ABSTRACT

YAN, SONG. Joint Modeling of Primary Binary Outcome and Longitudinal Covariates Measured at Informative Observation Times. (Under the direction of Dr. Wenbin Lu and Dr. Daowen Zhang.)

In some biomedical studies, we are interested in the relationship between a primary binary outcome and longitudinal data repeatedly measured over the study period. A common joint modeling approach to this objective is to use subject-specific normal random effects in a mixed model for longitudinal covariates as predictors in a model (e.g., logistic model) for the final binary outcome. Meanwhile, in many cases, the observation times for longitudinal data may be correlated with the longitudinal covarites and the distribution of random effects for longitudinal measurements may not be normally distributed. To take into account the discrete informative observation and the non-normal random effects, in Chapter 1 of this dissertation, we propose to introduce a third model in the joint model for the informative discrete observation times, and relax the normality distributional assumption of random effects using the semi-nonparametric (SNP) approach of Gallant and Nychka (1987) [8]. An EM algorithm is developed for parameter estimation. Extensive simulation designed to evaluate the proposed method indicates that ignoring either informative observation times or distributional assumption of the random effects would lead to invalid and/or inefficient inference. Applying our new approach to the New York University(NYU) in vitro fertilization (IVF) data reveals some interesting findings the traditional approach failed to discover.

In Chapter 2, we extend our joint model to accommodate the cases in which the longitudinal profiles are nonlinear and the informative times of longitudinal covariates are continuous. The extended joint model consists of three components: (1) a semiparametric mixed model (SPMM) for longitudinal covariates; (2) a multiplicative frailty model for informative observation times; and (3) a logistic model for primary binary outcomes. These three submodels are correlated via shared subject-specific random effects. We develop an EM algorithm for computing the maximum likelihood estimates (MLEs) of the parameters, and use an EM-aided numerical differentiation method to estimate the variance-covariance matrix of MLEs.

Joint Modeling of Primary Binary Outcome and Longitudinal Covariates Measured at Informative Observation Times

by
Song Yan

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2011

APPROVED BY:

Dr. Marie Davidian	Dr. Anastasios A. Tsiatis
Dr. Wenbin Lu Co-Chair of Advisory Committee	Dr. Daowen Zhang Chair of Advisory Committee

DEDICATION

To My Loving Family.

BIOGRAPHY

Song Yan was born in Anqin, Anhui Province, China. He obtained a bachelor degree in Information and Electronics Engineering and a master degree in Applied Mathematics from Zhejiang University, China. He came to United States and continued his graduate study in Statistics at North Carolina State University in 2006. He had a very good time at North Carolina State University and will complete his Ph.D in Statistics in November, 2011. In his spare time, he likes movie, swimming and traveling with friends.

ACKNOWLEDGEMENTS

I am very grateful to my advisors, Dr. Daowen Zhang and Dr. Wenbin Lu. This dissertation could not be possible without their guidance. Their encouragement, valuable ideas and consistent support are always the driving forces during my Ph.D. study. They not only help me on specific problems, but more importantly also tell me the right way to do research. I am privileged to have worked with them for more than three years.

I would like to thank Dr. Marie Davidian and Dr. Anastasios Tsiatis for being on my thesis committee and providing useful suggestions to improve the quality of this work. I also appreciate the Department of Statistics for creating a very supportive environment for our graduate students. I am thankful to Dr. Pam Arroway, Dr. Jacqueline M. Hughes-Oliver, Dr. John Monahan and Dr. Sujit Kumar Ghosh for their help during my graduate study.

I am deeply indebted to my parents, my wife and my wife's parents for their endless love and support. Special thanks to my wife, Hancao, for her patience, independence and continuous support. She always stands by me during my ups and downs.

Finally, I offer my regards and blessings to all my friends in Raleigh. I am missing the good time with you in the past five years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

A Seminonparametric Approach to Joint Modeling of A Primary Binary Outcome and Longitudinal Data Measured at Discrete Informative Times

1.1 Introduction

For patients undergoing in vitro fertilization and embryo transfer (IVF-ET), the early determination of pregnancy prognosis is of critical importance [9]. From a medical standpoint, there are increased risks of adverse outcomes in IVF pregnancies compared with natural conceptions, including ectopic pregnancies and spontaneous abortions [9]. The incidence of ectopic pregnancy after IVF is nearly 2 to 5 times higher than that in natural pregnancies [25]. In naturally conceived cycles, among a variety of markers, human chorionic gonadotropin β subunit(β-hCG) has been found highly predictive of normal pregnancy [6]. In fact, because implantation can be timed more accurately in IVF pregnancies, β-hCG curves for IVF pregnancies should be even more accurate [6]. In the fertility literature, various cutoff levels of initial β-hCG measured some specific number of days after ET were proposed to predict viability (e.g., Lambers et al.(2006) [13]). However, a significant number of normal pregnancies often may have β-hCG levels below the established cutoffs [23]. Others suggested that the rising rate of early β-hCG after ET may also have a very strong positive correlation with the pregnancy outcome [5, 23]. Shamonki et al. (2009) [23] depicted logarithmic curves of initial β-hCG level and the rise of early β-hCG and the live delivery outcomes with IVF, and suggested that using the β-hCG curves could give both the clinician and the patient a more accurate assessment of the pregnancy since both the initial β-hCG value and the rising rate of β-hCG curve seem highly correlated with the final pregnancy outcome. Furthermore, Shamonki et al. (2009) [23] demonstrated there is a strong correlation among age, early β-hCG values and pregnancy outcome.

To understand the relationship between final pregnancy outcomes and early antenatal hormonal characteristics among IVF patients, a study conducted at the New York University Fertility Center collected data from patients who underwent IVF treatment between 2001 and 2003 (This retrospective study was approved by the Institutional Review Board at the NYU School of Medicine). For each patient, besides the baseline covariates and the primary binary pregnancy outcome (viable or nonviable) at the end of the study, β-hCG values are repeatedly measured at six potential follow-up time intervals between day 7 and day 23 after ET. Fig. 1.1 shows the observed log β-hCG profiles of 30 randomly selected patients with viable pregnancies and another 30 randomly se-

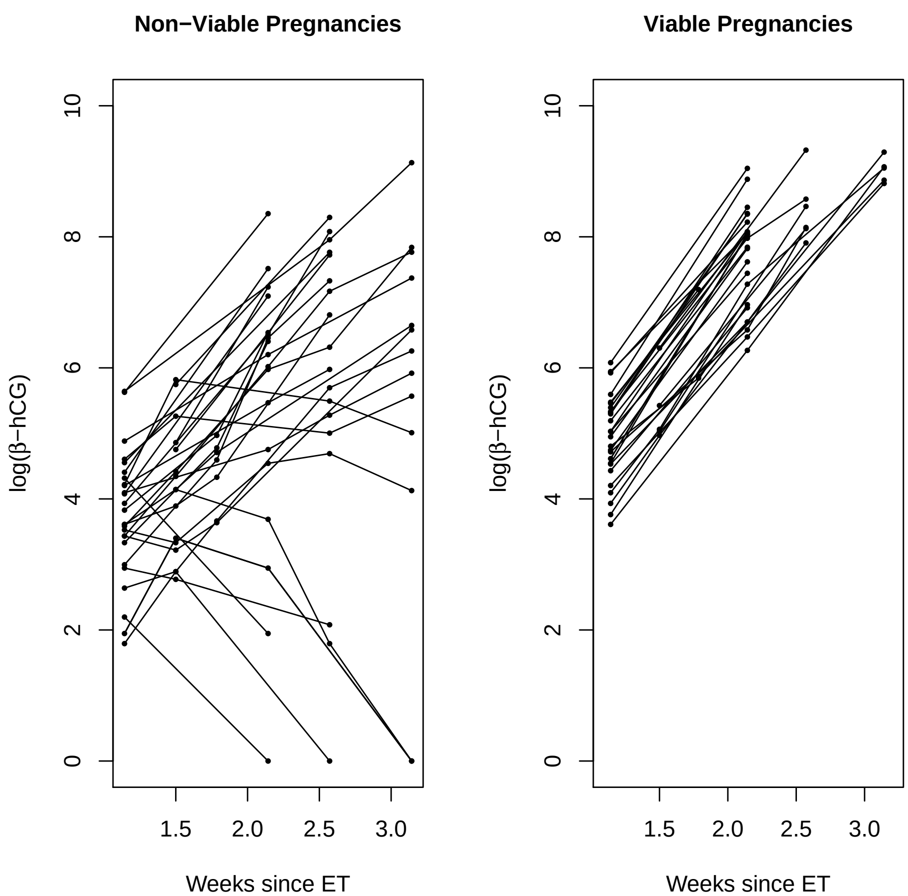

Figure 1.1: Log transformed $\beta-$hCG profiles for 30 randomly selected patients from the non-viable pregnancy group (left) and viable pregnancy group (right), respectively.

lected patients with non-viable pregnancies. From Fig. 1.1, we can observe that the log β-hCG profiles of viable pregnancies show a clearly increasing trend, while those of nonviable pregnancies have a very diverse pattern, a mixture of increasing and decreasing trends, which suggests that patients' β-hCG profiles may be correlated with their final pregnancy outcomes. In this paper, we take a joint modeling approach to studying the relationship between the pregnancy outcome and repeated measures of early β-hCG values after ET. In the statistical literature, there have been abundant works on joint models for a primary outcome (discrete or continuous) and longitudinal covariates. A popular approach is the latent variable approach that uses subject-specific random effects characterizing longitudinal covariate profiles as covariates in the model for the primary outcome ([11, 30, 10, 29, 27, 14, 28], among many others).

One particular challenge in the analysis of NYU-IVF data is that the measurement mechanism of β-hCG profiles might also be informative for the primary pregnancy outcome. Table 1.1 shows the proportion of women having β-hCG measurements at each of the six intervals in the viable and non-viable pregnancy group separately. From Table 1.1 and Fig. 1.1, we observe that women with viable pregnancies tend to have less β-hCG measurements at intermittent time intervals 2 (days 10-11), 3 (days 12-13), 5 (days 17-19) and 6 (days 20-23) compared to those with non-viable pregnancies. To take into account the possible informative pattern of β-hCG measurements, we introduce extra subject-specific logistic submodels, one at each potential observation time (the middle point of a time interval) to model the probability that a β-hCG value will be measured at that time point for a study subject. The correlation among the β-hCG profile, binary pregnancy outcome and discrete observation times are then characterized via latent subject-specific random variables.

In addition, most joint modeling strategies usually assume a normal or other para-

Table 1.1: Proportion of women having β-hCG measures at each time interval.

	1	2	3	4	5	6
Viable	91.7%	15.9%	7.3%	87.0%	15.7%	5.1%
Non-viable	91.7%	50.8%	25.0%	73.5%	37.9%	36.4%

metric distributions for unobserved subject-specific random effects shared by submodels. This assumption can be restrictive and easily violated in many applications. To informally explore the distributions of subject-specific random effects in the linear mixed model for the β-hCG measurements from NYU-IVF data, we fit individual β-hCG profiles by simple linear regression over observation times (here we define the first observation time as time 0 and the unit is week). Fig. 1.2 presents the Q-Q plots of subject-specific intercept and slope estimates from individual least squares fits. Both plots suggest some discrepancy from the normal distribution, while the Q-Q plot of subject-specific slopes shows a larger deviation from the straight line. To take into account this departure from normality, we adopt the seminonparametric (SNP) approach of Gallant and Nychka (1987) [8] to model the distributions of subject-specific random effects. Similar approaches were also used by Zhang and Davidian (2001) [33] and Song, Davidian, and Tsiatis (2002) [24] in other contexts. We develop a maximum likelihood estimation method for the parameters in the joint model via an expectation maximization (EM) algorithm [7], and use an EM-aided numerical differentiation method to compute the variance-covariance matrix of the estimators of interest.

In this chapter we propose a joint model to study the association between a primary binary outcome and longitudinal covariates that are measured at informative discrete occasions. In this joint model, the random effects can follow a flexible distribution as well. To the best of our knowledge, this has not been studied in the literature. The rest of this chapter is organized as follows. Section 1.2 introduces notation and describes the proposed joint model. Section 1.3 presents the inference procedure using an EM algorithm. Section 1.4 applies our method to NYU-IVF data, in which we also investigate the consequences of ignoring the violation of normality assumption and the informative observation times. Section 1.5 provides simulation studies to further justify our method, followed by some discussion in Section 1.6.

1.2 Joint Model

1.2.1 Data Notation

For NYU-IVF data, let Y_i denote the binary pregnancy outcome of subject i, $i = 1, \cdots, n$, obtained at the end of the study, taking the value of 1 if a viable pregnancy is achieved and 0 otherwise. Let X_i denote the baseline covariates, including age and BMI (appropriately standardized for computational stability). Denote the potential measurement occasions of β-hCG profiles by $\mathbf{t} = (t_1, \cdots, t_m)'$, where $t_1 = 0$ is the first observation time point, m is the maximum number of measurements and $m = 6$ for NYU-IVF data. Moreover, denote the complete log β-hCG profile by $\mathbf{Z}_i = (Z_{i1}, \cdots, Z_{im})'$. Define $\mathbf{R}_i = (R_{i1}, \cdots, R_{im})'$, where R_{ij} indicates whether a measurement of the β-hCG value occurs at the jth time point: $R_{ij} = 1$ if a measurement is taken and $R_{ij} = 0$ otherwise. Therefore, Z_{ij} is observed only if $R_{ij} = 1$. The observed data consist of $\{Y_i, X_i, R_{ij}, Z_{ij}R_{ij} : i = 1, \cdots, n; j = 1, \cdots, m\}$.

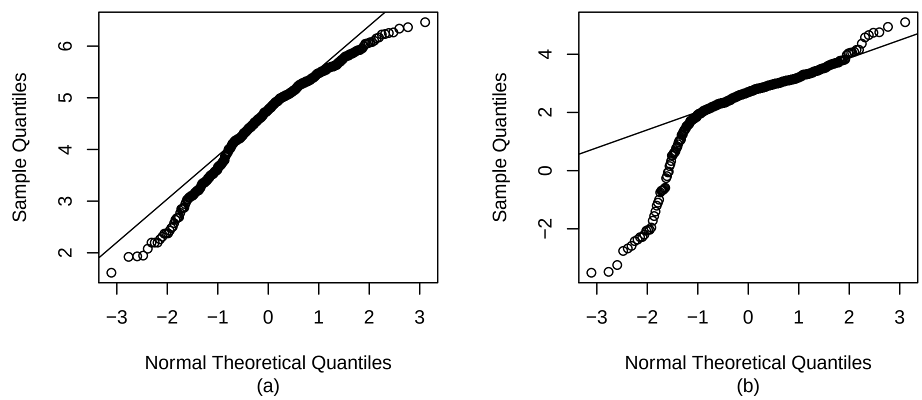

Figure 1.2: (a) Q-Q plot of subject-specific intercept estimates from individual least squares fits. (b) Q-Q plot of subject-specific slope estimates from individual least squares fits.

1.2.2 Model Specifications

Fig. 1.1 indicates that there might be a quadratic trend over time in the longitudinal log β-hCG profile. To allow for the possible quadratic trend, we consider the following linear mixed model for the longitudinal log β-hCG profile Z_{ij} in the ideal situation of no missingness,

$$Z_{ij} = \gamma_0 + \gamma_1 t_j + \gamma_2 t_j^2 + \gamma_3' X_i + u_{i0} + u_{i1} t_j + \epsilon_{ij}, \qquad j = 1, \cdots, m, \tag{1.1}$$

where $\mathbf{u}_i \equiv (u_{i0}, u_{i1})'$ are mean-zero random effects, measuring deviation in the intercept and the changing rate of subject-specific log β-hCG profile at $t_1 = 0$ from the population profile. It is assumed that the error term vector $\boldsymbol{\epsilon}_i \equiv (\epsilon_{i1}, \cdots, \epsilon_{im})'$ is independent of $\mathbf{u}_i$ and follows a multivariate normal distribution $N(0, \Sigma)$, where the variance-covariance matrix Σ has the form $\Sigma = \sigma_\epsilon^2 \Gamma(\rho)$ with σ_ϵ^2 being a positive scalar and $\Gamma(\rho)$ a $m \times m$ Markov structured correlation matrix indexed by ρ. Define $\gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3')'$.

Note that we could add $u_{i2} t_j^2$ in (1.1) to model subject-specific random effect u_{i2} of t_j^2. A preliminary analysis of NYU-IFV data indicates that u_{i2} may not be necessary (the P-value from the correct likelihood ratio test assuming normal distribution for $\mathbf{u}_i$ is 0.29). Furthermore, adding u_{i2} will make the computation challenging for the SNP approach for $\mathbf{u}_i$. Therefore, we restrict two random effects in model (1.1).

For the binary pregnancy outcome Y_i, we assume a logistic model,

$$\text{logit}\{P(Y_i = 1 | X_i, \mathbf{u}_i)\} = \beta_0 + \beta_1' X_i + \alpha_1 u_{i0} + \alpha_2 u_{i1}, \tag{1.2}$$

where α_1 and α_2 are effects of u_{i0} and u_{i1} on the binary outcome Y_i, respectively. Define $\beta = (\beta_0, \beta_1')'$. Finally, to account for the informative measurement pattern of β-hCG profiles as discussed before, we introduce extra logistic models for the measurement indicators R_{ij}'s

$$\text{logit}\{P(R_{ij} = 1|X_i, \mathbf{u}_i)\} = \eta_{0j} + \eta_{1j}' X_i + \alpha_{3j} u_{i0} + \alpha_{4j} u_{i1}, \qquad j = 1, \cdots, m, \tag{1.3}$$

where α_{3j} and α_{4j} are the time-specific effects of u_{i0} and u_{i1} on the measurement indicator. Define $\eta_j = (\eta_{0j}, \eta_{1j}')'$ and $\lambda_j = (\eta_j', \alpha_{3j}, \alpha_{4j})'$, $j = 1, \cdots, m$.

Note that the three submodels (1.1), (1.2) and (1.3) are correlated through shared subject-specific random effects $\mathbf{u}_i$, and the magnitude of the association is controlled by the parameters α_1, α_2, α_{3j} and α_{4j}. Besides $\mathbf{u}_i$, extra randomness may exist. For example, the Markov structured variance-covariance matrix of the error terms in submodel (1.1) describes the additional association in repeated β-hCG measurements that can not be explained by the random effects $\mathbf{u}_i$. Following the convention in joint modeling, we assume that $\mathbf{Z}_i$, $\mathbf{R}_i$ and Y_i are mutually independent given X_i and $\mathbf{u}_i$.

To flexibly model the distribution of the random effects $\mathbf{u}_i$, we adopt the SNP approach proposed by Gallant and Nychka (1987) [8]. This approach assumes that the random effects $\mathbf{u}_i$ belongs to a class of densities that are sufficiently smooth so that they do not exhibit unusual behavior such as kinks, jumps, or oscillation. However the densities are very flexible and are allowed to be skewed, multi-modal, and fat- or thin-tailed (relative to the normal distribution); furthermore, the class includes the normal distribution as a special case [33]. Specifically, SNP approach re-formulates the random effects $\mathbf{u}_i$ as

$$\mathbf{u}_i = \mu + D\mathbf{b}_i \tag{1.4}$$

where $\mu = (\mu_0, \mu_1)'$, D is a 2×2 lower triangular matrix with $\xi \equiv \text{vech}(D) = (d_{00}, d_{10}, d_{11})'$ and the "transformed" random effects $\mathbf{b}_i = (b_{i1}, b_{i2})'$ have a smooth density function

$$h_K(\mathbf{b}) = P_K^2(\mathbf{b})\varphi(\mathbf{b}) = \Bigg(\sum_{0\le i_1+i_2\le K} a_{i_1 i_2} b_1^{i_1} b_2^{i_2}\Bigg)^2 \varphi(\mathbf{b}), \tag{1.5}$$

with $\mathbf{b} = (b_1, b_2)'$, where $\varphi(\cdot)$ is the standard bivariate normal density, K is a non-negative integer, $P_K(\mathbf{b})$ is a bivariate polynomial function of order K with coefficients $a_{i_1 i_2}, i_1, i_2 = 0, 1, \cdots, K$, and $0 \le i_1 + i_2 \le K$. To ensure $h_K(\mathbf{b})$ to be a proper bivariate density function, the coefficients $\{a_{ij}\}$ of $P_K(\mathbf{b})$ must be chosen to satisfy the equality $\int h_K(\mathbf{b})d\mathbf{b} = 1$. Zhang and Davidian (2001) [33] showed that the above condition is equivalent to $E\{P_K^2(\mathbf{U})\} = \mathbf{a}'A\mathbf{a} = 1$, where $\mathbf{U} = (U_1, U_2)'$ is a standard bivariate normal random variable, $\mathbf{a}$ is a $d \times 1$ vector containing all the coefficients $a_{i_1 i_2}$'s, and A is the corresponding matrix with the elements of $E(U_1^{i_1+j_1} U_2^{i_2+j_2})$ with $0 \le i_1 + i_2 \le K$ and $0 \le j_1 + j_2 \le K$. In Appendix A.2 we give more details on the computation of the matrix A. Note that $P_0(\cdot) \equiv 1$, which corresponds to the standard bivariate normal density of $\mathbf{b}_i$. As demonstrated in Zhang and Davidian (2001) [33], in many cases the SNP approach with a K as small as one or two can adequately approximate complicated shapes, including multimodality and skewness. Since A is a symmetric positive definite matrix, hence it has a decomposition $A = BB'$ with some square matrix B. Write $c = B'\mathbf{a}$, then we require $\mathbf{a}'A\mathbf{a} = \mathbf{c}'\mathbf{c} = 1$. As in [33], we consider a polar coordinate transformation of $\mathbf{c} = (c_1, \cdots, c_d)'$, i.e. $c_1 = \sin(\phi_1), c_2 = \cos(\phi_1)\sin(\phi_2), \cdots, c_{d-1} = \cos(\phi_1)\cos(\phi_2)\cos(\phi_{d-2})\sin(\phi_{d-1})$, and $c_d = \cos(\phi_1)\cos(\phi_2)\cos(\phi_{d-2})\cos(\phi_{d-1})$, where $-\pi/2 < \phi_l \le \pi/2$ for $l = 1, 2, \cdots, d-1$. Then $\mathbf{c}'\mathbf{c} = 1$ and thus $\int h_K(\mathbf{b})d\mathbf{b} = 1$ is automatically satisfied by parameterizing $h_K(\mathbf{b})$ in terms of $\phi = (\phi_1, \cdots, \phi_{d-1})'$.

For illustration, let us consider a univariate SNP density with $K = 2$. Using the results in [4], we have $a_0 = \cos(\phi_1) - \sin(\phi_1)\sin(\phi_2)/\sqrt{2}$, $a_1 = \sin(\phi_1)\cos(\phi_2)$ and $a_2 = \sin(\phi_1)\sin(\phi_2)/\sqrt{2}$. Fig. 1.3 plots the resulting SNP densities for some selected values of (ϕ_1, ϕ_2). It is seen clearly that the SNP density can be used to approximate many different densities with distinct features (such as multi-modal, skewed, etc).

1.3 Maximum Likelihood Estimation and Inference

For subject i, let $m_i = \sum_{j=1}^{m} R_{ij}$, and $\mathbf{t}_i^* = (t_{i1}, \cdots, t_{im_i})'$ be the time points where β-hCG values are measured, and denote by $\mathbf{Z}_i^* = (Z_{i1}, \cdots, Z_{im_i})'$ the observed log β-hCG values. The observed data for subject i is denoted as $\mathbf{O}_i$ and write $\mathbf{O} = (\mathbf{O}_i; i = 1, \cdots, n)$. Let $\Theta = \{\gamma, \sigma_\epsilon^2, \rho, \beta, \alpha_1, \alpha_2, \lambda_1, \cdots, \lambda_m, \mu, \xi, \phi\}$ denote the unknown parameters in the proposed joint model. Then the joint likelihood contributed by the observed data is given by

$$L(\Theta) = \prod_{i=1}^{n} \left[\int f_L(\mathbf{Z}_i^*|\mathbf{u}_i) \;\; f_P(Y_i|\mathbf{u}_i) \;\; f_R(\mathbf{R}_i|\mathbf{u}_i) \;\; f_{\mathbf{u}}(\mathbf{u}_i) d\mathbf{u}_i \right], \tag{1.6}$$

where $f_L(\mathbf{Z}_i^*|\mathbf{u}_i)$, $f_P(Y_i|\mathbf{u}_i)$ and $f_R(\mathbf{R}_i|\mathbf{u}_i)$ are conditional density functions (given random effects $\mathbf{u}_i$) of the observed longitudinal log β-hCG values, primary binary outcome and measurement indicators for subject i, respectively, and $f_{\mathbf{u}}(\cdot)$ is the density function of the random effects $\mathbf{u}_i$ derived from the SNP representation (1.4). Here for simplicity, we omit the dependence of these density functions on the observed covariates. Specifically,

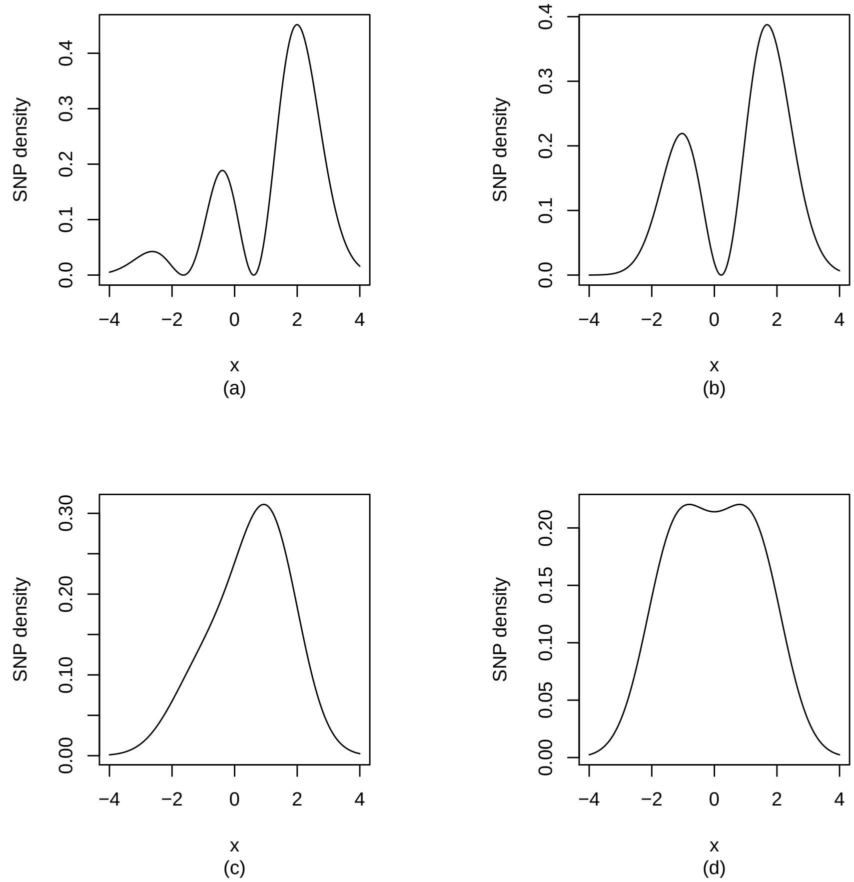

Figure 1.3: Univariate SNP densities with some selected ϕ's for $K = 2$: (a) $\phi_1 = \pi/2$, $\phi_2 = 3\pi/10$; (b) $\phi_1 = \pi/2$, $\phi_2 = \pi/10$; (c) $\phi_1 = \pi/10$, $\phi_2 = 3\pi/10$; (d) $\phi_1 = \pi/10$, $\phi_2 = \pi/2$.

we have

$$
\begin{cases}
f_L(Z_i^*|\mathbf{u}_i;\gamma,\Sigma_i) & = (2\pi)^{-m_i/2}|\Sigma_i|^{-1/2}\exp\Big\{-(\mathbf{Z}_i^*-\mu_i^*)'\Sigma_i^{-1}(\mathbf{Z}_i^*-\mu_i^*)\Big\}, \\
f_P(Y_i|\mathbf{u}_i;\beta,\alpha_1,\alpha_2) & = \frac{\exp\{Y_i(\beta_0+\beta_1'X_i+\alpha_1u_{i0}+\alpha_2u_{i1})\}}{1+\exp\{\beta_0+\beta_1'X_i+\alpha_1u_{i0}+\alpha_2u_{i1}\}}, \\
f_R(\mathbf{R}_i|\mathbf{u}_i;\lambda_1,\cdots,\lambda_m) & = \prod_{j=1}^m \frac{\exp\{R_{ij}(\eta_{0j}+\eta_{1j}'X_i+\alpha_{3j}u_{i0}+\alpha_{4j}u_{i1})\}}{1+\exp\{\eta_{0j}+\eta_{1j}'X_i+\alpha_{3j}u_{i0}+\alpha_{4j}u_{i1}\}}, \\
f_{\mathbf{u}}(\mathbf{u}_i;\mu,\xi,\phi) & = P_K^2\{D^{-1}(\mathbf{u}_i-\mu)\}\varphi\{D^{-1}(\mathbf{u}_i-\mu)\}|D|^{-1},
\end{cases}
\tag{1.7}
$$

where

$$\mu_i^* = \gamma_0 + \gamma_1\mathbf{t}_i^* + \gamma_2\mathbf{t}_i^{*2} + \gamma_3'X_i + u_{i0} + u_{i1}\mathbf{t}_i^* = \mathrm{E}(\mathbf{Z}_i^*|\mathbf{u}_i, X_i),$$

$\mathbf{t}_i^{*2}$ is the vector formed by squaring each element of $\mathbf{t}_i^*$, $\Sigma_i = \sigma_\epsilon^2\Gamma_i(\rho)$ is the associated Markov structured variance-covariance matrix for the observed $\mathbf{Z}_i^*$. Note that it is different from Σ for $\boldsymbol{\epsilon}_i$ in submodel (1.1).

The direct maximization of the likelihood function (1.6) is very challenging, mainly due to two reasons: (a) the integration in (1.6) does not have an analytical form; (b) the number of model parameters is large and may cause numerical instability. This motivates us to develop an EM algorithm to maximize the likelihood function (for a given K). A striking property of the EM algorithm is that the observed data likelihood will always increase during the parameter update process. In addition, the parameter update can be separately carried over for different but smaller subsets of parameters and some parameter updates may even have closed forms.

The details of E-step and M-step of the EM algorithm can be found in Appendix A.1 and A.2. The proposed EM algorithm requires specifying initial values of the parameters. Here we use the naive regression calibration (RC) method to produce the initial values for Θ. To be more specific, we first fit the linear mixed submodel for longitudinal covariates without considering the informative observation times, to obtain the initial estimators $(\widehat{\gamma}^{(0)}, \widehat{\sigma}_{\epsilon}^{2(0)}, \widehat{\rho}^{(0)})$ and the best linear unbiased predictors (BLUP) $\widehat{u}_{i0}^{(0)}$ and $\widehat{u}_{i1}^{(0)}$, $i = 1, \cdots, n$. This can be easily done using standard software packages, e.g., PROC MIXED in SAS or lme in R. Then we fit the logistic models for the binary outcome and measurement indicators with X_i and the BLUP estimates $\widehat{u}_{i0}^{(0)}$ and $\widehat{u}_{i1}^{(0)}$ as covariates to obtain the initial estimators of parameters in (1.2) and (1.3), which can also be implemented directly by standard software packages. As suggested by Zhang (2001)[33], it is crucial to have a good initial value for $(\mu', \xi', \phi')'$ and especially ϕ for any optimization approach. Similar to Song, Davidian, and Tsiatis (2002) [24], we calculate the maximum-likelihood estimate of $(\mu', \xi', \phi')'$ to obtain $(\mu^{(0)}, \xi^{(0)}, \phi^{(0)})'$ by treating the BLUP $\widehat{\mathbf{u}}_i^{(0)} = (\widehat{u}_{i0}^{(0)}, \widehat{u}_{i1}^{(0)})$ as observed data and maximizing the density corresponding to $f_{\mathbf{u}_i}(\widehat{\mathbf{u}}_i^{(0)})$ in (1.7). Standard optimization software package is used to maximize $f_{\mathbf{u}_i}(\widehat{\mathbf{u}}_i^{(0)})$. That optimization procedure for EM initial values needs starting values too, for which we carry out a grid search for starting value of ϕ over $(-\pi/2, \pi/2]$ to ensure that the maximum has been found ,and choose as starting values μ and ξ calculated by moment methods for the given ϕ. The resulting estimators are denoted by $(\widehat{\mu}^{(0)}, \widehat{\xi}^{(0)}, \widehat{\phi}^{(0)})$. The details of this procedure can be found in the Appendix A.3. For the convergence criterion of the proposed EM algorithm, we consider $\max|\widehat{\Theta}^{(k+1)} - \widehat{\Theta}^{(k)}| < \delta$ with $\delta = 0.0001$ in our numerical studies. We use the Gaussian quadrature method with 25 quadrature knots to approximate the two dimensional integrations used in the E-step.

In the SNP representation, the parameter K is a tuning parameter, controlling the flexibility of the random effects distribution, which is needed to be chosen based on the data. Here we select K based on various information criteria that all take the form of

$-l(\Theta)/N + C(N)p_{net}/N$, where $l(\Theta)$ is the log likelihood, $N = \sum_{i=1}^{n} m_i$ and p_{net} is the number of free parameters in the joint model. For example, $C(N) = 1$ is for the Akaike Information Criterion (AIC), $C(N) = 0.5\log N$ for the Schwarz Information Criterion (BIC), and $C(N) = \log\log(N)$ for the Hannan-Quinn criterion (HQ). As discussed by Zhang and Davidian (2001) [33], HQ criterion is often preferred in the selection of K. We will evaluate the empirical performance of these information criteria via simulations in Section 1.5.

For parametric models, Louis's formula [21] can be used to compute the variance of the estimates obtained by the EM algorithm. However, for our proposed SNP likelihood approach, the number of parameters in the joint model can be large and our interest mainly focuses on the regression parameters for the effects of the baseline predictors and random effects. The direct calculation of the variance for the EM estimates based on Louis's formula may be unstable. Here we calculate the variance for the estimates of the regression parameters by inverting the observed information matrix based on the corresponding profile log likelihood function. In general, the observed information matrix does not have a closed analytical form. We compute it using an EM-aided numerical differentiation method, which was studied by Meilijson (1989) [22] for parametric models and then extended to the proportional hazards model with missing covariates by Chen and Little (1999) [3]. Zeng and Cai (2005) [31] used a similar approach to variance estimation based on a profile likelihood function. The details of the EM-aided numerical differentiation method are given in Appendix A.4.

1.4 Application to NYU-IVF Data

The NYU-IVF data consist of 540 pregnancies obtained after IVF treatment at the New York University Fertility Center from 2001 to 2003. In this study, viable pregnancies are defined as pregnancies reaching the second trimester, including singleton, twins, higher order multiples and stillbirths; non-viable pregnancies include biochemical pregnancies, ectopic pregnancies and first trimester abortions [2]. The average age of participants was 35.27 years (SD= 4.40 years) and the average body mass index (BMI) was 23.71 (SD= 5.05). So for computational stability and ease of interpretation, age is centered at 35 and divided by 10, BMI is centered at 23 and divided by 10. The possible observation times of β-hCG are rounded to the median days of each six potential time intervals from day 7 to day 23 after ET. Interval 1 is days 7-9, interval 2 is days 10-11, interval 3 is days 12-13, interval 4 is days 14-16, interval 5 is days 17-19, and interval 6 is days 20-23. We define the first observation time as time 0 and transform the day unit to week unit. The total number of β-hCG observations is $N = 1325$.

We fit the NYU-IVF data by our proposed joint model using the SNP likelihood-based approach with $K = 0$, 1 and 2 ($K = 0$ means normality assumption for $\mathbf{u}_i$). For comparison, we also analyzed the NYU-IVF data by a reduced joint model in which the informative measurement mechanism of the β-hCG profiles is ignored. When choosing the K in the SNP representation, we use the AIC, BIC and HQ criteria. The log likelihood for full model for $K = 0, 1, 2$ are $-1732.86, -1641.57$ and -1635.49; the AIC values are $1.3380, 1.2706$ and 1.2683; the HQ values are $1.3674, 1.3014$ and 1.3013; the BIC values are $1.4163, 1.3528$ and 1.3564. The log likelihood for reduced model for $K = 0, 1, 2$ are $-770.39, -678.03$ and -673.38; the AIC values are $0.5927, 0.5245$ and 0.5233; the HQ values are $0.6037, 0.5370$ and 0.5379; the BIC values are $0.6221, 0.5578$ and 0.5624. Two of three criteria choose $K = 2$ for the full joint models and $K = 1$ for the reduced joint models, which again supports our previous discovery that the subject-specific random effects are not normally distributed. The parameter estimates with their estimated standard errors computed based on the proposed EM-aided numerical differentiation method for the full and reduced joint models are summarized in Tables 1.2, 1.3 and 1.4, where Tables 1.2 and 1.3 summarize the results for the linear mixed model and the logistic model for the primary binary outcome while Table 1.4 for the logistic models for the measurement indicators.

Columns (A) and (B) in Table 1.2 are the results for our proposed joint model with $K = 0$ and 2, respectively. We make the following observations: (1) age is significantly negatively associated with the final pregnancy outcome, while BMI is not; (2) BMI is significantly negatively associated with the β-hCG profile, while age is not; (3) greater baseline value and stronger increasing trend of the β-hCG profile at baseline are associated with higher chance of viable pregnancy (e.g., $\widehat{\alpha}_1 = 1.610$ and $\widehat{\alpha}_2 = 1.527$ when $K = 2$, with the p-values of both < 0.05); (4) the point estimates for $K = 0$ and 2 are very comparable except for estimates of α_2, while the estimates for $K = 2$ show certain efficiency gain for some parameters. We will further investigate the efficiency of estimates for $K = 0$ and 2 by simulations in Section 1.5.

Columns (C) and (D) in Table 1.3 are the results for the reduced model ignoring the informative measurement times with $K = 0$ and 1, respectively. Comparing column (D) (Table 1.3) to column (B) (Table 1.2), we observe that there is some discrepancy in the estimated population curves of the log β-hCG profiles. This is further supported by the results from the simulation in Section 1.5, which demonstrates that ignoring the informative measurement pattern of longitudinal data may cause an estimation bias of the mean slope parameter in the linear mixed model. More importantly, we find that the estimate of α_1 (the effect of the random intercept on the final pregnancy outcome) becomes insignificant in the reduced model, due to a small point estimate and an increased standard error estimate.

From the results in Table 1.4 we observe that the estimates for the random intercepts (α_{3j}'s) are negatively significant at intermediate times t_2, t_3, t_5 and t_6, while those for the random slopes (α_{4j}'s) are not significant. This suggests that doctors may require more intermediate measurements of β-hCG values for women with lower β-hCG values at t_1. Furthermore, age and BMI do not have significant effects on most of the measurement indicators.

The estimated joint and marginal density functions of the random effects $\mathbf{u}_i$ for the full model with $K = 2$ are given in Fig. 1.4: (a) plots the estimated bivariate density function, which shows a bimodal distribution; (b) gives the contour plot corresponding to the bivariate density in (a) together with the posterior estimates of the $\mathbf{u}_i$'s, which also shows clearly two clusters; (c) and (d) are estimated marginal densities of u_{i0} and u_{i1} respectively, together with their histograms. It is seen that the estimated density of u_{i1}'s also shows a clear bimodal, implying the deviation from the normal distribution. The bimodal distribution of the subject-specific random effects indicates that the underlying study population may be a mixture of two populations, separated most by the latent variable u_{i1}. According to the results for the primary pregnancy outcome, we infer that one sub-population has a higher likelihood to have a viable pregnancy, while the other has a lower likelihood.

Table 1.2: Estimation of parameters in the β-hCG longitudinal submodel and binary pregnancy outcome submodel in the proposed full model for NYU-IVF data

	Full model K=0 (A)			Full model K=2 (B)		
Para	Est	SE	p-value	Est	SE	p-value
longitudinal submodel for β-hCG						
γ_0(intercept)	4.605	0.071	<0.0001	4.606	0.066	<0.0001
γ_1 (slope)	2.952	0.170	<0.0001	2.881	0.144	<0.0001
γ_2 (quadratic)	-0.432	0.122	0.0003	-0.481	0.111	<0.0001
$\gamma_{3,1}$(age)	-0.185	0.191	0.33	-0.190	0.155	0.22
$\gamma_{3,2}$(BMI)	-0.385	0.156	0.01	-0.371	0.124	0.003
σ_e^2	0.395			0.385		
ρ	0.276	0.061	<0.0001	0.303	0.066	<0.0001
binary outcome submodel for viable pregnancy						
β_0	1.633	0.216	<0.0001	1.583	0.199	<0.0001
$\beta_{1,1}$(age)	-1.180	0.403	0.003	-1.104	0.361	0.002
$\beta_{1,2}$(BMI)	-0.352	0.260	0.17	-0.359	0.229	0.12
α_1	1.523	0.447	0.0006	1.610	0.324	<0.0001
α_2	2.059	0.720	0.004	1.527	0.415	0.0002

SE is the estimated standard error.

Table 1.3: Estimation of parameters in the β-hCG longitudinal submodel and binary pregnancy outcome submodel in the reduced model for NYU-IVF data

	Reduced model K=0 (C)			Reduced model K=1 (D)		
Para	Est	SE	p-value	Est	SE	p-value
longitudinal submodel for β-hCG						
γ_0(intercept)	4.613	0.057	<0.0001	4.729	0.048	<0.0001
γ_1 (slope)	2.593	0.153	<0.0001	2.610	0.126	<0.0001
γ_2 (quadratic)	-0.357	0.101	0.0004	-0.349	0.093	0.0001
$\gamma_{3,1}$(age)	-0.189	0.167	0.25	-0.183	0.141	0.19
$\gamma_{3,2}$(BMI)	-0.396	0.132	0.002	-0.378	0.115	0.001
σ_e^2	0.392			0.367		
ρ	0.288	0.063	<0.0001	0.305	0.058	<0.0001
binary outcome submodel for viable pregnancy						
β_0	1.609	0.211	<0.0001	1.642	0.209	<0.0001
$\beta_{1,1}$(age)	-1.283	0.398	0.001	-1.083	0.342	0.001
$\beta_{1,2}$(BMI)	-0.347	0.253	0.17	-0.360	0.223	0.10
α_1	1.415	0.747	0.058	1.278	0.692	0.064
α_2	1.762	0.639	0.005	1.609	0.627	0.010

SE is the estimated standard error.

Table 1.4: Estimation of parameters in β-hCG observation times submodel of the proposed full model for NYU-IVF data

	Full model K=0 (A)			Full model K=2 (B)		
Para	Est	SE	p-value	Est	SE	p-value
observation times model						
η_{02}	-2.233	0.745	0.002	-2.412	0.415	<0.0001
$\eta_{12,1}$(age)	0.110	0.429	0.79	0.175	0.436	0.68
$\eta_{12,2}$(BMI)	0.258	0.350	0.46	0.367	0.327	0.26
α_{32}	-5.066	1.470	0.0005	-4.684	1.022	<0.0001
α_{42}	1.320	0.513	0.01	0.636	0.345	0.07
η_{03}	-2.668	0.259	<0.0001	-2.674	0.221	<0.0001
$\eta_{13,1}$(age)	0.384	0.380	0.31	0.397	0.361	0.27
$\eta_{13,2}$(BMI)	-0.259	0.318	0.41	-0.249	0.308	0.41
α_{33}	-2.649	1.059	0.012	-2.321	0.476	<0.0001
α_{43}	0.771	0.610	0.20	0.383	0.305	0.20
η_{04}	2.697	0.921	0.003	2.178	0.217	<0.0001
$\eta_{14,1}$(age)	-0.401	0.409	0.32	-0.366	0.310	0.24
$\eta_{14,2}$(BMI)	-0.532	0.338	0.11	-0.492	0.275	0.073
α_{34}	4.083	1.434	0.004	2.137	0.460	<0.0001
α_{44}	-1.413	1.763	0.42	-0.331	0.211	0.11
η_{05}	-2.053	0.553	0.0002	-1.894	0.190	<0.0001
$\eta_{15,1}$(age)	0.837	0.399	0.035	0.698	0.318	0.028
$\eta_{15,2}$(BMI)	0.562	0.286	0.049	0.531	0.207	0.01
α_{35}	-3.539	1.675	0.03	-2.068	0.513	<0.0001
α_{45}	1.180	1.402	0.40	0.664	0.441	0.13
η_{06}	-2.438	0.228	<0.0001	-2.425	0.213	<0.0001
$\eta_{16,1}$(age)	0.148	0.345	0.66	0.190	0.317	0.55
$\eta_{16,2}$(BMI)	0.237	0.255	0.35	0.286	0.243	0.23
α_{36}	-2.059	0.491	<0.0001	-1.994	0.527	0.0001
α_{46}	0.242	0.273	0.37	0.188	0.187	0.31

SE is the estimated standard error.

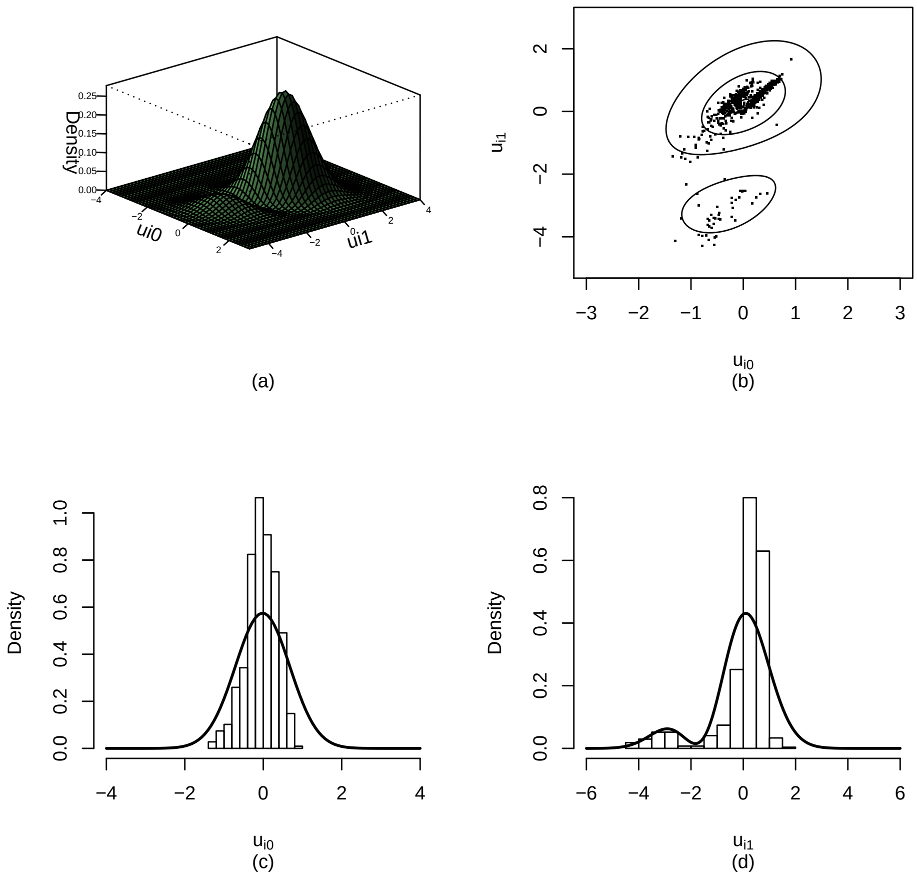

Figure 1.4: (a) Estimated bivariate density of $\mathbf{u}_i$; (b) Contour plots of the density in (a) with posterior estimates of $\mathbf{u}_i$ superimposed (contours are 5, 50 and 95%); (c) and (d) Estimated marginal densities for the random intercept and slope, superimposed by the histograms for posterior estimates of $\mathbf{u}_i$, respectively.

1.5 Simulation Studies

Simulation was conducted to evaluate the performance of the proposed method under practical settings. We used the same linear mixed model for longitudinal covariates (without a quadratic term of t_j for simplicity) and logistic model for the primary binary outcome. For the informative measurement indicators, we consider the following simpler logistic models:

$$\text{logit}\{P(R_{ij} = 1|X_i, \mathbf{u}_i)\} = \eta_{0j} + \eta_1' X_i + \alpha_3 u_{i0} + \alpha_4 u_{i1}, \qquad j = 1, \cdots, m, \tag{1.8}$$

where the effects of covariates and subject-specific random variables on the measurement indicators are assumed to be the same over all time points. We consider two baseline covariates: a binary covariate X_{i1} sampled from the Bernoulli distribution with a success probability of 0.5 and a continuous covariate X_{i2} sampled from the standard normal distribution. Longitudinal covariates $\mathbf{Z}_i$ were generated at six possible observation times $(t_1, \cdots, t_6) = (1, 2, 3, 4, 5, 6)$ in weeks. The subject-specific random effects $\mathbf{u}_i$'s were generated from a mixture of two bivariate normal distributions: $F(\mathbf{u}_i) = 0.7\Phi(\mathbf{u_i}; \mu_1, V) + 0.3\Phi(\mathbf{u_i}; \mu_2, V)$ with $\mu_1 = (1.56, 0)'$, $\mu_2 = (-3.64, 0)'$ and the elements of the covariance matrix V given by $v_{00} = 0.81, v_{01} = v_{10} = -0.0456$ and $v_{11} = 1.96$. The true values of the regression parameters are shown in Tables 1.5. We run 100 simulation with sample size of $n = 300$.

We consider the SNP representation with $K = 0, 1$ and 2, and choose the optimal K values by the AIC, HQ and BIC criterion respectively. Simulation results are summarized in Table 1.5 and Table 1.7. Here we report the results for $K = 0$ corresponding to the bivariate normal distribution of the random effects, and the optimal estimates chosen by AIC, HQ and BIC information criterion. For comparison, we also report the estimation results based on the reduced model ignoring the informative measurement process in Table 1.6 (for $K = 0$ and HQ) and Table 1.8 (for AIC and BIC). We choose to use HQ results as illustration for our simulation since the results by AIC, HQ and BIC criteria are very similar. For our proposed method (Table 1.5), the estimates of all the parameters are nearly unbiased in all cases, the SE's are close to the SD's, and all the CP's are close to the nominal level. This implies that the HQ information criterion performs well in selecting the tuning parameter K. Similar observations were also found in the literature [24]. However, the estimators assuming the incorrect normality assumption for the random effects ($K = 0$) may be less efficient than those estimated by SNP with the K value chosen by the HQ criterion (for example, the SD of $\widehat{\gamma}_{2,1}$ is 0.156 for $K = 0$, but 0.079 for HQ; the SD for $\widehat{\gamma}_{2,2}$ is 0.327 for $K = 0$, but 0.160 for HQ). In terms of the selection of the tuning parameter K by HQ, the values 0, 1 and 2 were chosen 2%, 84% and 14% out of 100 runs, indicating the ability of the HQ criterion to detect the departure of the random effects distribution from normality. These selection proportions by AIC and BIC are 1%, 78%, 21% and 2%, 88%, 10% respectively.

The results for the reduced model are shown Table 1.6. We can see that the estimates of the mean slope parameter (γ_1) in the linear mixed model for longitudinal covariates show bigger biases compared to those of the full model and the associated empirical coverage probability (CP) of the 95% Wald-type confidence interval is significantly below the nominal level ($\leq 87\%$). We also see that the paramter estimates in the primary binary outcome model exhibit some degree of bias. In particular, there is about 10% bias in the regression parameter estimate $\widehat{\beta}_{1,1}$, and 7% to 12% bias in the parameter estimate $\widehat{\alpha}_1$. In addition, these parameter estimates tend to have greater variability than those from our proposed method. This suggests that ignoring the informative measurement process may cause invalid or less efficient inference for some parameters in the joint model. Table 1.8

gives similar results for AIC and BIC criteria.

In Fig. 1.5, Fig. 1.6 and Fig. 1.7, we plot the true and estimated (based on the means of 100 simulations) joint and marginal densities of the subject-specific random effects $\mathbf{u}_i$ by HQ, AIC and BIC criteria: (a) and (b) are for the true and estimated bivariate densities, respectively; (c) and (d) are for the marginal densities of the random intercept and baseline slope, respectively. It can be seen that the SNP approach can estimate the true bivariate and marginal densities of random effects very well, demonstrating the flexibility of the SNP representation to capture the complex distribution of random effects that are different from the normal.

1.6 Discussion

In this chapter, we propose a joint model that can naturally study the association between a primary binary outcome and longitudinal covariates that are possibly measured at discrete informative observation times. The association is described by latent subject-specific random variables whose distribution is flexibly modeled by the SNP representation, allowing the departure from normal assumption. Our approach provides a great insight into the relationship between early β-hCG profiles and the final pregnancy outcomes after IVF treatment. The results demonstrate the power of using the latent characteristics of early stage β-hCG profiles for the prediction of viable pregnancies achieved by IVF treatment, and the importance to take into account the informative measurement process and non-normally distributed random effects for data analysis.

Other modeling strategies proposed in the literature may also be used to analyze NYU-IVF data. A particular one tries to use the β-hCG value measured at some specific time point as a covariate in the logistic model for the primary pregnancy outcome. This

Table 1.5: Simulation results: $K = 0$ denotes the estimation by assuming normal random effects. HQ is the estimation when K is selected by the HQ criterion.

		K=0				HQ			
Para	true	Est	SD	SE	CP	Est	SD	SE	CP
					Proposed full joint model:				
longitudinal submodel									
γ_0	1	0.987	0.237	0.220	93 %	0.986	0.175	0.177	96 %
γ_1	0.5	0.498	0.080	0.087	97 %	0.502	0.080	0.085	96 %
$\gamma_{2,1}(X_1)$	0.5	0.483	0.156	0.157	94 %	0.502	0.079	0.077	93 %
$\gamma_{2,2}(X_2)$	-0.5	-0.490	0.327	0.312	94 %	-0.504	0.160	0.161	96 %
σ_e^2	1	0.993				1.005			
ρ	0.3	0.296				0.304			
binary outcome submodel									
β_0	0.8	0.783	0.396	0.375	93 %	0.814	0.373	0.361	96 %
$\beta_{1,1}(X_1)$	-1	-1.048	0.277	0.286	98 %	-1.050	0.237	0.245	97 %
$\beta_{1,2}(X_2)$	-1.5	-1.568	0.522	0.531	95 %	-1.550	0.434	0.442	93 %
α_1	-1	-1.071	0.173	0.171	96 %	-1.041	0.152	0.156	98 %
α_2	1	1.061	0.222	0.209	97 %	1.048	0.202	0.204	92 %
informative times submodel									
η_{01}	1.6	1.637	0.213	0.216	96 %	1.617	0.199	0.205	96 %
η_{02}	1	1.037	0.218	0.206	93 %	1.015	0.212	0.201	93 %
η_{03}	1.5	1.520	0.209	0.213	95 %	1.498	0.195	0.189	94 %
η_{04}	1	1.006	0.222	0.205	95 %	0.984	0.199	0.193	94 %
η_{05}	1	1.040	0.219	0.206	94 %	1.019	0.189	0.190	96 %
η_{06}	1.4	1.402	0.220	0.211	93 %	1.385	0.192	0.183	94 %
$\eta_{1,1}(X_1)$	1	0.991	0.104	0.110	95 %	1.002	0.082	0.087	96 %
$\eta_{1,2}(X_2)$	-0.5	-0.502	0.228	0.204	94 %	-0.513	0.157	0.149	94 %
α_3	0.5	0.526	0.043	0.041	94 %	0.510	0.040	0.040	96 %
α_4	-0.5	-0.504	0.052	0.058	98 %	-0.502	0.051	0.059	98 %
$\%K$						(2, 84, 14)			

SD is the sample standard deviation; SE is the mean of estimated standard errors; CP is the empirical coverage probability of 95% Wald-type confidence intervals; $\%K$ represents the proportions of $K = 0$, 1 or 2 out of 100 runs selected by the HQ criterion.

Table 1.6: Simulation results: $K = 0$ denotes the estimation by assuming normal random effects. HQ is the estimation when K is selected by the HQ criterion.

		K=0				HQ			
Para	true	Est	SD	SE	CP	Est	SD	SE	CP
					Reduced model:				
longitudinal submodel									
γ_0	1	1.014	0.237	0.218	91%	0.982	0.178	0.169	90%
γ_1	0.5	0.401	0.079	0.085	84%	0.408	0.078	0.086	87%
$\gamma_{2,1}(X_1)$	0.5	0.479	0.157	0.156	94%	0.501	0.080	0.085	96%
$\gamma_{2,2}(X_2)$	-0.5	-0.488	0.326	0.310	94%	-0.504	0.164	0.175	97%
σ_e^2	1	0.992				1.022			
ρ	0.3	0.290				0.295			
binary outcome submodel									
β_0	0.8	0.724	0.395	0.383	93 %	0.819	0.381	0.330	88%
$\beta_{1,1}(X_1)$	-1	-1.096	0.312	0.304	99 %	-1.102	0.249	0.256	99%
$\beta_{1,2}(X_2)$	-1.5	-1.564	0.535	0.551	96 %	-1.527	0.448	0.457	94%
α_1	-1	-1.124	0.212	0.197	100 %	-1.073	0.157	0.169	98%
α_2	1	1.088	0.260	0.234	96 %	1.074	0.210	0.210	96%
$\%K$						(1, 77, 22)			

SD is the sample standard deviation; SE is the mean of estimated standard errors; CP is the empirical coverage probability of 95% Wald-type confidence intervals; $\%K$ represents the proportions of $K = 0$, 1 or 2 out of 100 runs selected by the HQ criterion.

Table 1.7: Simulation results: AIC and BIC are estimation of K's preferred by AIC and BIC, respectively.

		AIC				BIC			
Para	true	Est	SD	SE	CP	Est	SD	SE	CP
				Proposed full joint model:					
longitudinal submodel									
γ_0	1	0.984	0.173	0.176	96 %	0.986	0.176	0.177	96 %
γ_1	0.5	0.502	0.080	0.084	96 %	0.502	0.080	0.086	95 %
$\gamma_{2,1}(X_1)$	0.5	0.501	0.079	0.076	93 %	0.503	0.078	0.077	93 %
$\gamma_{2,2}(X_2)$	-0.5	-0.504	0.159	0.160	95 %	-0.503	0.162	0.161	96 %
σ_e^2	1	1.006				1.005			
ρ	0.3	0.302				0.305			
binary outcome submodel									
β_0	0.8	0.816	0.375	0.364	96 %	0.815	0.372	0.362	96 %
$\beta_{1,1}(X_1)$	-1	-1.051	0.237	0.241	97 %	-1.050	0.237	0.245	97 %
$\beta_{1,2}(X_2)$	-1.5	-1.547	0.434	0.440	93 %	-1.553	0.434	0.443	93 %
α_1	-1	-1.040	0.150	0.156	98 %	-1.042	0.152	0.157	97 %
α_2	1	1.050	0.201	0.202	93 %	1.047	0.203	0.203	92 %
informative times submodel									
η_{01}	1.6	1.615	0.199	0.203	96 %	1.616	0.199	0.204	96 %
η_{02}	1	1.014	0.212	0.201	93 %	1.015	0.211	0.202	94 %
η_{03}	1.5	1.497	0.194	0.187	93 %	1.499	0.196	0.190	95 %
η_{04}	1	0.983	0.198	0.193	94 %	0.984	0.198	0.192	94 %
η_{05}	1	1.018	0.189	0.190	95 %	1.019	0.189	0.191	96 %
η_{06}	1.4	1.389	0.192	0.181	94 %	1.386	0.192	0.183	94 %
$\eta_{1,1}(X_1)$	1	1.001	0.082	0.085	96 %	1.003	0.082	0.088	97 %
$\eta_{1,2}(X_2)$	-0.5	-0.515	0.155	0.149	94 %	-0.511	0.159	0.149	94 %
α_3	0.5	0.509	0.039	0.041	95 %	0.510	0.040	0.040	96 %
α_4	-0.5	-0.501	0.052	0.058	98 %	-0.502	0.051	0.060	97 %
$\%K$		(1, 78, 21)				(2, 88, 10)			

Table 1.8: Simulation results: AIC and BIC are estimation of K's preferred by AIC and BIC, respectively.

		AIC				BIC			
Para	true	Est	SD	SE	CP	Est	SD	SE	CP
					Reduced model:				
longitudinal submodel									
γ_0	1	0.981	0.175	0.168	90%	0.982	0.180	0.171	91%
γ_1	0.5	0.408	0.078	0.086	87%	0.409	0.078	0.087	87%
$\gamma_{2,1}(X_1)$	0.5	0.500	0.080	0.084	96%	0.502	0.080	0.085	96%
$\gamma_{2,2}(X_2)$	-0.5	-0.506	0.163	0.176	97%	-0.504	0.164	0.175	96%
σ_e^2	1	1.023				1.021			
ρ	0.3	0.294				0.295			
binary outcome submodel									
β_0	0.8	0.826	0.385	0.330	89%	0.816	0.381	0.330	88%
$\beta_{1,1}(X_1)$	-1	-1.099	0.243	0.256	99%	-1.104	0.251	0.256	99%
$\beta_{1,2}(X_2)$	-1.5	-1.523	0.450	0.458	95%	-1.527	0.449	0.456	93%
α_1	-1	-1.072	0.155	0.166	98%	-1.073	0.159	0.170	98%
α_2	1	1.076	0.208	0.209	95%	1.073	0.212	0.211	97%
$\%K$		(1, 70, 29)				(1, 88, 11)			

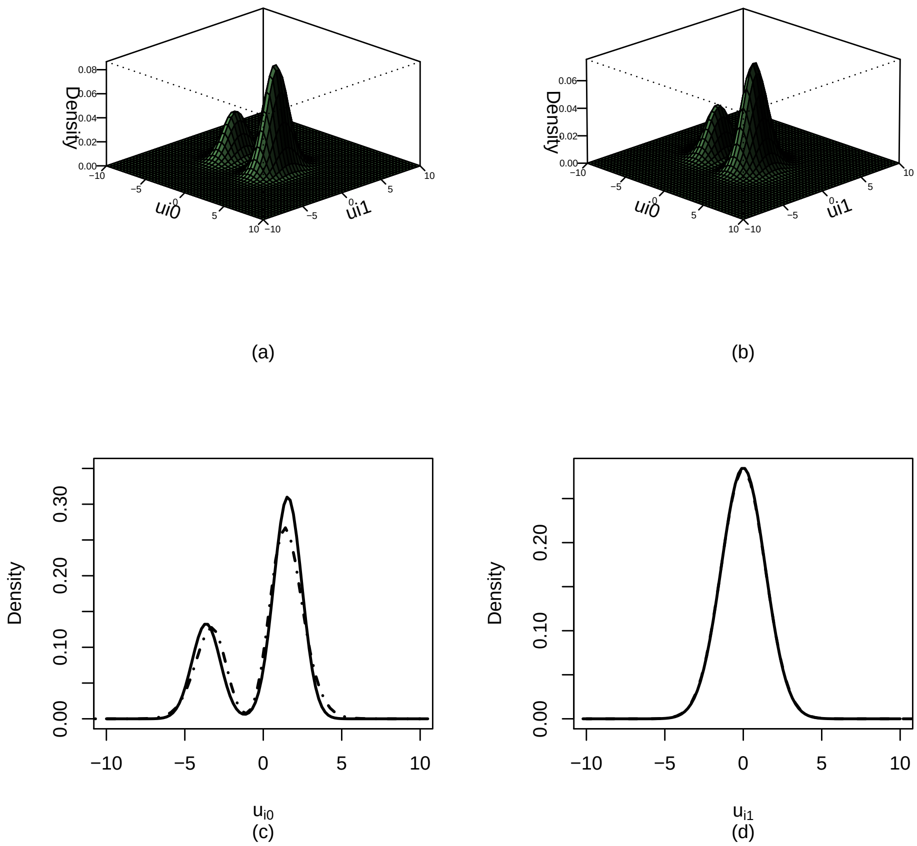

Figure 1.5: (a) and (b) True and estimated bivariate density of $\mathbf{u}_i$ by HQ; (c) and (d) True marginal density (solid line) and estimated marginal density chosen by HQ (dash-dotted line) for the random intercept an slope, respectively.

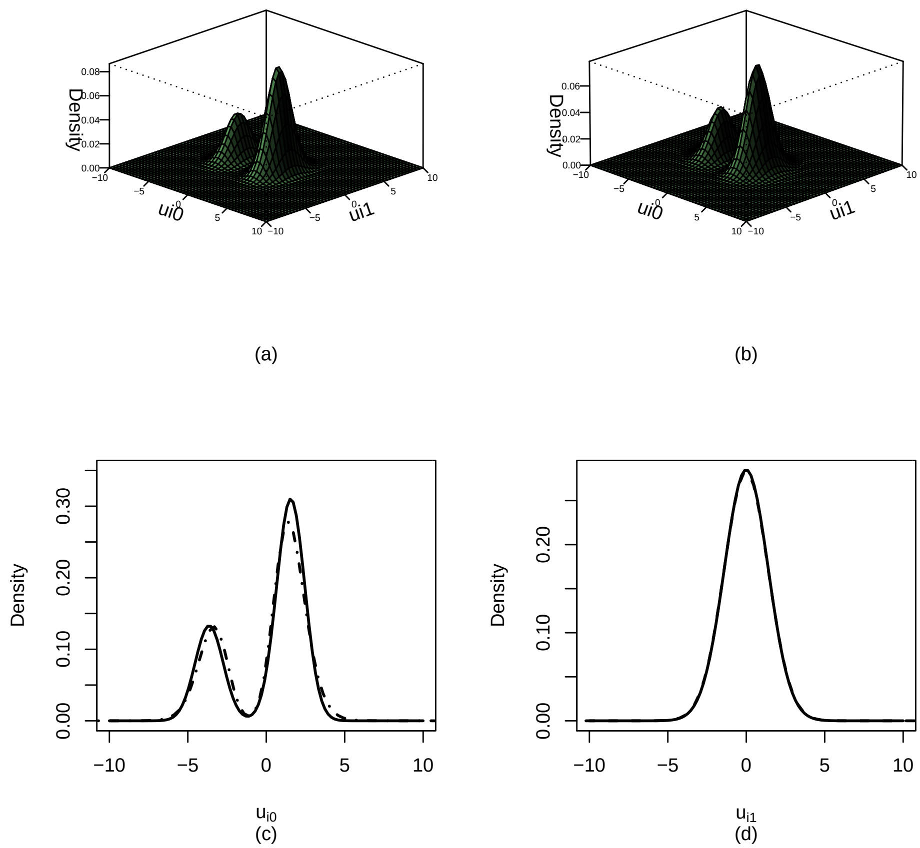

Figure 1.6: (a) and (b) True and estimated bivariate density of $\mathbf{u}_i$ by AIC; (c) and (d) True marginal density (solid line) and estimated marginal density chosen by AIC (dash-dotted line) for the random intercept an slope, respectively.

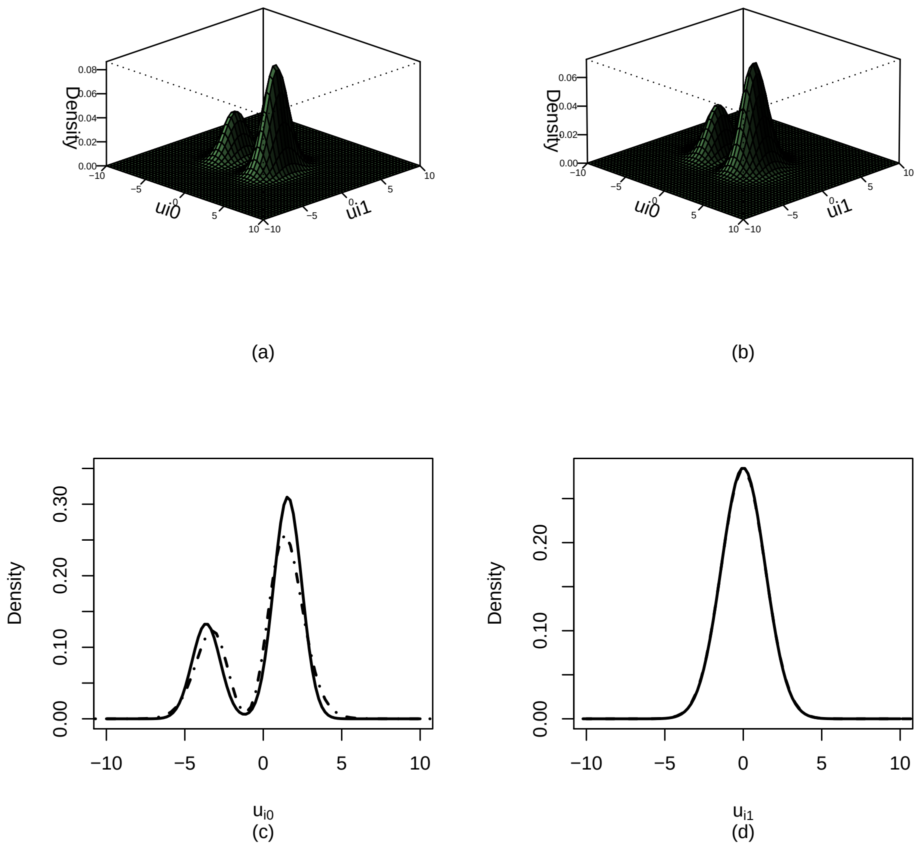

Figure 1.7: (a) and (b) True and estimated bivariate density of $\mathbf{u}_i$ by BIC; (c) and (d) True marginal density (solid line) and estimated marginal density chosen by BIC (dash-dotted line) for the random intercept an slope, respectively.

approach may be problematic for NYU-IVF data for the following reasons. First, the observation time points are different for different women. Therefore, it is not possible to select a common time point where all women had β-hCG measured. Second, if we could find a common time point at which β-hCG measurements were available for all women, using the β-hCG value at this particular time point as a covariate may not have a scientific and/or clinical justification. Third, suppose we had β-hCG data at a clinically meaningful time point, there is usually some degree of measurement error in this hormone data. It is well-known that putting an error-prone variable as a covariate in a logistic model will lead to a biased result.

If we follow the convention in mixed model literature to interpret $Z_i(t) = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \gamma_3^T X_i + u_{i0} + u_{i1} t$ as the true log β-hCG of subject i (with covariates X_i) at time point t, our joint model then implies that the viable pregnancy probability has a logistic regression model as follows

$$\text{logit}\{P(Y_i = 1|X_i, t, \mathbf{u}_i)\} = \beta_0(t) + \tilde{\beta}_1' X_i + \alpha_1 Z_i(t) + (\alpha_2 - \alpha_1 t) Z_i'(t).$$

That is, given the results from our joint model, we can simultaneously investigate the effects of the true log β-hCG and its changing rate at any time on the primary pregnancy outcome. Since both $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ are positive, changing rate of the true log β-hCG at baseline is most predictive of the viable pregnancy outcome. Similarly, we can show from our joint model that the logit of a viable pregnancy probability is (linearly) related to the average of the true log β-hCG and its changing rate in any specific interval $[0, T]$.

Chapter 2

Joint Modeling of Primary Binary Outcome and Longitudinal Covariates Measured at Continuous Informative Observation Times

2.1 Introduction

In this chapter, we will extend the joint model of chapter 1 to continuous observation times cases and use a semiparametric mixed model to characterize the nonlinear longitudinal profiles. In the statistical literature, there have been abundant works on joint models for a primary outcome (discrete or continuous) and longitudinal covariates. A popular approach is to use latent variables that are random parameters characterizing the subject-specific longitudinal covariate profile and accounting for the dependence between the primary outcome and longitudinal covariates [11, 30, 10, 29, 27, 14, 28]. In these approaches, longitudinal covariates are usually assumed to be measured at prescheduled time points, or the measurement times are independent of the longitudinal covariates and the primary outcome. However, in many applications, such assumptions are restrictive. To take into account for irregular and possibly subject-specific observation times, Lin and Ying (2001) [16] proposed to use a counting process and treat observation times as recurrent events. In their approach, observation times are assumed to be independent of longitudinal outcomes given observed covariates, which may still be restrictive in practice. To allow more flexible dependence between observation times and longitudinal outcomes, a multiplicative frailty model [1] (Andersen et al. (1993)) for observation times has attracted much attention recently [26, 15, e.g.]. The dependence between observation times and longitudinal outcomes is explained by the correlation of the frailty and subject-specific random effects in modeling longitudinal outcomes. Moreover, such a modeling strategy has been widely used in the framework of joint modeling of recurrent events and a survival time [19, 12] and joint modeling of longitudinal data with informative observation times and a terminal event [18, 17].

In this chapter, we propose a joint model to study the association between a primary binary outcome and longitudinal covariates that are measured at informative occasions. To the best of our knowledge, this has not been studied in the literature. The correlation among longitudinal covariates, observation times and primary binary outcomes are modeled via latent random variables shared by three submodels: (1) a semiparametric mixed model (SPMM) for longitudinal covariates; (2) a multiplicative frailty model for informative observation times; and (3) a logistic model for primary binary outcomes. Here the SPMM proposed by Zhang and Lin (1998) [34] can naturally handle nonlinear longitudinal trajectories and account for subject heterogeneity in repeated measures. We develop a maximum likelihood estimation method for the parameters using the expectation maximization (EM) algorithm [7], and use an EM-aided numerical differentiation method to compute the variance-covariance matrix of our estimators.

The rest of this chapter is organized as follows. Section 2.2 introduces notations and describes the proposed joint model. Section 2.3 presents the inference procedure using the EM algorithm. Section 2.4 provides simulation studies in which our method is also compared to the regression calibration (RC) method. Section 2.5 applies our method to the NYU-IVF data, followed by discussions in Section 2.6.

2.2 Joint Model

Consider a longitudinal study involving n independent subjects on the time interval $[0, \tau]$. For the ith subject, let Y_i denote the primary binary outcome obtained at the end of the study. Moreover, let X_i be the p-dimensional baseline covariates and $Z_i(\cdot)$ be the univariate longitudinal covariate process measured over the study period. Practically, $Z_i(\cdot)$ are only available at subject-specific observation times. Let $0 < T_{i1} < \cdots < T_{im_i} \leq \tau$ denote the m_i observation times for subject i, and define $Z_{ij} = Z_i(T_{ij})$. Then the observed data consist of $\{Y_i, X_i, T_{ij}, Z_{ij} : i = 1, \cdots, n; j = 1, \cdots, m_i\}$. Moreover, we use a counting process, defined as $N_i(t) = \Sigma_j I(T_{ij} \leq t)$, to denote the number of observations taken on subject i up to time t, where $I(\cdot)$ is the indicator function. The proposed joint model is specified by the following three submodels:

(M1) A semiparametric mixed model (SPMM) for the longitudinal covariate process

$$Z_i(t) = f(t) + \gamma' X_i + u_{i0} + u_{i1} t + \epsilon_i(t), \tag{2.1}$$

where $\mathbf{u}_i \equiv (u_{i0}, u_{i1})'$ are subject-specific random effects, $\epsilon_i(t)$ is independent normal process with mean 0 and variance σ_ϵ^2, γ is a p-dimensional coefficient vector of X_i and $f(\cdot)$ is an unspecified smooth function.

(M2) A multiplicative frailty model for $N_i(\cdot)$, i.e., given X_i and $\mathbf{u}_i$, the intensity function of $N_i(\cdot)$ is given by

$$\lambda(t|X_i, \mathbf{u}_i) = \lambda_0(t)e^{\eta' X_i + \alpha_1 u_{i0} + \alpha_2 u_{i1}}, \qquad 0 \leq t \leq \tau, \tag{2.2}$$

where $\lambda_0(t) \geq 0$ is the unspecified baseline intensity function, η is a p-dimensional coefficient vector of X_i, α_1 and α_2 are effects of u_{i0} and u_{i1} on the intensify function.

(M3) A logistic model for the binary outcome Y_i

$$P[Y_i = 1|X_i, \mathbf{u}_i] = \frac{e^{\beta' H_i + \alpha_3 u_{i0} + \alpha_4 u_{i1}}}{1 + e^{\beta' H_i + \alpha_3 u_{i0} + \alpha_4 u_{i1}}}, \tag{2.3}$$

where $H_i = (1, X_i')'$, β is a $(p+1)$-dimensional coefficient vector, α_3 and α_4 are effects of u_{i0} and u_{i1} on the binary outcome Y.

Note that the above three submodels are correlated through shared subject-specific random effects $\mathbf{u}_i$, and the magnitude of the association is controlled by the parameters α_1, α_2, α_3 and α_4. Here we assume that $\mathbf{u}_i$'s are independent bivariate normal random variables with mean 0 and variance-covariance matrix:

$$D = \begin{pmatrix} \sigma_{00} & \sigma_{01} \\ \sigma_{01} & \sigma_{11} \end{pmatrix}. \tag{2.4}$$

In addition, we assume that given X_i and $\mathbf{u}_i$, $Z_i(\cdot)$, $N_i(\cdot)$ and Y_i are all independent from each other. For computational convenience, we assume that the smooth function $f(\cdot)$

can be represented as a linear combination of cubic B-spline basis functions $\{B_k(\cdot); 1 \leq k \leq K\}$, such that

$$f(t) = \sum_{k=1}^{K} B_k(t)\zeta_k = \zeta' B(t), \tag{2.5}$$

where $B(t) = \{B_1(t), \cdots, B_K(t)\}'$ and $\zeta = (\zeta_1, \cdots, \zeta_K)'$. Such a representation has been widely used in the statistical literature for nonparametric function estimation. In practice, the number of basis functions K may also depend on the sample size, and can be selected using some information criterion, such as AIC or BIC.

Compared with the joint model defined in equations (1.1), (1.2) and (1.3) in Chapter 1. The joint model defined in (2.1), (2.2) and (2.3) is obviously more general: (i) model (2.1) can characterize more general nonlinear longitudinal trend than model (1.1) (the latter only consider quadratic trend); (ii) counting process is more suitable when the observation times of repeated measurements are irregular and unequally spaced.

2.3 Maximum Likelihood Estimation and Inference

Let $\Theta = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta, \gamma, \eta, \zeta, D, \sigma_\epsilon^2, \lambda_0(\cdot)\}$ denote the unknown parameters defined in the proposed joint model, and write $\mathbf{T}_i = (T_{i1}, \cdots, T_{im_i})'$ and $\mathbf{Z}_i$=$(Z_{i1}, ..., Z_{im_i})'$, $i = 1, \cdots, n$. The observed data for subject i is denoted as $\mathbf{O}_i = (\mathbf{T}_i, \mathbf{Z}_i, X_i, Y_i)$ and write $\mathbf{O} = (\mathbf{O}_i; i = 1, \cdots, n)$. Then the joint likelihood from the observed data is given by

$$L(\Theta) = \prod_{i=1}^{n} \left[\int f_{ri} \quad f_{li} \quad f_{pi} \quad f_{\mathbf{u}}(\mathbf{u}_i) d\mathbf{u}_i \right], \tag{2.6}$$

where f_{ri}, f_{li} and f_{pi} are the density functions of the observation times, longitudinal covariates and primary binary outcome for subject i, respectively, and $f_{\mathbf{u}}(\cdot)$ is the density function of the random effects $\mathbf{u}_i$. Moreover, we have

$$\begin{cases} f_{ri} = \left\{\prod_{j=1}^{m_i} \lambda_0(T_{ij}) e^{\eta' X_i + \alpha_1 u_{i0} + \alpha_2 u_{i1}}\right\} \exp\left\{-\Lambda_0(\tau) e^{\eta' X_i + \alpha_1 u_{i0} + \alpha_2 u_{i1}}\right\} \\ f_{li} = (2\pi\sigma_\epsilon^2)^{-\frac{m_i}{2}} \exp\left\{ -\frac{1}{2\sigma_\epsilon^2} (\mathbf{Z}_i - \mathbf{B}_i \zeta - \gamma' X_i - u_{i0} - u_{i1} \mathbf{T}_i)^2 \right\} \\ f_{pi} = \frac{e^{Y_i(\beta' H_i + \alpha_3 u_{i0} + \alpha_4 u_{i1})}}{1 + e^{\beta' H_i + \alpha_3 u_{i0} + \alpha_4 u_{i1}}} \end{cases}, \quad (2.7)$$

where $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$ is the baseline cumulative intensity function, $\mathbf{B}_i = \{B_k(T_{ij})\}_{jk}$, $j = 1, \cdots, m_i$, $k = 1, \cdots, K$, is an $m_i \times K$ matrix, and a^2 is defined to be $a^2 = a'a$ for a column vector a.

The direct maximization of the likelihood function (2.6) is very challenging, mainly due to three reasons: (a) the integration in (2.6) does not have an analytical form; (b) the baseline intensity function $\lambda_0(\cdot)$ is nonparametric; (c) the number of model parameters is large. Various approaches have been proposed in the literature for maximum joint likelihood estimation for similar but less complex problems. For example, [18] and [17] proposed to use numerical integration techniques, such as Gaussian quadrature, to approximate their likelihood functions. They adopted a piecewise constant function for the baseline intensity function. Then the maximization can be done using standard softwares, such as the SAS Proc NLMIXED. Such approaches are easy to use, but they usually work well only for a small number of random effects, say a single shared latent variable, and may loose the flexibility of estimating $\lambda_0(\cdot)$ nonparametrically.

This motivates us to develop a nonparametric maximum likelihood estimation (NPMLE) method via an EM algorithm. A striking property of the EM algorithm is the likelihood will always increase during parameter update process. And often time the parameter update can be separately carried over for different sets of parameters and some parameters may even have closed form updates. Let $t_1 < \cdots < t_d$ denote the total d distinguished observation times across all subjects. As in NPMLE literature for censored survival data [32], the maximizer of the baseline cumulative intensity function $\Lambda_0(\cdot)$ is an increasing step function with jumps only at the observed observation times t_q, $q = 1, \cdots, d$.

2.3.1 The EM Algorithm

The log complete-data likelihood is given by

$$l_C(\Theta) = \sum_{i=1}^{n} [\log(f_{ri}) + \log(f_{li}) + \log(f_{pi}) + \log\{f_{\mathbf{u}}(\mathbf{u}_i)\}]. \tag{2.8}$$

In the kth E-step, we need to calculate the conditional expectation of $l_C(\Theta)$ with respect to the random effects $\mathbf{u}_i$ given the observed data $\mathbf{O}$ and the current parameter estimates $\widehat{\Theta}^{(k)}$, commonly called the Q function and denoted as $Q(\Theta|\widehat{\Theta}^{(k)})$. From equation (A.1), the conditional expectation $Q(\Theta|\widehat{\Theta}^{(k)})$ can be expressed as $Q(\Theta|\widehat{\Theta}^{(k)}) = Q_1(\alpha_1, \alpha_2, \eta, \lambda_0; \mathbf{O}, \widehat{\Theta}^{(k)}) + Q_2(\zeta, \gamma, \sigma_\epsilon^2; \mathbf{O}, \widehat{\Theta}^{(k)}) + Q_3(\beta, \alpha_3, \alpha_4; \mathbf{O}, \widehat{\Theta}^{(k)}) + Q_4(D; \mathbf{O}, \widehat{\Theta}^{(k)})$, where

$$\begin{aligned}
Q_1(\alpha_1, \alpha_2, \eta, \lambda_0; \mathbf{O}, \widehat{\Theta}^{(k)}) &= \sum_{i=1}^{n} E_i\{\log(f_{ri})|\mathbf{O}_i, \widehat{\Theta}^{(k)}\} \\
Q_2(\zeta, \gamma, \sigma_\epsilon^2; \mathbf{O}, \widehat{\Theta}^{(k)}) &= \sum_{i=1}^{n} E_i\{\log(f_{li})|\mathbf{O}_i, \widehat{\Theta}^{(k)}\} \\
Q_3(\beta, \alpha_3, \alpha_4; \mathbf{O}, \widehat{\Theta}^{(k)}) &= \sum_{i=1}^{n} E_i\{\log(f_{pi})|\mathbf{O}_i, \widehat{\Theta}^{(k)}\} \\
Q_4(D; \mathbf{O}, \widehat{\Theta}^{(k)}) &= \sum_{i=1}^{n} E_i[\log\{f_{\mathbf{u}}(\mathbf{u}_i)\}|\mathbf{O}_i, \widehat{\Theta}^{(k)}],
\end{aligned} \tag{2.9}$$

and the expectation $E_i(\cdot|\mathbf{O}_i, \widehat{\Theta}^{(k)})$ is taken with respect to the conditional density function $f(\mathbf{u}_i|\mathbf{O}_i, \widehat{\Theta}^{(k)})$ of $\mathbf{u}_i$ given $\mathbf{O}_i$ evaluated at $\widehat{\Theta}^{(k)}$. By simple calculation, we have

$$f(\mathbf{u}_i|\mathbf{O}_i, \widehat{\Theta}^{(k)}) = \frac{f(\mathbf{O}_i, \mathbf{u}_i|\widehat{\Theta}^{(k)})}{\int f(\mathbf{O}_i, \mathbf{u}_i|\widehat{\Theta}^{(k)}) d\mathbf{u}_i} = \frac{f_{pi}(Y_i|\mathbf{u}_i, \widehat{\Theta}^{(k)}) f_{ri}(\mathbf{T}_i|\mathbf{u}_i, \widehat{\Theta}^{(k)}) g(\mathbf{u}_i|\mathbf{Z}_i, \widehat{\Theta}^{(k)})}{\int f_{pi}(Y_i|\mathbf{u}_i, \widehat{\Theta}^{(k)}) f_{ri}(\mathbf{T}_i|\mathbf{u}_i, \widehat{\Theta}^{(k)}) g(\mathbf{u}_i|\mathbf{Z}_i, \widehat{\Theta}^{(k)}) d\mathbf{u}_i}, \tag{2.10}$$

where $g(\mathbf{u}_i|\mathbf{Z}_i, \widehat{\Theta}^{(k)})$ is the density of a bivariate normal distribution with mean

$$\mu_i^{(k)} = \mathbf{W}_i^{(k)} \mathbf{G}_i' (\mathbf{Z}_i - \mathbf{B}_i \widehat{\zeta}^{(k)} - \widehat{\gamma}^{(k)\prime} X_i) / (\widehat{\sigma}_\epsilon^2)^{(k)},$$

and variance-covariance matrix

$$\mathbf{W}_i^{(k)} = \left\{ (\widehat{D}^{(k)})^{-1} + \frac{\mathbf{G}_i' \mathbf{G}_i}{(\widehat{\sigma}_\epsilon^2)^{(k)}} \right\}^{-1},$$

with $\mathbf{G}_i = (\mathbf{1}, \mathbf{T}_i)$ being an $m_i \times 2$ matrix. This distribution can be viewed as the "working" conditional distribution of $\mathbf{u}_i$ given $\mathbf{Z}_i$ and $\widehat{\Theta}^{(k)}$. Then the integrations in (2.8) and (2.9) can be computed using numerical integration techniques. Here we use the Gaussian quadrature method.

Next, in the M-step, it is clear that the maximization of $Q(\Theta|\widehat{\Theta}^{(k)})$ can be conducted separately for each component in $Q(\Theta|\widehat{\Theta}^{(k)})$. For $Q_1(\alpha_1, \alpha_2, \eta, \lambda_0; \mathbf{O}, \widehat{\Theta}^{(k)})$, the maximizer of $\lambda_0(t_q)$ with fixed α_1, α_2 and η can be obtained by a Breslow-type estimator

$$\widehat{\lambda}_0^{(k+1)}(t_q; \eta, \alpha_1, \alpha_2) = \frac{1}{\sum_{i=1}^n e^{\eta' X_i} E_i\{e^{\alpha_1 u_{i0} + \alpha_2 u_{i1}} |\mathbf{O}_i, \widehat{\Theta}^{(k)}\}}, \qquad q = 1, \cdots, d. \tag{2.11}$$

Plugging the above $\widehat{\lambda}_0^{(k+1)}(\cdot; \eta, \alpha_1, \alpha_2)$ into $Q_1(\alpha_1, \alpha_2, \eta, \lambda_0; \mathbf{O}, \widehat{\Theta}^{(k)})$ yields the profile Q function for η, α_1, α_2. Then the updated estimators $(\widehat{\eta}^{(k+1)}, \widehat{\alpha}_1^{(k+1)}, \widehat{\alpha}_2^{(k+1)})$ at $(k+1)$th step can be obtained by solving the "score function" of this profile Q function, which is given in the Appendix B.1. The updated estimator for $\lambda_0(\cdot)$ is then $\widehat{\lambda}_0^{(k+1)}(\cdot) = \widehat{\lambda}_0^{(k+1)}(\cdot; \widehat{\eta}^{(k+1)}, \widehat{\alpha}_1^{(k+1)}, \widehat{\alpha}_2^{(k+1)})$. For $Q_2(\zeta, \gamma, \sigma_\epsilon^2; \mathbf{O}, \widehat{\Theta}^{(k)})$ and $Q_4(D; \mathbf{O}, \widehat{\Theta}^{(k)})$, it can be shown by simple algebra that the maximizers $\widehat{\gamma}^{(k+1)}$ and $\widehat{\zeta}^{(k+1)}$ are the solutions of the following equations

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i}\begin{pmatrix} X_iX_i' & X_iB_{ij}' \\ B_{ij}X_i' & B_{ij}B_{ij}' \end{pmatrix}\begin{pmatrix} \gamma \\ \zeta \end{pmatrix} = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\begin{pmatrix} X_i\Big(Z_{ij} - u_{i0}^{(k)} - u_{i1}^{(k)}T_{ij}\Big) \\ B_{ij}\Big(Z_{ij} - u_{i0}^{(k)} - u_{i1}^{(k)}T_{ij}\Big) \end{pmatrix}, \tag{2.12}$$

where $B_{ij} = \{B_1(T_{ij}), \cdots, B_K(T_{ij})\}'$, $u_{i0}^{(k)} = E_i(u_{i0}|\mathbf{O}_i, \widehat{\Theta}^{(k)})$ and $u_{i1}^{(k)} = E_i(u_{i1}|\mathbf{O}_i, \widehat{\Theta}^{(k)})$. Moreover, we have

$$\begin{aligned}(\widehat{\sigma}_\epsilon^2)^{(k+1)} \quad = \quad & \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\big\{Z_{ij} - (\widehat{\zeta}^{(k+1)})'B_{ij} - (\widehat{\gamma}^{(k+1)})'X_i\big\}^2}{\sum_{i=1}^{n} m_i} \\ & - \frac{2\sum_{i=1}^{n}\sum_{j=1}^{m_i}\big\{Z_{ij} - (\widehat{\zeta}^{(k+1)})'B_{ij} - (\widehat{\gamma}^{(k+1)})'X_i\big\}\big(u_{i0}^{(k)} + u_{i1}^{(k)}T_{ij}\big)}{\sum_{i=1}^{n} m_i} \\ & + \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i} E_i\{\big(u_{i0} + u_{i1}t_{ij}\big)^2|\mathbf{O}_i, \widehat{\Theta}^{(k)}\}}{\sum_{i=1}^{n} m_i},\end{aligned} \tag{2.13}$$

and

$$\widehat{D}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} E_i(\mathbf{u}_i\mathbf{u}_i'|\mathbf{O}_i, \widehat{\Theta}^{(k)}). \tag{2.14}$$

Finally, the updated estimators $\widehat{\beta}^{(k+1)}$, $\widehat{\alpha}_3^{(k+1)}$ and $\widehat{\alpha}_4^{(k+1)}$ can be obtained by the Newton-Raphson (or one-step Newton-Raphson) method based on the "score function" and "information matrix" of $Q_3(\beta, \alpha_3, \alpha_4; \mathbf{O}, \widehat{\Theta}^{(k)})$. We iterate the E-step and M-step until a pre-specified convergence criterion is met. Let $\widehat{\Theta}$ denote the final estimator at convergence.

2.3.2 Variance Estimation

For parametric models, Louis's formula [21] can be used to compute the variance estimates of MLEs obtained by the EM algorithm. However, for our problem, this approach would be intimidating due to the presence of the infinitely dimensional parameter $\lambda_0(\cdot)$, which is estimated nonparametrically. Here, instead, we compute the variance estimates of the regression parameter estimators by inverting the observed information matrix based on the corresponding profile log likelihood function. In general, the observed information matrix does not have a closed form. Thus, we compute it using an EM-aided differentiation method, which was first studied by Meilijson (1989) [22] for parametric models and then extended to the proportional hazards model with missing covariates by [3]. Similar method was also used by Liu and Lu (2006) [20] for variance estimation in their joint modeling approach to the interval mapping of quantitative trait loci for time-to-event data. Specifically, define $\theta = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta, \gamma, \zeta, \eta)$, the regression parameters of our main interest, and ρ the remain nuisance parameters in Θ. Let $\widehat{\theta}$ denote the MLE of θ obtained by the proposed EM algorithm. Moreover, let $\widehat{\rho}(\theta) = \{\widehat{D}(\theta), \widehat{\sigma}^2_\epsilon(\theta), \widehat{\lambda}_0(\cdot;\theta)\}$ denote the maximizer of the observed likelihood function over ρ with θ fixed, which can be obtained by a similar EM algorithm as proposed. Then we calculate the information matrix for θ as follows:

i Perturb the jth component of $\widehat{\theta}$ by a small amount h from both sides, i.e. $\widehat{\theta}_j + h$ and $\widehat{\theta}_j - h$, while keeping other components of $\widehat{\theta}$ unchanged; denote the resulting parameter vectors by $\widehat{\theta}^{j+}$ and $\widehat{\theta}^{j-}$ respectively, and run the EM algorithm to obtain $\widehat{\rho}(\widehat{\theta}^{j+})$ and $\widehat{\rho}(\widehat{\theta}^{j-})$ accordingly.

ii Compute the jth row of the information matrix of θ by

$$\frac{1}{2h}\left[E\left\{l_C^\theta(\widehat{\theta}^{j+},\rho(\widehat{\theta}^{j+}))\Big|\mathbf{O},\widehat{\theta}^{j+},\widehat{\rho}(\widehat{\theta}^{j+})\right\}-E\left\{l_C^\theta(\widehat{\theta}^{j-},\rho(\widehat{\theta}^{j-}))\Big|\mathbf{O},\widehat{\theta}^{j-},\widehat{\rho}(\widehat{\theta}^{j-})\right\}\right]$$

where $l_C^\theta(\widehat{\theta}^{j+},\rho(\widehat{\theta}^{j+}))$ and $l_C^\theta(\widehat{\theta}^{j-},\rho(\widehat{\theta}^{j-}))$ denote the derivatives of the complete-data log likelihood $l_C(\theta,\rho)$ with respect to θ evaluated at $(\widehat{\theta}^{j+},\rho(\widehat{\theta}^{j+}))$ and $(\widehat{\theta}^{j-},\rho(\widehat{\theta}^{j-}))$ respectively.

The derivation of $l_C^\theta(\theta,\rho)$ is straightforward and hence omitted here. For the perturbation size h, Chen (1999) [3] suggested to use $h=c/n$, where c is a positive constant. We found that $c=0.15$ would produce satisfactory results in all of our numerical studies.

2.4 Simulation Studies

Simulation was conducted to evaluate the performance of the proposed method under practical settings. We consider the proposed joint model with two covariates: a binary covariate X_{i1} sampled from the Bernoulli distribution with a success probability of 0.5 and a continuous covariate X_{i2} sampled from the standard normal distribution. The regression parameters are chosen as $\eta'=(\eta_1,\eta_2)=(0.5,-0.5), \gamma'=(\gamma_1,\gamma_2)=(-0.5,0.5), \beta'=(\beta_1,\beta_2,\beta_3)=(0.8,-1,-1.5)$ and $(\alpha_1,\alpha_2,\alpha_3,\alpha_4)=(-0.5,0.5,-1,1)$. The error terms ϵ_{ij}'s are generated independently from a standard normal distribution. For the covariance matrix D of the random effects $\mathbf{u}_i$, we set $\sigma_{00}=1$, $\sigma_{01}=0.5$, and $\sigma_{11}=1$. The nonparametric component $f(t)$ is assumed to take the form of $\sin(2\pi t/\tau), t\in[0,\tau]$, with $\tau=6$. We choose $\lambda_0(t)=at$, with $a=0.11$ and 0.215, corresponding to the averages of $\bar{m}=2$ and 4 observation times per subject during the study period, respectively. We conducted 100 runs of simulation with sample size of $n=300$ for each setting.

The proposed EM algorithm requires specifying initial values of the parameters. Here we use the naive regression calibration (RC) method to produce those initial values. To be more specific, we first fit the semiparametric mixed model (2.1) with the cubic B-splines representation for $f(\cdot)$, ignoring the informative observation times, to obtain the initial estimators $(\widehat{\gamma}^{(0)}, \widehat{\zeta}^{(0)}, \widehat{D}^{(0)}, (\widehat{\sigma}_\epsilon^2)^{(0)})$ and the best linear unbiased predictors (BLUP) $\widehat{u}_{i0}^{(0)}$ and $\widehat{u}_{i1}^{(0)}$, $i = 1, \cdots, n$. This can be easily done using standard software packages, e.g. PROC MIXED in SAS or lme in R. Then we fit a Cox proportional hazards model for T_{ij}'s with covariates X_{i1}, X_{i2}, $\widehat{u}_{i0}^{(0)}$ and $\widehat{u}_{i1}^{(0)}$ to obtain the initial estimators $(\widehat{\eta}^{(0)}, \widehat{\alpha}_1^{(0)}, \widehat{\alpha}_2^{(0)}, \widehat{\lambda}_0^{(0)})$, and fit a logistic model for Y_i's to obtain $(\widehat{\beta}^{(0)}, \widehat{\alpha}_3^{(0)}, \widehat{\alpha}_4^{(0)})$, which can also be implemented directly by standard software packages.

For the convergence criterion of the proposed EM algorithm, we consider $\max|\widehat{\Theta}^{(k+1)} - \widehat{\Theta}^{(k)}| < \delta$ with $\delta = 0.0001$ in our numerical studies. We use the Gaussian quadrature method with 15 quadrature knots to approximate the two dimensional integrations used in the E-step. For the cubic B-splines representation of the smooth function $f(\cdot)$, we choose three equally spaced internal knots, giving 7 basis functions in total. In the EM-aided numerical differentiation procedure for variance estimation, it might not be enough to stop EM algorithm only at point $\max|\widehat{\theta}^{k+1} - \widehat{\theta}^{k}| < \epsilon$, more EM steps might be needed until the scores of conditional expectation of complete-data log-likelihood become stable. We choose the perturbation value of $h = 0.15/n = 0.0005$.

Tables 2.1 and 2.2 summarize the bias of the estimators of regression parameters, the sample standard deviation (SD) of the estimators, the average of the standard errors (SE) obtained using the proposed EM-aided numerical differentiation method, and the empirical coverage probabilities (CP) of 95% Wald-type confidence intervals, for $\bar{m} = 2$ and 4, respectively. For comparison purpose, we also include the results for the initial estimators obtained using the RC method.

Table 2.1: Simulation results for setting 1: $\bar{m} = 2$, for comparing the performance between the proposed method and regression calibration (RC), based on 100 simulation runs with sample size of $n = 300$

		Proposed Method				RC			
Parameter	True	Bias	SD	SE	CP(%)	Bias	SD	SE	CP(%)
recurrent event model									
η_1	0.5	0.004	0.050	0.051	93 %				
η_2	-0.5	0.014	0.097	0.101	95 %				
α_1	-0.5	0.002	0.080	0.086	96 %				
α_2	0.5	-0.010	0.078	0.076	93 %				
longitudinal model									
γ_1	0.5	0.008	0.075	0.078	95 %	0.019	0.075	0.078	95 %
γ_2	-0.5	0.007	0.157	0.155	94 %	-0.003	0.154	0.156	97 %
binary outcome model									
β_0	0.8	0.030	0.228	0.235	96 %	-0.070	0.193	0.195	93 %
β_1	-1	-0.055	0.200	0.195	94 %	0.089	0.165	0.155	87 %
β_2	-1.5	-0.043	0.369	0.342	97 %	0.140	0.305	0.277	90 %
α_3	-1	-0.048	0.357	0.349	94 %	0.064	0.315	0.276	88 %
α_4	1	0.060	0.303	0.317	97 %	-0.066	0.285	0.259	93 %

SD is the sample standard deviation; SE is the mean of estimated standard errors; CP is the empirical coverage probability of 95% Wald-type confidence intervals.

Table 2.2: Simulation results for setting 2: $\bar{m} = 4$, for comparing the performance between the proposed method and regression calibration (RC), based on 100 simulation runs with sample size of $n = 300$

		Proposed Method				RC			
Parameter	True	Bias	SD	SE	CP(%)	Bias	SD	SE	CP(%)
recurrent event model									
η_1	0.5	0.007	0.044	0.040	92 %				
η_2	-0.5	0.019	0.078	0.080	95 %				
α_1	-0.5	-0.002	0.063	0.064	97 %				
α_2	0.5	0.004	0.052	0.058	95 %				
longitudinal model									
γ_1	0.5	0.002	0.072	0.074	96 %	0.017	0.074	0.074	95%
γ_2	-0.5	-0.004	0.146	0.148	95 %	-0.019	0.147	0.149	95%
binary outcome model									
β_0	0.8	0.043	0.232	0.231	96 %	-0.042	0.206	0.197	93%
β_1	-1	-0.040	0.212	0.189	93 %	0.084	0.176	0.157	87%
β_2	-1.5	-0.055	0.317	0.333	96 %	0.100	0.273	0.279	91%
α_3	-1	-0.065	0.366	0.308	95 %	0.003	0.367	0.277	88%
α_4	1	0.045	0.272	0.276	97 %	-0.030	0.272	0.245	92%

SD is the sample standard deviation; SE is the mean of estimated standard errors; CP is the empirical coverage probability of 95% Wald-type confidence intervals.

Based on the results, we observe that our proposed estimators have negligible biases for all the parameters, the SE's are close to the SD's, and all the CP's are close to the nominal level; while for the RC estimators, some show severe biases and the corresponding CP's are usually lower than 95% with some even below 90%. To study the performance of our method for estimating the nonparametric function $f(\cdot)$, in Figure 2.1, we present the average estimated nonparametric functions over 100 simulation runs, for both $\bar{m} = 2$ and $\bar{m} = 4$. It can be seen that our method works very well for estimating the nonparametric function $f(\cdot)$. In contrast, the estimated $\hat{f}(\cdot)$ by RC method shows some obvious bias.

2.5 Application

The NYU-IVF data consist of 540 pregnancies obtained after IVF treatment at the New York University Fertility Center from 2001 to 2003. In this study, viable pregnancies are defined as pregnancies reaching the second trimester, including singleton, twins, higher order multiples and stillbirths; non-viable pregnancies include biochemical pregnancies, ectopic pregnancies and first trimester abortions [2]. The average age of participants was 35.27 years (SD= 4.40 years) and the average body mass index (BMI) was 23.71 (SD= 5.05).

We apply the proposed approach to the NYU-IVF data. For subject i, let $Y_i = 1/0$ denote the final pregnancy outcome (1 = viable pregnancy; 0 = nonviable pregnancy) and let Z_{ij} stand for the logarithm of the β-HCG values observed at the observation times T_{ij}, $j = 1, \cdots, m_i$. Note that the maximum value of m_i is 6. Besides the primary binary outcome and the longitudinal covariate, the data also include two baseline covariates, age and BMI. In our analysis, we transform age and BMI into (0,1) range by their maximal values, and denote them by X_{i1} and X_{i2}, respectively. For comparison, we also fit the

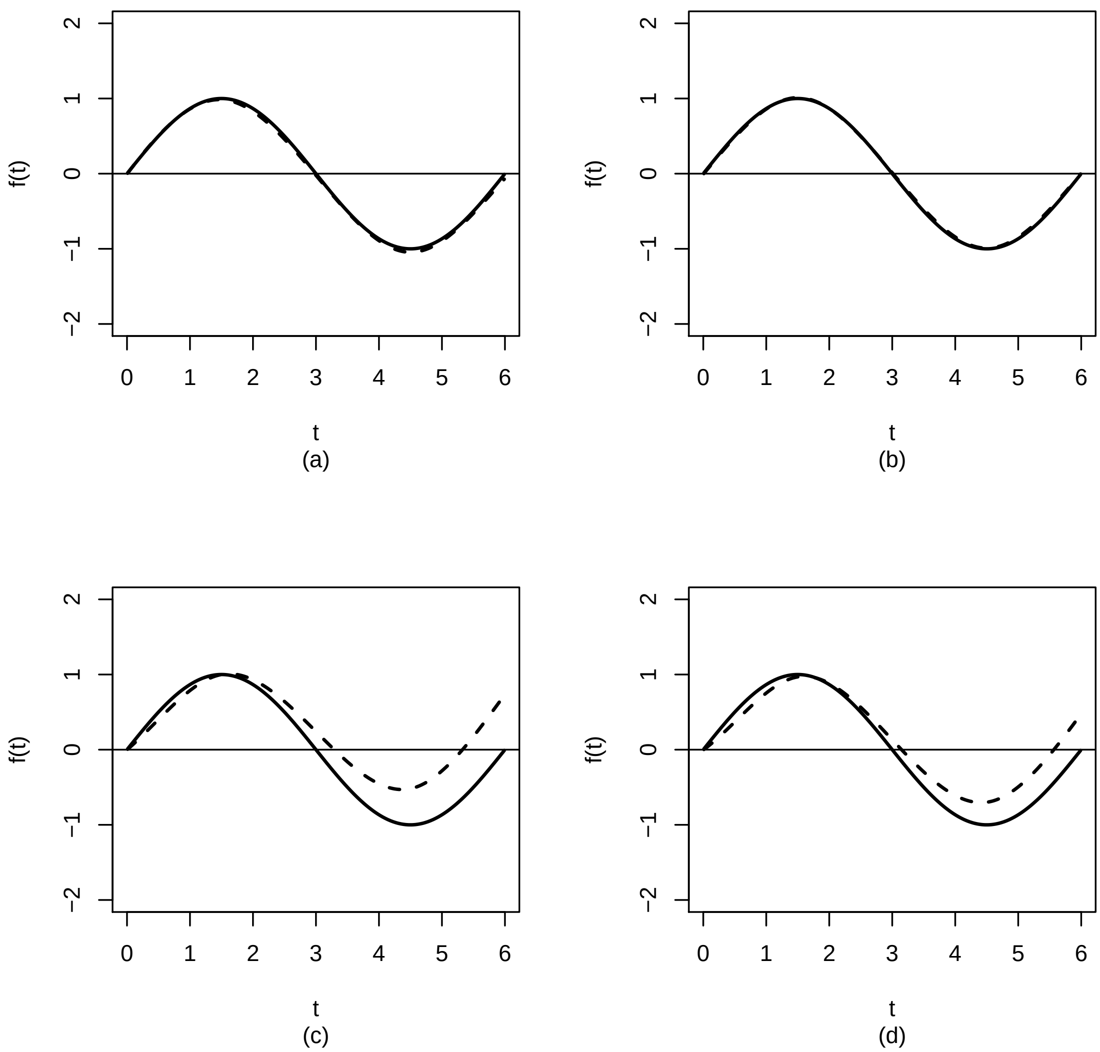

Figure 2.1: Estimation results for the nonparametric function $f(\cdot)$: (a) estimated by full joint model and $\bar{m} = 2$; (b) estimated by full joint model and $\bar{m} = 4$; (c) estimated by RC method and $\bar{m} = 2$; (d) estimated by RC method and $\bar{m} = 4$. The solid lines are the true function, the dashed lines are the average estimated curves using our method based on 100 simulation runs.

NYU-IVF data by the RC method ignoring the possibly informative measurement mechanism of the β-HCG profiles. The estimation results for both methods are summarized in Table 2.3.

We find that both methods indicate that greater baseline values and stronger increasing trends of the β-HCG profiles are associated with higher chances of viable pregnancy (for example, for our method, the estimators $\widehat{\alpha}_3 = 1.521$ and $\widehat{\alpha}_4 = 3.423$, with the p-values of both < 0.0001), and the baseline age is significantly negatively associated with the probability of viable pregnancy. Moreover, the results also show that the subject-specific random intercepts and slopes of the β-HCG profiles have significant and negative associations with the β-HCG observation frequencies (for our method, the estimators $\widehat{\alpha}_1 = -0.202$ and $\widehat{\alpha}_2 = -0.260$, with the p-values of both < 0.0001), which is consistent with our empirical observations shown in Figure 1.1. We also give our estimation for the smooth function $f(\cdot)$. The resulting fitted smooth curve of $f(\cdot)$ along with the 95% pointwise confidence intervals are shown in Figure 2.2. The estimator of $f(\cdot)$ shows certain nonlinear trend for the baseline β-HCG profiles.

2.6 Discussion

In this article, we propose a joint model that can naturally study the association of a primary binary outcome and longitudinal covariates that are possibly measured at informative observation times. Our simulation studies show that the EM algorithm provides stable estimates for the model parameters, and that the EM-aided differentiation approach yields reliable variance estimates for the parameter estimators of interest. When applied to the IVF study, the proposed approach provides a great insight into the relationship between early β-HCG levels and the final pregnancy outcomes after IVF treatment.

Table 2.3: Results on regression parameters in the three sub-models from our method and regression calibration (RC) for NYU-IVF data

	Our Method		RC	
Parameter	Est	SE	Est	SE
recurrent event model				
η_1	0.199	0.296		
η_2	0.178	0.295		
α_1	-0.202	0.038		
α_2	-0.260	0.040		
longitudinal model				
γ_1	-0.391	0.402	-0.399	0.401
γ_2	-1.489	0.400	-1.498	0.399
binary outcome model				
β_0	6.364	1.510	5.981	1.323
β_1	-4.991	1.667	-4.728	1.459
β_2	-2.311	1.475	-2.140	1.259
α_3	1.521	0.242	1.433	0.235
α_4	3.423	0.431	3.182	0.349

Est, the estimates of parameters;
SE, the estimated standard errors.

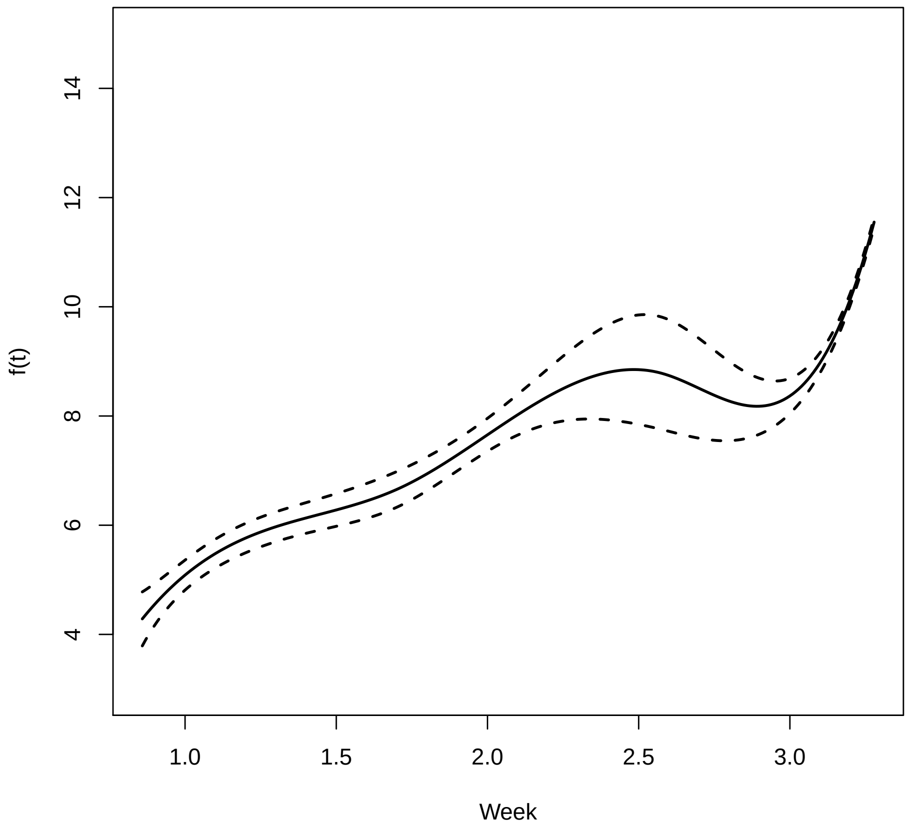

Figure 2.2: Estimated curve of $f(\cdot)$ (solid line) and its 95% pointwise of confidence intervals (dashed lines) for the logarithm of β-HCG.

Our results demonstrate the predictive power of using the latent characteristics of early stage β-HCG profiles for the prediction of viable pregnancies achieved by IVF treatment.

We considered a binary primary outcome in this paper. The joint model can be easily extended to other kinds of primary outcomes, such as outcomes usually fit by a generalized linear model. Then the EM algorithm and the EM-aided differentiation approach can be easily modified for the parameter and variance estimation for the model parameters.

The large sample properties of the proposed estimators, including the consistency, asymptotic normality and efficiency can be studied using the nonparametric maximum likelihood estimation theory as in [32]. This will be investigated in our future work.

REFERENCES

[1] P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. New York: Springer-Verlag, 1993.

[2] S. Bjercke, T. Tanbo, P. O. Dale, L. Morkrid, and T. Abyholm. Human chorionic gonadotrophin concentrations in early pregnancy after in-vitro fertilization. *Human Reproduction*, 14:1642–1646, 1999.

[3] H. Y. Chen and R. J. A. Little. Proportional hazards regression with missing covariates. *Journal of the American Statistical Association*, 94:896–908, 1999.

[4] J. Chen, D. Zhang, and M. Davidian. A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution. *Biostatistics*, 2001:1–27, 2001.

[5] K. Chung, M. D. Sammel, C. Coutifaris, R. Chalian, K. Lin, A. J. Castelbaum, M. F. Freedman, and K. T. Barnhart. Defining the rise of serum hCG in viable pregnancies achieved through use of IVF. *Human Reproduction*, 21:823–828, 2006.

[6] E. Confino, R. H. Demir, J. Friberg, and N. Gleicher. The predictive value of hCG β subunit levels in pregnancies achieved by in vitro fertilization and embryo transfer: an international collaborative study. *Fertility and Sterility*, 45:526–531, 1986.

[7] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–38, 1977.

[8] A. R. Gallant and D. W. Nychka. Seminonparametric maximum likelihood estimation. *Econometrica*, 55:363–390, 1987.

[9] I. Z. Glatstein, M. D. Hornstein, M. J. Kahana, K. V. Jackson, and A. J. Friedman. The predictive value of discriminatory human chorionic gonadotropin levels in the diagnosis of implantation outcome in vitro fertilization cycles. *Fertility and Sterility*, 63:350–356, 1995.

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

[11] J. Hogan and N. Laird. Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine*, 16:259–272, 1997.

[12] X. Huang and L. Liu. A joint frailty model for survival time and gap times between recurrent events. *Biometrics*, 63:389–397, 2007.

[13] M. J. Lambers, H. G. I. V. Weering, M. D. V. Grunewold, C. B. Lambalk, R. Homburg, R. Schats, and P. G. A. Hopes. Optimizing hCG cut-off values: A single determination on day 14 or 15 is sufficient for a reliable prediction of pregnancy outcome. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 127:94–98, 2006.

[14] E. Li, D. Zhang, and M. Davidian. Conditional estimation for generalized linear models when covariates are subject-specific parameters in a mixed model for longitudinal measurements. *Biometrics*, 60:1–7, 2004.

[15] Y. Liang, W. Lu, and Z. Ying. Joint Modeling and Analysis of Longitudinal Data with Informative Observation Times. *Biometrics*, 65:377–384, 2009.

[16] D. Y. Lin and Z. Ying. Semiparametric and nonparametric regression analysis of longitudinal data(with discussion). *Journal of the American Statistical Association*, 96:103–126, 2001.

[17] L. Liu and X. Huang. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society, Series C*, 58:65–81, 2009.

[18] L. Liu, X. Huang, and J. O'Quigley. Analysis of longitudinal dta in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics*, 64:950–958, 2008.

[19] L. Liu, R. A. Wolfe, and X. Huang. Shared frailty models for recurrent events and a terminal event. *Biometrics*, 60:747–756, 2004.

[20] M. Liu, W. Lu, and Y. Shao. Interval Mapping of Quantitative Trait Loci for Time-to-Event Data with the Proportional Hazards Mixture Cure Model. *Biometrics*, 62:1053–1061, 2006.

[21] T. A. Louis. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 44:226–233, 1982.

[22] E. Meilijson. A fast improvement to the EM algorithm on its own terms. *Journal of the Royal Statistical Society, Series B*, 51:127–138, 1989.

[23] M. Shamonki, J. L. Frattarelli, P. A. Bergh, and R. T. Scott. Logarithmic curves depicting initial level and rise of serum beta human chorionic gonadotropin and live delivery outcomes with in vetro fertilization: an analysis of 6021 pregnancies. *Fertility and Sterility*, 91:1760–1764, 2009.

[24] X. Song, M. Davidian, and A. A. Tsiatis. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, 58:742–753, 2002.

[25] A. Strandell, J. Thorburn, and L. Hamberger. Risk factors for ectopic pregnancy in assisted reproduction. *Fertility and Sterility*, 71:282–286, 1999.

[26] J. Sun, L. Sun, and D. Liu. Regression analysis of longitudinal data in the presence of informative observation and censoring times. *Journal of the American Statistical Association*, 102:1397–1406, 2007.

[27] A. A. Tsiatis and M. Davidian. A semiparametric estimator for the proportional hazards model with l ongitudinal covariates measured with error. *Biometrika*, 88:447–458, 2001.

[28] E. F. Vonesh, T. Greene, and M. D. Schluchter. Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in Medicine*, 25:143–163, 2006.

[29] C. Y. Wang, N. Wang, and S. Wang. Regression analysis when covariates are regression parameters of a random-effects model for observed longitudinal measurements. *Biometrics*, 56:487–495, 2000.

[30] M. S. Wulfsohn and A. A. Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, 53:330–339, 1997.

[31] D. Zeng and J. Cai. Simultaneous modeling of survival and longitudinal data with an application to repeated quality of life measures. *Lifetime Data Analysis*, 11:151–174, 2005.

[32] D. Zeng and D. Y. Lin. Maximum likelihood estimation in semiparametric regression models with censored data (with discussion). *Journal of the Royal Statistical Society, Series B*, 69:507–564, 2007.

[33] D. Zhang and M. Davidian. Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*, 57:795–802, 2001.

[34] D. Zhang, X. Lin, J. Raz, and M. Sowers. Semiparametric stochastic mixed models for longitudinal data. *Journal of The American Statistical Association*, 93:710–719, 1998.

APPENDICES

Appendix A

Appendix of Chapter 1

A.1 The details of EM algorithm for Our Joint Model

The log complete-data likelihood is given by

$$l_C(\Theta) = \sum_{i=1}^{n} [\log\{f_L(\mathbf{Z}_i^*|\mathbf{u}_i)\} + \log\{f_P(Y_i|\mathbf{u}_i)\} + \log\{f_R(\mathbf{R}_i|\mathbf{u}_i)\} + \log\{f_{\mathbf{u}}(\mathbf{u}_i)\}]. \quad \text{(A.1)}$$

In the kth E-step, we need to calculate the conditional expectation of $l_C(\Theta)$ with respect to the random effects $\mathbf{u}_i$ given the observed data $\mathbf{O}$ and the current parameter estimates $\widehat{\Theta}^{(k)}$, commonly called the Q function and denoted as $Q(\Theta|\widehat{\Theta}^{(k)})$. From equation (A.1), the conditional expectation $Q(\Theta|\widehat{\Theta}^{(k)})$ can be expressed as

$$\begin{aligned} Q(\Theta|\widehat{\Theta}^{(k)}) &= Q_1(\gamma, \sigma_\epsilon^2, \rho; \mathbf{O}, \widehat{\Theta}^{(k)}) + Q_2(\beta, \alpha_1, \alpha_2; \mathbf{O}, \widehat{\Theta}^{(k)}) \\ &+ Q_3(\lambda_1, \cdots, \lambda_m; \mathbf{O}, \widehat{\Theta}^{(k)}) + Q_4(\mu, \xi, \phi; \mathbf{O}, \widehat{\Theta}^{(k)}) \end{aligned} \quad \text{(A.2)}$$

where

$$
\begin{aligned}
Q_1(\gamma, \sigma_\epsilon^2, \rho; \mathbf{O}, \widehat{\Theta}^{(k)}) &= \sum_{i=1}^{n} E[\log\{f_L(\mathbf{Z}_i^*|\mathbf{u}_i)\}|\mathbf{O}_i, \widehat{\Theta}^{(k)}] \\
Q_2(\beta, \alpha_1, \alpha_2; \mathbf{O}, \widehat{\Theta}^{(k)}) &= \sum_{i=1}^{n} E[\log\{f_P(Y_i|\mathbf{u}_i)\}|\mathbf{O}_i, \widehat{\Theta}^{(k)}] \\
Q_3(\lambda_1, \cdots, \lambda_m; \mathbf{O}, \widehat{\Theta}^{(k)}) &= \sum_{i=1}^{n} E[\log\{f_R(\mathbf{R}_i|\mathbf{u}_i)\}|\mathbf{O}_i, \widehat{\Theta}^{(k)}] \\
Q_4(\mu, \xi, \phi; \mathbf{O}, \widehat{\Theta}^{(k)}) &= \sum_{i=1}^{n} E[\log\{f_{\mathbf{u}}(\mathbf{u}_i)\}|\mathbf{O}_i, \widehat{\Theta}^{(k)}], \qquad \text{(A.3)}
\end{aligned}
$$

and the expectation $E(\cdot|\mathbf{O}_i, \widehat{\Theta}^{(k)})$ is taken with respect to the conditional density function $f(\mathbf{u}_i|\mathbf{O}_i, \widehat{\Theta}^{(k)})$ of $\mathbf{u}_i$ given $\mathbf{O}_i$ evaluated at $\widehat{\Theta}^{(k)}$. By simple calculation, we have

$$
\begin{aligned}
f(\mathbf{u}_i|\mathbf{O}_i, \widehat{\Theta}^{(k)}) &= \frac{f(\mathbf{O}_i, \mathbf{u}_i|\widehat{\Theta}^{(k)})}{\int f(\mathbf{O}_i, \mathbf{u}_i|\widehat{\Theta}^{(k)})d\mathbf{u}_i} \\
&= \frac{f_P(Y_i|\mathbf{u}_i, \widehat{\Theta}^{(k)})f_R(\mathbf{R}_i|\mathbf{u}_i, \widehat{\Theta}^{(k)})P_K^2\{D^{-1}(\mathbf{u}_i - \mu)\}g(\mathbf{u}_i|\mathbf{Z}_i^*, \widehat{\Theta}^{(k)})}{\int f_P(Y_i|\mathbf{u}_i, \widehat{\Theta}^{(k)})f_R(\mathbf{R}_i|\mathbf{u}_i, \widehat{\Theta}^{(k)})P_K^2\{D^{-1}(\mathbf{u}_i - \mu)\}g(\mathbf{u}_i|\mathbf{Z}_i^*, \widehat{\Theta}^{(k)})d\mathbf{u}_i}, \qquad \text{(A.4)}
\end{aligned}
$$

where $g(\mathbf{u}_i|\mathbf{Z}_i^*, \widehat{\Theta}^{(k)})$ is the density of a bivariate normal distribution with mean

$$\mu_i^{(k)} = \mathbf{W}_i\Big\{G_i'\Sigma_i^{-1}(\mathbf{Z}_i^* - \gamma_2'X_i) + (DD')^{-1}\mu\Big\}$$

and variance-covariance matrix

$$\mathbf{W}_i^{(k)} = \Big\{(DD')^{-1} + \mathbf{G}_i'\Sigma_i^{-1}\mathbf{G}_i\Big\}^{-1},$$

with $\mathbf{G}_i = (\mathbf{1}, \mathbf{t}_i^*)$ being an $m_i \times 2$ matrix. This distribution can be viewed as the "working" conditional distribution of $\mathbf{u}_i$ given $\mathbf{Z}_i^*$ and $\widehat{\Theta}^{(k)}$. Then the integrations in (A.3) and (A.4) can be computed using numerical integration techniques. Here we use the Gaussian quadrature method.

Next, in the M-step, it is clear that the maximization of $Q(\Theta|\widehat{\Theta}^{(k)})$ can be conducted separately for each component in $Q(\Theta|\widehat{\Theta}^{(k)})$.

Specifically, $Q_1(\gamma, \sigma_\epsilon^2, \rho; \mathbf{O}, \widehat{\Theta}^{(k)})$, the conditional expectation of log-likelihood of $(\gamma, \sigma_\epsilon^2, \rho)$, takes the following form:

$$\begin{aligned}&-\frac{1}{2}\sum_{i=1}^{n}\log|\Sigma_i| - \frac{1}{2}\sum_{i=1}^{n} E\big\{(\mathbf{Z}_i^* - \gamma_0 - \gamma_1\mathbf{t}_i^* - \gamma_2' X_i - u_{i0} - u_{i1}\mathbf{t}_i^*)'\Sigma_i^{-1} \\ &\times(\mathbf{Z}_i^* - \gamma_0 - \gamma_1\mathbf{t}_i^* - \gamma_2' X_i - u_{i0} - u_{i1}\mathbf{t}_i^*)|\mathbf{O}_i, \widehat{\Theta}^{(k)}\big\}\end{aligned} \tag{A.5}$$

where $\Sigma_i = \sigma_\epsilon^2\Gamma_i(\rho)$. it can be shown by simple algebra that

$$\widehat{\gamma}^{(k+1)} = \Big(\sum_{i=1}^{n} X_i'(\widehat{\Sigma}_i^{(k)})^{-1}Xi\Big)^{-1}\sum_{i=1}^{n} X_i'(\widehat{\Sigma}_i^{(k)})^{-1}\big(\mathbf{Z}_i^* - u_{i0}^{(k)} - u_{i1}^{(k)}\mathbf{t}_i^*\big) \tag{A.6}$$

where $(\widehat{\Sigma}_i^{(k)})^{-1} = (\hat{\sigma}_\epsilon^2)^{(k)}\Gamma_i(\hat{\rho}^{(k)})$, $u_{i0}^{(k)} = E(u_{i0}|\mathbf{O}_i, \widehat{\Theta}^{(k)})$ and $u_{i1}^{(k)} = E(u_{i1}|\mathbf{O}_i, \widehat{\Theta}^{(k)})$.

The updated estimators $(\widehat{\sigma}_\epsilon^{2(k+1)}, \widehat{\rho}^{(k+1)})$ at $(k+1)$th step can be obtained by solving the "score equation" of $Q_1(\gamma, \sigma_\epsilon^2, \rho; \mathbf{O}, \widehat{\Theta}^{(k)})$ with respect to σ_ϵ^2 and ρ. For example, if subject i has $m_i = 4$ observations. Let $d_{jk} = |t_{ij} - t_{ik}|$ be the length of time between times t_{ij} and t_{ik} for all $j, k = 1, \cdots, m_i$, then

$$\Gamma_i(\rho) = \begin{pmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} \\ \rho^{d_{12}} & 1 & \rho^{d_{23}} & \rho^{d_{24}} \\ \rho^{d_{13}} & \rho^{d_{23}} & 1 & \rho^{d_{34}} \\ \rho^{d_{14}} & \rho^{d_{24}} & \rho^{d_{34}} & 1 \end{pmatrix}.$$

The derivative of $Q_1(\gamma, \sigma_\epsilon^2, \rho; \mathbf{O}, \hat{\Theta}^{(k)})$ with respect to σ_ϵ^2 and ρ have the following expressions:

$$\begin{aligned} S_{(k)}(\sigma_\epsilon^2) &= -\frac{1}{2}\sum_{i=1}^{n} \mathrm{tr}\big\{\Sigma_i^{-1}\Gamma_i(\rho)\big\} + \frac{1}{2}\sum_{i=1}^{n} E\big\{(\mathbf{Z}_i^* - \gamma_0 - \gamma_1\mathbf{t}_i^* - \gamma_2' X_i - u_{i0} - u_{i1}\mathbf{t}_i^*)' \\ &\quad \times \Sigma_i^{-1}\Gamma_i(\rho)\Sigma_i^{-1}(\mathbf{Z}_i^* - \gamma_0 - \gamma_1\mathbf{t}_i^* - \gamma_2' X_i - u_{i0} - u_{i1}\mathbf{t}_i^*)|\mathbf{O}_i, \widehat{\Theta}^{(k)}\big\} \end{aligned} \tag{A.7}$$

$$\begin{aligned} S_{(k)}(\rho) &= -\frac{1}{2}\sum_{i=1}^{n} \mathrm{tr}\Big\{\Sigma_i^{-1}\sigma_\epsilon^2\frac{\partial\Gamma_i(\rho)}{\partial\rho}\Big\} + \frac{1}{2}\sum_{i=1}^{n} E\big\{(\mathbf{Z}_i^* - \gamma_0 - \gamma_1\mathbf{t}_i^* - \gamma_2' X_i - u_{i0} - u_{i1}\mathbf{t}_i^*)' \\ &\quad \times \Sigma_i^{-1}\sigma_\epsilon^2\frac{\partial\Gamma_i(\rho)}{\partial\rho}\Sigma_i^{-1}(\mathbf{Z}_i^* - \gamma_0 - \gamma_1\mathbf{t}_i^* - \gamma_2' X_i - u_{i0} - u_{i1}\mathbf{t}_i^*)|\mathbf{O}_i, \widehat{\Theta}^{(k)}\big\} \end{aligned} \tag{A.8}$$

Let $S_{(k)}(\sigma_\epsilon^2, \rho) = (S_{(k)}(\sigma_\epsilon^2), S_{(k)}(\rho))'$. We then solve the following equation

$$S_{(k)}(\sigma_\epsilon^2, \rho) = 0 \tag{A.9}$$

to obtain the update for $\psi = (\sigma_\epsilon^2, \rho)$. However, the "score equation" (A.9) does not have a closed form solution. We then propose to update $\psi = (\sigma_\epsilon^2, \rho)$ by one-step Fisher scoring algorithm

$$\widehat{\psi}^{(k+1)} = \widehat{\psi}^{(k)} + I_{(k)}^{-1}(\widehat{\psi}^{(k)})S_{(k)}(\widehat{\psi}^{(k)}), \tag{A.10}$$

where $S_{(k)}(\widehat{\psi}^{(k)}) = S_{(k)}(\widehat{\sigma}_\epsilon^{2(k)}, \widehat{\rho}^{(k)})$ is $S_{(k)}(\sigma_\epsilon^2, \rho)$ evaluated at $\widehat{\psi}^{(k)} = (\widehat{\sigma}_\epsilon^{2(k)}, \widehat{\rho}^{(k)})$, and $I_{(k)}(\psi) = I_{(k)}(\sigma_\epsilon^2, \rho)$ is the pseudo Fisher score matrix of ψ and calculated by

$$I(\psi) = \begin{pmatrix} I_{11} & I_{12} \\ I_{12} & I_{22} \end{pmatrix} \tag{A.11}$$

where

$$I_{11} = \frac{1}{2}\mathrm{tr}(\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \sigma^2}\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \sigma_\epsilon^2})$$

$$I_{12} = \frac{1}{2}\mathrm{tr}(\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \sigma_\epsilon^2}\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \rho})$$

$$I_{22} = \frac{1}{2}\mathrm{tr}(\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \rho}\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \rho})$$

The updated estimators $\widehat{\beta}^{(k+1)}$, $\widehat{\alpha}_1^{(k+1)}$ and $\widehat{\alpha}_2^{(k+1)}$ can be obtained by the Newton-Raphson (or one-step Newton-Raphson) method based on the "score function" and "information matrix" of $Q_2(\beta, \alpha_3, \alpha_4; \mathbf{O}, \widehat{\Theta}^{(k)})$. The updated estimators $(\widehat{\lambda}_1^{(k+1)}, \cdots, \widehat{\lambda}_m^{(k+1)})$ can be obtained in a similar way based on the "score function" and "information matrix" of $Q_3(\lambda_1, \cdots, \lambda_m; \mathbf{O}, \widehat{\Theta}^{(k)})$. The $(k+1)$th update of μ, ξ and ϕ can be found by using any optimization software, e.g., nlm in R, to maximize $Q_4(\mu, \xi, \phi; \mathbf{O}, \widehat{\Theta}^{(k)})$ under the constraint $E(\mathbf{u}_i) = 0$. The details can be found in the Appendix A.2. We iterate the E-step and M-step until a pre-specified convergence criterion is met. Let $\widehat{\Theta}$ denote the final estimator at convergence.

A.2 Updating Parameters μ, D and ϕ in the EM Algorithm

At the kth step in the EM algorithm, $Q_4(\mu, \xi, \phi; \mathbf{O}, \widehat{\Theta}^{(k)})$ takes the following form:

$$\sum_{i=1}^{n} E\left[\log\left\{P_K^2\{D^{-1}(\mathbf{u}_i - \mu)\}\right\}\right] - \frac{1}{2}\sum_{i=1}^{n} E\left[(\mathbf{u}_i - \mu)'(DD')^{-1}(\mathbf{u}_i - \mu)\right] - n\log|D| \tag{A.12}$$

To satisfy $E(\mathbf{u}_i) = 0$, let $E(\mathbf{u}_i) = \mu + D \cdot E(\mathbf{b}_i) = 0$, then

$$\widehat{\mu}(D, \phi) = -D \cdot E(\mathbf{b}_i) \tag{A.13}$$

Plugging the above $\widehat{\mu}(D, \phi)$ into (A.12) we have

$$\begin{aligned} Q_4(\widehat{\mu}(D,\phi), \xi, \phi; \mathbf{O}, \widehat{\Theta}^{(k)}) &= \sum_{i=1}^{n} E\left[\log\left\{P_K^2\{D^{-1}(\mathbf{u}_i - \widehat{\mu}(D,\phi))\}\right\}\right] \\ &- \frac{1}{2}\sum_{i=1}^{n} E\left[(\mathbf{u}_i - \widehat{\mu}(D,\phi))'(DD')^{-1}(\mathbf{u}_i - \widehat{\mu}(D,\phi))\right] - n\log|D| \end{aligned} \tag{A.14}$$

then the updated estimators $(\hat{D}^{(k+1)}, \hat{\phi}^{(k+1)})$ at $(k+1)$th step can be obtained by using standard optimization software to maximize $Q_4(\widehat{\mu}(D,\phi), \xi, \phi; \mathbf{O}, \widehat{\Theta}^{(k)})$. The updated estimator for μ is then $\widehat{\mu}^{(k+1)} = \widehat{\mu}(\widehat{D}^{(k+1)}, \widehat{\phi}^{(k+1)})$.

$E(\mathbf{b}_i)$ varies for different K values. To ensure $\int g_K(\mathbf{b})d\mathbf{b} = 1$, we need the quadratic constraint $a^T A a = 1$. Because A is positive definite, there exists B positive definite such that $A = BB^T$. Let $c = B'a$, then $c^T c = 1$ and $a = (B')^{-1}c$.

For $K = 1$,

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{A.15}$$

and

$$P_1^2(\mathbf{b}) = (a_{00} + a_{10}b_1 + a_{01}b_2)^2 \tag{A.16}$$

Let $a = (a_{00}, a_{10}, a_{01})'$, define $c = (c_1, c_2, c_3)'$, where $c_1 = \sin\phi_1, c_2 = \cos\phi_1\sin\phi_2$ and $c_3 = \cos\phi_1\cos\phi_2$. $a = c$ and

$$E(\mathbf{b}_i) = E\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 2a_{00}a_{10} \\ 2a_{00}a_{01} \end{pmatrix} \tag{A.17}$$

For $K = 2$,

$$A = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 3 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 3 \end{pmatrix}, \tag{A.18}$$

and

$$P_2^2(\mathbf{b}) = (a_{00} + a_{10}b_1 + a_{01}b_2 + a_{20}b_1^2 + a_{11}b_1b_2 + a_{02}b_2^2)^2 \tag{A.19}$$

Let $a = (a_{00}, a_{10}, a_{01}, a_{20}, a_{11}, a_{02})'$, define $c = (c_1, c_2, c_3, c_4, c_5, c_6)'$ where $c_1 = \sin\phi_1, c_2 = \cos\phi_1\sin\phi_2, c_3 = \cos\phi_1\cos\phi_2\sin\phi_3, \cdots, c_6 = \cos\phi_1\cos\phi_2\cdots\cos\phi_5$. Let $c = B'a$, then $a =$

$(B')^{-1}c$ and

$$E(\mathbf{b}_i) = E\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 2a_{00}a_{10} + 2a_{10}a_{02} + 2a_{01}a_{11} + 6a_{10}a_{20} \\ 2a_{00}a_{01} + 2a_{10}a_{11} + 2a_{01}a_{20} + 6a_{01}a_{02} \end{pmatrix} \tag{A.20}$$

A.3 The Procedure to Obtain the Initial Values for $(\mu', \xi', \phi')'$ in the EM Algorithm

We denote the initial values for $(\mu', \xi', \phi')'$ in the EM algorithm by $(\widehat{\mu}^{(0)}, \widehat{\xi}^{(0)}, \widehat{\phi}^{(0)})$. $(\widehat{\mu}^{(0)}, \widehat{\xi}^{(0)}, \widehat{\phi}^{(0)})$ are obtained from the following procedures:

1. For each element in the vector ϕ, we carry out a grid search for its starting value over $(-\pi/2, \pi/2]$. For example when $K = 1$, $\phi = (\phi_1, \phi_2)$, for each of ϕ_1, ϕ_2, we then choose d numbers equally spaced in the interval $(-\pi/2, \pi/2]$. Consequently, we need to search d^2 pairs of starting values for (ϕ_1, ϕ_2).

2. Since $\mathbf{u}_i = \mu + D\mathbf{b}_i$, by moment method we have

$$\begin{cases} E(u_{i1}) &= \mu_1 + E(b_{i1})D_{00} \\ E(u_{i2}) &= \mu_2 + E(b_{i1})D_{01} + E(b_{i2})D_{11} \end{cases} \tag{A.21}$$

and

$$\begin{pmatrix} \mathrm{var}(u_{i1}) & \mathrm{cov}(u_{i1}, u_{i2}) \\ \mathrm{cov}(u_{i1}, u_{i2}) & \mathrm{var}(u_{i2}) \end{pmatrix} = \begin{pmatrix} U_{11} & U_{12} \\ U_{12} & U_{22} \end{pmatrix}$$

where

$$
\begin{aligned}
U_{11} &= D_{00}^2 \text{var}(b_{i1}) \\
U_{12} &= D_{00} D_{01} \text{var}(b_{i1}) + D_{00} D_{11} \text{cov}(b_{i1}, b_{i2}) \\
U_{22} &= D_{01}^2 \text{var}(b_{i1}) + 2 D_{01} D_{11} \text{cov}(b_{i1}, b_{i2}) + D_{11}^2 \text{var}(b_{i2})
\end{aligned}
\tag{A.22}
$$

We assume $E(u_i) = 0$, then we have

$$
\hat{D}_{00} = \sqrt{\frac{\text{var}(u_{i1})}{\text{var}(b_{i1})}} \tag{A.23}
$$

and

$$
\hat{D}_{11} = \sqrt{\frac{\{\text{var}(u_{i1})\text{var}(u_{i2}) - \text{cov}^2(u_{i1}, u_{i2})\}\text{var}(b_{i1})}{\text{var}(u_{i1})\{\text{var}(b_{i1})\text{var}(b_{i2}) - \text{cov}^2(b_{i1}, b_{i2})\}}} \tag{A.24}
$$

and

$$
\hat{D}_{01} = \frac{\text{cov}(u_{i1}, u_{i2})}{\sqrt{\text{var}(u_{i1})\text{var}(b_{i1})}} - \frac{D_{11}\text{cov}(b_{i1}, b_{i2})}{\text{var}(b_{i1})} \tag{A.25}
$$

where if $K = 1$

$$
\text{var}(b_{i1}) = E(b_{i1}^2) - E^2(b_{i1}) = a_{00}^2 + 3a_{10}^2 + a_{01}^2 - 4a_{00}^2 a_{10}^2 \tag{A.26}
$$

$$
\text{var}(b_{i2}) = E(b_{i2}^2) - E^2(b_{i2}) = a_{00}^2 + a_{10}^2 + 3a_{01}^2 - 4a_{00}^2 a_{01}^2 \tag{A.27}
$$

$$\text{cov}(b_{i1}, b_{i2}) = E(b_{i1}b_{i2}) - E(b_{i1})E(b_{i2}) = 2a_{10}a_{01} - 4a_{00}^2 a_{10}a_{01} \tag{A.28}$$

when $K = 2$

$$\begin{aligned} \text{var}(b_{i1}) &= E(b_{i1}^2) - E^2(b_{i1}) \\ &= a_{00}^2 + 3a_{10}^2 + a_{01}^2 + 15a_{20}^2 + 3a_{11}^2 + 3a_{02}^2 + 6a_{00}a_{20} + 2a_{00}a_{02} + 6a_{20}a_{02} \\ &- (2a_{00}a_{10} + 2a_{10}a_{02} + 2a_{01}a_{11} + 6a_{10}a_{20})^2 \end{aligned} \tag{A.29}$$

$$\begin{aligned} \text{var}(b_{i2}) &= E(b_{i2}^2) - E^2(b_{i2}) \\ &= a_{00}^2 + a_{10}^2 + 3a_{01}^2 + 3a_{20}^2 + 3a_{11}^2 + 15a_{02}^2 + 2a_{00}a_{20} + 6a_{00}a_{02} + 6a_{20}a_{02} \\ &- (2a_{00}a_{01} + 2a_{10}a_{11} + 2a_{01}a_{20} + 6a_{01}a_{02})^2 \end{aligned} \tag{A.30}$$

$$\begin{aligned} \text{cov}(b_{i1}, b_{i2}) &= E(b_{i1}b_{i2}) - E(b_{i1})E(b_{i2}) \\ &= 2a_{00}a_{11} + 2a_{10}a_{01} + 6a_{20}a_{11} + 6a_{02}a_{11} \\ &- (2a_{00}a_{10} + 2a_{10}a_{02} + 2a_{01}a_{11} + 6a_{10}a_{20}) \\ &\quad \cdot (2a_{00}a_{01} + 2a_{10}a_{11} + 2a_{01}a_{20} + 6a_{01}a_{02}) \end{aligned} \tag{A.31}$$

$E(u_{i1}), E(u_{i2}), \text{var}(u_{i1}), \text{var}(u_{i2})$ and $\text{cov}(u_{i1}, u_{i2})$ can be estimated by BLUP estimators of $\mathbf{u}_i, i = 1, \cdots, n$.

3. Now we can obtain the maximum-likelihood estimate of $(\mu', \xi', \phi')'$ to obtain $(\mu^{(0)}, \xi^{(0)}, \phi^{(0)})'$ by treating the BLUP $\widehat{\mathbf{u}}_i^{(0)} = (\widehat{u}_{i0}^{(0)}, \widehat{u}_{i1}^{(0)})$ as observed data and maximizing the density corresponding to $f_{\mathbf{u}_i}(\widehat{\mathbf{u}}_i^{(0)})$ in (1.7). Standard optimization software package is used to maximize $f_{\mathbf{u}_i}(\widehat{\mathbf{u}}_i^{(0)})$. The starting values of that optimization procedure for EM initial values can be found from step 1 and 2 above, where a grid search is carried out for ϕ to ensure that the maximum has been found.

A.4 EM-Aided Algorithm for Variance Estimates

Define $\theta = (\gamma, \beta, \alpha_1, \alpha_2, \lambda_1, \cdots, \lambda_m)$, the regression parameters of our main interest, and ζ the remaining nuisance parameters in Θ. Let $\widehat{\theta}$ denote the MLE of θ obtained by the proposed EM algorithm. Moreover, let $\widehat{\zeta}(\theta) = \big(\widehat{\sigma}_\epsilon^2(\theta), \widehat{\rho}(\theta), \widehat{\mu}'(\theta), \widehat{\xi}'(\theta), \widehat{\phi}'(\theta)\big)'$ denote the maximizer of the observed likelihood function over ζ with θ fixed, which can be obtained by a similar EM algorithm as proposed. Then we calculate the information matrix for θ as follows:

i Perturb the jth component of $\widehat{\theta}$ by a small amount h from both sides, i.e. $\widehat{\theta}_j + h$ and $\widehat{\theta}_j - h$, while keeping other components of $\widehat{\theta}$ unchanged; denote the resulting parameter vectors by $\widehat{\theta}^{j+}$ and $\widehat{\theta}^{j-}$ respectively, and run the EM algorithm to obtain $\widehat{\zeta}(\widehat{\theta}^{j+})$ and $\widehat{\zeta}(\widehat{\theta}^{j-})$ accordingly.

ii Compute the jth row of the information matrix of θ by

$$\frac{1}{2h}\left[E\Big\{l_C^\theta(\widehat{\theta}^{j+}, \zeta(\widehat{\theta}^{j+}))\Big|\mathbf{O}, \widehat{\theta}^{j+}, \widehat{\zeta}(\widehat{\theta}^{j+})\Big\} - E\Big\{l_C^\theta(\widehat{\theta}^{j-}, \zeta(\widehat{\theta}^{j-}))\Big|\mathbf{O}, \widehat{\theta}^{j-}, \widehat{\zeta}(\widehat{\theta}^{j-})\Big\}\right]$$

where $l_C^\theta(\widehat{\theta}^{j+}, \zeta(\widehat{\theta}^{j+}))$ and $l_C^\theta(\widehat{\theta}^{j-}, \zeta(\widehat{\theta}^{j-}))$ denote the derivatives of the complete-data log likelihood $l_C(\theta, \zeta)$ with respect to θ evaluated at $(\widehat{\theta}^{j+}, \zeta(\widehat{\theta}^{j+}))$ and

$(\widehat{\theta}^{j-}, \zeta(\widehat{\theta}^{j-}))$ respectively.

The derivation of $l_C^{\theta}(\theta, \zeta)$ is straightforward and hence omitted here. For the perturbation size h, [3] suggested to use $h = c/n$, where c is a positive constant. We found that $c = 0.15$ would produce satisfactory results in all of our numerical studies.

Appendix B

Appendix of Chapter 2

B.1 Updating Parameters $\lambda_0(\cdot), \eta, \alpha_1, \alpha_2, \beta, \alpha_3, \alpha_4$

The profile Q function for η, α_1, α_2, $Q_1\{\eta, \alpha_1, \alpha_2, \widehat{\lambda}_0(\cdot; \eta, \alpha_1, \alpha_2)\}$, at the k-th step of the EM algorithm is

$$\sum_{i=1}^{n}\Bigg[\sum_{j=1}^{m_i}\Big[\log\{\widehat{\lambda}_0(t_{ij};\eta,\alpha_1,\alpha_2)\} + \eta^T X_i + \alpha_1 E_i(u_{i0}|\mathbf{O}_i,\widehat{\Theta}^{(k)}) + \alpha_2 E_i(u_{i1}|\mathbf{O}_i,\widehat{\Theta}^{(k)})\Big]$$
$$- \quad \Big(\sum_{q=1}^{d}\widehat{a}_q\Big)e^{\eta^T X_i}E_i(e^{\alpha_1 u_{i0}+\alpha_2 u_{i1}}|\mathbf{O}_i,\widehat{\Theta}^{(k)})\Bigg],$$

where $\widehat{a}_q = \widehat{\lambda}_0(t_q; \eta, \alpha_1, \alpha_2)$. Simple algebra shows that the derivative of $Q_1\{\eta, \alpha_1, \alpha_2, \widehat{\lambda}_0(\cdot; \eta, \alpha_1, \alpha_2)\}$ with respect to $(\eta, \alpha_1, \alpha_2)$ has the following expression

$$S_{(k)}(\eta,\alpha_1,\alpha_2) = \sum_{i=1}^{n} m_i\xi_{2i} - \frac{\sum_{q=1}^{d}\sum_{i=1}^{n} e^{\eta^T X_i}E_i(\xi_{1i}e^{\alpha_1 u_{i0}+\alpha_2 u_{i1}}\Big|\mathbf{O}_i,\widehat{\Theta}^{(k)})}{\sum_{i=1}^{n} e^{\eta^T X_i}E_i(e^{\alpha_1 u_{i0}+\alpha_2 u_{i1}}|\mathbf{O}_i,\widehat{\Theta}^{(k)})},$$

where $\xi_{1i} = (X_i', u_{i0}, u_{i1})'$, $\xi_{2i} = \{X_i', E_i(u_{i0}|\mathbf{O}_i, \widehat{\Theta}^{(k)}), E(u_{i1}|\mathbf{O}_i, \widehat{\Theta}^{(k)})\}'$. We then solve the following equation

$$S_{(k)}(\eta, \alpha_1, \alpha_2) = 0 \tag{B.1}$$

to obtain the update for $\psi = (\eta, \alpha_1, \alpha_2)$. However, the profile "score equation" (B.1) does not have a closed form solution. We then propose to update $\psi = (\eta, \alpha_1, \alpha_2)$ by one-step Newton-Raphson algorithm

$$\widehat{\psi}^{(k+1)} = \widehat{\psi}^{(k)} + I_{(k)}^{-1}(\widehat{\psi}^{(k)})S_{(k)}(\widehat{\psi}^{(k)}), \tag{B.2}$$

where $S_{(k)}(\widehat{\psi}^{(k)}) = S_{(k)}(\widehat{\eta}^{(k)}, \widehat{\alpha}_1^{(k)}, \widehat{\alpha}_2^{(k)})$ is $S_{(k)}(\eta, \alpha_1, \alpha_2)$ evaluated at $\widehat{\psi}^{(k)} = (\widehat{\eta}^{(k)}, \widehat{\alpha}_1^{(k)}, \widehat{\alpha}_2^{(k)})$, and $I_{(k)}(\psi) = I_{(k)}(\eta, \alpha_1, \alpha_2)$ is the derivative matrix of $S_{(k)}(\eta, \alpha_1, \alpha_2)$ defined by

$$\begin{aligned} I_{(k)}(\eta, \alpha_1, \alpha_2) &= -\frac{\partial S_{(k)}(\eta, \alpha_1, \alpha_2)}{\partial(\eta, \alpha_1, \alpha_2)} = \sum_{q=1}^{d} \widehat{a}_q \left[\sum_{i=1}^{n} e^{\eta' X_i} E_i(\xi_{1i}^{\otimes 2} e^{\alpha_1 u_{i0} + \alpha_2 u_{i1}} |\mathbf{O}_i, \widehat{\Theta}^{(k)}) \right. \\ & \left. -\widehat{a}_q \left\{ \sum_{i=1}^{n} e^{\eta' X_i} E_i(\xi_{1i} e^{\alpha_1 u_{i0} + \alpha_2 u_{i1}} |\mathbf{O}_i, \widehat{\Theta}^{(k)}) \right\}^{\otimes 2} \right], \end{aligned}$$

where $a^{\otimes 2}$ is defined to be $a^{\otimes 2} = aa'$ for a column vector a. After the updates $(\widehat{\eta}^{(k+1)}, \widehat{\alpha}_1^{(k+1)}, \widehat{\alpha}_2^{(k+1)})$ are available, the update for $\lambda_0(\cdot)$ can be obtained by

$$\widehat{\lambda}_0^{(k+1)}(\cdot) = \widehat{\lambda}_0^{(k+1)}(\cdot; \widehat{\eta}^{(k+1)}, \widehat{\alpha}_1^{(k+1)}, \widehat{\alpha}_2^{(k+1)}). \tag{B.3}$$

$(\beta, \alpha_3, \alpha_4)$ is updated in a similar way: let

$$
\begin{aligned}
S_C(\beta, \alpha_3, \alpha_4) &= \frac{\partial}{\partial(\beta, \alpha_3, \alpha_4)} \log f_{pi}\{Y_i|\mathbf{u}_i, \beta, \alpha_3, \alpha_4\} \\
&= Y_i \xi_{3i} - \frac{\xi_{3i} e^{\beta' H_i + \alpha_3 u_{i0} + \alpha_4 u_{i1}}}{1 + e^{\beta' H_i + \alpha_3 u_{i0} + \alpha_4 u_{i1}}}
\end{aligned} \tag{B.4}
$$

where $\xi_{3i} = (1, \xi'_{1i})'$, then

$$
\begin{aligned}
S_{(k)}(\beta, \alpha_3, \alpha_4) &= \sum_{i=1}^{n} E_i \left\{ S_C(\beta, \alpha_3, \alpha_4) \middle| \mathbf{O}_i, \widehat{\Theta}^{(k)} \right\} \\
&= \sum_{i=1}^{n} \left\{ Y_i \xi_{4i} - E_i \left(\frac{\xi_{3i} e^{\beta' H_i + \alpha_3 u_{i0} + \alpha_4 u_{i1}}}{1 + e^{\beta' H_i + \alpha_3 u_{i0} + \alpha_4 u_{i1}}} \middle| \mathbf{O}_i, \widehat{\Theta}^{(k)} \right) \right\},
\end{aligned} \tag{B.5}
$$

where $\xi_{4i} = (1, \xi'_{2i})'$, and

$$
I_{(k)}(\beta, \alpha_3, \alpha_4) = -\frac{\partial S_{(k)}(\beta, \alpha_3, \alpha_4)}{\partial(\beta, \alpha_3, \alpha_4)} = \sum_{i=1}^{n} E_i \left\{ \xi_{3i}^{\otimes 2} \frac{e^{\beta' H_i + \alpha_3 u_{i0} + \alpha_4 u_{i1}}}{(1 + e^{\beta' H_i + \alpha_3 u_{i0} + \alpha_4 u_{i1}})^2} \middle| \mathbf{O}_i, \widehat{\Theta}^{(k)} \right\},
$$

Then

$$
(\widehat{\beta}^{(k+1)}, \widehat{\alpha}_3^{(k+1)}, \widehat{\alpha}_4^{(k+1)}) = (\widehat{\beta}^{(k)}, \widehat{\alpha}_3^{(k)}, \widehat{\alpha}_4^{(k)}) + I_{(k)}^{-1}(\widehat{\beta}^{(k)}, \widehat{\alpha}_3^{(k)}, \widehat{\alpha}_4^{(k)}) \cdot S_{(k)}(\widehat{\beta}^{(k)}, \widehat{\alpha}_3^{(k)}, \widehat{\alpha}_4^{(k)}) \tag{B.6}
$$

B.2 Score Function in the EM-Aided Algorithm

For $\theta = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta, \gamma, \zeta, \eta)$.

$$
\frac{\partial l_C(\theta, \rho)}{\partial \theta} = \left\{ S'_C(\eta, \alpha_1, \alpha_2), S'_C(\gamma, \zeta), S'_C(\beta, \alpha_1, \alpha_2) \right\}' \tag{B.7}
$$

where $S_C(\eta, \alpha_1, \alpha_2)$ and $S_C(\beta, \alpha_1, \alpha_2)$ are defined as above. Moreover

$$\begin{aligned} S_C(\gamma, \zeta) &= \frac{\partial}{\partial(\gamma, \zeta)} \log f_{li}(Z_i | \mathbf{u}_i, \gamma, \zeta, \sigma_\epsilon^2) \\ &= \frac{1}{\sigma_\epsilon^2} \sum_{j=0}^{m_i} \left[(X_i', B_{ij}')' \cdot \Big(Z_{ij} - \zeta' B_{ij} - \gamma' X_i - u_{i0} - u_{i1} t_{ij} \Big) \right] \end{aligned} \quad \text{(B.8)}$$