly specifies the marginal relationship between the responses and covariates, and β_M is the $p \times 1$ marginal regression parameters of main interest. Next, the semiparametric linear transformation model (2.2) is chosen to provide a flexible survival model for the informative dropout time while it still can be easily incorporated into model (2.3) for the dependence of the longitudinal responses and informative dropouts. Model (2.2) includes the proportional hazards model [25], the proportional odds model [26], and the Box-Cox transformation model as special cases and has been studied intensively in survival analysis literature [27–29]. In addition, as we present in Section 3, the explicit assumption on the error distribution in (2.2) can facilitate the "marginalization" procedure for parameter estimation.

The conditional mean model (2.3) is motivated by the construction of the marginalized random-effects model [30, 31]. As a motivating example, we consider a continuous Gaussian process following a simple random-effects model, $Y_{ij} = \beta' X_{ij} + b_{0i} + b_{1i} t_{ij} + \varepsilon_{ij}$, where (b_{i0}, b_{i1}) are the random intercept and slope, and error terms ε_{ij}, $j = 1, \ldots, n_i$ are assumed to follow $N(\mathbf{0}, \Sigma_i)$ but independent of (b_{i0}, b_{i1}) or η_i. Note that ε_{ij}'s can still exhibit temporal dependence in addition to what has been accounted by the random effects, that is, Σ_i with AR(1) covariance structure. Furthermore, as in joint modeling approaches via latent variables, the joint distribution of (b_{0i}, b_{1i}, η_i) is assumed to be $N\left(\mathbf{0}, \left(\begin{smallmatrix} \Sigma_b & C \\ C' & 1 \end{smallmatrix}\right)\right)$. It is easy to see that the conditional mean has the expression as model (2.3),

$$\mathrm{E}(Y_{ij} \mid X_{ij}, \eta_i) = \Delta_{ij} + C_1 \eta_i + C_2 \eta_i t_{ij}. \tag{2.5}$$

We use model (2.3) primarily as a parsimonious model for the dependence structure between the longitudinal responses and informative dropout times. However, note that although model (2.3) takes a similar form as the marginalized random-effects model, it does not intend to fully specify the joint distribution of the repeated measurements since model (2.3) only specifies the conditional mean function and there is no conditional independence assumed.

Note that Δ_{ij} is the solution of (2.4), and thus its value implicitly depends on β_M, α, the formulation of $b_{ij}(\eta_i)$ and the distribution of η_i. The specification of $b_{ij}(\eta_i)$ reflects investigator's assumptions on the dependence structure among the longitudinal responses and their association with the dropout times. It is well known that the dependence assumptions between longitudinal measurements and informative dropouts are usually unverifiable from the observed data, but it intrinsically affects the inference about β_M. Thus, a sensitivity analysis under various assumptions is always warranted. It is clear that the sensitivity analysis can be easily conducted within the framework of model (2.3). For example, the analysis may start with a large model in the specification for $\alpha' b_{ij}(\eta_i)$, for example, $\alpha_1 \eta_i + \alpha_2 \eta_i t_{ij} + \alpha_3' \eta_i X_{ij}$, and then examine the statistical significance of estimates for α's to further simplify the model. As shown in the next subsection, complex structure can be imposed on $b_{ij}(\eta_i)$ without introducing much extra computation. Lastly, we note that the marginal interpretation of longitudinal parameters in model (2.1) is invariant under different specifications of the conditional mean model (2.3).

3. Estimation and Asymptotic Properties

3.1. Conditional Generalized Estimating Equation

First, assume that η_i is known. We construct a "conditional" generalized estimating equation for $\mathcal{B} = (\beta_M', \alpha')'$. More specifically, the estimating function $U(\mathcal{B})$ is specified as

$$\sum_{i=1}^{n} \mathbf{A}_{i\cdot}'(\eta_i)\mathbf{W}_i\left[\mathbf{Y}_{i\cdot} - \boldsymbol{\mu}_{i\cdot}(\eta_i)\right] = \sum_{i=1}^{n} \mathbf{A}_{i\cdot}'(\eta_i)\mathbf{W}_i\left[\mathbf{Y}_{i\cdot} - g^{-1}\left\{\boldsymbol{\Delta}_{i\cdot} + \mathbf{b}_{i\cdot}(\eta_i)\alpha\right\}\right], \tag{3.1}$$

where $\mathbf{Y}_{i\cdot}$ denotes the $m_i \times 1$ vector of observed responses of subject i; $\boldsymbol{\Delta}_{i\cdot} = (\Delta_{i1}, \ldots, \Delta_{im_i})'$; $\mathbf{b}_{i\cdot}(\eta_i) = \{b_{i1}(\eta_i), \ldots, b_{im_i}(\eta_i)\}'$; $\boldsymbol{\mu}_{i\cdot}(\eta_i) = \{\mu_{i1}(\eta_i), \ldots, \mu_{im_i}(\eta_i)\}'$; $\mathbf{A}_{i\cdot}(\eta_i) = [\partial\boldsymbol{\mu}_{i\cdot}(\eta_i)/\partial\beta_M', \partial\boldsymbol{\mu}_{i\cdot}(\eta_i)/\partial\alpha']$; $\mathbf{W}_i$ is a $m_i \times m_i$ weight matrix.

It is easy to see that $U(\mathcal{B})$ has mean zero at the true parameter values $\mathcal{B}_0$ under model (2.3). Note that the vector of marginal parameters β_M is implicitly present in $U(\mathcal{B})$ with $\boldsymbol{\Delta}_{i\cdot}$ through the constrain equation (2.4). Thus, the Jacobian matrix $\mathbf{A}_{i\cdot}(\eta_i)$ needs to be derived using both the constrain (2.4) and models (2.1) and (2.3), which is different from the ordinary GEE. More specifically, entries of the Jacobian matrix $\mathbf{A}_{i\cdot}(\eta_i)$ are given by

$$\begin{aligned}
\frac{\partial\mu_{ij}(\eta_i)}{\partial\beta_M} &= X_{ij}\Upsilon_{ij}\left[\mathrm{E}\{\Upsilon_{ij}(\eta_i)\}\right]^{-1}\Upsilon_{ij}(\eta_i),\\
\frac{\partial\mu_{ij}(\eta_i)}{\partial\alpha} &= \left\{b_{ij}(\eta_i) - \mathrm{E}\{b_{ij}(\eta_i)\Upsilon_{ij}(\eta_i)\}\left[\mathrm{E}\{\Upsilon_{ij}(\eta_i)\}\right]^{-1}\right\}\Upsilon_{ij}(\eta_i),
\end{aligned} \tag{3.2}$$

where $\Upsilon_{ij} = 1/\dot{g}\{g^{-1}(\beta_M' X_{ij})\}$, $\Upsilon_{ij}(\eta_i) = 1/\dot{g}[g^{-1}\{\Delta_{ij} + \alpha' b_{ij}(\eta_i)\}]$, and we use $\dot{a}(x)$ to denote the derivative of a function $a(x)$ throughout this paper. In particular, we have $\Upsilon_{ij} = \mu_{ij}(1-\mu_{ij})$ and $\Upsilon_{ij}(\eta_i) = \mu_{ij}(\eta_i)\{1 - \mu_{ij}(\eta_i)\}$ under the logit-link function for binary data; $\Upsilon_{ij} = \mu_{ij}$ and $\Upsilon_{ij}(\eta_i) = \mu_{ij}(\eta_i)$ under the log-link function for count data. Thus, under these canonical link functions, $\Upsilon_{ij} = \mathrm{Var}(Y_{ij} \mid X_{ij})$ and $\Upsilon_{ij}(\eta_i) = \mathrm{Var}(Y_{ij} \mid X_{ij}, \eta_i)$ are the marginal variance and conditional variance of the responses, respectively. In addition, these formulations also facilitate our selection of the weight matrix $\mathbf{W}_i$. For example, for binary longitudinal data with logit-link function, we can choose a weight matrix as $\mathbf{W}_i^{-1} = \mathrm{diag}\{\mathrm{Var}(Y_{ij} \mid X_{ij}, \eta_i),\ j = 1, \ldots, m_i\}$.

It is clear that the implementation of the estimating function (3.1) requires the knowledge of η_i, which is an unknown quantity and has to be estimated first. The estimation of the semiparametric linear transformation model (2.2) has been studied by many authors [27–29]. In particular, Chen et al. [27] proposed a class of martingale-based estimating equations,

$$\begin{aligned}
&\sum_{i=1}^{n}\int_0^{\infty} Z_i\left[dN_i(t) - I\left(T_i^* \geq t\right)d\Lambda\left\{\theta' Z_i + H(t)\right\}\right] = 0,\\
&\sum_{i=1}^{n}\left[dN_i(t) - I\left(T_i^* \geq t\right)d\Lambda\left\{\theta' Z_i + H(t)\right\}\right] = 0, \quad \forall t \geq 0,
\end{aligned} \tag{3.3}$$

where $N_i(t) = I(T_i^* \le t, \delta_i = 1)$. Then an iterative algorithm can be carried out to solve θ and H simultaneously. We estimate (θ, H) using the approach of Chen et al. [27] and shall denote the estimates as $\widehat{\Theta} \triangleq (\widehat{\theta}, \widehat{H})$.

3.2. Estimation Procedure for Pure Informative Dropouts

We first consider the situation of pure informative dropouts, that is, $\delta_i \equiv 1$. Define $\widehat{\eta}_i = \widehat{H}(T_i) + \widehat{\theta}' Z_i$ and replace η_i's in (3.1) with their estimated counterparts $\widehat{\eta}_i$'s. Denote the resulting estimating function by $U(\mathcal{B}; \widehat{\Theta})$ and define the estimator of $\mathcal{B}$ as the solution to $U(\mathcal{B}; \widehat{\Theta}) = 0$. The estimation of $\mathcal{B}$ entails an iteration between solving nonlinear equations for Δ_{ij} and updating a Newton-Raphson equation for $\mathcal{B}$. More specifically, given the current estimated value of $\mathcal{B}^{(J)}$ at the Jth step, we first estimate $\Delta_{ij}^{(J)}$ from

$$g^{-1}\left\{\beta_M^{(J)'} X_{ij}\right\} = \int g^{-1}\left\{\Delta_{ij} + \alpha^{(J)'} b_{ij}(\eta)\right\} dF(\eta), \tag{3.4}$$

and then update the parameters $\mathcal{B}$ by

$$\mathcal{B}^{(J+1)} = \mathcal{B}^{(J)} + \left\{\sum_{i=1}^{n} \mathbf{A}_i^{(J)'}(\widehat{\eta}_i) \mathbf{W}_i^{(J)} \mathbf{A}_i^{(J)}(\widehat{\eta}_i)\right\}^{-1} \left\{\sum_{i=1}^{n} \mathbf{A}_i^{(J)'}(\widehat{\eta}_i) \mathbf{W}_i^{(J)} \left[\mathbf{Y}_{i\cdot} - \boldsymbol{\mu}_{i\cdot}^{(J)}(\widehat{\eta}_i)\right]\right\}, \tag{3.5}$$

where $\mathbf{A}_i^{(J)}(\widehat{\eta}_i)$, $\mathbf{W}_i^{(J)}$, and $\boldsymbol{\mu}_{i\cdot}^{(J)}(\widehat{\eta}_i)$ are evaluated at the current parameter values $\mathcal{B}^{(J)}$ and $\Delta_{ij}^{(J)}$. The algorithm is iterated until it converges. Because η_i is assumed to follow an explicit parametric distribution F, it greatly simplifies the marginalization procedure (3.4). We propose to use the Gaussian-quadrature approach [32] to numerically evaluate (3.4) and $\mathbf{A}_i^{(J)}(\widehat{\eta}_i)$. Since the integrand of (3.4) is monotonic in $\Delta_{ij}^{(J)}$ and so is the whole integral, it is easy to calculate a large number of $\Delta_{ij}^{(J)}$, $i = 1, \ldots, n$, $j = 1, \ldots, m_i$, in all iterative steps. Moreover, the numerical integration is only upon the one-dimensional space of η_i and requires light computation even with complex structure assumed on $b_{ij}(\eta_i)$. The proposed iterative algorithm has been implemented using "R" codes, which are available from the authors upon request.

3.3. Estimation Procedure for Mixed Types of Dropouts

We generalize the proposed estimation function (3.1) to accommodate the situation where there are mixed informative dropouts and random censoring. More specifically, the modified estimating equation is given by

$$U^*(\mathcal{B}; \Theta) = \sum_{i=1}^{n} \mathbf{A}_{i\cdot}^{*\prime}(\eta_i^*, \delta_i) \mathbf{W}_i^* \left[\mathbf{Y}_{i\cdot} - \boldsymbol{\mu}_{i\cdot}^*(\eta_i^*, \delta_i)\right] = 0, \tag{3.6}$$

where $\eta_i^* = H(T_i^*) + \theta' Z_i$; the jth component of $\boldsymbol{\mu}_{i\cdot}^*(\eta_i^*, \delta_i)$ is

$$\mu_{ij}^*(\eta_i^*, \delta_i) = \delta_i g^{-1}\left\{\Delta_{ij} + \alpha' b_{ij}(\eta_i^*)\right\} + (1 - \delta_i) \mathrm{E}\left[g^{-1}\left\{\Delta_{ij} + \alpha' b_{ij}(\eta)\right\} \mid \eta \ge \eta_i^*\right], \tag{3.7}$$

and the Jacobian matrix $\mathbf{A}_{i\cdot}^{*}(\eta_i^{*},\delta_i) = [\partial\boldsymbol{\mu}_{i\cdot}^{*}(\eta_i^{*},\delta_i)/\partial\beta_M', \partial\boldsymbol{\mu}_{i\cdot}^{*}(\eta_i^{*},\delta_i)/\partial\alpha']$. When $\delta_i = 1$, the ith component of $U^{*}(\mathcal{B};\Theta)$ is the same as the one in (3.1). For $\delta_i = 0$, the entries of $\mathbf{A}_{i\cdot}^{*}(\eta_i^{*},0)$ are given by

$$\frac{\partial\mu_{ij}^{*}(\eta_i^{*},0)}{\partial\beta_M} = X_{ij}\Upsilon_{ij}\left[\mathrm{E}\{\Upsilon_{ij}(\eta_i)\}\right]^{-1}\mathrm{E}\{\Upsilon_{ij}(\eta) \mid \eta \geq \eta_i^{*}\},$$

$$\frac{\partial\mu_{ij}^{*}(\eta_i^{*},0)}{\partial\alpha} = \mathrm{E}\{b_{ij}(\eta)\Upsilon_{ij}(\eta) \mid \eta \geq \eta_i^{*}\} - \mathrm{E}\{b_{ij}(\eta_i)\Upsilon_{ij}(\eta_i)\}\left[\mathrm{E}\{\Upsilon_{ij}(\eta_i)\}\right]^{-1}\mathrm{E}\{\Upsilon_{ij}(\eta) \mid \eta \geq \eta_i^{*}\}. \tag{3.8}$$

In addition, the entries of the weight matrix can be changed to $\mathrm{Var}\{Y_{ij} \mid X_{ij}, \eta_i^{*}, \delta_i\}$ accordingly. Conditional expectations of various functions given $\eta \geq \eta_i^{*}$ are computed using the Gaussian-quadrature method. Let $\widehat{\eta}_i^{*} = \widehat{H}(T_i^{*}) + \widehat{\theta}' Z_i$ and replace η_i^{*}'s in (3.6) with their estimated counterparts $\widehat{\eta}_i^{*}$'s. Denote the resulting estimating function by $U^{*}(\mathcal{B};\widehat{\Theta})$. Then the estimator $\widehat{\mathcal{B}}$ of $\mathcal{B}$ can be obtained from the equation $U^{*}(\mathcal{B};\widehat{\Theta}) = 0$ using the same iterative algorithm described in the previous subsection.

3.4. Asymptotic Properties of $\widehat{\mathcal{B}}$

In this subsection, we establish the asymptotic properties of $\widehat{\mathcal{B}}$. Towards this end, we need the following assumptions.

(C1) The covariates X_{ij}'s are bounded with probability 1.

(C2) The true parameter values $\mathcal{B}_0$ and θ_0 belong to the interior of a known compact set, and the true transformation function H_0 has a continuous and positive derivative.

(C3) Let $\Lambda(\cdot)$ denote the cumulative hazard function of η_i. Define $\lambda(t) = \dot{\Lambda}(t)$ and $\psi(t) = \dot{\lambda}(t)/\lambda(t)$. Then $\lambda(\cdot)$ is positive, $\psi(\cdot)$ is continuous, and $\lim_{t\to-\infty}\lambda(t) = 0 = \lim_{t\to-\infty}\psi(t)$.

(C4) τ is finite and satisfies $P(T > \tau) > 0$ and $P(C = \tau) > 0$.

(C5) The matrix $\Omega \equiv \mathrm{E}\{\mathbf{A}_{1\cdot}^{*\prime}(\eta_1^{*},\delta_1)\mathbf{W}_1\mathbf{A}_{1\cdot}^{*}(\eta_1^{*},\delta_1)\}$ is positive finite, and the number of repeated measurements $m_i \ll N$.

The regularity conditions (C1)–(C4) are also used by Chen et al. [27] to derive the consistency and asymptotically normality of the estimators $\widehat{\Theta}$. Condition (C5) is needed to establish the consistency and asymptotic normality of $\widehat{\mathcal{B}}$, which is given in the following theorem.

Theorem 3.1. *Under conditions (C1)–(C5), with probability 1, $|\widehat{\mathcal{B}} - \mathcal{B}_0| \to 0$. In addition, one has, as $n \to \infty$,*

$$\sqrt{n}\left(\widehat{\mathcal{B}} - \mathcal{B}_0\right) \longrightarrow_D N\left(0, \Omega^{-1}V\Omega^{-1}\right). \tag{3.9}$$

The definition of V and a sketch of the proof for Theorem 3.1 are given in the appendix. The asymptotic variance-covariance matrix can be consistently estimated by its empirical counterpart $\widehat{\Omega}^{-1}\widehat{V}\widehat{\Omega}^{-1}$, which can be easily obtained using the usual plug-in rule.

4. Numerical Studies

4.1. Simulations

We conducted a series of simulation studies to evaluate the finite-sample performance of our proposed approach. Consider a binary longitudinal process with the marginal probability of success as $g^{-1}(\beta_0 + \beta_1 t + \beta_2 Z)$, where g was the logit-link function; observations occurred at $\mathbf{t}_i = \{t_{ij} = j, j = 0, 1, \ldots, 5\}$; Z was generated from a Bernoulli distribution with the success probability of 0.5, and $(\beta_0, \beta_1, \beta_2) = (-1.5, 0.3, 1)$. The informative dropout time T_i was generated from a linear transformation model $H(T_i) = -\theta Z_i + \eta_i$, where $\theta = -0.5$. We considered three distributions for η_i: the standard normal distribution (N.), the extreme value distribution (E.), and the logistic distribution (L.), corresponding to the normal transformation model, the Cox proportional hazard model, and the proportional odds model, respectively, for the informative dropout time. We then generated the binary response Y_{ij} independently from a Bernoulli distribution with the success probability of $g^{-1}\{\Delta_{ij} + \alpha b_{ij}(\eta_i)\}$, where Δ_{ij} was calculated to match the marginal mean value as in (2.4) and α indicated the level of dependence. We considered several combinations to specify the dependence between longitudinal outcomes and informative dropout times. More specifically, when $\alpha = 0$, there was no informative dropouts; when $\alpha = 0.5$ (or 0.25) and $b_{ij}(\eta_i) = \eta_i$, the dependence existed and was linear in the latent variable η_i; when $\alpha = 0.25$ (or 0.5) and $b_{ij}(\eta_i) = \eta_i t_{ij}$, the dependence was present through an interaction between the latent variable and the observation time.

For each scenario, we considered samples of size 100 and 200 and conducted 500 runs of simulations. The Gaussian-quadrature approximation was calculated using 50 grid points. We first considered the situation of pure informative dropouts and generated the dropout time T_i from the transformation model with $H_0(t) = 2\{\arctan(t) + \pi/2\}$. Under the assumptions of η_i following the normal, the extreme value, and the logistic distributions, the average numbers of repeated measurements were 3.91, 3.37, and 3.94, respectively. The estimation results on β_1 and α are summarized in Table 1. The proposed estimators are unbiased under all simulated scenarios, and the Wald-type 95% confidence intervals all have reasonable empirical coverage probabilities. The performance of the proposed method is consistent with different distributional assumptions of η_i and different specifications of the dependence structure, and the results improve as the sample size increases.

Next, we consider the situation where there are mixed informative dropouts and random censoring. For simplicity, let C_i be an administrative censoring at the end of the study, that is, $\tau = 6$. The informative dropout time T_i was generated from the transformation model with $H(t) = \log(t) - 1$ and η_i followed the standard normal distribution. This yielded the proportion of informative dropouts of 69.6% and the average number of repeated measurements about 4. Other settings were kept the same as in the previous simulations. The simulation results are presented in Table 2. Again, the proposed approach gives unbiased parameter estimates and reasonable coverage probabilities under all the scenarios. For comparison, we also implemented the ordinary GEE method [16]. When the informative dropout is absent, that is, $\alpha = 0$, the GEE method yields consistent parameter estimates of β_1 as expected. But when there is informative dropout ($\alpha \neq 0$), the performance of the GEE method deteriorates quickly as the magnitude of the dependence between the longitudinal data and informative dropout increases.

Last, we conducted sensitivity analysis for the proposed approach and our simulations consisted of two parts. First, as discussed in Section 2, to better characterize the dependence

Table 1: Simulation results for pure informative dropouts.

F	α	$b_{ij}(\eta_i)$		β_1				α			
			N	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
N.	0	η_i	100	0.001	0.084	0.082	0.948	0.004	0.130	0.127	0.954
			200	0.006	0.057	0.058	0.952	–0.002	0.090	0.091	0.942
	0.5	η_i	100	0.008	0.083	0.082	0.952	0.008	0.148	0.142	0.936
			200	0.005	0.056	0.058	0.954	0.001	0.102	0.101	0.944
	0.25	$\eta_i t_{ij}$	100	0.004	0.102	0.101	0.940	0.006	0.083	0.081	0.938
			200	0.009	0.071	0.071	0.948	–0.001	0.060	0.058	0.940
E.	0	η_i	100	0.007	0.107	0.097	0.920	0.001	0.150	0.140	0.942
			200	0.002	0.069	0.069	0.950	–0.003	0.010	0.097	0.958
	0.5	η_i	100	0.010	0.097	0.097	0.948	0.025	0.161	0.161	0.962
			200	0.009	0.071	0.069	0.932	0.002	0.111	0.113	0.950
	0.25	$\eta_i t_{ij}$	100	0.026	0.138	0.135	0.936	0.001	0.124	0.114	0.926
			200	0.009	0.089	0.095	0.954	0.001	0.079	0.082	0.952
L.	0	η_i	100	0.004	0.079	0.077	0.932	0.001	0.076	0.077	0.950
			200	0.002	0.057	0.055	0.952	–0.001	0.058	0.054	0.938
	0.5	η_i	100	0.009	0.079	0.076	0.944	0.007	0.114	0.105	0.928
			200	0.002	0.053	0.054	0.964	0.006	0.074	0.075	0.956
	0.25	$\eta_i t_{ij}$	100	0.003	0.096	0.096	0.954	0.008	0.069	0.067	0.942
			200	0.002	0.067	0.068	0.964	0.004	0.047	0.047	0.958

In Tables 1–3: F: error distribution of the semiparametric transformation model; N: sample size; SSE: sample standard deviations of estimates; SEE: mean of estimates standard errors; CP: 95% coverage probability of Wald-type confidence interval.

Table 2: Simulation results for mixed types of dropouts.

α	$b_{ij}(\eta_i)$	N	Proposed β_1			Proposed α			GEE β_1		
			Bias	SSE	CP	Bias	SSE	CP	Bias	SSE	CP
0	η_i	100	0.003	0.075	0.958	–0.007	0.158	0.956	0.002	0.068	0.946
		200	0.005	0.056	0.954	–0.003	0.114	0.944	0.005	0.049	0.956
0.25	η_i	100	0.006	0.073	0.948	0.009	0.156	0.964	0.061	0.068	0.878
		200	0.004	0.055	0.950	–0.003	0.116	0.936	0.057	0.050	0.768
0.50	η_i	100	0.008	0.076	0.948	0.010	0.174	0.952	0.116	0.070	0.660
		200	0.006	0.051	0.960	–0.008	0.121	0.956	0.113	0.049	0.380
0.25	$\eta_i t_{ij}$	100	0.005	0.106	0.948	0.005	0.106	0.960	0.233	0.074	0.110
		200	0.009	0.078	0.936	–0.007	0.077	0.940	0.233	0.054	0.000
0.50	$\eta_i t_{ij}$	100	0.007	0.097	0.954	0.010	0.141	0.950	0.397	0.081	0.000
		200	0.008	0.071	0.940	–0.008	0.104	0.936	0.395	0.059	0.000

structure between longitudinal responses and informative dropouts, we would suggest to start with a large model in the specification for $\alpha' b_{ij}(\eta_i)$ and then examine the statistical significance of estimates for α's to further simplify the model. We simulated data from a simple model with either $\alpha' b_{ij}(\eta_i) = \alpha_1 \eta_i$ or $\alpha' b_{ij}(\eta_i) = \alpha_2 \eta_i t_{ij}$, and then applied the proposed approach by assuming a bigger model (2.3) as $\Delta_{ij} + \alpha_1 \eta_i + \alpha_2 \eta_i t_{ij}$. The simulation results are summarized in the top panel in Table 3. The proposed method can reasonably well estimate all the parameters, and in particular, could correctly indicate the unnecessary zero term. Second, we simulated data from $\Delta_{ij} + \alpha_1 \eta_i + \alpha_2 \eta_i t_{ij}$ but fitted misspecified models that omitted

Table 3: Sensitivity analysis for misspecified models under mixed types of dropouts.

True	Fitted		Proposed β_1		Proposed α_1		Proposed α_2	
		N	Bias	CP	Bias	CP	Bias	CP
$0.25\eta_i$	$\alpha_1\eta_i + \alpha_2\eta_i t_{ij}$	100	0.009	0.946	0.014	0.956	0.002	0.950
		200	0.009	0.964	−0.002	0.948	0.002	0.968
$0.25\eta_i t_{ij}$	$\alpha_1\eta_i + \alpha_2\eta_i t_{ij}$	100	0.005	0.946	−0.004	0.954	0.007	0.956
		200	0.009	0.936	0.001	0.956	−0.007	0.952
$0.5\eta_i + 0.1\eta_i t_{ij}$	$\alpha_1\eta_i$	100	0.071	0.860	0.119	0.918		
		200	0.069	0.788	0.103	0.896		
$0.2\eta_i + 0.25\eta_i t_{ij}$	$\alpha_2\eta_i t_{ij}$	100	−0.017	0.924			0.086	0.850
		200	−0.017	0.902			0.078	0.692

some terms. The results are summarized in the lower panel of Table 3. It is evident that misspecified models lead to biased estimates for both longitudinal regression coefficients and dependence parameters.

4.2. Application to Renal Disease Data from MDRD Study

Here we considered a subgroup of 129 patients with low-protein diet in MDRD study B, among whom, 62 patients were randomized to the group of normal-blood-pressure control and 67 patients were randomized to the group of low-blood-pressure control. Besides the randomized intervention, other covariates included time in study (`time`), baseline disease progression status (`Prog`), baseline blood pressure (`Bp`), and log-transformed baseline urine protein level (`log.Pt`). There were 52 (40.3%) patients left the study prematurely for kidney transplant or dialysis and were treated as informative dropouts.

We applied the proposed approach to estimate the marginal effects of covariates on GFR values. To account for the possible informative dropouts, we assumed that the dependence term $\alpha' b_{ij}(\eta_i)$ had a form of $\alpha_1\eta_i + \alpha_2\eta_i t_{ij}$, analogous to the joint modeling approach with latent random intercept and random slope used in Schluchter et al. [33]. We considered the situations of η_i following the standard normal, the extreme value, or the standard logistic distributions. Because the outcome was a continuous variable, we used the identity link function.

Our results are presented in Table 4 and compared with the results from the ordinary GEE [16] with an independent working correlation matrix. More specifically, the slope estimates from the proposed approach indicate a much faster decreasing rate of GFR (e.g., `time` Est = −0.27, SE = 0.03, under the normality assumption for η_i) than the result from the ordinary GEE method (`time` Est = −0.14, SE = 0.03). A possible explanation is that those patients remaining under observation usually have better kidney functions and thus higher GFR values. The ordinary GEE approach that treats the observed patients as random representatives of the population tends to underestimate the degressive trend of GFR.

The estimates for the intervention on blood pressure control show positive effect of the low-blood-pressure control on the longitudinal GFR development. Although the results are not statistically significant, the estimates from the proposed method (e.g., `Intervention` Est = 0.82, SE = 1.07, under the normality assumption for η_i) are about twice large of the values from the ordinary GEE method (`Intervention` Est = 0.35, SE = 0.90). Moreover, for the proposed approach, the results under different distributional assumptions for η_i are quite similar. The estimates of the dependency parameters (α_1, α_2) are positive and statistically significant.

Table 4: Estimates of regression coefficients for the MDRD study.

Variable	Proposed						GEE	
	Normal		EV		Logistic			
	Est	SE	Est	SE	Est	SE	Est	SE
`Intercept`	18.54	0.96	18.57	0.91	18.58	1.11	18.57	0.78
`Time`	−0.27	0.03	−0.29	0.04	−0.28	0.03	−0.14	0.03
`Intervention`	0.82	1.06	0.74	1.01	0.71	1.17	0.35	0.90
`Prog`	−0.14	1.07	−0.08	1.02	−0.14	1.19	−0.14	0.91
`Bp`	−0.15	1.38	−0.20	1.34	−0.07	1.48	−0.36	0.49
`log.Pt`	−1.09	0.39	−1.09	0.37	−1.12	0.42	−0.61	0.38
η_i	1.91	0.50	1.38	0.38	1.11	0.28		
$\eta_i t_{ij}$	0.14	0.04	0.14	0.04	0.08	0.03		

`Intervention`: blood pressure control (1: low and 0: normal); `Prog`: baseline disease progression status (1: yes and 0: no); `Bp`: baseline blood pressure; `log.Pt`: baseline log-transformed urine protein level.

This indicates that higher GFR values are positively associated with longer dropout times in the study. In addition, our proposed approach shows that the baseline urine protein level is significantly associated with the longitudinal GFR development, but the ordinary GEE method does not show such significance. The results obtained using our proposed method are also consistent with those reported in Schluchter et al. [33].

5. Discussion

In this paper, we propose a semiparametric marginalized model for marginal inference of the relationship between longitudinal responses and covariates in the presence of informative dropouts. The regression parameters represent the covariate effects on the population level. The proposed estimators are expected to be insensitive to misspecification of the latent variable distribution [31], which is desirable pertaining to the sensitivity analysis on unverifiable assumptions for the informative dropouts. In practice, the choice between marginal models and other types of joint modeling approaches should be determined by study objective.

To estimate the regression parameters in the proposed marginalized model, we proposed a class of simple conditional generalized estimating equations and demonstrated its computational convenience. In general, a likelihood-based approach can be used to achieve more efficient inference and is also of great interest. For example, a marginalized random effects model [30, 31] can be used for the longitudinal process and a frailty model [34] can be used for the dropout time. Furthermore, latent variables (b_{ij}, η_i) can be modeled by a copula distribution or non-Gaussian distributions [35]. The likelihood-based methods enjoy the high efficiency and facilitate the implementation of classical model selection procedures, such as AIC or BIC; however, intensive computations are often involved when the dimension of random effects is high.

Appendix

A. Proof of Theorem 3.1

Under the conditions (C1)–(C4), Chen et al. [27] established the consistency of $\widehat{\theta}$ and the uniform consistency of a transformed $\widehat{H}(\cdot)$, that is, $\sup_{t\in[0,\tau]}|\exp\{\widehat{H}(t)\}-\exp\{H_0(t)\}| = o_p(1)$.

We first derive the consistency of the proposed estimator $\widehat{\mathcal{B}}$, which is the solution of the equation $U^*(\mathcal{B};\widehat{\Theta}) = 0$. Note that $-(1/n)(\partial U^*(\mathcal{B};\widehat{\Theta})/\partial\mathcal{B}) = (1/n)\sum_n \mathbf{A}_{i\cdot}^{*\prime}(\widehat{\eta}_i^*,\delta_i)\mathbf{W}_i\mathbf{A}_{i\cdot}^*(\widehat{\eta}_i^*,\delta_i)$ is a positive definite matrix, and by the law of large numbers and the consistency of $\widehat{\theta}$ and $\widehat{H}$, it converges uniformly to a deterministic positive definite matrix $\Omega(\mathcal{B})$ over a compact set of $\mathcal{B}$. In addition, we have that $(1/n)U^*(\mathcal{B};\widehat{\Theta})$ converges uniformly to a deterministic function $u^*(\mathcal{B};\Theta_0)$ satisfying $u^*(\mathcal{B}_0;\Theta_0) = 0$ and $-(\partial u^*(\mathcal{B};\Theta_0)/\partial\mathcal{B})|_{\mathcal{B}=\mathcal{B}_0} = \Omega(\mathcal{B}_0) = \Omega$. Thus, the estimating equation $U^*(\mathcal{B};\widehat{\Theta}) = 0$ exists a unique solution $\widehat{\mathcal{B}}$. Since $\mathcal{B}_0$ is the unique solution of $u^*(\mathcal{B}_0;\Theta_0) = 0$, the consistency of $\widehat{\mathcal{B}}$ easily follows.

To prove the asymptotic normality, by the Taylor expansion, we have

$$\sqrt{n}\left(\widehat{\mathcal{B}} - \mathcal{B}_0\right) = \left\{-\frac{1}{n}\frac{\partial U^*\left(\mathcal{B}^*;\widehat{\Theta}\right)}{\partial\mathcal{B}}\right\}^{-1}\frac{1}{\sqrt{n}}U^*\left(\mathcal{B}_0;\widehat{\Theta}\right), \tag{A.1}$$

where $\mathcal{B}^*$ lies between $\widehat{\mathcal{B}}$ and $\mathcal{B}_0$. Since $\widehat{\mathcal{B}}$ is consistent and $-(1/n)(\partial U^*(\mathcal{B};\widehat{\Theta})/\partial\mathcal{B})$ converges uniformly to $\Omega(\mathcal{B})$, we have that $-(1/n)(\partial U^*(\mathcal{B}^*;\widehat{\Theta})/\partial\mathcal{B})$ converges to Ω. Furthermore, the Taylor expansion of $U^*(\mathcal{B}_0;\widehat{\Theta})$ around Θ_0 gives

$$\begin{aligned}
\frac{1}{\sqrt{n}}U^*\left(\mathcal{B}_0;\widehat{\Theta}\right) &= \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n}\mathbf{A}_{i\cdot}^{*\prime}(\eta_i^*,\delta_i)\mathbf{W}_i\left\{\mathbf{Y}_{i\cdot} - \boldsymbol{\mu}_{i\cdot}^*(\eta_i^*,\delta_i)\right\}\right.\\
&\qquad\left.-\sum_{i=1}^{n}\mathbf{A}_{i\cdot}^{*\prime}(\eta_i^*,\delta_i)\mathbf{W}_i\left\{\boldsymbol{\mu}_{i\cdot}^*(\widehat{\eta}_i^*,\delta_i) - \boldsymbol{\mu}_{i\cdot}^*(\eta_i^*,\delta_i)\right\}\right]\\
&= \frac{1}{\sqrt{n}}\left[U^*(\mathcal{B}_0;\Theta_0) - \sum_{i=1}^{n}\mathbf{A}_{i\cdot}^{*\prime}(\eta_i^*,\delta_i)\mathbf{W}_i\left\{\left.\frac{\partial\boldsymbol{\mu}_{i\cdot}^*(\eta,\cdot)}{\partial\eta}\right|_{(\eta_i^*,\delta_i)}\right\}\right.\\
&\qquad\left.\times\left\{\left(\widehat{\theta} - \theta_0\right)'(Z_i + O_i) + \widehat{H}(T_i^*;\theta_0) - H_0(T_i^*)\right\}\right] + o_p(1),
\end{aligned} \tag{A.2}$$

where $O_i = \partial\widehat{H}(T_i^*;\theta)/\partial\theta|_{\theta=\theta_0}$ and its asymptotic representation can be found in the appendix of Chen et al. [27]. The asymptotic representations of $\widehat{\theta}$ and $\widehat{H}$ have also been derived by Chen et al. [27],

$$\begin{aligned}
\sqrt{n}\left(\widehat{\theta} - \theta\right) &= \Sigma_*^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\left\{Z_i - \mu_b(t)\right\}dM_i(t) + o_p(1),\\
\Lambda^*\left\{\widehat{H}(t;\theta_0)\right\} - \Lambda^*\left\{H_0(t)\right\} &= n^{-1}\sum_{i=1}^{n}\int_0^{t}\frac{\lambda^*\left\{H_0(s)\right\}}{B_2(s)}dM_i(s) + o_p\left(n^{-1/2}\right),
\end{aligned} \tag{A.3}$$

where $N_i(t) = \delta_i I(T_i^* \le t)$ and $\xi_i(t) = I(T_i^* \ge t)$ are the counting and at-risk processes, respectively, $M_i(t) = N_i(t) - \int_0^t \xi_i(s)d\Lambda\{\theta_0'Z_i + H_0(s)\}$ is the mean-zero martingale process, and

the definitions of Σ_* and the functions $\Lambda^*(\cdot)$, $\lambda^*(\cdot)$, $B_2(\cdot)$, and $\mu_b(\cdot)$ are given in the appendix of Chen et al. [27].

Plugging these terms back to the expansion of $(1/\sqrt{n})U^*(\mathcal{B}_0;\widehat{\Theta})$ specified in (A.2), some rearrangement yields that it is equal to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathbf{A}_{i\cdot}^{*\prime}(\eta_i^*,\delta_i)\mathbf{W}_i\{\mathbf{Y}_{i\cdot}-\boldsymbol{\mu}_{i\cdot}^*(\eta_i^*,\delta_i)\}-\int_0^\tau\left[Q_1\{Z_i-\mu_b(t)\}+Q_2(t)\right]dM_i(t)\right)+o_p(1), \tag{A.4}$$

where

$$\begin{aligned}
Q_1&=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\left[\mathbf{A}_{i\cdot}^{*\prime}(\eta_i^*,\delta_i)\mathbf{W}_i\left\{\left.\frac{\partial\boldsymbol{\mu}_{i\cdot}^*(\eta,\cdot)}{\partial\eta}\right|_{(\eta_i^*,\delta_i)}\right\}\right](Z_i+O_i)'\Sigma_*^{-1},\\
Q_2(t)&=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\left[\mathbf{A}_{i\cdot}^{*\prime}(\eta_i^*,\delta_i)\mathbf{W}_i\left\{\left.\frac{\partial\boldsymbol{\mu}_{i\cdot}^*(\eta,\cdot)}{\partial\eta}\right|_{(\eta_i^*,\delta_i)}\right\}\right]\frac{B(t,T_i^*)\xi_i(t)}{B_2(t)}.
\end{aligned} \tag{A.5}$$

The definition of $B(t,s)$ can be found in Chen et al. [27].

Hence, $(1/\sqrt{n})U^*(\mathcal{B}_0;\widehat{\Theta})$ has been written as a standardized summation of independent terms with mean zero. By the central limit theorem, it is asymptotically equivalent to a multivariate Gaussian variable with zero mean and covariance matrix V, which is the limit of

$$\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{A}_{i\cdot}^{*\prime}(\eta_i^*,\delta_i)\mathbf{W}_i\{\mathbf{Y}_{i\cdot}-\boldsymbol{\mu}_{i\cdot}^*(\eta_i^*,\delta_i)\}-\int_0^\tau\left[Q_1\{Z_i-\mu_b(t)\}+Q_2(t)\right]dM_i(t)\right)^{\otimes 2}. \tag{A.6}$$

From (A.1), it is easy to see that the estimator $\widehat{\mathcal{B}}$ is asymptotically normal with mean zero and the variance-covariance matrix $\Omega^{-1}V\Omega^{-1}$, which can be consistently estimated by its empirical counterpart $\widehat{\Omega}^{-1}\widehat{V}\widehat{\Omega}^{-1}$ using the usual plug-in rule.

Acknowledgments

The authors would like to thank the editor and the referees for their instructive comments. This research was supported partially by NIH grants RO1 CA-140632 and RO3 CA-153083.

References

[1] G. J. Beck, R. L. Berg, C. H. Coggins et al., "Design and statistical issues of the modification of diet in renal disease trial. The modification of diet in renal disease study group," *Controlled Clinical Ttrials*, vol. 12, pp. 566–586, 19991.

[2] S. Klahr, A. S. Levey, G. J. Beck et al., "The effects of dietary protein restriction and blood-pressure control on the progression of chronic renal disease," *The New England Journal of Medicine*, vol. 330, no. 13, pp. 877–884, 1994.

[3] J. W. Hogan and N. M. Laird, "Mixture models for the joint distribution of repeated measures and event times," *Statistics in Medicine*, vol. 16, no. 1–3, pp. 239–257, 1997.

[4] J. W. Hogan and N. M. Laird, "Model-based approaches to analysing incomplete longitudinal and failure time data," *Statistics in Medicine*, vol. 16, no. 1–3, pp. 259–272, 1997.

[5] J. W. Hogan, J. Roy, and C. Korkontzelou, "Tutorial in biostatistics. Handling drop-out in longitudinal studies," *Statistics in Medicine*, vol. 23, no. 9, pp. 1455–1497, 2004.

[6] J. G. Ibrahim and G. Molenberghs, "Missing data methods in longitudinal studies: a review," *Test*, vol. 18, no. 1, pp. 1–43, 2009.

[7] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, John Wiley & Sons, New York, NY, USA, 1987.

[8] A. A. Tsiatis and M. Davidian, "Joint modeling of longitudinal and time-to-event data: an overview," *Statistica Sinica*, vol. 14, no. 3, pp. 809–834, 2004.

[9] V. De Gruttola and X. M. Tu, "Modelling progression of CD4-lymphocyte count and its relationship to survival time," *Biometrics*, vol. 50, no. 4, pp. 1003–1014, 1994.

[10] M. D. Schluchter, "Methods for the analysis of informatively censored longitudinal data," *Statistics in Medicine*, vol. 11, no. 14-15, pp. 1861–1870, 1992.

[11] M. C. Wu and R. J. Carroll, "Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process," *Biometrics*, vol. 44, no. 1, pp. 175–188, 1988.

[12] R. J. A. Little, "Pattern-mixture models for multivariate incomplete data," *Journal of the American Statistical Association*, vol. 88, pp. 125–134, 1993.

[13] B. F. Kurland and P. J. Heagerty, "Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths," *Biostatistics*, vol. 6, no. 2, pp. 241–258, 2005.

[14] C. Dufouil, C. Brayne, and D. Clayton, "Analysis of longitudinal studies with death and drop-out: a case study," *Statistics in Medicine*, vol. 23, no. 14, pp. 2215–2226, 2004.

[15] D. B. Rubin, "Inference and missing data," *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976.

[16] K. Y. Liang and S. L. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika*, vol. 73, no. 1, pp. 13–22, 1986.

[17] J. M. Robins, A. Rotnitzky, and L. P. Zhao, "Analysis of semiparametric regression-models for repeated outcomes in the presence of missing data," *Journal of the American Statistical Association*, vol. 90, pp. 106–121, 1995.

[18] G. Y. Yi and R. J. Cook, "Marginal methods for incomplete longitudinal data arising in clusters," *Journal of the American Statistical Association*, vol. 97, no. 460, pp. 1071–1080, 2002.

[19] J. J. Heckman, "Sample selection bias as a specification error," *Econometrica*, vol. 47, pp. 153–162, 1979.

[20] G. M. Fitzmaurice, G. Molenberghs, and S. R. Lipsitz, "Regression-models for longitudinal binary responses with informative drop-outs," *Journal of the Royal Statistical Society Series B*, vol. 57, pp. 691–704, 1995.

[21] M. G. Kenward, "Selection models for repeated measurements with non-random dropout: an illustration of sensitivity," *Statistics in Medicine*, vol. 17, no. 23, pp. 2723–2732, 1998.

[22] G. Molenberghs, M. G. Kenward, and E. Lesaffre, "The analysis of longitudinal ordinal data with nonrandom drop-out," *Biometrika*, vol. 84, no. 1, pp. 33–44, 1997.

[23] B. F. Kurland and P. J. Heagerty, "Marginalized transition models for longitudinal binary data with ignorable and non-ignorable drop-out," *Statistics in Medicine*, vol. 23, no. 17, pp. 2673–2695, 2004.

[24] K. J. Wilkins and G. M. Fitzmaurice, "A marginalized pattern-mixture model for longitudinal binary data when nonresponse depends on unobserved responses," *Biostatistics*, vol. 8, no. 2, pp. 297–305, 2007.

[25] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society Series B*, vol. 34, pp. 187–220, 1972.

[26] S. Bennett, "Analysis of survival data by the proportional odds model," *Statistics in Medicine*, vol. 2, no. 2, pp. 273–277, 1983.

[27] K. Chen, Z. Jin, and Z. Ying, "Semiparametric analysis of transformation models with censored data," *Biometrika*, vol. 89, no. 3, pp. 659–668, 2002.

[28] S. C. Cheng, L. J. Wei, and Z. Ying, "Analysis of transformation models with censored data," *Biometrika*, vol. 82, no. 4, pp. 835–845, 1995.

[29] D. Clayton and J. Cuzick, "Multivariate generalizations of the proportional hazards model," *Journal of the Royal Statistical Society Series A*, vol. 148, pp. 82–117, 1985.

[30] P. J. Heagerty, "Marginally specified logistic-normal models for longitudinal binary data," *Biometrics*, vol. 55, no. 3, pp. 688–698, 1999.

[31] P. J. Heagerty and S. L. Zeger, "Marginalized multilevel models and likelihood inference," *Statistical Science*, vol. 15, no. 1, pp. 1–26, 2000.
[32] W. H. Press, *Numerical Recipes in C : The Art of Scientific Computing*, Cambridge University Press, New York, NY, USA, 2nd edition, 1992.
[33] M. D. Schluchter, T. Greene, and G. J. Beck, "Analysis of change in the presence of informative censoring: application to a longitudinal clinical trial of progressive renal disease," *Statistics in Medicine*, vol. 20, no. 7, pp. 989–1007, 2001.
[34] P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding, *Statistical Models Based on Counting Processes*, Springer Series in Statistics, Springer, New York, NY, USA, 1993.
[35] R. Tsonaka, G. Verbeke, and E. Lesaffre, "A semi-parametric shared parameter model to handle nonmonotone nonignorable missingness," *Biometrics*, vol. 65, no. 1, pp. 81–87, 2009.