STATISTICAL INFERENCE FOR JOINT MODELLING OF LONGITUDINAL AND SURVIVAL DATA

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Engineering and Physical Sciences

2014

Qiuju Li
School of Mathematics
Contents

Abstract 16

Declaration 17

Copyright Statement 18

Dedication 19

Acknowledgements 20

1 Introduction 21
   1.1 Models for longitudinal data 21
      1.1.1 General overview 21
      1.1.2 Linear mixed-effects model 22
      1.1.3 Generalised linear mixed-effects model 23
   1.2 Models for survival data 24
   1.3 Models for joint analysis of longitudinal and survival data 26
   1.4 Covariance estimation and modelling 29
   1.5 Outline of the subsequent chapters 33

2 Bayesian inference for joint modelling of longitudinal continuous, binary and ordinal events 36
   2.1 Introduction 36
   2.2 Statistical Models and Methods 39
      2.2.1 Statistical Models 39
      2.2.2 Statistical Methods 41
      2.2.3 Model selection 44
3 Survival analysis with longitudinal covariates measured with correlated error

3.1 Introduction ................................................. 64
3.2 Model Definition .......................................... 67
3.3 Conditional Score Estimator ............................... 68
3.4 Bias Analysis .............................................. 69
3.5 Generalised Conditional Score Estimator .................. 72
3.6 Covariance Estimator and the Inference ................. 77
   3.6.1 Software Implementation ............................. 78
   3.6.2 Modified Cholesky Decomposition .................... 89
3.7 Application to ACTG 175 Data ............................. 93
3.8 Discussion .............................................. 99
3.9 Chapter appendices ..................................... 101

4 Survival analysis with cumulative information of longitudinal covariates

4.1 Introduction .............................................. 111
4.2 Models ................................................... 114
4.3 Statistical inference ................................... 118
4.4 Estimation procedure .................................. 122
4.5 Real data analysis .................................... 126
   4.5.1 Schizophrenia trial data .............................. 126
   4.5.2 Liver cirrhosis trial data ............................ 130
4.6 Simulation study ....................................... 136
   4.6.1 General simulation settings ......................... 136
   4.6.2 Bias analysis ....................................... 138
   4.6.3 Case 1: statistical inference for the cumulative models . . . . . . 142
   4.6.4 Case 2: cumulative analysis for default current models ........... 150
4.7 Discussion ........................................................................ 155
4.8 Comments: why not other ways to characterize the curve \( \gamma(u) \) ? .... 157
4.9 Derivatives and log-concavity of parameters and Laplace approximation 162

5 Conclusions and future work ................................................. 182
  5.1 Summary of the thesis ...................................................... 182
  5.2 Future work ................................................................... 186
    5.2.1 Model selection ...................................................... 186
    5.2.2 Informative visiting process .................................... 186
    5.2.3 Measure theory to serve the joint modelling framework .... 187

6 Additional materials ............................................................... 199

Word count xxxxx
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Estimates of Fixed Effect Coefficients–Scenario 1</td>
<td>46</td>
</tr>
<tr>
<td>2.2</td>
<td>Estimates of Fixed Effect Coefficients–Scenario 2</td>
<td>46</td>
</tr>
<tr>
<td>2.3</td>
<td>Simulation output results for scenario 3</td>
<td>53</td>
</tr>
<tr>
<td>2.4</td>
<td>Percentage of $\text{DIC}<em>{\text{Joint}} \leq \text{DIC}</em>{\text{Sep}}$ when varying $\rho$ in CS of $G$–Scenario[4]</td>
<td>53</td>
</tr>
<tr>
<td>2.5</td>
<td>Efficiencies of fixed effects in the three models each with random intercept and random slope</td>
<td>53</td>
</tr>
<tr>
<td>2.6</td>
<td>Covariance Parameters Estimators</td>
<td>54</td>
</tr>
<tr>
<td>2.7</td>
<td>Statistical Inference for the NorStOP Data</td>
<td>57</td>
</tr>
<tr>
<td>3.1</td>
<td><strong>Scenario I(a)(b): AR(1).</strong> Simulation results for $\rho = -0.6$ and $\sigma^2 = 0.6$ of AR(1) within-subject covariance. All the possible measured time points are at weeks of ${0,2,4,8,16,24,32,40,48,56,64,72,80}$ and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.</td>
<td>81</td>
</tr>
<tr>
<td>3.2</td>
<td><strong>Scenario I(c)(d): AR(1).</strong> Simulation results for $\rho = -0.6$ and $\sigma^2 = 0.6$ of AR(1) within-subject covariance. All the possible measured time points are at weeks of ${0,2,4,8,16,24,32,40,48,56,64,72,80}$ and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.</td>
<td>82</td>
</tr>
</tbody>
</table>
3.3 Scenario II(a)(b): AR(1) Simulation results for $\rho = 0.25$ and $\sigma^2 = 0.6$ of AR(1) within-subject covariance. All the possible measured time points are at weeks of \{0, 2, 4, 6, 8, 10, 12, 14, 16, 24, 32, 40, 48, 56, 64, 72, 80\} and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.

3.4 Scenario II(c)(d): AR(1) Simulation results for $\rho = 0.25$ and $\sigma^2 = 0.6$ of AR(1) within-subject covariance. All the possible measured time points are at weeks of \{0, 2, 4, 6, 8, 10, 12, 14, 16, 24, 32, 40, 48, 56, 64, 72, 80\} and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.

3.5 Scenario III(a)(b): AR(1) Simulation results for $\rho = 0.5$ and $\sigma^2 = 0.6$ of AR(1) within-subject covariance. All the possible measured time points are at weeks of \{0, 2, 4, 6, 8, 10, 12, 14, 16, 24, 32, 40, 48, 56, 64, 72, 80\} and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.

3.6 Scenario III(c)(d): AR(1) Simulation results for $\rho = 0.5$ and $\sigma^2 = 0.6$ of AR(1) within-subject covariance. All the possible measured time points are at weeks of \{0, 2, 4, 6, 8, 10, 12, 14, 16, 24, 32, 40, 48, 56, 64, 72, 80\} and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.
3.7 **Scenario IV(a)(b): AR(1)** Simulation results for $\rho = 0.75$ and $\sigma^2 = 0.6$ of AR(1) within-subject covariance. All the possible measured time points are at weeks of $\{0,2,4,6,8,10,12,14,16,24,32,40,48,56,64,72,80\}$ and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(\%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(\%), Monte Carlo coverage probability of 95% Wald confidence interval.

3.8 **Scenario IV(c)(d): AR(1)** Simulation results for $\rho = 0.75$ and $\sigma^2 = 0.6$ of AR(1) within-subject covariance. All the possible measured time points are at weeks of $\{0,2,4,6,8,10,12,14,16,24,32,40,48,56,64,72,80\}$ and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(\%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(\%), Monte Carlo coverage probability of 95% Wald confidence interval.

3.9 **Scenario V: Modified Cholesky Decomposition** Simulation results for Cholesky decomposition of within-subject covariance. All the possible measured time points are at weeks of $\{0,2,4,8,16,24,32,40,48,56,64,72,80\}$ and with 10% chance of missing after week 16. RB(\%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(\%), Monte Carlo coverage probability of 95% Wald confidence interval. GCS-ChoD, the MEst output of generalized conditional score estimation with Cholesky decomposition of within-subject covariance; GCS-ChoD(mle), the MLE output of generalized conditional score estimation with Cholesky decomposition of within-subject covariance.
3.10 **Scenario V: Modified Cholesky Decomposition** Simulation results for Cholesky decomposition of within-subject covariance. All the possible measured time points are at weeks of \{0, 2, 4, 8, 16, 24, 32, 40, 48, 56, 64, 72, 80\} and with 10% chance of missing after week 16. RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval. GCS-ChoD, the MEst output of generalized conditional score estimation with Cholesky decomposition of within-subject covariance; GCS-ChoD(mle), the MLE output of generalized conditional score estimation with Cholesky decomposition of within-subject covariance.

3.11 **Statistical Inference for ACTG175 data** Analysis results are showed for the CS and GCS approaches, and for the GCS approach, the within-subject covariance was captured by AR(1) with lme R command and by the data-driven method of modified Cholesky decomposition. Both MEst and MLE estimations are presented for each approach, except for the GCS with Cholesky decomposition covariance. Values inside the parenthesis are the standard deviation of corresponding estimators. sd, standard deviation of estimators, ** indicates the significance of covariate of 95% Wald confidence interval.

4.1 15-point Gauss-Kronrod quadrature rules

4.2 Analysis outputs for the PANSS data: ‘est.’ estimates of the parameters; ‘sd’ standard deviation; \( \rho_{ij} = \text{cor}(\alpha_i, \alpha_j) \). JM outputs are from R package with current default values of longitudinal process as covariate of Cox regression model with method=“spline-PH-aGH”, other model settings are the same with that of cumulative analysis.
4.3 Analysis outputs for liver cirrhosis trial data: ‘est.’ estimates of the parameters; ‘sd’ standard deviation; $\rho_{ij} = \text{cor}(\alpha_i, \alpha_j)$. JM outputs are from R package with current default values of longitudinal process as covariate of Cox regression model with method=“spline-PH-aGH”, other model settings are the same with that of cumulative analysis. Note: cumulative term only for the cumulative analysis and the parameter $a$ is not comparable between two analysis.

4.4 Bias analysis $B^c(a, b, t)$. $M_0$ indicates the underlying models, from which analysis data are generated; $B^0(a, -, t)$ and $B^c(a, b, t)$ indicate the underlying models come along with hazard function of (4.16) and (4.17). Par: parameters; Truval.: true values of parameters; Mean: mean of the parameter estimators from Monte Carlo datasets. SD: standard deviation of the Monte Carlo estimators; SE: mean of approximations of standard deviation for each Monte Carlo dataset; RB(%): relative bias; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval. And the simulated data are fitted by $M_1$: $B^0(a, -, t)$.

4.5 Bias analysis $B(a, b, t)$. $M_0$ indicates the underlying models, from which analysis data are generated; $B^0(a, -, t)$ and $B(a, b, t)$ indicate the underlying models come along with hazard function of (4.16) and (4.18). Par: parameters; Truval.: true values of parameters; Mean: mean of the parameter estimators from Monte Carlo datasets. SD: standard deviation of the Monte Carlo estimators; SE: mean of approximations of standard deviation for each Monte Carlo dataset; RB(%): relative bias; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval. And the simulated data are fitted by $M_1$: $B^0(a, -, t)$.

4.6 Cumulative analysis of models with $B^c(a, b, t)$, moderate baseline hazards; $\lambda_0(t) = 1$ when $b = 0.5$, otherwise $\lambda_0(t) = 0.5$. Par: parameters; Truval.: true values of parameters; Mean: mean of the parameter estimators from Monte Carlo datasets. SD: standard deviation of the Monte Carlo estimators; SE: mean of approximations of standard deviation for each Monte Carlo dataset; RB(%): relative bias; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.
4.7 Cumulative analysis of models with $B^c(a, b, t)$, small baseline hazards; 
$\lambda_0(t)=1e-07$ when $b \in \{0.5, 1.0, 1.5\}$; $\lambda_0(t)=1e-06$ when $b \in \{2, 2.5, 3.0\}$. 
Par: parameters; Truval.: true values of parameters; Mean: mean of the parameter estimators from Monte Carlo datasets. SD: standard deviation of the Monte Carlo estimators; SE: mean of approximations of standard deviation for each Monte Carlo dataset; RB(%): relative bias; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.

4.8 Cumulative analysis of models with $B(a, b, t)$. Par: parameters; Truval.: true values of parameters; Mean: mean of the parameter estimators from Monte Carlo datasets. SD: standard deviation of the Monte Carlo estimators; SE: mean of approximations of standard deviation for each Monte Carlo dataset; RB(%): relative bias; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.

4.9 Estimations for the distinct elements of covariance matrix $G$ of the simulation scenario with $B^c(a = 1, b = 1, t)$ in Table 4.6; Trueval.: the true underlying values; Mean: mean of the Monte Carlo estimators.

4.10 Cumulative analysis represents default current models. $M_0$ denotes the underlying models Monte Carlo data generated from; $M_1$ denotes the fitting models; $B(a, b = 6, t)$ indicates that data are fitted by model with $B(a, b, t)$ but with $b = 6$ fixed; “$M_1: B(a, -, t)$” indicates that inference outputs for the default current model are obtained from JM package with argument of method=“spline-PH-aGH” and method=“piecewise-PH-aGH” presented by top and middle panel, respectively. Note: analysis of three models/approaches are applied to the same Monte Carlo datasets and $\lambda_0(t) = 1$ of the underlying setting.

6.1 Bias analysis. $M_0$ denote the underlying models and they are fitted by the default current model $B^9(a, -, t)$. Inference are obtained by using the JM package with argument of method=“piecewise-PH-GH”.

6.2 PANSS data from APTS short course (blank is missing) - Table 1/4

6.3 PANSS data from APTS short course (blank is missing) - Table 2/4

6.4 PANSS data from APTS short course (blank is missing) - Table 3/4
6.5 PANSS data from APTS short course (blank is missing) - Table 4/4
List of Figures

2.1 Comparison of stability of estimates sd of fixed-effect coefficients between separated and joint models in the simulation studies of scenario 2 with 100 Monte Carlo datasets. .............................................. 49

2.2 The efficiencies of fixed effects against (a) the correlation coefficient $\rho$ in CS, by assuming $\text{diag}(G) = (16, 4, 2.25)^T$; (b) (part of) the variance of random effects when $\rho = 0.75$ in CS, by assuming $\text{diag}(G) = \sigma_b^2(4, 1, 0.72)^T$; (c) when $\rho = 0.5$; (d) when $\rho = 0.25$, where the curves are fitted using nonparametric loess method in R. ............................................. 50

3.1 Conditional score (CS) estimators of survival parameters, varying with correlation $\rho$, with the mutual independent assumption on $\varepsilon_i$ for each subject $i$ regardless underlying covariance of $\varepsilon_i$ of AR(1)=$\sigma^2|\rho|i - j|$ with $\sigma^2 = 0.6$; x-axis is for $\rho$, y-axis are CS estimators for (a) $\gamma$; dotted horizontal line indicates true parameter value, solid black line is CS estimates with point estimators surrounding, and the dash-dot line is for Naive estimators; 500 Monte Carlo datasets and 300 sample size for each $\rho$. ................................................................. 73

3.2 Conditional score (CS) estimators of survival parameters, varying with correlation $\rho$, with the mutual independent assumption on $\varepsilon_i$ for each subject $i$ regardless underlying covariance of $\varepsilon_i$ of AR(1)=$\sigma^2|\rho|i - j|$ with $\sigma^2 = 0.6$; x-axis is for $\rho$, y-axis are CS estimators for (b) $\eta_1$; dotted horizontal line indicates true parameter value, solid black line is CS estimates with point estimators surrounding, and the dash-dot line is for Naive estimators; 500 Monte Carlo datasets and 300 sample size for each $\rho$. ................................................................. 74
3.3 Conditional score (CS) estimators of survival parameters, varying with correlation $\rho$, with the mutual independent assumption on $\varepsilon_i$ for each subject $i$ regardless underlying covariance of $\varepsilon_i$ of AR(1)=$\sigma^2|\rho|^{i-j}$ with $\sigma^2 = 0.6$; x-axis is for $\rho$, y-axis are CS estimators for (c) $\eta_2$; dotted horizontal line indicates true parameter value, solid black line is CS estimates with point estimators surrounding, and the dash-dot line is for Naive estimators; 500 Monte Carlo datasets and 300 sample size for each $\rho$.

4.1 Possible shapes for weighted curves $\gamma(t-u)$

4.2 Possible shapes for time-lag weighted curves $\gamma(u)$

4.3 Survival part of the log-likelihood function of complete data with respect to pair $(a, b)$, other parameters are given as discussed in the section; the red round point indicates the maximum point of log-likelihood given other parameters setting.

4.4 Analysis for PANSS data, from left to right they are for weighted curve, $\gamma(t-u)$, past trajectory of PANSS score $m_i(u)$, and the integral of cumulative information $\gamma(t-u)m_i(u)$, where $t$ stands for the survival time and $u$ is the time before survival time. The grey area is the integration value $\int_0^t \gamma(t-u)m_i(u)du$. The dash lines in first column are 95% confidence interval for the weighted curve, and each row is for different subject.

4.5 Illustration of grid search for initial values of $(a, b)$ when both current and cumulative terms are taken into account the Cox model for liver data.

4.6 Illustrations of grid search for initial values of $(a, b)$ when cumulative term (without default current values) is taken into account the Cox model for liver data.
4.7 Analysis for liver cirrhosis trial data, from left to right they are for weighted curve, \( \gamma(t - u) \), past trajectory of prothrombin index \( m_i(u) \), and the integral of cumulative information \( \gamma(t - u)m_i(u) \), where \( t \) stands for the survival time and \( u \) is the time before survival time. The grey area is the integration value \( \int_0^t \gamma(t - u)m_i(u)du \). The dash lines in first column are 95% confidence interval for the weighted curve, and each row is for different subject.

4.8 Bias analysis \( B^c(a, b, t) \). The solid black lines are the underlying time-varying effects of cumulative analysis with \( B^c(a, b, t) \); the dash-black lines are estimators of associations from default current value of hazard model with \( B^0(a, -, t) \), which are time-constant. Top panel is for the scenario when survival data are from underlying model with \( B^c(a = 0.8, b = 0.5, t) \), middle panel with \( B^c(a = 0.8, b = 1, t) \) and bottom panel with \( B^c(a = 0.8, b = 4, t) \). No dash lines in plots of second and third columns due to that the default current value model does not consider the derivative effects of \( m_i(t) \).

4.9 Bias analysis \( B(a, b, t) \). The solid black lines are the underlying time-varying effects of cumulative analysis with \( B(a, b, t) \); the dash-black lines are estimators of associations from default current value of hazard model with \( B^0(a, -, t) \), which are time-constant. Top panel is for the scenario when survival data are from underlying model with \( B(a = 1, b = 0.5, t) \), middle panel with \( B(a = 1, b = 1, t) \) and bottom panel with \( B(a = 1, b = 4, t) \). No dash lines in plots of second and third columns due to that the default current value model does not consider the derivative effects of \( m_i(t) \).

4.10 Density functions of estimators of survival parameters \( (\eta_1, \eta_2, a, b)^T \) from 300 Monte Carlo datasets and 500 sample size each, with the underlying parameter setting \( (\eta_1, \eta_2, a, b)^T = (-1.5, -1, 1.2, 2.5)^T \) and the hazard model is with term \( B^c(a, b, t) \); moderate baseline hazards, \( \lambda_0(t) = 0.5 \).
4.11 Density functions of estimators of survival parameters \((\eta_1, \eta_2, a, b)^T\) from 300 Monte Carlo datasets and 500 sample size each, with the underlying parameter setting \((\eta_1, \eta_2, a, b)^T = (1.5, -1, 1, 0.5)^T\) and the hazard model is with term \(B^c(a, b, t)\); small baseline hazards \(\lambda_0(t) = 1e - 07\).  

4.12 Basis functions of B-spline with order 4, i.e. cubic splines.  

4.13 Integral functions of B-spline basis and time polynomial with degree 1 (upper panel) and degree 2 (lower panel).  

4.14 Collinearity of covariates for the B-spline coefficients, where we have \(\int \text{Bk}[j] = \sum_{\ell=0}^q \alpha_{\ell \xi} \int_0^t B_{j,k}(u)(t-u)^\ell du\), the covariate for coefficient parameter \(\xi_j\).  

6.1 Density functions of estimators of survival parameters of default current value model \(B^0(a, -, t)\) with correct model fitting.
In longitudinal studies, data collected within a subject or cluster are somewhat correlated by their very nature and special cares are needed to account for such correlation in the analysis of data. Under the framework of longitudinal studies, three topics are being discussed in this thesis.

In chapter 2, the joint modelling of multivariate longitudinal process consisting of different types of outcomes are discussed. In the large cohort study of UK north Staffordshire osteoarthritis project, longitudinal trivariate outcomes of continuous, binary and ordinary data are observed at baseline, year 3 and year 6. Instead of analysing each process separately, joint modelling is proposed for the trivariate outcomes to account for the inherent association by introducing random effects and the covariance matrix G. The influence of covariance matrix G on statistical inference of fixed-effects parameters has been investigated within the Bayesian framework. The study shows that by joint modelling the multivariate longitudinal process, it can reduce the bias and provide with more reliable results than it does by modelling each process separately.

Together with the longitudinal measurements taken intermittently, a counting process of events in time is often being observed as well during a longitudinal study. It is of interest to investigate the relationship between time to event and longitudinal process, on the other hand, measurements taken for the longitudinal process may be potentially truncated by the terminated events, such as death. Thus, it may be crucial to jointly model the survival and longitudinal data. It is popular to propose linear mixed-effects models for the longitudinal process of continuous outcomes and Cox regression model for survival data to characterize the relationship between time to event and longitudinal process, and some standard assumptions have been made. In chapter 3, we try to investigate the influence on statistical inference for survival data when the assumption of mutual independence on random error of linear mixed-effects models of longitudinal process has been violated. And the study is conducted by utilising conditional score estimation approach, which provides with robust estimators and shares computational advantage. Generalised sufficient statistic of random effects is proposed to account for the correlation remaining among the random error, which is characterized by the data-driven method of modified Cholesky decomposition. The simulation study shows that, by doing so, it can provide with nearly unbiased estimation and efficient statistical inference as well.

In chapter 4, it is trying to account for both the current and past information of longitudinal process into the survival models of joint modelling. In the last 15 to 20 years, it has been popular or even standard to assume that longitudinal process affects the counting process of events in time only through the current value, which, however, is not necessary to be true all the time, as recognised by the investigators in more recent studies. An integral over the trajectory of longitudinal process, along with a weighted curve, is proposed to account for both the current and past information to improve inference and reduce the under estimation of effects of longitudinal process on the risk hazards. A plausible approach of statistical inference for the proposed models has been proposed in the chapter, along with real data analysis and simulation study.
Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s Policy on Presentation of Theses.
Dedication

To my family
Acknowledgements

I would like to express sincere thankfulness and appreciation to my supervisor, Professor Jianxin Pan, for his excellent supervision throughout my PhD study and the financial support from his grants for the maintenance of my first two-year study. Also thank Professor Gilbert MacKenzie and Dr Christiana Charalambous for agreeing to be examiners.

It is an honour to study in the University of Manchester and I would like to sincerely thank the school of mathematics for being tremendously supportive in the past few years, particularly on the scholarship and financial support for PG students attending conference which made my treasure attendances to the Joint Statistical Meetings (Montreal, Canada, 2013) and ICSA conference (Hong Kong, 2013). My appreciation also goes to university IT service team for the computing support. I would also like to thank all the officemates for being nicely around, thanks to Kun, Hannah, Lina, Tarak, Ioanna, Mario, Qingze and Yi, especially thank Qingze for sharing his computer running programming codes.

And my deep appreciation to Prof. Jinglong Wang for his tremendous and elegant knowledge on statistics which guided me to build up the way of thinking statistics, and to Prof. Qinfeng Xu for his constructive mentoring during my master study. Also I would like to particularly thank my best friend, Miss Yang Xiang, who is an excellent lady and has an important impact on my life, especially the 20s. She is the person who motived and inspired me to work harder to pursue something good in our life.

Last, but certainly not least, I would like to thank my parents, sister and brothers for all of the support and would like to say I must have inherited my mother’s intelligence. My deepest thanks and love to my husband, Kunpeng, without his love, patience, support and guidance, completion of this thesis would not have been possible.
Chapter 1

Introduction

1.1 Models for longitudinal data

1.1.1 General overview

Longitudinal studies nowadays play an important role in several fields, such as in the study of clinical trials, medical research, social and economics, and they are indispensable to the study of change in outcomes over time. In longitudinal studies, repeated measurements of some outcomes/variables are usually taken routinely on each participant, which results in the collected data being correlated. In traditional statistical methodology, one fundamental assumption is that the data collected are random samples from the whole sample space and share the mutual independence, which, however, is obviously violated by the data collected within subjects in longitudinal studies. Repeated measurements arising from longitudinal studies, by their very nature, are usually correlated and have a complex random-error structure, and special care is needed to analyse the data. In past decades, there have been many remarkable developments in statistical methodology for longitudinal data analysis. Longitudinal studies vary in the types of outcomes of interest, for example, when the outcomes are continuous and may be approximately normally distributed, the linear mixed-effects models (LMM) have been the dominant approach for the analysis of longitudinal data. For the discrete and non-Gaussian longitudinal data, instead, the method has been extended to generalised linear mixed-effects models (GLMM). In addition, Liang and Zeger (1986) proposed the generalised estimating equations (GEE) approach for analysing discrete
longitudinal data, which has been used with advantage.

1.1.2 Linear mixed-effects model

In longitudinal studies, repeated measurements of an outcome for each subject are taken over time. The CD4+ cell counts, for example, are observed at different time points for each individual of 369 HIV-infected men enrolled in the Multicenter AIDS Cohort Study to monitor the disease progression, together with other covariates informations (Diggle et.al., 2002). Denote $Y_{ij}$ the $j^{th}$ measurement of the $i^{th}$ subject taken at the pre-specified time point of $t_{ij}$. For the $i^{th}$ of $m$ subjects, number of $n_i$ measurements are taken, that is, we have $j = 1, 2, \ldots, n_i$ and $i = 1, 2, \ldots, m$. Laird and Ware (1982) proposed the linear mixed-effects model for the analysis of longitudinal data,

$$Y_i = X_i \beta + Z_i b_i + \varepsilon_i, \quad i = 1, 2, \ldots, m,$$

where $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})^T$, $\beta$ is $p \times 1$ fixed-effects coefficient, $b_i$s are the $q \times 1$ subject-level random effects, and $X_i$ and $Z_i$ are $n_i \times p$ and $n_i \times q$ design matrices for fixed and random effects, respectively, which may or may not share common elements. Both design matrices can include baseline and/or time-dependent covariates, such as the age, gender and the measured timepoints, $t_i = (t_{i1}, t_{i2}, \ldots, t_{in_i})^T$. Random errors $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in_i})^T$ are typically assumed that they are from multivariate normal distribution with mean zero and covariance $\Sigma_i$ and that they are mutually independent from the random effects $b_i$s. Regarding the random effects $b_i$s, it is also popular to assume that they are normally distributed with $b_i \sim N(0, G)$. The random effects are un-observed and it is somewhat difficult to verify the normally distributed assumption, some studies have been conducted to investigate the performance of the model (1.1) when the normal assumptions or the covariance structure $G$ have been violated and concluded that the influence of the misspecification on the fixed effects coefficients might not be significant in terms of the estimations for parameters, but it might loose some efficiency and lead to biased inference for the individual trajectories of the longitudinal process (Daniels and Zhao, 2003, Song, Davidian, and Tsiatis, 2002b, for instance). By taking a multivariate normality assumption on the random effects and
random errors, the conditional and marginal distributions are,

\[ Y_i | b_i \sim N(X_i\beta + Z_i b_i, \Sigma_i), \]
\[ Y_i \sim N(X_i\beta, R_i), \quad R_i = Z_i G Z_i^T + \Sigma_i. \] (1.2)

As we know, design matrix \( Z_i \) may be time-dependent, which indicates that the variance of response \( Y_{ij} \) may be time varying. In the meantime, the covariance of random errors, \( \Sigma_i \), can have any types of covariance structures, such as compound symmetry, AR(1) or simply the diagonal matrix \( \sigma^2 I_{n_i} \), where \( I_{n_i} \) is the identity matrix of size \( n_i \).

Laird and Ware (1982) showed that the expectation-maximization (EM) algorithm (Dempster, Laird, and Rubin, 1977) could be utilised to obtain the statistical inference for model (1.1), and soon after, Jennrich and Schluchter (1986) proposed Newton-Raphson and Fisher scoring algorithms, among others, to fit the model(1.1). To date, the maximum likelihood and restricted maximum likelihood estimation are the most popular methods for estimation and inference (Verbeke and Molenberghs, 2000).

### 1.1.3 Generalised linear mixed-effects model

When the outcomes of interest are discrete-type data, such as the binary and ordinal types of responses introduced in chapter 2 are collected in a longitudinal study, linear mixed-effects model can not be utilised directly, instead, it has been extended to the generalised linear mixed-effects model (GLMM). And the GLMM has three components,

1. Distribution assumption for random effects, e.g., \( b_i \sim N(0, G) \), and given the random effects, responses are from exponential family, that is,

\[ f_{Y_{ij}}(y_{ij}|b_i) = \exp \left( \frac{y_{ij} \theta_i - b(\theta_i)}{a(\phi)} + c(y_{ij}, \phi) \right); \]

2. Linear predictor, \( \eta_{ij} = x_{ij}^T \beta + z_{ij}^T b_i \);

3. Link function, \( g(\mu_{ij}) = \eta_{ij} = x_{ij}^T \beta + z_{ij}^T b_i, \ j = 1, 2, \ldots, n_i, \ i = 1, 2, \ldots, m, \)

where \( \mu_{ij} = E(Y_{ij}|b_i) \), \( a(\cdot) \), \( b(\cdot) \) and \( c(\cdot) \) are known functions and \( g(\cdot) \) is the link function to connect covariates and responses, such as the probit and logit link functions. \( x_{ij} \) and \( z_{ij} \) are the vectors of covariates associated with the fixed and random effects, \( \beta \)
CHAPTER 1. INTRODUCTION

and $b_i$, respectively. The conditional variance is assumed to depend on the conditional mean according to $\text{Var}(Y_{ij}|b_i) = \phi \nu(\mu_{ij})$, where $\nu(\cdot)$ is a known variance function and $\phi$ is a scale dispersion parameter that may be known or may need to be estimated.

Together with the generalised linear mixed-effects model and others, generalised estimating equations approach provides an excellent alternative for analysing the longitudinal discrete data, although which we will not pursue this approach in the thesis.

1.2 Models for survival data

Time-to-event data are of particular interest in a number of applied fields, such as in biology and medicine, time to occurrence of certain events or time to death from the onset of a disease, for example. Let $\tilde{T}$ denote the random variable of time to event, a fundamental quantity of interest in survival analysis is the hazard function, defined by

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{Pr(t \leq \tilde{T} < t + \Delta t|\tilde{T} \geq t)}{\Delta t}. \quad (1.3)$$

If $\tilde{T}$ is a continuous random variable with probability and survival functions of $f(t)$ and $S(t)$, respectively, then $\lambda(t) = f(t)/S(t) = -d\log(S(t))/dt$, and the cumulative hazard function $\Lambda(t)$, defined by

$$\Lambda(t) = \int_0^t \lambda(u)du = -\log(S(t)).$$

Thus, the survival function for the continuous survival time has

$$S(t) = Pr(\tilde{T} > t) = \exp(-\Lambda(t)) = \exp\left(-\int_0^t \lambda(u)du\right). \quad (1.4)$$

Time-to-event data present themselves in different ways, which lead to special problems in analysing such data. Censoring, for instance, can subject to left censoring, right censoring or censoring within certain intervals. A right-censoring mechanism is said to be independent or non-informative if the hazard rates that apply to individuals on trial at each time are the same as those that would have applied had there been no censoring (Kalbfleisch and Prentice, 2002). Throughout this thesis, non-informative censoring is assumed with respect to the censoring process of observed data. However, when the censoring is informative, additional care is required in the data analysis.
CHAPTER 1. INTRODUCTION

When time-to-event characters are of interest, the observed response of \(i^{th}\) subject consists of \((T_i, \delta_i)\), where \(T_i = \min(\tilde{T}_i, C_i)\), \(\delta_i = I(\tilde{T}_i \leq C_i)\). Furthermore, \(\tilde{T}_i\) is the underlying time to event, which may or may not be observed, and \(C_i\) is the censoring time which is independent of \(\tilde{T}_i\) and the covariates \(X_i\) if applicable. Denote the distinct failure times by \(V_{(1)} < V_{(2)} < \cdots < V_{(k)}\), where \(k\) is the number of distinct failure times of the observed data, the survival function \(S(t)\) can be estimated by the Kaplan-Meier estimator (Kaplan and Meier, 1958), defined by

\[
\hat{S}(t) = \prod_{j: V_{(j)} \leq t} (1 - \frac{d_j}{r_j}),
\]

where \(d_j\) is the number of subjects with events occurring at time \(V_{(j)}\) and \(r_j\) is the number of individual that are at risk at time \(V_{(j)}\).

One of the most popular approach for analysing survival data is to utilise the semi-parametric proportional hazards model (PH model) proposed by Cox (1972) to capture the relationship of hazard rates with covariates on a log scale,

\[
\lambda_i(t) = \lim_{\Delta t \to 0} \frac{Pr(t \leq \tilde{T}_i < t + \Delta t | \tilde{T}_i \geq t)}{\Delta t} = \lambda_0(t) \exp(X_i^T \beta),
\]

(1.5)

where \(\lambda_0(t)\) is arbitrary unspecified non-negative baseline hazard function, which makes the model semi-parametric, \(X_i\) is the covariates vector and may include time-independent and/or time-dependent covariates, and \(\beta\) is the associated regression coefficient. When \(\beta\) is time-varying, denoting \(\beta(t)\), for instance, the proportional hazard assumption is violated and additional special techniques are required in order to obtain statistical inference for model (1.5). One can also refer to the accelerated failure model, which is an important alternative to the PH model.

With the independence assumption of data from different subjects, the likelihood function of observed data turns out to be

\[
L(\theta; T_i, \delta_i, X_i, i = 1, m) = \prod_{i=1}^{m} f(T_i)^{\delta_i} S(T_i)^{1-\delta_i} = \prod_{i=1}^{m} \lambda_i(T_i)^{\delta_i} S(T_i),
\]

where \(\theta\) denote the parameters set of model (1.5), \(i = 1, m\) stands for \(i = 1, 2, \cdots, m\) and \(m\) is the total number of subjects as aforementioned. Thus, the log-likelihood

\[
\ell(\theta; T_i, \delta_i, X_i, i = 1, m) = \sum_{i=1}^{m} \left( \delta_i \left( \log(\lambda_0(T_i)) + X_i^T \beta \right) - \int_{0}^{T_i} \lambda_i(u) du \right),
\]
the maximum likelihood estimation can be obtained for parameters based on the log-likelihood function. Alternatively, partial likelihood is utilised instead,

$$L_p(\theta; T_i, \delta_i, X_i, i = 1, m) = \prod_{i=1}^{m} \left( \frac{\exp(X_i^T \beta)}{\sum_{j \in R(T_i)} \exp(X_j^T \beta)} \right)^{\delta_i},$$

where $R(T_i)$ is the risk set at time $T_i$, that is, $R(T_i) = \{j : T_i \leq T_j\}$. Thus the log partial likelihood function turns out to be

$$\ell_p(\theta; T_i, \delta_i, X_i, i = 1, m) = \sum_{i=1}^{m} \delta_i \left( X_i^T \beta - \log \left( \sum_{j \in R(T_i)} \exp(X_j^T \beta) \right) \right),$$

the maximum estimation based on the partial likelihood function has a huge advantage as it can eliminate the nuisance baseline hazard function $\lambda_0(t)$ when obtaining statistical inference for coefficient parameters $\beta$.

Furthermore, the introduction of counting processes and martingales theory into survival analysis has much extended the applications of Cox model (Andersen et al., 1993). The stochastic process $N_i$ defined by $N_i(t) = I(T_i \leq t)$ is a univariate counting process with intensity function given by

$$\lambda_i(t) = \lambda_0(t) \exp(X_i^T \beta)Y_i(t),$$

where $Y_i(t) = I(t \leq T_i)$ is a predictable process. We have $E(dN_i(t)|F_{t-}) = \lambda_i(t)dt$ and the filtration $F_{t-}$ represents the available data just before time $t$. For more details, one can refer to Fleming and Harrington (1991) and Andersen et al. (1993).

### 1.3 Models for joint analysis of longitudinal and survival data

It is increasingly common that two types of data have been collected in a longitudinal study for each participant, that is, a sequence of longitudinal measurements during the follow-up times and a point process of events in time, together with other covariates. Statistical interest of the data may include but not limited to: 1) the change of the longitudinal process over time and its relationship with other covariates; 2) time-to-event process and its association with other covariates, including the profile level of longitudinal process; 3) statistical inference for both processes (Henderson,
Diggle and Dobson, 2000). On one hand, the repeated measurements of a longitudinal process may not be taken beyond death or drop out of a participant, for instance, and to some extent, the death or drop out may be caused by inadequate level of the longitudinal process. For example, participants with high score of positive and negative syndrome scale tend to drop out in a schizophrenia study. If the scheduled repeated measurements are truncated by the terminated event, erroneous inference with respect to the longitudinal process may be concluded subsequently since the data collected are biased. The event process information is therefore required to account for when analysing data. On the other hand, when the survival analysis of time-to-event data is of interest, it is popular to introduce a Cox regression model to investigate the relationship of time to event with the longitudinal process, that is, the time-dependent covariate. In order to implement the statistical inference, the complete knowledge of trajectory of the longitudinal process is required, which however is usually measured intermittently and often with error (Wulfsohn and Tsiatis, 1997; Tsiatis and Davidian, 2001). The joint modelling of longitudinal and survival data has been therefore motivated. Here, we would like to study the framework of a likelihood-based approach for the joint modelling of longitudinal and survival data, where the longitudinal process is assessed by linear mixed-effects models and the time to event by Cox regression model. Furthermore, only the current value of longitudinal process is accounted for into the Cox regression model, as proposed by several articles in the joint modelling literature (Song, Davidian and Tsiatis, 2002; Tsiatis and Davidian, 2004, among others).

Denote \((T_i, \delta_i, Y_i)\) the observed outcomes of interest of the \(i\)th subject, \(i = 1, 2, \ldots, m\), where \(T_i = \min(\tilde{T}_i, C_i)\), \(\delta_i = I(\tilde{T}_i \leq C_i)\) and \(C_i\) is the censoring time and \(\tilde{T}_i\) is the underlying event time which may or may not be observed. \(Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in})^T\) are the longitudinal measurements taken at time \(t_i = (t_{i1}, t_{i2}, \ldots, t_{in})^T\). The censoring process is assumed to be non-informative and also the time points of repeated measurements taken, that is, the schedule for taking longitudinal measurements may be pre-specified in the study and are independent of the values of the longitudinal process. For the continuous outcomes of longitudinal process, linear mixed-effects models are commonly used and polynomial of times may be covariates as well. More specifically,

\[
Y_i(t) = \alpha_0 + \alpha_{1i}t + \alpha_{2i}t^2 + \cdots + \alpha_{qi}t^q + X_i^T\beta + \varepsilon_i(t) = m_i(t) + \varepsilon_i(t), \tag{1.6}
\]

where \(\alpha_i = (\alpha_{0i}, \alpha_{1i}, \ldots, \alpha_{qi})^T\) are the random effects, \(X_i\) is the covariate vector which
may contain time-dependent covariates as well but they are fully observed, in particular, it is usually referred to baseline covariates in the studies of this thesis. \(\beta\) is the corresponding regression coefficient vector and \(\varepsilon_i(t)\) is the random error at time \(t\). The proportional hazards regression model is proposed for the time to event, by accounting for the longitudinal process through \(m_i(t)\) and other covariates \(Z_i\),

\[
\lambda_i(t) = \lim_{dt \to 0} dt^{-1} Pr\left(t \leq \tilde{T}_i < t + dt | \tilde{T}_i \geq t, Z_i, \bar{m}_i(t), C_i, \bar{Y}_i(t), \bar{\varepsilon}_i(t)\right) \\
= \lim_{dt \to 0} dt^{-1} Pr\left(t \leq \tilde{T}_i < t + dt | \tilde{T}_i \geq t, Z_i, m_i(t)\right) \\
= \lambda_0(t) \exp\left(\gamma m_i(t) + Z_i^T \eta\right),
\]

where \(\bar{m}_i(t) = \{m_i(u) : u \leq t\}\), \(\bar{Y}_i(t) = \{Y_{ij} : t_{ij} \leq t\}\), and \(\bar{\varepsilon}_i(t) = \{\varepsilon_{ij} : t_{ij} \leq t\}\). \(m_i(t)\) sometime is mentioned as the underlying unobservable “true” value of the longitudinal process at time \(t\), which, in practice, are measured intermittently and often with measurement error \(\varepsilon_i(t)\), which can be due to the measuring instrument and/or the biological variation. The likelihood function of joint modelling of the observed data,

\[
L(\theta; T_i, \delta_i, Y_i, X_i, Z_i, i = 1, m) = \prod_{i=1}^{m} \int Pr(T_i, \delta_i | \alpha_i, X_i, Z_i) Pr(Y_i | \alpha_i, X_i) Pr(\alpha_i) d\alpha_i,
\]

where \(Pr(\alpha_i)\) is the probability density function of random effects \(\alpha_i\) and \(\theta\) is the parameters vector of the joint model. Given the unobservable “true” current value \(m_i(t)\) of the longitudinal process, the observed time to event and longitudinal measurements are assumed to be mutually independent. In details, we have

\[
Pr(T_i, \delta_i | \alpha_i, X_i, Z_i) = \lambda_i(T_i) \delta_i S(T_i) \\
= \left(\lambda_0(T_i) \exp(\gamma m_i(T_i) + Z_i^T \eta)\right)^{\delta_i} \exp\left(-\int_0^{T_i} \lambda_i(u) du\right).
\]

Under the normally distributed assumption of measurement error, i.e, \(\varepsilon_i \sim N(0, \Sigma_i)\), given random effect \(\alpha_i\), we have

\[
Pr(Y_i | \alpha_i, X_i) \propto |\Sigma_i|^{-1/2} \exp\left(-\frac{1}{2}(Y_i - A_i \alpha_i - 1_{n_i} X_i^T \beta)^T \Sigma_i^{-1}(Y_i - A_i \alpha_i - 1_{n_i} X_i^T \beta)\right),
\]

where \(A_i\) is the design matrix associated with random effect \(\alpha_i\) and \(1_{n_i}\) is the ones vector of size \(n_i\).
Maximum likelihood estimation is one of the standard approaches to obtain the parameters estimates and therefore statistical inference for the joint modelling, in which the integration with respect to random effects is required. Due to the complexity of the joint model, the integration usually turns out to be quite of challenge. Wulfsohn and Tsiatis (1997) showed that the EM algorithm can be successfully applied to the joint modelling for inference, and the expectation of any function of random effects is evaluated by using an $m$-points Gauss-Hermite quadrature formula (Press et al., 1992). A Monte Carlo approach is also quite straightforward to assist in obtaining the expectation of functions of random effects (Henderson et al. 2000; Hsieh, Tseng, and Wang, 2006). Rizopoulos, Verbeke and Lesaffre (2009) utilised fully exponential Laplace approximation for these integrations with respect to random effects, which also has been proposed in the work of Chapter 4 of this thesis. Alternatively, when the primary interest is in the statistical inference for survival model, such as the estimation of parameters $\gamma$ and $\eta$ of model (1.7), Tsiatis and Davidian (2001) proposed a conditional score estimation approach, in which the random effects were treated as nuisance parameters so as to rule out any distribution assumptions on the random effects, and showed, through heuristic proof and simulation, that their method produces nearly unbiased estimates. In the meantime, the conditional score estimation approach provides a huge computational advantage, as no integration with respect to random effects is required to implement statistical inference. Details will be discussed in chapter 3 of the thesis.

1.4 Covariance estimation and modelling

Longitudinal data collected within subjects are correlated by their very nature, and it inevitably leads to the issue of covariance estimation. The misspecification of the covariance may lead to the lost of efficiency or even bias of the estimation of interest (Daniels and Zhao, 2003). However, it is well known that it is challenging work to obtain appropriate statistical inference for the covariance matrix. For the model (1.1), for example, under the assumption of $\text{Cov}(\alpha_i) = G$, $\text{Cov}(\varepsilon_i) = \Sigma_i$ and the mutual independence between random effects and random error, it has

$$
\text{Cov}(Y_i) = R_i(\theta) = Z_iGZ_i^T + \Sigma_i,
$$
where $\theta$ denotes the parameter set that involves in the covariance matrix $R_i$. By taking $\Sigma_i = \sigma^2_{\varepsilon} I_{n_i}$ and compound symmetry structure for matrix $G = \sigma^2_b (\rho)$, for example, we have $\theta = (\sigma^2_b, \rho, \sigma^2_{\varepsilon})^T$. It is popular to utilise maximum likelihood approach for estimating variance components, which, however, often leads to biased estimates and underestimate the covariance. Patterson and Thompson (1971) proposed a restricted maximum likelihood (REML) approach which takes into account the loss in degrees of freedom resulting from estimating fixed effects and the bias can be reduced. As also showed by Harville (1977), the restricted log-likelihood function turns out to be, up to constant,

$$
\ell_{\text{REML}}(\theta; Y_i, X_i, Z_i, i = 1, m) = -\frac{1}{2} \sum_{i=1}^{m} \log |R_i(\theta)| - \frac{1}{2} \sum_{i=1}^{m} \log |X_i^T R_i^{-1}(\theta) X_i| - \frac{1}{2} \sum_{i=1}^{m} (Y_i - X_i \tilde{\beta}(\theta))^T R_i^{-1}(\theta) (Y_i - X_i \tilde{\beta}(\theta)),
$$

where

$$
\tilde{\beta}(\theta) = \left( \sum_{i=1}^{m} X_i^T R_i^{-1}(\theta) X_i \right)^{-1} \left( \sum_{i=1}^{m} X_i^T R_i^{-1}(\theta) Y_i \right).
$$

Due to the complicated form of the restricted log-likelihood function and/or covariance structure, there is usually no analytical form available for the covariance parameters and therefore some numerical algorithms, such as Newton-Raphson algorithm, are required.

Regarding the selection of covariance structure, AIC or BIC criteria can be used to help with choosing the structures, such as among the options of compound symmetry, AR(1), Toeplitz or unstructured covariances. Often that there may be no covariance of special structure suitable for analysing the data, and an unstructured covariance may have to be proposed. On one hand, the number of parameters increases quadratically with the dimension of covariance matrix, but on the other hand, covariance matrices have to be positive definite. When $p > n$, no one can guarantee such positive definiteness, where $p$ is the dimension of covariance matrix and $n$ is the sample size.

The early work about estimating covariance can be tracked back to Anderson (1973), in which the covariance matrix is proposed to be a linear combination of known symmetric matrices and the linear coefficients are the unknown parameters to help capture the underlying covariance. If a sequence of matrices basis exists and can be defined, such as the concept of basis in the study of B-spline methodology.
(de Boor, 2001), it should be perfectly able to capture the possible covariance for the analysing data. However, any method adapted must ensure the positive definiteness of the covariance matrix. Many others approaches have also been proposed in the literature, such as the matrix-logarithmic covariance model (Chiu, Leonard and Tsui, 1996), and variance-correlation decomposition, spectral or eigenvalue-eigenvector decomposition and Cholesky decomposition. More specifically, assume that a sequence of random variables \( Y = (Y_1, Y_2, \ldots, Y_n)^T \) and that \( \text{Var}(Y) = \Sigma \), it has the variance-correlation decomposition that can provide an excellent and straightforward statistical interpretation

\[
\Sigma = \Lambda^{\frac{1}{2}} R \Lambda^{\frac{1}{2}},
\]

where \( \Lambda \) is diagonal matrix, whose diagonal entries are the variance of each elements of \( Y_i \), that is, \( \text{diag}(\Lambda) = \text{diag}(\Sigma) \), and \( R \) is the correlation matrix, such as demonstrated by the work of Diggle (1990), Diggle and Verbyla (1998) and Huang and Fitzmaurice (2005) among others. However, the correlation matrix \( R \) still has to work with the positive definiteness constraint. On the other hand, the spectral decomposition is given by

\[
\Sigma = A V A^T,
\]

where \( A \) is the matrix consisting of eigenvectors of \( \Sigma \) and \( V \) is diagonal matrix of eigenvalues of \( \Sigma \). The matrix logarithm of \( \Sigma \) is defined as \( \log(\Sigma) = A \log(V) A^T \) which is a symmetric matrix with unconstrained entries (Chiu et al., 1996). However, as aforementioned, it does not have a straightforward statistical interpretation.

In mathematics, a positive definite matrix always has the unique way of Cholesky decomposition and the covariance matrix, which is born to possess the positive definiteness, perfectly belongs to such category, that is, we have

\[
\Sigma = L L^T, \tag{1.8}
\]

where \( L \) is a lower triangular matrix and unique. Pinheiro and Bates (1996) presented three unconstrained parametrization methods based on the Cholesky decomposition (1.8). Meanwhile, a closely related variant of the classical Cholesky decomposition is the modified Cholesky decomposition \( \Sigma = T D T^T \), where \( T \) here is a lower triangular matrix with diagonal elements of being ones and \( D \) is a diagonal matrix. It is obvious that we can rearrange and have \( \Sigma^{-1} = (T^T)^{-1} D^{-1} T^{-1} \). If we re-denote the \( L = T^{-1} \),
then we have the modified Cholesky decomposition

$$\Sigma^{-1} = L^T D^{-1} L,$$  \hspace{1cm} (1.9)

where $L$ is still a lower triangular matrix with ones diagonal elements. Denote

$$L = \begin{pmatrix}
1 & 0 & 0 & \ldots & 0 \\
-\phi_{21} & 1 & 0 & \ldots & 0 \\
-\phi_{31} & -\phi_{32} & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
-\phi_{n1} & -\phi_{n2} & -\phi_{n3} & \ldots & 1
\end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix}
\sigma_1^2 & 0 & 0 & \ldots & 0 \\
0 & \sigma_2^2 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & \sigma_n^2
\end{pmatrix}.
$$

As assumed earlier $\text{Var}(Y) = \Sigma$ and $E(Y) = \mu$, thus we have $E\left(L(Y - \mu)\right) = 0$ and $\text{Var}\left(L(Y - \mu)\right) = L\Sigma L^T = D$. The fact of that $D$ is a diagonal matrix indicates the elements of vector $L(Y - \mu)$ are mutually independent. And the expansion of the term $L(Y - \mu)$ shows

$$L(Y - \mu) = \begin{pmatrix}
1 & 0 & 0 & \ldots & 0 \\
-\phi_{21} & 1 & 0 & \ldots & 0 \\
-\phi_{31} & -\phi_{32} & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
-\phi_{n1} & -\phi_{n2} & -\phi_{n3} & \ldots & 1
\end{pmatrix}
\begin{pmatrix}
Y_1 - \mu_1 \\
Y_2 - \mu_2 - \phi_{21}(Y_1 - \mu_1) \\
Y_3 - \mu_3 - \phi_{31}(Y_1 - \mu_1) - \phi_{32}(Y_2 - \mu_2) \\
\vdots \\
Y_n - \mu_n - \sum_{k=1}^{n-1} \phi_{nk}(Y_k - \mu_k)
\end{pmatrix}
$$

that is, if we denote

$$Y_i - \mu = \sum_{j=1}^{i-1} \phi_{ij}(Y_j - \mu_j) + \epsilon_i \hspace{1cm} (1.10)$$

where, in particular, we denote $Y_1 - \mu_1 = \epsilon_1$, thus $\text{var}(\epsilon) = D$ with $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^T$. Equation (1.10) has a very useful statistical interpretation for the modified Cholesky decomposition, that is, the below-diagonal entries of $L$ are the negatives of autoregression coefficients when $Y_i$ is regressed on its predecessors $Y_1, Y_2, \ldots, Y_{i-1}$ and the diagonal elements of $D$ are the corresponding innovation variances, that is, $\text{var}(\epsilon_i) = \sigma_i^2$. 


In particular, for the data collected from longitudinal studies, measurements of each participant are taken chronologically.

The modified Cholesky decomposition shows that the positive definiteness of matrix $\Sigma$ is guaranteed, meanwhile, the component values of $\phi_{ij}$ and $\log(\sigma_i^2)$ are unbounded. Furthermore, regression models may be proposed for the autoregression coefficients and innovation variances inspired by the statistical interpretation, that is

$$\phi_{ij} = z_{ij}^T \zeta, \quad \log(\sigma_i^2) = h_i^T \xi,$$

where $z_{ij}$ and $h_i$ are covariates vectors associated with $\zeta$ and $\xi$, respectively, which may be proposed as, for example, polynomial functions of time lags and times for $z_{ij}$ and $h_i$, respectively,

$$1, (t_i - t_j)^1, (t_i - t_j)^2, \ldots, (t_i - t_j)^q;$$

$$1, t_i, t_i^2, \ldots, t_i^d,$$

together with some other baseline covariates, such as the treatment factor and/or the interaction with time and/or time lags. The parametrizing approach by proposing regression models for the components $\phi_{ij}$ and $\log(\sigma_i^2)$ can substantially reduce the number of covariance parameters and mitigate the high dimension issue rising in the covariance estimation. Based on this modified Cholesky decomposition and the corresponding regression models, many research work have been well established and demonstrated, such as Pourahmadi (1999, 2000), Smith and Kohn (2002), Daniels and Zhao (2003), Pan and MacKenzie (2003, 2006, 2007), Ye and Pan (2006) among others. Furthermore, Xu and MacKenzie (2012) combine the method of Chiu et al. (1996) and the modified Cholesky decomposition to handle multivariate longitudinal responses.

1.5 Outline of the subsequent chapters

Chapter 2 studies the joint modelling of mixed multivariate longitudinal responses. It is motivated by the large cohort study of UK North Staffordshire Osteoarthritis Project, where triple-type outcomes of longitudinal processes are observed, i.e., continuous-type BMI, binary-type depression index and ordinal-type pain interference. Measurements of trivariate process are taken at baseline, year three and year
six. Random-effects joint models (LMM and GLMM) are proposed to characterize the underlying inherent associations of mixed-type outcomes of longitudinal processes and the correlation of within-subject measurements. The inherent association is characterized by the covariance matrix \( G \) of random effects. The influence of covariance matrix \( G \) on the statistical inference of fixed-effects parameters is investigated following a Bayesian approach. The simulation studies show that by ignoring inherent association between longitudinal processes, that is, the data of each process are analysed separately, may lead to bias and the statistical inference may lack efficiency, particularly in the case of regression coefficients of subject-specific covariates that are not changing over time (e.g. gender). As a result, more reliable estimations are provided by joint modelling the multivariate longitudinal process.

Mixed random-effects models are frequently adapted to capture the changing pattern of a longitudinal process over time. Cox proportional hazards models are also popular in dealing with the point process of events in time in order to investigate the relationship of hazard rate with covariates. Therefore, it may be desirable to jointly model the repeated measurements and survival data both collected in a longitudinal study. Some assumptions are routinely made on joint modelling, for example, a normality assumption on the random effects. It is common to assume that random errors are normally distributed and mutually independent, however, in practice, this is not necessarily always true. In chapter 3, we investigate the impact on statistical inference of survival parameters when some assumption is violated, in particular, the mutual independence of random errors. Due to the complexity of a likelihood-based approach under the joint modelling framework, the conditional score estimator proposed by Tsiatis and Davidian (2001) is utilised to investigate the influence of a violation of the independence assumption. A sufficient statistic is proposed to rule out any distribution assumptions on random effects, making this conditional score approach has a great computational advantage. In Chapter 3, we show that, when the assumption of mutual independence on random errors is violated, this may lead to bias. Apart from dealing with the random effects, the generalised conditional score approach has been proposed to account for any correlation presenting in the random errors, which is characterized by a covariance structure for the random errors. A data-driven method
called the modified Cholesky decomposition is proposed to characterize the underlying covariance structure. Both theoretical results and simulation studies show that, by utilising the generalised conditional score estimator, roughly unbiased estimate for both survival and covariance parameters can be obtained.

In chapter 4, we try to also account for the past trajectory information of longitudinal process when investigating its relationship with the point process of events in time. In the first place of jointly analysing repeated-measurement and time-to-event outcomes, it is typical to model the relation of the longitudinal process with the time-to-event process via the baseline or mean of repeated measurements (e.g., Hu, Tsiatis and Davidian, 1998). Techniques for reducing bias caused by measurement error (Carroll et al., 2006) have been introduced to deal with such topics in survival analysis. Tsiatis, Degruittola and Wulfsohn (1995), among other authors, later identified that the present values of the longitudinal process, which are often time-dependent, would be more appreciate to use in survival analysis models. Nowadays this is the standard framework for jointly modelling the longitudinal and survival data, along with joint models with shared random effects (e.g., Henderson, Diggle and Dobson, 2000). However, in more recent studies, investigators have recognized that both past and recent values of a longitudinal process, blood pressure, for example, may have instructive impact on the present risk of events, for instance, of heart disease or stroke mortality (Boshuizen et al., 2006). In chapter 4, the past information is accounted for by taking an integral over the trajectory of the longitudinal process, which is identified by longitudinal measurements with linear mixed-effects models. A likelihood-based approach is proposed for statistical inference through the EM algorithm. Meanwhile, a fully exponential Laplace approximation approach is proposed to approximate the integrations with respect to the random effects. Statistical inference is provided and demonstrated by real data analysis and simulation studies.

The final chapter summarizes the thesis and discusses some further work. Note notations may be slightly different from chapter to chapter in order to better serve the different topics explored in each chapter, for example, transpose of a matrix $A$ may be denoted by either $A'$ or $A^T$. 
Chapter 2

Bayesian inference for joint modelling of longitudinal continuous, binary and ordinal events

2.1 Introduction

In medical studies, repeated measurements of different types of outcomes may be routinely collected from the same patient. For example, the UK North Staffordshire Osteoarthritis Project (NorStOP), a large cohort study of adults aged 50 years and over who were sampled from the general population in North Staffordshire between 2002 and 2008, involves collection of longitudinal outcomes of continuous, binary and ordinal data from the same patient. Approval for all stages of the study was granted by the UK North Staffordshire Research Ethics Committee. The aim of the project is to study osteoarthritis, the clinical syndrome of joint pain and stiffness, which is the common cause of chronical pain and frequent reason for restricted activity in older people.

The details of the NorStOP survey were reported in Thomas et al.(2007). At baseline, all patients aged 50 years and over who registered with three general practices in North Staffordshire were sent a Health Questionnaire in 2002 ($n = 11,230$). Baseline respondents who gave permission to be re-contacted were mailed similar surveys three
and six years later. The population in our study comprises 2533 people in the NorStOP cohort who completed and returned the Health Questionnaires at all baseline, three and six years. In the study, the trivariate responses consisted of Body Mass Index (BMI) calculated from self-reported height and weight, Depression (Yes/No) ascertained with cut-off score \( \geq 8 \) at the Hospital Anxiety and Depression Scale (HADS) (Zigmond and Snaith, 1983), and Pain Interference (PI) generated from the Medical Outcomes Study (MOS) 36-item short-form health survey (SF-36) with values returned on an ordinal scale 1-5 (Ware and Kosinski, 2009). The latter values reflect responses to the question: “During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?”. The five response categories correspond to “extremely”, “quite a bit”, “moderately”, “a little bit” and “not at all” as having no pain interference. Continuous, binary and ordinal values were recorded at all the three time points. The baseline covariates include age, gender, adequacy of income, alcohol use and perception of health. A single question assessed the perception of adequacy of income: “Thinking about the cost of living as it affects you, which of these descriptions best describes your situation?”). The four response categories were: “find it a strain to get by from week to week”, “have to be careful with money”, “able to manage without much difficulty” and “quite comfortably off”. Participants’ perception of the control they had over their health was hypothesized to have links with disability; this question asked “In your opinion, is it a matter of luck whether you are well or ill, or is it something which can be controlled?”. The five response categories were: “all luck”, “mostly luck”, “bit of both”, “mostly under control” and “almost all under control”.

Explanatory statistical analysis of the NorStOP data was made by Thomas et al. (2004, 2007), and it was concluded there that the persistence of the joint pain that interferes with life increases with age. In this chapter, we investigate the age-related association with the multivariate responses. In the literature, there are many well-established statistical methods to model such longitudinal data separately. Laird and Ware (1982), for example, developed linear mixed models for longitudinal continuous data. Breslow and Clayton (1993) proposed generalized linear mixed models implemented with penalized quasi-likelihood approximation for longitudinal continuous/discrete data. For ordinal categorical outcomes, cumulative link models are
usually utilized (Agresti, 1990), of which random-effects models can be used to handle longitudinal ordinal data. These methods are all subject-specific statistical methods. Alternatively, a population-specific method, generalized estimating equations (GEE) proposed by Liang and Zeger (1986), can be used to model both continuous and discrete longitudinal data.

In addition to the within-subject correlation of such data, it is important to take into account the association among responses of mixed types when modelling such data sets. Failure to do so may cause parameter estimation problems in the models for longitudinal binary and ordinal data. We therefore analyze such responses of mixed types jointly, rather than separately, in order to improve statistical inferences. Examples of modelling multivariate longitudinal data with mixed types were considered by many authors, including Regan and Catalano (2002), Molenberghs and Verbeke (2006), and Fitzmaurice, Davidian, Verbeke and Molenberghs (2008) among others. Catalano and Ryan (1992) and Fitzmaurice and Laird (1995) proposed conditional models for combined continuous and binary outcomes using the GEE method, and Catalano (1997) extended such conditional models to combined continuous and ordinal outcomes. Regan and Catalano (1999a,1999b) proposed a probit-normal model for joint analysis of clustered bivariate outcomes of binary and continuous data. Geys et al. (2001) utilized the Plackett-Dale approach to deal with such bivariate outcomes, which was later extended to ordinal and continuous outcomes by Faes et al. (2004). On the other hand, the generalized linear mixed models are also commonly used in analyzing combined continuous and binary/ordinal outcomes in longitudinal or clustered data, and PROC NLMIXED in SAS may help to calculate the estimates of parameters. Fieuws and Verbeke et al. (2006) proposed pairwise fitting methods for multivariate longitudinal profiles, of which the types of outcomes are usually only binary and/or continuous. Jorgensen et al. (1996) proposed a class of state-space models for multivariate longitudinal data with mixed types where the outcomes may have different distributions.

As stated previously, the trivariate longitudinal mixed responses of the NorStOP cohort are continuous, binary and ordinal in nature, respectively. To our knowledge, there was no article yet that addressed the joint analysis of such trivariate responses or even just combined binary and ordinal outcomes. One may choose to use the pairwise
approach in this case, but it inevitably leads to a loss of efficiency for the estimates of parameters. In addition, our experience in the joint analysis of the binary and ordinal responses in the NorStOP cohort study is that the PROC NLMIXED routines in SAS stop and return error messages due to non-convergence problem. Dunson (2000) provided a general framework for modelling clustered mixed outcomes using Bayesian latent variable models, in which different outcomes share the same random effects, but modelling for longitudinal ordinal data was not considered there. In this chapter, we propose to use joint random effects models to model the trivariate responses motivated by the NorStOP cohort study. Bayesian analysis methods are used to make statistical inferences.

2.2 Statistical Models and Methods

2.2.1 Statistical Models

Motivated by the NorStOP data, below we focus on the trivariate longitudinal outcomes of being continuous ($Y$), binary ($Z$) and ordinal ($T$) in nature respectively. In the study, we assume there are $m$ subjects and $n_i$ repeated measurements for the $i^{th}$ subject. Suppose the ordinal response $T$ has $K$ levels. Denote $C_{ij} = (Y_{ij}, Z_{ij}, T_{ij})^T$ as the mixed outcome measured at the $j^{th}$ time point of subject $i$. To reflect the within-subject correlation and the mutual association between the mixed outcomes, we propose the following joint random effects models,

$$
\begin{align*}
Y_{ij} &= X_{ij}^T \beta_1 + D_{ij}^T b_{1i} + \epsilon_{ij}, \\
Pr(Z_{ij} = 1) &= H_1(X_{2ij}^T \beta_2 + D_{2ij}^T b_{2i}), \\
Pr(T_{ij} \leq k) &= H_2(\alpha_k + X_{3ij}^T \beta_3 + D_{3ij}^T b_{3i}),
\end{align*}
$$

(2.1)

where $j = 1, 2, ..., n_i; i = 1, 2, ..., m; k = 1, ..., K - 1$, and

$$
\epsilon_i = \begin{pmatrix} 
\epsilon_{i1} \\
\vdots \\
\epsilon_{in_i}
\end{pmatrix} \overset{iid}{\sim} N(0, \sigma^2 \epsilon I_{n_i}), \quad b_i = \begin{pmatrix} 
b_{1i} \\
b_{2i} \\
b_{3i}
\end{pmatrix} \overset{iid}{\sim} N(0, G),
$$

$\{Y_{ij}, Z_{ij}, T_{ij}\}$ are the continuous, binary and ordinal responses, respectively, and $(X_{\ell ij}, D_{\ell ij})$, $(\ell = 1, 2, 3)$ are covariate vectors. $\beta = (\beta_1, \beta_2, \beta_3)^T$ are unknown regression coefficients,
and $b_i$s are random effects which are independent with the within-subject random errors $\varepsilon_i$s. $H_1(\cdot)$ and $H_2(\cdot)$ are the inverse of link functions, which in principle can be taken as any monotone increasing functions mapping the unit interval $(0, 1)$ onto $(-\infty, \infty)$ (see, e.g., McCullagh, 1980). In particular, the logistic link function and the probit link function (i.e., the inverse of the cumulative distribution function of $N(0, 1)$) are commonly used. Note for the longitudinal ordinal responses $T_{ij}$ the cumulative probabilities $Pr(T_{ij} \leq k) = \sum_{\nu=1}^{k} Pr(T_{ij} = \nu)$ are usually of interest (Albert and Chib, 1993), which can be modeled by a function of the linear predictor. As the longitudinal responses are ordinal, it is natural to impose order constraints to the cutpoint parameters $\alpha_k$, that is, $\alpha_1 < \alpha_2 < \cdots < \alpha_{K-1}$. A further constraint on either $\alpha_k$ ($k = 1, 2, \ldots, K - 1$) or $\beta_3$ needs to be made, in order to make the model identifiable. Without loss of generality, we assume $\alpha_1 = 0$ and have any intercept parameter included in $\beta_3$.

It is noted that the covariance matrix $G$ of the random effects $b_i = (b_{i1}, b_{i2}, b_{i3})^T$ actually reflects the inherent association between the longitudinal continuous, binary and ordinal outcomes. A very special case is that $G$ is block-diagonal, i.e., $\text{Cov}(b_{i1}, b_{i2}) = 0$, $\text{Cov}(b_{i1}, b_{i3}) = 0$ and $\text{Cov}(b_{i2}, b_{i3}) = 0$, implying that joint modelling for the three different type outcomes is not necessary in this case.

It is well known that the logistic model is highly related with the probit model, due to the fact that $\Phi(\nu) \approx H(c\nu)$ with $c = 15\pi/(16\sqrt{3}) \approx 1.7$, where $\Phi(\cdot)$ and $H(\cdot)$ are the cumulative distribution functions of the standard normal distribution $N(0, 1)$ and the logistic distribution, respectively (Wang et al., 2000). Therefore, the logistic and probit models perform very similarly. The motivated NorStOP project involves the ordinal PI outcomes and so it is natural to consider to use the probit link. On the other hand, for convenience of using MCMC sampling method the functions $H_1(\cdot)$ and $H_2(\cdot)$ are naturally taken to be the cumulative distribution function of $N(0, 1)$, i.e., $H_1(\cdot) = H_2(\cdot) = \Phi(\cdot)$. In order to draw random samples, latent variables are introduced to make statistical inferences and the details will be given later. To avoid negative value of the mean of the normal distribution of the latent variables for ordinal responses $T_{ij}$, we change the sign of regression coefficients for the ordinal outcomes in the model (2.1), as suggested by Albert and Chib (1993). Hence, the models we study
are,
\[
\begin{aligned}
Y_{ij} &= X_{1ij}^T \beta_1 + D_{1ij}^T b_{1i} + \varepsilon_{ij}, \\
Pr(Z_{ij} = 1) &= \Phi(X_{2ij}^T \beta_2 + D_{2ij}^T b_{2i}), \\
Pr(T_{ij} \leq k) &= \Phi(\alpha_k - X_{3ij}^T \beta_3 - D_{3ij}^T b_{3i}),
\end{aligned}
\] (2.2)

where \( j = 1, 2, ..., n_i; i = 1, 2, ..., m; k = 1, ..., K - 1, 0 = \alpha_1 < \alpha_2 < ... < \alpha_{K-1}, \)
\( \varepsilon_{ij} \sim i.d. N(0, \sigma^2_\varepsilon) \) and \( b_i = (b_{1i}, b_{2i}, b_{3i})^T \sim i.d. N(0, G). \)

### 2.2.2 Statistical Methods

Likelihood-based methods may be used for estimation of parameters in the model (2.1) and/or model (2.2), but they are analytically challenging and computationally demanding. For example, the maximum likelihood estimation for the model (2.1) or model (2.2) involves integrating out random effects \( b_i \)s from the joint likelihood function of responses \( C_i \) and random effects \( b_i \). When \( b_i \)s are high-dimensional, the integral problem becomes extremely difficult unless the functions \( H_1(\cdot) \) and \( H_2(\cdot) \) are both identity links, making the maximum likelihood estimates of parameters in the models analytically intractable. To avoid the complicated and high-dimensional integration problem here, we propose to consider Bayesian sampling methods to make statistical inferences for the model (2.1) or model (2.2). By using a Bayesian approach, complete conditional distributions of all parameters are required, which are not straightforward in respect of proposed model (2.2). However, it is not too difficult to develop conditional distributions by using data augmentation approach (Tanner and Wong, 1987), that is, latent variables are introduced with respect to the binary and ordinal response data.

In the spirit of Albert and Chib (1993), we use a latent variable approach for Bayesian analysis of binary and polychotomous response data. Let \( X_{\ell i} = (X_{\ell i1}, ..., X_{\ell im_i})^T \), and \( D_{\ell i} = (D_{\ell i1}, ..., D_{\ell im_i})^T \), \((\ell = 2, 3)\). For the binary responses in the model (2.2), we introduce latent variables \( \lambda_{1i} = (\lambda_{1i1}, ..., \lambda_{1im_i})^T \) and assume \( \lambda_{1i} \)s are independent random variables from \( N(X_{2i}^T \beta_2 + D_{2i}^T b_{2i}, I_{n_i}) \) provided that \( b_{2i} \)s are given, where \( I_{n_i} \) is an \((n_i \times n_i)\) identity matrix. Define \( Z_{ij} = 1 \) if \( \lambda_{1ij} > 0 \) and \( Z_{ij} = 0 \) otherwise. For the ordinal response in model (2.2), we introduce latent variables \( \lambda_{2i} = (\lambda_{2i1}, ..., \lambda_{2im_i})^T \) and assume that \( \lambda_{2i} \)s are independent random variables from \( N(X_{3i}^T \beta_3 + D_{3i}^T b_{3i}, I_{n_i}) \)
provided that \( b_{3i} \)s are given. Define \( T_{ij} = k \) if \( \alpha_{k-1} < \lambda_{2ij} \leq \alpha_k \) (\( k = 1, 2, \ldots, K \)), where \( \alpha_1 = 0 \) and \( -\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_{K-1} < \alpha_K = +\infty \). We treat the latent variables \( \{\lambda_{1i}, \lambda_{2i}\} \) and the random effects \( b_i \) as missing data when using Gibbs sampling, and develop their conditional distributions as well as other distributions.

Regarding prior distributions of parameters, for the convenient use of the conjugate distribution we assume the following prior distributions,

\[
\begin{align*}
\beta_1 &\sim N(0, B_1), \quad \beta_2 \sim N(0, B_2), \quad \beta_3 \sim N(0, B_3) \\
W &= G^{-1} \sim W_q(V, \text{df}), \quad \tau = 1/\sigma^2_e \sim Gamma(a, b) \\
\pi(\alpha_k) &\propto 1 \quad \text{and} \quad 0 < \alpha_2 < \cdots < \alpha_{K-1}
\end{align*}
\]

In other words, the prior distributions of the regression coefficients \( \beta_1, \beta_2 \) and \( \beta_3 \) are multivariate normal distribution with mean 0 and covariance matrices \( B_1, B_2 \) and \( B_3 \), respectively. The covariance matrices \( B_1, B_2 \) and \( B_3 \) can be taken as any positive definite symmetric matrices, but for simplicity they are chosen to be diagonals with a large value on diagonal entries, for example, \( B_1 = B_2 = B_3 = \text{diag}(100, \ldots, 100) \), representing non-informative prior information on the regression coefficients. The inverse covariance matrix of random effects, \( W = G^{-1} \), is distributed as \( W_q(V, \text{df}) \), a Wishart distribution with the degree of freedom \( \text{df} \) and the positive definite scale matrix \( V \).

The prior distribution of \( \tau \), the inverse of the within-subject random error variance \( \sigma^2_e \), is assumed to be a Gamma distribution with shape \( a \) and rate \( b \). Finally, \( \pi(\cdot) \) is a density function and the prior of \( \alpha_k \) is assumed to be an uniform distribution but with constraints \( 0 < \alpha_2 < \cdots < \alpha_{K-1} \). In addition, apriori independence assumption is assumed for the parameters. We then have the joint density function of the responses
\[ C = (Y, Z, T), \text{ random effects } b \text{ and latent variables } \lambda \text{ as} \]
\[
f(Y, Z, T, b_1, \ldots, b_m, \lambda_1, \lambda_2, \beta_1, \beta_2, \beta_3, \alpha_2, \ldots, \alpha_{K-1}, \tau, W)\]
\[
\propto \prod_{i=1}^{m} \left[ \prod_{j=1}^{n_i} (2\pi)^{-1/2} \tau^{1/2} \exp \left\{ -\frac{\tau}{2} \left( y_{ij} - x_{1ij}^T \beta_1 - d_{1ij}^T b_{1i} \right)^2 \right\} \right]
\cdot (2\pi)^{-\frac{p_1}{2}} |B_1|^{-1/2} \exp \left\{ -\frac{1}{2} \beta_{1}^T B_1^{-1} \beta_{1} \right\}
\cdot (2\pi)^{-\frac{p_2}{2}} |B_2|^{-1/2} \exp \left\{ -\frac{1}{2} \beta_{2}^T B_2^{-1} \beta_{2} \right\}
\cdot (2\pi)^{-\frac{p_3}{2}} |B_3|^{-1/2} \exp \left\{ -\frac{1}{2} \beta_{3}^T B_3^{-1} \beta_{3} \right\}
\cdot \frac{b^{a}}{\Gamma(a)} r^{a-1} \exp \{-br\} \cdot \frac{|W|^{-\frac{df-(m+q_2+q_3)}{2}} \exp \left\{ -\frac{1}{2} \text{trace}(V^{-1}W) \right\} \exp \left\{ -\frac{1}{2} \text{trace}(V^{-1}W) \right\}}{2^{\frac{df(q_1+q_2+q_3)}{2}} |V|^{\frac{df}{2}} \Gamma_{q_1+q_2+q_3}(\frac{df}{2})}
\]

where \(1_{\cdot} \) is the indicator function, \( \tau = 1/\sigma^2 \), \( W = G^{-1}, b_i = (b_{1i}, b_{2i}, b_{3i})^T \), and \( b_{1i}, b_{2i} \) and \( b_{3i} \) are \( q_1 \), \( q_2 \) and \( q_3 \)-variant vectors of random effects, respectively. \( \beta_1, \beta_2 \) and \( \beta_3 \) are \( p_1 \), \( p_2 \) and \( p_3 \)-variant vectors of regression coefficients, respectively. We assume that \( \{Y_{ij}, Z_{ij}, T_{ij}\} \) are mutually independent given random effects \( b_i \) \( (i = 1, ..., m; j = 1, ..., n_i) \).

By taking advantage of the conjugate distribution, it is not difficult to construct the posterior distributions of the parameters, random effects and latent variables, see the last section of this chapter for details. We utilize Gibbs sampling method to generate random samples from the posterior distributions and make statistical inferences for the model (2.1) or model (2.2). Note that for the cutpoints \( \alpha = (\alpha_2, \ldots, \alpha_{K-1})^T \), Albert and Chib (1993) suggested to set the conditional distribution of \( \alpha_k \) given \( \{T_{ij}, \lambda_{2ij}, j = 1, ..., n_i, i = 1, ..., m\} \) and \( \{\alpha_i, \ell \neq k\} \) as the uniform distribution on the interval \( (\max\{\min \{\lambda_{2ij} : T_{ij} = k\}, \alpha_{k-1}\}, \min\{\min \{\lambda_{2ij} : T_{ij} = k + 1\}, \alpha_{k+1}\}) \). However, this uniform distribution may lead to a slow convergence because the interval may be too narrow (Cowles, 1996). In this chapter, we propose to use the reparameterizing method of Nandram and Chen (1996) to improve the convergence, that is, we consider the reparameterization \( r^* = 1/\alpha_{K-1} \) and \( \alpha^*_k = r^* \alpha_k, k = 1, \ldots, K - 2 \).
2.2.3 Model selection

We propose to use Spiegelhalter et al.’s (2002) deviance information criterion (DIC) to select models, defined by

\[ DIC = D(\bar{\theta}) + 2p_D = \overline{D(\theta)} + p_D \]

where \( D(\theta) = -2\ell(\theta | \text{Data}) \) and \( \ell(\cdot) \) is the log-likelihood function. \( \bar{\theta} \) is the posterior mean of parameter \( \theta \), and \( p_D = \overline{D(\theta)} - D(\bar{\theta}) \) is the effective number of parameters, where \( \overline{D(\theta)} \) is the posterior mean of the deviance. In our case,

\[ \theta = (\beta_1, \beta_2, \beta_3, \alpha, \sigma^2, G, \lambda_1, \lambda_2, b = \{b_i, i = 1, m\}) \]

The DIC value can be easily obtained from MCMC analysis. Obviously, the smaller the DIC value, the better the model proposed.

2.3 Simulation Study

In this section, we conduct several simulation studies to assess the performance of the proposed estimation approach, and compare the joint and separate modelling strategies. We firstly generate mixed outcomes from model (2.2) by including random intercept and random slope for the continuous outcome but only random intercept for both of the binary and ordinal outcomes. More specifically, we consider the following models in our first simulation study,

\[
\begin{align*}
Y_{ij} &= \beta_{10} + \beta_{11} t_{ij} + \beta_{12} t_{ij}^2 + b_{1i1} + b_{1i2} t_{ij} + \epsilon_{ij} \\
Pr(Z_{ij} = 1) &= \Phi(\beta_{20} + \beta_{21} t_{ij} + \beta_{22} X_{2i} + b_{2i}) \\
Pr(T_{ij} \leq k) &= \Phi(\alpha_k - \beta_{30} - \beta_{31} t_{ij} - \beta_{32} X_{3i} - b_{3i}), k = 1, \ldots, K - 1
\end{align*}
\]  

(2.4)

where \{\(t_{ij}, X_{2i}, X_{3i}\) \(j = 1, \ldots, n_i, i = 1, \ldots, m\)\} are covariates, \(b_i = (b_{1i1}, b_{1i2}, b_{2i}, b_{3i})^T\) are random effects that are multivariate and normally distributed with \(N(0, G)\), and the within-subject random errors \(\epsilon_{ij} \sim \text{iid} N(0, \sigma_\epsilon^2)\). Meanwhile, the random effects and the within-subject random errors are mutually independent.

The models and statistical methods we proposed are able to handle both balanced and unbalanced longitudinal data. For illustration, we generate unbalanced data to study the performance of the proposed models and methods. Firstly, we assume that
the number of repeated measurements $n_i$ is mutually independent and it follows a Poisson distribution, more specifically, $n_i \overset{iid}{\sim} \text{Pois}(5)$ but it is truncated by $3 \leq n_i \leq 10$. The number of subjects is $m = 400$. Regarding the measurement time points $t_i = (t_{i1}, t_{i2}, ..., t_{in})^T$, we randomly sample $n_i$ points from the possible values $\{1, 2, 3, ..., 10\}$. The covariates $X_{2i}$ and $X_{3i}$ are taken from $X_{2i} \overset{iid}{\sim} N(0, 3)$ and $X_{3i} \overset{iid}{\sim} N(0, 2)$. For the binary outcomes, we take $Z_{ij} = 1$ if $\lambda_{1ij} > 0$ and $Z_{ij} = 0$ otherwise where $\lambda_{1ij} \overset{iid}{\sim} N(\beta_20 + \beta_{21} t_{ij} + \beta_{22} X_{2i} + b_{2i}, 1)$ provided that $b_{2i}$s are given. Similarly, the ordinal outcomes take 1-5 scales and in fact $T_{ij} = k$ if $\alpha_{k-1} < \lambda_{2ij} \leq \alpha_k$, ($k = 1, 2, ..., 5$) where $\lambda_{2ij} \overset{iid}{\sim} N(\beta_{30} + \beta_{31} t_{ij} + \beta_{32} X_{3i} + b_{3i}, 1)$ provided that $b_{3i}$s are given. We choose the true values of the parameters as $\beta_{1} = (2, 0.2, 0.4)^T$, $\beta_{2} = (-2, 0.3, 0.6)^T$, $\beta_{3} = (-1, 0.4, 0.5)^T$, $(\alpha_2, \alpha_3, \alpha_4) = (1, 2, 3, 5)^T$ and $\sigma_{\varepsilon}^2 = 2$. The true values of the covariance matrix $G$ are presented in the top panel of Table 2.6, where the random effects have small variances but large correlation coefficients, for example, $\rho_{32} = 0.9$. We call this study as Scenario 1.

Non-informative prior information is specified. We choose $B_i = \text{diag}(100, 100, 100)$, $i = 1, 2, 3$, $df=5$, $V = \text{diag}(5, 5, 5, 5)$ and $a=b=0.1$ as the values of the parameters in the prior distributions. We then generate 100 Monte Carlo data sets and run a chain of length 3000 with burn-in number 1500 for each simulated data set. The simulation results of the fixed effects $\beta$ are presented in Table 2.1, where BaySep. and BayJoint. (or BSep. and BJoint.) refer to the separate modelling and joint modelling, respectively. Also, ‘Mean’ is the posterior mean, ‘SD’ is the standard deviation of the 100 Monte Carlo Means, and ‘MSD’ is the average of the 100 Monte Carlo SDs. We compute the absolute bias of the estimated fixed effects, and find that the biases using both two modelling methods are very small, implying that the two modelling methods yield unbiased estimates of the fixed effects. However, in general the joint modelling strategy leads to a smaller standard deviation than the separate modelling method. This can be measured by efficiency (Ye and Pan, 2006) of fixed effect $\beta_k$, for instance, defined by

$$\text{Efficiency}(\beta_k) = \frac{\text{BSep.var}(\beta_k)}{\text{BJoint.var}(\beta_k)}$$

where $\text{BSep.var}(\beta_k)$ and $\text{BJoint.var}(\beta_k)$ are the posterior variances of $\beta_k$ within the framework of the separated modelling and the joint modelling, respectively. As showed in Table 2.1, all the values of Monte Carlo mean of $\text{Efficiency}(\beta_k)$ are greater than
### Table 2.1: Estimates of Fixed Effect Coefficients—Scenario 1

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\beta_{10}$</th>
<th>$\beta_{11}$</th>
<th>$\beta_{12}$</th>
<th>$\beta_{20}$</th>
<th>$\beta_{21}$</th>
<th>$\beta_{22}$</th>
<th>$\beta_{30}$</th>
<th>$\beta_{31}$</th>
<th>$\beta_{32}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Val.</td>
<td>2</td>
<td>0.2</td>
<td>0.4</td>
<td>-2</td>
<td>0.3</td>
<td>0.6</td>
<td>-1</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>BaySep. Mean</td>
<td>2.0149</td>
<td>0.1921</td>
<td>0.4006</td>
<td>-2.0157</td>
<td>0.3027</td>
<td>0.6089</td>
<td>-0.9960</td>
<td>0.4001</td>
<td>0.5007</td>
</tr>
<tr>
<td>BaySep. SD</td>
<td>0.1050</td>
<td>0.0465</td>
<td>0.0038</td>
<td>0.1254</td>
<td>0.0185</td>
<td>0.0353</td>
<td>0.0686</td>
<td>0.0112</td>
<td>0.0237</td>
</tr>
<tr>
<td>BaySep. MSD</td>
<td>0.1066</td>
<td>0.0475</td>
<td>0.0040</td>
<td>0.1139</td>
<td>0.0165</td>
<td>0.0306</td>
<td>0.0648</td>
<td>0.0096</td>
<td>0.0217</td>
</tr>
<tr>
<td>BayJoint. Mean</td>
<td>2.0163</td>
<td>0.1917</td>
<td>0.4006</td>
<td>-2.0042</td>
<td>0.3009</td>
<td>0.6040</td>
<td>-0.9932</td>
<td>0.3997</td>
<td>0.5015</td>
</tr>
<tr>
<td>BayJoint. SD</td>
<td>0.1040</td>
<td>0.0454</td>
<td>0.0037</td>
<td>0.1288</td>
<td>0.0189</td>
<td>0.0309</td>
<td>0.0686</td>
<td>0.0110</td>
<td>0.0178</td>
</tr>
<tr>
<td>BayJoint. MSD</td>
<td>0.1065</td>
<td>0.0472</td>
<td>0.0039</td>
<td>0.1085</td>
<td>0.0159</td>
<td>0.0257</td>
<td>0.0638</td>
<td>0.0096</td>
<td>0.0170</td>
</tr>
<tr>
<td>abs(BSep.bias)</td>
<td>0.0149</td>
<td>0.0079</td>
<td>0.0006</td>
<td>0.0157</td>
<td>0.0027</td>
<td>0.0089</td>
<td>0.0040</td>
<td>0.0001</td>
<td>0.0007</td>
</tr>
<tr>
<td>abs(BJoint.bias)</td>
<td>0.0163</td>
<td>0.0083</td>
<td>0.0006</td>
<td>0.0042</td>
<td>0.0009</td>
<td>0.0040</td>
<td>0.0068</td>
<td>0.0009</td>
<td>0.0015</td>
</tr>
<tr>
<td>Efficiency($\beta_k$)</td>
<td>1.0038</td>
<td>1.0100</td>
<td>1.0129</td>
<td>1.1095</td>
<td>1.0852</td>
<td>1.4496</td>
<td>1.0321</td>
<td>1.0202</td>
<td>1.6452</td>
</tr>
</tbody>
</table>

### Table 2.2: Estimates of Fixed Effect Coefficients—Scenario 2

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\beta_{10}$</th>
<th>$\beta_{11}$</th>
<th>$\beta_{12}$</th>
<th>$\beta_{20}$</th>
<th>$\beta_{21}$</th>
<th>$\beta_{22}$</th>
<th>$\beta_{23}$</th>
<th>$\beta_{24}$</th>
<th>$\beta_{30}$</th>
<th>$\beta_{31}$</th>
<th>$\beta_{32}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Val.</td>
<td>2</td>
<td>0.2</td>
<td>0.4</td>
<td>-2</td>
<td>0.3</td>
<td>0.6</td>
<td>-0.24</td>
<td>-0.05</td>
<td>-1</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>BaySep. Mean</td>
<td>1.9592</td>
<td>0.2043</td>
<td>0.3990</td>
<td>-2.0857</td>
<td>0.3090</td>
<td>0.6315</td>
<td>-0.2410</td>
<td>-0.0554</td>
<td>-1.0196</td>
<td>0.4012</td>
<td>0.4947</td>
</tr>
<tr>
<td>BaySep. SD</td>
<td>0.2965</td>
<td>0.0903</td>
<td>0.0061</td>
<td>0.2034</td>
<td>0.0216</td>
<td>0.0658</td>
<td>0.0656</td>
<td>0.0482</td>
<td>0.1133</td>
<td>0.0128</td>
<td>0.0412</td>
</tr>
<tr>
<td>BaySep. MSD</td>
<td>0.2627</td>
<td>0.0928</td>
<td>0.0066</td>
<td>0.2055</td>
<td>0.0222</td>
<td>0.0591</td>
<td>0.0660</td>
<td>0.0421</td>
<td>0.1040</td>
<td>0.0117</td>
<td>0.0440</td>
</tr>
<tr>
<td>BayJoint. Mean</td>
<td>1.9645</td>
<td>0.2053</td>
<td>0.3990</td>
<td>-2.0194</td>
<td>0.3018</td>
<td>0.6064</td>
<td>-0.2409</td>
<td>-0.0506</td>
<td>-1.0133</td>
<td>0.3997</td>
<td>0.4969</td>
</tr>
<tr>
<td>BayJoint. SD</td>
<td>0.2779</td>
<td>0.0899</td>
<td>0.0060</td>
<td>0.2002</td>
<td>0.0228</td>
<td>0.0436</td>
<td>0.0413</td>
<td>0.0279</td>
<td>0.1107</td>
<td>0.0126</td>
<td>0.0271</td>
</tr>
<tr>
<td>BayJoint. MSD</td>
<td>0.2611</td>
<td>0.0924</td>
<td>0.0065</td>
<td>0.1768</td>
<td>0.0214</td>
<td>0.0404</td>
<td>0.0383</td>
<td>0.0237</td>
<td>0.0991</td>
<td>0.0116</td>
<td>0.0268</td>
</tr>
<tr>
<td>abs(BSep.bias)</td>
<td>0.0408</td>
<td>0.0043</td>
<td>0.0011</td>
<td>0.0857</td>
<td>0.0090</td>
<td>0.0315</td>
<td>0.0010</td>
<td>0.0054</td>
<td>0.0196</td>
<td>0.0012</td>
<td>0.0053</td>
</tr>
<tr>
<td>abs(BJoint.bias)</td>
<td>0.0355</td>
<td>0.0053</td>
<td>0.0010</td>
<td>0.0194</td>
<td>0.0018</td>
<td>0.0064</td>
<td>0.0009</td>
<td>0.0006</td>
<td>0.0133</td>
<td>0.0003</td>
<td>0.0031</td>
</tr>
<tr>
<td>Efficiency($\beta_k$)</td>
<td>1.0234</td>
<td>1.0232</td>
<td>1.0097</td>
<td>1.3940</td>
<td>1.0907</td>
<td>2.2375</td>
<td>3.0245</td>
<td>3.1897</td>
<td>1.1185</td>
<td>1.0191</td>
<td>2.7232</td>
</tr>
</tbody>
</table>
1.0, especially, 1.4496 and 1.6452 for $\beta_{22}$ and $\beta_{32}$, respectively, implying that the joint modelling strategy is more efficient than the separate modelling method when estimating the fixed effects $\beta$’s.

From Table 2.1, it seems that for the continuous longitudinal outcome, of which the model involves time-dependent covariates, there is no much significant difference between the joint and separate modelling strategies in terms of the estimates of regression coefficients $\beta_{10}, \beta_{11}, \beta_{12}$. For the binary and ordinal longitudinal outcomes, however, there are apparent differences in the estimates of the fixed effects and the associated standard deviations. To understand more about the phenomenon, below we conduct more simulation studies for different model scenarios.

**Scenario 2**: We add two additional covariates in the binary outcome model. In other words, in addition to covariate $X_{2i1} \overset{iid}{\sim} N(0, 3)$ we add $X_{2i2} \overset{iid}{\sim} N(1, 2)$ and $X_{2i3} \overset{iid}{\sim} N(1, 3)$ in the model in Scenario 1. We also take large values for some elements of the covariance matrix of the random effects. See the details in the second panel in Table 2.6.

**Scenario 3**: We reconsider the models in Scenario 2 but this time we only include a random intercept to the models for the continuous and ordinal outcomes. For the binary outcomes, random intercept and slope are included in the model. We also look at some cases of no inherent association between the models, e.g., $G_{31} = 0$.

**Scenario 4**: We investigate how the covariance matrix $G$ influences the efficiencies of the fixed effects. We reconsider the models in Scenario 2 but, for simplicity this time we only consider a random intercept model for all of the three outcomes. We assume that the covariance matrix $G$ is of a structure of compound symmetry (CS) and we consider two cases: (a) the correlation coefficient $\rho$ in CS ranges from -0.45 to 0.85 with a step of 0.1, together with $\rho = 0$ and $\rho = 0.9$, but $\text{diag}(G) = (16, 4, 2.25)^T$ are the variances of the random effects $b_{1i}, b_{2i}$ and $b_{3i}$; and (b) the variances of the random effects are set to be $\text{diag}(G) = \sigma^2_b(4, 1, 0.72)^T$ with $\sigma^2_b$ ranging from 0.09 to 4.6 with a step of 0.3, but the correlation coefficient $\rho$ in CS is fixed at $\rho = 0.25, 0.5$ and 0.75, respectively, representing weak, moderate and strong correlation of the three longitudinal
outcomes, respectively.

The simulation studies in Scenario 2 aim to investigate how the joint models achieve gains in terms of efficiency of estimates of the fixed effects when the mixed types of data are strongly associated and with large variances of random effects. Scenario 3 studies the models with more random effects in the model for the binary data and assesses the performance of the proposed joint modelling approach when certain model components do not have an inherent association. Scenario 4 assesses how the degrees of the correlation and variance of random effects affect the results of efficiency.

Comparing to Scenario 1, Scenario 2 has large elements of the covariance matrix $G$ for the random effects, especially, large diagonal entries. The simulation results of the regression coefficients $\beta$ are presented in Table 2.2, where, similarly, there is no much significant difference between the joint and separate modelling strategies results for continuous outcomes, as we saw in Scenario 1. Table 2.2 shows that, in Scenario 2, the biases of the regression coefficients in the models for discrete outcomes by the separate modelling are larger than these by the joint modelling. Also, the values of Efficiency($\beta_k$) for some regression coefficients are quite large. For example, the values of Efficiency for $\beta_{22}$ and $\beta_{32}$ are greater than 2.0, and those for $\beta_{23}$ and $\beta_{24}$ are even greater than 3.0. These results indicate that the joint modelling analysis produces more efficient estimates of the regression coefficients than the separate modelling method. In addition, our MCMC analysis shows that the standard deviations for the estimated regression coefficients by the joint modelling method are much more stable than these by the separate modelling approach, see Figure 2.1, for example.

It is interesting to see how the proposed joint modelling method works when there is no association between different mixed outcomes. In the simulation studies for Scenario 3, where random intercept and slope are included in the model for the binary responses but only random intercept is included in the model for the continuous and ordinal outcomes, we set some covariance elements to be zero, for example, $G_{31} = 0$, $G_{41} = 0$ and $G_{43} = 0$. As shown in the bottom panel of Table 2.6, the proposed joint modelling approach yields the estimates of $G_{31}$, $G_{41}$ and $G_{43}$ that are not significantly different from zero. Also, the performance of the joint modelling method is very similar to the separate modelling approach in the case of no inherent association between different mixed outcomes, see Table 2.3 for details.
In Scenario 4 we aim to study how the degrees of the inherent association between the three longitudinal outcomes and the heterogeneity of the different outcomes affect statistical inferences of the joint modelling, particularly the estimates and efficiencies of the fixed effects. We reconsider Scenario 2 but for simplicity this time we only include a random intercept for each of the three models. The covariance matrix $G$ is set to be a structure of compound symmetry. The case (a) in Scenario 4 aims to assess the influence of the correlation coefficient $\rho$ in CS on the parameter estimates, while the case (b) studies the effects of variances of random effects on statistical inferences in the joint modelling. For illustration, the efficiencies of some fixed effects corresponding to the cases (a) and (b) are displayed in Figure 2.2(a) and Figure 2.2(b-d), respectively. Note that in case (a) the correlation coefficient $\rho$ in CS must not be smaller than -0.5 because the covariance matrix $G$ is positive definite.

Intuitively, the advantage of the joint modelling over the separate modelling may vanish when the correlation coefficients (in absolute value) between the random effects $b_{1i}$, $b_{2i}$, and $b_{3i}$ decrease toward to zero. This is confirmed in our simulation studies, see
Figure 2.2: The efficiencies of fixed effects against (a) the correlation coefficient $\rho$ in CS, by assuming $\text{diag}(G) = (16, 4, 2.25)^T$; (b) (part of) the variance of random effects when $\rho = 0.75$ in CS, by assuming $\text{diag}(G) = \sigma_b^2(4, 1, 0.72)^T$; (c) when $\rho = 0.5$; (d) when $\rho = 0.25$, where the curves are fitted using nonparametric loess method in R.

Figure 2.2. Figure 2.2(a) shows that the stronger the inherent association between the mixed longitudinal outcomes, the greater the gain of efficiencies of fixed effects by the joint modelling method. On the other hand, Figure 2.2(b), which gives the plot of the efficiencies of fixed effects against the values of $\sigma_b^2$ when $\rho = 0.75$, indicates that the larger the variances of the random effects, the greater the loss of efficiencies of fixed effects by the separate modelling approach. Similar results are also obtained when $\rho = 0.5$ as showed by Figure 2.2(c). When $\rho = 0.25$, representing weak correlation, there may be no that much significant gain by utilising joint analysis as it does with strong correlation, but we can also see that the benefit increases with variation,
Figure 2.2(d) for details. In summary, the joint modelling for the mixed longitudinal outcomes is really necessary if the association and/or heterogeneity of the mixed outcomes is strong.

Note that the simulation studies in Scenarios 1-4 all show that, in general the proposed joint modelling approach through MCMC can estimate the covariance matrix of the random effects $b_i$ quite well. But we also notice that when the dimension of the random effects in the models for discrete outcomes increases, and/or the variances of the random effects and random errors are large, the estimates of some variance components, e.g., $G_{22}$ and $G_{33}$ in Scenario 3, may not be very accurate, see Table 2.6. In this case, in the spirit of restricted maximum likelihood estimation certain penalties may be added to the joint likelihood function in Equation (2.3) in order to reduce the biases of the estimates of the variance components.

In this chapter, the DIC is utilized for the purpose of model selection, in particular, it is used to select models between separated and joint modelling for mixed longitudinal outcomes. As pointed out by Robert and Titterington (2002), however, the DIC may suffer from over fitting for certain models. As an alternative, Bayesian predictive information criterion (BPIC) was proposed by Ando (2007, 2011) for evaluation of hierarchical Bayesian and empirical Bayes models, and it generally improves the DIC-based model selection method in terms of resolving the over fitting problem. However, the BPIC method involves much extra computational efforts as it works on the marginalized log-likelihood function. In our case it is to integrate out random effects from the joint likelihood function of the three mixed longitudinal outcomes and the random effects, which is actually analytically intractable. The calculation of the BPIC is thus very challenging in particular for the models with high-dimensional random effects. Monte Carlo approximation to the integrals may be a solution but largely increases the computational loads. In contrast, the DIC-based method is computationally easy and performs quite well for the purpose of model selection between the joint modelling and separate modelling strategies. To verify this, we carry out a simulation study which uses the same simulation setup as that in Scenario 4. In Table 2.4 we report the number of $DIC_{Joint} \leq DIC_{Sep}$ among 100 simulation runs for a range of the correlation coefficient $\rho$ in CS of the random effects. Table 2.4 shows that, when the inherent association between the mixed longitudinal outcomes is strong and
even moderate, the DIC has a very great chance to select the joint modelling strategy. When the inherent association is weak, e.g., between -0.15 and 0.15, the DIC much prefers to choose the separate modelling. Therefore, the DIC method works well in model selection between the joint modelling and separate modelling strategies.

People may raise the issue of multivariate random effects for each of the three models. We therefore conduct a simulation study by including both random intercept and random slope in each of the three models for continuous, binary and ordinal longitudinal data. For the simulation model setup, we choose the same model components as those in Scenario 2 except for the six-dimensional random effects vector $b_i = (b_{i11}, b_{i12}, b_{i21}, b_{i22}, b_{i31}, b_{i32})^T$, where $b_{i11}$ and $b_{i21}$ are the random intercept and random slope associated with time $t_{ij}$ for the $\ell$th model ($\ell = 1, 2, 3$).

We assume that the covariance matrix $G$ of the random effects $b_i$ is in CS with $\sqrt{\text{diag}(G)} = (4, 1.2, 3, 0.8, 2, 1.5)^T$ and $\rho = 0.75$, and we call this model setup as **Case 1**. Again, the simulation run is 100 and the resulting averaged efficiencies of the regression coefficients are reported in Table 2.5. We further consider **Case 2**, in which all model components are the same as Case 1 except for random intercept only for three response models with $\sqrt{\text{diag}(G)} = (4, 2, 1.5)^T$ and two additional covariates that are included in the model for the continuous outcomes. One covariate, say $X_{i11}$, is a binary variable generated from the Bernoulli distribution $\text{Bin}(1, 0.6)$, and the other covariate $X_{i12}$ is generated from $N(0, 4^2)$. The associated regression coefficients are denoted by $\beta_{13}$ and $\beta_{14}$, respectively, and the resulting efficiencies of the regression coefficients are reported in Table 2.5, see Case 2. We notice that Case 1 in Table 2.5 shows very similar results to Scenario 2, that is, the gain of efficiency for the regression coefficients mainly focuses on the models for longitudinal binary and ordinal outcomes. Case 2 shows a very interesting phenomenon, that is, a strong inherent association between mixed longitudinal outcomes or a large heterogeneity of the different outcomes has substantial influences not only on the efficiencies of fixed effects in the models for binary and ordinal longitudinal outcomes, but also on the fixed effects in the model for continuous outcome such as $\beta_{13}$ and $\beta_{14}$. Note that the corresponding covariates $X_{i11}$ and $X_{i12}$ are time-independent here, that is, cluster-level covariates, and so the conclusion is somewhat different from Scenario 1 because the covariates there are all time-dependent.
### Table 2.3: Simulation output results for scenario 3

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\beta_{10}$</th>
<th>$\beta_{11}$</th>
<th>$\beta_{12}$</th>
<th>$\beta_{20}$</th>
<th>$\beta_{21}$</th>
<th>$\beta_{22}$</th>
<th>$\beta_{23}$</th>
<th>$\beta_{24}$</th>
<th>$\beta_{30}$</th>
<th>$\beta_{31}$</th>
<th>$\beta_{32}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Val.</td>
<td>2</td>
<td>0.2</td>
<td>0.4</td>
<td>-2</td>
<td>0.3</td>
<td>0.6</td>
<td>-0.24</td>
<td>-0.05</td>
<td>-1</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>BaySep. Mean</td>
<td>1.9709</td>
<td>0.2026</td>
<td>0.3996</td>
<td>-1.8764</td>
<td>0.2902</td>
<td>0.5750</td>
<td>-0.2288</td>
<td>-0.0502</td>
<td>-0.9925</td>
<td>0.4012</td>
<td>0.5072</td>
</tr>
<tr>
<td>BaySep. SD</td>
<td>0.3027</td>
<td>0.0732</td>
<td>0.0067</td>
<td>0.3996</td>
<td>0.1295</td>
<td>0.1254</td>
<td>0.1258</td>
<td>0.0864</td>
<td>0.1267</td>
<td>0.0160</td>
<td>0.0437</td>
</tr>
<tr>
<td>BaySep. MSD</td>
<td>0.3027</td>
<td>0.0815</td>
<td>0.0072</td>
<td>0.3693</td>
<td>0.0998</td>
<td>0.1007</td>
<td>0.1168</td>
<td>0.0789</td>
<td>0.1207</td>
<td>0.0137</td>
<td>0.0475</td>
</tr>
<tr>
<td>BayJoint. Mean</td>
<td>1.9716</td>
<td>0.2030</td>
<td>0.3996</td>
<td>-1.8525</td>
<td>0.2861</td>
<td>0.5651</td>
<td>-0.2355</td>
<td>-0.0469</td>
<td>-0.9890</td>
<td>0.4005</td>
<td>0.5054</td>
</tr>
<tr>
<td>BayJoint. SD</td>
<td>0.2983</td>
<td>0.0734</td>
<td>0.0067</td>
<td>0.3701</td>
<td>0.1210</td>
<td>0.1022</td>
<td>0.0905</td>
<td>0.0576</td>
<td>0.1266</td>
<td>0.0159</td>
<td>0.0397</td>
</tr>
<tr>
<td>BayJoint. MSD</td>
<td>0.3013</td>
<td>0.0815</td>
<td>0.0072</td>
<td>0.2986</td>
<td>0.0915</td>
<td>0.0736</td>
<td>0.0818</td>
<td>0.0540</td>
<td>0.1202</td>
<td>0.0136</td>
<td>0.0412</td>
</tr>
<tr>
<td>abs(BSep.bias)</td>
<td>0.0291</td>
<td>0.0026</td>
<td>0.0004</td>
<td>0.1236</td>
<td>0.0098</td>
<td>0.0250</td>
<td>0.0112</td>
<td>0.0002</td>
<td>0.0075</td>
<td>0.0012</td>
<td>0.0072</td>
</tr>
<tr>
<td>abs(BJoint.bias)</td>
<td>0.0284</td>
<td>0.0030</td>
<td>0.0004</td>
<td>0.1475</td>
<td>0.0140</td>
<td>0.0349</td>
<td>0.0045</td>
<td>0.0031</td>
<td>0.0110</td>
<td>0.0005</td>
<td>0.0054</td>
</tr>
<tr>
<td>Efficiency($\beta_k$)</td>
<td>1.0139</td>
<td>1.0010</td>
<td>1.0019</td>
<td>1.6505</td>
<td>1.4072</td>
<td>2.0750</td>
<td>2.1471</td>
<td>2.2604</td>
<td>1.0114</td>
<td>1.0075</td>
<td>1.3384</td>
</tr>
</tbody>
</table>

### Table 2.4: Percentage of DIC$_{Joint} \leq$ DIC$_{Sep}$ when varying $\rho$ in CS of G–Scenario[4]

<table>
<thead>
<tr>
<th>Cor.$\rho$</th>
<th>-0.45</th>
<th>-0.35</th>
<th>-0.25</th>
<th>-0.15</th>
<th>-0.05</th>
<th>0.00</th>
<th>0.05</th>
<th>0.15</th>
<th>0.25</th>
<th>0.35</th>
<th>0.45</th>
<th>0.55</th>
<th>0.65</th>
<th>0.75</th>
<th>0.85</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIC(%)</td>
<td>100</td>
<td>98</td>
<td>62</td>
<td>30</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>19</td>
<td>40</td>
<td>81</td>
<td>96</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

### Table 2.5: Efficiencies of fixed effects in the three models each with random intercept and random slope

<table>
<thead>
<tr>
<th>Efficiency($\beta_k$)</th>
<th>$\beta_{10}$</th>
<th>$\beta_{11}$</th>
<th>$\beta_{12}$</th>
<th>$\beta_{13}$</th>
<th>$\beta_{14}$</th>
<th>$\beta_{20}$</th>
<th>$\beta_{21}$</th>
<th>$\beta_{22}$</th>
<th>$\beta_{23}$</th>
<th>$\beta_{24}$</th>
<th>$\beta_{30}$</th>
<th>$\beta_{31}$</th>
<th>$\beta_{32}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>1.224</td>
<td>1.141</td>
<td>0.995</td>
<td>—</td>
<td>—</td>
<td>2.113</td>
<td>1.410</td>
<td>2.856</td>
<td>2.591</td>
<td>2.616</td>
<td>1.489</td>
<td>1.370</td>
<td>2.015</td>
</tr>
<tr>
<td>Case 2</td>
<td>1.431</td>
<td>1.003</td>
<td>1.002</td>
<td>2.224</td>
<td>2.165</td>
<td>1.218</td>
<td>1.063</td>
<td>1.734</td>
<td>2.026</td>
<td>2.139</td>
<td>1.041</td>
<td>1.022</td>
<td>2.012</td>
</tr>
</tbody>
</table>
### Table 2.6: Covariance Parameters Estimators

#### Variance Estimators from model: random intercept/slope+intercept+intercept–Scenario[1]

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\sigma^2$</th>
<th>$G_{11}$</th>
<th>$G_{21}$</th>
<th>$G_{31}$</th>
<th>$G_{41}$</th>
<th>$G_{12}$</th>
<th>$G_{22}$</th>
<th>$G_{32}$</th>
<th>$G_{42}$</th>
<th>$G_{33}$</th>
<th>$G_{43}$</th>
<th>$G_{44}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Cor.</td>
<td>$\rho$</td>
<td>1.0</td>
<td>0.2</td>
<td>0.2</td>
<td>1.0</td>
<td>0.9</td>
<td>1.0</td>
<td>0.8</td>
<td>1.0</td>
<td>0.8</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>True Val.</td>
<td></td>
<td>2.0000</td>
<td>0.2500</td>
<td>0.0400</td>
<td>0.0800</td>
<td>0.1050</td>
<td>0.1600</td>
<td>0.2880</td>
<td>0.2240</td>
<td>0.6400</td>
<td>0.4480</td>
<td>0.4900</td>
</tr>
<tr>
<td>BaySep.</td>
<td>Mean</td>
<td>1.9911</td>
<td>0.2833</td>
<td>0.0406</td>
<td>1.0</td>
<td>0.1609</td>
<td>1.0</td>
<td>0.2812</td>
<td>1.0</td>
<td>0.6672</td>
<td>1.0</td>
<td>0.4908</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.0691</td>
<td>0.0687</td>
<td>0.0265</td>
<td>0.0149</td>
<td>0.0122</td>
<td>0.0132</td>
<td>0.0110</td>
<td>0.0124</td>
<td>0.1255</td>
<td>0.0593</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MSD</td>
<td>0.0630</td>
<td>0.0974</td>
<td>0.0284</td>
<td>0.0132</td>
<td>0.1122</td>
<td>0.0522</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BayJoint.</td>
<td>Mean</td>
<td>1.9964</td>
<td>0.2575</td>
<td>0.0433</td>
<td>0.1039</td>
<td>0.1049</td>
<td>0.1608</td>
<td>0.2812</td>
<td>0.2227</td>
<td>0.6393</td>
<td>0.4468</td>
<td>0.4784</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.0691</td>
<td>0.0651</td>
<td>0.0255</td>
<td>0.0145</td>
<td>0.0309</td>
<td>0.0247</td>
<td>0.0110</td>
<td>0.0224</td>
<td>0.1101</td>
<td>0.0538</td>
<td>0.0580</td>
</tr>
<tr>
<td></td>
<td>MSD</td>
<td>0.0625</td>
<td>0.0883</td>
<td>0.0274</td>
<td>0.0131</td>
<td>0.0284</td>
<td>0.0204</td>
<td>0.0966</td>
<td>0.0509</td>
<td>0.0505</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Variance Estimators from model: random intercept/slope+intercept+intercept–Scenario[2]

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\sigma^2$</th>
<th>$G_{11}$</th>
<th>$G_{21}$</th>
<th>$G_{31}$</th>
<th>$G_{41}$</th>
<th>$G_{12}$</th>
<th>$G_{22}$</th>
<th>$G_{32}$</th>
<th>$G_{42}$</th>
<th>$G_{33}$</th>
<th>$G_{43}$</th>
<th>$G_{44}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Cor.</td>
<td>$\rho$</td>
<td>1.0</td>
<td>0.6</td>
<td>0.5</td>
<td>0.7</td>
<td>0.1000</td>
<td>0.0094</td>
<td>0.0145</td>
<td>0.0309</td>
<td>0.0247</td>
<td>0.0110</td>
<td>0.0284</td>
</tr>
<tr>
<td>True Val.</td>
<td></td>
<td>4.0000</td>
<td>16.0000</td>
<td>4.0000</td>
<td>4.2000</td>
<td>1.4400</td>
<td>2.1600</td>
<td>1.4400</td>
<td>4.0000</td>
<td>2.4000</td>
<td>2.2500</td>
<td></td>
</tr>
<tr>
<td>BaySep.</td>
<td>Mean</td>
<td>4.0034</td>
<td>16.0035</td>
<td>2.9130</td>
<td>0.4000</td>
<td>0.0400</td>
<td>0.0328</td>
<td>0.0131</td>
<td>0.0284</td>
<td>0.0204</td>
<td>0.0966</td>
<td>0.0509</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.1541</td>
<td>1.4215</td>
<td>0.2710</td>
<td>0.0980</td>
<td>0.0132</td>
<td>0.0110</td>
<td>0.0522</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MSD</td>
<td>0.1516</td>
<td>1.4468</td>
<td>0.3041</td>
<td>0.1103</td>
<td>0.1467</td>
<td>0.0522</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BayJoint.</td>
<td>Mean</td>
<td>4.0064</td>
<td>15.8556</td>
<td>2.9334</td>
<td>3.9944</td>
<td>4.2362</td>
<td>1.4318</td>
<td>2.1769</td>
<td>1.4368</td>
<td>4.0314</td>
<td>2.4267</td>
<td>2.2377</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.1536</td>
<td>1.4129</td>
<td>0.2710</td>
<td>0.5255</td>
<td>0.4265</td>
<td>0.0994</td>
<td>0.2178</td>
<td>0.1168</td>
<td>0.6130</td>
<td>0.2558</td>
<td>0.2011</td>
</tr>
<tr>
<td></td>
<td>MSD</td>
<td>0.1546</td>
<td>1.4355</td>
<td>0.3038</td>
<td>0.5947</td>
<td>0.4276</td>
<td>0.1102</td>
<td>0.1459</td>
<td>0.1264</td>
<td>0.5859</td>
<td>0.2608</td>
<td>0.2052</td>
</tr>
</tbody>
</table>

#### Variance Estimators from model: random intercept+intercept/slope+intercept–Scenario[3]

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\sigma^2$</th>
<th>$G_{11}$</th>
<th>$G_{21}$</th>
<th>$G_{31}$</th>
<th>$G_{41}$</th>
<th>$G_{12}$</th>
<th>$G_{22}$</th>
<th>$G_{32}$</th>
<th>$G_{42}$</th>
<th>$G_{33}$</th>
<th>$G_{43}$</th>
<th>$G_{44}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Cor.</td>
<td>$\rho$</td>
<td>1.0</td>
<td>0.6</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>-0.4</td>
<td>0.6</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>True Val.</td>
<td></td>
<td>4.0000</td>
<td>16.0000</td>
<td>7.2000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>9.0000</td>
<td>-1.8000</td>
<td>2.7000</td>
<td>2.2500</td>
<td>0.0000</td>
<td>2.2500</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.1466</td>
<td>1.3037</td>
<td>0.5255</td>
<td>0.4265</td>
<td>0.0994</td>
<td>0.2178</td>
<td>0.1168</td>
<td>0.6130</td>
<td>0.2558</td>
<td>0.2011</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MSD</td>
<td>0.1567</td>
<td>1.3921</td>
<td>0.6540</td>
<td>0.4494</td>
<td>0.2494</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BayJoint.</td>
<td>Mean</td>
<td>4.0169</td>
<td>15.9553</td>
<td>6.8967</td>
<td>0.0223</td>
<td>0.0328</td>
<td>7.3334</td>
<td>-1.2141</td>
<td>2.5620</td>
<td>1.8235</td>
<td>0.0008</td>
<td>2.2385</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.1475</td>
<td>1.2951</td>
<td>0.4311</td>
<td>0.3939</td>
<td>2.5137</td>
<td>0.7407</td>
<td>0.5855</td>
<td>0.5998</td>
<td>0.1689</td>
<td>0.2652</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MSD</td>
<td>0.1570</td>
<td>1.3783</td>
<td>0.4213</td>
<td>0.3832</td>
<td>1.8235</td>
<td>0.5211</td>
<td>0.5125</td>
<td>0.3986</td>
<td>0.1664</td>
<td>0.2435</td>
<td></td>
</tr>
</tbody>
</table>
2.4 Application

In the NorStOP data, the trivariate responses consist of Body Mass Index (BMI), Depression (Yes/No) and Pain Interference with an ordinal scale from 1 to 5. The covariates include Age, Gender, Alcohol, Income and Time. Note that the NorStOP project is a large cohort study and in this chapter we only consider statistical analysis for completed observations. In total, there are 2313 patients who completed the six-year follow-up study. The patients are aged 50 years and over, and their trivariate responses were collected at baseline, year 3 and year 6, say, Time ∈ \{0, 3, 6\}. We propose to use Age2 = Age + Time - 50 as a time-dependent covariate rather than use Age and Time separately, so as to avoid a potential confounding problem. In addition, Alcohol and Income were collected as covariates of ordinal categories. Our aim is to verify the existence of association between the mixed outcomes together with quantifying the effects of the various covariates, and to compare with the separate modelling method in order to provide evidence of improvements of the joint modelling approach.

In the first instance, random intercept and random slope are both included into the models for the trivariate outcomes. We then apply the proposed joint modelling approach to analyze the NorStOp data and find that the estimate of the variance of the random slope in each model is very small. For example, for the model for continuous response BMI it is about 0.01, which is rather small in comparison to the variance estimate of the random intercept (17.04), so that random slope is excluded from each model. Below we only consider a random intercept model for each of the trivariate responses. We use the DIC value to select the best model, which is given as follows

$$\begin{align*}
\text{BMI}_{ij} &= \beta_{10} + \beta_{11}\text{Age2}_{ij} + \text{Income}_i + b_{1i} + \epsilon_{ij}, \quad (j = 1, 2, 3, \quad i = 1, 2, ..., 2313) \\
\Pr(\text{Depression}_{ij} = 1) &= \Phi(\beta_{20} + \beta_{21}\text{Age2}_{ij} + \text{Alcohol}_i + \text{Income}_i + b_{2i}) \\
\Pr(\text{PI}_{ij} \leq k) &= \Phi(\alpha_k - \beta_{30} - \beta_{31}\text{Age2}_{ij} - \text{Gender}_i - \text{Alcohol}_i - \text{Income}_i - b_{3i})
\end{align*}$$

(2.5)

where \(k = 1, ..., 4\). Note that in the NorStOP data analysis, the covariate Age2 is taken as continuous covariate, while Income, Alcohol and Gender are all treated as categorical variables or factors. Specifically, the factors Income and Alcohol are classified into four and five classes/levels, respectively, representing the cost of living “1-Strain”, “2-Have to be careful”, “3-Able to manage” and “4- Quite comfortable” for Income, and the
Alcohol consumption on the basis of “1-Daily/most days”, “2-Once/twice a week”, “3-Once/twice a month”, “4-Once or twice a year” and “5-Never”. We choose the first level as the reference level. For the factor Gender, female is chosen as the reference level. The random effects $b_i = (b_{i1}, b_{i2}, b_{i3})^T$ are multivariate-normally distributed with $N(0, G)$, and the random errors are $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2_\varepsilon)$. In this real data analysis, we run a MCMC chain of length 6000 with burn-in number 5000 in order to ensure the stability of the posterior mean of the parameters. Statistical results for the NorStOP data are presented in Table 2.7. It is clear that the DIC value (52681) by the separate modelling method is larger than the DIC value (52480) by the joint modelling approach. In fact, for all the models we have analyzed to achieve the best model (2.5), the DIC values by the joint modelling approach are always smaller than that by the separate modelling method, implying that the inherent association between mixed longitudinal outcomes exists.

On the other hand, the estimates of the covariance and correlation matrices of the random effects $b_i = (b_{i1}, b_{i2}, b_{i3})^T$ are given by

$$
\hat{G}_{Joint} = \begin{pmatrix}
17.0429 & 1.1378 & 1.4270 \\
1.1378 & 2.4881 & 1.3406 \\
1.4270 & 1.3406 & 1.7239
\end{pmatrix}, \quad 
\hat{Cor}_{Joint} = \begin{pmatrix}
1.0000 & 0.1747 & 0.2633 \\
0.1747 & 1.0000 & 0.6473 \\
0.2633 & 0.6473 & 1.0000
\end{pmatrix}
$$

respectively. This further shows the evidence of the existence of the inherent association between mixed longitudinal outcomes. For example, the estimate of the correlation coefficient between $b_{i2}$ and $b_{i3}$ is $\hat{\rho}_{32} = 0.6473$. It implies that the joint modelling strategy is indeed necessary in order to make statistical analysis accurate and reliable.

In order to make a comparison, we also separately model the three longitudinal outcomes BMI, Depression and PI. The modelling results together with those by the joint modelling strategy are summarized in Table 2.7. From Table 2.7, for continuous response BMI it seems that there is no significant difference in parameter estimates made by the separate and joint modelling approaches. For the discrete outcomes Depression and Pain Interference, however, there are differences for the estimates of the regression coefficients by the separate and joint modelling strategies. Importantly, the SD of the joint analysis is smaller than that of the separate analysis for Binary and Ordinal outcomes, so that the joint modelling yields more efficient estimates of the regression
<table>
<thead>
<tr>
<th>Continuous Model: BMI</th>
<th>Separate Analysis</th>
<th>Joint Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>28.7691</td>
<td>28.7823</td>
</tr>
<tr>
<td></td>
<td>0.4782</td>
<td>0.4901</td>
</tr>
<tr>
<td></td>
<td>27.8342</td>
<td>27.8345</td>
</tr>
<tr>
<td></td>
<td>29.6966</td>
<td>29.7368</td>
</tr>
<tr>
<td>Age2</td>
<td>-0.0150</td>
<td>-0.0175</td>
</tr>
<tr>
<td></td>
<td>0.0080</td>
<td>0.0080</td>
</tr>
<tr>
<td></td>
<td>-0.0310</td>
<td>-0.0332</td>
</tr>
<tr>
<td></td>
<td>0.0009</td>
<td>-0.0032</td>
</tr>
<tr>
<td>Income2</td>
<td>-1.1360</td>
<td>-1.1227</td>
</tr>
<tr>
<td></td>
<td>0.4907</td>
<td>0.5111</td>
</tr>
<tr>
<td></td>
<td>-2.1154</td>
<td>-2.1199</td>
</tr>
<tr>
<td></td>
<td>-0.1989</td>
<td>-0.1373</td>
</tr>
<tr>
<td>Income3</td>
<td>-1.9265</td>
<td>-1.9125</td>
</tr>
<tr>
<td></td>
<td>0.4961</td>
<td>0.4951</td>
</tr>
<tr>
<td></td>
<td>-2.9054</td>
<td>-2.8706</td>
</tr>
<tr>
<td></td>
<td>-0.9459</td>
<td>-0.9508</td>
</tr>
<tr>
<td>Income4</td>
<td>-2.3179</td>
<td>-2.2743</td>
</tr>
<tr>
<td></td>
<td>0.5097</td>
<td>0.5284</td>
</tr>
<tr>
<td></td>
<td>-3.3259</td>
<td>-3.3225</td>
</tr>
<tr>
<td></td>
<td>-1.3438</td>
<td>-1.2453</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Binary Model: Depression</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Age2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Alcohol2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Alcohol3</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Alcohol4</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Alcohol5</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Income2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Income3</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Income4</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ordinal Model: PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Age2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>GenderMale</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Alcohol2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Alcohol3</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Alcohol4</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Alcohol5</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Income2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Income3</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Income4</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
coefficients in the models for binary outcome Depression and ordinal outcome Pain Interference. Note for the threshold parameters \( \alpha \), we used the Metropolis-Hastings algorithm and the resulting estimates are \( \hat{\alpha} = (\hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4)^T = (1.1729, 2.0865, 3.6151)^T \) in the joint models.

From Table 2.7, Age2 variable seems to be insignificant in the separate approach, but significant in the joint approach, so potentially there could be issues of underfitting the model due to inflation of standard errors or bias in estimates obtained with the separate approach. Also it seems that there is no much difference in efficiency between the two approaches, which may be due to weak correlation as measurements are only taken at baseline, year three and year six, or may also be due to the large sample size. As further demonstrated by Table 2.7, when people are aged 50 years and over their BMI values tend to decrease slightly with age, but joint pain may become more severe as their age increases. In addition, it seems that “Age” does not play a very important role in modelling for binary outcomes Depression. Interestingly, people with higher incomes seem to have a lower BMI value and a reduced probability of being depressed. They are also less likely to suffer from severe pain interference. People who drink alcohol once/twice a week have less probability of being depression when comparing to people who drink daily/most days, once/twice a month, once/twice a year, or never drink. Furthermore, comparing to the people who drink daily/most days, people who never drink have less probability of suffering from severe joint pain. Finally, female seems to have a higher risk of suffering from pain interference than male.

### 2.5 Further discussion

There are already many well-established statistical methods for longitudinal data, especially for one type of outcomes. Multivariate outcomes of mixed types, however, such as continuous, ordinal and binary longitudinal data, are collected routinely from the same subject. These mixed-type longitudinal outcomes are associated mutually in nature. Separate analysis of the mixed-type longitudinal outcomes is possible, but the estimates of parameters in the models for discrete longitudinal data may be very biased and inefficient, in particular, when the inherent association of the mixed-type outcomes is strong. In this chapter, we proposed a joint random-effect model to analyze
CHAPTER 2. THREE-EVENT

the mixed-typed longitudinal outcomes, simultaneously. Compared to the separate modelling strategy, the joint modelling approach leads to very efficient estimates of parameters in all the models. The proposed estimation method is implemented using MCMC resampling methods and it works very well in our simulation studies and real data analysis.

The key point in the chapter is that inherent association of mixed longitudinal outcomes can be characterized by covariance matrix $G$ of random effects $b_i$. This idea can be applied to joint modelling of several mixed-type longitudinal outcomes. When the dimension of the random effects is high, however, it may be quite challenging to estimate the covariance matrix $G$ because $G$ has a constraint of positive definiteness. In the meantime, there are a number of parameters in $G$ to estimate. For example, if random intercept and slope are both included in the models for the three responses considered in this chapter, there are 21 parameters in $G(6 \times 6)$ to estimate. Clearly, the number of parameters increases quadratically with the dimension of the random effects. A possible solution to this issue is to use latent variables method, for instance, proposed by Fitzmaurice et al. (2008) among others. Alternatively, matrix-logarithm (Chiu, Leonard and Tsui, 1996) or modified Cholesky decomposition based covariance modelling approaches may be applied to the estimation of the covariance matrix $G$ (Pourahmadi, 1999, 2000 and Pan and MacKenzie, 2003, among others).

In this chapter, the DIC is used to select either joint modelling or separate modelling strategy for the three mixed-type longitudinal outcomes. In the context of variable selection, it is likely to suffer from overfitting problem for certain models. As an alternative, BPIC was proposed by Ando (2007, 2011) to resolve the overfitting issue from predictive point of view. However, the calculation of the BPIC score is not an easy task for certain models as it involves the use of the marginal log-likelihood function. In our cases here we have to integrate out random effects from the joint likelihood function of the three longitudinal responses and random effects, which in general is analytically intractable. In our study, the DIC suggests to choose the separate modelling strategy when the inherent association is close to zero, and select the joint modelling method when the inherent association is moderate and large. We therefore prefer to use the DIC as a criterion of model selection here.

By utilizing Bayesian approach, there is no need to integrate out random effects
but as a compromise its computation may be intensive especially when the sample size is very large. For example, in our analysis for the NorStOP data there are 2313 subjects with three repeated measurements each. It takes about six hours to have a MCMC chain with length of 11,000 converges.

Finally, we would like to mention some open issues in the joint modelling framework, for example, covariance modelling of the covariance matrix $G$ and variable selection in the joint models of continuous, binary and ordinal longitudinal outcomes. On the other hand, variable selection with respect to random effects in the joint models is also an open issue and deserves a further investigation. Regarding the NorStOP data, the population probabilities of Depression and/or Pain Interference of each category may also be of interest. Finally, we think that marginal generalised linear models (Liang and Zeger, 1986) may also be useful in this context of joint models.
2.6 Chapter appendices

Posterior distributions of parameters

According to the joint density function (3) and the prior distributions assumed, we have

\[
\beta_1 \mid Y, Z, T, X_1, X_2, X_3, b_1, \ldots, b_m, \lambda_1, \lambda_2, \beta_2, \beta_3, \alpha_2, \ldots, \alpha_{K-1}, \tau, W
\]

\[
\propto \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - x_{1ij}^T \beta_1 - d_{1ij}^T b_{1j})^2 - \frac{1}{2} \beta_1^T B_1^{-1} \beta_1 \right\}
\]

\[
= \exp \left\{ -\frac{\tau}{2} (Y - X_1 \beta_1 - D_1 b_1)^T (Y - X_1 \beta_1 - D_1 b_1) - \frac{1}{2} \beta_1^T B_1^{-1} \beta_1 \right\}
\]

\[
\propto \exp \left\{ -\frac{1}{2} \left[ \beta_1^T (\tau X_1^T X_1 + B_1^{-1}) \beta_1 - 2 \tau \beta_1^T X_1^T (Y - D_1 b_1) \right] \right\}
\]

Therefore, we obtain the posterior distribution of

\[
\beta_1 \mid \text{data}, \Theta^{-\beta_1} \sim N \left( (\tau X_1^T X_1 + B_1^{-1})^{-1} \tau X_1^T (Y - D_1 b_1), \ (\tau X_1^T X_1 + B_1^{-1})^{-1} \right)
\]

where \(\Theta^{-\beta_1}\) stands for all parameters involved but \(\beta_1\), and \(\text{data}\) include observed and unobserved data,

\[
Y = \begin{pmatrix}
y_{11} \\
y_{12} \\
\vdots \\
y_{1n_1} \\
y_{21} \\
\vdots \\
y_{m1} \\
y_{m2} \\
\vdots \\
y_{mn_m}
\end{pmatrix}, \quad X_1 = \begin{pmatrix}
x_{111} & x_{112} & \cdots & x_{11p_1} \\
x_{121} & x_{122} & \cdots & x_{12p_1} \\
\vdots & \vdots & \ddots & \vdots \\
x_{1n_11} & x_{1n_12} & \cdots & x_{1n_1p_1} \\
x_{211} & x_{212} & \cdots & x_{21p_1} \\
\vdots & \vdots & \ddots & \vdots \\
x_{m11} & x_{m12} & \cdots & x_{m1p_1} \\
x_{m21} & x_{m22} & \cdots & x_{m2p_1} \\
\vdots & \vdots & \ddots & \vdots \\
x_{mn_11} & x_{mn_12} & \cdots & x_{mn_1p_1}
\end{pmatrix}, \quad b_1 = \begin{pmatrix}
b_{111} \\
b_{112} \\
\vdots \\
b_{1n_11} \\
b_{121} \\
\vdots \\
b_{1m1} \\
b_{1m2} \\
\vdots \\
b_{1mq_1}
\end{pmatrix},
\]

\[
\beta_1 = \begin{pmatrix}
\beta_{11} \\
\beta_{12} \\
\vdots \\
\beta_{1p_1}
\end{pmatrix}, \quad D_1 = \begin{pmatrix}
d_{111} & d_{112} & \cdots & d_{11q_1} \\
d_{121} & d_{122} & \cdots & d_{12q_1} \\
\vdots & \vdots & \ddots & \vdots \\
d_{1n_11} & d_{1n_12} & \cdots & d_{1n_1q_1}
\end{pmatrix}, \quad \text{and } D_1 = \text{diag}(D_{11}, D_{12}, \ldots, D_{1m})
\]
Similarly, we can get the posterior distributions for $\beta_2$ and $\beta_3$,

$$
\beta_2 \mid data, \Theta^{-\beta_2} \sim N \left( (X_T^T X_2 + B_2^{-1})^{-1} X_2^T (\lambda_1 - D_2 b_2), \left( X_2^T X_2 + B_2^{-1} \right)^{-1} \right)
$$

$$
\beta_3 \mid data, \Theta^{-\beta_3} \sim N \left( (X_3^T X_3 + B_3^{-1})^{-1} X_3^T (\lambda_2 - D_3 b_3), \left( X_3^T X_3 + B_3^{-1} \right)^{-1} \right)
$$

Here $\lambda_k = (\lambda_{k1}, \lambda_{k2}, \cdots, \lambda_{km})^T$, $\lambda_{ki} = (\lambda_{k1i}, \lambda_{k2i}, \lambda_{kmi})^T$, $k=1,2$ are the latent variables introduced. The other notations, such as $X_2$, $D_2$, $b_2$, have similar definitions with that mentioned above.

Regarding $\tau$, conditioned on the data and other parameters, we have

$$
\tau \mid Y, Z, T, X_1, X_2, X_3, b_1, \ldots, b_m, \lambda_1, \lambda_2, \beta_1, \beta_2, \beta_3, \alpha_2, \ldots, \alpha_{K-1}, W
$$

$$
\propto \tau^{\frac{1}{2} \sum_{i=1}^m \alpha_i} e^{-br} e^{-\frac{1}{2} (Y - X_1 \beta_1 - D_1 b_1)^T (Y - X_1 \beta_1 - D_1 b_1)}
$$

Therefore, the conditional posterior distribution of $\tau$ is Gamma distribution and

$$
\tau \mid data, \Theta^{-\tau} \sim \text{Gamma} \left( \frac{1}{2} \sum_{i=1}^m \alpha_i + a, b + \frac{1}{2} (Y - X_1 \beta_1 - D_1 b_1)^T (Y - X_1 \beta_1 - D_1 b_1) \right)
$$

Regarding $W$, we have

$$
W \mid Y, Z, T, X_1, X_2, X_3, b_1, \ldots, b_m, \lambda_1, \lambda_2, \beta_1, \beta_2, \beta_3, \alpha_2, \ldots, \alpha_{K-1}, \tau
$$

$$
\propto \left| W \right|^{\frac{m}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^m b_i^T W b_i \right\} \left| W \right|^{\frac{d}{2} - (q_1 + q_2 + q_3)} \exp \left\{ -\frac{1}{2} \text{trace} (V^{-1} W) \right\}
$$

$$
= \left| W \right|^{\frac{m + df - (q_1 + q_2 + q_3)}{2}} \exp \left\{ -\frac{1}{2} \text{trace} \left( \sum_{i=1}^m b_i^T W \right) \right\}
$$

Therefore, the conditional posterior distribution of $W$ is Wishart distribution and we have

$$
W \mid data, \Theta^{-W} \sim W_{q_1 + q_2 + q_3} \left( \sum_{i=1}^m b_i^T W \right)^{-1}, m + df
$$

That is, $W$ is a $(q_1 + q_2 + q_3) \times (q_1 + q_2 + q_3)$ symmetric matrix of random variables, and the conditional posterior distribution is Wishart distribution with $m + df$ as degree of freedom and $\left( \sum_{i=1}^m b_i^T W \right)^{-1}$ as scale matrix. Here, we define $b_i = (b_{1i1}, \ldots, b_{1iq_1}, b_{2i1}, \ldots, b_{2iq_2}, b_{3i1}, \ldots, b_{3iq_3})^T$. 


Regarding the random variable $b_i$, we have

$$b_i \mid Y, Z, T, X_1, X_2, X_3, \lambda_1, \lambda_2, \beta_1, \beta_2, \beta_3, \alpha_2, \ldots, \alpha_K, \tau, W$$

$$\propto \exp\left\{-\frac{\tau}{2}(y_i - X_{1i}\beta_1 - D_{1i}b_{1i})^T(y_i - X_{1i}\beta_1 - D_{1i}b_{1i})\right\} \cdot \exp\left\{-\frac{1}{2}(\lambda_{1i} - X_{2i}\beta_2 - D_{2i}b_{2i})^T(\lambda_{1i} - X_{2i}\beta_2 - D_{2i}b_{2i})\right\} \cdot \exp\left\{-\frac{1}{2}(\lambda_{2i} - X_{3i}\beta_3 - D_{3i}b_{3i})^T(\lambda_{2i} - X_{3i}\beta_3 - D_{3i}b_{3i})\right\} \cdot \exp\left\{-\frac{1}{2}b_i^T W b_i\right\}$$

Therefore, we obtain

$$b_i \mid data^{-b_i}, \Theta \sim N\left((M_i + W)^{-1}E_i, (M_i + W)^{-1}\right)$$

Where

$$M_i = \begin{pmatrix} \tau D_{1i}^T D_{1i} & 0 & 0 \\ 0 & D_{2i}^T D_{2i} & 0 \\ 0 & 0 & D_{3i}^T D_{3i} \end{pmatrix}, \quad E_i = \begin{pmatrix} \tau D_{1i}^T (y_i - X_{1i}\beta_1) \\ D_{2i}^T (\lambda_{1i} - X_{2i}\beta_2) \\ D_{3i}^T (\lambda_{2i} - X_{3i}\beta_3) \end{pmatrix}$$

The conditional posterior distribution of $\lambda_1$ is

$$\lambda_1 \mid data^{-\lambda_1}, \Theta \sim N\left(X_2\beta_2 + D_2 b_2, I_N\right)$$

where $\lambda_{1ij} > 0$ if $Z_{ij} = 1$ otherwise $\lambda_{1ij} \leq 0$, $I_N$ is $N \times N$ identity matrix and $N = \sum_{i=1}^m n_i$

The conditional posterior distribution of $\lambda_2$ is

$$\lambda_2 \mid data^{-\lambda_2}, \Theta \sim N\left(X_3\beta_3 + D_3 b_3, I_N\right)$$

where $\alpha_{k-1} < \lambda_{2ij} \leq \alpha_k$ if $T_{ij} = k$, $k = 1, 2, \ldots, K$ and $-\infty = \alpha_0 < \alpha_1 < \ldots < \alpha_{K-1} < \alpha_K = \infty$. 
Chapter 3

Survival analysis with longitudinal covariates measured with correlated error

3.1 Introduction

It has become increasingly common in clinical research to collect information both on longitudinal measurements on a continuous response and a possibly censored time to event for each participant, along with additional covariates information. AIDS Clinical Trials Group (ACTG) Study 175, for example, in addition to the baseline covariates information, CD4 counts were collected on each participant per protocol about every 12 weeks after randomization and the number of days until the first occurrence of a \( \geq 50 \) percent decline in the CD4 cell count, development of the AIDS, or death were collected as well. It is often of interest to capture the trajectory of the longitudinal process and/or to characterize the relationship between the survival and the longitudinal process and other covariates. One popular framework is to assume that the longitudinal process follows a linear mixed effects model (Laird and Ware, 1982) and that the survival depends on covariates through a proportional hazards model (Cox, 1972).

Longitudinal process usually varies with time, which makes the time-dependent
covariate when modelling survival data. To implement the Cox model with time-
dependent covariates, complete knowledge of the true covariates history for each sub-
ject is required. Time-dependent covariates, however, are generally measured inter-
mittently and often with error, which much complicates the survival modelling since it
is well known that it leads to biased estimation if substituting mismeasured values for
true covariates in the Cox model (Prentice, 1982). The naive approach, for instance,
which may simply impute the unknown value by its ordinary least square estimate,
tends to introduce biased inference as well. Some research methods have been advan-
ced to reduce the bias with respect to the inference for failure time parameter, which
is due to the measurement error and/or to the fact that longitudinal measurements
are available and collected intermittently only. Regression calibration, whose basis is
the replacement of unknown covariates by their regression on available observed data
(Carroll et al., 2006, Ch.4), for example, is used to reduce the bias of the naive ap-
proach but may still yield biased estimators and therefore give erroneous results for the
large measurement error or when the relative risk of interest is large (Wang, Wang, and
Wang., 2000; Tsiatis and Davidian, 2001). A corrected score approach (Nakamura,
1992; Huang and Wang 2000; Song and Huang, 2005; Wang, 2006) has been proposed
to reduce the bias as well, which mainly deals with the cases when the covariates of
interest are time-independent, such as random intercept and/or slope, although it usual-
ly leads to consistent and robust estimators. Likelihood-based approaches (Faucett
and Thomas, 1996; Wulfsohn and Tsiatis, 1997; Henderson, Diggle, and Dobson, 2000;
Rizopoulos, 2011) are consistent and efficient with normality assumption on both the
underlying random effects and the measurement error. The semi- or nonparametric
likelihood approaches release distribution assumptions on random effects, instead they
use seminonparametric or nonparametric approach to characterize the density func-
tion of random effects (Hu, Tsiatis, and Davidian, 1998; Song, Davidian, and Tsiatis,
2002b). Due to the requirement of integrating out random effects, likelihood-based
approaches can be very computationally intensive. Alternatively, motivated by the
conditional score estimation of Stefanski and Carroll (1987) for the generalized linear
model, which proposed sufficient statistic with respect to unknown covariates and con-
ditional estimating equations for the parameter inference, Tsiatis and Davidian (2001)
proposed the conditional score estimator with respect to the Cox regression parameter
to rule out any assumptions on the underlying random effects. The method is easy to implement and the estimates have been shown to be consistent and robust (Song and Huang, 2005; Wang, 2006).

A normality assumption is commonly made on unobservable random effects and measurement error when jointly modelling longitudinal and survival data. Some literature has pointed out that we may still produce robust and consistent estimates with distributional departure of random effects from normality (Song et al., 2002b, among others). In addition, the majority of work in the literature tends to take it for granted that the measurement errors (i.e. within-subject random errors) are normally distributed and mutually independent. It may be reasonable with respect to the distributional assumption of normality, however, instead of being independent, the measurement error may still correlate to some extent and the proposed random effects may not precisely account for all the inherent association. Many authors have noted that misspecification of the covariance structure could lead to erroneous statistical inference, similar to other examples of assumption violations (Wang, Carroll, and Liang, 1996; Henderson et al., 2000; Daniels and Zhao, 2003; Tsiatis and Davidian, 2004). In this chapter, we investigate how a violating assumption of mutual independence of within-subject errors could influence estimation, by utilising the conditional score approach proposed by Tsiatis and Davidian (2001). The approach is plausible in terms of the inference implementation and it generates consistent and robust estimators for the proportional hazards model as discussed previously. In this study, our primary interest is to characterize the relationship between survival and the covariates, which includes the covariate consisting of current/present value of the relative time-dependent longitudinal process and other covariates which may or may not be time-dependent.

Note, in this chapter, in order to try to make formulas look slightly tidier, instead of notation $A^T$, we use $A'$ to denote the transpose of a matrix $A$. The “true” values of relative longitudinal process are denoted by $X_i(t)$ and the corresponding measurements by $W_i(t)$ to emphasise that, in this chapter, the current/present value of longitudinal process is treated as one of important covariates of the survival regression model.
3.2 Model Definition

For simplicity, we assume a single longitudinal process, but generalisation to multiple cases should be feasible and without much difficulties (Song, Davidian, and Tsiatis, 2002a). Together with some ordinary covariates denoted by $Z_i$, which can be either time-dependent or time-independent but true values are collected exactly, measurements for the longitudinally observed process are taken at different time points $t_{ij}$ with $j = 1, 2, \ldots, n_i$ and $i = 1, 2, \ldots, m$ per protocol until the event or censor happens for each subject $i$, that is, $t_{ij} \leq V_i$, where $V_i$ denotes the observed time-to-event. $W_i(t_{ij})$, hereafter abbreviated as $W_{ij}$, is the $j$th measurement for subject $i$ at time $t_{ij}$, and thus $W_i = (W_{i1}, W_{i2}, \ldots, W_{in_i})'$ denote all the available longitudinal observations for each subject, which are measured intermittently and often with error of the of interest longitudinal process $X_i(t)$. The longitudinal data are assumed to follow a linear mixed effects model with time polynomial function, but the generalisation to that with additional baseline covariates is straightforward, that is, denote $X_i(t) = \alpha_0i + \alpha_1i t + \cdots + \alpha_pit^p$, then we have $W_i(t) = X_i(t) + \varepsilon_i(t)$, where the error term $\varepsilon_i(t)$ can be time-dependent or not. Denote $\varepsilon_{ij}$ the deviation of the measurement from $X_i(t)$ at time $t_{ij}$, that is, $W_{ij} = X_{ij} + \varepsilon_{ij}$. Given $t_i = (t_{i1}, \ldots, t_{in_i})'$, the within-subject error $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in_i})'$ are assumed to be normally distributed with mean zero and covariance $\Sigma_i$ and they are independent of random effects $\alpha_i = (\alpha_{0i}, \alpha_{1i}, \ldots, \alpha_{pi})'$. More precisely, we have

$$(\varepsilon_i|T_i, C_i, \alpha_i, Z_i, t_i, n_i) \sim N_{n_i}(0, \Sigma_i),$$

(3.1)

where $N_{n_i}$ denotes the $n_i$-variate normal distribution. On the other hand, let $T_i$ and $C_i$ denote the survival and censor time, respectively, that is, we have $V_i = \min(T_i, C_i)$, the indicator function $\delta_i = I(T_i \leq C_i)$ indicates the occurrence of event or censoring. We assume the hazard of failure is related to covariates $X_i(t)$ and $Z_i$ through a proportional hazards regression model, that is,

$$\lambda_i(t) = \lim_{dt \to 0} dt^{-1} Pr(t \leq T_i < t + dt|T_i \geq t, \alpha_i, Z_i, C_i, t_i(t), \varepsilon_i(t)) = \lambda_0(t) \exp (\gamma X_i(t) + \eta'Z_i),$$

(3.2)

where $\lambda_0(t)$ denotes an unspecified baseline hazard function, and $\eta$ is the unknown $q \times 1$ vector of regression coefficients of covariates $Z_i$. The collection of longitudinal
measurement times up to and including time $t$ is denoted by $\bar{t}_i(t) = \{t_{ij} : t_{ij} \leq t\}$ and similarly $\bar{\varepsilon}_i(t) = \{\varepsilon_{ij} : t_{ij} \leq t\}$. The proposed Cox model indicates that the censoring is noninformative, and given random effects $\alpha_i$, failure time $T_i$ is independent of the observed longitudinal measurements.

Regarding the statistical inference for linear mixed effect models, different potential structures for within-subject covariance $\Sigma_i$ have been considered by many researchers and are even applicable in some standard statistics softwares, such as R and SAS. However, when studying the survival regression modelling with longitudinal covariates, most of the literatures routinely assume that $\varepsilon_i$s are mutually independent and the variance is constant, that is, $\varepsilon_{ij} \sim \text{iid } N(0, \sigma^2)$, which, however, is not necessary to be true all the time. In this chapter, we would like to investigate the influence of violating the assumption of mutual independence of $\varepsilon_i$ on the inference for survival parameters and no distribution assumption is made on random effects $\alpha_i$s, as discussed by Tsiatis and Davidian (2001).

### 3.3 Conditional Score Estimator

Tsiatis and Davidian (2001) proposed a conditional score (CS) estimator in respect of unknown regression parameters in the proportional hazards model when the underlying time-dependent covariate follows a linear mixed model. More specifically, let $\hat{X}_i(t)_{ols}$ denote the ordinary least squares (OLS) estimator of $X_i(t)$ using all the $i^{th}$ longitudinal data up to and including time $t$, that is, $\hat{X}_i(t)_{ols} = \bar{\beta}(A_{i,t}'A_{i,t})^{-1}A_{i,t}'W_{i,t}$, where vector $\bar{\beta} = (1, t, \ldots, t^p)'$, $A_{i,t}$ is the design matrix with $\ell^{th}$ column $(\bar{t}_i(t))^{\ell}$, $\ell = 0, 1, \ldots, p$ and $W_{i,t}$ is longitudinal measurements of the $i^{th}$ subject taken up to and including time $t$. A sufficient statistic of $X_i(t)$ is proposed, provided the parameters are known here,

$$S_i(t, \gamma) = \gamma\sigma^2_{\hat{X}_i(t)_{ols}} dN_i(t) + \hat{X}_i(t)_{ols},$$

where $\sigma^2_{\hat{X}_i(t)_{ols}} = \text{Var}(\hat{X}_i(t)_{ols}|\alpha_i)$ and $dN_i(t) = I(V_i \leq t, t_{i(p+1)} \leq t, \delta_i = 1)$. Given the sufficient statistic, the conditional hazard function turns out to be

$$\lambda_i(t|S_i(t, \gamma)) = \lambda_0(t) \exp(\gamma S_i(t, \gamma) - \frac{1}{2} \gamma^2 \sigma^2_{\hat{X}_i(t)_{ols}} + \eta'Z_i).$$
Unbiased estimating equations for \((\gamma, \eta)\)' are proposed based on the conditional hazard,

\[
\sum_{i=1}^{m} \int \left( \frac{S_i(t, \gamma)}{Z_i} \right) \left( dN_i(t) - \lambda_i(t|S_i(t, \gamma))dt \right) = 0,
\]

and therefore statistical inference with respect to the survival regression model is obtained. By utilising the conditional score estimator, random effects are actually treated as nuisance parameters to some extent and therefore no distribution assumption is required. The estimates have been shown to be consistent and robust under certain assumptions. One can refer to Tsiatis and Davidian (2001) for more details.

3.4 Bias Analysis

In the work of Tsiatis and Davidian (2001), it was assumed that the within-subject errors are mutually independent and normally distributed. However, violating the independence assumption may lead to bias since the variance of the ordinary least squares estimators of \(X_i(t)\) will have changed. Under the assumption that, given random effects \(\alpha_i\), the within-subject errors \(\varepsilon_i\) of the \(i^{th}\) subject are normally distributed with zero mean and \(\sigma^2 I_n\) covariance matrix, we have \(\hat{X}_i(t)_{\text{ols}}\) normally distributed with \(E(\hat{X}_i(t)_{\text{ols}}) = X_i(t)\) and \(\text{Var}(\hat{X}_i(t)_{\text{ols}}) = \sigma^2 \hat{\theta}(A'_{i,t}A_{i,t})^{-1}\tilde{\tau}\), which will appear in conditional intensity to assist the statistical inference for survival analysis under the framework of the conditional score approach. In practice, the pooled estimator provides a consistent estimator for unknown \(\sigma^2\) (Tsiatis and Davidian, 2001),

\[
\hat{\sigma}^2 = \frac{\sum_{i=1}^{m} I(n_i > (p + 1)) SSE_i}{\sum_{i=1}^{m} I(n_i > (p + 1)) (n_i - p - 1)},
\]

where \(SSE_i\) is the residual sum of squares for the least squares fit to all \(n_i\) observations of \(i^{th}\) subject. However, when the assumption of mutual independence has been violated with general covariance \(\Sigma_i\), the above variance rule for \(\hat{X}_i(t)_{\text{ols}}\) does not hold any more. The underlying variance turns out to be,

\[
\text{Var}(\hat{X}_i(t)_{\text{ols}}) = \hat{\tau}'(A'_{i,t}A_{i,t})^{-1}A_{i,t}^{'}\Sigma_{i,t}A_{i,t}(A_{i,t}^{'}A_{i,t})^{-1}\hat{\tau}, \quad (3.3)
\]

where \(\Sigma_{i,t} = \text{Cov}(\varepsilon_{ij} : t_{ij} \leq t)\), that is, the covariance matrix of the measurement errors up to and including time \(t\) of the \(i^{th}\) subject. If one assumes that the within-subject
error are independent, regardless underlying covariance matrix of \( \Sigma_{i,t} \), and would use the above algorithm to show that, given \( n_i \),

\[
\text{Var}(\hat{X}_i(t)_{\text{obs}}) = \hat{\sigma}^2 \tilde{I}(A_i'A_i)^{-1} \tilde{t},
\]

where

\[
\hat{\sigma}^2 = E(\hat{\sigma}^2) = \frac{\sum_{j=1}^{m} I(n_j > (p+1))(\text{tr}(\Sigma_j) - \text{tr}((A_j'A_j)^{-1}A_j'\Sigma_j A_j))}{\sum_{j=1}^{m} I(n_j > (p+1))(n_j - p - 1)},
\]

one can refer to chapter appendix D for more details. We will not theoretically investigate its properties and the relationship with underlying variance (3.3), but it seems that when \( \Sigma_i \neq \sigma^2 I_{n_i} \), variances (3.3) and (3.4) may not even be close. We study the influence of misspecification of covariance of measurement error on the survival parameters inference by a numerical study, where a larger difference between variance (3.3) and (3.4) may lead to a larger bias for the conditional score estimator.

A simulation study is conducted under the following scenario, which is a modification of the one in Tsiatis and Davidian (2001) and the crucial part is that the within-subject error \( \varepsilon_i \) of \( i^\text{th} \) subject are from multivariate normal distribution with mean zero but AR(1) structure covariance with \( (i,j)^{th} \) element of \( \sigma^2 \rho^{|i-j|} \), rather than \( \sigma^2 I_{n_i} \). Simulated longitudinal observations are from \( W_i(t_{ij}) = X_i(t_{ij}) + \varepsilon_{ij} \), where \( X_i(t_{ij}) = \alpha_0 + \alpha_1 t_{ij} \). Both random error \( \varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{im})' \) and random effects \( \alpha_i = (\alpha_{i0}, \alpha_{i1})' \) of \( i^\text{th} \) subject are generated from multivariate normal distributions, and \( \varepsilon_i \) and \( \alpha_i \) are independent. Regarding the hazard function, apart from the time-dependent covariate \( X_i(t) \), we consider two additional baseline covariates \( Z_i = (Z_{i1}, Z_{i2})' \) in the Cox regression model, that is, \( \lambda_i(t) = \lambda_0(t) \exp\{\gamma X_i(t) + \eta_1 Z_{i1} + \eta_2 Z_{i2}\} \), where \( Z_{i1} \) is continuous covariate and simulated from normal distribution \( N(1,1) \), while \( Z_{i2} \) is factor or categorical variable with value of 1 or 0, treatment indicators, for example, and we generate \( Z_{i2} \) from a Bernoulli distribution of Bin(1,0.6). Similarly, we have \( E(\alpha_i) = (4.173, -0.0103)' \) and \( \text{Cov}(\alpha_i) = D \), where \( D \) has distinct elements \( (D_{11}, D_{12}, D_{22}) = (1.24, -0.0114, 0.003) \). In addition, the underlying baseline hazards function \( \lambda_0(t) = 1, \gamma = -1, \eta_1 = 1 \) and \( \eta_2 = -2 \), and all the possible measurement time points are at weeks of \{0, 2, 4, 6, 8, 10, 12, 14, 16, 24, 32, 40, 48, 56, 64, 72, 80\}. Measurements scheduled after week 16 have 10% chance of missing and no measurement is taken beyond the observed survival time, that is, \( t_{ij} \leq V_i \), where \( V_i = \min(T_i, C_i) \),
CHAPTER 3. MEASUREMENT ERROR

the underlying failure time $T_i$ are simulated according to hazards function and the
censoring time $C_i$ are generated from exponential distribution with mean 110, and no
event or censoring before and including week 16.

In the numerical study, we consider a sequence of $\rho$ with values from -0.9 to 0.9
by setting $step = 0.1$ and $\sigma^2 = 0.6$ to construct the AR(1) within-subject covariance.
Each scenario for different value of $\rho$, 500 Monte Carlo datasets are generated with
a sample size of $m = 300$ for each dataset. Conditional score approach proposed by
Tsiatis and Davidian (2001) is utilised for the statistical inference of survival param-
eters with the mutually independent assumption on $\varepsilon_i$, regardless of the underlying
AR(1) covariance structure. Meanwhile, we also present the naive estimators, where
$X_i(t)$ are imputed at each failure time from a single least squares fit to all the available
longitudinal data for each subject $i$ with $n_i \geq 2$ and substituted in the usual partial
likelihood, as discussed by Tsiatis and Davidian (2001). Simulation results are demon-
strated by Figure 3.1 to Figure 3.3 with respect to parameters $(\gamma, \eta_1, \eta_2)'$, respectively,
where the smoothing curves are generated by using the loess command of R software.
Figures illustrate a clear pattern of bias that the stronger the correlation, the larger
the bias with respect to not just $\gamma$ but also $\eta_1$ and $\eta_2$, though they are nearly unbiased
when $\rho = 0$. And figures also show that when $\rho$ goes to 1, conditional score approach
gradually performs toward the naive method, while when $\rho$ goes to -1, it tends to
severely overestimate the survival parameters, which calls for additional caution when
using the proposed conditional score approach. It is interesting to note that the naive
estimation seems to have a good performance when $\rho$ goes to -1. And our guess is
that, for some extreme cases, for example $\rho = -0.9$, and that when the number of
measurements may not be large enough, such strong negative correlation may be ac-
counted for in analysis by regression models of longitudinal measurements, that is, by
random effects $\alpha_i$, when the least square estimation is utilised for each subject. As
illustrated by Figure 3.1 to Figure 3.3, the absolute CS bias is less when $\rho$ is positive
than that when it is negative, which may be due to the fact that the difference be-
tween variance (3.3) and (3.4) is less when the $\varepsilon_i$ are positively correlated than when
negatively correlated. To some extent, some positive inherent association introduced
by the AR(1) covariance structure has been accounted for by the proposed least square
estimate $\hat{X}_i(t)_{ols}$. Hence the bias on survival parameters, due to the misspecification of
within-subject covariance, concerns us and it may also happen when other approaches
are utilised to the joint modelling of longitudinal and survival data, such as the stand-
ard classical likelihood-based approach. It may be common to lead to bias inference
for survival parameters when within-subject covariance has been misspecified and the
concern motives the work of this study.

3.5 Generalised Conditional Score Estimator

Bias analysis shows that the violation of mutual independence assumption may lead to
misuse of variance of OLS estimator of $X_i(t)$, which actually causes biased inference
for survival parameters by utilising the conditional score approach. To correct the
inference, one can replace the erroneous variance of estimator by equation (3.3), that
is, the exact variance of OLS estimator of $X_i(t)$, to access the statistical inference for
survival models. Alternatively, we propose generalised least squares estimator for the
unknown underlying covariate $X_i(t)$, which has smaller variation than that of OLS
estimator. Let $\hat{X}_i(t)_{\text{gls}}$ denote the generalised least squares (GLS) estimator of $X_i(t)$
for the $i^{th}$ subject with at least $p+1$ measurements taken by using all the longitudinal
data up to and including time $t$, that is, we have,

$$\hat{X}_i(t)_{\text{gls}} = \hat{\gamma}(A_{i,t}'\Sigma_{i,t}^{-1}A_{i,t})^{-1}A_{i,t}'\Sigma_{i,t}^{-1}W_{i,t}. $$

Given $\alpha_i$, along with the assumption that $\varepsilon_i$ are from multivariate normal distribution
with mean zero and covariance $\Sigma_i$, the GLS estimators $\hat{X}_i(t)_{\text{gls}}$ are normally distributed
with mean $X_i(t)$ and variance $\hat{\sigma}^2(X_i(t)_{\text{gls}})$, denoted by $\hat{\sigma}^2(X_i(t)_{\text{gls}})$. We assume that
$\Sigma_i(V_i) = \{\Sigma_{i,t} : t \leq V_i\}$ are all known for now, and they will be relaxed subsequently.

Define the counting process $N_i(t) = I(\delta_i = 1, V_i \leq t, t_{i(p+1)} \leq t)$ and the at risk
process $Y_i(t) = I(V_i \geq t, t_{i(p+1)} \leq t)$, where $t_{i(p+1)}$ is the $(p+1)^{th}$ measurement time
point of $i^{th}$ subject as aforementioned. It is straightforward to show that, conditional
on $Y_i(t) = 1$, the complete sufficient statistic for $X_i(t)$ is,

$$S_i(t, \gamma, \Sigma_{i,t}) = \gamma\sigma^2_{X_i(t)_{\text{gls}}} dN_i(t) + \hat{X}_i(t)_{\text{gls}},$$

where $dN_i(t)$ is the corresponding counting process increment of $N_i(t)$. Similarly, the
Figure 3.1: Conditional score (CS) estimators of survival parameters, varying with correlation $\rho$, with the mutual independent assumption on $\varepsilon_i$ for each subject $i$ regardless underlying covariance of $\varepsilon_i$ of AR(1)=$\sigma^2|\rho| i - j$ with $\sigma^2 = 0.6$; x-axis is for $\rho$, y-axis are CS estimators for (a) $\gamma$; dotted horizontal line indicates true parameter value, solid black line is CS estimates with point estimators surrounding, and the dash-dot line is for Naive estimators; 500 Monte Carlo datasets and 300 sample size for each $\rho$. 
Figure 3.2: Conditional score (CS) estimators of survival parameters, varying with correlation ρ, with the mutual independent assumption on ε_i for each subject i regardless underlying covariance of ε_i of AR(1)=σ²ρ|i−j| with σ² = 0.6; x-axis is for ρ, y-axis are CS estimators for (b) η_1; dotted horizontal line indicates true parameter value, solid black line is CS estimates with point estimators surrounding, and the dash-dot line is for Naive estimators; 500 Monte Carlo datasets and 300 sample size for each ρ.
Figure 3.3: Conditional score (CS) estimators of survival parameters, varying with correlation $\rho$, with the mutual independent assumption on $\varepsilon_i$ for each subject $i$ regardless underlying covariance of $\varepsilon_i$ of AR(1)=$\sigma^2|\rho| i - j|$ with $\sigma^2 = 0.6$; x-axis is for $\rho$, y-axis are CS estimators for (c) $\eta_2$; dotted horizontal line indicates true parameter value, solid black line is CS estimates with point estimators surrounding, and the dash-dot line is for Naive estimators; 500 Monte Carlo datasets and 300 sample size for each $\rho$. 
The conditional intensity process is defined as
\[
\lim_{dt \to 0} dt^{-1} Pr(dN_i(t) = 1 | S_i(t, \gamma, \Sigma_{i,t}), Y_i(t)) = \lambda_0(t) \exp \left( \gamma S_i(t, \gamma, \Sigma_{i,t}) - \frac{1}{2} \gamma^2 \sigma_{\hat{X}_i(t)}^2 + \eta' Z_i \right) Y_i(t),
\] (3.5)
and one can refer to chapter appendix C for more details. The conditional intensity of 
\[
dN_i(t) = \sum_{i=1}^m dN_i(t)
\]
given \{S_i(t, \gamma, \Sigma_{i,t}), Z_i, \tilde{t}_i(t), Y_i(t), i = 1, \ldots, m\}
is \(\lambda_0(t) E_0(t, \gamma, \eta, \Sigma_t)\), where denote \(\Sigma_t = \{\Sigma_{i,t} : i = 1, \cdots, m\}\) and
\[
E_0(t, \gamma, \eta, \Sigma_t) = \sum_{i=1}^m E_{0i}(t, \gamma, \eta, \Sigma_{i,t});
\]
\[
E_{0i}(t, \gamma, \eta, \Sigma_{i,t}) = \exp \left( \gamma S_i(t, \gamma, \Sigma_{i,t}) - \frac{1}{2} \gamma^2 \sigma_{\hat{X}_i(t)}^2 + \eta' Z_i \right) Y_i(t),
\]
which suggests that a reasonable estimator for \(\lambda_0(t) dt\) is given by
\[
\hat{\lambda}_0(t) dt = dN(t)/E_0(t, \gamma, \eta, \Sigma_t).
\]
Denote \(\theta_s = (\gamma, \eta)'\) and let \(K_i(t, \theta_s)\) be any deterministic functions or, more generally, predictable, locally bounded process. And denote intensity function of \(i^{th}\) subject by \(\lambda_i(t; \theta_s)\) to emphasise that it is function of parameter \(\theta_s\). By analogy with M-Estimators (Andersen et al., 1993, p433), we can define an estimator \(\hat{\theta}_s\) as a solution to the equation,
\[
U_1(\theta_s) = \sum_{i=1}^m \int K_i(t, \theta_s) \left\{ dN_i(t) - \lambda_i(t; \theta_s) dt \right\}
\]
\[
= \sum_{i=1}^m \int K_i(t, \theta_s) \left\{ dN_i(t) - E_{0i}(t, \gamma, \eta, \Sigma_{i,t}) \lambda_0(t) dt \right\} = 0, \quad (3.6)
\]
which, upon substitution of \(\hat{\lambda}_0(t) dt\) for \(\lambda_0(t) dt\), may be written as
\[
U_1(\theta_s) = \sum_{i=1}^m \int K_i(t, \theta_s) \left\{ dN_i(t) - \frac{E_{0i}(t, \gamma, \eta, \Sigma_{i,t})}{E_0(t, \gamma, \eta, \Sigma_t)} dN(t) \right\} = 0.
\]
Define \(E_1(t, \gamma, \eta, \Sigma_t) = \sum_{i=1}^m E_{1i}(t, \gamma, \eta, \Sigma_{i,t})\) with
\[
E_{1i}(t, \gamma, \eta, \Sigma_{i,t}) = K_i(t, \theta_s) E_{0i}(t, \gamma, \eta, \Sigma_{i,t}),
\]
therefore, similarly, the estimating equations can be expressed as

\[ U_1(\theta_s) = \sum_{i=1}^{m} \int \left\{ K_i(t, \theta_s) - \frac{E_1(t, \gamma, \eta, \Sigma_t)}{E_0(t, \gamma, \eta, \Sigma_t)} \right\} dN_i(t) = 0. \]  

(3.7)

In the spirit of M-estimator and also maximum likelihood estimator (Andersen et al., 1993, p433), we propose two forms of \( K_i(t, \theta_s) \), one is \( K_i(t, \theta_s) = (S_i(t, \gamma, \Sigma_{i,t}), Z_i)' \) as proposed by Tsiatis and Davidian (2001) and refered as MEst estimation thereafter. The other one is \( K_i(t, \theta_s) = (S_i(t, \gamma, \Sigma_{i,t}) - \gamma \sigma^2_{X_i(t)_s}, Z_i)' \) and referred as MLE estimation thereafter. Statistical inference for \( \theta_s = (\gamma, \eta)' \) is the primary interest of this study. Accordingly, Tsiatis and Davidian (2001) sketched that the conditional score approach generated consistent and approximately normally distributed estimators, which is also straightly applicable to the proposed generalized conditional score approach. Numerical methods, such as Newton-Raphson algorithm, are required to obtain the solution \( \hat{\theta}_s \) of equations (3.7), which although may have several roots, using the naive estimator as the starting value (as suggested by Stefanski and Carroll, 1987) would be a practical strategy for locating the consistent root. Our simulation studies have also verified this phenomenon.

### 3.6 Covariance Estimator and the Inference

We are aware of that the survival estimating equations (3.7) are functions of covariance matrices \( \Sigma_{i,t}s \), that is, the implementation of proposed generalised conditional score (GCS) approach requires the knowledge of within-subject covariance, which, however, usually are unknown in practice and require inference as well. Denote \( \theta_c \) the parameters of within-subject covariance and the corresponding proposed estimating equations \( U_2(\theta_c) = 0 \), which will be discussed in details later on. We obtain statistical inference for within-subject covariances \( \Sigma_i, i = 1, \ldots, m \), based on information provided by the measurements taken from longitudinal process, that is, the proposed estimating equations with respect to \( \theta_c \), i.e \( U_2(\theta_c) = 0 \), can be free of survival parameters \( \theta_s = (\gamma, \eta)' \). Denote \( \theta = (\gamma, \eta, \theta_c)' \), then estimating equations of parameters \( \theta \) can be denoted by

\[ U(\theta) = \begin{pmatrix} U_1(\theta_s, \theta_c) \\ U_2(\theta_c) \end{pmatrix} = 0, \]  

(3.8)
where $U_1(\theta_s, \theta_c)$ relaxes the required knowledge of covariances $\Sigma_{i,t}$s in equations (3.7). Due to the complexity of models, we apply a “two-stage” estimation method rather than a simultaneous method to obtain inference for $\theta$. The estimator for $\hat{\theta}_c$ is obtained according to equations $U_2(\theta_c) = 0$ first, then we replace the covariance parameters by $\hat{\theta}_c$ in $U_1(\theta_s, \theta_c) = 0$ to approach the survival inference. Furthermore, the usual sandwich rule is used to obtain the approximate standard error for the estimates $\hat{\theta} = (\hat{\gamma}, \hat{\eta}, \hat{\theta}_c)'$, more specifically, $\text{Cov}(\hat{\theta}) = \left(\frac{\partial U(\theta)}{\partial \theta}\right)^{-1} \left(\sum_{i=1}^{m} u_i(\theta)u_i'(\theta)\right)\left(\frac{\partial U(\theta)}{\partial \theta}\right)^{-1}$, where we denote that $U(\theta) = \sum_{i=1}^{m} u_i(\theta)$ and $u_i(\theta)$ denotes the first derivative with respect to $\theta$ of $i^{th}$ subject loglikelihood function. The following simulation studies also demonstrate that this strategy yields reliable estimates of uncertainty. The details of estimating equations and derivatives for sandwich rule can be found in the chapter appendix A and B with respect to equations $U_2(\theta_c) = 0$ and $U_1(\theta_s, \theta_c) = 0$, respectively.

3.6.1 Software Implementation

Due to the constraint of positive definiteness and potential high dimensions of parameters of covariance matrix, it is well known that usually it is not easy to develop inference for a covariance matrix, especially the within-subject covariance matrices $\Sigma_i$, $i=1,\ldots,m$, in our study. On the other hand, with respect to linear mixed-effects models, there already have been some plug-in packages available to obtain the estimators of within-subject covariance, such as lme by R software. In this section, we replace the unknown covariances $\Sigma_i$s in survival estimating equations (3.7) with estimators obtained straightly from lme output, which can be somewhat consistent and unbiased with correct model fitting, as demonstrated by simulation results, especially when the number of longitudinal measurements is relatively large enough. Note that when software line lme is being utilised for covariance inference, it goes with the assumption that random effects $\alpha_i$s are from multivariate normal distributions, which may not be true, as illustrated by the simulation scenarios. However, the simulation study shows that even the underlying random effects are actually generated from other distributions rather than the normal ones, lme may still output plausible inference for the within-subject covariance under certain framework. We carry out simulation scenarios under the settings of bias analysis but with additional investigation aspects. Regarding the random effects, four different underlying distributions are being considered here and
they are the modified cases in Tsiatis and Davidian (2001): (a) normal distribution as detailed in bias analysis; (b) bimodal mixture of normals with mixing proportion 0.5 generated as described in Davidian and Gallant (1993), we set \( sep = 4 \) and choose upper triangular matrix to yield \( \text{cov}(\alpha_i) \) with the same standard deviations of random effects in the normal case but \( D_{12} = 0.039 \); (c) bivariate skew-normal distribution (Azzalini and Dalla Valle, 1996) with coefficients of skewness of -0.1 and 0.95 for \( \alpha_{0i} \) and \( \alpha_{1i} \), respectively; (d) bivariate t distribution with five degree of freedom. Details on how to generate random effects from distributions (b)-(d), one can refer to Chapter 6 of the thesis. Random effects \( \alpha_i \)s are generated to have same mean and covariance as the norm case unless specified. Along with distribution set of random effects, the simulated measurement error \( \varepsilon_i \)s are normally distributed with mean zero and AR(1) structure of covariance with \( \sigma^2 = 0.6 \). With respect to the correlation parameter \( \rho \) of AR(1), four scenarios are being conducted:

- Scenario I: negative correlation \( \rho = -0.6 \);
- Scenario II: weak positive correlation \( \rho = 0.25 \);
- Scenario III: moderate positive correlation \( \rho = 0.5 \);
- Scenario IV: strong positive correlation \( \rho = 0.75 \).

When \( \rho = -0.6 \), data are generated with \( t_i \in \{0, 2, 4, 8, 16, 24, 32, 40, 48, 56, 64, 72, 80\} \), while when the measurement error are positively correlated, the \( t_i \) settings are the same as that in bias analysis, that is, more longitudinal measurements tend to be taken in these scenarios, in order to make the AR(1) covariance structure more identifiable. Within the investigation of each distribution of each scenario, simulations are conducted with 500 Monte Carlo datasets and 500 sample size for each dataset. Simulation results are presented through Table 3.1 to Table 3.8, where inference for survival and AR(1) covariance parameters are illustrated by utilising four approaches: ideal, naive, conditional score (CS) and generalised conditional score (GCS). The ideal outputs are obtained by placing the exact \( X_i(t) \) values in partial likelihood function to obtain the traditional partial likelihood estimators and standards deviation for the survival parameters, which serve as benchmark for perfect performance of inference. In tables, we only present GCS inference of MEst estimations since MLE approach has
very similar performance with that of MEst approach. And note in tables RB(%) is
the relative bias calculated by percentage of differences between mean of Monte Carlo
estimates and true values over underlying true values.

Consistent with findings presented in bias analysis, CS approach tends to overesti-
mate the Cox regression coefficients of $\gamma$ and $\eta$ when measurement error are negatively
correlated and leads to substantial bias when the correlation is strong, for instance,
there is even more than 70% relative bias for $\gamma$ estimators when $\rho = -0.6$, as presented
by Table 3.1 and Table 3.2. While it tends to attenuate the regression coefficients when
measurement error are positively correlated and the larger the correlation, the larger
the bias, as presented by Table 3.3 to Table 3.8. When $\rho = 0.75$, the statistical perfor-
mane of CS approach goes down to that of naive approach, a bit even worse that with
slightly larger variation, in terms of that in tables the SDs of CS method are slightly
larger than that of naive approach. On the other hand, the naive approach always
tends to attenuate the regression coefficients. Due to the biased inference, both CS
and naive approaches tend to go with pretty low coverage probabilities with respect to
the estimates of survival parameters. In the meantime, it is interesting to note that,
amongst four different distributions of random effects, the one with bimodal mixture
tends to have smaller bias with respect to association parameter $\gamma$.

Regarding the proposed GCS approach, in general, with a good covariance estima-
tion with respect to measurement error, it can largely improve statistical inference for
survival parameters and lead to nearly unbiased estimates, regardless underlying dis-
tribution of random effects. And the mean of Monte Carlo estimators are quite close to
that of ideal statistical performance, especially when there is an excellent performance
of statistical inference for AR(1) covariance of random error, as presented by tables
from Table 3.1 to Table 3.4. Because GCS approach takes into account within-subject
covariance for inference, the corresponding survival estimators tend to have larger vari-
ation in terms of that the SDs are larger than those of other approaches, which seems
to be acceptable here since the GCS approach has substantially reduce the bias for
return. And therefore the inference goes with plausible coverage probability as well.
Though we also note that SEs tend to slightly smaller than SDs, which may be partly
due to the inference procedure of “two-stage” with respect to covariance and survival
parameters, and/or which may also due to the relative value of $\sigma^2$. As demonstrated
Table 3.1: **Scenario I(a)(b): AR(1)**. Simulation results for $\rho = -0.6$ and $\sigma^2 = 0.6$ of AR(1) within-subject covariance. All the possible measured time points are at weeks of \{0,2,4,8,16,24,32,40,48,56,64,72,80\} and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.

<table>
<thead>
<tr>
<th>$\rho = -0.6$</th>
<th>Method</th>
<th>$\alpha_1$ Normal</th>
<th>$\alpha_1$ Bimodal mixture</th>
<th>$\gamma(-1)$</th>
<th>Mean</th>
<th>RB(%)</th>
<th>SD</th>
<th>SE</th>
<th>CP(%)</th>
<th>Mean</th>
<th>RB(%)</th>
<th>SD</th>
<th>SE</th>
<th>CP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ideal</td>
<td>-1.004</td>
<td>0.4</td>
<td>0.051</td>
<td>0.051</td>
<td>94.2</td>
<td></td>
<td></td>
<td></td>
<td>-1.002</td>
<td>0.2</td>
<td>0.046</td>
<td>0.046</td>
<td>94.8</td>
</tr>
<tr>
<td></td>
<td>Naive</td>
<td>-0.841</td>
<td>15.9</td>
<td>0.046</td>
<td>0.043</td>
<td>7.4</td>
<td></td>
<td></td>
<td></td>
<td>-0.869</td>
<td>13.1</td>
<td>0.043</td>
<td>0.037</td>
<td>11.2</td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>-1.663</td>
<td>66.3</td>
<td>0.261</td>
<td>0.200</td>
<td>6.6</td>
<td></td>
<td></td>
<td></td>
<td>-1.399</td>
<td>39.9</td>
<td>0.176</td>
<td>0.140</td>
<td>17.0</td>
</tr>
<tr>
<td></td>
<td>GCS</td>
<td>-1.012</td>
<td>12.2</td>
<td>0.072</td>
<td>0.068</td>
<td>95.0</td>
<td></td>
<td></td>
<td></td>
<td>-1.008</td>
<td>0.8</td>
<td>0.062</td>
<td>0.058</td>
<td>93.4</td>
</tr>
<tr>
<td></td>
<td>Ideal</td>
<td>1.002</td>
<td>0.2</td>
<td>0.066</td>
<td>0.068</td>
<td>96.2</td>
<td></td>
<td></td>
<td></td>
<td>0.998</td>
<td>0.2</td>
<td>0.073</td>
<td>0.071</td>
<td>93.6</td>
</tr>
<tr>
<td></td>
<td>Naive</td>
<td>0.902</td>
<td>9.8</td>
<td>0.068</td>
<td>0.064</td>
<td>68.2</td>
<td></td>
<td></td>
<td></td>
<td>0.888</td>
<td>11.2</td>
<td>0.071</td>
<td>0.066</td>
<td>60.2</td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>1.404</td>
<td>40.4</td>
<td>0.203</td>
<td>0.165</td>
<td>32.8</td>
<td></td>
<td></td>
<td></td>
<td>1.336</td>
<td>33.6</td>
<td>0.184</td>
<td>0.151</td>
<td>40.0</td>
</tr>
<tr>
<td></td>
<td>GCS</td>
<td>1.007</td>
<td>0.7</td>
<td>0.085</td>
<td>0.079</td>
<td>93.2</td>
<td></td>
<td></td>
<td></td>
<td>1.006</td>
<td>0.6</td>
<td>0.087</td>
<td>0.081</td>
<td>92.8</td>
</tr>
<tr>
<td></td>
<td>Ideal</td>
<td>-2.018</td>
<td>0.9</td>
<td>0.140</td>
<td>0.137</td>
<td>94.4</td>
<td></td>
<td></td>
<td></td>
<td>-2.006</td>
<td>0.3</td>
<td>0.146</td>
<td>0.143</td>
<td>94.6</td>
</tr>
<tr>
<td></td>
<td>Naive</td>
<td>-1.814</td>
<td>9.3</td>
<td>0.139</td>
<td>0.128</td>
<td>68.2</td>
<td></td>
<td></td>
<td></td>
<td>-1.779</td>
<td>11.1</td>
<td>0.146</td>
<td>0.133</td>
<td>62.8</td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>-2.822</td>
<td>41.1</td>
<td>0.406</td>
<td>0.344</td>
<td>33.4</td>
<td></td>
<td></td>
<td></td>
<td>-2.680</td>
<td>34.0</td>
<td>0.385</td>
<td>0.311</td>
<td>42.8</td>
</tr>
<tr>
<td></td>
<td>GCS</td>
<td>-2.027</td>
<td>1.3</td>
<td>0.165</td>
<td>0.163</td>
<td>94.6</td>
<td></td>
<td></td>
<td></td>
<td>-2.013</td>
<td>0.6</td>
<td>0.184</td>
<td>0.167</td>
<td>92.2</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>CS</td>
<td>0.7402</td>
<td>23.4</td>
<td>0.026</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.7372</td>
<td>22.9</td>
<td>0.027</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2(0.6)$</td>
<td>GCS</td>
<td>0.6009</td>
<td>0.2</td>
<td>0.019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.6008</td>
<td>0.2</td>
<td>0.020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho(-0.6)$</td>
<td>GCS</td>
<td>-0.5985</td>
<td>0.3</td>
<td>0.015</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.6000</td>
<td>0.0</td>
<td>0.015</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.2: Scenario I(c)(d): AR(1). Simulation results for $\rho = -0.6$ and $\sigma^2 = 0.6$ of AR(1) within-subject covariance. All the possible measured time points are at weeks of $\{0,2,4,8,16,24,32,40,48,56,64,72,80\}$ and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\alpha_1$ Skew-normal</th>
<th>Mean</th>
<th>RB(%)</th>
<th>SD</th>
<th>SE</th>
<th>CP</th>
<th>$\alpha_i$ Bivariate $t_5$</th>
<th>Mean</th>
<th>RB(%)</th>
<th>SD</th>
<th>SE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma(-1)$</td>
<td>Ideal</td>
<td>-1.010</td>
<td>1.0</td>
<td>0.052</td>
<td>0.052</td>
<td>94.8</td>
<td>-1.004</td>
<td>0.4</td>
<td>0.053</td>
<td>0.052</td>
<td>94.8</td>
<td></td>
</tr>
<tr>
<td>Naive</td>
<td>-0.838</td>
<td>16.2</td>
<td>0.051</td>
<td>0.043</td>
<td>7.8</td>
<td>-0.838</td>
<td>16.2</td>
<td>0.048</td>
<td>0.045</td>
<td>9.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>-1.773</td>
<td>77.3</td>
<td>0.321</td>
<td>0.226</td>
<td>7.0</td>
<td>-1.713</td>
<td>71.3</td>
<td>0.270</td>
<td>0.200</td>
<td>5.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCS</td>
<td>-1.013</td>
<td>1.3</td>
<td>0.076</td>
<td>0.070</td>
<td>94.8</td>
<td>-1.019</td>
<td>1.9</td>
<td>0.077</td>
<td>0.071</td>
<td>92.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_1(1)$</td>
<td>Ideal</td>
<td>1.009</td>
<td>0.9</td>
<td>0.069</td>
<td>0.067</td>
<td>95.4</td>
<td>1.010</td>
<td>1.0</td>
<td>0.068</td>
<td>0.067</td>
<td>94.2</td>
<td></td>
</tr>
<tr>
<td>Naive</td>
<td>0.908</td>
<td>9.2</td>
<td>0.070</td>
<td>0.064</td>
<td>68.6</td>
<td>0.916</td>
<td>8.4</td>
<td>0.069</td>
<td>0.064</td>
<td>71.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>1.441</td>
<td>44.1</td>
<td>0.236</td>
<td>0.173</td>
<td>30.8</td>
<td>1.437</td>
<td>43.7</td>
<td>0.227</td>
<td>0.167</td>
<td>30.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCS</td>
<td>1.012</td>
<td>1.2</td>
<td>0.085</td>
<td>0.080</td>
<td>95.6</td>
<td>1.021</td>
<td>2.1</td>
<td>0.089</td>
<td>0.079</td>
<td>90.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_2(-2)$</td>
<td>Ideal</td>
<td>-2.018</td>
<td>0.9</td>
<td>0.136</td>
<td>0.136</td>
<td>95.0</td>
<td>-2.010</td>
<td>0.5</td>
<td>0.141</td>
<td>0.135</td>
<td>94.2</td>
<td></td>
</tr>
<tr>
<td>Naive</td>
<td>-1.812</td>
<td>9.4</td>
<td>0.142</td>
<td>0.127</td>
<td>65.8</td>
<td>-1.818</td>
<td>9.1</td>
<td>0.138</td>
<td>0.127</td>
<td>65.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>-2.881</td>
<td>44.0</td>
<td>0.490</td>
<td>0.354</td>
<td>35.6</td>
<td>-2.857</td>
<td>42.9</td>
<td>0.424</td>
<td>0.340</td>
<td>32.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCS</td>
<td>-2.024</td>
<td>1.2</td>
<td>0.172</td>
<td>0.162</td>
<td>94.0</td>
<td>-2.030</td>
<td>1.5</td>
<td>0.173</td>
<td>0.161</td>
<td>93.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>CS</td>
<td>0.7405</td>
<td>23.4</td>
<td>0.027</td>
<td>-</td>
<td>-</td>
<td>0.7388</td>
<td>23.1</td>
<td>0.025</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2(0.6)$</td>
<td>GCS</td>
<td>0.5984</td>
<td>0.3</td>
<td>0.020</td>
<td>-</td>
<td>-</td>
<td>0.5992</td>
<td>0.1</td>
<td>0.019</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>$\rho(-0.6)$</td>
<td>GCS</td>
<td>-0.6012</td>
<td>0.2</td>
<td>0.016</td>
<td>-</td>
<td>-</td>
<td>-0.5983</td>
<td>0.3</td>
<td>0.016</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.3: Scenario II(a)(b): AR(1) Simulation results for $\rho = 0.25$ and $\sigma^2 = 0.6$ of AR(1) within-subject covariance. All the possible measured time points are at weeks of $\{0, 2, 4, 6, 8, 10, 12, 14, 16, 24, 32, 40, 48, 56, 64, 72, 80\}$ and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.

<table>
<thead>
<tr>
<th>$\rho = 0.25$</th>
<th>Method</th>
<th>$\alpha_i$ Normal</th>
<th>RB(%)</th>
<th>SD</th>
<th>SE</th>
<th>CP(%)</th>
<th>$\alpha_i$ Bimodal mixture</th>
<th>Mean</th>
<th>RB(%)</th>
<th>SD</th>
<th>SE</th>
<th>CP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma(-1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>1.006</td>
<td>0.6</td>
<td>0.052</td>
<td>0.051</td>
<td>93.8</td>
<td>1.002</td>
<td>0.2</td>
<td>0.045</td>
<td>0.045</td>
<td>95.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive</td>
<td>-0.780</td>
<td>22.0</td>
<td>0.046</td>
<td>0.041</td>
<td>0.4</td>
<td>-0.835</td>
<td>16.5</td>
<td>0.041</td>
<td>0.035</td>
<td>2.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>-0.876</td>
<td>12.4</td>
<td>0.069</td>
<td>0.063</td>
<td>44.8</td>
<td>-0.920</td>
<td>8.0</td>
<td>0.059</td>
<td>0.053</td>
<td>62.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCS</td>
<td>-1.028</td>
<td>2.8</td>
<td>0.103</td>
<td>0.091</td>
<td>93.4</td>
<td>-1.020</td>
<td>2.0</td>
<td>0.079</td>
<td>0.070</td>
<td>91.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_1(1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>1.003</td>
<td>0.3</td>
<td>0.067</td>
<td>0.068</td>
<td>95.4</td>
<td>1.002</td>
<td>0.2</td>
<td>0.070</td>
<td>0.070</td>
<td>96.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive</td>
<td>0.874</td>
<td>12.6</td>
<td>0.066</td>
<td>0.064</td>
<td>50.8</td>
<td>0.864</td>
<td>13.6</td>
<td>0.069</td>
<td>0.065</td>
<td>41.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>0.936</td>
<td>6.4</td>
<td>0.081</td>
<td>0.078</td>
<td>84.6</td>
<td>0.934</td>
<td>6.6</td>
<td>0.085</td>
<td>0.079</td>
<td>84.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCS</td>
<td>1.022</td>
<td>2.2</td>
<td>0.101</td>
<td>0.094</td>
<td>92.8</td>
<td>1.017</td>
<td>1.7</td>
<td>0.103</td>
<td>0.093</td>
<td>92.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_2(-2)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>2.015</td>
<td>0.7</td>
<td>0.137</td>
<td>0.137</td>
<td>94.2</td>
<td>2.002</td>
<td>0.1</td>
<td>0.134</td>
<td>0.142</td>
<td>96.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive</td>
<td>-1.745</td>
<td>12.7</td>
<td>0.135</td>
<td>0.125</td>
<td>45.2</td>
<td>-1.726</td>
<td>13.7</td>
<td>0.135</td>
<td>0.131</td>
<td>41.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>-1.867</td>
<td>6.7</td>
<td>0.168</td>
<td>0.160</td>
<td>83.8</td>
<td>-1.864</td>
<td>6.8</td>
<td>0.165</td>
<td>0.163</td>
<td>86.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCS</td>
<td>-2.042</td>
<td>2.1</td>
<td>0.212</td>
<td>0.193</td>
<td>93.6</td>
<td>-2.031</td>
<td>1.5</td>
<td>0.197</td>
<td>0.191</td>
<td>93.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>CS</td>
<td>0.5346</td>
<td>10.9</td>
<td>0.012</td>
<td>-</td>
<td>0.5363</td>
<td>10.6</td>
<td>0.011</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCS</td>
<td>0.6017</td>
<td>0.3</td>
<td>0.017</td>
<td>-</td>
<td>0.6028</td>
<td>0.5</td>
<td>0.016</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2(0.6)$</td>
<td>GCS</td>
<td>0.2520</td>
<td>0.8</td>
<td>0.019</td>
<td>-</td>
<td>0.2527</td>
<td>1.1</td>
<td>0.019</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho(0.25)$</td>
<td>GCS</td>
<td>0.2520</td>
<td>0.8</td>
<td>0.019</td>
<td>-</td>
<td>0.2527</td>
<td>1.1</td>
<td>0.019</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.4: Scenario II(c)(d): AR(1) Simulation results for $\rho = 0.25$ and $\sigma^2 = 0.6$ of AR(1) within-subject covariance. All the possible measured time points are at weeks of \{0,2,4,6,8,10,12,14,16,24,32,40,48,56,64,72,80\} and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.

<table>
<thead>
<tr>
<th>$\rho = 0.25$ Method</th>
<th>$\alpha_i$ Skew-normal Mean</th>
<th>RB(%)</th>
<th>SD</th>
<th>SE</th>
<th>CP(%)</th>
<th>$\alpha_i$ Bivariate $t_5$ Mean</th>
<th>RB(%)</th>
<th>SD</th>
<th>SE</th>
<th>CP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma(-1)$ Ideal</td>
<td>-1.001</td>
<td>0.1</td>
<td>0.052</td>
<td>0.052</td>
<td>95.4</td>
<td>-1.006</td>
<td>0.6</td>
<td>0.052</td>
<td>0.052</td>
<td>95.0</td>
</tr>
<tr>
<td>Naive</td>
<td>-0.767</td>
<td>23.3</td>
<td>0.044</td>
<td>0.040</td>
<td>0.006</td>
<td>-0.768</td>
<td>23.2</td>
<td>0.045</td>
<td>0.043</td>
<td>0.0</td>
</tr>
<tr>
<td>CS</td>
<td>-0.866</td>
<td>13.4</td>
<td>0.069</td>
<td>0.064</td>
<td>41.2</td>
<td>-0.866</td>
<td>13.4</td>
<td>0.069</td>
<td>0.065</td>
<td>43.6</td>
</tr>
<tr>
<td>GCS</td>
<td>-1.012</td>
<td>1.2</td>
<td>0.103</td>
<td>0.091</td>
<td>94.4</td>
<td>-1.022</td>
<td>2.2</td>
<td>0.104</td>
<td>0.092</td>
<td>93.2</td>
</tr>
<tr>
<td>$\eta_1(1)$ Ideal</td>
<td>1.000</td>
<td>0.0</td>
<td>0.069</td>
<td>0.067</td>
<td>94.0</td>
<td>1.009</td>
<td>0.9</td>
<td>0.068</td>
<td>0.068</td>
<td>94.2</td>
</tr>
<tr>
<td>Naive</td>
<td>0.868</td>
<td>13.2</td>
<td>0.067</td>
<td>0.063</td>
<td>47.0</td>
<td>0.881</td>
<td>11.9</td>
<td>0.066</td>
<td>0.063</td>
<td>54.0</td>
</tr>
<tr>
<td>CS</td>
<td>0.929</td>
<td>7.1</td>
<td>0.081</td>
<td>0.077</td>
<td>82.8</td>
<td>0.943</td>
<td>5.7</td>
<td>0.080</td>
<td>0.077</td>
<td>85.8</td>
</tr>
<tr>
<td>GCS</td>
<td>1.006</td>
<td>0.6</td>
<td>0.099</td>
<td>0.092</td>
<td>92.6</td>
<td>1.026</td>
<td>2.6</td>
<td>0.100</td>
<td>0.093</td>
<td>93.4</td>
</tr>
<tr>
<td>$\eta_2(-2)$ Ideal</td>
<td>-1.999</td>
<td>0.1</td>
<td>0.122</td>
<td>0.135</td>
<td>97.6</td>
<td>-2.016</td>
<td>0.8</td>
<td>0.129</td>
<td>0.136</td>
<td>96.0</td>
</tr>
<tr>
<td>Naive</td>
<td>-1.734</td>
<td>13.3</td>
<td>0.127</td>
<td>0.124</td>
<td>43.2</td>
<td>-1.753</td>
<td>12.3</td>
<td>0.133</td>
<td>0.125</td>
<td>46.4</td>
</tr>
<tr>
<td>CS</td>
<td>-1.858</td>
<td>7.1</td>
<td>0.165</td>
<td>0.159</td>
<td>82.8</td>
<td>-1.871</td>
<td>6.4</td>
<td>0.167</td>
<td>0.159</td>
<td>84.4</td>
</tr>
<tr>
<td>GCS</td>
<td>-2.016</td>
<td>8.8</td>
<td>0.202</td>
<td>0.190</td>
<td>94.6</td>
<td>-2.035</td>
<td>1.8</td>
<td>0.204</td>
<td>0.191</td>
<td>92.8</td>
</tr>
<tr>
<td>AR(1) $\sigma^2$</td>
<td>CS</td>
<td>0.5341</td>
<td>11.0</td>
<td>0.012</td>
<td>-</td>
<td>0.5342</td>
<td>11.0</td>
<td>0.012</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCS</td>
<td>0.5949</td>
<td>0.8</td>
<td>0.016</td>
<td>-</td>
<td>0.6009</td>
<td>0.1</td>
<td>0.016</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2(0.6)$</td>
<td>CS</td>
<td>0.5341</td>
<td>11.0</td>
<td>0.012</td>
<td>-</td>
<td>0.5342</td>
<td>11.0</td>
<td>0.012</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCS</td>
<td>0.5949</td>
<td>0.8</td>
<td>0.016</td>
<td>-</td>
<td>0.6009</td>
<td>0.1</td>
<td>0.016</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>$\rho(0.25)$</td>
<td>CS</td>
<td>0.2437</td>
<td>2.5</td>
<td>0.019</td>
<td>-</td>
<td>0.2511</td>
<td>0.4</td>
<td>0.019</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.5: Scenario III(a)(b): AR(1) Simulation results for $\rho = 0.5$ and $\sigma^2 = 0.6$ of AR(1) within-subject covariance. All the possible measured time points are at weeks of $\{0, 2, 4, 6, 8, 10, 12, 14, 16, 24, 32, 40, 48, 56, 64, 72, 80\}$ and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.

<table>
<thead>
<tr>
<th>$\rho = 0.5$</th>
<th>Method</th>
<th>$\alpha_i$</th>
<th>Normal Mean</th>
<th>RB(%)</th>
<th>SD</th>
<th>SE</th>
<th>CP(%)</th>
<th>$\alpha_i$</th>
<th>Bimodal mixture Mean</th>
<th>RB(%)</th>
<th>SD</th>
<th>SE</th>
<th>CP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma(-1)$</td>
<td>Ideal</td>
<td>-1.006</td>
<td>0.6</td>
<td>0.052</td>
<td>0.051</td>
<td>93.8</td>
<td></td>
<td>-1.002</td>
<td>0.2</td>
<td>0.045</td>
<td>0.045</td>
<td>95.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Naive</td>
<td>-0.723</td>
<td>27.7</td>
<td>0.044</td>
<td>0.039</td>
<td>0.0</td>
<td></td>
<td>-0.790</td>
<td>21.0</td>
<td>0.040</td>
<td>0.033</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>-0.743</td>
<td>25.7</td>
<td>0.056</td>
<td>0.052</td>
<td>0.8</td>
<td></td>
<td>-0.826</td>
<td>17.4</td>
<td>0.051</td>
<td>0.045</td>
<td>6.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCS</td>
<td>-1.042</td>
<td>4.2</td>
<td>0.127</td>
<td>0.109</td>
<td>95.2</td>
<td></td>
<td>-1.029</td>
<td>2.9</td>
<td>0.093</td>
<td>0.081</td>
<td>91.8</td>
<td></td>
</tr>
<tr>
<td>$\eta_1(1)$</td>
<td>Ideal</td>
<td>1.003</td>
<td>0.3</td>
<td>0.067</td>
<td>0.068</td>
<td>95.4</td>
<td></td>
<td>1.002</td>
<td>0.2</td>
<td>0.070</td>
<td>0.070</td>
<td>96.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Naive</td>
<td>0.844</td>
<td>15.6</td>
<td>0.066</td>
<td>0.063</td>
<td>29.6</td>
<td></td>
<td>0.827</td>
<td>17.3</td>
<td>0.069</td>
<td>0.064</td>
<td>22.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>0.865</td>
<td>13.5</td>
<td>0.075</td>
<td>0.073</td>
<td>55.6</td>
<td></td>
<td>0.858</td>
<td>14.2</td>
<td>0.079</td>
<td>0.074</td>
<td>52.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCS</td>
<td>1.033</td>
<td>3.3</td>
<td>0.116</td>
<td>0.105</td>
<td>93.4</td>
<td></td>
<td>1.025</td>
<td>2.5</td>
<td>0.115</td>
<td>0.102</td>
<td>92.4</td>
<td></td>
</tr>
<tr>
<td>$\eta_2(-2)$</td>
<td>Ideal</td>
<td>-2.015</td>
<td>0.7</td>
<td>0.137</td>
<td>0.137</td>
<td>94.2</td>
<td></td>
<td>-2.002</td>
<td>0.1</td>
<td>0.134</td>
<td>0.142</td>
<td>96.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Naive</td>
<td>-1.682</td>
<td>15.9</td>
<td>0.134</td>
<td>0.123</td>
<td>29.2</td>
<td></td>
<td>-1.652</td>
<td>17.4</td>
<td>0.135</td>
<td>0.129</td>
<td>25.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>-1.722</td>
<td>13.9</td>
<td>0.156</td>
<td>0.149</td>
<td>51.4</td>
<td></td>
<td>-1.710</td>
<td>14.5</td>
<td>0.156</td>
<td>0.153</td>
<td>50.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCS</td>
<td>-2.061</td>
<td>3.0</td>
<td>0.246</td>
<td>0.216</td>
<td>93.4</td>
<td></td>
<td>-2.045</td>
<td>2.3</td>
<td>0.222</td>
<td>0.210</td>
<td>92.8</td>
<td></td>
</tr>
<tr>
<td>AR(1) $\sigma^2$</td>
<td>CS</td>
<td>0.4330</td>
<td>27.8</td>
<td>0.011</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.4354</td>
<td>27.4</td>
<td>0.011</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCS</td>
<td>0.6056</td>
<td>0.9</td>
<td>0.026</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.6057</td>
<td>0.9</td>
<td>0.025</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>AR(1) $\sigma^2(0.6)$</td>
<td>CS</td>
<td>0.5038</td>
<td>0.8</td>
<td>0.021</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.5034</td>
<td>0.7</td>
<td>0.020</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.6: Scenario III(c)(d): AR(1) Simulation results for $\rho = 0.5$ and $\sigma^2 = 0.6$ of AR(1) within-subject covariance. All the possible measured time points are at weeks of $\{0,2,4,6,8,10,12,14,16,24,32,40,48,56,64,72,80\}$ and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.

<table>
<thead>
<tr>
<th>$\rho = 0.5$ Method</th>
<th>Mean</th>
<th>RB(%)</th>
<th>SD</th>
<th>SE</th>
<th>CP(%)</th>
<th>Mean</th>
<th>RB(%)</th>
<th>SD</th>
<th>SE</th>
<th>CP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma(-1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>-1.001</td>
<td>0.1</td>
<td>0.052</td>
<td>0.052</td>
<td>95.4</td>
<td>-1.006</td>
<td>0.6</td>
<td>0.052</td>
<td>0.052</td>
<td>95.0</td>
</tr>
<tr>
<td>Naive</td>
<td>-0.710</td>
<td>29.0</td>
<td>0.042</td>
<td>0.038</td>
<td>0.0</td>
<td>-0.709</td>
<td>29.1</td>
<td>0.044</td>
<td>0.041</td>
<td>0.0</td>
</tr>
<tr>
<td>CS</td>
<td>-0.733</td>
<td>26.7</td>
<td>0.056</td>
<td>0.052</td>
<td>1.0</td>
<td>-0.728</td>
<td>27.2</td>
<td>0.058</td>
<td>0.054</td>
<td>0.6</td>
</tr>
<tr>
<td>GCS</td>
<td>-1.014</td>
<td>1.4</td>
<td>0.123</td>
<td>0.106</td>
<td>95.2</td>
<td>-1.033</td>
<td>3.3</td>
<td>0.125</td>
<td>0.110</td>
<td>94.4</td>
</tr>
<tr>
<td>$\eta_1(1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>1.000</td>
<td>0.0</td>
<td>0.069</td>
<td>0.067</td>
<td>94.0</td>
<td>1.009</td>
<td>0.9</td>
<td>0.068</td>
<td>0.068</td>
<td>94.2</td>
</tr>
<tr>
<td>Naive</td>
<td>0.838</td>
<td>16.2</td>
<td>0.067</td>
<td>0.062</td>
<td>29.0</td>
<td>0.852</td>
<td>14.8</td>
<td>0.066</td>
<td>0.062</td>
<td>32.6</td>
</tr>
<tr>
<td>CS</td>
<td>0.860</td>
<td>14.0</td>
<td>0.074</td>
<td>0.072</td>
<td>51.0</td>
<td>0.874</td>
<td>12.6</td>
<td>0.075</td>
<td>0.072</td>
<td>59.4</td>
</tr>
<tr>
<td>GCS</td>
<td>1.007</td>
<td>0.7</td>
<td>0.110</td>
<td>0.100</td>
<td>91.4</td>
<td>1.034</td>
<td>3.4</td>
<td>0.115</td>
<td>0.103</td>
<td>93.4</td>
</tr>
<tr>
<td>$\eta_2(-2)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>-1.999</td>
<td>0.1</td>
<td>0.122</td>
<td>0.135</td>
<td>97.6</td>
<td>-2.016</td>
<td>0.8</td>
<td>0.129</td>
<td>0.136</td>
<td>96.0</td>
</tr>
<tr>
<td>Naive</td>
<td>-1.673</td>
<td>16.4</td>
<td>0.127</td>
<td>0.122</td>
<td>27.8</td>
<td>-1.694</td>
<td>15.3</td>
<td>0.133</td>
<td>0.123</td>
<td>28.4</td>
</tr>
<tr>
<td>CS</td>
<td>-1.718</td>
<td>14.1</td>
<td>0.154</td>
<td>0.148</td>
<td>49.0</td>
<td>-1.735</td>
<td>13.3</td>
<td>0.156</td>
<td>0.148</td>
<td>56.2</td>
</tr>
<tr>
<td>GCS</td>
<td>-2.020</td>
<td>1.0</td>
<td>0.229</td>
<td>0.208</td>
<td>93.8</td>
<td>-2.050</td>
<td>2.5</td>
<td>0.232</td>
<td>0.212</td>
<td>93.4</td>
</tr>
<tr>
<td>AR(1) $\sigma^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>0.4318</td>
<td>28.0</td>
<td>0.011</td>
<td>-</td>
<td>-</td>
<td>0.4323</td>
<td>27.9</td>
<td>0.011</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma^2(0.6)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCS</td>
<td>0.5910</td>
<td>1.5</td>
<td>0.024</td>
<td>-</td>
<td>-</td>
<td>0.6036</td>
<td>0.6</td>
<td>0.026</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\rho(0.5)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCS</td>
<td>0.4924</td>
<td>1.5</td>
<td>0.020</td>
<td>-</td>
<td>-</td>
<td>0.5024</td>
<td>0.5</td>
<td>0.021</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 3.7: Scenario IV(a)(b): AR(1) Simulation results for $\rho = 0.75$ and $\sigma^2 = 0.6$ of AR(1) within-subject covariance. All the possible measured time points are at weeks of \{0, 2, 4, 6, 8, 10, 12, 14, 16, 24, 32, 40, 48, 56, 64, 72, 80\} and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.

<table>
<thead>
<tr>
<th>$\rho = 0.75$</th>
<th>Method</th>
<th>$\alpha_i$ Normal</th>
<th></th>
<th></th>
<th></th>
<th>$\alpha_i$ Bimodal mixture</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{(-1)}$</td>
<td>Ideal</td>
<td>-1.006</td>
<td>0.6</td>
<td>0.052</td>
<td>0.051</td>
<td>93.8</td>
<td>-1.002</td>
<td>0.2</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>Naive</td>
<td>-0.674</td>
<td>32.6</td>
<td>0.042</td>
<td>0.037</td>
<td>0.0</td>
<td>-0.751</td>
<td>24.9</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>-0.645</td>
<td>35.5</td>
<td>0.046</td>
<td>0.045</td>
<td>0.0</td>
<td>-0.750</td>
<td>25.0</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>GCS</td>
<td>-1.072</td>
<td>7.2</td>
<td>0.171</td>
<td>0.158</td>
<td>97.0</td>
<td>-1.046</td>
<td>4.6</td>
<td>0.115</td>
</tr>
<tr>
<td>$\eta_1(1)$</td>
<td>Ideal</td>
<td>1.003</td>
<td>0.3</td>
<td>0.067</td>
<td>0.068</td>
<td>95.4</td>
<td>1.002</td>
<td>0.2</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>Naive</td>
<td>0.819</td>
<td>18.1</td>
<td>0.066</td>
<td>0.062</td>
<td>18.8</td>
<td>0.795</td>
<td>20.5</td>
<td>0.069</td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>0.814</td>
<td>18.6</td>
<td>0.071</td>
<td>0.069</td>
<td>21.6</td>
<td>0.796</td>
<td>20.4</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>GCS</td>
<td>1.052</td>
<td>5.2</td>
<td>0.142</td>
<td>0.131</td>
<td>96.8</td>
<td>1.038</td>
<td>3.8</td>
<td>0.132</td>
</tr>
<tr>
<td>$\eta_2(-2)$</td>
<td>Ideal</td>
<td>-2.015</td>
<td>0.7</td>
<td>0.137</td>
<td>0.137</td>
<td>94.2</td>
<td>-2.002</td>
<td>0.1</td>
<td>0.134</td>
</tr>
<tr>
<td></td>
<td>Naive</td>
<td>-1.631</td>
<td>18.4</td>
<td>0.134</td>
<td>0.122</td>
<td>19.2</td>
<td>-1.589</td>
<td>20.5</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>-1.618</td>
<td>19.1</td>
<td>0.148</td>
<td>0.142</td>
<td>25.6</td>
<td>1.586</td>
<td>20.7</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>GCS</td>
<td>-2.098</td>
<td>4.9</td>
<td>0.268</td>
<td>0.268</td>
<td>95.4</td>
<td>-2.072</td>
<td>3.6</td>
<td>0.257</td>
</tr>
<tr>
<td>AR(1)</td>
<td>$\sigma^2$</td>
<td>CS</td>
<td>0.2652</td>
<td>55.8</td>
<td>0.008</td>
<td>-</td>
<td>-</td>
<td>0.2679</td>
<td>55.4</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2(0.6)$</td>
<td>GCS</td>
<td>0.6214</td>
<td>3.6</td>
<td>0.064</td>
<td>-</td>
<td>-</td>
<td>0.6162</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>$\rho(0.75)$</td>
<td>GCS</td>
<td>0.7560</td>
<td>0.8</td>
<td>0.023</td>
<td>-</td>
<td>-</td>
<td>0.7542</td>
<td>0.6</td>
</tr>
</tbody>
</table>
Table 3.8: **Scenario IV(c)(d): AR(1)** Simulation results for $\rho = 0.75$ and $\sigma^2 = 0.6$ of AR(1) within-subject covariance. All the possible measured time points are at weeks of {0, 2, 4, 6, 8, 10, 12, 14, 16, 24, 32, 40, 48, 56, 64, 72, 80} and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.

<table>
<thead>
<tr>
<th>$\rho = 0.75$ Method</th>
<th>$\gamma(-1)$ $\alpha_i$ Skew-normal</th>
<th>$\eta_1(1)$ $\alpha_i$ Bivariate $t_5$</th>
<th>$\eta_2(-2)$ $\alpha_i$ Skew-normal</th>
<th>AR(1) $\sigma^2$</th>
<th>$\sigma^2(0.6)$ GCS</th>
<th>$\rho(0.75)$ GCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideal</td>
<td>-1.001 0.1 0.052 0.052 95.4</td>
<td>1.000 0.0 0.069 0.067 94.0</td>
<td>-1.999 0.1 0.122 0.135 97.6</td>
<td>0.2638 56.1 0.008 -</td>
<td>0.5786 3.6 0.053 -</td>
<td>0.7395 1.4 0.023 -</td>
</tr>
<tr>
<td>Naive</td>
<td>-0.662 33.8 0.041 0.036 0.0</td>
<td>0.814 18.6 0.067 0.062 20.0</td>
<td>-1.624 18.8 0.128 0.121 16.0</td>
<td>0.2638 55.9 0.008 -</td>
<td>0.5786 2.9 0.064 -</td>
<td>0.7395 0.6 0.024 -</td>
</tr>
<tr>
<td>CS</td>
<td>-0.636 36.4 0.046 0.044 0.0</td>
<td>0.811 18.9 0.070 0.069 25.2</td>
<td>-1.617 19.1 0.146 0.140 24.0</td>
<td>0.2638 55.9 0.008 -</td>
<td>0.5786 2.9 0.064 -</td>
<td>0.7395 0.6 0.024 -</td>
</tr>
<tr>
<td>GCS</td>
<td>-1.009 0.9 0.144 0.125 92.8</td>
<td>1.006 0.6 0.123 0.109 92.8</td>
<td>-2.016 0.8 0.129 0.136 96.0</td>
<td>0.2643 55.9 0.008 -</td>
<td>0.6172 2.9 0.064 -</td>
<td>0.7545 0.6 0.024 -</td>
</tr>
</tbody>
</table>

The possible measured time points are at weeks of {0, 2, 4, 6, 8, 10, 12, 14, 16, 24, 32, 40, 48, 56, 64, 72, 80} and with 10% chance of missing after week 16. Mean: mean of Monte Carlo estimators; RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.
by Tsiatis and Davidian (2001), where $\sigma^2 = 0.3$, the value of SEs are almost that of SDs, especially when the random effects are from skew-normal distribution, while when $\sigma^2 = 0.6$, as presented by Song, Davidian and Tsiatis (2002b), the SEs are also somewhat slightly smaller than SDs with respect to CS approach.

On the other hand, the simulation studies also raise us a concern. When the statistical inference performance with respect to within-subject covariance is not good enough, in terms of bias and variation of the covariance parameter estimators, proposed GCS approach may still lead to certain biased inference, although significant improvement has been provided, comparing with the naive and CS approaches. Simulation results presented in Table 3.7 and Table 3.8, for example, where $\rho = 0.75$, except for the skew-normal distribution scenario, GCS estimators are with some slight bias due to the slightly poor statistical performance of $\sigma^2$ and $\rho$ for $\sigma^2 \rho^{|i-j|}$ of AR(1) of the lme R software output. When the $\rho$ of AR(1) is positive and relatively large, such as $\rho = 0.75$ in our case, and also the number of longitudinal measurements is not huge enough, it seems the variations between random effects and measurement error of linear mixed-effect models will be slightly mixed up numerically in algorithm. Thus, lme may be no longer able to capture well the AR(1) covariance, which may less statistical performance of GCS method, in terms of slightly larger relative bias and SDs/SEs, but along with larger coverage probability subsequently, as presented by Table 3.7 and Table 3.8. Furthermore, for the plug-in packages of softwares, the choices for covariance structure are limited and it may happen that none of them are suitable for the real application data and misspecification may lead to poor covariance inference. This concern motives us to model covariance with more flexibility and in a data-driven way.

### 3.6.2 Modified Cholesky Decomposition

A data-driven method of modified Cholesky decomposition approach is proposed to capture the within-subject covariance of longitudinal covariate process modelling with linear mixed effects model $W_i(t) = \alpha_{0i} + \alpha_{1i} t + \cdots + \alpha_{pi} t^p + \varepsilon_i(t)$. For each $i^{th}$ subject, given random effects $\alpha_i$, the within-subject error $\varepsilon_i = (\varepsilon_{i1}, \cdots, \varepsilon_{im})'$ are assumed from multivariate normal distribution with mean zero and a general covariance $\Sigma_i$. A modified Cholesky decomposition is adopted to model the positive definite covariance
matrix $\Sigma_i$. More specifically, we have $L_i\Sigma_i L_i' = D_i$, or $\Sigma_i^{-1} = L_i D_i^{-1} L_i'$ equally, where $L_i$ is a unique lower triangular matrix with ones as diagonal entries and $D_i$ is a unique diagonal matrix with positive diagonals. This modified Cholesky decomposition has a clear statistical interpretation, given random effects $\alpha_i$, the below-diagonal entries of $L_i$ are the negatives of the autoregressive coefficients, $\phi_{ijk}$, in the autoregressive model $W_{ij} = X_{ij} + \sum_{k=1}^{j-1} \phi_{ijk} (W_{ik} - X_{ik}) + \epsilon_{ij}$, where $X_{ij}/X_{ik}$ are the underlying values of longitudinal measurements $W_{ij}/W_{ik}$, and the diagonal entries of $D_i$ are the prediction error/innovation variance $\sigma^2_{ij} = \text{var}(\epsilon_{ij} | \alpha_i)$ (Pourahmadi, 1999). We model the unconstrained parameters $\phi_{ijk}$ and $\log \sigma^2_{ij}$ by

$$\phi_{ijk} = z'_{ijk} \zeta, \quad \log \sigma^2_{ij} = h'_{ij} \xi.$$  

Thus, we have the parameters of within-subject covariance $\Sigma_i$ denoted as $\theta_c = (\zeta, \xi)'$, and $z_{ijk}$ and $h_{ij}$ are covariates vectors, which may contain baseline covariates, polynomial terms in time, and their interactions. When we use polynomials in time to model the autoregressive coefficients and innovation variances, for example, the covariates may take the forms

$$z_{ijk} = (1, (t_{ij} - t_{ik}), (t_{ij} - t_{ik})^2, \cdots, (t_{ij} - t_{ik})^{q-1})',$$

$$h_{ij} = (1, t_{ij}, t_{ij}^2, \cdots, t_{ij}^{d-1})',$$

where it usually has $t_{ij} > t_{ik}$ since longitudinal measurements are taken in time order.

Likewise, GCS approach actually is treating random effects $\alpha_i$ as nuisance parameters, which are in the order of sample size and can be potentially of high dimension, hence the maximum likelihood estimation of $\theta_c$ tends to lead to biased estimation (Neyman and Scott 1948). Along with all available longitudinal observations, we turn to modified profile log-likelihood method instead, which is also called restricted log-likelihood for $\theta_c$,

$$\ell_{mp}(\theta_c) = -\frac{1}{2} \sum_{i=1}^{m} \log |\Sigma_i(\theta_c)| - \frac{1}{2} \sum_{i=1}^{m} \log |A_i' \Sigma_i^{-1}(\theta_c) A_i|$$

$$- \frac{1}{2} \sum_{i=1}^{m} ((W_i - A_i \tilde{\alpha}_i)' \Sigma_i^{-1}(\theta_c) (W_i - A_i \tilde{\alpha}_i)),$$

where $\Sigma_i = \text{Cov}(\varepsilon_i)$, $W_i = (W_{i1}, \cdots, W_{in})'$ and $\tilde{\alpha}_i = (A_i' \Sigma_i^{-1}(\theta_c) A_i)^{-1} A_i' \Sigma_i^{-1}(\theta_c) W_i$. The restricted log-likelihood function serves linear regression model $W_i = A_i \alpha_i + \varepsilon_i$.
given $\alpha_i, i = 1, \ldots, m$. We maximize the restricted log-likelihood with respect to $\theta_c$ to obtain the restricted maximum likelihood (REML) estimators and numerical methods are required. In the first instance, we have been trying the standard Newton-Raphson algorithm, however, due to the complexity of restricted log-likelihood function and that of Cholesky decomposition regression models as well as the potential highly dimensional subject-level random effects, our experimental experience tells that the Newton-Raphson algorithm does not have a good enough performance in capturing the REML estimators of $\theta_c = (\zeta, \xi)'$. Instead, numerical Nelder-Mead algorithm (Nelder and Mead, 1965) is utilised to obtain the REML estimators for $\theta_c$ and the sandwich rule is used to calculate the standard deviation of $\hat{\theta}_c$. The first and second derivatives of restricted log-likelihood function with respect to $\theta_c$ are provided in appendices of the chapter. By letting first derivatives equal zeroes, they are unbiased estimating equations under the assumption that observed data from different subjects are mutually independent, which satisfy the standard M-estimator arguments as presented by Carroll et al. (2006, A.6). Therefore, the restricted maximum likelihood estimators are consistent with asymptotic normal distribution under certain regular conditions, the simulation results presented in Table 3.9 and Table 3.10 also demonstrate such properties.

A simulation study is conducted carrying the same scenario settings of last session, except for the underlying covariance of measurement error. Instead of AR(1) structure, the underlying within-subject covariances are constructed by modified Cholesky decomposition. Orthogonal polynomial functions of form (3.10) are for $z_{ijk}$ and $h_{ij}$ with degree of $q = 4$ and $d = 4$, respectively, along with underlying values $\zeta = (0, -2, 2, -1)'$ and $\xi = (-0.5, 1, 1.5, -1)'$. Scenarios of four different random-effects distributions are again under our investigation and each scenario is conducted with 500 Monte Carlo datasets and 500 sample size for each. Again, both of MEst and MLE estimations of generalised conditional score approaches have been investigated, along with other aforementioned approaches. When the number of valid longitudinal observations is relatively large, $t_i \in \{0, 2, 4, 6, 8, 10, 12, 14, 16, 24, 32, 40, 48, 56, 64, 72, 80\}$ for example, that is, $9 \leq n_i \leq 17$, MEst and MLE estimations of GCS approach have very similar performance and provide nearly unbiased estimators and plausible statistical
CHAPTER 3. MEASUREMENT ERROR

inference as well. However, when some of $n_i$ are not large enough, due to the complexity of covariance modelling and also the conditional intensity, GCS approach of MLE estimation tends to have more reliable performance than that of MEst. Here we present simulation results of scenarios with smaller possible $n_i$, that is, potential measurements may be taken at $t_i \in \{0, 2, 4, 8, 16, 24, 32, 40, 48, 56, 64, 72, 80\}$ and we have $5 \leq n_i \leq 13$. Simulation results are presented in Table 3.9 and Table 3.10, inference results of five estimation approaches are demonstrated. Method GCS-ChoD indicates the modified Cholesky decomposition approach is utilised to capture the covariance of measurement error and the GCS of MEst approach for the survival parameters, that is, $K_i(t, \theta_s) = (S_i(t, \gamma, \Sigma_{i,t}), Z_i)'$ appears in the estimating equations (3.7). Method GCS-ChoD(mle) utilises the modified Cholesky decomposition for covariance but the GCS of MLE approach for survival parameters, that is, in equations (3.7), we propose $K_i(t, \theta_s) = (S_i(t, \gamma, \Sigma_{i,t}) - \gamma \sigma^2_{X_{i(t)gl}}), Z_i)'$ to further account for the adjusting variation term $\sigma^2_{X_{i(t)gl}}$ with respect to the $\gamma$ estimating equation.

Again, simulation results presented by Table 3.9 and Table 3.10 show that both naive and CS approaches can lead to substantial bias. Here, the statistical performance of CS approach goes down to the naive one and both lead to extremely poor coverage probability, subsequently. In the meantime, GCS approach of MEst tends to slightly overestimate the survival parameters, but GCS approach of MLE has a very good statistical performance and provides virtually unbiased estimates, and the coverage probabilities also show the validity of approach. When the underlying covariance matrix of $\varepsilon_i$ is of complexity, such as with variant variation, it may have more reliable inference by also accounting for the adjusting variation term $\sigma^2_{X_{i(t)gl}}$. For instance, a subject with longitudinal measurements taken at times $\{0,2,4,8,16\}$, the underlying covariance $\Sigma_i$ constructed by modified Cholesky decomposition is

$$
\begin{pmatrix}
0.7383 \\
0.9662 & 2.4785 \\
0.9634 & 0.9941 & 8.5965 \\
-0.9279 & -0.9247 & -0.9170 & 2.0243 \\
-0.7209 & -0.7502 & -0.7578 & 0.6005 & 3.2797
\end{pmatrix},
$$

where the diagonal entries are underlying $\sigma^2_{i}$s, off-diagonal are correlations and the elements of upper triangular matrix are omitted. Here the variation of random error,
var(ε_{ij}), is variant, while they are invariant within the AR(1) setting of last session. On the other hand, the proposed REML approach has pretty good statistical performance with respect to covariance parameters θ_c = (ζ, ξ)', though we also note that, regarding the survival parameters, the SEs seems again slightly underestimated in this simulation study and further improvement may be developed.

Like with many other numerical algorithms, a good initial value sometimes can largely improve the algorithm performance. In the simulation study, the naive estimators are used for the starting points for CS and GCS approaches for the survival inference. Regarding the initial values for (ζ, ξ)', we randomly generate values nearby the underlying values as starting points, say within the range of ±0.5. In practice, without the knowledge of underlying values, we can firstly set zeroes as the starting point, then use the converge values with some slight modifications as the new starting points, which will usually converge to the consistent root. In the meantime, the outputs with several different initial values can be used to check the convergence for the real data application.

### 3.7 Application to ACTG 175 Data

We apply the proposed generalized conditional score approach to ACTG 175 data, which include a total of 2467 HIV-1 infected adults with CD4 cell counts from 200 to 500 per cubic millimetre, recruited between December 1991 and October 1992 and the study ended in April 1995 (Hammer et al., 1996). Each participant was randomly assigned to one of the four regimens: zidovudine alone, zidovudine plus didanosine, zidovudine plus zalcitabine and didanosine alone to compare the treatment effects, and the primary end point was a ≥ 50 percent decline in the CD4 cell count, development of the AIDS, or death. Along with time to progression to the primary end point and some other baseline covariates, such as gender, weight and age at randomization, CD4 counts were also collected on each subject per protocol about every 12 weeks after randomization. It has been pointed out that the CD4 may perform as good surrogate marker of treatment effect but subject to considerable measurement error (Tsiatis, DeGruttola, and Wulfsohn, 1995). Likewise, we assume \( X_i(t) = \alpha_{0i} + \alpha_{1i} t \) represents inherent log_{10}CD4 count for subject \( i \) at time \( t \) and \( W_i(t) = X_i(t) + \varepsilon_i(t) \) for the
Table 3.9: **Scenario V: Modified Cholesky Decomposition** Simulation results for Cholesky decomposition of within-subject covariance. All the possible measured time points are at weeks of \{0,2,4,8,16,24,32,40,48,56,64,72,80\} and with 10% chance of missing after week 16. RB(%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval. GCS-ChoD, the MEst output of generalized conditional score estimation with Cholesky decomposition of within-subject covariance; GCS-ChoD(mle), the MLE output of generalized conditional score estimation with Cholesky decomposition of within-subject covariance.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Method</th>
<th>$\alpha_i$ Normal</th>
<th></th>
<th>$\alpha_i$ Bimodal mixture</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>Ideal</td>
<td>-1.006 0.6 0.051 0.051 95.6</td>
<td></td>
<td>-1.002 0.2 0.046 0.045 94.4</td>
<td></td>
</tr>
<tr>
<td>(-1)</td>
<td>Naive</td>
<td>-0.296 70.4 0.030 0.025 0.0</td>
<td></td>
<td>-0.405 59.5 0.027 0.018 0.0</td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>GCS-ChoD</td>
<td>-1.140 14.0 0.130 0.110 80.4</td>
<td></td>
<td>-1.058 5.8 0.089 0.078 92.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCS-ChoD(mle)</td>
<td>-1.019 1.9 0.093 0.081 93.2</td>
<td></td>
<td>-1.004 0.4 0.073 0.066 93.0</td>
<td></td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>Ideal</td>
<td>1.003 0.3 0.070 0.071 95.6</td>
<td></td>
<td>1.000 0.0 0.071 0.074 96.0</td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>Naive</td>
<td>0.618 38.2 0.065 0.062 0.0</td>
<td></td>
<td>0.526 47.4 0.067 0.063 0.0</td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>GCS-ChoD</td>
<td>1.087 8.7 0.128 0.110 87.2</td>
<td></td>
<td>1.052 5.2 0.113 0.104 91.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCS-ChoD(mle)</td>
<td>1.013 1.3 0.103 0.093 92.0</td>
<td></td>
<td>1.005 0.5 0.100 0.095 92.8</td>
<td></td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>Ideal</td>
<td>-2.006 0.3 0.138 0.136 94.8</td>
<td></td>
<td>-2.008 0.4 0.137 0.142 96.2</td>
<td></td>
</tr>
<tr>
<td>(-2)</td>
<td>Naive</td>
<td>-1.211 39.4 0.119 0.112 0.0</td>
<td></td>
<td>-1.055 47.3 0.135 0.117 0.0</td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>GCS-ChoD</td>
<td>-1.178 41.1 0.130 0.130 0.0</td>
<td></td>
<td>-1.024 48.8 0.177 0.150 0.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCS-ChoD(mle)</td>
<td>-2.170 8.5 0.257 0.217 88.0</td>
<td></td>
<td>-2.105 5.2 0.226 0.205 91.2</td>
<td></td>
</tr>
<tr>
<td>Modified</td>
<td>$\zeta_0$ (0)</td>
<td>0.000 - 0.001 0.001 95.4</td>
<td></td>
<td>0.000 - 0.001 0.001 95.8</td>
<td></td>
</tr>
<tr>
<td>Cholesky</td>
<td>$\zeta_1(-2)$</td>
<td>-1.999 0.1 0.029 0.029 94.0</td>
<td></td>
<td>-2.000 0.0 0.029 0.029 93.8</td>
<td></td>
</tr>
<tr>
<td>Decom-</td>
<td>$\zeta_2$ (2)</td>
<td>1.998 0.1 0.022 0.023 95.0</td>
<td></td>
<td>1.998 0.1 0.022 0.023 95.2</td>
<td></td>
</tr>
<tr>
<td>-position</td>
<td>$\zeta_3(-1)$</td>
<td>-0.998 0.2 0.035 0.034 94.6</td>
<td></td>
<td>-1.001 0.1 0.033 0.034 94.8</td>
<td></td>
</tr>
<tr>
<td>Inferences</td>
<td>$\xi_0(-0.5)$</td>
<td>-0.501 0.3 0.010 0.010 95.0</td>
<td></td>
<td>-0.502 0.3 0.010 0.010 92.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\xi_1$ (1)</td>
<td>0.990 1.0 0.111 0.107 93.8</td>
<td></td>
<td>0.987 1.3 0.102 0.107 95.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\xi_2$ (1.5)</td>
<td>1.499 0.1 0.119 0.119 95.2</td>
<td></td>
<td>1.494 0.4 0.123 0.118 94.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\xi_3$ (-1)</td>
<td>-1.005 0.5 0.123 0.118 94.2</td>
<td></td>
<td>-1.007 0.7 0.115 0.116 95.2</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.10: **Scenario V: Modified Cholesky Decomposition** Simulation results for Cholesky decomposition of within-subject covariance. All the possible measured time points are at weeks of \{0,2,4,8,16,24,32,40,48,56,64,72,80\} and with 10% chance of missing after week 16. RB(\%), the percentage of estimated relative bias; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP(\%), Monte Carlo coverage probability of 95% Wald confidence interval. GCS-ChoD, the MEst output of generalized conditional score estimation with Cholesky decomposition of within-subject covariance; GCS-ChoD(mle), the MLE output of generalized conditional score estimation with Cholesky decomposition of within-subject covariance.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Method</th>
<th>(\alpha_i) Skew-normal</th>
<th>(\alpha_i) Bivariate (t_5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\gamma)</td>
<td>Ideal</td>
<td>-1.004 0.4 0.054 0.052 95.0</td>
<td>-0.999 0.1 0.052 0.052 95.6</td>
</tr>
<tr>
<td>Naive</td>
<td>-0.293 70.7 0.028 0.024 0.0</td>
<td>-0.278 72.2 0.028 0.026 0.0</td>
<td></td>
</tr>
<tr>
<td>(-1) CS</td>
<td>-0.231 76.9 0.035 0.027 0.0</td>
<td>-0.214 78.6 0.032 0.028 0.0</td>
<td></td>
</tr>
<tr>
<td>GCS-ChoD</td>
<td>-1.131 13.1 0.133 0.113 85.4</td>
<td>-1.126 12.6 0.128 0.110 84.8</td>
<td></td>
</tr>
<tr>
<td>GCS-ChoD(mle)</td>
<td>-1.006 0.6 0.091 0.083 92.8</td>
<td>-1.002 0.2 0.090 0.082 93.2</td>
<td></td>
</tr>
<tr>
<td>(\eta_1)</td>
<td>Ideal</td>
<td>1.007 0.7 0.073 0.071 93.4</td>
<td>0.999 0.1 0.072 0.071 94.8</td>
</tr>
<tr>
<td>Naive</td>
<td>0.623 37.7 0.066 0.062 0.0</td>
<td>0.641 35.9 0.064 0.062 0.0</td>
<td></td>
</tr>
<tr>
<td>(1) CS</td>
<td>0.608 39.2 0.074 0.072 0.0</td>
<td>0.6252 37.5 0.071 0.070 0.00</td>
<td></td>
</tr>
<tr>
<td>GCS-ChoD</td>
<td>1.076 7.6 0.118 0.109 91.4</td>
<td>1.076 7.6 0.121 0.108 89.2</td>
<td></td>
</tr>
<tr>
<td>GCS-ChoD(mle)</td>
<td>1.006 0.6 0.095 0.092 95.0</td>
<td>1.004 0.4 0.098 0.091 94.6</td>
<td></td>
</tr>
<tr>
<td>(\eta_2)</td>
<td>Ideal</td>
<td>-2.014 0.7 0.131 0.135 96.0</td>
<td>-2.002 0.1 0.134 0.134 95.0</td>
</tr>
<tr>
<td>Naive</td>
<td>-1.235 38.3 0.122 0.111 0.0</td>
<td>-1.277 36.2 0.120 0.112 0.0</td>
<td></td>
</tr>
<tr>
<td>(-2) CS</td>
<td>-1.205 39.8 0.136 0.131 0.0</td>
<td>-1.248 37.6 0.135 0.130 0.0</td>
<td></td>
</tr>
<tr>
<td>GCS-ChoD</td>
<td>-2.161 8.1 0.235 0.214 90.6</td>
<td>-2.148 7.4 0.223 0.210 91.2</td>
<td></td>
</tr>
<tr>
<td>GCS-ChoD(mle)</td>
<td>-2.018 0.9 0.186 0.179 94.2</td>
<td>-2.008 0.4 0.179 0.176 94.4</td>
<td></td>
</tr>
<tr>
<td>Modified Cholesky</td>
<td>(\zeta_0(0))</td>
<td>0.000 – 0.001 0.001 92.6</td>
<td>0.000 – 0.001 0.001 95.0</td>
</tr>
<tr>
<td>(\zeta_1(-2))</td>
<td>-2.000 0.0 0.031 0.029 93.2</td>
<td>-1.999 0.1 0.029 0.029 95.2</td>
<td></td>
</tr>
<tr>
<td>Decom-(\zeta_2 (2))</td>
<td>1.999 0.1 0.023 0.023 95.0</td>
<td>1.998 0.1 0.023 0.023 94.4</td>
<td></td>
</tr>
<tr>
<td>-position (\zeta_3(-1))</td>
<td>-1.001 0.1 0.037 0.034 93.0</td>
<td>-0.997 0.3 0.035 0.034 93.4</td>
<td></td>
</tr>
<tr>
<td>Inferences</td>
<td>(\xi_0(-0.5))</td>
<td>-0.502 0.5 0.010 0.010 95.0</td>
<td>-0.501 0.3 0.010 0.010 95.2</td>
</tr>
<tr>
<td>(\xi_1(1))</td>
<td>0.990 1.0 0.108 0.108 95.6</td>
<td>0.987 1.3 0.107 0.107 95.4</td>
<td></td>
</tr>
<tr>
<td>(\xi_2(1.5))</td>
<td>1.497 0.2 0.128 0.120 94.0</td>
<td>1.493 0.5 0.118 0.119 95.6</td>
<td></td>
</tr>
<tr>
<td>(\xi_3(-1))</td>
<td>-1.009 0.9 0.122 0.119 95.0</td>
<td>-1.005 0.5 0.119 0.118 95.2</td>
<td></td>
</tr>
</tbody>
</table>
observed measurements, where \( \varepsilon_i(t) \) can be time-dependent or not. In this application analysis, we consider AIDS or death only as our primary event which leads to 308 events, and the hazard is modelled by proportional hazard model (3.2) with \( X_i(t) \) and some additional baseline covariates \( Z_i \) of age, gender, weight at randomization, for example.

We pre-analyse the longitudinal log_{10} CD4 data with the R command of \textit{lme} with REML method, and have tried to fit the within-subject covariance with structures of Independent, Compound Symmetry and AR(1). The fitted BIC values are -26451, -26441, -27881, respectively, that is, the linear mixed model with AR(1) within-subject covariance has smallest BIC amongst them and the corresponding \( \hat{\rho} = 0.35 \) when AR(1), which suggested that there may still exist some correlation among the within-subject errors after the linear mixed effects model fitting with random intercept and slope on log_{10}CD4. Meanwhile, the density plot for residuals is visibly symmetric with respect to origin and normally distributed, thus it seems reasonable to assume that \( \varepsilon_i \) is normally distributed with mean zero and a general covariance \( \Sigma_i \). Therefore, for the analysis of ACTG 175 data, we model the longitudinal log_{10}CD4 with AR(1) within-subject covariance. Furthermore, data-driven approach of modified Cholesky decomposition has been applied to data as well for the general case. The covariance inference for CS and GCS approaches are presented in the bottom panel of Table 3.11. Regarding the modified Cholesky decomposition approach, the covariates of \( z_{ijk} \) and \( h_{ij} \) are the orthogonal polynomial functions of form (3.10), we have tried both the degree pair of (\( q=3,d=3 \)) and (\( q=4,d=4 \)) and find that all the estimators of covariance parameters of \( \theta_c \) are significant for the former degree pair, however, for the later degree pair, the estimator of \( \zeta_3 \) is not significant and all others are significant, thus, we decide to choose the degree pair of (\( q=3,d=4 \)) in the end for the ACTG 175 data. Certain criteria can be applied to obtain the optimal polynomial degree pair (Pan and MacKenzie, 2003), however, this is not the focus of this study, so we choose it simply base its selection on its significance of regression effect. As discussed before, for the numerical algorithm, we firstly set zeroes as the starting points of \( \theta_c \), which is not necessary to converge to the optimal point of maximizing the modified profile log-likelihood function. We make some changes on suggested point, such as adding some noise, and re-set it as start point, then it usually converges to the optimal point.
In the meantime, we randomly choose three other different starting points to check convergence, and in the ACTG data analysis, they all converged to the same estimating values of \((\zeta, \xi)'\).

Regarding the survival analysis of the ACTG data, Song et al. (2002b) reported that, in the present of log_{10}CD4, treatment effect was not significant any more, which suggested that log_{10}CD4 might be proposed as surrogate marker. In addition, our GSC approach also indicates its non-significance. Therefore, our final hazard model is without the treatment covariate. Together with log_{10}CD4, we take some other baseline covariates of interest into the hazard model, such as gender, weight and age at randomization, and the analysis shows that the weight is not significant. Thus, the final hazard model we present here is with the longitudinal log_{10}CD4 covariate and age as well as the gender. Age has significant positive effect on the hazard rate, that is, the older the adults infected HIV-1, the higher possibility of experiencing AIDS or death. Out of the 2187 analysed participants, 1807 are male and 380 are female. The analysis shows the gender does not have significant effect on the hazard rate, but it is still kept in the final model to illustrated its non-significance. Our simulation studies show that the MEst approach of Cholesky decomposition might slightly overestimate the parameters, so we only present MLE approach for Cholesky decomposition here for the application. The final analysis outputs are presented in details in Table 3.11 and standard deviations are calculated with the sandwich rule as discussed in simulation study. It shows that GCS approaches, especially the one with Cholesky decomposition, provide stronger coefficients of log_{10}CD4 than that of CS approach which, as showed in simulation studies, tends to attenuate the regression coefficients if there does exist some positive correlation among the within-subject error. Meanwhile, we also notice that the estimators of GCS with AR(1) covariance are more like that of CS approach, but the coefficient of log_{10}CD4 of GCS approach with modified Cholesky decomposition of covariance are much stronger, which suggests that it should be more accurate to capture the covariance by using Cholesky decomposition regression method than that simply by AR(1) structure.
Table 3.11: **Statistical Inference for ACTG175 data** Analysis results are showed for the CS and GCS approaches, and for the GCS approach, the within-subject covariance was captured by AR(1) with *lme* R command and by the data-driven method of modified Cholesky decomposition. Both MEst and MLE estimations are presented for each approach, except for the GCS with Cholesky decomposition covariance. Values inside the parenthesis are the standard deviation of corresponding estimators. sd, standard deviation of estimators, ** indicates the significance of covariate of 95% Wald confidence interval.

<table>
<thead>
<tr>
<th>Statistical Inference for Survival Analysis</th>
<th>CS MEst</th>
<th>CS MLE</th>
<th>GCS-AR(1) MEst</th>
<th>GCS-AR(1) MLE</th>
<th>GCS-CholD MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log10cd4</td>
<td>-2.097 (0.114)</td>
<td>-2.090 (0.113)</td>
<td>-2.171 (0.111)</td>
<td>-2.158 (0.109)</td>
<td>-2.331 (0.144)**</td>
</tr>
<tr>
<td>age</td>
<td>0.023 (0.008)</td>
<td>0.023 (0.008)</td>
<td>0.023 (0.008)</td>
<td>0.023 (0.008)</td>
<td>0.023 (0.008)**</td>
</tr>
<tr>
<td>sexM</td>
<td>0.307 (0.192)</td>
<td>0.306 (0.191)</td>
<td>0.315 (0.195)</td>
<td>0.315 (0.194)</td>
<td>0.203 (0.181)</td>
</tr>
</tbody>
</table>

Inference for within-subject Covariance parameters

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Independent ( \Sigma_i = \sigma^2 I_{n_i} ): ( \sigma^2 = 0.11^2 )</th>
<th>AR(1) ( \Sigma_i = \sigma^2 \rho^{i-j} I_{n_i} ): ( \sigma^2 = 0.12^2, \rho = 0.35 )</th>
<th>ChoD ( \zeta_0 \ zeta_1 \ zeta_2 \ zeta_0 \ zeta_1 \ zeta_2 \ zeta_3 )</th>
<th>Estimator</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \zeta_0 ) \ zeta_1 \ zeta_2 \ zeta_0 \ zeta_1 \ zeta_2 \ zeta_3</td>
<td>( \zeta_0 ) \ zeta_1 \ zeta_2 \ zeta_0 \ zeta_1 \ zeta_2 \ zeta_3</td>
<td>( \zeta_0 ) \ zeta_1 \ zeta_2 \ zeta_0 \ zeta_1 \ zeta_2 \ zeta_3</td>
<td>( \zeta_0 ) \ zeta_1 \ zeta_2 \ zeta_0 \ zeta_1 \ zeta_2 \ zeta_3</td>
<td>( \zeta_0 ) \ zeta_1 \ zeta_2 \ zeta_0 \ zeta_1 \ zeta_2 \ zeta_3</td>
</tr>
</tbody>
</table>
3.8 Discussion

To date, when jointly modelling the longitudinal and survival data by the LMM and the Cox regression model, the random errors are commonly assumed to be independent and identically distributed from a normal distribution. For instance, the well-developed R package JM (Rizopoulos, 2010) can not adapt the covariance setting on the random errors for data analysis when a non-identity correlation structure is defined. It may be reasonable to assume normality on the random errors, however, the mutual independence is not necessarily always true. Our simulation studies also show it may lead to bias when the independence assumption is violated. Furthermore, inference for all survival regression parameters may be influenced by this violation, that is, bias may be introduced on regression parameters of both longitudinal and baseline covariates. In this study, apart from the random effects, the remaining correlation of the longitudinal measurements is further accounted for data analysis by proposing a non-identity covariance structure on random errors.

Due to the complexity of joint models, likelihood-based approach tends to be very computationally intensive, in terms of the integration requirement for the random effects and the survival function. Thus we utilise the conditional score estimator, which provides consistent and robust estimates for survival parameters and also has a huge computational advantage. However, we are also aware of that, likewise CS approach conducted and pointed out by Tsiatis and Davidian (2001), the proposed generalized conditional score approach will sacrifice some efficiency, since no assumption is made on the random effects, that is, the population feature of random effects has not been used for statistical inference.

In addition, statistical inference for the covariance structure of random errors and survival analysis is implemented by a “two-stage” approach. Statistical inference for the covariance structure is obtained first according to the repeated measurements of the longitudinal process, then we replace the unknown covariance by its estimate when attempting the inference for survival data. On one hand, when the pre-scheduled measurements of the longitudinal process are informatively truncated by the point process of event, some covariance information may be also contained in the survival data and the proposed “two-stage” approach will miss such information. On the other
hand, erroneous inference for the covariance structure will subsequently lead to biased estimates of the survival parameters. The proposed conditional score approach will be more attractive if we can model and infer the covariance structure of random errors and the survival data simultaneously. Furthermore, when analysing the ACTG 175 data, we simply use the rule of significance to choose the polynomial degree of time, when modelling the covariance structure by utilising the modified Cholesky decomposition, and the variable selection for the Cox regression model. However, it is still an open question with respect to the model and/or variable selection under the joint modelling framework of this chapter.

Finally, some caution may be needed when utilising the conditional score estimation approach, including the proposed generalised approach, for survival analysis under the joint modelling framework of longitudinal and survival data. An adjusting term of variation of sufficient statistic is introduced to the exponential power of the hazard function to assist the inference, which can be a little bit tricky sometimes since the exponential function is much more sensitive to the change of power (e.g. $e^2 << e^{10}$). Without proper inference for the covariance of random errors, the generalised conditional score approach may also lead to biased inference for the survival parameters. It should be worth utilising classical likelihood-based approach to further investigate the impact on survival analysis if the assumption of independence of random errors is violated.
3.9 Chapter appendices

Appendix A:
REML estimation for the within-subject covariance of the longitudinal process modelling with linear mixed effect model with modified Cholesky decomposition.

In this section, some notations are different from that of the paper, and the typical notations for longitudinal data/models and modified Cholesky decomposition are used.

Models and Methods

Let

\[ Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})' , \]

where \( Y_{ij} \) denote the \( j^{th} \) measurement for subject \( i \) with \( i = 1, \ldots, m \) and \( j = 1, \ldots, n_i \), and \( X_i \) denote all the covariates information for subject \( i \). Linear mixed effect models are proposed to fit the longitudinal process \( Y_i, Y_i = X_i\beta_i + \varepsilon_i, i = 1, 2, \ldots, m \),

\[ \varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in_i})' , \]

and \( \varepsilon_i|\beta_i \sim N(0, \Sigma_i) \).

Observed data from different subjects are independent. Our primary interest is to capture the within-subject covariance \( \Sigma_i \), which has to be a positive definite and symmetric matrix, and a data-driven method of modified Cholesky decomposition is proposed to characterize the covariance by regression modelling. That is, we have \( T_i\Sigma_iT_i' = D_i \), or equally \( \Sigma_i^{-1} = T_i'D_i^{-1}T_i \) where \( T_i \) is a unique lower triangular matrix with ones as diagonal entries and \(-\phi_{ijk} \) as the \((j,k)\)th entry if \( k < j \). And \( D_i \) is a unique diagonal matrix with positive diagonals, that is \( D_i = diag(\sigma_{i1}^2, \sigma_{i2}^2, \ldots, \sigma_{in_i}^2) \).

We model \( \phi_{ijk} \) and \( \log \sigma_{ij}^2 \) by,

\[ \phi_{ijk} = z'_{ijk}\zeta, \quad \log \sigma_{ij}^2 = h'_{ij}\xi, \]

where \( \zeta \) is a \( d \times 1 \) vector and \( \xi \) is a \( q \times 1 \) vector. Denote \( \theta = (\zeta, \xi)' \), the modified profile log-likelihood function after multiplying (-2), with respect to \( \theta = (\zeta, \xi)' \),

\[ \ell_{mp}(\theta) = \sum_{i=1}^{m} \log |\Sigma_i| + \sum_{i=1}^{m} \log |X_i'T_i\Sigma_i^{-1}X_i| + \sum_{i=1}^{m} \left( Y_i'T_i\Sigma_i^{-1}Y_i - Y_i'T_i\Sigma_i^{-1}X_i(X_i'T_i\Sigma_i^{-1}X_i)^{-1}X_i'T_i\Sigma_i^{-1}Y_i \right). \]
Then we have derivatives,

\[
\frac{\partial \ell_{mp}}{\partial \zeta} = 2 \sum_{i=1}^{m} \text{Trace} \left( (X_i^T \Sigma_i^{-1} X_i)^{-1} X_i^T \frac{\partial T_i^T}{\partial \zeta} D_i^{-1} T_i X_i \right) + 2 \sum_{i=1}^{m} Y_i^T A_i \frac{\partial T_i^T}{\partial \zeta} D_i^{-1} T_i A_i Y_i,
\]

\[
\frac{\partial \ell_{mp}}{\partial \xi} = \sum_{i=1}^{m} \text{Trace} (H_{id} B_i) - \sum_{i=1}^{m} Y_i^T A_i^T T_i^T D_i^{-1} H_{id} T_i A_i Y_i,
\]

where \( \zeta \) and \( \xi \) denote the \( s \)th element of \( \zeta \) and \( \ell \)th element of \( \xi \), respectively, \( H_{id} = \text{diag}(h_{i1\ell}, h_{i2\ell}, \cdots, h_{in\ell}) \), \( \text{Trace}(\cdot) \) is the trace function of a matrix, and that,

\[
A_i = I_{n_i} - X_i (X_i^T \Sigma_i^{-1} X_i)^{-1} X_i^T \Sigma_i^{-1},
\]

\[
B_i = I_{n_i} - T_i X_i (X_i^T \Sigma_i^{-1} X_i)^{-1} X_i^T T_i^T D_i^{-1} = T_i A_i T_i^{-1},
\]

where \( I_{n_i} \) is the \( n_i \times n_i \) identity matrix. Therefore, we have \( A_i X_i = 0 \), and \( B_i T_i X_i = 0 \).

\section*{Properties}

Firstly we have,

\[
E(Y_i Y_i^T | \beta_i) = \text{Var}(Y_i | \beta_i) + E(Y_i | \beta_i) E(Y_i | \beta_i)^T = \Sigma_i + X_i \beta_i \beta_i^T X_i^T.
\]

Denote

\[
\Psi_{i\zeta} = \frac{\partial \ell_{mpi}}{2 \partial \zeta}, \quad \Psi_{i\xi} = \frac{\partial \ell_{mpi}}{2 \partial \xi},
\]

then we have,

\[
E(\Psi_{i\zeta} | \beta_i) = \text{Trace} \left( (X_i^T \Sigma_i^{-1} X_i)^{-1} X_i^T \frac{\partial T_i^T}{\partial \zeta} D_i^{-1} T_i X_i \right) + \text{Trace} \left( A_i^T \frac{\partial T_i^T}{\partial \zeta} D_i^{-1} T_i A_i E \left( Y_i Y_i^T | \beta_i \right) \right)
\]

\[
= \text{Trace} \left( (X_i^T \Sigma_i^{-1} X_i)^{-1} X_i^T \frac{\partial T_i^T}{\partial \zeta} D_i^{-1} T_i X_i \right) + \text{Trace} \left( A_i^T \frac{\partial T_i^T}{\partial \zeta} D_i^{-1} T_i A_i \Sigma_i \right)
\]

\[
= \text{Trace} \left( \frac{\partial T_i^T}{\partial \zeta} D_i^{-1} T_i \Sigma_i \right) = \text{Trace} \left( T_i^{-1} \frac{\partial T_i^T}{\partial \zeta} \right) = 0,
\]

\[
2E(\Psi_{i\xi} | \beta_i) = \text{Trace} (H_{id} B_i) - \text{Trace} \left( A_i^T T_i^T D_i^{-1} H_{id} T_i A_i E \left( Y_i Y_i^T | \beta_i \right) \right)
\]

\[
= \text{Trace} (H_{id} B_i) - \text{Trace} \left( A_i^T T_i^T D_i^{-1} H_{id} T_i A_i \Sigma_i \right)
\]

\[
= \text{Trace}(H_{id}) - \text{Trace}(T_i^T D_i^{-1} H_{id} T_i \Sigma_i) = \text{Trace}(H_{id}) - \text{Trace}(H_{id} T_i \Sigma_i T_i^T D_i^{-1}) = \text{Trace}(H_{id}) - \text{Trace}(H_{id}) = 0.
\]
Therefore, we have,
\[ E(\Psi_{\xi_s}) = E\left( E(\Psi_{\xi_i}|\beta_i) \right) = 0, \]
\[ E(\Psi_{\xi_t}) = E\left( E(\Psi_{\xi_i}|\beta_i) \right) = 0. \]

Denote,
\[ \Psi_{\xi_s} = \frac{1}{m} \sum_{i=1}^{m} \left[ \text{Trace} \left( (X_i^T \Sigma_i^{-1} X_i)^{-1} X_i^T \frac{\partial T_i^T}{\partial \xi_s} D_i^{-1} T_i X_i \right) + Y_i^T A_i^T \frac{\partial T_i^T}{\partial \xi_s} D_i^{-1} T_i A_i Y_i \right], \]
\[ \Psi_{\xi_t} = \frac{1}{2m} \sum_{i=1}^{m} \left[ \text{Trace} \left( H_i B_i - Y_i^T A_i^T D_i^{-1} H_i T_i A_i Y_i \right) \right]. \]

The second derivatives,
\[
\frac{\partial \Psi_{\xi_s}}{\partial \xi_{s'}} = \frac{1}{m} \sum_{i=1}^{m} \left[ \text{Trace} \left( (X_i^T \Sigma_i^{-1} X_i)^{-1} X_i^T \frac{\partial T_i^T}{\partial \xi_{s'}} D_i^{-1} X_i \right) - \text{Trace} \left( (X_i^T \Sigma_i^{-1} X_i)^{-1} X_i^T \left( \frac{\partial T_i^T}{\partial \xi_{s'}} D_i^{-1} T_i + T_i^T D_i^{-1} \frac{\partial T_i}{\partial \xi_{s'}} \right) X_i (X_i^T \Sigma_i^{-1} X_i)^{-1} X_i^T D_i^{-1} T_i X_i \right) + Y_i^T A_i^T \frac{\partial T_i^T}{\partial \xi_{s'}} D_i^{-1} T_i A_i Y_i + Y_i^T A_i^T \frac{\partial T_i^T}{\partial \xi_{s'}} D_i^{-1} (\frac{\partial T_i}{\partial \xi_{s'}} A_i + T_i \frac{\partial A_i}{\partial \xi_{s'}}) Y_i \right],
\]
\[
\frac{\partial \Psi_{\xi_t}}{\partial \xi_s} = \frac{\partial \Psi_{\xi_t}}{\partial \xi_s} = \frac{1}{m} \sum_{i=1}^{m} \left[ Y_i^T A_i^T \left( T_i^T D_i^{-1} \frac{\partial T_i}{\partial \xi_s} + \frac{\partial T_i^T}{\partial \xi_s} D_i^{-1} T_i \right) X_i (X_i^T \Sigma_i^{-1} X_i)^{-1} X_i^T D_i^{-1} H_i T_i A_i Y_i \right.
\]
\[
- \text{Trace} \left( (X_i^T \Sigma_i^{-1} X_i)^{-1} X_i^T \frac{\partial T_i^T}{\partial \xi_s} D_i^{-1} B_i H_i T_i X_i \right) - Y_i^T A_i^T \frac{\partial T_i^T}{\partial \xi_s} D_i^{-1} H_i T_i A_i Y_i \right],
\]
\[
\frac{\partial \Psi_{\xi_t}}{\partial \xi_{s'}} = \frac{1}{2m} \sum_{i=1}^{m} \left[ \text{Trace} \left( H_i T_i X_i (X_i^T \Sigma_i^{-1} X_i)^{-1} X_i^T D_i^{-1} H_i T_i A_i Y_i + Y_i^T A_i^T D_i^{-1} H_i T_i A_i Y_i \right) \right],
\]

where we have,
\[
\frac{\partial A_i}{\partial \xi_{s'}} = -X_i (X_i^T \Sigma_i^{-1} X_i)^{-1} X_i^T \left( \frac{\partial T_i^T}{\partial \xi_{s'}} D_i^{-1} T_i + T_i^T D_i^{-1} \frac{\partial T_i}{\partial \xi_{s'}} \right) A_i,
\]
\[
\frac{\partial A_i}{\partial \xi_{s'}} = X_i (X_i^T \Sigma_i^{-1} X_i)^{-1} X_i^T D_i^{-1} H_i T_i A_i.
\]

Denote \( \Psi_{\xi}(\theta) = (\Psi_{\xi_1}, \cdots, \Psi_{\xi_m})' \) and \( \Psi_{\xi}(\theta) = (\Psi_{\xi_1}, \cdots, \Psi_{\xi_m})' \), then we have the estimating equations
\[
U(\theta) = \begin{pmatrix} \Psi_{\xi}(\theta) \\ \Psi_{\xi}(\theta) \end{pmatrix} = 0
\]
are unbiased. And $\hat{\theta}$, the solution of $U(\theta) = 0$, are the consistent estimators of $\theta$ and are asymptotic normal distributed under certain regular conditions. The usual sandwich rule can be used to calculate the standard deviation of the estimators.

Appendix B:

Proportional hazards model for survival analysis modelling with time-independent and time-dependent longitudinal covariate measured with correlated error.

Along with the time-to-event data and some other baseline covariates information, measurements on a longitudinal process were also collected. Our primary interest is to characterize the relationship between time-to-event and covariates which include the current value of the time-dependent longitudinal process. It is popular to capture the relationship with proportional hazards model, of which the implementation requires the complete knowledge of the trajectory of the longitudinal process. However, the longitudinal process are usually measured intermittently and often with measurement error. Linear mixed effect model is a standard framework to capture the continuous longitudinal process. Let $W_i$ denote all the available longitudinal measurements for each subject $i$ and $W_i = (W_{i1}, W_{i2}, \cdots, W_{in_i})'$, where $W_{ij}$ is the $j^{th}$ measurement for subject $i$ at time $t_{ij}$, then we propose the linear mixed effect model

$$W_i = A_i \alpha_i + \epsilon_i,$$

where $A_i$ is a $n_i \times p$ design matrix, $\alpha_i$ is the random effect, a subject-level coefficient and given $\alpha_i$, we have $\epsilon_i \sim N(0, \Sigma_i)$. Regarding the survival data, let $V_i$ denote the observed survival time, which can be potential right-censored, and we have that $V_i = \min(T_i, C_i)$ where $T_i$ is the underlying time-to-event and $C_i$ is the underlying censor time and we assume the censor is independent of the time-to-event $T_i$. And indicator $\delta_i = I(T_i \leq C_i)$ denotes the occurrence of event or censoring. The proportional hazards model is proposed to characterize the relationship between hazard and covariates,

$$\lambda_i(t) = \lim_{dt \to 0} dt^{-1} Pr(t \leq T_i < t + dt | T_i \geq t, Z_i, C_i, X_i(t))$$

$$= \lambda_0(t) \exp(\gamma X_i(t) + \eta' Z_i),$$
where $X_i(t) = \hat{\beta}\alpha_i = \alpha_{0i} + \alpha_{1i}t + \cdots + \alpha_{pi}t^p$ and $W_i(t) = X_i(t) + \epsilon_i(t)$, and $\epsilon_i(t)$ denotes the measurement error for $X_i(t)$ at time $t$ which can be time-dependent or not. Therefore, we have the observed data $(V_i, \delta_i, W_i, Z_i, n_i, t_i, A_i)$ where $t_i = (t_{i1}, \cdots, t_{in_i})'$. All the observed data from different subjects are assumed to be independent, and given $X_i(t)$, $V_i$ and $W_i$ are mutually independent. Our primary interest is on the survival parameters $\theta_s = (\gamma, \eta)'$ and the within-subject covariance parameters $\theta_c$ to characterize the covariance $\Sigma_i$ with $L_i\Sigma_i L_i' = D_i$, or equally $\Sigma_i^{-1} = L_iD_i^{-1}L_i'$, and $L_i$ is a unique lower triangular matrix with ones as diagonal entries and $-\phi_{ijk}$ as the $(j, k)$th entry if $k < j$, and $D_i$ is a unique diagonal matrix with positive diagonals. Covariance is captured with the regression model,

$$
\phi_{ijk} = z_{ijk}'\zeta, \quad \log \sigma_{ij}^2 = h_{ij}'\xi,
$$

where $\zeta$ is a $d \times 1$ vector and $\xi$ is a $q \times 1$ vector, and $\theta_c = (\zeta, \xi)'$. The inference will be for $\theta = (\theta_s, \theta_c)'$.

### Score Functions of Survival Analysis

The generalized least square estimators for $X_i(t)$ is $\hat{X}_i(t) = \hat{\beta}(A_i'\Sigma_i^{-1}A_i)^{-1}A_i\Sigma_i^{-1}W_{i,t}$ and the corresponding variance is $\sigma_{\hat{X}_i(t)}^2 = \hat{\beta}(A_i'\Sigma_i^{-1}A_i)^{-1}\tilde{\eta}^t$, then, accordingly, the sufficient statistic for $X_i(t)$ is,

$$
S_i(t, \gamma, \Sigma_i,t) = \gamma\sigma_{\hat{X}_i(t)}^2dN_i(t) + \hat{X}_i(t).
$$

We have the conditional intensity process, give the sufficient statistic $S_i(t, \gamma, \Sigma_i,t),

$$
\lambda_i(t) = \lambda_0(t) \exp \left\{ \gamma S_i(t, \gamma, \Sigma_i,t) - \frac{1}{2}\gamma^2\sigma_{\hat{X}_i(t)}^2 + \eta'Z_i \right\} Y_i(t) \\
= \lambda_0(t) \exp \left\{ \gamma S_i(t, \gamma, \Sigma_i,t) - \frac{1}{2}\gamma^2\hat{\beta}(A_i'\Sigma_i^{-1}A_i)^{-1}\tilde{\eta} + \eta'Z_i \right\} Y_i(t),
$$

where $Y_i(t) = I(V_i \geq t, t_{i(p+1)} \leq t)$ is the at risk process. And denote

$$
E_0(t, \gamma, \eta, \Sigma_i,t) = \sum_{i=1}^m E_{0i}(t) = \sum_{i=1}^m \exp \left\{ \gamma S_i(t, \gamma, \Sigma_i,t) - \frac{1}{2}\gamma^2\hat{\beta}(A_i'\Sigma_i^{-1}A_i)^{-1}\tilde{\eta} + \eta'Z_i \right\} Y_i(t),
$$

$$
E_1(t, \gamma, \eta, \Sigma_i,t) = \sum_{i=1}^m E_{1i}(t) \\
= \sum_{i=1}^m K_i(t, \theta) \exp \left\{ \gamma S_i(t, \gamma, \Sigma_i,t) - \frac{1}{2}\gamma^2\hat{\beta}(A_i'\Sigma_i^{-1}A_i)^{-1}\tilde{\eta} + \eta'Z_i \right\} Y_i(t).
$$
CHAPTER 3. MEASUREMENT ERROR

And the score function,

$$U_1(\theta) = \sum_{i=1}^{m} \int \left\{ K_i(t, \theta) - \frac{E_i(t, \gamma, \eta, \Sigma_i, t)}{E_0(t, \gamma, \eta, \Sigma_i, t)} \right\} dN_i(t) = 0,$$

where $K_i(t, \theta)$ is a vector and will be detailed later on.

Method1: MEst proposed by Tsiatis & Davidian (2001)

let $K_i^a(t, \theta) = \begin{pmatrix} S_i(t) \\ Z_i \end{pmatrix}$, then we have

$$E_i^a(t, \gamma, \eta, \Sigma_i, t) = \sum_{i=1}^{m} E_i^a(t) = \sum_{i=1}^{m} \begin{pmatrix} S_i(t) \\ Z_i \end{pmatrix} \exp \left\{ \gamma S_i(t) - \frac{1}{2} \gamma^2 \left( A_i^t \Sigma_i^{-1} A_i \right)^{-1} \tilde{t} + \eta' Z_i \right\} Y_i(t).$$

The differentiation of the score function, which is a $p_1 \times p_1$ matrix, where $p_1$ is the dimension of $\theta_s$,

$$\frac{\partial U_1(\theta)}{\partial \theta_s'} = \sum_{i=1}^{m} \int \left( -\sum_{i=1}^{m} \frac{K_i^a(t, \theta)}{E_0(t, \gamma, \eta, \Sigma_i, t)} \frac{E_i^a(t)}{E_0(t, \gamma, \eta, \Sigma_i, t)} + \frac{E_i^a(t, \gamma, \eta, \Sigma_i, t)}{E_0(t, \gamma, \eta, \Sigma_i, t)^2} \right) dN_i(t).$$

Denote $D(t, \zeta, \xi) = \tilde{t}' \left( A_i^t \Sigma_i^{-1} A_i, t \right)^{-1} \tilde{t} = \text{Trace} \left( \left( A_i^t \Sigma_i^{-1} A_i, t \right)^{-1} \tilde{t} \tilde{t}' \right)$, and we have,

$$\frac{\partial D(t, \zeta, \xi)}{\partial \zeta_s} = -\text{Trace} \left( \left( A_i^t \Sigma_i^{-1} A_i, t \right)^{-1} \tilde{t} \left( \frac{\partial T_i^t}{\partial \zeta_s} D_i^{-1} T_i + T_i D_i^{-1} \frac{\partial T_i^t}{\partial \zeta_s} D_i^{-1} \tilde{t} \right) A_i, t \left( A_i^t \Sigma_i^{-1} A_i, t \right)^{-1} \tilde{t} \right),$$

$$\frac{\partial D(t, \zeta, \xi)}{\partial \xi_t} = \text{Trace} \left( \left( A_i^t \Sigma_i^{-1} A_i, t \right)^{-1} A_i, t \left( T_i D_i^{-1} H_i T_i \right) A_i, t \left( A_i^t \Sigma_i^{-1} A_i, t \right)^{-1} \tilde{t} \right),$$

where $\zeta_s$ is the $s^{th}$ element of the $\zeta$, and $\xi_t$ is the $t^{th}$ element of $\xi$. Denote $\frac{\partial D(t, \zeta, \xi)}{\partial \theta}$ = \left( \frac{\partial D(t, \zeta, \xi)}{\partial \zeta}, \frac{\partial D(t, \zeta, \xi)}{\partial \xi} \right)'$, then the differentiation of the score function, which is a $p_1 \times p_2$ matrix, where $p_2 = d + q$,

$$\frac{\partial U_1(\theta)}{\partial \theta_c} = \sum_{i=1}^{m} \int \left( -\sum_{i=1}^{m} \frac{K_i^a(t, \theta)}{E_0(t, \gamma, \eta, \Sigma_i, t)} \frac{\partial D(t, \zeta, \xi)}{\partial \theta} E_0(t) + \frac{E_i^a(t, \gamma, \eta, \Sigma_i, t)}{E_0(t, \gamma, \eta, \Sigma_i, t)^2} E_0(t) \right) dN_i(t).$$

Method2: MLE

As for the maximum likelihood estimate, we have

$$K_i^b(t, \theta) = \begin{pmatrix} S_i - \gamma \tilde{t} \left( A_i^t \Sigma_i^{-1} A_i, t \right)^{-1} \tilde{t} \\ Z_i \end{pmatrix} = \begin{pmatrix} S_i - \gamma D(t, \zeta, \xi) \\ Z_i \end{pmatrix}. $$
CHAPTER 3. MEASUREMENT ERROR

And

\[ E^b_1(t, \gamma, \eta, \Sigma_{i,t}) = \sum_{i=1}^{m} E^b_i(t) \]

\[ = \sum_{i=1}^{m} \left( S_i(t) - \gamma D(t, \zeta, \xi) \right) \exp \left\{ \gamma S_i(t) - \frac{1}{2} \gamma^2 D(t, \zeta, \xi) + \eta' Z_i \right\} Y_i(t). \]

The differentiation of the score function, which is a \( p_1 \times p_1 \) matrix,

\[
\frac{\partial U_1(\theta)}{\partial \theta_s} = \sum_{i=1}^{m} \int \left( \begin{pmatrix} -D(t, \zeta, \xi) & 0 \\ 0 & 0 \end{pmatrix} - \sum_{i=1}^{m} \begin{pmatrix} -D(t, \zeta, \xi) & 0 \\ 0 & 0 \end{pmatrix} \right) \frac{E_0(t)}{E_0(t, \gamma, \eta, \Sigma_{i,t})} \frac{E_0(t, \gamma, \eta, \Sigma_{i,t})'}{E_0(t, \gamma, \eta, \Sigma_{i,t})^2} \right) dN_i(t),
\]

and the differentiation with respect to the covariance parameters \( \theta_c = (\zeta', \xi')' \),

\[
\frac{\partial U_1(\theta)}{\partial \theta_c'} = \sum_{i=1}^{m} \int \left( \begin{pmatrix} -\gamma \frac{\partial D(t, \zeta, \xi)'}{\partial \theta} & 0 \\ 0 & 0 \end{pmatrix} - \sum_{i=1}^{m} \begin{pmatrix} -\gamma \frac{\partial D(t, \zeta, \xi)'}{\partial \theta} & 0 \\ 0 & 0 \end{pmatrix} \right) \frac{E_0(t)}{E_0(t, \gamma, \eta, \Sigma_{i,t})} \frac{E_0(t, \gamma, \eta, \Sigma_{i,t})'}{E_0(t, \gamma, \eta, \Sigma_{i,t})^2} \right) dN_i(t).
\]

Second Derivatives

The first derivatives refer to the corresponding estimating score functions. Let \( U_1(\theta) \) denote the score function for the survival analysis, and \( U_2(\theta_c) \) that for the covariance of the Cholesky decomposition, then we have

\[
\partial \begin{pmatrix} U_1(\theta) \\ U_2(\theta_c) \end{pmatrix} / \partial \begin{pmatrix} \theta_s' \\ \theta_c' \end{pmatrix} \rightarrow \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix},
\]

where \( U_{11}, U_{12} \) and \( U_{22} \) are the corresponding derivatives, and \( \theta_s \) denotes the survival parameters \((\gamma', \eta')'\) and \( \theta_c \) denote the covariance parameters \((\zeta', \xi')'\). And the sandwich rule will be used to calculate the standard deviations for parameters.
Appendix C:
Sufficient Statistic and Conditional Intensity

In this session, we would like to give the details for finding the sufficient statistic and conditional intensity function. Models are proposed by

\[
W_i(t_{ij}) = \alpha_0i + \alpha_1t_{ij} + \alpha_2t_{ij}^2 + \cdots + \alpha_pt_{ij}^p + \varepsilon_{ij},
\]

\[
\lambda_i(t) = \lambda_0(t) \exp(\gamma X_i(t) + \eta^T Z_i) \quad i = 1, 2, \ldots, m; \quad j = 1, 2, \ldots, n_i,
\]

where \(X_i(t) = \alpha_0i + \alpha_1t + \alpha_2t^2 + \cdots + \alpha_pt^p\) and \(\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in})' \sim N(0, \Sigma_i)\).

Denote \(\alpha_i = (\alpha_{0i}, \alpha_{1i}, \ldots, \alpha_{pi})'\) and the generalised least square estimator of \(X_i(t)\)

\[
\hat{X}_i(t)_{\text{gls}} = \hat{\theta}(A'_{i,t} \Sigma^{-1}_{i,t} A_{i,t})^{-1} A'_{i,t} \Sigma^{-1}_{i,t} W_{i,t},
\]

where \(A_{i,t}\) is the design matrix with \(\ell\)th column \((\tilde{t}_i(t))'\), where \(\tilde{t}_i(t) = (t_{i1}, \ldots, t_{is})'\) with \(t_{is} \leq t < t_{i(s+1)}\). Thus, we have

\[
E(\hat{X}_i(t)_{\text{gls}}|\alpha_i) = \hat{\theta}(A'_{i,t} \Sigma^{-1}_{i,t} A_{i,t})^{-1} A'_{i,t} \Sigma^{-1}_{i,t} \alpha_i = \hat{\theta} \alpha_i,
\]

\[
\text{Var}(\hat{X}_i(t)_{\text{gls}}|\alpha_i) = \hat{\theta}(A'_{i,t} \Sigma^{-1}_{i,t} A_{i,t})^{-1} A'_{i,t} \Sigma^{-1}_{i,t} \Sigma^{-1}_{i,t} A_{i,t} (A'_i \Sigma^{-1}_{i,t} A_{i,t})^{-1} \hat{\theta} = \hat{\theta}(A'_{i,t} \Sigma^{-1}_{i,t} A_{i,t})^{-1} \hat{\theta} = \sigma^2_{\hat{X}_i(t)_{\text{gls}}}.
\]

The estimator \(\hat{X}_i(t)_{\text{gls}}\) conditional on \(\{\alpha_i, \tilde{t}_i(t), Y_i(t) = 1, Z_i\}\) is normally distributed with mean \(X_i(t) = \sum_{l=0}^p \alpha_l t^l\) and variance \(\hat{\theta}(A'_{i,t} \Sigma^{-1}_{i,t} A_{i,t})^{-1} \hat{\theta}\), where \(\hat{\theta} = (1, t, \ldots, t^p)\).

For now, we assume that the \(\Sigma^{-1}_{i,t}\) or \(\Sigma_{i,t}\) is known,

\[
\{dN_i(t) = r, \hat{X}_i(t)_{\text{gls}} = x|Y_i(t) = 1, \alpha_i, W_i(t_{ij})s\} \text{ taken up to and including time } t \text{ at } \tilde{t}_i(t)\}
\]

\[
= \Pr(dN_i(t) = r|Y_i(t) = 1, \hat{X}_i(t)_{\text{gls}} = x, \alpha_i, Z_i, \tilde{t}_i(t)) \Pr(\hat{X}_i(t)_{\text{gls}} = x|Y_i(t) = 1, \alpha_i, Z_i, \tilde{t}_i(t))
\]

\[
= \left[\lambda_0(t) dt \exp(\gamma X_i(t) + \eta^T Z_i)\right]' \left[1 - \lambda_0(t) dt \exp(\gamma X_i(t) + \eta^T Z_i)\right]^{-r} \frac{\exp\left(-\frac{(x-X_i(t))}{2\sigma_{\hat{X}_i(t)_{\text{gls}}}}^2\right)}{(2\pi\sigma_{\hat{X}_i(t)_{\text{gls}}})^{1/2}}.
\]

Thus the conditional likelihood of \(\{dN_i(t) = r, \hat{X}_i(t)_{\text{gls}}\}\) given \(\{Y_i(t) = 1, \alpha_i, Z_i, \tilde{t}_i(t)\}\), up to order \(dt\), is

\[
\left[\lambda_0(t) dt \exp(\gamma X_i(t) + \eta^T Z_i)\right]^{dN_i(t)} \frac{1}{(2\pi\sigma_{\hat{X}_i(t)_{\text{gls}}})^{1/2}} \exp\left(-\frac{\hat{X}_i(t)_{\text{gls}}^2}{2}\right) \exp\left(-\frac{\hat{X}_i(t)_{\text{gls}}^2 - 2\hat{X}_i(t)_{\text{gls}} X_i(t) + X_i(t)^2}{2\sigma^2_{\hat{X}_i(t)_{\text{gls}}}}\right)
\]

\[
= \exp\left(X_i(t) \left(\gamma dN_i(t) + \frac{\hat{X}_i(t)_{\text{gls}}}{\sigma_{\hat{X}_i(t)_{\text{gls}}}}\right) \frac{[\lambda_0(t) dt \exp(\eta Z_i)]^{dN_i(t)} (2\pi\sigma_{\hat{X}_i(t)_{\text{gls}}})^{1/2}}{\sigma_{\hat{X}_i(t)_{\text{gls}}}} \exp\left(-\frac{\hat{X}_i(t)_{\text{gls}}^2 + X_i(t)^2}{2\sigma^2_{\hat{X}_i(t)_{\text{gls}}}}\right)\right).
\]
CHAPTER 3. MEASUREMENT ERROR

This representation implies that, conditional on \( Y_i(t) = 1 \),

\[
S_i(t, \gamma, \Sigma_{i,t}) = \gamma \sigma^2_{X_i(t)_{gls}} dN_i(t) + \tilde{X}_i(t)_{gls}
\]
is a complete sufficient statistic for \( \alpha_i \) and denoted by \( S_i \) for short.

Equation (3.12) turns out to be,

\[
\frac{[\lambda_0(t) dt \exp(\eta' Z_i)]^{dN_i(t)}}{(2\pi \sigma^2_{\tilde{X}_i(t)_{gls}})^{1/2}} \exp \left( -\frac{S_i^2 - 2S_i \gamma \sigma^2_{X_i(t)_{gls}} dN_i(t) + \gamma^2 \sigma^4_{X_i(t)_{gls}} dN_i(t) + X_i(t)^2}{2\sigma^2_{\tilde{X}_i(t)_{gls}}} + \frac{X_i(t) S_i}{\sigma^2_{\tilde{X}_i(t)_{gls}}} \right) = \frac{[\lambda_0(t) dt \exp(\eta' Z_i)]^{dN_i(t)}}{(2\pi \sigma^2_{\tilde{X}_i(t)_{gls}})^{1/2}} \exp \left( -\frac{S_i^2}{2\sigma^2_{\tilde{X}_i(t)_{gls}}} + \gamma S_i dN_i(t) - \frac{1}{2} \gamma^2 \sigma^2_{X_i(t)_{gls}} dN_i(t) + \frac{2S_i X_i(t) - X_i(t)^2}{2\sigma^2_{\tilde{X}_i(t)_{gls}}} \right)
\]

\[
= \lambda_0(t) dt \exp(\eta' Z_i)^{dN_i(t)} \exp \left( -\frac{S_i^2}{2\sigma^2_{\tilde{X}_i(t)_{gls}}} + \gamma S_i dN_i(t) - \frac{1}{2} \gamma^2 \sigma^2_{X_i(t)_{gls}} dN_i(t) \right) \kappa_0
\]

\[
= Pr \left( dN_i(t), \tilde{X}_i(t)_{gls} | Y_i(t) = 1, \alpha_i, Z_i, \tilde{t}_i(t) \right),
\]

where \( \kappa_0 = \exp \left( \frac{2S_i X_i(t) - X_i(t)^2}{2\sigma^2_{\tilde{X}_i(t)_{gls}}} \right) \). Thus we have

\[
Pr \left( dN_i(t) = 1 | S_i(t, \gamma, \Sigma_{i,t}), Z_i, \tilde{t}_i(t), Y_i(t) = 1 \right) = \frac{Pr \left( dN_i(t) = 1, S_i | Z_i, \tilde{t}_i(t), Y_i(t) = 1 \right)}{Pr \left( dN_i(t) = 1, S_i | Z_i, \tilde{t}_i(t), Y_i(t) = 1 \right) + Pr \left( dN_i(t) = 0, S_i | Z_i, \tilde{t}_i(t), Y_i(t) = 1 \right)}
\]

Thus we have the numerator

\[
um = \lambda_0(t) dt \exp(\eta' Z_i) \exp \left( -\frac{S_i^2}{2\sigma^2_{\tilde{X}_i(t)_{gls}}} + \gamma S_i - \frac{1}{2} \gamma^2 \sigma^2_{X_i(t)_{gls}} \right) \int \kappa_0 Pr(\alpha_i | \cdots) d\alpha_i,
\]

and the denominator

\[
dem = \num + \exp \left( -\frac{S_i^2}{2\sigma^2_{\tilde{X}_i(t)_{gls}}} \right) \int \kappa_0 Pr(\alpha_i | \cdots) d\alpha_i,
\]

where \( Pr(\alpha_i | \cdots) \) is the density function for random effects \( \alpha_i \). Therefore we have,

\[
Pr \left( dN_i(t) = 1 | S_i(t, \gamma, \Sigma_{i,t}), Z_i, \tilde{t}_i(t), Y_i(t) = 1 \right) \approx \lambda_0(t) dt \exp(\eta' Z_i) \exp \left( -\frac{S_i^2}{2\sigma^2_{\tilde{X}_i(t)_{gls}}} + \gamma S_i - \frac{1}{2} \gamma^2 \sigma^2_{X_i(t)_{gls}} \right)
\]

\[
\approx \lambda_0(t) dt \exp(\eta' Z_i) \exp \left( \gamma S_i - \frac{1}{2} \gamma^2 \sigma^2_{X_i(t)_{gls}} \right) = \lambda_0(t) \exp \left( \gamma S_i - \frac{1}{2} \gamma^2 \sigma^2_{X_i(t)_{gls}} + \eta' Z_i \right) dt.
\]

Thus, the conditional intensity function turns out to be

\[
\lambda_i(t | S_i(t, \gamma, \Sigma_{i,t})) \approx \lambda_0(t) \exp \left( \gamma S_i(t, \gamma, \Sigma_{i,t}) - \frac{1}{2} \gamma^2 \sigma^2_{X_i(t)_{gls}} + \eta' Z_i \right) Y_i(t).
\]
Appendix D: Pooled Estimator for Unknown $\sigma^2$

If we define $W_i = A_i \alpha_i + \varepsilon_i$, where $A_i$ is the $n_i \times (p + 1)$ design matrix with $\ell^{th}$ column $t_i^{\ell}$, where $t_i = (t_{i1}, t_{i2}, \ldots, t_{in_i})^T$ and $\ell = 0, 1, \ldots, p$, and $\hat{\sigma}^2$ is the solution to the estimating equation

\[ \sum_{i=1}^{m} I(n_i > (p + 1)) \left( (W_i - A_i \hat{\alpha}_i)^T (W_i - A_i \hat{\alpha}_i) - (n_i - p - 1)\sigma^2 \right) = 0, \]

where $\hat{\alpha}_i = (A_i^T A_i)^{-1} A_i^T W_i = \alpha_i + (A_i^T A_i)^{-1} A_i^T \varepsilon_i$. As

\[ (W_i - A_i \hat{\alpha}_i)^T (W_i - A_i \hat{\alpha}_i) = \varepsilon_i^T (I_{n_i} - A_i (A_i^T A_i)^{-1} A_i^T) \varepsilon_i, \]

that this is an unbiased estimating equation for $\sigma^2$ follows by noting that

\[
E(I(n_i > (p + 1))((W_i - A_i \hat{\alpha}_i)^T (W_i - A_i \hat{\alpha}_i) - (n_i - p - 1)\sigma^2)) \\
= E(E[I(n_i > (p + 1))((W_i - A_i \hat{\alpha}_i)^T (W_i - A_i \hat{\alpha}_i) - (n_i - p - 1)\sigma^2)|T_i, C_i, \alpha_i, Z_i, t_i, n_i]) \\
= E(E[I(n_i > (p + 1))(\varepsilon_i^T (I_{n_i} - A_i (A_i^T A_i)^{-1} A_i^T) \varepsilon_i - (n_i - p - 1)\sigma^2)|T_i, C_i, \alpha_i, Z_i, t_i, n_i]).
\]

Since, given $E(\varepsilon_i|T_i, C_i, \alpha_i, Z_i, t_i, n_i) = 0$ and $Var(\varepsilon_i|T_i, C_i, \alpha_i, Z_i, t_i, n_i) = \sigma^2 I_{n_i}$,

\[
E[I(n_i > (p + 1))(\varepsilon_i^T (I_{n_i} - A_i (A_i^T A_i)^{-1} A_i^T) \varepsilon_i)|T_i, C_i, \alpha_i, Z_i, t_i, n_i] \\
= I(n_i > (p + 1)) \left( E[\varepsilon_i^T (I_{n_i} - A_i (A_i^T A_i)^{-1} A_i^T) \varepsilon_i|T_i, C_i, \alpha_i, Z_i, t_i, n_i] \right) \\
= I(n_i > (p + 1)) \left( tr\{ (I_{n_i} - A_i (A_i^T A_i)^{-1} A_i^T) E(\varepsilon \varepsilon^T|T_i, C_i, \alpha_i, Z_i, t_i, n_i) \} \right) \\
= I(n_i > (p + 1)) \left( tr\{ (I_{n_i} - A_i (A_i^T A_i)^{-1} A_i^T) \sigma^2 I_{n_i} \} \right) \\
= I(n_i > (p + 1)) \left( (n_i - tr\{ A_i (A_i^T A_i)^{-1} A_i^T) \}) \sigma^2 \right) \\
= I(n_i > (p + 1)) \left( (n_i - p - 1) \sigma^2 \right),
\]

showing that the inner conditional expectation is zero, which indicates that the estimating equation is unbiased, thus, $\hat{\sigma}^2$ is consistent estimate of $\sigma^2$.

However, if $E(\varepsilon_i|T_i, C_i, \alpha_i, Z_i, t_i, n_i) = 0$ and $Var(\varepsilon_i|T_i, C_i, \alpha_i, Z_i, t_i, n_i) = \Sigma_i$, then,

\[
E[I(n_i > (p + 1))(\varepsilon_i^T (I_{n_i} - A_i (A_i^T A_i)^{-1} A_i^T) \varepsilon_i)|T_i, C_i, \alpha_i, Z_i, t_i, n_i] \\
= I(n_i > (p + 1)) \left( tr\{ \Sigma_i \} - tr\{ (A_i^T A_i)^{-1} A_i^T \Sigma_i A_i \} \right).
\]

Hence if $\Sigma_i \neq \sigma^2 I_{n_i}$, the estimate $\hat{\sigma}^2$ is often biased or indicates different interpretation.
Chapter 4

Survival analysis with cumulative information of longitudinal covariates

4.1 Introduction

In terms of joint modelling of longitudinal and survival data, to date, it is popular to either link them by introducing current/present value of longitudinal process to the survival Cox regression model (Tsiatis and Davidian, 2004, for instance) or by proposing the shared random-effects models (Henderson et al., 2000, 2002) to investigate the association between time to event and the underlying longitudinal process, which is usually unobserved or unobservable and the complete knowledge of the process is somewhat unavailable but is required accordingly by the statistical inference implementation of survival models. In more recent studies, investigators have recognized that both past and current levels of longitudinal process, blood pressure, for example, may have instructive impact on the present hazards of heart disease and stroke mortality (Boshuizen et al., 2007). And it is reasonable that if a change in the risk factor would change the disease rate importantly, the analysis should not only relate disease hazards to the present levels of risk factors at the time of event, death, for example, but may also relate to that of a period before or some other periods (Clarke et al., 2001). It may lead to bias and/or under estimation of the impact of levels of longitudinal process on the risk hazards by failing to account for the past information.
in the analysis of data of such longitudinal study.

Instead of the standard default current value models, in the study of this chapter, the trajectory information of longitudinal process is proposed to introduce to the Cox regression model to investigate relationship between the point process of events in time and longitudinal process. An weighted integration of trajectory information is introduced to attempt the impact of longitudinal process on the point process of events in time. More specifically,

\[ \int_{0}^{t} \gamma(t-u)m_{i}(u)du, \]

where \( m_{i}(\cdot) \) is the trajectory of longitudinal process, that is, the underlying values of longitudinal process, \( \gamma(\cdot) \) is the corresponding weighted function, and \( \gamma(t-u) \) indicates that it is only a function of time lag. The levels of longitudinal profile at different past periods may have different influence on the present risk hazards. As illustrated by Figure 4.1, for instance, the weight for current levels of underlying longitudinal process may be fixed and being a constant, but it tends to be weaker as time lag increases. Figure 4.1(a) represents the scenario of longitudinal process having positive effects on the risk hazards, while Figure 4.1(b) represents that of negative effects. And \( u \) represents the time point of past and \( t-u \) is the corresponding time lag to present. Different curves illustrate the different possible weighted curves of \( \gamma(t-u) \).

Figure 4.1: Possible shapes for weighted curves \( \gamma(t-u) \)

![Figure 4.1](image)

(a) positive effects  (b) negative effects

Without integration, statistical inference for the models with covariate (4.1) seems to turn to the problems which are similar to those arising in statistical inference for the Cox regression models with time-varying coefficient \( \gamma(t) \) for \( m_{i}(t) \), where many
well-established methods have been proposed, such as by simply proposing piecewise function to parametrize the coefficient or the local linear polynomial approach. It is also popular and classical to utilise a B-splines approach to characterize the time-varying coefficient. However, with integration, all these approaches seem to be not appropriate and more details of discussion on this issue can be found in section 4.8 of this chapter.

In the study of this chapter, an exponential function is proposed for the weighted curve \( \gamma(t - u) \), more specifically,

\[
\gamma(t - u) = ae^{-b(t-u)}, \quad 0 \leq u \leq t, b > 0, \text{ and } a \in \mathbb{R},
\]

(4.2)

where \( a \) and \( b \) are parameters of interest and \( b \) measures impact of time dependence. Alternatively, the weighted function \( \gamma(\cdot) \) can be represented by

\[
\gamma(u) = ae^{-bu}, \quad 0 \leq u \leq t, b > 0, \text{ and } a \in \mathbb{R},
\]

(4.3)

as illustrated by Figure 4.2, where the x-axis denotes the time lag to present.

Figure 4.2: Possible shapes for time-lag weighted curves \( \gamma(u) \)

(a) positive effects

(b) negative effects

The justification for choosing exponential function for the weighted effects \( \gamma(\cdot) \) is that, the effects of past levels of longitudinal process on the present risk hazard rate may tend to be weaker as time lag increases, an underlying assumption on the impact pattern of longitudinal process on the time-to-event process, which may be commonly seen in practice. It is also the motivation for proposing a Gaussian process for serially correlated components, as proposed by Henderson et al.(2002), where the exponential function is used to characterize the correlation of the stationary Gaussian process.
CHAPTER 4. CUMULATIVE INFORMATION

4.2 Models

For simplicity, a single longitudinal process is considered in the study and linear mixed-effects models with polynomial function of time are proposed for the longitudinal measurements. The framework should be applicable to extend to generalised linear mixed-effects models and/or a multivariate longitudinal process. Denote \( Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{im})^T \) the longitudinal measurements for the \( i^{th} \) subject of \( m \) taken at pre-specified time points \( t_i = (t_{i1}, t_{i2}, \ldots, t_{im})^T \), respectively, and \( i = 1, 2, \ldots, m \).

Some other covariates may have been collected as well during the study and denoted by \( X_i \), which may be time-dependent or time-independent, but they are completely observable and exact values have been collected. However, the time-independent baseline covariates are particularly discussed here. Linear mixed-effects model is proposed for the longitudinal process

\[
Y_i(t) = \alpha_{0i} + \alpha_{1i}t + \cdots + \alpha_{qi}t^q + X_i^T \beta + \varepsilon_i(t), \quad i = 1, 2, \cdots, m, \tag{4.4}
\]

where random effects \( \alpha_i = (\alpha_{0i}, \alpha_{1i}, \cdots, \alpha_{qi})^T \) are assumed to be normally distributed with mean \( \alpha \) and covariance matrix \( G \). Unknown parameter \( \beta \) is \( p \times 1 \) vector of regression coefficients of covariates \( X_i \). And \( \varepsilon_i(t) \) is the random error for \( i^{th} \) subject at time \( t \) and \( \varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{im})^T \) are random errors of longitudinal measurements of \( i^{th} \) subject. It is routine to assume \( \varepsilon_i \sim N(0, \Sigma_i) \) and they are mutually independent of random effect \( \alpha_i \)s. As discussed by chapter 3, covariance \( \Sigma_i \) can be any positive definite covariance matrix, however, for simplicity without loss of generality, it is assumed to be \( \sigma^2 I_{ni} \) in this study, where \( I_{ni} \) denotes the identity matrix of size \( ni \). In matrix form, model (4.4) is denoted by

\[
Y_i = A_i \alpha_i + 1_{ni} X_i^T \beta + \varepsilon_i, \quad i = 1, 2, \cdots, m, \tag{4.5}
\]

where \( 1_{ni} \) is the ones vector of size \( ni \), covariates \( X_i \)s are measured at baseline and/or time-independent variables. When time-dependent covariates are included, special care may need in calculating the cumulative term. Denote \( m_i(t) = \alpha_{0i} + \alpha_{1i}t + \cdots + \alpha_{qi}t^q + X_i^T \beta \) the underlying value of longitudinal process at time \( t \) of \( i^{th} \) subject.

Regarding the point process of events in time, denote \((T_i, \delta_i, Z_i)\) the corresponding observed data of \( i^{th} \) subject, \( i = 1, 2, \cdots, m \), where \( T_i = \min(\hat{T}_i, C_i) \), \( \delta_i = I(\hat{T}_i \leq C_i) \), and \( \hat{T}_i \) and \( C_i \) denote the underlying time to event and censoring time, respectively.
The standard framework of Cox regression model is proposed to investigate the relationship of time to event with covariates. And a cumulative term accounting for the trajectory information of longitudinal process is proposed by

\[ \int_{t_L}^{t} \gamma(t-u)m_i(u)du = \int_0^{t_0(t)} \gamma(u)m_i(t-u)du, \]  

(4.6)

where \( t_L \in \{0, t-c, \max(0, t-c)\} \) represents the scenarios that the cumulative information is calculated ever since the baseline, or only certain last period ever since time \( t-c \), or certain last period but no earlier than the baseline, thus, \( t_0(t) \in \{t, c, \min(t, c)\} \) subsequently. Therefore, given the assumption of linear mixed-effects model (4.4) for longitudinal process, formula (4.6) accounting for the cumulative information of longitudinal process can be expanded by

\[
\begin{align*}
\int_0^{t_0(t)} & \gamma(u)m_i(t-u)du \\
= & \int_0^{t_0(t)} \gamma(u) \left( \sum_{\ell=0}^{q} \alpha_{\ell i}(t-u)^{\ell} \right) du + X_i^T \beta \int_0^{t_0(t)} \gamma(u)du \\
= & \sum_{\ell=0}^{q} \alpha_{\ell i} \left( \int_0^{t_0(t)} \gamma(u)(t-u)^{\ell}du \right) + X_i^T \beta \int_0^{t_0(t)} \gamma(u)du \\
= & \sum_{\ell=0}^{q} \alpha_{\ell i} \left( \int_0^{t_0(t)} \gamma(u) \left( \sum_{k=0}^{\ell} C_{\ell}^k (-1)^k u^k \right)du \right) + X_i^T \beta \int_0^{t_0(t)} \gamma(u)du \\
= & \sum_{\ell=0}^{q} \sum_{k=0}^{\ell} \alpha_{\ell i} (-1)^k C_{\ell}^k \int_0^{t_0(t)} \gamma(u)u^kdu + X_i^T \beta \int_0^{t_0(t)} \gamma(u)du \\
= & \int_0^{t_0(t)} \gamma(u)du + \sum_{k=1}^{q} \frac{(-1)^k C_{\ell}^k (t)}{k!} m_i^{(k)}(t) \int_0^{t_0(t)} \gamma(u)u^kdu, \\
\end{align*}
\]

(4.7)

where \((t-u)^{\ell} = \sum_{k=0}^{\ell} C_{\ell}^k (-1)^k u^k\), \(C_{\ell}^k = \frac{\ell!}{(\ell-k)!k!}\) and \(m_i^{(k)}(t)\) denotes the \(k\)th derivative of \(m_i(t)\) with respect to time \(t\). Equation (4.7) indicates that, by taking into account the cumulative information, the regression coefficient of current value \(m_i(t)\) of longitudinal process turns out to be \(\int_0^{t_0(t)} \gamma(u)du\) in the Cox regression model. Interestingly, if the boundary \(t_0(t)\) is a function of \(t\), such as by proposing \(t_0(t) = t\), current value of longitudinal process has a time-varying effect on the present hazard. The current levels of blood pressure, for instance, might have time-varying effects on the present risk of cardiovascular event, that is, the coefficient of current value
might be time-dependent. Furthermore, \( m_i^{(k)}(t) \) is the \( k \)th derivative of \( m_i(t) \) with respect to time \( t \), that is, the cumulative term (4.6) takes into account all other features of trajectory. When the LMM models only include the random intercept and random slope, \( m_i(t) = \alpha_{i0} + \alpha_{i1}t + X_i^T \beta \), for example, then \( q = 1 \), \( m_i^{(1)}(t) = \alpha_{i1} \), and the second part of the equation (4.7) turns out to be \( -\alpha_{i1} \int_0^{t_0(t)} u \gamma(u) du \), linear combination of random slope, which, to some extent, explains why a model with current value and random slope sometimes may provide better statistical inference than a model without covariate of random slope does (e.g., Henderson et al., 2000, among others). Equation (4.7) shows that, by taking into account the cumulative term, it actually has been taking into account all the trajectory features of longitudinal process, the current value and steepness, for instance.

As mentioned in this chapter’s introduction, the exponential function is proposed for the weighted curve, i.e., \( \gamma(u) = ae^{-bu} \), \( u \geq 0 \), thus the regression coefficient of current value of longitudinal process in the Cox regression model is \( \int_0^{t_0(t)} ae^{-bu} du \), and the cumulative term (4.6) subsequently turns out to be

\[
\int_0^{t_0(t)} \gamma(u)m_i(t - u)du = \int_0^{t_0(t)} ae^{-bu} \left( \sum_{\ell=0}^{q} \alpha_{\ell i}(t - u)^\ell + X_i^T \beta \right) du \\
= a \left( \sum_{\ell=0}^{q} \alpha_{\ell i} \int_0^{t_0(t)} e^{-bu}(t - u)^\ell du + X_i^T \beta \int_0^{t_0(t)} e^{-bu} du \right).
\]

Denote \( B(a, b, t) = (B_0(a, b, t), B_1(a, b, t), \ldots, B_q(a, b, t))^T \) and \( B(b, t) = (B_0(b, t), B_1(b, t), \ldots, B_q(b, t))^T \), and we have \( B_\ell(a, b, t) = aB_\ell(b, t) \), where

\[
B_\ell(b, t) = \int_0^{t_0(t)} e^{-bu}(t - u)^\ell du \\
= \sum_{\nu=1}^{\ell+1} (-1)^\nu \frac{\ell!}{b^\nu (\ell + 1 - \nu)!} \left( e^{-b(t_0(t)\ell+1-\nu) - t^{\ell+1-\nu}} \right),
\]

in which the theorem of integration by parts is utilised, and \( t_0(t) \) is denoted by \( t_0 \) for
short. Define \( 0^\theta = 1 \), we have \( B_0(b, t) = -\frac{1}{b}(e^{-bt} - 1) \), even when \( t_0 = t \). Thus,

\[
\int_0^{t_0(t)} \gamma(u)m_i(t-u)\,du \\
= a \left( \sum_{q=0}^{\ell} \beta_i \sum_{\nu=1}^{\ell+1} \frac{(-1)^\nu \ell!}{b^\nu (\ell + 1 - \nu)!} \left( e^{-bt_0(t-t_0)} \ell + 1 - \nu \right) + \frac{1}{b}(1 - e^{-bt_0(t-t_0)} X^T_i \beta) \right) \\
= B(a, b, t)^T \alpha_i + B_0(a, b, t) X^T_i \beta = a \left( B(b, t)^T \alpha_i + B_0(b, t) X^T_i \beta \right). 
\]

Therefore, by accounting for the cumulative information of longitudinal process, Cox regression model for the hazard rate may be proposed by

\[
\lambda_i(t) = \lim_{dt \to 0} P(r(t) \leq \tilde{T}_i < t + dt | \tilde{T}_i \geq t, \bar{m}_i(t), C_i, \alpha_i, \bar{r}_i(t)) \\
= \lambda_0(t) \exp \left( \int_{1_t}^{t} \gamma(t-u)m_i(u)\,du + Z^T_i \eta_i \right) \\
= \lambda_0(t) \exp \left( \int_{t_0(t)}^{t} \gamma(u)m_i(t-u)\,du + Z^T_i \eta_i \right) \\
= \lambda_0(t) \exp \left( B(a, b, t)^T \alpha_i + B_0(a, b, t) X^T_i \beta + Z^T_i \eta_i \right), 
\]

where \( \lambda_0(t) \) is an arbitrary unspecific non-negative function, referred to baseline hazard function. And \( m_i(t) = \{ m_i(u) : u \leq t \} \) denotes the complete trajectory information of longitudinal process at and before time \( t \) of the \( i^{th} \) subject, \( \bar{r}_i(t) = \{ \bar{r}_i(t_{ij}) : t_{ij} \leq t \} \) and \( C_i \) is the underlying censoring time and the non-informative censoring is assumed as well in the study, as indicated by the model.

Alternatively, the current value \( m_i(t) \) of longitudinal process may be being particularly of interest. Furthermore, as discussed earlier, when the cumulative term is proposed, the effect of current value on the hazard rate turns out to be \( \int_0^{t_0(t)} \gamma(u)\,du \), for example, it is \( \int_0^{t_0(t)} ae^{-bu} \,du \) in our study, which is obviously dependent on the choice of \( \gamma(u) \) function and decreasing rate \( b \) will affect the initial effect parameter \( a \) as well, also the interpretation may not be directly straightforward. Therefore we would like to propose an alternative Cox regression model with slight change of model (4.9), that is, the effect of the longitudinal process on the point process of events in time is accounted for firstly by the current value, then the remaining effect is accounted for by the cumulative term. More specifically,

\[
\lambda_i(t) = \lambda_0(t) \exp \left( \gamma(0)m_i(t) + \int_{0}^{t_0(t)} \gamma(u)m_i(t-u)\,du + Z^T_i \eta_i \right), 
\]
where \( \gamma(0) \) is proposed to keep the continuity of the weighted effects on hazard rate.

Cumulative term has somewhat accounted for certain effects of the current levels \( m_i(t) \), due to features of integration, current levels in first and second terms actually play different roles, thus, it can assure identifiability of the model, our numerical experiments also have shown such property. Therefore, by utilising an exponential function for the weighted curve, we have that \( \gamma(0) = a \) and that the Cox regression model turns out to be

\[
\lambda_i(t) = \lim_{dt \to 0} Pr(t \leq \hat{T}_i < t + dt | \hat{T}_i \geq t, \tilde{m}_i(t), Z_i, C_i, \alpha_i, \bar{\varepsilon}_i(t))
\]

\[= \lambda_0(t) \exp \left( \gamma(0)m_i(t) + \int_0^{t_0(t)} \gamma(u)m_i(t-u)du + Z_i^T \eta \right) \]

\[
= \lambda_0(t) \exp \left( B^c(a, b, t)^T \alpha_i + B^c_0(a, b, t)X_i^T \beta + Z_i^T \eta \right),
\]

where \( B^c(a, b, t) = (B^c_0(a, b, t), B^c_1(a, b, t), \ldots, B^c_q(a, b, t))^T \) and that

\[
B^c_\ell(a, b, t) = \gamma(0) + \int_0^{t_0(t)} \gamma(u)du \]

which indicates that the regression of coefficient of current value on the event hazard rate turns out to be \( \gamma(0) + \int_0^{t_0(t)} \gamma(u)du \), that is, \( a(1 + \int_0^{t_0(t)} e^{-bu}du) \) of the case of exponential weighted curve for \( \gamma(u) \).

In this study, the joint models are in terms of joint modelling the model (4.4) and the model (4.9) or model (4.10). In particular, the statistical inference in the next of this chapter are based on the joint models of model (4.4) and model (4.10), and only a slight change will make the statistical inference suitable for the joint models of model (4.4) and model (4.9) as well.

### 4.3 Statistical inference

Maximum likelihood estimation approach tends to give robust estimates, thus it is utilised here to develop statistical inference for the joint models of model (4.4) and
model (4.9) or model (4.10). Some traditional assumptions are made here, for instance, longitudinal process and point process of events in time are assumed to be mutually independent given random effects $\alpha_i$ and common covariates. The visiting process is pre-specified and independent of the values of measurements taken. And the underlying censoring process of participants are assumed to be independent, that is, non-informative censoring process. Observed data and random effects from different subjects are assumed to be mutually independent as well. Thus, likelihood function for the joint models turns out to be

$$L(\theta; \text{data}) = \int Pr(\text{data}, \text{random effects}|\theta)d(\text{random effects})$$

$$= \int Pr(\text{data}|\text{random effects}, \theta)Pr(\text{random effects}|\theta)d(\text{random effects})$$

$$= \int \left( \prod_{i=1}^{m} Pr(T_i, \delta_i, Y_i | \alpha_i, \theta) \right) \left( \prod_{i=1}^{m} Pr(\alpha_i|\theta) \right) d\alpha_1d\alpha_2 \ldots d\alpha_m$$

$$= \prod_{i=1}^{m} \int Pr(T_i, \delta_i|\alpha_i, \theta)Pr(Y_i|\alpha_i, \theta)Pr(\alpha_i|\theta)d\alpha_i,$$

where $\theta$ denotes the parameters vector of joint models, and $\text{data}$ denotes observed data from both processes of all participants. Covariates are suppressed here for short. And the log-likelihood function,

$$\ell(\theta; \text{data}) = \sum_{i=1}^{m} \log \left( \int Pr(T_i, \delta_i|\alpha_i, \theta)Pr(Y_i|\alpha_i, \theta)Pr(\alpha_i|\theta)d\alpha_i \right). \tag{4.11}$$

In more details, we have the complete data likelihood function of $i^{th}$ subject be

$$Pr(T_i, \delta_i|\alpha_i, \theta)Pr(Y_i|\alpha_i, \theta)Pr(\alpha_i|\theta)$$

$$= \left( \lambda_i(T_i) \right)^{\delta_i} S_i(T_i) (2\pi)^{-\frac{d+1}{2}} |G|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (\alpha_i - \alpha)^T G^{-1} (\alpha_i - \alpha) \right)$$

$$\times (2\pi)^{-\frac{d}{2}} |\Sigma_i|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (Y_i - A_i\alpha_i - 1_nX_i^T \beta)^T \Sigma_i^{-1} (Y_i - A_i\alpha_i - 1_nX_i^T \beta) \right),$$
where $S_i(\cdot)$ denotes the survival function of $i^{th}$ subject. And,

$$
\log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right)
$$

$$
= \delta_i \left( \log(\lambda_0(T_i)) + B_c(a, b, T_i) T \alpha_i + B_0(a, b, T_i) X_i T \beta + Z_i T \eta \right)
$$

$$
- \int_0^{T_i} \lambda_0(u) \exp \left( B_c(a, b, u) T \alpha_i + B_0(a, b, u) X_i T \beta + Z_i T \eta \right) du
$$

$$
- \frac{1}{2} \log |\Sigma_i| - \frac{1}{2} (Y_i - A_i \alpha_i - 1_n X_i T \beta)^T \Sigma_i^{-1} (Y_i - A_i \alpha_i - 1_n X_i T \beta)
$$

$$
- \frac{1}{2} \log |G| - \frac{1}{2} (\alpha_i - \alpha)^T G^{-1} (\alpha_i - \alpha) - \frac{n_i}{2} \log(2\pi) - \frac{q + 1}{2} \log(2\pi).
$$

Maximum likelihood estimation approach is utilised to develop statistical inference for the joint models, thus, estimators for the parameters $\theta$ are proposed to maximize the log-likelihood function (4.11), that is, estimations are obtained by solving equations

$$
\frac{\partial \ell(\theta; \text{data})}{\partial \theta} = 0 \quad (4.12)
$$

and we have that,

$$
\frac{\partial \ell(\theta; \text{data})}{\partial \theta} = \sum_{i=1}^m \frac{\partial}{\partial \theta} \left( \int Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) d\alpha_i \right)
$$

$$
= \sum_{i=1}^m \int \frac{\partial}{\partial \theta} \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right) Pr(T_i, \delta_i, Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) d\alpha_i
$$

$$
= \sum_{i=1}^m \int \frac{\partial}{\partial \theta} \left( \log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right) \right) Pr(T_i, \delta_i, Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) d\alpha_i
$$

$$
= \sum_{i=1}^m \int \frac{\partial}{\partial \theta} \left( \log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right) \right) Pr(T_i, \delta_i, Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) d\alpha_i
$$

where we assume the conditions for derivative commuting over integral are satisfied, which is also the case of study models.

Under the maximum likelihood estimation framework, the corresponding estimates for $\lambda_0(t)$ usually lead to under estimation of the standard deviations of parameter estimations (Hsieh, Tseng and Wang, 2006, for instance), and additional efforts are required in order to obtain proper standard deviations for the inference, such as bootstrap methods to approach the standard deviations, which is usually quite computationally intensive, especially within the framework of joint modelling. In the study,
we propose piecewise function to characterize the unspecific baseline hazard function, \( \lambda_0(t) \), which, on one hand, can turn the non-parametric issues into parametric ones, on the other hand, it may be able to avoid the issue of under estimation of standard deviations (Rizopoulos, Verbeke and Lesallfre, 2009). More specifically,

\[
\lambda_0(t) = \sum_{k=1}^{K} \zeta_k I(v_{k-1} < t \leq v_k),
\]

where \( v_0 < v_1 < \cdots < v_K \) are knots, which can be simply chosen as equally spaced follow-up time or quantile points of observed survival times. And \( \zeta = (\zeta_1, \zeta_2, \cdots, \zeta_K)^T \) are the parameters of piecewise function for charactering the baseline hazard function.

With the maximum likelihood estimation approach, inverse of Fisher information is proposed to calculate the covariance matrix for estimators, and the negative of Hessian matrix is used to estimate the Fisher information,

\[
\frac{\partial^2 \ell(\theta, \text{data})}{\partial \theta^T \partial \theta} = \sum_{i=1}^{m} \int \left( \frac{\partial^2}{\partial \theta^T \partial \theta} \log \Pr(T_i, \delta_i, Y_i, \alpha_i|\theta) \Pr(\alpha_i|T_i, \delta_i, Y_i, \theta) \right) d\alpha_i
\]

\[
+ \sum_{i=1}^{m} \int \frac{\partial}{\partial \theta} \log \left( \Pr(T_i, \delta_i, Y_i, \alpha_i|\theta) \right) \frac{\partial}{\partial \theta^T} \log \left( \Pr(T_i, \delta_i, Y_i, \alpha_i|\theta) \right) \Pr(\alpha_i|T_i, \delta_i, Y_i, \theta) d\alpha_i
\]

\[
- \sum_{i=1}^{m} \left( \int \frac{\partial}{\partial \theta} \log \left( \Pr(T_i, \delta_i, Y_i, \alpha_i|\theta) \right) \Pr(\alpha_i|T_i, \delta_i, Y_i, \theta) d\alpha_i \right)
\]

\[
\int \frac{\partial}{\partial \theta^T} \log \left( \Pr(T_i, \delta_i, Y_i, \alpha_i|\theta) \right) \Pr(\alpha_i|T_i, \delta_i, Y_i, \theta) d\alpha_i \right)
\]

\[
= \sum_{i=1}^{m} E \left( \frac{\partial^2}{\partial \theta^T \partial \theta} \log \left( \Pr(T_i, \delta_i|\alpha_i, \theta) \Pr(Y_i|\alpha_i, \theta) \Pr(\alpha_i|\theta) \right) \big| T_i, \delta_i, Y_i \right)
\]

\[
+ \sum_{i=1}^{m} \text{Var} \left( \frac{\partial}{\partial \theta} \log \left( \Pr(T_i, \delta_i|\alpha_i, \theta) \Pr(Y_i|\alpha_i, \theta) \Pr(\alpha_i|\theta) \right) \big| T_i, \delta_i, Y_i \right).
\]

That is, covariance matrix for parameters estimators \( \hat{\theta} \) is approximated by

\[
\hat{\text{cov}}(\hat{\theta}) = - \left( \frac{\partial^2 \ell(\theta, \text{data})}{\partial \theta^T \partial \theta} \right)^{-1}
\]

and therefore statistical inference for the parameters obtained. Our later on simulation study also verifies the approach applicable.

In this study, EM-algorithm is proposed to find the MLE of \( \theta \) and fully exponential Laplace approximation is proposed to approximate the posterior expectations of functions of random effects, appearing in the estimating equations (4.12), conditional on the observed data and current estimated of the parameters.
4.4 Estimation procedure

Together with piecewise function for baseline hazard $\lambda_0(t)$, the parameters vector for joint models is denoted by $\theta = (\alpha, \beta, G, \sigma^2, a, b, \eta, \zeta_k: k = 1, K)^T$ and we have that,

$$\log \left( Pr(T_i, \delta_i|\alpha_i, \theta) Pr(Y_i|\alpha_i, \theta) Pr(\alpha_i|\theta) \right)$$

$$= \delta_i \left( \log \left( \sum_{k=1}^{K} \zeta_k I(v_{k-1} < T_i \leq v_k) + B^c(a, b, T_i)^T \alpha_i + B^0_0(a, b, T_i)X_i^T \beta + Z_i^T \eta \right) \right)$$

$$- \sum_{k=1}^{K} \zeta_k e^{Z_i^T \eta} \int_{v_{k-1}}^{\min(v_k, T_i)} \exp \left( B^c(a, b, u)^T \alpha_i + B^0_0(a, b, u)X_i^T \beta \right) du$$

$$- \frac{1}{2} \log |\Sigma_i| - \frac{1}{2}(Y_i - A_i\alpha_i - 1_n, X_i^T \beta)^T \Sigma_i^{-1}(Y_i - A_i\alpha_i - 1_n, X_i^T \beta)$$

$$- \frac{1}{2} \log |G| - \frac{1}{2}(\alpha_i - \alpha)^T G^{-1}(\alpha_i - \alpha) - \frac{n_i}{2} \log(2\pi) - \frac{q+1}{2} \log(2\pi),$$

where $K_i = \min\{k: T_i \leq v_k\}$. As noticed, there is no analytical form for the integration of cumulative hazard,

$$\int_{v_{k-1}}^{\min(v_k, T_i)} \exp \left( B^c(a, b, u)^T \alpha_i + B^0_0(a, b, u)X_i^T \beta \right) du,$$

which also complicates the implementation of statistical inference, along with the complexity of joint models. Numerical methods are applied to approximate the integration for the cumulative hazards, specifically, we use the 15-points Gauss-Kronrod quadrature rules (Kahaner et al., 1989, p153), that is

$$\int_{-1}^{1} f(x) dx \approx \sum_i \omega_i f(x_i),$$

where $x_i$ and $\omega_i$ are nodes and the corresponding weights, respectively. And the details of nodes and weights for 15-points Gauss-Kronrod quadrature rules are presented in Table 4.1. The integral interval of cumulative hazards is not between -1 and 1, hence the change of interval is applied here (Kahaner et al., 1989, p148)

$$\int_{v_{k-1}}^{v_k} g(u) du = \frac{v_k - v_{k-1}}{2} \int_{-1}^{1} g \left( \frac{(v_k - v_{k-1})x + v_{k-1} + v_k}{2} \right) dx$$

$$\approx \frac{v_k - v_{k-1}}{2} \sum_i \omega_i g \left( \frac{(v_k - v_{k-1})x_i + v_{k-1} + v_k}{2} \right).$$

Denote

$$u_{kj} = \frac{(v_k - v_{k-1})x_j + v_{k-1} + v_k}{2}, dt_{kj} = \frac{v_k - v_{k-1}}{2}, \omega_{kj} = \omega_j, 1 \leq j \leq 15,$$
Table 4.1: 15-point Gauss-Kronrod quadrature rules

<table>
<thead>
<tr>
<th>Kronrod nodes</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>±0.99145 53711 20813</td>
<td>0.02293 53220 10529</td>
</tr>
<tr>
<td>±0.94910 79123 42759</td>
<td>0.06309 20926 29979</td>
</tr>
<tr>
<td>±0.86486 44233 59769</td>
<td>0.10479 00103 22250</td>
</tr>
<tr>
<td>±0.74153 11855 99394</td>
<td>0.14065 32597 15525</td>
</tr>
<tr>
<td>±0.58608 72354 67691</td>
<td>0.16900 47266 39267</td>
</tr>
<tr>
<td>±0.40584 51513 77397</td>
<td>0.19035 05780 64785</td>
</tr>
<tr>
<td>±0.20778 49550 07898</td>
<td>0.20443 29400 75298</td>
</tr>
<tr>
<td>0.00000 00000 00000</td>
<td>0.20948 21410 84728</td>
</tr>
</tbody>
</table>

then we have,

\[ \sum_{k=1}^{K_i} \zeta_k e^{Z_i^T \eta} \int_{v_{k-1}}^{\min(v_k,T_i)} \exp \left( B^c(a, b, u)T \alpha_i + B_0^c(a, b, u)X_i^T \beta \right) du \]

\[ \approx \sum_{k=1}^{K_i} \zeta_k e^{Z_i^T \eta} \sum_{j \in \Omega_{ik}} dt_{kj} \omega_{kj} \exp \left( B^c(a, b, u_{kj})X_i^T \beta \right), \]

where \( \Omega_{ik} = \{ j : v_{k-1} \leq u_{kj} \leq \min(v_k, T_i) \} \).

Parameters are estimated using the expectation-maximization (EM) algorithm by maximizing the likelihood of the observed data, \((T_i, \delta_i, Y_i : i = 1, m)\), which is done by iterating between an E-step, where we compute the expected log-likelihood of the complete data, \((T_i, \delta_i, Y_i, \alpha_i : i = 1, m)\), conditional on the observed data and the current estimated of the parameters, and an M-step, where new parameter estimates are computed by maximizing this expected log-likelihood. As illustrated in the last section of this chapter, some parameters, such as \( \alpha, G, \) and \( \sigma^2 \), have analytical forms for the M-step. For some other parameters, however, there is no analytically closed forms, such as the survival parameters of \( a, b \) and \( \eta \), and one-step Newton-Raphson approach is proposed. Therefore, each M-step of iterating, estimation equations are

\[ \sum_{i=1}^{m} \int \frac{\partial}{\partial \theta} \left( \log P_r(T_i, \delta_i, Y_i, \alpha_i|\theta) \right) P_r(\alpha_i|T_i, \delta_i, Y_i, \theta_0) d\alpha_i = 0, \quad (4.13) \]

where \( \theta_0 \) stands for the current estimated of the parameters and new parameter estimates are expected for the next iteration, until convergence.

In the meantime, expectation of functions of random effects conditional on the
observed data and current estimated parameter $\theta_0$ are required, that is,

$$E(h(\alpha_i)|T_i, \delta_i, Y_i, \theta_0) = \int h(\alpha_i) Pr(\alpha_i|T_i, \delta_i, Y_i, \theta_0) d\alpha_i$$

$$= \int \frac{h(\alpha_i) Pr(T_i, \delta_i, Y_i, \alpha_i|\theta_0)}{Pr(T_i, \delta_i, Y_i|\theta_0)} d\alpha_i$$

$$= \frac{\int h(\alpha_i) Pr(T_i, \delta_i, Y_i, \alpha_i|\theta_0) d\alpha_i}{\int Pr(T_i, \delta_i|\theta_0) Pr(Y_i|\theta_0) Pr(\alpha_i|\theta_0) d\alpha_i}$$

which are under the assumption that: (1) data that are from different subjects are mutually independent; (2) given the random effects and common covariates, longitudinal process and the point process of events are assumed to be independent. The expectations are not tractable usually and numerical approaches are required. Gauss-Hermite quadrature approach has been successfully used to deal with the inference for joint models, especially with low dimension of random effects, such as by Wulfsohn and Tsiatis (1997) and Henderson et al. (2000), among others. However, when the dimension of the random effects goes large, even with only 3 dimensions, that is, $\alpha_i = (\alpha_{0i}, \alpha_{1i}, \alpha_{2i})^T$ and relative large sample size, say $m = 500$, turns out to be quite computationally intensive and our experience shows that an ordinary computer tends to have difficulty dealing with such computation burden. And the adaptive Gauss-Hermite quadrature approach (Liu and Pierce, 1994) has recently been proposed to improve it consequently. Another straightforward approach is the Monte Carlo approach, for example, proposed by Henderson et al. (2000) and Hsieh et al. (2006), which, however, is also quite computationally intensive and Monte Carlo error, that is, variation introduced by employing a Monte Carlo approach, is inevitable to be introduced in the statistical inference.

Here we would like to propose the fully exponential Laplace approximations to approximate the posterior expectations. As demonstrated by equation (4.14), the expectations can be treated as posterior expectations of the functions of random effects. Laplace’s method typically has an error of order $O(n^{-1})$, while by applying the same approximation technique, Laplace’s method, to the numerator and the denominator, the error is of order $O(n^{-2})$, as discussed by Tierney and Kadane (1986). In our case, the random effects function $h(\alpha_i)$ is not necessary to be positive all the time, and can be nonpositive function as well, for example, the simplest case of $h(\alpha_i) = \alpha_i$, we
have \( h(\alpha_i) \in R \). Therefore, in order to obtain a second-order approximation to the expectation \( E(h(\alpha_i)|T_i, \delta_i, Y_i) \), the fully exponential method is used to approximate the moment-generating function \( E(\exp(sh(\alpha_i))) \), whose integrand is positive and then differentiate the result, details are referred to the last section of the chapter. In summary, we have,

\[
E(h(\alpha_i)|T_i, \delta_i, Y_i) = h(\hat{\alpha}_i) - \frac{1}{2} \text{Trace} \left( \left( -L''(\hat{\alpha}_i) \right)^{-1} \left( -h''(\hat{\alpha}_i) \right) \right) + O(n^{-2})
\]

where \( \hat{\alpha}_i = \text{argmax}_{\alpha_i} \log Pr(T_i, \delta_i, Y_i|\theta_0) \), \( L(\alpha_i) = \log Pr(T_i, \delta_i, Y_i, \alpha_i|\theta_0) \) and \( h''(\cdot) \) denotes the second order derivative of the \( h(\cdot) \) with respect to \( \alpha_i \), and \( L''(\cdot) \) similarly.

Newton-Raphson algorithm is proposed to find the maximization point of \( \alpha_i \), which actually is quite fast according to our experience. When calculating Hessian matrix, the conditional variances of functions of random effects are required as well, one can also apply the Laplace approximate to obtain the posterior variance. Alternatively, we apply the Monte Carlo approach to reach the approximation of the posterior variances and size of 2000 Monte Carlo is used. It may be slightly time consuming but we only apply it once for the standard deviations in the end of iterations, hence it is acceptable.

Regarding choosing initial values for the parameters, starting values for longitudinal process can easily obtained from the outputs of linear mixed regression models when the analysis includes only the longitudinal process. The starting values for the piecewise parameters \( \zeta \) and regression coefficients \( \eta \) are not that sensitive to results, and one can use some wide values, such as 1 for all \( \zeta \) and 0 or 0.1 for \( \eta \), and after several iterations of joint models proposed by the study, they usually will approach to certain radian of the destined values, and these values can be used as initial values. Also the separate analysis of the survival data can also provide us with some helpful thoughts. In the meantime, the R package JM can be utilised too, especially when the current values of longitudinal process have been included in the models, that is \( B^c(a, b, t) \), instead of \( B(a, b, t) \), is proposed in the Cox regression model. The most challenging part is to choose an appropriate initial value for the parameter \( b \), whose information has not been demonstrated in some straightforward way yet. And a bad starting value for \( b \) usually leads to long running-time to get convergence or even lack of convergence. Here we would like to propose the way of grid search to obtain the pair starting values for \( (a, b) \). Given proper starting values for parameters \( \zeta, \beta \) and \( \eta \) and the BLUE random effects estimations of LMM for the \( \alpha_i \). And define sequences
of \( a \) and \( b \) with \( a \in [a_1, a_2] \) and \( b \in [b_1, b_2] \), where range for \( a \) can be chosen according to the outputs of JM, and \( b \) can blindly be some positive values, such as \( b \in [0.1, 25] \). The pair of \((a, b)\) is chosen by maximizing the function

\[
\sum_{i=1}^{m} \left[ \delta_i \left( \log \left( \sum_{k=1}^{K} \zeta_k I(v_{k-1} < T_i \leq v_k) \right) + B^c(a, b, T_i)^T \alpha_i + B_0^c(a, b, T_i)X_i^T \beta + Z_i^T \eta \right) \right. \\
- \sum_{k=1}^{K_i} \zeta_k e^{Z_i^T \eta} \int_{v_{k-1}}^{\min(v_k, T_i)} \exp \left( B^c(a, b, u)^T \alpha_i + B_0^c(a, b, u)X_i^T \beta \right) du \]

and therefore the initial values for parameters \( a \) and \( b \), more details are demonstrated in the real data analysis the next section of this chapter. And our experience shows that decent initial values can be approached by this procedure.

### 4.5 Real data analysis

#### 4.5.1 Schizophrenia trial data

The data were obtained from a randomized clinical trial comparing different drug regimes in the treatment of chronic schizophrenia and have been previously discussed by Diggle (1998), Henderson et al. (2000) and others. In the study, patients were randomly allocated amongst the treatments of placebo, haloperidol and risperidone. Haloperidol is regarded as a standard therapy. Risperidone is described as ‘a novel chemical compound with useful pharmacological characteristics, as has been demonstrated in the in vitro and in vivo experiments’. The primary response variable was the total score obtained on the Positive and Negative Symptom Rating Scale (PANSS), a measure of psychiatric disorder. The study design specified that this score should be taken at weeks -1, 0, 2, 4, 6 and 8, where -1 refers to selection into the trial and 0 refers to baseline. A reduction of 20% in the mean PANSS score is considered a worthwhile clinical improvement. The study, however, had a serious drop-out rate of patients, which might or might not be due to the inadequate response.

Due to the confidential issue, so far we do not have the complete clinical trial data. Here we would like to analyse subset of the data which was circulated out in an ATPS training course in England. The data are available from 150 patients. For the data set, the event is considered to be drop-out from the study for causes related with the longitudinal PANSS profile. Of 150 patients, 68 completed the study, and from
those that drop-out are known to have drop-out because of inadequate response, denoted as events in the data set. The details of the data have been attached in the end the thesis, where treatment 1 is the standard therapy haloperidol, treatment 2 is placebo, and treatment 3 is the new treatment risperidone. The interest lies on the change of PANSS score with time and its association with treatments. Logarithm of PANSS score is applied here to shrink the scale for the integration purpose. Our simple preliminary analysis with longitudinal PANSS score shows that the model with random effects of measurement times with quadratic term has better fitting than it does with linear trend only. Hence the model proposed for longitudinal process is by

$$Y_i(t) = \alpha_0i + \alpha_1i t + \alpha_2i t^2 + \beta^T trt_i + \varepsilon_i(t),$$

where $trt_i$ denotes treatment factor and $\beta$ denotes the vector of corresponding treatment effects. Random effects $\alpha_i = (\alpha_0i, \alpha_1i, \alpha_2i)^T$ are assumed to be normally distributed with mean $\alpha = (\alpha_0, \alpha_1, \alpha_2)^T$ and covariance $G$. Random errors for the measurements of $i^{th}$ subject, $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in_i})^T$, are assumed from multivariate normal distribution as well with zero mean and identity covariance matrix $\sigma^2 I_{n_i}$. And we denote $m_i(t) = \alpha_0i + \alpha_1i t + \alpha_2i t^2 + \beta^T trt_i$. Regarding the point process of events in time, Cox regression model is proposed to characterize the relationship of informative drop-out with the PANSS profile and treatment regimes as well. More specifically, by proposing $\gamma(u) = ae^{-bu}, u > 0$, we have,

$$\lambda_i(t) = \lim_{dt \to 0} \frac{dt}{dt}Pr(t \leq \tilde{T}_i < t + dt|\tilde{T}_i \geq t, m_i(t), C_i, trt_i, \alpha_i, \varepsilon_i(t))$$

$$= \lambda_0(t) \exp \left( \gamma(0)m_i(t) + \int_0^t \gamma(t-u)m_i(u)du + \eta^T trt_i \right)$$

$$= \lambda_0(t) \exp \left( am_i(t) + \int_0^t ae^{-bu}m_i(t-u)du + \eta^T trt_i \right)$$

$$= \lambda_0(t) \exp \left( B^c(a, b, t)^T \alpha_i + B^c_0(a, b, t)\beta^T trt_i + \eta^T trt_i \right),$$

where we apply $\lambda_0(t) = \sum_{k=1}^{K} \zeta_k(v_{k-1} < t \leq v_k)$, and 15 knots of equal spaced follow-up time are used. Treatment factor is also considered in Cox model since PANSS profile may not be a good enough marker and it is of interest to find out. One may wonder if it might cause the issue of model identification, however, our experience shows that it works in this way since the former $trt_i$ is introduced by the longitudinal process. Regarding the choices of initial values for parameters, the grid search approach is
utilised. As demonstrated by Figure 4.3, we start algorithm with initial values of 
\( a = 4.5 \) and \( b = 2.56 \) and those for other parameters are setted up as discussed earlier.

Analysis outputs of the proposed method are summarized in Table 4.2, in which the statistical inference results by using R package JM are presented as well. Both analysis have the same model settings except that there is not cumulative term for the JM models and that only default current values of PANSS profile and treatment factor have been accounted for in the analysis of drop-out process due to inadequate response. The analysis results show that, to some extend, the statistical inference from these two approaches share some similarity. However, by ignoring past trajectory information of the longitudinal process, it may under estimate the influence of different levels of PANSS score on the drop-out rate due to inadequate response. And it may also lead to biased inference for the changing pattern of longitudinal process and treatment effects. For instance, the estimated regression coefficient of treatment risperidone is -0.0676, which is slightly stronger than -0.0448 obtained by ignoring the cumulative effects. The former one is significant under the 10% Wald confidence level, the latter one, however, is not. And the issue of under estimating seems to also appear in the inference for association between different levels of PANSS profile and the informative drop-out. As illustrated by formula (4.7) and the analysis results presented by Table 4.2, under the framework of cumulative analysis, the regression coefficient of current value \( m_i(t) \) is 
\[
a(1 + \int_0^t e^{-bu}du) \approx 4.48(1 + \int_0^t e^{-2.12u}du) > 3.39, \]
where 3.39 is the estimated association obtained from the joint models by ignoring all the past trajectory information of PANSS process, presented by Table 4.2. Henderson et al. (2000) stated that the events of drop out appeared to be affected not only by the current value but also the steepness of the trajectory, thus it seems to be more reasonable to take into account the past trajectory information, which also somewhat justifies the inference of cumulative analysis. In the meantime, weighted curve \( \gamma(t - u) \), PANSS score trajectory \( m_i(u) \) and the cumulative information \( \int_0^t \gamma(t - u)m_i(u)du \) for some random subjects are displayed in Figure 4.4, in order to demonstrate how the trajectory information of the longitudinal process is accounted for analysing data. The dash lines are 95% confidence interval for weighted curve and simply calculated by \( \pm 1.96\text{sd} \). The standard deviation of weight curve \( \gamma(u) = ae^{-bu} \) is approximately estimated by utilising delta method,
Figure 4.3: Survival part of the log-likelihood function of complete data with respect to pair \((a, b)\), other parameters are given as discussed in the section; the red round point indicates the maximum point of log-likelihood given other parameters setting.
Table 4.2: Analysis outputs for the PANSS data: ‘est.’ estimates of the parameters; ‘sd’ standard deviation; $\rho_{ij} = \text{cor}(\alpha_i, \alpha_j)$. JM outputs are from R package with current default values of longitudinal process as covariate of Cox regression model with method="spline-PH-aGH", other model settings are the same with that of cumulative analysis.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Cumulative analysis</th>
<th>JM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est.</td>
<td>sd</td>
</tr>
<tr>
<td>longitudinal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>4.5173</td>
<td>0.0174</td>
</tr>
<tr>
<td>survival</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>4.4772</td>
<td>0.0051</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.0128</td>
<td>0.0010</td>
</tr>
</tbody>
</table>

$G: (sd, \rho)$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\sigma_{\alpha_0}$</th>
<th>$\sigma_{\alpha_1}$</th>
<th>$\sigma_{\alpha_2}$</th>
<th>$\rho_{01}$</th>
<th>$\rho_{02}$</th>
<th>$\rho_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative</td>
<td>0.1733</td>
<td>0.0982</td>
<td>0.0085</td>
<td>-0.0150</td>
<td>0.0409</td>
<td>-0.9341</td>
</tr>
<tr>
<td>JM</td>
<td>0.1721</td>
<td>0.0946</td>
<td>0.0084</td>
<td>-0.007</td>
<td>0.0066</td>
<td>-0.9395</td>
</tr>
</tbody>
</table>

that is,

$$\text{Var}(\hat{\gamma}(u)) = \left( \begin{array}{c} \frac{\partial \gamma(u)}{\partial a} \\ \frac{\partial \gamma(u)}{\partial b} \end{array} \right)^T \text{Cov} \left( \begin{array}{c} \hat{a} \\ \hat{b} \end{array} \right) \left( \begin{array}{c} \frac{\partial \gamma(u)}{\partial a} \\ \frac{\partial \gamma(u)}{\partial b} \end{array} \right) \bigg|_{a=\hat{a}, b=\hat{b}}.$$  

### 4.5.2 Liver cirrhosis trial data

Liver cirrhosis trial data has previously been described by Andersen et al. (1993, page 19) and others. In the study, from year 1962 to 1969, patients with histologically verified liver cirrhosis at several hospitals in Copenhagen were included in a randomized clinical trial with random allocation either to treatment with the hormone prednisone or to an inactive placebo treatment. And the main purpose of the trial was to ascertain whether prednisone prolonged survival for patients with cirrhosis. The patients were followed until death or end of the study and survival time was the time since randomization. And the visits were scheduled to take place after 3, 6 and 12 months of treatment and, thereafter, once a year, though the achieved times and numbers varied considerably between patients, and the number of measurements of each patient, $n_i$, varies from 1 to 17. A number of variables were recorded at entry and throughout
Figure 4.4: Analysis for PANSS data, from left to right they are for weighted curve, $\gamma(t-u)$, past trajectory of PANSS score $m_i(u)$, and the integral of cumulative information $\gamma(t-u)m_i(u)$, where $t$ stands for the survival time and $u$ is the time before survival time. The grey area is the integration value $\int_0^t \gamma(t-u)m_i(u)du$. The dash lines in first column are 95% confidence interval for the weighted curve, and each row is for different subject.
the study, though here we focus on just two, treatment and prothrombin index measurements, as others do. The liver cirrhosis trial data can be downloaded from the R software package “joineR” (Philipson et al., 2012), where the data are stored in the balanced format and contain longitudinal follow-up information of prothrombin index measurements of each patient, the respective therapy arm, and also the time-to-event information. And the data are available for 488 patients, 251 received prednisone and 237 placebo.

In the analysis of liver cirrhosis trial data, longitudinal measurements of prothrombin index is again considered with log transformation. In the work by Henderson et al. (2002), linear mixed-effects model proposed for the longitudinal process is characterized by visiting times, baseline measurement indicators and their interaction term with treatment arm, plus the indicator of treatment regimes. Our simple preliminary analysis for prothrombin index measurements shows that, by taking into account the baseline measurements of prothrombin, it seems to have a much better model fitting in terms of smaller AIC/BIC, and it does so as well by accounting for the quadratic term of visiting time. Hence, for the longitudinal measurements of prothrombin index, the linear mixed-effects is proposed by

$$Y_i(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \beta_1 \text{trt}_i + \beta_2 \text{baseline}_i + \varepsilon_i(t),$$

where $\text{trt}_i$ is the 0-1 treatment arm indicator and $\text{baseline}_i$ is the prothrombin index measurements of $i^{th}$ subject taken at baseline. And the assumption settings for the model are the same with that for linear mixed-effects model of modelling the PANSS data. Similarly, we have $m_i(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \beta_1 \text{trt}_i + \beta_2 \text{baseline}_i$. When incorporating both of current and cumulative terms to Cox regression model, as illustrated by Figure 4.5, the grid search for initial values of $(a, b)$ suggests that $b = 16.7$ and our attempt to fit the model also indicates $b$ goes to large value. When $b$ goes large value, the cumulative value of integration turns out to be ignorable, which makes it lack of benefit doing so, meanwhile, it may lead to divergence of the algorithm as appearing in our attempt. Therefore we do not use the models with both of current and cumulative terms for the survival analysis here. The default current model should work fine with the liver data in terms of the joint modelling, and there already have some well-established analysis and discussion of survival analysis with current default value, Henderson et al. (2002) for instance. However, here we would like to try an
alternative approach and propose the hazard function of death by

\[
\lambda_i(t) = \lambda_0(t) \exp \left( \int_0^t \gamma(t - u)m_i(u)du + \eta rt_i \right)
\]

\[
= \lambda_0(t) \exp \left( a \int_0^t e^{-bu}m_i(t - u)du + \eta rt_i \right)
\]

\[
= \lambda_0(t) \exp \left( B(a, b, t)^T \alpha_i + B_0(a, b, t)(\beta_1 trt_i + \beta_2 baseline_0 + \eta rt_i) \right).
\]

Note that the \( B(a, b, t) \) is proposed in the survival model, instead of \( B^c(a, b, t) \). Again, \( \lambda_0(t) = \sum_{k=1}^K \zeta_k(v_{k-1} < t \leq v_k) \) and 20 knots of equal spaced follow-up time are used for the liver data. Two settings of initial values for \( \zeta, \beta \) and \( \eta \) have been tried during our attempt of analysing the data and two different initial pairs for \((a, b)\) have been suggested, as illustrated in Figure 4.6 (a) and (b). Interestingly, the algorithm with these two different initial settings converge to the same estimation for parameters, although, as expected, the closer the initials to the convergent points, the quicker the algorithm converges.

Analysis outputs are summarized in Table 4.3. Again inference from the classical joint models with default current values setting, denoted by JM approach, are provided as well. On one hand, the statistical performance of the common parts, for example the treatment effects and the change of prothrombin with time, by utilising two approaches
is of interest. As indicated by the analysis outputs, both approaches provide quite similar results and they suggest that the treatment of prednisone may be able to prolong survival for patients with cirrhosis, since the treatment has positive effect on the prothrombin index profile and the higher prothrombin index indicates the lower hazards rate of death of patients with liver cirrhosis. On the other hand, within the cumulative analysis, regression effect of current value $m_i(t)$ of survival model turns out to be $\int_0^t ab^{-bu}du$ and we have $\hat{b} = 0.2692$, which shows that the effect of present level of prothrombin index on the present hazard rate of death may be time-varying, and further investigation is required. Also, the weighted curve $\gamma(t - u)$, trajectory of prothrombin $m_i(u)$, and the cumulative information $\int_0^t \gamma(t - u)m_i(u)du$ for some random subjects are displayed in Figure 4.7. As expected, the shapes of integration area are a bit different from that of models with both of current and cumulative values of longitudinal process for the survival analysis, and see Figure 4.4 and 4.7 for details.
Figure 4.7: Analysis for liver cirrhosis trial data, from left to right they are for weighted curve, $\gamma(t - u)$, past trajectory of prothrombin index $m_i(u)$, and the integral of cumulative information $\gamma(t - u)m_i(u)$, where $t$ stands for the survival time and $u$ is the time before survival time. The grey area is the integration value $\int_0^t \gamma(t - u)m_i(u)du$. The dash lines in first column are 95% confidence interval for the weighted curve, and each row is for different subject.
Table 4.3: Analysis outputs for liver cirrhosis trial data: ‘est.’ estimates of the parameters; ‘sd’ standard deviation; $\rho_{ij} = \text{cor}(\alpha_i, \alpha_j)$. JM outputs are from R package with current default values of longitudinal process as covariate of Cox regression model with method=“spline-PH-aGH”, other model settings are the same with that of cumulative analysis. Note: cumulative term only for the cumulative analysis and the parameter $a$ is not comparable between two analysis.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Cumulative analysis</th>
<th>JM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est.</td>
<td>sd</td>
</tr>
<tr>
<td>longitudinal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.9942</td>
<td>0.0086</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.0159</td>
<td>0.0107</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.0030</td>
<td>0.0012</td>
</tr>
<tr>
<td>trt</td>
<td>0.1086</td>
<td>0.0176</td>
</tr>
<tr>
<td>baseline0</td>
<td>0.7644</td>
<td>0.0020</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.0656</td>
<td>0.0019</td>
</tr>
<tr>
<td>survival</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>-0.4845</td>
<td>0.0048</td>
</tr>
<tr>
<td>$b$</td>
<td>0.2692</td>
<td>0.0966</td>
</tr>
<tr>
<td>trt</td>
<td>-0.0326</td>
<td>0.1115</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>G: (sd, $\rho$)</th>
<th>$\sigma_{\alpha_0}$</th>
<th>$\sigma_{\alpha_1}$</th>
<th>$\sigma_{\alpha_2}$</th>
<th>$\rho_{01}$</th>
<th>$\rho_{02}$</th>
<th>$\rho_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative</td>
<td>0.1050</td>
<td>0.1436</td>
<td>0.0132</td>
<td>0.6044</td>
<td>-0.6441</td>
<td>-0.9908</td>
</tr>
<tr>
<td>JM</td>
<td>0.1117</td>
<td>0.1477</td>
<td>0.0135</td>
<td>0.6260</td>
<td>-0.6683</td>
<td>-0.9906</td>
</tr>
</tbody>
</table>

4.6 Simulation study

4.6.1 General simulation settings

Simulation study of the section is based on the scenarios of real data analysis. Quadrature polynomial function of time is proposed in the linear mixed-effects model to characterize the longitudinal process, more specifically,

$$Y_i(t) = m_i(t) + \varepsilon_i(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \varepsilon_i(t). \quad (4.15)$$

Cox regression model is proposed to capture the relationship of point process of events in time with covariates, and three Cox regression models are investigated in the study.

- The Cox regression model of default current value assumes that the hazard rate relates to longitudinal process only through the current value, denoted by

$$\lambda_i(t) = \lambda_0(t) \exp(\alpha m_i(t) + Z_i^T \eta)$$

$$= \lambda_0(t) \exp(B^0(a, -, t)^T \alpha_i + Z_i^T \eta), \quad (4.16)$$

where $B^0(a, -, t) = a(1, t, t^2)^T$, $a$ is the association parameter that links the longitudinal process and event hazard by current value $m_i(t)$, and $\eta$ is the regression
coefficients vector of covariates $Z_i$, which can be measured and recorded exactly, or of which the complete knowledge is available.

- Cumulative analysis model with current value represented by

$$
\lambda_i(t) = \lambda_0(t) \exp \left( \gamma(0)m_i(t) + \int_0^t \gamma(t-u)m_i(u)du + Z_i^T \eta \right)
$$

$$
= \lambda_0(t) \exp \left( \alpha_m(t) + \int_0^t \alpha_{b}u(\alpha_{0i} + \alpha_{1i}(t-u) + \alpha_{2i}(t-u)^2)du + Z_i^T \eta \right)
$$

$$
= \lambda_0(t) \exp(B^c(a, b, t)^T \alpha_i + Z_i^T \eta), \quad (4.17)
$$

where $B^c(a, b, t) = a(1 + \int_0^t e^{-bu}du, t + \int_0^t (t-u)e^{-bu}du, t^2 + \int_0^t (t-u)^2 e^{-bu}du)^T$.

- Cumulative analysis model without current value given by

$$
\lambda_i(t) = \lambda_0(t) \exp \left( \int_0^t \gamma(t-u)m_i(u)du + Z_i^T \eta \right)
$$

$$
= \lambda_0(t) \exp \left( \int_0^t \alpha_{b}u(\alpha_{0i} + \alpha_{1i}(t-u) + \alpha_{2i}(t-u)^2)du + Z_i^T \eta \right)
$$

$$
= \lambda_0(t) \exp(B(a, b, t)^T \alpha_i + Z_i^T \eta), \quad (4.18)
$$

where $B(a, b, t) = a(\int_0^t e^{-bu}du, \int_0^t (t-u)e^{-bu}du, \int_0^t (t-u)^2 e^{-bu}du)^T$.

The joint models of model (4.15) with model (4.16)/(4.17)/(4.18) are investigated. For simplicity and without lost of generality, the case of $Z_i = (Z_{i1}, Z_{i2})^T$ is proposed in the Cox model, where $Z_{i1}$ is the continuous covariate and $Z_{i2}$ is the 0-1 covariate, thus, we have $\eta = (\eta_1, \eta_2)^T$, subsequently.

With respect to the generating of Monte Carlo datasets, $\alpha_i$s are from multivariate normal distribution with mean $\alpha = (3.5, -0.3, 0.1)^T$ and covariance matrix $G$ with diagonal elements $(1.24, 0.2, 0.1)^T$ and correlations $(\rho_{10}, \rho_{20}, \rho_{21})^T = (-0.19, -0.19, 0.1)^T$, where $\rho_{ij} = \text{cor}(\alpha_i, \alpha_j)$ with $i, j \in \{0, 1, 2\}$, in most of the scenario settings. However, throughout the study, we may slightly adjust the parameter values from time to time in order to have around 45% to 60% of subjects occurring failure events of each Monte Carlo dataset, otherwise, in some scenarios, simulating data may be generated with extreme censoring rates, such as with censoring rate of 95% or 1%. Complete space of visiting time points for longitudinal measurements taken is pre-specified and are located at $seq(0, 8, length = 15)$, that is, 15 points are equally-spaced located between starting and ending points of 0 and 8, respectively. Throughout the study, there is not
event or censoring happening before time 2 and 10% chance of no-show-up missing measuring per schedule is assumed after the baseline. Also, the measurements will not be taken beyond the observed time of event process, that is, we have $t_{ij} \leq T_i$. Random error $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{im})^T$ for the $i^{th}$ subject is normally distributed with mean zero and covariance $\sigma^2 I_n$ with $\sigma^2 = 0.3$ and independent from random effects $\alpha_i$'s.

Regarding the survival data, the underlying time to event $\tilde{T}_i, i = 1, 2, \ldots, m$, are generated based on the hazard function (4.16), (4.17) or (4.18) in different scenarios, accordingly. And the underlying censoring time of $i^{th}$ subject $C_i, i = 1, 2, \ldots, m$, is from exponential distribution with mean 20 and censoring only may happen after time 2. Therefore, the observed survival data for the $i^{th}$ subject is $(T_i, \delta_i)$ with $T_i = \min(\tilde{T}_i, C_i)$ and $\delta_i = I(\tilde{T}_i \leq C_i)$. As aforementioned, covariate $Z_{i1}$ is continuous-type variable and generated from normal distribution $N(8, 3)$ in most of the time, and the mean value is chosen to assist the ratio of events occurring, hence it may vary as well in different scenarios, for example, $Z_{i1}$ is from $N(4,1)$ for the scenario of Table 4.7. And $Z_{i2}$ is 0-1 variable from binomial distribution with $B(1,0.6)$. The true values of $a$, $b$ and $\eta$ are specified in the tables presenting the simulation results and baseline hazard function $\lambda_0(t)$ is setted to be constant, 0.5, for example, within the scenario of Table 4.6. The ratio of failures/events has been setted up to be varying round 45% to 60% as aforementioned. Under this framework, the average number of measurements taken is around 9 to 10 for each Monte Carlo dataset.

### 4.6.2 Bias analysis

In this section, we would like to investigate how the joint models of model (4.15) and default current value model (4.16) perform when the underlying data generating mechanism is actually the joint models of model (4.15) and cumulative analysis models. Different scenarios are considered here within the misspecified modelling framework, in which simulated data are actually from the joint models having Cox model either (4.17) or (4.18), and each scenario is conducted with 500 Monte Carlo datasets and 300 sample size for each dataset. The underlying Cox regression models are indicated in corresponding output tables, along with the underlying settings for model parameters. Simulated data are being modelled by joint models with default current value and their statistical inference are obtained by utilising the JM package of software R, developed
CHAPTER 4. CUMULATIVE INFORMATION

by Rizopoulos (2010), with argument setting of method="spline-PH-aGH", that is, the adaptive Gauss-Hermite approach is used to approximate the integration of functions of random effects in the E-step of EM algorithm with respect to the JM outputs.

Simulation results presented in Table 4.4/4.5 are corresponding to the underlying Cox regression model of model (4.17)/(4.18), respectively. In the tables, $M_0$ indicates the underlying models, from which the corresponding analysing data are generated. More specifically, $M_0$: $B^0(a, -, t)/B^c(a, b, t)/B(a, b, t)$ indicates the observed survival data are generated according to hazard function (4.16)/(4.17)/(4.18), respectively, while $M_1$ indicates the fitting models. The left-top panels of Table 4.4 and Table 4.5 present the statistical inference of joint models with default current value when the analysis is conducted with correct model fitting. In general, the default current value model has a good performance under the specified parameter settings. The inference produces nearly unbiased estimators and reasonable coverage probability for the parameters of both the longitudinal process and the point process of events in time, especially the inference presented in Table 4.4. However, we also note that the estimation for longitudinal parameter $\alpha_2$ seems quite under estimated and has poor coverage probability subsequently, though the underlying value is relatively small. On the other hand, it seems the JM approach tends to slightly under estimated the survival parameters, especially as demonstrated by the results presented in Table 4.5, where the coverage probabilities for $\eta_1$ and association parameter $a$ are also relatively lower than the expected value of 95%, which suggests the further development is required.

In the meantime, the simulation results for scenarios with misspecified model fitting are presented in other panels of Table 4.4/4.5 and the underlying Cox regression models are model (4.17)/(4.18), respectively, however, Cox model (4.16) is applied to analyse the simulated data and the JM approach is utilised. Both tables show that the misspecification may severely under estimate the covariate effects of Cox models. Not only the effects of longitudinal process, as illustrated by Figure 4.9, but it may also tend to attenuate the regression coefficients of other baseline covariates, $\eta_1$ and $\eta_2$ presented in Table 4.4/4.5 for example, especially when the failure process relates pass information of longitudinal process closely, that is, when $b$ is relatively small. When the past level of longitudinal profile has long effect on the present event hazard and the weighted curve $\gamma(u)$ does not drop dramatically with respect to the time lag,
the misspecification may lead to severe bias and erroneous inference therefore. As demonstrated by Table 4.4, for instance, due to the misspecification, it leads to about 26% bias for both estimations of \( \eta_1 \) and \( \eta_2 \) when \( b = 0.5 \), and more than 13% bias when \( b = 1 \), and also extremely low of coverage probability subsequently. Table 4.5 indicates the similar issue pattern of misspecification.

With respect to the statistical inference for the influence of longitudinal process on the point process of events in time, equation (4.7) of the chapter shows that, by introducing the cumulative term of longitudinal process, it is not only taking into account the impact of current level of process, but also the trajectory information. The derivatives of longitudinal trajectory with respect to time also have been introduced to the Cox regression model and serve the inference. In the simulation study, the derivatives are up to second order, that is, the effects of \( m_i(t) \), \( m'_i(t) \) and \( m''_i(t) \) have been accounted into the cumulative analysis, where \( m'_i(t) \) and \( m''_i(t) \) are the first and second derivatives of \( m_i(t) \) with respect to time \( t \), along with the underlying regression coefficients

\[
a \left( 1 + \int_0^t e^{-bu} du \right), \quad -a \int_0^t u e^{-bu} du, \quad \frac{a}{2} \int_0^t u^2 e^{-bu} du,
\]

where the underlying Cox regression model is the hazard function (4.17) and we have \( \gamma(u) = ae^{-bu} \). When the hazard rate is from function (4.18), the underlying effect of current value of longitudinal process is slightly different. More specifically, the corresponding regression coefficients are

\[
a \int_0^t e^{-bu} du, \quad -a \int_0^t u e^{-bu} du, \quad \frac{a}{2} \int_0^t u^2 e^{-bu} du.
\]

It shows that, by taking cumulative analysis, the inference for time-varying influence may be under investigation as well with respect to the current value of longitudinal process and derivatives of its trajectory, as illustrated by the Figures 4.8/4.9, where solid black lines represent the underlying regression effects of \( m_i(t) \), \( m'_i(t) \) and \( m''_i(t) \), the dash-black lines are the estimates of associations by using JM approach, that is, the inference for the impact of longitudinal process on the failure time. As demonstrated by the figures, on one hand, the misspecified default current value model does not account for past longitudinal trajectory information in analysis and completely ignores the effects its derivatives may have. On the other hand, it may severely under estimate the
Table 4.4: Bias analysis $B^c(a, b, t)$. $M_0$ indicates the underlying models, from which analysis data are generated; $B^0(a, -, t)$ and $B^c(a, b, t)$ indicate the underlying models come along with hazard function of (4.16) and (4.17). Par: parameters; Truval.: true values of parameters; Mean: mean of the parameter estimators from Monte Carlo datasets. SD: standard deviation of the Monte Carlo estimators; SE: mean of approximations of standard deviation for each Monte Carlo dataset; RB(%): relative bias; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval. And the simulated data are fitted by $M_1$: $B^0(a, -, t)$.

<table>
<thead>
<tr>
<th>Par</th>
<th>Truval</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>RB(%)</th>
<th>CP(%)</th>
<th>Truval</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>RB(%)</th>
<th>CP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>3.5</td>
<td>3.5028</td>
<td>0.2713</td>
<td>0.0663</td>
<td>0.1</td>
<td>92.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.2</td>
<td>-0.2002</td>
<td>0.0294</td>
<td>0.0300</td>
<td>0.1</td>
<td>95.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.1</td>
<td>0.0887</td>
<td>0.0195</td>
<td>0.0066</td>
<td>11.3</td>
<td>41.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.3</td>
<td>0.2992</td>
<td>0.0097</td>
<td>0.0159</td>
<td>0.3</td>
<td>99.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>-1.0</td>
<td>-0.9644</td>
<td>0.0735</td>
<td>0.0723</td>
<td>3.6</td>
<td>90.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>1.0</td>
<td>0.9720</td>
<td>0.1933</td>
<td>0.2022</td>
<td>2.8</td>
<td>95.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>0.8</td>
<td>0.7821</td>
<td>0.0537</td>
<td>0.0517</td>
<td>2.2</td>
<td>91.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>3.5</td>
<td>3.5053</td>
<td>0.3074</td>
<td>0.0662</td>
<td>0.2</td>
<td>92.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.3</td>
<td>-0.3029</td>
<td>0.0373</td>
<td>0.0359</td>
<td>1.0</td>
<td>93.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.1</td>
<td>0.0901</td>
<td>0.0196</td>
<td>0.0068</td>
<td>9.9</td>
<td>45.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.3</td>
<td>0.3017</td>
<td>0.0094</td>
<td>0.0160</td>
<td>0.6</td>
<td>99.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>-1.0</td>
<td>-1.2403</td>
<td>0.1174</td>
<td>0.1024</td>
<td>17.3</td>
<td>33.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>1.0</td>
<td>0.8632</td>
<td>0.2257</td>
<td>0.2188</td>
<td>13.7</td>
<td>88.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>-</td>
<td>1.0082</td>
<td>0.0980</td>
<td>0.0861</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>3.5</td>
<td>3.5169</td>
<td>0.3790</td>
<td>0.0662</td>
<td>0.5</td>
<td>93.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.3</td>
<td>-0.3046</td>
<td>0.0396</td>
<td>0.0382</td>
<td>1.5</td>
<td>93.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.1</td>
<td>0.0802</td>
<td>0.0203</td>
<td>0.0065</td>
<td>19.8</td>
<td>32.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.3</td>
<td>0.3002</td>
<td>0.0095</td>
<td>0.0168</td>
<td>0.1</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>-1.0</td>
<td>-0.9276</td>
<td>0.0766</td>
<td>0.0722</td>
<td>7.3</td>
<td>75.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>1.0</td>
<td>0.9398</td>
<td>0.1958</td>
<td>0.2021</td>
<td>6.0</td>
<td>94.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>-</td>
<td>1.0668</td>
<td>0.0788</td>
<td>0.0744</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
impact of current level of longitudinal profile on the failure time as well, as illustrated by the first column of the first two rows of Figure 4.8 and Figure 4.9, there is a significant gap between the dash lines and the solid lines. The effects may be time varying, though they may turn to constant after some time point, especially when the value of $b$ is large. Note Figure 4.8 and Figure 4.9 have very similar performance pattern, where the underlying hazard functions are the Cox regression model (4.17) and model (4.18), respectively.

Therefore, by ignoring the past information longitudinal process, the misspecified model fitting may severely underestimate its influence on the failure time and it may also attenuate the regression coefficients of other covariates in Cox regression model for the inference of survival data. Improvements for the statistical inference are required.

4.6.3 Case 1: statistical inference for the cumulative models

In this section, we would like to investigate the numerical performance of proposed estimation approach for cumulative analysis of the joint models of model (4.15) and model (4.17)/(4.18) within the framework of correct model fitting. In real data analysis, it shows that the value of baseline hazard function $\lambda_0(t)$ sometimes can be either quite small or relatively large. For example, the estimates of parameter $\zeta_{k,s}$ for PANSS data are around 1e-14 (which is tiny, maybe because of the relatively large value of CD4+ counts and its strong effect on the drop-out process, further investigation may be required), while they all are greater than 1 for liver data. And these estimators are consistent with those obtained by JM approach with argument setting of method=“piecewise-PH-aGH”. Thus, the joint models of cumulative analysis are investigated under several scenarios here, of which the models come along with moderate or small values of baseline hazards, along with different value settings for $b$ parameter. Each scenario of simulations in this section is conducted with 300 Monte Carlo datasets and 500 sample size for each dataset. Statistical inference for joint models of (4.15) and (4.17) are presented in Table 4.6 and Table 4.7 with respect to moderate and small baseline hazards, respectively. Results for joint models of (4.15) and (4.18) are presented in Table 4.8. In general, the proposed estimation approach for cumulative analysis has a pretty good statistical inference performance for longitudinal parameters, even including the inference for parameter $\alpha_2$. As aforementioned, when the JM
Figure 4.8: Bias analysis $B^c(a, b, t)$. The solid black lines are the underlying time-varying effects of cumulative analysis with $B^c(a, b, t)$; the dash-black lines are estimators of associations from default current value of hazard model with $B^0(a, - , t)$, which are time-constant. Top panel is for the scenario when survival data are from underlying model with $B^c(a = 0.8, b = 0.5, t)$, middle panel with $B^c(a = 0.8, b = 1, t)$ and bottom panel with $B^c(a = 0.8, b = 4, t)$. No dash lines in plots of second and third columns due to that the default current value model does not consider the derivative effects of $m_i(t)$. 

\[ \begin{align*}
\text{Top panel: } & B^c(a = 0.8, b = 0.5, t) \\
\text{Middle panel: } & B^c(a = 0.8, b = 1, t) \\
\text{Bottom panel: } & B^c(a = 0.8, b = 4, t)
\end{align*} \]
Table 4.5: Bias analysis \( B(a, b, t) \). \( M_0 \) indicates the underlying models, from which analysis data are generated; \( B^0(a, -, t) \) and \( B(a, b, t) \) indicate the underlying models come along with hazard function of (4.16) and (4.18). Par: parameters; Truval.: true values of parameters; Mean: mean of the parameter estimators from Monte Carlo datasets. SD: standard deviation of the Monte Carlo estimators; SE: mean of approximations of standard deviation for each Monte Carlo dataset; RB(\%): relative bias; CP(\%), Monte Carlo coverage probability of 95% Wald confidence interval. And the simulated data are fitted by \( M_1: B^0(a, -, t) \).

<table>
<thead>
<tr>
<th>Par</th>
<th>Truval.</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>RB(%)</th>
<th>CP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_0 )</td>
<td>3.5</td>
<td>3.5020</td>
<td>0.0695</td>
<td>0.0663</td>
<td>0.1</td>
<td>93.1</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>-0.3</td>
<td>-0.2974</td>
<td>0.0391</td>
<td>0.0371</td>
<td>0.9</td>
<td>94.6</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>0.1</td>
<td>0.0847</td>
<td>0.0204</td>
<td>0.0068</td>
<td>15.3</td>
<td>40.4</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>0.3</td>
<td>0.2991</td>
<td>0.0101</td>
<td>0.0164</td>
<td>0.3</td>
<td>100.0</td>
</tr>
<tr>
<td>( \eta_1 )</td>
<td>-1.5</td>
<td>-1.4189</td>
<td>0.1190</td>
<td>0.1100</td>
<td>5.4</td>
<td>82.8</td>
</tr>
<tr>
<td>( \eta_2 )</td>
<td>-1.0</td>
<td>-0.9260</td>
<td>0.1972</td>
<td>0.1997</td>
<td>7.4</td>
<td>94.1</td>
</tr>
<tr>
<td>( a )</td>
<td>1.0</td>
<td>0.9668</td>
<td>0.0697</td>
<td>0.0665</td>
<td>3.3</td>
<td>88.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Par</th>
<th>Truval.</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>RB(%)</th>
<th>CP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_0 )</td>
<td>3.5</td>
<td>3.5030</td>
<td>0.0707</td>
<td>0.0667</td>
<td>0.1</td>
<td>93.8</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>-0.3</td>
<td>-0.3179</td>
<td>0.0388</td>
<td>0.0378</td>
<td>6.0</td>
<td>91.2</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>0.1</td>
<td>0.0987</td>
<td>0.0205</td>
<td>0.0072</td>
<td>1.4</td>
<td>46.8</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>0.3</td>
<td>0.3022</td>
<td>0.0101</td>
<td>0.0167</td>
<td>0.7</td>
<td>99.6</td>
</tr>
<tr>
<td>( \eta_1 )</td>
<td>-1.2</td>
<td>-0.8760</td>
<td>0.0673</td>
<td>0.0614</td>
<td>27.0</td>
<td>0.8</td>
</tr>
<tr>
<td>( \eta_2 )</td>
<td>1.0</td>
<td>0.7449</td>
<td>0.1961</td>
<td>0.1825</td>
<td>25.5</td>
<td>68.2</td>
</tr>
<tr>
<td>( a )</td>
<td>-</td>
<td>-0.7460</td>
<td>0.0532</td>
<td>0.0465</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Par</th>
<th>Truval.</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>RB(%)</th>
<th>CP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_0 )</td>
<td>3.5</td>
<td>3.5035</td>
<td>0.0704</td>
<td>0.0667</td>
<td>0.1</td>
<td>93.8</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>-0.3</td>
<td>-0.3098</td>
<td>0.0360</td>
<td>0.0359</td>
<td>3.3</td>
<td>92.6</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>0.1</td>
<td>0.0973</td>
<td>0.0203</td>
<td>0.0075</td>
<td>2.7</td>
<td>53.2</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>0.3</td>
<td>0.3013</td>
<td>0.0099</td>
<td>0.0161</td>
<td>0.4</td>
<td>100.0</td>
</tr>
<tr>
<td>( \eta_1 )</td>
<td>-0.8968</td>
<td>0.0668</td>
<td>0.0633</td>
<td>10.3</td>
<td>57.8</td>
<td></td>
</tr>
<tr>
<td>( \eta_2 )</td>
<td>1.5</td>
<td>1.3421</td>
<td>0.2067</td>
<td>0.2018</td>
<td>10.5</td>
<td>85.8</td>
</tr>
<tr>
<td>( a )</td>
<td>-</td>
<td>0.6295</td>
<td>0.0405</td>
<td>0.0396</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Par</th>
<th>Truval.</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>RB(%)</th>
<th>CP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_0 )</td>
<td>3.5</td>
<td>3.5060</td>
<td>0.0693</td>
<td>0.0668</td>
<td>0.2</td>
<td>94.6</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>-0.3</td>
<td>-0.3067</td>
<td>0.0356</td>
<td>0.0360</td>
<td>2.2</td>
<td>93.8</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>0.1</td>
<td>0.0972</td>
<td>0.0196</td>
<td>0.0082</td>
<td>2.8</td>
<td>58.0</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>0.3</td>
<td>0.3006</td>
<td>0.0099</td>
<td>0.0161</td>
<td>0.2</td>
<td>100.0</td>
</tr>
<tr>
<td>( \eta_1 )</td>
<td>-0.7671</td>
<td>0.0537</td>
<td>0.0543</td>
<td>4.1</td>
<td>89.2</td>
<td></td>
</tr>
<tr>
<td>( \eta_2 )</td>
<td>1.5</td>
<td>1.4318</td>
<td>0.2013</td>
<td>0.1986</td>
<td>4.6</td>
<td>93.0</td>
</tr>
<tr>
<td>( a )</td>
<td>-</td>
<td>0.5049</td>
<td>0.0328</td>
<td>0.0318</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Par</th>
<th>Truval.</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>RB(%)</th>
<th>CP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_0 )</td>
<td>3.5</td>
<td>3.5076</td>
<td>0.0669</td>
<td>0.0677</td>
<td>0.2</td>
<td>95.2</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>-0.3</td>
<td>-0.3057</td>
<td>0.0384</td>
<td>0.0369</td>
<td>2.5</td>
<td>93.0</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>0.1</td>
<td>0.0949</td>
<td>0.0217</td>
<td>0.0138</td>
<td>5.1</td>
<td>69.9</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>0.3</td>
<td>0.2992</td>
<td>0.0094</td>
<td>0.0167</td>
<td>0.3</td>
<td>100.0</td>
</tr>
<tr>
<td>( \eta_1 )</td>
<td>-0.9166</td>
<td>0.0586</td>
<td>0.0576</td>
<td>8.3</td>
<td>64.9</td>
<td></td>
</tr>
<tr>
<td>( \eta_2 )</td>
<td>1.0</td>
<td>0.9371</td>
<td>0.1675</td>
<td>0.1771</td>
<td>6.3</td>
<td>94.0</td>
</tr>
<tr>
<td>( a )</td>
<td>-</td>
<td>0.3940</td>
<td>0.0255</td>
<td>0.0256</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 4.9: Bias analysis $B(a, b, t)$. The solid black lines are the underlying time-varying effects of cumulative analysis with $B(a, b, t)$; the dash-black lines are estimators of associations from default current value of hazard model with $B^0(a, - , t)$, which are time-constant. Top panel is for the scenario when survival data are from underlying model with $B(a = 1, b = 0.5, t)$, middle panel with $B(a = 1, b = 1, t)$ and bottom panel with $B(a = 1, b = 4, t)$. No dash lines in plots of second and third columns due to that the default current value model does not consider the derivative effects of $m_i(t)$. 

\[ m(t) \]

\[ m_i(t) \]

\[ m''_i(t) \]
approach is utilised and the simulated data is analysed with correct model fitting, the relative bias for parameter $\alpha_2$ may still be about 15%, along with only 40% of seriously poor coverage probability, as presented by the top left panels of Table 4.4/4.5. In the meantime, the proposed estimation approach also provides plausible statistical inference for the survival parameters. For instance, nearly unbiased estimations or relatively small bias and reasonable coverage probability are obtained. The estimations are also quite robust in terms of small variation. But we also note that, similarly with the performance of the JM approach for joint models with standard default current value, the coverage probability is a bit lower than the expected value 95% and further development on statistical inference is required.

The cumulative analysis of joint models with $B^c(a, b, t)$ are presented in Table 4.6/4.7. More specifically, the simulation results for scenarios with moderate baseline hazards are presented in Table 4.6, which numerically verifies the validity of the proposed approach with nearly unbiased estimations and decent coverage probability. But we also note that the coverage probabilities for parameters $\eta_1$ and $a$ are sometimes lower than the expected value of 95% and that their corresponding standard deviations are slightly over estimated in terms of that the SEs are slightly greater than SDs, that is, the means of the estimated standard deviations tend to be larger than the standard deviation of the Monte Carlo estimators, which are expected to be close if applicable. With respect to the scenarios with small baseline hazards, the simulation results are presented in Table 4.7. It is impressive to note that the statistical inference for survival parameter $\eta_2$ is perfectly good in terms of that the visibly unbiased and efficient estimation is provided, SEs and SDs are close in all scenarios of the table, along with efficient 95% coverage probability. However, the statistical inference performance for other survival parameters may not be as ideal as that for $\eta_2$. For example, the SEs of $a$ and $\eta_1$ tend to be less than the values of corresponding SDs and their coverage probabilities may tend to be quite low subsequently. The estimations for all parameters are numerically unbiased in all scenarios presented by the table, except for the parameter $b$ when its underlying value is large. As aforementioned, when the underlying value of $b$ is large, which indicates the past information of longitudinal process may only have relatively weak influence on or barely relate to the present event hazards, the joint models of model (4.15) and (4.17)/(4.18) may be lack of convergence. The estimation
distribution of parameter $b$ seems to be skewed, also as illustrated by Figure 4.10/4.11.

The statistical inference performance demonstrated by Table 4.6/4.7 shows that a little bit care may be required about the typically granted property of asymptotic normality for the maximum likelihood estimation with respect to survival parameters under the joint modelling framework of longitudinal and survival data, especially regarding the cumulative analysis. We also plot the density functions of Monte Carlo estimators of survival parameters, as illustrated by Figure 4.10/4.11, with respect to the scenarios with moderate and small baseline hazards, respectively. It seems it is very likely that the estimations of survival parameters, obtained by maximizing the joint likelihood of longitudinal and survival data, may not have the asymptotic normal distribution in some cases, and they may come from some distribution with heavy tail and/or from some skewed distribution. Our another experiment with 500 Monte Carlo datasets and 300 sample size each, which although is not presented here any more, also shows the similar pattern of density functions. Best to our knowledge, there is no work in the literature to date discussing the asymptotic properties of the maximum likelihood estimation of survival parameters under the joint modelling framework of longitudinal and survival data. Further study and investigation is required with respect to this topic.

We also investigate the statistical performance of joint models of model (4.15) and model (4.18) with $B(a, b, t)$. The simulation results are presented in Table 4.8. These indicate the proposed estimation approach works well within this framework in the sense that the estimations are nearly unbiased, the values of SDs are close to that of SEs and the coverage probabilities are around 95%. The density functions of estimation for survival parameters, which are not presented here, show the same pattern with that presented by Figure 4.10/4.11. Our simulation experiment suggests the estimations under this framework are always roughly unbiased and efficient regardless of the value of $b > 0$. It suggests the joint models of model (4.15) and model (4.18) could be an alternative way of jointly modelling the longitudinal and survival data. However, as discussed in the next section, the interpretation of these models may not be straightforward, especially the impact pattern of the longitudinal process on the point process of events in time, which may be of primary interest. Because even just the current value of the longitudinal process $m_i(t)$ has impact on the risk hazard, joint
Table 4.6: Cumulative analysis of models with $B^c(a, b, t)$, moderate baseline hazards; $\lambda_0(t) = 1$ when $b = 0.5$, otherwise $\lambda_0(t) = 0.5$. Par: parameters; Truval.: true values of parameters; Mean: mean of the parameter estimators from Monte Carlo datasets. SD: standard deviation of the Monte Carlo estimators; SE: mean of approximations of standard deviation for each Monte Carlo dataset; RB(%): relative bias; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.

<table>
<thead>
<tr>
<th>Par</th>
<th>Truval.</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>RB(%)</th>
<th>CP(%)</th>
<th>Truval.</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>RB(%)</th>
<th>CP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B^c(a = 0.5, b = 0.5, t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$B^c(a = 1, b = 1, t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>3.5</td>
<td>3.4983 0.0513</td>
<td>0.0532 0.1</td>
<td>96.3</td>
<td>3.5</td>
<td>3.4991 0.0534</td>
<td>0.0533 0.0</td>
<td>95.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.3</td>
<td>-0.3004 0.0266</td>
<td>0.0263 0.1</td>
<td>93.7</td>
<td>-0.3</td>
<td>-0.3067 0.0279</td>
<td>0.0272 2.2</td>
<td>93.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.1</td>
<td>0.1007 0.0157</td>
<td>0.0151 0.7</td>
<td>94.0</td>
<td>0.1</td>
<td>0.1059 0.0161</td>
<td>0.0159 5.9</td>
<td>93.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.3</td>
<td>0.2995 0.0073</td>
<td>0.0071 0.2</td>
<td>94.3</td>
<td>0.3</td>
<td>0.2992 0.0070</td>
<td>0.0073 0.3</td>
<td>96.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>-1.5</td>
<td>-1.4611 0.0924</td>
<td>0.0865 2.6</td>
<td>89.0</td>
<td>-1.5</td>
<td>-1.3901 0.1013</td>
<td>0.1763 7.3</td>
<td>78.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>1.0</td>
<td>0.9621 0.1646</td>
<td>0.1719 3.8</td>
<td>96.0</td>
<td>-1.0</td>
<td>-0.9520 0.1910</td>
<td>0.2473 4.8</td>
<td>93.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>0.5</td>
<td>0.4917 0.0344</td>
<td>0.0342 1.7</td>
<td>91.0</td>
<td>1.0</td>
<td>0.9408 0.0678</td>
<td>0.1523 5.9</td>
<td>91.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$b$</td>
<td>0.5</td>
<td>0.5120 0.0993</td>
<td>0.0983 2.4</td>
<td>93.0</td>
<td>1.0</td>
<td>1.0333 0.1936</td>
<td>0.2058 3.3</td>
<td>86.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B^c(a = 1, b = 1, t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$B^c(a = 1, b = 2, t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>3.5</td>
<td>3.5030 0.0560</td>
<td>0.0531 0.1</td>
<td>94.3</td>
<td>3.5</td>
<td>3.5046 0.0544</td>
<td>0.0531 0.1</td>
<td>94.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.3</td>
<td>-0.3079 0.0267</td>
<td>0.0267 2.6</td>
<td>93.3</td>
<td>-0.3</td>
<td>-0.3083 0.0295</td>
<td>0.0263 2.8</td>
<td>89.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.1</td>
<td>0.1031 0.0167</td>
<td>0.0154 3.1</td>
<td>92.0</td>
<td>0.1</td>
<td>0.1033 0.0151</td>
<td>0.0153 3.3</td>
<td>95.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.3</td>
<td>0.2990 0.0077</td>
<td>0.0072 0.4</td>
<td>92.3</td>
<td>0.3</td>
<td>0.2993 0.0075</td>
<td>0.0071 0.2</td>
<td>93.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>-1.5</td>
<td>-1.4227 0.0983</td>
<td>0.1831 5.2</td>
<td>90.3</td>
<td>-1.5</td>
<td>-1.4578 0.1038</td>
<td>0.1813 2.8</td>
<td>92.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>-1.0</td>
<td>-0.9564 0.1905</td>
<td>0.2319 4.4</td>
<td>92.0</td>
<td>-1.0</td>
<td>-0.9719 0.1989</td>
<td>0.2263 2.8</td>
<td>94.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>1.0</td>
<td>0.9576 0.0796</td>
<td>0.1634 4.2</td>
<td>92.7</td>
<td>1.0</td>
<td>0.9680 0.0808</td>
<td>0.1602 3.2</td>
<td>94.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$b$</td>
<td>1.5</td>
<td>1.5454 0.3432</td>
<td>0.3910 3.0</td>
<td>93.3</td>
<td>2.0</td>
<td>1.9861 0.4382</td>
<td>0.5696 0.7</td>
<td>95.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B^c(a = 1.2, b = 2.5, t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$B^c(a = 1.2, b = 3, t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>3.5</td>
<td>3.5006 0.0504</td>
<td>0.0533 0.0</td>
<td>95.7</td>
<td>3.5</td>
<td>3.5025 0.0511</td>
<td>0.0532 0.1</td>
<td>95.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.3</td>
<td>-0.3095 0.0297</td>
<td>0.0268 3.2</td>
<td>90.3</td>
<td>-0.3</td>
<td>-0.3092 0.0292</td>
<td>0.0266 3.1</td>
<td>89.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.1</td>
<td>0.1072 0.0154</td>
<td>0.0155 7.2</td>
<td>90.3</td>
<td>0.1</td>
<td>0.1057 0.0158</td>
<td>0.0155 5.7</td>
<td>91.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.3</td>
<td>0.2991 0.0073</td>
<td>0.0072 0.3</td>
<td>92.3</td>
<td>0.3</td>
<td>0.2998 0.0076</td>
<td>0.0072 0.1</td>
<td>93.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>-1.5</td>
<td>-1.4149 0.0985</td>
<td>0.1506 5.7</td>
<td>82.7</td>
<td>-1.5</td>
<td>-1.4224 0.1029</td>
<td>0.1677 5.2</td>
<td>85.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>-1.0</td>
<td>-0.9445 0.1843</td>
<td>0.2458 5.6</td>
<td>95.0</td>
<td>-1.0</td>
<td>-0.9434 0.1992</td>
<td>0.2388 7.5</td>
<td>93.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>1.2</td>
<td>1.1397 0.0867</td>
<td>0.1500 5.0</td>
<td>87.7</td>
<td>1.2</td>
<td>1.1433 0.0851</td>
<td>0.1704 4.7</td>
<td>92.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$b$</td>
<td>2.5</td>
<td>2.5724 0.5687</td>
<td>0.6020 2.9</td>
<td>92.7</td>
<td>3.0</td>
<td>3.0673 0.7484</td>
<td>0.8047 2.2</td>
<td>93.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.7: Cumulative analysis of models with $B^c(a, b, t)$, small baseline hazards; $\lambda_0(t)=1\times10^{-07}$ when $b \in \{0.5, 1.0, 1.5\}$; $\lambda_0(t)=1\times10^{-06}$ when $b \in \{2, 2.5, 3.0\}$. Par: parameters; Truval.: true values of parameters; Mean: mean of the parameter estimators from Monte Carlo datasets. SD: standard deviation of the Monte Carlo estimators; SE: mean of approximations of standard deviation for each Monte Carlo dataset; RB(%): relative bias; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.

<table>
<thead>
<tr>
<th>Par</th>
<th>Truval.</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>RB(%)</th>
<th>CP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B^c(a=1, b=0.5, t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>2.173</td>
<td>2.1784</td>
<td>0.0546</td>
<td>0.0533</td>
<td>0.3</td>
<td>95.7</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.500</td>
<td>-0.512</td>
<td>0.0303</td>
<td>0.0277</td>
<td>2.5</td>
<td>89.7</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.150</td>
<td>0.1579</td>
<td>0.0170</td>
<td>0.0157</td>
<td>5.3</td>
<td>90.0</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.300</td>
<td>0.2994</td>
<td>0.0075</td>
<td>0.0074</td>
<td>0.2</td>
<td>94.3</td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>1.500</td>
<td>1.4236</td>
<td>0.1030</td>
<td>0.0844</td>
<td>5.1</td>
<td>67.7</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>-1.000</td>
<td>-0.9876</td>
<td>0.1856</td>
<td>0.2345</td>
<td>1.2</td>
<td>97.3</td>
</tr>
<tr>
<td>$a$</td>
<td>1.000</td>
<td>0.9667</td>
<td>0.0532</td>
<td>0.0538</td>
<td>3.3</td>
<td>76.3</td>
</tr>
<tr>
<td>$b$</td>
<td>0.500</td>
<td>0.5212</td>
<td>0.0672</td>
<td>0.0924</td>
<td>4.2</td>
<td>95.7</td>
</tr>
</tbody>
</table>

| $B^c(a=1, b=1.5, t)$ |          |           |          |          |       |       |
| $\alpha_0$ | 2.173   | 2.1736    | 0.0539   | 0.0530   | 0.0   | 95.0  |
| $\alpha_1$ | -0.500  | -0.5120   | 0.0285   | 0.0265   | 2.4   | 90.7  |
| $\alpha_2$ | 0.150   | 0.1556    | 0.0159   | 0.0152   | 3.7   | 92.0  |
| $\sigma^2$ | 0.300   | 0.2988    | 0.0073   | 0.0071   | 0.1   | 94.7  |
| $\eta_1$  | 1.500   | 1.4734    | 0.1162   | 0.0438   | 1.8   | 44.7  |
| $\eta_2$  | -1.000  | -1.0036   | 0.1938   | 0.2007   | 0.4   | 97.7  |
| $a$       | 1.000   | 0.9990    | 0.0766   | 0.0575   | 0.1   | 84.0  |
| $b$       | 1.500   | 1.6367    | 0.4601   | 0.4134   | 9.1   | 93.0  |

| $B^c(a=1, b=2.5, t)$ |          |           |          |          |       |       |
| $\alpha_0$ | 2.173   | 2.1744    | 0.0495   | 0.0531   | 0.1   | 95.3  |
| $\alpha_1$ | -0.500  | -0.5107   | 0.0280   | 0.0269   | 2.1   | 92.3  |
| $\alpha_2$ | 0.150   | 0.1548    | 0.0146   | 0.0154   | 3.2   | 95.7  |
| $\sigma^2$ | 0.300   | 0.2998    | 0.0071   | 0.0072   | 0.4   | 94.7  |
| $\eta_1$  | 1.500   | 1.5135    | 0.1173   | 0.0874   | 0.9   | 71.0  |
| $\eta_2$  | -1.000  | -1.0036   | 0.1938   | 0.2007   | 0.4   | 97.7  |
| $a$       | 1.000   | 1.0025    | 0.0873   | 0.0567   | 0.3   | 70.7  |
| $b$       | 2.500   | 2.7526    | 1.2723   | 0.9455   | 10.1  | 88.3  |
modelling can also produce reasonable statistical inference for the longitudinal process and the Cox regression coefficient $\eta$.

At last, we would like to mention that statistical inference for the covariance matrix $G$ of the proposed joint models is always with good performance under our simulation scenarios of these studies. An example is presented in Table 4.9, which is the corresponding estimation for the scenario of $B^c(a = 1, b = 1, t)$ in Table 4.6.

### 4.6.4 Case 2: cumulative analysis for default current models

In this section, we would like the investigate the statistical performance of joint models of model (4.15) and model (4.18) while the analysis data are actually generated from the joint models with default current value model, that is, the underlying models for the simulated data are joint models of model (4.15) and model (4.16). One possible way is to apply the proposed models and estimation approach to simulated data without any
Table 4.8: Cumulative analysis of models with $B(a, b, t)$. Par: parameters; Truval.: true values of parameters; Mean: mean of the parameter estimators from Monte Carlo datasets. SD: standard deviation of the Monte Carlo estimators; SE: mean of approximations of standard deviation for each Monte Carlo dataset; RB(%): relative bias; CP(%), Monte Carlo coverage probability of 95% Wald confidence interval.

<table>
<thead>
<tr>
<th>Par</th>
<th>Truval.</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>RB(%)</th>
<th>CP(%)</th>
<th>Truval.</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>RB(%)</th>
<th>CP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B(a = 0.8, b = 0.3, t), \lambda_0(t)=0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>3.5</td>
<td>3.5016</td>
<td>0.0538</td>
<td>0.0533</td>
<td>0.1</td>
<td>93.3</td>
<td>2.173</td>
<td>2.1763</td>
<td>0.0481</td>
<td>0.0532</td>
<td>0.2</td>
<td>97.3</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.3</td>
<td>-0.2986</td>
<td>0.0292</td>
<td>0.0262</td>
<td>0.5</td>
<td>91.3</td>
<td>-0.500</td>
<td>-0.4982</td>
<td>0.0278</td>
<td>0.0265</td>
<td>0.4</td>
<td>95.0</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.1</td>
<td>0.1007</td>
<td>0.0136</td>
<td>0.0149</td>
<td>0.7</td>
<td>96.0</td>
<td>0.150</td>
<td>0.1504</td>
<td>0.0142</td>
<td>0.0148</td>
<td>0.3</td>
<td>95.0</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.3</td>
<td>0.2998</td>
<td>0.0073</td>
<td>0.0071</td>
<td>0.1</td>
<td>95.3</td>
<td>0.300</td>
<td>0.2995</td>
<td>0.0072</td>
<td>0.0071</td>
<td>0.2</td>
<td>95.0</td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>-1.5</td>
<td>-1.5266</td>
<td>0.0917</td>
<td>0.0918</td>
<td>1.8</td>
<td>94.3</td>
<td>1.500</td>
<td>1.4786</td>
<td>0.1085</td>
<td>0.1754</td>
<td>1.4</td>
<td>96.3</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>-1.0</td>
<td>-1.0290</td>
<td>0.1693</td>
<td>0.1722</td>
<td>2.9</td>
<td>95.3</td>
<td>-1.000</td>
<td>-1.0024</td>
<td>0.1756</td>
<td>0.1683</td>
<td>0.2</td>
<td>91.3</td>
</tr>
<tr>
<td>$a$</td>
<td>0.8</td>
<td>0.8158</td>
<td>0.0578</td>
<td>0.0547</td>
<td>2.0</td>
<td>93.3</td>
<td>1.000</td>
<td>1.0037</td>
<td>0.0617</td>
<td>0.0557</td>
<td>0.4</td>
<td>67.7</td>
</tr>
<tr>
<td>$b$</td>
<td>0.3</td>
<td>0.3014</td>
<td>0.0463</td>
<td>0.0414</td>
<td>0.5</td>
<td>93.7</td>
<td>0.300</td>
<td>0.3101</td>
<td>0.0401</td>
<td>0.0364</td>
<td>3.4</td>
<td>91.0</td>
</tr>
</tbody>
</table>

| $B(a = 1, b = 0.3, t), \lambda_0(t)=1e-07$ |         |          |          |        |        |       |         |          |          |        |        |       |  
| $\alpha_0$ | 3.5     | 3.5016   | 0.0538   | 0.0533 | 0.1    | 93.3  | 2.173   | 2.1763   | 0.0481   | 0.0532 | 0.2    | 97.3  |
| $\alpha_1$ | -0.3    | -0.2986  | 0.0292   | 0.0262 | 0.5    | 91.3  | -0.500  | -0.4982  | 0.0278   | 0.0265 | 0.4    | 95.0  |
| $\alpha_2$ | 0.1     | 0.1007   | 0.0136   | 0.0149 | 0.7    | 96.0  | 0.150   | 0.1504   | 0.0142   | 0.0148 | 0.3    | 95.0  |
| $\sigma^2$ | 0.3     | 0.2998   | 0.0073   | 0.0071 | 0.1    | 95.3  | 0.300   | 0.2995   | 0.0072   | 0.0071 | 0.2    | 95.0  |
| $\eta_1$ | -1.5    | -1.5266  | 0.0917   | 0.0918 | 1.8    | 94.3  | 1.500   | 1.4786   | 0.1085   | 0.1754 | 1.4    | 96.3  |
| $\eta_2$ | -1.0    | -1.0290  | 0.1693   | 0.1722 | 2.9    | 95.3  | -1.000  | -1.0024  | 0.1756   | 0.1683 | 0.2    | 91.3  |
| $a$      | 0.8     | 0.8158   | 0.0578   | 0.0547 | 2.0    | 93.3  | 1.000   | 1.0037   | 0.0617   | 0.0557 | 0.4    | 67.7  |
| $b$      | 0.3     | 0.3014   | 0.0463   | 0.0414 | 0.5    | 93.7  | 0.300   | 0.3101   | 0.0401   | 0.0364 | 3.4    | 91.0  |

Table 4.9: Estimations for the distinct elements of covariance matrix G of the simulation scenario with $B^c(a = 1, b = 1, t)$ in Table 4.6; Trueval.: the true underlying values; Mean: mean of the Monte Carlo estimators.

<table>
<thead>
<tr>
<th>$G_{11}$</th>
<th>$G_{12}$</th>
<th>$G_{13}$</th>
<th>$G_{22}$</th>
<th>$G_{23}$</th>
<th>$G_{33}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trueval.</td>
<td>1.24000</td>
<td>-0.09462</td>
<td>-0.06691</td>
<td>0.20000</td>
<td>0.01414</td>
</tr>
<tr>
<td>Mean</td>
<td>1.24046</td>
<td>-0.09283</td>
<td>-0.06663</td>
<td>0.19866</td>
<td>0.01246</td>
</tr>
</tbody>
</table>
Figure 4.11: Density functions of estimators of survival parameters \((\eta_1, \eta_2, a, b)^T\) from 300 Monte Carlo datasets and 500 sample size each, with the underlying parameter setting \((\eta_1, \eta_2, a, b)^T = (1.5, -1, 1, 0.5)^T\) and the hazard model is with term \(B^a(a, b, t)\); small baseline hazards \(\lambda_0(t) = 1e-07\).

constraints, as demonstrated by the second example of real data analysis. However, the algorithm may suffer from lacking of convergence, hence we fix \(b\) to be some large value throughout the algorithm. As illustrated by Figure 4.8/4.9, the coefficients turn out to be almost time constant when \(b = 4\), rather than time varying. Thus, we set up \(b = 6\) throughout this simulation study, which is conducted with 300 Mote Carlo datasets and 300 sample size for each. Simulation results are presented in Table 4.10, where \(M_0\) indicates the underlying hazard model, while \(M_1\) indicates the fitting model. The simulation results presented on the top panel are obtained by utilising JM approach and setting argument of method=“spline-PH-aGH”, while method=“piecewise-PH-aGH” for the middle panel. Note that the JM approach with method=“piecewise-PH-aGH” may lead to seriously biased estimates for the survival parameters of the joint models. Our numerical experiments show that when there are only two-variates random effects, the R software lines with argument of method=“piecewise-PH-aGH”
can provide a good statistical performance, in terms of unbiased estimates, however, when the dimensions of random effects are three, such as in our simulation cases, the software lines of JM package may delivery outputs with substantial bias. And our experiments also show that the JM approach with option “piecewise-PH-GH” has the same problem, some improvements may needed with respect to the corresponding algorithm. Again, inference presented by the top panel shows that the issue of low coverage probability may also exist for the popular default current model. If the asymptotic properties hold with respect to the maximum likelihood estimation of survival parameters within the joint modelling framework, the coverage probability would be expected to be around 95%. However, it is only 74.5% for parameter $\eta_1$ and 83.6% for the association parameter $a$.

Simulation results presented in the top and bottom panels of Table 4.10 suggest similar statistical performance with respect to the statistical inference for the longitudinal process and the Cox regression coefficient $\eta$. It is interesting to note that the inference by using cumulative analysis may even be slightly better than that of JM approach. For instance, with respect to parameter $\alpha_2$, the JM approach leads to 19% relative bias, an issue aforementioned, while the cumulative analysis provides virtually unbiased estimation. The relative bias for $\eta_1$ and $\eta_2$ are also slightly less than that by utilising the JM approach. In addition, cumulative analysis has larger coverage probabilities for the parameters $\eta_1$ and $\eta_2$. In the meantime, the estimation for standard deviations of parameter estimators also seems quite reasonable, in the sense that the values of SDs and SEs presented in the bottom panel are close. It seems that the joint models of cumulative analysis with $B(a, b, t)$ may be a good alternative for the joint models of linear mixed-effects model and the default current value hazards model with $B^0(a, -, t)$, even when the point process of events in time is actually only affected by the current value of the longitudinal process. Nonetheless, some transformation may be required for the interpretation of models and parameters, since sometimes the influence pattern of longitudinal process on the failure time may be of particular interest.
Table 4.10: Cumulative analysis represents default current models. $M_0$ denotes the underlying models Monte Carlo data generated from; $M_1$ denotes the fitting models; $B(a, b = 6, t)$ indicates that data are fitted by model with $B(a, b, t)$ but with $b = 6$ fixed; “$M_1: B(a, -, t)$” indicates that inference outputs for the default current model are obtained from JM package with argument of method=“spline-PH-aGH” and method=“piecewise-PH-aGH” presented by top and middle panel, respectively. Note: analysis of three models/approaches are applied to the same Monte Carlo datasets and $\lambda_0(t) = 1$ of the underlying setting.

<table>
<thead>
<tr>
<th>Par</th>
<th>Truval.</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>RB(%)</th>
<th>CP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$M_0: B^0(a = 1, -, t)$; $M_1: B^0(a, -, t)$; (spline-)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>3.5</td>
<td>3.5024</td>
<td>0.0661</td>
<td>0.0662</td>
<td>0.1</td>
<td>95.3</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.3</td>
<td>-0.2971</td>
<td>0.0435</td>
<td>0.0379</td>
<td>0.8</td>
<td>88.6</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.1</td>
<td>0.0811</td>
<td>0.0221</td>
<td>0.0069</td>
<td>19.0</td>
<td>32.9</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.3</td>
<td>0.3005</td>
<td>0.0095</td>
<td>0.0167</td>
<td>0.2</td>
<td>100.0</td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>-1.5</td>
<td>-1.3936</td>
<td>0.1145</td>
<td>0.1055</td>
<td>7.1</td>
<td>74.5</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>-1.0</td>
<td>-0.9288</td>
<td>0.2163</td>
<td>0.1983</td>
<td>7.1</td>
<td>92.3</td>
</tr>
<tr>
<td>$a$(JM)</td>
<td>1.0</td>
<td>0.9558</td>
<td>0.0701</td>
<td>0.0656</td>
<td>4.4</td>
<td>83.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$M_0: B^0(a = 1, -, t)$; $M_1: B^0(a, -, t)$; (piecewise-)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>3.5</td>
<td>3.4965</td>
<td>0.0658</td>
<td>0.0657</td>
<td>0.1</td>
<td>96.0</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.3</td>
<td>-0.2972</td>
<td>0.0433</td>
<td>0.0377</td>
<td>0.9</td>
<td>89.0</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.1</td>
<td>0.0813</td>
<td>0.0222</td>
<td>0.0067</td>
<td>18.7</td>
<td>31.4</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.3</td>
<td>0.3001</td>
<td>0.0094</td>
<td>0.0167</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>-1.5</td>
<td>-1.0374</td>
<td>0.0836</td>
<td>0.0773</td>
<td>30.8</td>
<td>0.0</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>-1.0</td>
<td>-0.7095</td>
<td>0.1838</td>
<td>0.1826</td>
<td>29.1</td>
<td>59.5</td>
</tr>
<tr>
<td>$a$(JM)</td>
<td>1.0</td>
<td>0.8133</td>
<td>0.0548</td>
<td>0.0517</td>
<td>18.7</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$M_0: B^0(a = 1, -, t)$; $M_1: B(a, b = 6, t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>3.5</td>
<td>3.4942</td>
<td>0.0661</td>
<td>0.0688</td>
<td>0.2</td>
<td>96.7</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.3</td>
<td>-0.2963</td>
<td>0.0436</td>
<td>0.0363</td>
<td>1.2</td>
<td>87.0</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.1</td>
<td>0.0993</td>
<td>0.0234</td>
<td>0.0205</td>
<td>0.7</td>
<td>90.3</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.3</td>
<td>0.3001</td>
<td>0.0094</td>
<td>0.0099</td>
<td>0.0</td>
<td>97.0</td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>-1.5</td>
<td>-1.4431</td>
<td>0.1076</td>
<td>0.1216</td>
<td>3.8</td>
<td>90.3</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>-1.0</td>
<td>-0.9637</td>
<td>0.2034</td>
<td>0.2071</td>
<td>3.6</td>
<td>96.3</td>
</tr>
<tr>
<td>$a$(cum)</td>
<td>-6.3</td>
<td>-6.3034</td>
<td>0.4228</td>
<td>-</td>
<td>- &lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>$\hat{a} \int_0^t e^{-bt} du$</td>
<td>1.0</td>
<td>1.0506</td>
<td>0.0705</td>
<td>-</td>
<td>5.1</td>
<td>-</td>
</tr>
</tbody>
</table>
4.7 Discussion

In more recent studies, it has been recognized by investigators that sometimes not just the recent but also the past information of a longitudinal process may have instructive impact on the corresponding point process of events in time. In this study, the trajectory information of a longitudinal process has been accounted for in the analysis of time to event by taking integration of the trajectory. A special case is particularly studied here when the influence of past information tends to be weaker as time lag of present increases. The integral with a weighted curve is proposed by

\[
\int_{t_L}^{t} \gamma(t-u)m_i(u)du,
\]

where the weighted curve \( \gamma(t-u) \) is function of time lag. And the exponential function \( \gamma(u) = ae^{-bu} \) is proposed to characterize the specified influence pattern. When \( t_L \) is taken to be 0, the trajectory information ever since the baseline has been considered, the case we have been particularly discussing in the study. However, in some circumstances, only a certain period of past may need to consider and how to choose the \( t_L \) may also depend on the background knowledge of application.

The cumulative term of integration actually can be transformed into linear combination of derivatives of longitudinal trajectory with respect to time, as illustrated by formula (4.7). Some literatures also have pointed out that it may improve statistical inference by introducing not just the current value but also the random slope of longitudinal process into covariates of the survival regression model, where the linear mixed random-effects model with random intercept and slope is proposed for the longitudinal process, such as the study by Henderson et al. (2000) and the help document of JM package by Rizopoulos (2014) among others. They share some common points, for example, they introduce derivative functions of trajectory (random slope, for instance) into survival models, however, the proposed cumulative models with integration of longitudinal trajectory are in a more general form. On one hand, the cumulative analysis automatically accounts for all features of the longitudinal trajectory in the analysis of survival data. On the other hand, to some extent, the cumulative models are able to deal with the issue of time-varying coefficients, as illustrated by Figure 4.8/4.9. Furthermore, if the random effects of longitudinal process are directly introduced into survival regression model as covariates, such as the framework proposed by Wang et al. (2000), models may suffer from lack of reliability when the random effects have a
zero mean and a small variance, that is, the design matrix may be roughly singular. However, the proposed cumulative analysis framework would not have such problem.

In the study of simulation, it implies that the underlying values of baseline hazards may have influenced the statistical inference performance of proposed methodology, especially in terms of the plausible estimation of standard deviations for the parameter estimators. In the study, piecewise function is proposed to characterize the nonparametric baseline hazards function and it may be one of the reasons to cause such issue since, as demonstrated by Table 4.10, the JM approach has better performance by setting argument of method=“spline-PH-aGH” than that the argument of method=“piecewise-PH-aGH” does. It may be worthwhile to try to characterize the nonparametric baseline function by utilising B-splines or some other approaches to see if the inference can be improved therefore. On the other hand, in the last step of algorithm, Monte Carlo simulation approach has been used to approximate the posterior variance of functions of random effects given the observed data and parameter estimates, then the approximation of minus Hessian matrix and estimation of standard deviations for parameters estimators are obtained thereafter. It is somewhat inevitable for Monte Carlo simulation approach to introduce Monte Carlo error at the same time, and the adaptive Gauss-Hermite approach may be a good alternative, although they both are computationally intensive, but it is acceptable since they are only going to be needed once at the last iteration of proposed algorithm.

In the last section of simulation study, it is interesting to find that the cumulative models actually can somehow represent the default current model, in terms of statistical inference for the parameters of $\alpha$ and $\eta$, and also the association parameter after transformation, as demonstrated by Table 4.10. It seems pretty worthwhile to develop some further explore in this topic since this kind of cumulative models may be a quality alternative for the recent popular default current value models. On the other hand, in the study, we choose model between the default current and cumulative models by simply investigating the value of parameter $b$. When the $b$ value is large, we conclude that the past levels of longitudinal process may only have weak or even none effects on the present hazard rate and the default current value model should be a good choice. While when $b$ value is small, it suggests that the past levels of longitudinal process may have important impact on the present hazard and the effects may
be time varying as well. However, it is an interesting topic and also an open question to develop a systematic approach to choose which one is more appropriate model.

The special form of exponential function has been proposed to characterize the weighted curve due to the reasons that have been discussed earlier or after. However, it is still an open question to proposed other proper forms of function for the weighted curve $\gamma(t-u)$, or even some other framework to account for both the recent and past information of longitudinal process in the analysis of time-to-event data. In the meantime, it may also happen in practice that the levels of longitudinal profile of a certain period of past may have particularly important impact on the present hazards, rather than the levels of current, that is, the exponential function would no longer be able to capture such feature. The proposed models may incur some technical problems when there are some other time-dependent covariates in the longitudinal linear mixed random-effects models, apart from the time variables themselves. Furthermore, theoretic properties of estimations and model selection within the joint modelling framework, for example, are far from well-developed.

The fully exponential Laplace approximation has been proposed to approximate the posterior mean of functions of random effects given the observed data and current parameter estimates. According to our experimental experience, it seems that we may not have the issue of algorithm not being able to converge, like mentioned by other literatures. It usually takes around 30 minutes to get algorithm converged for a dataset of 500 sample size and three-dimensions of random effects for the linear mixed-effects model for the longitudinal process. When the parameters are close to the underlying values, the converge speed of iterations tends to be pretty slow and if it could be speeded up, the algorithm would be much compelling.

4.8 Comments: why not other ways to characterize the curve $\gamma(u)$?

In more recent studies, investigators have recognized that not only the current values of longitudinal processes but also the past levels may have influence on the disease. Both past and recent blood pressures, for example, may have instructive impact on the risk of on-going cardiovascular events. In the study, in terms of joint modelling,
we would like to also account for the past trajectory information into the survival analysis. The integration of trajectory profile of longitudinal process is proposed,
\[
\int_0^t \gamma(t - u)m_i(u)du = \int_0^t \gamma(u)m_i(t - u)du,
\]
(4.19)
where \(m_i(u)\) is the past trajectory of longitudinal process, history profile of blood pressure for example, and \(t\) is the present time. \(\gamma(t - u)\) is the weighted curve for the integration due to the fact that, in some or most of circumstances, the association becomes weaker as the time lag since present increases. At first glance, equation (4.19) seems to share the similar form of functional regression analysis for longitudinal data. On the other hand, without the algorithm of integration, it seems to turn to the time-varying coefficients problem that appears in Cox model frequently. And, in this section, we would like to discuss in more details why none of these framework works for our purpose of this study.


Yao, Müller and Wang (2005) introduced functional linear regression analysis for longitudinal data, where the underlying but unobservable sample consists of pairs of random trajectories \((X_i, Y_i), i = 1, 2, \ldots, m\), with square integrable predictor trajectories \(X_i\) and response trajectories \(Y_i\). The observations with measurement errors are taken at a random number of random measurement times for \(X_i\) and/or \(Y_i\) for the \(i^{th}\) subject. In that paper, they considered a functional linear regression model in which both the predictor \(X\) and response \(Y\) are smooth random functions,
\[
E(Y(t)|X) = \alpha(t) + \int_S \beta(s, t)X(s)ds,
\]
then a nonparametric functional principle component analysis was proposed for the model inference. And Müller and Stadtmüller (2005) also discussed a related topic but with response variable of scalar and proposed a generalized functional linear models with linear predictors
\[
\eta = \alpha + \int \beta(t)X(t)d\omega(t),
\]
and an orthonormal basis is proposed to expand the predictor process \(X(t)\) and the parameter function \(\beta(t)\). Back to our study, first, the longitudinal
process is characterized by linear mixed random-effects model and we would like to keep it in that way, which means we will not be able to expand the $m_i(u)$ and $\gamma(t-u)$ by an orthogonal basis. Second, more importantly, in functional data analysis, the integration is all over the whole space of time, for example. For the equation (4.19), however, the integral area is up to present time and the present hazards rate is of interest, which makes the essential difference with the functional data analysis. Therefore, the functional data analysis framework seems not applicable here, although they share some similarities in the form of integration.


- Local linear polynomial approach
  A popular approach to deal with varying-coefficient problem is to use the local linear techniques, for example, by Fan, Lin and Zhou (2006), Cai et al. (2007) and Song and Wang (2008), among others. In the local linear polynomial approach, it admits Taylor expansion, that is, for each given time $u_0$ and $u$ around $u_0$, $\gamma(u) \approx \gamma(u_0) + \gamma'(u_0)(u-u_0)$, and substitutes $\gamma(u)$ by the approximation into estimating equations to obtain the inference for $\gamma(u_0)$. And statistical inference for the whole curve $\gamma(u)$ is obtained by varying $u_0$ in the whole range space of $u$. However, in our study, the whole trajectory knowledge of weighted curve is required, or at least part of the trajectory information, and point inference only at each round does not help with the integration for the cumulative information at all. Therefore, the local linear polynomial approach cannot been applied here neither.

- B-splines approach
  It seems a thrilling approach to characterize the weighted curve $\gamma(u)$ by a sequence of B spline basis
  $$\gamma(u) = \sum_j \xi_j B_{j,k}(u),$$
  where $k$ is the order and $B_{j,k}(u)$, $j = 1, 2, .., K_\xi$ are basis, and the number $K_\xi$ of elements of parameter vector $\xi$ is decided by the number of knots applied in the analysis. A sequence of basis functions with order 4 and 10
inner knots of $K_\xi = 14$ are demonstrated in Figure 4.12. By doing so, we have
\[
\int_0^t \gamma(t-u)m_i(u)du = \int_0^t \gamma(u)m_i(t-u)du = \sum_j \xi_j \int_0^t B_{j,k}(u)m_i(t-u)du
\]
\[
= \sum_j \xi_j \int_0^t B_{j,k}(u) \left( \sum_{\ell=0}^q \alpha_{\ell i}(t-u)^\ell + X_i^T \beta \right) du
\]
\[
= \sum_j \xi_j \left( \sum_{\ell=0}^q \alpha_{\ell i} \int_0^t B_{j,k}(u)(t-u)^\ell du \right) + \left( \sum_j \xi_j \int_0^t B_{j,k}(u)du \right) X_i^T \beta
\]
\[
= \xi^T \tilde{B}_k(t)\alpha_i + \xi^T \tilde{B}_{k[:0]}(t)X_i^T \beta,
\]
where $\tilde{B}_k(t) = \{ \tilde{B}_{k[j,\ell]}(t) : j = 1, 2, \ldots, K_\xi, \ell = 0, 1, \ldots, q \}$ is a $K_\xi \times (q+1)$ matrix with elements of $\tilde{B}_{k[j,\ell]}(t) = \int_0^t B_{j,k}(u)(t-u)^\ell du$ and $\tilde{B}_{k[:0]}(t)$ is the first column of matrix $\tilde{B}_k(t)$, then survival Cox regression model turns out to be with linear regression predictors,
\[
\lambda_i(t) = \lambda_0(t) \exp \left( \int_0^t \gamma(t-u)m_i(u)du + \eta^T Z_i \right)
\]
\[
= \lambda_0(t) \exp \left( \xi^T \tilde{B}_k(t)\alpha_i + \xi^T \tilde{B}_{k[:0]}(t)X_i^T \beta + Z_i^T \eta \right),
\]
which is a really likeable transformation. However, very unfortunately, the design matrix $\tilde{B}_k(t)\alpha_i$ for the basis coefficients $\xi$ turns out to be collinear as illustrated by Figure 4.14, unless only small number of knots is being used, such as $K_\xi = 2$, however, by doing so, there is no accuracy at all for the weighted curve $\gamma(u)$ and it is pointless. It should have been expected according to the knowledge of integral function $B_{j,k}(u)(t-u)^\ell$, as illustrated by Figure 4.13 for instance, and the B-spline basis functions are actually periodic in some sense. The classical method of B-spline approach, which is full of flexibility, however, does not work for the integration framework.

- Simplest form of piecewise function for weighted curve $\gamma(u)$

One way is to go back to the simplest case, that is, by taking the piecewise function for the weighted curve,
\[
\gamma(u) = \sum_k \xi_k I(t_{k-1} < u \leq t_k),
\]
and the cumulative term turns out to be
\[
\int_0^t \gamma(u)m_i(t-u)du = \sum_k \xi_k \int_{t_{k-1}}^{\min(t_k,t)} m_i(t-u)du.
\]
However, it has the same problem of collinearity of design matrix for the coefficient parameters $\xi$, as B-spline approach does, since for each subject, we only have

$$\int_0^{T_i} \gamma(u) m_i(T_i - u) du = \sum_k \xi_k \int_{t_{k-1}}^{\min(t_k, T_i)} m_i(T_i - u) du.$$ 

Therefore, in the end, a reasonable form for the weighted curve $\gamma(u)$ is proposed by

$$\gamma(u) = ae^{-bu}, \quad u \geq 0, \ b > 0 \ and \ a \in R.$$ 

Other forms or ways to characterize the weighted curve function is still an open question.
4.9 Derivatives and log-concavity of parameters and Laplace approximation

Note: The formula in this section are working on the joint models of model (4.4) and model (4.9) for the reason to keep formulas short, and some slight change will make them to the joint models of model (4.4) and model (4.10).

Part A: Derivatives formulae for integration and EM algorithm with respect to \((\alpha_i; \alpha, \beta, \sigma^2, G; \zeta, a, \eta, b)\), respectively.
Figure 4.14: Collinearity of covariates for the B-spline coefficients, where we have that $\text{intBk}[j] = \sum_{i=0}^{q} \alpha_i \int_0^t B_{j,k}(u)(t - u)du$, the covariate for coefficient parameter $\xi_j$. 
with respect to $\alpha_i$, for the Laplace approximation purpose.

$$\frac{\partial}{\partial \alpha_i} \log \left( \Pr(T_i, \delta_i|\alpha_i, \theta) \Pr(Y_i|\alpha_i, \theta) \Pr(\alpha_i|\theta) \right) := U(\alpha_i)$$

$$= \delta_i B(a, b, T_i) - \int_0^{T_i} \lambda_0(u) \exp \left( B(a, b, u)^T \alpha_i + B_0(a, b, u) X_i^T \beta + Z_i^T \eta \right) B(a, b, u) du$$

$$+ A_i^T \Sigma_i^{-1} (Y_i - A_i \alpha_i - 1_n, X_i^T \beta) - G^{-1}(\alpha_i - \alpha)$$

$$= \delta_i B(a, b, T_i) - \int_0^{T_i} \lambda_0(u) \exp \left( B(a, b, u)^T \alpha_i + B_0(a, b, u) X_i^T \beta + Z_i^T \eta \right) B(a, b, u) du$$

$$+ \frac{1}{\sigma^2} A_i^T (Y_i - A_i \alpha_i - 1_n, X_i^T \beta) - G^{-1}(\alpha_i - \alpha),$$

$$\frac{\partial^2}{\partial \alpha_i^T \partial \alpha_i} \log \left( \Pr(T_i, \delta_i|\alpha_i, \theta) \Pr(Y_i|\alpha_i, \theta) \Pr(\alpha_i|\theta) \right) := \partial U(\alpha_i, \alpha_i^T)$$

$$= - \int_0^{T_i} \lambda_0(u) \exp \left( B(a, b, u)^T \alpha_i + B_0(a, b, u) X_i^T \beta + Z_i^T \eta \right) B(a, b, u) B(a, b, u)^T du$$

$$- A_i^T \Sigma_i^{-1} A_i - G^{-1}$$

$$= - \int_0^{T_i} \lambda_0(u) \exp \left( B(a, b, u)^T \alpha_i + B_0(a, b, u) X_i^T \beta + Z_i^T \eta \right) B(a, b, u) B(a, b, u)^T du$$

$$- \frac{1}{\sigma^2} A_i^T A_i - G^{-1}.$$

Newton-Raphson’s method is proposed to find the $\hat{\alpha}_i$ maximizing the quasi-loglikelihood

$$\log \left( \Pr(T_i, \delta_i|\alpha_i, \theta) \Pr(Y_i|\alpha_i, \theta) \Pr(\alpha_i|\theta) \right).$$

And

$$\alpha_i^{(i+1)} = \alpha_i^{(i)} - \partial U^{-1}(\alpha_i, \alpha_i^T)_{|\alpha_i = \alpha_i^{(i)}} U(\alpha_i^{(i)}).$$

with respect to $\alpha$

$$\frac{\partial}{\partial \alpha} \log \left( \Pr(T_i, \delta_i|\alpha_i, \theta) \Pr(Y_i|\alpha_i, \theta) \Pr(\alpha_i|\theta) \right)$$

$$= G^{-1}(\alpha_i - \alpha).$$

Thus, we have

$$\frac{\partial \ell(\theta; \text{data}, \theta_0)}{\partial \alpha} = \sum_{i=1}^{m} \left[ G^{-1}(\alpha_i - \alpha) \Pr(\alpha_i|T_i, \delta_i, Y_i, \theta_0) d\alpha_i \right].$$
and
\[
\hat{\alpha} = \frac{1}{n} \sum_{i=1}^{m} \alpha_i P_r(\alpha_i | T_i, \delta_i, Y_i, \theta_0) d\alpha_i
\]
\[
= \frac{1}{n} \sum_{i=1}^{m} E(\alpha_i | T_i, \delta_i, Y_i, \theta_0).
\]

• with respect to \( \beta \)
\[
\frac{\partial}{\partial \beta} \log \left( P_r(T_i, \delta_i | \alpha_i, \theta) P_r(Y_i | \alpha_i, \theta) P_r(\alpha_i | \theta) \right)
\]
\[
= \delta_i B_0(a, b, T_i) X_i + X_i 1_{n_i} \Sigma_i^{-1} (Y_i - A_i \alpha_i - 1_{n_i} X_i^T \beta)
\]
\[
- X_i \int_0^{T_i} \lambda_0(u) \exp \left( B(a, b, u)^T \alpha_i + B_0(a, b, u) X_i^T \beta + Z_i^T \eta \right) B_0(a, b, u) du
\]
\[
\approx \delta_i B_0(a, b, T_i) X_i + \frac{1}{\sigma^2} X_i 1_{n_i} (Y_i - A_i \alpha_i - 1_{n_i} X_i^T \beta)
\]
\[
- X_i \sum_{k=1}^{K_i} \zeta_k e^{Z_i^T \eta} \sum_{j \in \Omega_{ik}} dt_{kj} \omega_{kj} e^{B(a, b, u_{kj})^T \alpha_i + B_0(a, b, u_{kj}) X_{kj}^T \beta} B_0(a, b, u_{kj}),
\]

\[
\frac{\partial^2}{\partial \beta^T \partial \beta} \log \left( P_r(T_i, \delta_i | \alpha_i, \theta) P_r(Y_i | \alpha_i, \theta) P_r(\alpha_i | \theta) \right)
\]
\[
= - X_i X_i^T \left( \int_0^{T_i} \lambda_0(u) e^{B(a, b, u)^T \alpha_i + B_0(a, b, u) X_i^T \beta + Z_i^T \eta} B_0(a, b, u)^2 du + 1_{n_i} \Sigma_i^{-1} 1_{n_i} \right)
\]
\[
\approx - X_i X_i^T \left( \sum_{k=1}^{K_i} \zeta_k e^{Z_i^T \eta} \sum_{j \in \Omega_{ik}} dt_{kj} \omega_{kj} e^{B(a, b, u_{kj})^T \alpha_i + B_0(a, b, u_{kj}) X_{kj}^T \beta} B_0(a, b, u_{kj})^2 + \frac{n_i}{\sigma^2} \right),
\]

where \( K_i := \{ k : v_{k-1} < T_i \leq v_k \} \) and \( \Omega_{ik} = \{ j : v_{k-1} \leq u_{kj} \leq \min(v_k, T_i) \} \).

Thus, we have
\[
\sum_{i=1}^{m} \int \frac{\partial}{\partial \beta} \log \left( P_r(T_i, \delta_i | \alpha_i, \theta) P_r(Y_i | \alpha_i, \theta) P_r(\alpha_i | \theta) \right) P_r(\alpha_i | T_i, \delta_i, Y_i, \theta_0) d\alpha_i := U(\beta; \theta_0)
\]
\[
\approx \delta_i B_0(a, b, T_i) X_i + \frac{1}{\sigma^2} X_i 1_{n_i} (Y_i - A_i E(\alpha_i | \text{data, } \theta_0) - 1_{n_i} X_i^T \beta)
\]
\[
- X_i \sum_{k=1}^{K_i} \zeta_k e^{Z_i^T \eta} \sum_{j \in \Omega_{ik}} dt_{kj} \omega_{kj} e^{B_0(a, b, u_{kj}) X_{kj}^T \beta} E(e^{B(a, b, u_{kj})^T \alpha_i} | \text{data, } \theta_0) B_0(a, b, u_{kj}),
\]

\[
\sum_{i=1}^{m} \int \frac{\partial^2}{\partial \beta^T \partial \beta} \log \left( P_r(T_i, \delta_i, Y_i, \alpha_i | \theta) \right) P_r(\alpha_i | T_i, \delta_i, Y_i, \theta_0) d\alpha_i := \partial U(\beta, \beta^T; \theta_0)
\]
\[
\approx - \sum_{i=1}^{m} \left( \sum_{k=1}^{K_i} \zeta_k e^{Z_i^T \eta} \sum_{j \in \Omega_{ik}} dt_{kj} \omega_{kj} e^{B_0(a, b, u_{kj}) X_{kj}^T \beta} E(e^{B(a, b, u_{kj})^T \alpha_i} | \text{data, } \theta_0) B_0(a, b, u_{kj})^2 + \frac{n_i}{\sigma^2} \right) X_i X_i^T,
\]
where $E(e^{B(a,b,u_k)^T\alpha_i}) := E(e^{B(a,b,u_k)^T\alpha_i|data, \theta_0})$ and same for the following.

One-step Newton-Raphson method

$$\beta^{(t+1)} = \beta^{(t)} - \partial U(\beta, \beta^T; \theta_0)^{-1}U(\beta; \theta_0).$$

- with respect to $\sigma^2$

$$\frac{\partial}{\partial \sigma^2} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta)Pr(Y_i|\alpha_i, \theta)Pr(\alpha_i|\theta) \right)$$

$$= \frac{1}{2\sigma^4} \left( (Y_i - A_i\alpha_i - 1_nX_i^T\beta)^T(Y_i - A_i\alpha_i - 1_nX_i^T\beta) - \frac{n_i}{2\sigma^2} \right)$$

Thus, we have

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{m} \left( (Y_i - 1_nX_i^T\beta)^T(Y_i - 1_nX_i^T\beta - 2A_iE(\alpha_i)) + \text{Trace}(A_i^T A_i E(\alpha_i \alpha_i^T)) \right),$$

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

- with respect to $G$

Due to the fact that

$$\frac{\partial \log |det(X)|}{\partial X} = (X^{-1})^T = (X^T)^{-1},$$

$$\frac{\partial a^TX^{-1}b}{\partial X} = -X^{-T}ab^TX^{-T},$$

we have,

$$\frac{\partial}{\partial G} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta)Pr(Y_i|\alpha_i, \theta)Pr(\alpha_i|\theta) \right)$$

$$= \frac{1}{2} G^{-T}(\alpha_i - \alpha)(\alpha_i - \alpha)^TG^{-T} - \frac{1}{2} G^{-T},$$

and

$$\sum_{i=1}^{m} \int \frac{\partial}{\partial G} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta)Pr(Y_i|\alpha_i, \theta)Pr(\alpha_i|\theta) \right) Pr(\alpha_i|T_i, \delta_i, Y_i, \theta_0)d\alpha_i$$

$$= \sum_{i=1}^{m} \int \left( \frac{1}{2} G^{-T}(\alpha_i - \alpha)(\alpha_i - \alpha)^TG^{-T} - \frac{1}{2} G^{-T} \right) Pr(\alpha_i|T_i, \delta_i, Y_i, \theta_0)d\alpha_i$$

$$= \frac{1}{2} G^{-T}\left( \sum_{i=1}^{m} \left( E(\alpha_i \alpha_i^T|data, \theta_0) - E(\alpha_i|data, \theta_0)\alpha^T - \alpha E(\alpha_i^T|data, \theta_0) + \alpha \alpha^T \right)G^{-T} - nI \right)$$

Thus, we have

$$\dot{G} = \frac{1}{n} \sum_{i=1}^{m} \left( E(\alpha_i \alpha_i^T|data, \theta_0) - E(\alpha_i|data, \theta_0)\alpha^T - \alpha E(\alpha_i^T|data, \theta_0) \right) + \alpha \alpha^T.$$
• with respect to $\zeta$

\[
\frac{\partial}{\partial \zeta_k} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta) Pr(Y_i|\alpha_i, \theta) Pr(\alpha_i|\theta) \right) \\
= \delta_i(v_{k-1} < T_i \leq v_k) - e^{Z_i^T \eta} \int_{v_{k-1}}^{\min(v_k, T_i)} e^{B(b, u)T \alpha_i + B_0(b, u)X_i^T \beta} du \\
\approx \delta_i(v_{k-1} < T_i \leq v_k) - e^{Z_i^T \eta} \sum_{j \in \Omega_k} dt_{kj} \omega_{kj} e^{B(b, u)X_i^T \beta} \left( B(b, u)X_i^T \beta \right) I(v_{k-1} < T_i).
\]

Thus,

\[
\hat{\zeta}_k = \frac{\sum_{i=1}^{m} \delta_i(v_{k-1} < T_i \leq v_k)}{\sum_{i=1}^{m} e^{Z_i^T \eta} \sum_{j \in \Omega_k} dt_{kj} \omega_{kj} e^{B(b, u)X_i^T \beta} \left( B(b, u)X_i^T \beta \right) I(v_{k-1} < T_i)}.
\]

• with respect to $a$ and $\eta$

\[
\frac{\partial}{\partial a} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta) Pr(Y_i|\alpha_i, \theta) Pr(\alpha_i|\theta) \right) \\
= \delta_i \left( x(b, T_i)^T \alpha_i + B_0(b, T_i)X_i^T \beta \right) \\
- \int_0^{T_i} \lambda_0(u)e^{B(a, b, u)T \alpha_i + B_0(a, b, u)X_i^T \beta + Z_i^T \eta} \left( x(b, u)^T \alpha_i + B_0(b, u)X_i^T \beta \right) du \\
\approx K_i \left( x(b, T_i)^T \alpha_i + B_0(b, T_i)X_i^T \beta \right) \\
- \sum_{k=1}^{K_i} \zeta_k e^{Z_i^T \eta} \sum_{j \in \Omega_k} dt_{kj} \omega_{kj} e^{B(a, b, u)X_i^T \beta} \left( x(b, u)X_i^T \beta \right) \left( x(b, u)^T \alpha_i + B_0(b, u)X_i^T \beta \right),
\]

\[
\frac{\partial^2}{\partial a^2} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta) Pr(Y_i|\alpha_i, \theta) Pr(\alpha_i|\theta) \right) \\
= -\int_0^{T_i} \lambda_0(u)e^{B(a, b, u)T \alpha_i + B_0(a, b, u)X_i^T \beta + Z_i^T \eta} \left( x(b, u)^T \alpha_i + B_0(b, u)X_i^T \beta \right)^2 du \\
\approx -\sum_{k=1}^{K_i} \zeta_k e^{Z_i^T \eta} \sum_{j \in \Omega_k} dt_{kj} \omega_{kj} e^{B(a, b, u)X_i^T \beta} \left( x(b, u)X_i^T \beta \right)^2 \left( x(b, u)^T \alpha_i + B_0(b, u)X_i^T \beta \right)^2,
\]

\[
\frac{\partial^2}{\partial \eta^T \partial a} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta) Pr(Y_i|\alpha_i, \theta) Pr(\alpha_i|\theta) \right) \\
= -\left( \int_0^{T_i} \lambda_0(u)e^{B(a, b, u)T \alpha_i + B_0(a, b, u)X_i^T \beta + Z_i^T \eta} \left( x(b, u)^T \alpha_i + B_0(b, u)X_i^T \beta \right) du \right) Z_i^T \\
\approx -\sum_{k=1}^{K_i} \zeta_k e^{Z_i^T \eta} \sum_{j \in \Omega_k} dt_{kj} \omega_{kj} e^{B(a, b, u)X_i^T \beta} \left( x(b, u)X_i^T \beta \right) Z_i^T \left( x(b, u)^T \alpha_i + B_0(b, u)X_i^T \beta \right),
\]
and that
\[
\frac{\partial}{\partial \eta} \log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right) \\
= \left( \delta_i - \int_0^{T_i} \lambda_0(u) e^{B(a,b,u)^T \alpha_i + B_0(a,b,u) X_i^T \beta + Z_i^T \eta} du \right) Z_i \\
\approx \left( \delta_i - \sum_{k=1}^{K_i} \zeta_k e^{Z_i^T \eta} \sum_{j \in \Omega_i} dt_{kj} \omega_{kj} e^{B(a,b,u_k)^T \alpha_i + B_0(a,b,u_k) X_i^T \beta} \right) Z_i,
\]

\[
\frac{\partial^2}{\partial \eta^2} \log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right) \\
= - \left( \int_0^{T_i} \lambda_0(u) e^{B(a,b,u)^T \alpha_i + B_0(a,b,u) X_i^T \beta + Z_i^T \eta} \left( B(b,u)^T \alpha_i + B_0(b,u) X_i^T \beta \right) du \\
+ \delta_i \left( B(b,T_i)^T \alpha_i + B_0(b,T_i) X_i^T \beta \right) \right) Pr(\alpha_i | T_i, \delta_i, Y_i, \theta) d\alpha_i \\
\approx \delta_i \left( B(b,T_i)^T E(\alpha_i) + B_0(b,T_i) X_i^T \beta \right) - \left( \sum_{k=1}^{K_i} \zeta_k e^{Z_i^T \eta} \sum_{j \in \Omega_i} dt_{kj} \omega_{kj} e^{B_0(a,b,u_k) X_i^T \beta} \\
( B(b,u_k)^T E(\alpha_i e^{B(a,b,u_k)^T \alpha_i}) + B_0(b,u_k) X_i^T \beta E(e^{B(a,b,u_k)^T \alpha_i}) ) \right),
\]

\[
\sum_{i=1}^{m} \int \frac{\partial}{\partial a} \log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right) Pr(\alpha_i | T_i, \delta_i, Y_i, \theta) d\alpha_i := U(a; \theta_0) \\
= \sum_{i=1}^{m} \int \left( - \int_0^{T_i} \lambda_0(u) e^{B(a,b,u)^T \alpha_i + B_0(a,b,u) X_i^T \beta + Z_i^T \eta} \left( B(b,u)^T \alpha_i + B_0(b,u) X_i^T \beta \right) du \\
+ \delta_i \left( B(b,T_i)^T \alpha_i + B_0(b,T_i) X_i^T \beta \right) \right) Pr(\alpha_i | T_i, \delta_i, Y_i, \theta) d\alpha_i \\
\approx \delta_i \left( B(b,T_i)^T E(\alpha_i) + B_0(b,T_i) X_i^T \beta \right) - \left( \sum_{k=1}^{K_i} \zeta_k e^{Z_i^T \eta} \sum_{j \in \Omega_i} dt_{kj} \omega_{kj} e^{B_0(a,b,u_k) X_i^T \beta} \\
( B(b,u_k)^T E(\alpha_i e^{B(a,b,u_k)^T \alpha_i}) + B_0(b,u_k) X_i^T \beta E(e^{B(a,b,u_k)^T \alpha_i}) ) \right),
\]

\[
\sum_{i=1}^{m} \int \frac{\partial}{\partial \eta} \log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right) Pr(\alpha_i | T_i, \delta_i, Y_i, \theta) d\alpha_i := U(\eta; \theta_0) \\
= \sum_{i=1}^{m} Z_i \int \left( \delta_i - \int_0^{T_i} \lambda_0(u) e^{B(a,b,u)^T \alpha_i + B_0(a,b,u) X_i^T \beta + Z_i^T \eta} du \right) Pr(\alpha_i | T_i, \delta_i, Y_i, \theta) d\alpha_i \\
\approx \sum_{i=1}^{m} \left( \delta_i - \sum_{k=1}^{K_i} \zeta_k e^{Z_i^T \eta} \sum_{j \in \Omega_i} dt_{kj} \omega_{kj} e^{B_0(a,b,u_k) X_i^T \beta} E(e^{B(a,b,u_k)^T \alpha_i}) \right) Z_i,
\]
\[ \sum_{i=1}^{m} \int \frac{\partial^2}{\partial a^2} \log \left( P_r(T_i, \delta_i, Y_i, \alpha_i|\theta) \right) P_r(\alpha_i|T_i, \delta_i, Y_i, \theta_0) d\alpha_i := \partial U(a, a; \theta_0) \]
\[ \approx - \sum_{i=1}^{K_i} \sum_{k=1}^{K_i} \zeta_k \zeta_i \eta_i \eta_k \sum_{j \in \Omega_k} dt_{kj} \omega_{kj} e^{B_0(b, u_k)} X_i^T \beta \sum_{k=1}^{K_i} \sum_{i=1}^{m} \zeta_k \zeta_i \eta_i \eta_k \sum_{j \in \Omega_k} dt_{kj} \omega_{kj} e^{B_0(b, u_k)} X_i^T \beta \sum_{k=1}^{K_i} \sum_{i=1}^{m} \zeta_k \zeta_i \eta_i \eta_k \sum_{j \in \Omega_k} dt_{kj} \omega_{kj} e^{B_0(b, u_k)} X_i^T \beta \]

where \( g(b, t) = B(b, t)^T \alpha_i + B_0(b, t) X_i^T \beta \) and \( g(a, b, t) = B(a, b, t)^T \alpha_i + B_0(b, a, t) X_i^T \beta \),
that is, \( g(a, b, t) = a g(b, t) \).

\[ \sum_{i=1}^{m} \int \frac{\partial^2}{\partial \eta^2} \log \left( P_r(T_i, \delta_i, Y_i, \alpha_i|\theta) \right) P_r(\alpha_i|T_i, \delta_i, Y_i, \theta_0) d\alpha_i := \partial U(a, \eta^T; \theta_0) \]
\[ = - \sum_{i=1}^{m} \left( \int \int_0^{T_i} \lambda_0(u) e^{B_0(a, b, u) T \alpha_i + B_0(a, b, u) X_i^T \beta + Z_i \eta} P_r(\alpha_i|T_i, \delta_i, Y_i, \theta_0) dud\alpha_i \right) Z_i Z_i^T \]
\[ \approx - \sum_{i=1}^{m} \left( \sum_{k=1}^{K_i} \zeta_k \zeta_i \eta_i \eta_k \sum_{j \in \Omega_k} dt_{kj} \omega_{kj} e^{B_0(b, u_k)} X_i^T \beta \sum_{k=1}^{K_i} \sum_{i=1}^{m} \zeta_k \zeta_i \eta_i \eta_k \sum_{j \in \Omega_k} dt_{kj} \omega_{kj} e^{B_0(b, u_k)} X_i^T \beta \sum_{k=1}^{K_i} \sum_{i=1}^{m} \zeta_k \zeta_i \eta_i \eta_k \sum_{j \in \Omega_k} dt_{kj} \omega_{kj} e^{B_0(b, u_k)} X_i^T \beta \right) Z_i Z_i^T, \]

One-step Newton-Raphson method,
\[
\begin{pmatrix} a \\ \eta \end{pmatrix}^{(i+1)} = \begin{pmatrix} a \\ \eta \end{pmatrix}^{(i)} - \begin{pmatrix} \partial U(a, a; \theta_0) & \partial U(a, \eta^T; \theta_0) \\ \partial U(\eta, a; \theta_0) & \partial U(\eta, \eta^T; \theta_0) \end{pmatrix}^{-1} \left|_{\theta_0=\theta^{(i)}} \begin{pmatrix} U(a; \theta_0) \\ U(\eta; \theta_0) \end{pmatrix} \right|_{\theta_0=\theta^{(i)}}.
\]

* with respect to \( b \)
Firstly, with the assumption of the commutation of differentiation and integration,
\[ \frac{\partial}{\partial b} B_\ell(a, b, t) = \frac{\partial}{\partial b} \int_0^{t_0(t)} a e^{-bu(t - u)} \ell du = \frac{\partial}{\partial b} \left( e^{-bt} \int_0^{t_0(t)} a e^{b(t - u)} (t - u) \ell du \right) \\
= (-t) e^{-bt} \int_0^{t_0(t)} a e^{b(t - u)} (t - u) \ell du + e^{-bt} \int_0^{t_0(t)} a e^{b(t - u)} (t - u) \ell+1 du \\
= (-t) \int_0^{t_0(t)} a e^{-bu(t - u)} \ell du + \int_0^{t_0(t)} a e^{-bu(t - u)} \ell+1 du \\
= (-t) B_\ell(a, b, t) + B_{\ell+1}(a, b, t). \]

Denote \( ^*B(a, b, t) = \left( B_s(a, b, t), B_{s+1}(a, b, t), \ldots, B_{q+s}(a, b, t) \right)^T \), \( s = 1, 2 \) and \( q \) is the degree of polynomial function of time of linear mixed models for the longitudinal process. Thus, we have
\[ \frac{\partial}{\partial b} B(a, b, t) = (-t)B(a, b, t) + 1B(a, b, t), \]
\[ \frac{\partial^2}{\partial b^2} B(a, b, t) = 2B(a, b, t) - 1B(a, b, t)2t + B(a, b, t)t^2. \]

Denote \( g_s(a, b, t) = ^*B(a, b, t)\alpha_i + B_s(a, b, t)X_i^T\beta, s = 1, 2, \) that is, then we have,
\[ \frac{\partial}{\partial b} g(a, b, t) = \frac{\partial}{\partial b} \left( B(a, b, t)^T \alpha_i + B_0(a, b, t)X_i^T\beta \right) \\
= \left( 1B(a, b, t) - tB(a, b, t) \right) \alpha_i + \left( B_1(a, b, t) - tB_0(a, b, t) \right) X_i^T\beta \\
= g_1(a, b, t) - tg(a, b, t), \]
and
\[ \frac{\partial^2}{\partial b^2} g(a, b, t) = \left( 2B(a, b, t) - 1B(a, b, t)2t + B(a, b, t)t^2 \right) \alpha_i \\
+ \left( B_2(a, b, t) - B_1(a, b, t)2t + B(a, b, t)t^2 \right) X_i^T\beta \\
= g_2(a, b, t) - 2tg_1(a, b, t) + t^2g(a, b, t). \]

Note that, when the \( B_\ell(a, b, t) = a(t^\ell + \int_0^{t_0(t)} e^{-bu(t - u)} \ell du) \) is proposed in the models, that is, when the current value of longitudinal process is addressed out
for the event hazards, we have\( g^c(a, b, t) = B^c(a, b, t)\alpha_i + B_0^c(a, b, t)X_i^T\beta, \) and

\[
\frac{\partial}{\partial b} g^c(a, b, t) = \frac{\partial}{\partial b} \left( B(a, b, t)^T\alpha_i + B_0(a, b, t)X_i^T\beta \right) \\
= \left( B(a, b, t) - tB(a, b, t) \right)\alpha_i + \left( B_1(a, b, t) - tB_0(a, b, t) \right)X_i^T\beta \\
= g_1(t, a, b, t) - tg(a, b, t) = \frac{\partial}{\partial b} g(a, b, t),
\]

\[
\frac{\partial^2}{\partial b^2} g^c(a, b, t) = g_2(a, b, t) - 2tg_1(a, b, t) + t^2 g(a, b, t) = \frac{\partial^2}{\partial b^2} g(a, b, t).
\]

Thus, we have,

\[
\frac{\partial}{\partial b} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta)Pr(Y_i|\alpha_i, \theta)Pr(\alpha_i|\theta) \right) \\
= \delta_i \left( \frac{\partial}{\partial b} B(a, b, T_i)\alpha_i + \frac{\partial}{\partial b} B_0(a, b, T_i)X_i^T\beta \right) \\
- \int_0^{T_i} \lambda_0(u)e^{Z^TÎn}e^{B(a,b,u)^T\alpha_i + B_0(a,b,u)X_i^T\beta} \left( \frac{\partial}{\partial b} B(a, b, u)\alpha_i + \frac{\partial}{\partial b} B_0(a, b, u)X_i^T\beta \right) du \\
\approx \delta_i \left( g_1(a, b, T_i) - T_ig(a, b, T_i) \right) \\
- \sum_{k=1}^{K_i} \zeta_k e^{Z^TÎn} \sum_{j\in\Omega_{ik}} d_{kj}\omega_{kj} e^{B(a,b,u)^T\alpha_i + B_0(a,b,u)X_i^T\beta} \left( g_1(a, b, u_kj) - u_kjg(a, b, u_kj) \right).
\]

\[
\sum_{i=1}^{m} \int \frac{\partial}{\partial b} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta)Pr(Y_i|\alpha_i, \theta)Pr(\alpha_i|\theta) \right) Pr(\alpha_i|T_i, \delta_i, Y_i, \theta_0)d\alpha_i := U(b; \theta_0) \\
= \sum_{i=1}^{m} \left[ \delta_i \left( g_1(a, b, T_i) - T_ig(a, b, T_i) \right) \\
- \int_0^{T_i} \lambda_0(u)e^{Z^TÎn}e^{B(a,b,u)^T\alpha_i + B_0(a,b,u)X_i^T\beta} \left( g_1(a, b, u) - u_ig(a, b, u) \right) du \right] Pr(\alpha_i|T_i, \delta_i, Y_i, \theta_0)d\alpha_i \\
\approx \sum_{i=1}^{m} \left[ \delta_i \left( E(g_1(a, b, T_i)|data, \theta_0) - T_ig(a, b, T_i)|data, \theta_0) \right) \\
- \sum_{k=1}^{K_i} \zeta_k e^{Z^TÎn} \sum_{j\in\Omega_{ik}} d_{kj}\omega_{kj} e^{B(a,b,u_kj)X_i^T\beta} E\left( e^{B(a,b,u_kj)^T\alpha_i} \left( g_1(a, b, u_kj) - u_kjg(a, b, u_kj) \right) \right) \right].
\]
And,
\[
\frac{\partial^2}{\partial \theta^2} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta) Pr(Y_i|\alpha_i, \theta) Pr(\alpha_i|\theta) \right) \\
= \delta_i \left( g_2(a, b, T_i) - 2T_ig_1(a, b, T_i) + T_i^2g(a, b, T_i) \right) \\
- \int_0^{T_i} \lambda_0(u)e^{2ZT\eta e^{B(a,b,u)T\alpha + B_0(a,b,u)X_T\beta}} \left( g_1(a, b, u) - ug(a, b, u) \right)^2 du \\
- \int_0^{T_i} \lambda_0(u) e^{2ZT\eta e^{B(a,b,u)T\alpha + B_0(a,b,u)X_T\beta}} \left( g_2(a, b, u) - 2tg_1(a, b, u) + t^2g(a, b, u) \right) du \\
\approx \delta_i \left( g_2(a, b, T_i) - 2T_ig_1(a, b, T_i) + T_i^2g(a, b, T_i) \right) \\
- \sum_{k=1}^{K_i} \zeta_ke^{ZT\eta} \sum_{j \in \Omega_k} dt_k \omega_k \eta e^{B(a,b,u_k)T\alpha + B_0(a,b,u_k)X_T\beta} \left( g_1(a, b, u_k) - u_kg(a, b, u_k) \right)^2 \\
- \sum_{k=1}^{K_i} \zeta_ke^{ZT\eta} \sum_{j \in \Omega_k} dt_k \omega_k \eta e^{g(a,b,u_k)} \left( g_2(a, b, u_k) - 2u_kg_1(a, b, u_k) + u_k^2g(a, b, u_k) \right).
\]

\[
\sum_{i=1}^{m} \int \frac{\partial^2}{\partial \theta^2} \log \left( Pr(T_i, \delta_i, Y_i, \alpha_i|\theta) \right) Pr(\alpha_i|T_i, \delta_i, Y_i, \theta_0) d\alpha_i := \partial U(b, \theta_0)
\]
\[
= \sum_{i=1}^{m} \int \left[ \delta_i \left( g_2(a, b, T_i) - 2T_ig_1(a, b, T_i) + T_i^2g(a, b, T_i) \right) \\
- \int_0^{T_i} \lambda_0(u)e^{2ZT\eta e^{B(a,b,u)T\alpha + B_0(a,b,u)X_T\beta}} \left( g_2(a, b, u) - 2ug_1(a, b, u) + u^2g(a, b, u) \right) du \\
- \int_0^{T_i} \lambda_0(u) e^{2ZT\eta e^{g(a,b,u)}} \left( g_1(a, b, u) - ug(a, b, u) \right)^2 du \right] Pr(\alpha_i|T_i, \delta_i, Y_i, \theta_0) d\alpha_i
\]
\[
\approx \sum_{i=1}^{m} \left[ \delta_i \left( E(g_2(a, b, T_i)|data, \theta_0) - 2T_1E(g_1(a, b, T_i)|data, \theta_0) + T_i^2E(g(a, b, T_i)|data, \theta_0) \right) \\
- \sum_{k=1}^{K_i} \zeta_ke^{ZT\eta} \sum_{j \in \Omega_k} dt_k \omega_k \eta e^{B(a,b,u_k)T\alpha} \left( e^{B(a,b,u_k)T\alpha} (g_2(u_k) - 2u_kg_1(u_k) + u_k^2g(u_k)) \right) \\
- \sum_{k=1}^{K_i} \zeta_ke^{ZT\eta} \sum_{j \in \Omega_k} dt_k \omega_k \eta e^{B(a,b,u_k)T\alpha} \left( e^{B(a,b,u_k)T\alpha} (g_1(a, b, u_k) - u_kg(a, b, u_k))^2 \right),
\]

where $g(u_k)$ is short for $g(a, b, u_k)$ and $g_1(u_k)$ and $g_2(u_k)$ similarly. The one-step Newton-Raphson method is applied,
\[
b^{(t+1)} = b^{(t)} - (\partial U(b, \theta; \theta_0))^{-1}|_{\theta_0 = \theta^{(t)}} U(b, \theta_0)|_{\theta_0 = \theta^{(t)}}.
\]

Due to the complexity of log-likelihood function with respect to $b$, our programming experience shows that Newton-Raphson algorithm usually doesn’t serve
well with respect to the parameter $b$, and some other numerical method might be required. In this work, R command *nlminb* is utilised to assist obtaining the estimators for parameter $b$. And since the quasi-likelihood function

$$\ell_q(\theta) \propto \delta_i(B(a, b, T_i)^T \alpha_i + B_0(a, b, T_i)X_i^T \beta) - \int_0^{T_i} \lambda_0(u)e^{B(a, b, u)^T \alpha_i + B_0(a, b, u)X_i^T \beta + Z_i^T \eta} du,$$

By utilising the log-normal property, the maximizing likelihood function with respect to $b$ turns out to be,

$$\sum_i^m \delta_i \left( B(a, b, T_i)^T \tilde{\mu}_{\alpha_i} + B_0(a, b, T_i)X_i^T \beta \right)$$

$$- e^{Z_i^T \eta} \int_0^{T_i} \lambda_0(u)e^{B_0(a, b, u)X_i^T \beta} \exp \left( B(a, b, u)^T \tilde{\mu}_{\alpha_i} + \frac{1}{2} B(a, b, u)^T \tilde{\Sigma}_{\alpha_i} B(a, b, u) \right) du$$

$$\approx \sum_i^m \delta_i \left( B(a, b, T_i)^T \tilde{\mu}_{\alpha_i} + B_0(a, b, T_i)X_i^T \beta \right)$$

$$- \sum_{k=1}^{K_i} \zeta_k e^{Z_i^T \eta} \sum_{j \in \Omega_k} dt_{k,j} \omega_{k,j} \exp \left( B_0(a, b, u_{k,j})X_i^T \beta \right) \exp \left( B(a, b, u_{k,j})^T \tilde{\mu}_{\alpha_i} + \frac{1}{2} B(a, b, u_{k,j})^T \tilde{\Sigma}_{\alpha_i} B(a, b, u_{k,j}) \right),$$

where we define $\tilde{\mu}_{\alpha_i} := E(\alpha_i | T_i, \delta_i, Y_i, \theta)$ and $\tilde{\Sigma}_{\alpha_i} := Var(\alpha_i | T_i, \delta_i, Y_i, \theta)$

- The second derivatives of quasi-loglikelihood are all negative, we conclude that the maximum estimators of parameters $\theta$ exist. We also note that it is not obvious or easy to tell the sign of the second derivative with respective to $b$, but the curve of quasi-loglikelihood function drawn by numerical method shows the existence of the maximum point. Note,

$$g(a, b, t) = \sum_{\ell=0}^q \alpha_{\ell,i} \int_0^{t_0(t)} a e^{-bu} (t - u)^{\ell} du + X_i^T \beta \int_0^{t_0(t)} a e^{-bu} du,$$

$$\frac{\partial}{\partial b} g(a, b, t) = \sum_{\ell=0}^q \alpha_{\ell,i} \int_0^{t_0(t)} a e^{-bu} (-u) (t - u)^{\ell} du + X_i^T \beta \int_0^{t_0(t)} a e^{-bu} (-u) du,$$

$$\frac{\partial^2}{\partial b^2} g(a, b, t) = \sum_{\ell=0}^q \alpha_{\ell,i} \int_0^{t_0(t)} a e^{-bu} (-u)^2 (t - u)^{\ell} du + X_i^T \beta \int_0^{t_0(t)} a e^{-bu} (-u)^2 du.$$

**Part B: Derivatives formula for the Hessian matrix.**

As discussed before, the Fisher information matrix $I(\theta)$ is estimated by minus Hessian matrix, and we have that

$$\frac{\partial^2 \log L}{\partial \theta^T \partial \theta} = \sum_{i=1}^m \left( E(\partial U_i(\theta, \theta^T; \theta) | T_i, \delta_i, Y_i) + Cov(U_i(\theta; \theta) | T_i, \delta_i, Y_i) \right),$$
where
\[
\partial U_i(\theta, \theta^T; \theta) := \frac{\partial^2}{\partial \theta^T \partial \theta} \log \left( (T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right),
\]
\[
U_i(\theta; \theta) := \frac{\partial}{\partial \theta} \log \left( (T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right).
\]

Thus,
\[
\widehat{\text{Cov}}(\theta) = -\left( \frac{\partial^2 \log L}{\partial \theta^T \partial \theta} \right)^{-1}.
\]

In details, the minus Hessian matrix is calculated by
\[
\frac{\partial^2 \log L}{\partial \theta^T \partial \theta} = \begin{pmatrix}
\frac{\partial^2 \log L}{\partial \alpha \partial \alpha} & \frac{\partial^2 \log L}{\partial \alpha \partial \delta} & \frac{\partial^2 \log L}{\partial \alpha \partial \beta} & \frac{\partial^2 \log L}{\partial \alpha \partial \theta} & \frac{\partial^2 \log L}{\partial \delta \partial \alpha} & \frac{\partial^2 \log L}{\partial \delta \partial \delta} & \frac{\partial^2 \log L}{\partial \delta \partial \beta} & \frac{\partial^2 \log L}{\partial \delta \partial \theta} & \frac{\partial^2 \log L}{\partial \beta \partial \alpha} & \frac{\partial^2 \log L}{\partial \beta \partial \delta} & \frac{\partial^2 \log L}{\partial \beta \partial \beta} & \frac{\partial^2 \log L}{\partial \beta \partial \theta} & \frac{\partial^2 \log L}{\partial \theta \partial \alpha} & \frac{\partial^2 \log L}{\partial \theta \partial \delta} & \frac{\partial^2 \log L}{\partial \theta \partial \beta} & \frac{\partial^2 \log L}{\partial \theta \partial \theta} \\
\frac{\partial^2 \log L}{\partial \alpha \partial \alpha} & \frac{\partial^2 \log L}{\partial \alpha \partial \delta} & \frac{\partial^2 \log L}{\partial \alpha \partial \beta} & \frac{\partial^2 \log L}{\partial \alpha \partial \theta} & \frac{\partial^2 \log L}{\partial \delta \partial \alpha} & \frac{\partial^2 \log L}{\partial \delta \partial \delta} & \frac{\partial^2 \log L}{\partial \delta \partial \beta} & \frac{\partial^2 \log L}{\partial \delta \partial \theta} & \frac{\partial^2 \log L}{\partial \beta \partial \alpha} & \frac{\partial^2 \log L}{\partial \beta \partial \delta} & \frac{\partial^2 \log L}{\partial \beta \partial \beta} & \frac{\partial^2 \log L}{\partial \beta \partial \theta} & \frac{\partial^2 \log L}{\partial \theta \partial \alpha} & \frac{\partial^2 \log L}{\partial \theta \partial \delta} & \frac{\partial^2 \log L}{\partial \theta \partial \beta} & \frac{\partial^2 \log L}{\partial \theta \partial \theta} \\
\frac{\partial^2 \log L}{\partial \alpha \partial \alpha} & \frac{\partial^2 \log L}{\partial \alpha \partial \delta} & \frac{\partial^2 \log L}{\partial \alpha \partial \beta} & \frac{\partial^2 \log L}{\partial \alpha \partial \theta} & \frac{\partial^2 \log L}{\partial \delta \partial \alpha} & \frac{\partial^2 \log L}{\partial \delta \partial \delta} & \frac{\partial^2 \log L}{\partial \delta \partial \beta} & \frac{\partial^2 \log L}{\partial \delta \partial \theta} & \frac{\partial^2 \log L}{\partial \beta \partial \alpha} & \frac{\partial^2 \log L}{\partial \beta \partial \delta} & \frac{\partial^2 \log L}{\partial \beta \partial \beta} & \frac{\partial^2 \log L}{\partial \beta \partial \theta} & \frac{\partial^2 \log L}{\partial \theta \partial \alpha} & \frac{\partial^2 \log L}{\partial \theta \partial \delta} & \frac{\partial^2 \log L}{\partial \theta \partial \beta} & \frac{\partial^2 \log L}{\partial \theta \partial \theta} \\
\frac{\partial^2 \log L}{\partial \alpha \partial \alpha} & \frac{\partial^2 \log L}{\partial \alpha \partial \delta} & \frac{\partial^2 \log L}{\partial \alpha \partial \beta} & \frac{\partial^2 \log L}{\partial \alpha \partial \theta} & \frac{\partial^2 \log L}{\partial \delta \partial \alpha} & \frac{\partial^2 \log L}{\partial \delta \partial \delta} & \frac{\partial^2 \log L}{\partial \delta \partial \beta} & \frac{\partial^2 \log L}{\partial \delta \partial \theta} & \frac{\partial^2 \log L}{\partial \beta \partial \alpha} & \frac{\partial^2 \log L}{\partial \beta \partial \delta} & \frac{\partial^2 \log L}{\partial \beta \partial \beta} & \frac{\partial^2 \log L}{\partial \beta \partial \theta} & \frac{\partial^2 \log L}{\partial \theta \partial \alpha} & \frac{\partial^2 \log L}{\partial \theta \partial \delta} & \frac{\partial^2 \log L}{\partial \theta \partial \beta} & \frac{\partial^2 \log L}{\partial \theta \partial \theta} \\
\frac{\partial^2 \log L}{\partial \alpha \partial \alpha} & \frac{\partial^2 \log L}{\partial \alpha \partial \delta} & \frac{\partial^2 \log L}{\partial \alpha \partial \beta} & \frac{\partial^2 \log L}{\partial \alpha \partial \theta} & \frac{\partial^2 \log L}{\partial \delta \partial \alpha} & \frac{\partial^2 \log L}{\partial \delta \partial \delta} & \frac{\partial^2 \log L}{\partial \delta \partial \beta} & \frac{\partial^2 \log L}{\partial \delta \partial \theta} & \frac{\partial^2 \log L}{\partial \beta \partial \alpha} & \frac{\partial^2 \log L}{\partial \beta \partial \delta} & \frac{\partial^2 \log L}{\partial \beta \partial \beta} & \frac{\partial^2 \log L}{\partial \beta \partial \theta} & \frac{\partial^2 \log L}{\partial \theta \partial \alpha} & \frac{\partial^2 \log L}{\partial \theta \partial \delta} & \frac{\partial^2 \log L}{\partial \theta \partial \beta} & \frac{\partial^2 \log L}{\partial \theta \partial \theta} \\
\frac{\partial^2 \log L}{\partial \alpha \partial \alpha} & \frac{\partial^2 \log L}{\partial \alpha \partial \delta} & \frac{\partial^2 \log L}{\partial \alpha \partial \beta} & \frac{\partial^2 \log L}{\partial \alpha \partial \theta} & \frac{\partial^2 \log L}{\partial \delta \partial \alpha} & \frac{\partial^2 \log L}{\partial \delta \partial \delta} & \frac{\partial^2 \log L}{\partial \delta \partial \beta} & \frac{\partial^2 \log L}{\partial \delta \partial \theta} & \frac{\partial^2 \log L}{\partial \beta \partial \alpha} & \frac{\partial^2 \log L}{\partial \beta \partial \delta} & \frac{\partial^2 \log L}{\partial \beta \partial \beta} & \frac{\partial^2 \log L}{\partial \beta \partial \theta} & \frac{\partial^2 \log L}{\partial \theta \partial \alpha} & \frac{\partial^2 \log L}{\partial \theta \partial \delta} & \frac{\partial^2 \log L}{\partial \theta \partial \beta} & \frac{\partial^2 \log L}{\partial \theta \partial \theta}
\end{pmatrix} := \begin{pmatrix}
I_{11} \\
I_{12} \\
I_{21} \\
I_{22}
\end{pmatrix}.
\]

In the following context, for short, we denote
\[
\ell_{iq}(\theta) := \log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right).
\]

- \( \zeta = (\zeta_1, \zeta_2, \ldots, \zeta_K)^T \)

Since \( \log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right) \propto \delta_i \log(I_{iv}^T \zeta) - \omega_{iv}^T \zeta \), we have
\[
\frac{\partial}{\partial \zeta} \log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right) = \frac{\delta_i}{I_{iv}^T \zeta} I_{iv} - \omega_{iv},
\]

where \( I_{iv} := \{(v_{k-1} < T_i \leq v_k) : k = 1, K \}^T \), and \( I_{iv} = (1, 0, \cdots, 0)^T \), for instance, if \( v_0 < T_i \leq v_1 \). And
\[
\omega_{iv} := \begin{pmatrix}
e^{Z^T_i \eta \sum_{v_0}^{v_1} \exp(B(a, b, u) \alpha_i + B_0(a, b, u) X_i^T \beta) du} \\
e^{Z^T_i \eta \sum_{v_1}^{v_2} \exp(B(a, b, u) \alpha_i + B_0(a, b, u) X_i^T \beta) du} \\
\vdots \\
e^{Z^T_i \eta \sum_{v_{K-1}}^{v_{K-1}} \exp(B(a, b, u) \alpha_i + B_0(a, b, u) X_i^T \beta) du} \\
0 \\
\vdots \\
0
\end{pmatrix},
\]
where \( K_i := \{ k : v_{k-1} < T_i \leq v_k \} \) as mentioned in the context before. And

\[
\frac{\partial^2}{\partial \zeta^T \partial \zeta} \log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right) = -\frac{\delta_i}{(T_{iv} \zeta)^2} I_{iv} I_{iv}^T.
\]

Therefore,

\[
\frac{\partial^2}{\partial \zeta^2 \partial \zeta} \log L = \sum_{i=1}^m \left( -\frac{\delta_i}{(T_{iv} \zeta)^2} I_{iv} I_{iv}^T + \text{Var} \left( \frac{\partial}{\partial \zeta} \ell_{iq}(\theta)|T_i, \delta_i, Y_i, \theta) \right) \right)
\]

\[
= \sum_{i=1}^m \left( -\frac{\delta_i}{(T_{iv} \zeta)^2} I_{iv} I_{iv}^T + \text{Var} \left( \omega_{i\lambda}|T_i, \delta_i, Y_i, \theta) \right) \right).
\]

\[
\textbullet \ (\sigma^2; \zeta)
\]

\[
\frac{\partial}{\partial \sigma^2} \log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right) =: \frac{\partial}{\partial \sigma^2} \ell_{iq}(\theta)
\]

\[
= \frac{1}{\sigma^4} (Y_i - A_i \alpha_i - 1_n X_i^T \beta)^T(Y_i - A_i \alpha_i - 1_n X_i^T \beta) - \frac{n_i}{2\sigma^2}
\]

\[
= \frac{1}{\sigma^4} \left( (Y_i - 1_n X_i^T \beta)^T(Y_i - 1_n X_i^T \beta) - 2(Y_i - 1_n X_i^T \beta)^T A_i \alpha_i + A_i^T A_i \alpha_i - n_i\sigma^2 \right)
\]

\[
\frac{\partial^2}{\partial (\sigma^2)^2} \log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right) =: \frac{\partial^2}{\partial (\sigma^2)^2} \ell_{iq}(\theta)
\]

\[
= -\frac{1}{\sigma^6} (Y_i - A_i \alpha_i - 1_n X_i^T \beta)^T(Y_i - A_i \alpha_i - 1_n X_i^T \beta) + \frac{n_i}{2\sigma^4}
\]

\[
= -\frac{1}{\sigma^6} \left( (Y_i - 1_n X_i^T \beta)^T(Y_i - 1_n X_i^T \beta) - 2(Y_i - 1_n X_i^T \beta)^T A_i \alpha_i + A_i^T A_i \alpha_i \right) + \frac{n_i}{2\sigma^4}.
\]

Therefore,

\[
\frac{\partial^2}{\partial (\sigma^2)^2} \log L = \sum_{i=1}^m E \left( \frac{\partial^2}{\partial (\sigma^2)^2} \ell_{iq}(\theta)|T_i, \delta_i, Y_i, \theta) \right) + \text{Var} \left( \frac{\partial}{\partial \sigma^2} \ell_{iq}(\theta)|T_i, \delta_i, Y_i, \theta) \right).
\]

Since

\[
\frac{\partial^2}{\partial \zeta^T \partial \sigma^2} \log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right) = 0,
\]

we have,

\[
\frac{\partial^2}{\partial \zeta^T \partial \sigma^2} \log L = \sum_{i=1}^m \text{Cov} \left( \frac{\partial}{\partial \sigma^2} \ell_{iq}(\theta), \frac{\partial}{\partial \zeta^T} \ell_{iq}(\theta)|T_i, \delta_i, Y_i, \theta) \right)
\]

\[
= -\frac{1}{2\sigma^4} \sum_{i=1}^m \text{Cov} \left( (Y_i - A_i \alpha_i - 1_n X_i^T \beta)^T(Y_i - A_i \alpha_i - 1_n X_i^T \beta), \omega_{i\lambda}|T_i, \delta_i, Y_i, \theta) \right).
\]

\[
\textbullet \ (b; \sigma^2, \zeta)
\]

\[
\frac{\partial^2}{\partial b^2} \log L = \sum_{i=1}^m \left( E \left( \frac{\partial^2}{\partial b^2} \ell_{iq}(\theta)|T_i, \delta_i, Y_i, \theta) \right) + \text{Var} \left( \frac{\partial}{\partial b} \ell_{iq}(\theta)|T_i, \delta_i, Y_i, \theta) \right). \]
Since
\[ \frac{\partial^2}{\partial \sigma^2 \partial b} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta) Pr(Y_i|\alpha_i, \theta) Pr(\alpha_i|\theta) \right) = 0, \]
we have,
\[ \frac{\partial^2}{\partial \sigma^2 \partial b} \log L = \sum_{i=1}^{m} Cov \left( \frac{\partial}{\partial b} \ell_{iq}(\theta), \frac{\partial^2}{\partial \sigma^2 \partial b} \ell_{iq}(\theta) \right) |T_i, \delta_i, Y_i, \theta). \]

On the other hand,
\[
\begin{align*}
\frac{\partial^2}{\partial \zeta^2 \partial b} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta) Pr(Y_i|\alpha_i, \theta) Pr(\alpha_i|\theta) \right) &:= \frac{\partial^2}{\partial \zeta^2 \partial b} \ell_{iq}(\theta) \\
&= e^{Z_i T} \int_{v_1}^{v_2} \exp(B(a, b, u) \alpha_i + B_0(a, b, u) X_i^T \beta)(g_1(a, b, u) - u g(a, b, u)) du \\
&= e^{Z_i T} \int_{e^{K_1 - 1}}^{e^{K_1}} \exp(B(a, b, u) \alpha_i + B_0(a, b, u) X_i^T \beta)(g_1(a, b, u) - u g(a, b, u)) du \\
&= e^{Z_i T} \left( e^{Z_i T} \int_{v_1}^{v_2} \exp(B(a, b, u) \alpha_i + B_0(a, b, u) X_i^T \beta)(g_1(a, b, u) - u g(a, b, u)) du \right). \\
\end{align*}
\]

Therefore,
\[
\frac{\partial^2}{\partial \zeta^2 \partial b} \log L = \sum_{i=1}^{m} E \left( \frac{\partial^2}{\partial \zeta^2 \partial b} \ell_{iq}(\theta) |T_i, \delta_i, Y_i, \theta \right) + Cov \left( \frac{\partial}{\partial b} \ell_{iq}(\theta), \frac{\partial^2}{\partial \zeta^2 \partial b} \ell_{iq}(\theta) \right) |T_i, \delta_i, Y_i, \theta). \\
\]

• \((\theta_t; b, \sigma^2, \zeta)\), where \(\theta_t = (a, \eta)^T\)

\[
\begin{align*}
&\frac{\partial}{\partial \theta_t} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta) Pr(Y_i|\alpha_i, \theta) Pr(\alpha_i|\theta) \right) =: \frac{\partial}{\partial \theta_t} \ell_{iq}(\theta) \\
&= \delta_i \left( g(b, T_i) \right) - \left( \int_{0}^{T_i} \lambda_0(u) e^{Z_i^T \eta_i e^{B(a,b,u) T_a} + B_0(a,b,u) X_i^T \beta} g(b, u) du \right) \\
&\approx \delta_i \left( g(b, T_i) \right) - \left( \sum_{k=1}^{K} \zeta_k e^{Z_i^T \eta_i \sum_{j \in \Omega_k} dt_{k_j} \omega_{k_j} e^{B(a,b,u_k) T_a} + B_0(a,b,u_k) X_i^T \beta} g(b, u_k) \right). \\
\end{align*}
\]

Note: \(g(b, u) = B(b, u)^T a_i + B_0(b, u) X_i^T \beta\) and \(g(a, b, u) = a g(b, u)\).

\[
\begin{align*}
&\frac{\partial^2}{\partial \theta_t^2} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta) Pr(Y_i|\alpha_i, \theta) Pr(\alpha_i|\theta) \right) =: \frac{\partial^2}{\partial \theta_t^2} \ell_{iq}(\theta) \\
&= - \left( \int_{0}^{T_i} \lambda_0(u) e^{Z_i^T \eta_i e^{g(a,b,u) T_a} + B_0(a,b,u) X_i^T \beta} g(b, u) du \right) \left( \int_{0}^{T_i} \lambda_0(u) e^{Z_i^T \eta_i e^{g(a,b,u) T_a} + B_0(a,b,u) X_i^T \beta} g(b, u) du \right) \\
&\approx - \left( 2W_{i\lambda} \ 1W_{i\lambda} Z_i^T \\
&\ 1W_{i\lambda} Z_i \ 0W_{i\lambda} Z_i Z_i^T \right) \\
\end{align*}
\]
where we denote,

\[ 0W_{i\lambda} := \sum_{k=1}^{K_i} \zeta_{ik} e^{Z_i^T \eta} \sum_{j \in \Omega_{ik}} dt_{kj} \omega_{kj} e^{g(a,b,u_{kj})}, \]

\[ 1W_{i\lambda} := \sum_{k=1}^{K_i} \zeta_{ik} e^{Z_i^T \eta} \sum_{j \in \Omega_{ik}} dt_{kj} \omega_{kj} e^{g(a,b,u_{kj})} g(b, u_{kj}), \]

\[ 2W_{i\lambda} := \sum_{k=1}^{K_i} \zeta_{ik} e^{Z_i^T \eta} \sum_{j \in \Omega_{ik}} dt_{kj} \omega_{kj} e^{g(a,b,u_{kj})} g(b, u_{kj})^2. \]

Therefore,

\[ \frac{\partial^2 \log L}{\partial \theta^2} = \sum_{i=1}^{m} E \left( \frac{\partial^2}{\partial \theta^2} \ell_{iq}(\theta) \mid T_i, \delta_i, Y_i, \theta \right) + \text{Var} \left( \frac{\partial}{\partial \theta} \ell_{iq}(\theta) \mid T_i, \delta_i, Y_i, \theta \right). \]

Denote \( g_s(b, u) = sB(b, u)\alpha_i + B_s(b, u)X_i^T \beta, \) \( s = 1, 2, \) thus, \( g_s(a, b, u) = a g_s(b, u). \)

\[ \frac{\partial^2}{\partial b \partial \theta} \log \left( \text{Pr}(T_i, \delta_i|\alpha_i, \theta) \text{Pr}(Y_i|\alpha_i, \theta) \text{Pr}(\alpha_i|\theta) \right) =: \frac{\partial^2}{\partial b \partial \theta} \ell_{iq}(\theta) \]

\[ = \delta_i \left( g_1(b, T_i) - T_i g(b, T_i) \right) - \left( \int_{0}^{T_i} \lambda_0(u) e^{Z_i^T \eta \phi(a,b,u)} g'(b, u)(1 + g(a, b, u))du \right) \]

Note: \( g'(b, u) = g_1(b, u) - ug(b, u) \) and \( g'(a, b, u) = g_1(a, b, u) - ug(a, b, u). \)

Therefore,

\[ \frac{\partial^2 \log L}{\partial b \partial \theta} = \sum_{i=1}^{m} E \left( \frac{\partial^2}{\partial b \partial \theta} \ell_{iq}(\theta) \mid T_i, \delta_i, Y_i, \theta \right) + \text{Cov} \left( \frac{\partial}{\partial b} \ell_{iq}(\theta), \frac{\partial}{\partial \theta} \ell_{iq}(\theta) \mid T_i, \delta_i, Y_i, \theta \right). \]

Since

\[ \frac{\partial^2}{\partial \sigma^2 \partial \theta} \log \left( \text{Pr}(T_i, \delta_i|\alpha_i, \theta) \text{Pr}(Y_i|\alpha_i, \theta) \text{Pr}(\alpha_i|\theta) \right) = 0, \]

thus we have

\[ \frac{\partial^2 \log L}{\partial \sigma^2 \partial \theta} = \sum_{i=1}^{m} \text{Cov} \left( \frac{\partial}{\partial \theta} \ell_{iq}(\theta), \frac{\partial}{\partial \sigma^2} \ell_{iq}(\theta) \mid T_i, \delta_i, Y_i, \theta \right). \]

On the other hand,

\[ \frac{\partial^2}{\partial \zeta^2 \partial \theta} \log \left( \text{Pr}(T_i, \delta_i|\alpha_i, \theta) \text{Pr}(Y_i|\alpha_i, \theta) \text{Pr}(\alpha_i|\theta) \right) =: \frac{\partial^2}{\partial \zeta^2 \partial \theta} \ell_{iq}(\theta) \]

\[ = - \left( \int_{v_{K_i-1}}^{v_1} \lambda_0(u) e^{g(a,b,u)+Z_i^T \eta \phi(b,u)} du, \ldots, \int_{v_{K_i-1}}^{v_{K_i-1}} \lambda_0(u) e^{g(a,b,u)+Z_i^T \eta \phi(b,u)} du \right) \cdot 0 \cdot 0 \]

Therefore,

\[ \frac{\partial^2 \log L}{\partial \zeta^2 \partial \theta} = \sum_{i=1}^{m} E \left( \frac{\partial^2}{\partial \zeta^2 \partial \theta} \ell_{iq}(\theta) \mid T_i, \delta_i, Y_i, \theta \right) + \text{Cov} \left( \frac{\partial}{\partial \theta} \ell_{iq}(\theta), \frac{\partial}{\partial \zeta} \ell_{iq}(\theta) \mid T_i, \delta_i, Y_i, \theta \right). \]
\[(\beta; \theta_i, b, \sigma^2, \zeta)\]

\[
\frac{\partial}{\partial \beta} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta) Pr(Y_i|\alpha_i, \theta) Pr(\alpha_i|\theta) \right) =: \frac{\partial}{\partial \beta} \ell_{iq}(\theta)
\]

\[
= X_i \left( \delta_i B_0(a, b, T_i) - \int_0^{T_i} \lambda_0(u) e^{g(a,b,u)+\delta^T u} B_0(a, b, u) du + 1_n^T \Sigma_i^{-1} (Y_i - A_i \alpha_i - 1_n^T \beta) \right).
\]

And that,

\[
\frac{\partial^2}{\partial \beta^T \partial \beta} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta) Pr(Y_i|\alpha_i, \theta) Pr(\alpha_i|\theta) \right) =: \frac{\partial^2}{\partial \beta^T \partial \beta} \ell_{iq}(\theta)
\]

\[
= - \left( \int_0^{T_i} \lambda_0(u) e^{B(a,b,u)^T \alpha_i + B_0(a, b, u) X_i^T \beta + \delta^T u} B_0(a, b, u)^2 du + 1_n^T \Sigma_i^{-1} 1_n \right) X_i^T X_i
\]

\[
\approx - \left( \sum_{k=1}^{K_i} \zeta_k e^{Z_i^T \eta} \sum_{j \in \Omega_{k,j}} dt_{kj} \omega_{kj} e^{B(a,b,u_k)^T \alpha_i + B_0(a, b, u_k) X_i^T \beta} B_0(a, b, u_k)^2 + 1_n^T \Sigma_i^{-1} 1_n \right) X_i^T X_i.
\]

Therefore,

\[
\frac{\partial^2}{\partial \beta^T \partial \beta} \log L = \sum_{i=1}^{m} E \left( \frac{\partial^2}{\partial \beta^T \partial \beta} \ell_{iq}(\theta) | T_i, \delta_i, Y_i, \theta \right) + \text{Var} \left( \frac{\partial}{\partial \beta} \ell_{iq}(\theta) | T_i, \delta_i, Y_i, \theta \right).
\]

On the other hand,

\[
\frac{\partial^2}{\partial \theta_i^T \partial \beta} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta) Pr(Y_i|\alpha_i, \theta) Pr(\alpha_i|\theta) \right) = \left( \Psi_{1,\beta \theta_i}, \Psi_{2,\beta \theta_i} \right) =: \frac{\partial^2}{\partial \theta_i^T \partial \beta} \ell_{iq}(\theta),
\]

where we define

\[
\Psi_{1,\beta \theta_i} := \left( \delta_i B_0(b, T_i) - \int_0^{T_i} \lambda_0(u) e^{g(a,b,u)+\delta^T u} \left( B_0(b, u) + B_0(a, b, u) g(b, u) \right) du \right) X_i,
\]

\[
\Psi_{2,\beta \theta_i} := - \left( \int_0^{T_i} \lambda_0(u) e^{g(a,b,u)+\delta^T u} B_0(a, b, u) du \right) X_i Z_i^T.
\]

Therefore,

\[
\frac{\partial^2}{\partial \theta_i^T \partial \beta} \log L = \sum_{i=1}^{m} E \left( \frac{\partial^2}{\partial \theta_i^T \partial \beta} \ell_{iq}(\theta) | T_i, \delta_i, Y_i, \theta \right) + \text{Cov} \left( \frac{\partial}{\partial \theta_i} \ell_{iq}(\theta), \frac{\partial}{\partial \beta} \ell_{iq}(\theta) | T_i, \delta_i, Y_i, \theta \right).
\]

And

\[
\frac{\partial^2}{\partial \theta \partial \beta} \log \left( Pr(T_i, \delta_i|\alpha_i, \theta) Pr(Y_i|\alpha_i, \theta) Pr(\alpha_i|\theta) \right) =: \frac{\partial^2}{\partial \theta \partial \beta} \ell_{iq}(\theta)
\]

\[
= \left( \delta_i(B_1(a, b, T_i) - T_i B_0(a, b, T_i)) - \int_0^{T_i} \lambda_0(u) e^{g(a,b,u)+\delta^T u} \left( B_1(a, b, u) - u B_0(a, b, u) \right) du \\
- \int_0^{T_i} \lambda_0(u) e^{g(a,b,u)+\delta^T u} B_0(a, b, u) \left( g_1(a, b, u) - u g(a, b, u) \right) du \right) X_i,
\]

we have,

\[
\frac{\partial^2}{\partial \theta \partial \beta} \log L = \sum_{i=1}^{m} E \left( \frac{\partial^2}{\partial \theta \partial \beta} \ell_{iq}(\theta) | T_i, \delta_i, Y_i, \theta \right) + \text{Cov} \left( \frac{\partial}{\partial \beta} \ell_{iq}(\theta), \frac{\partial}{\partial \theta} \ell_{iq}(\theta) | T_i, \delta_i, Y_i, \theta \right).
\]
Meanwhile,

\[
\frac{\partial^2}{\partial \sigma^2 \partial \beta} \log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right) = \frac{\partial^2}{\partial \sigma^2 \partial \beta} \ell_{iq}(\theta)
\]

\[
= - \frac{1}{\sigma^4} X_i^T \eta_i (Y_i - A_i \alpha_i - 1_n, X_i^T \beta),
\]

thus

\[
\frac{\partial^2 \log L}{\partial \sigma^2 \partial \beta} = \sum_{i=1}^m E \left( \frac{\partial^2}{\partial \sigma^2 \partial \beta} \ell_{iq}(\theta) | T_i, \delta_i, Y_i, \theta \right) + \text{Cov} \left( \frac{\partial}{\partial \beta} \ell_{iq}(\theta), \frac{\partial}{\partial \beta} \ell_{iq}(\theta) | T_i, \delta_i, Y_i, \theta \right).
\]

At last,

\[
\frac{\partial^2}{\partial \zeta^T \partial \beta} \log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right) = \frac{\partial^2}{\partial \zeta^T \partial \beta} \ell_{iq}(\theta)
\]

\[
= - e^{Z_T \eta} X_i \left( \Psi_{1, \beta \zeta}, \Psi_{2, \beta \zeta}, \ldots, \Psi_{K_i, \beta \zeta}, 0, \ldots, 0 \right),
\]

where \( \Psi_{K_i, \beta \zeta} = \int_{v_{k-1}}^{v_k} \lambda_0(u) e^{g(a, b, u)} B_0(a, b, u) du, \ k = 1, K_i \). Hence,

\[
\frac{\partial^2 \log L}{\partial \zeta^T \partial \beta} = \sum_{i=1}^m E \left( \frac{\partial^2}{\partial \zeta^T \partial \beta} \ell_{iq}(\theta) | T_i, \delta_i, Y_i, \theta \right) + \text{Cov} \left( \frac{\partial}{\partial \beta} \ell_{iq}(\theta), \frac{\partial}{\partial \zeta} \ell_{iq}(\theta) | T_i, \delta_i, Y_i, \theta \right).
\]

- \((\alpha; \beta, \theta, b, \sigma^2, \zeta)\)

\[
\frac{\partial^2 \log L}{\partial \alpha^T \partial \alpha} = \sum_{i=1}^m \int (-G^{-1}) Pr(\alpha_i | T_i, \delta_i, Y_i, \theta) d\alpha_i + \int \text{Cov}(G^{-1} \alpha_i) Pr(\alpha_i | T_i, \delta_i, Y_i, \theta) d\alpha_i
\]

\[
= - n G^{-1} + G^{-1} \left( \sum_{i=1}^m \text{Var}(\alpha_i | T_i, \delta_i, Y_i, \theta) \right) G^{-1}.
\]

Since,

\[
\frac{\partial^2}{\partial \beta \partial \alpha} \log \left( Pr(T_i, \delta_i | \alpha_i, \theta) Pr(Y_i | \alpha_i, \theta) Pr(\alpha_i | \theta) \right) = 0,
\]

we have,

\[
\frac{\partial^2 \log L}{\partial \beta^T \partial \alpha} = \sum_{i=1}^m \text{Cov} \left( \frac{\partial}{\partial \alpha} \ell_{iq}(\theta), \frac{\partial}{\partial \beta} \ell_{iq}(\theta) | T_i, \delta_i, Y_i, \theta \right).
\]

Similarly,

\[
\frac{\partial^2 \log L}{\partial \alpha d \partial \alpha} = \sum_{i=1}^m \text{Cov} \left( \frac{\partial}{\partial \alpha} \ell_{iq}(\theta), \frac{\partial}{\partial \alpha} \ell_{iq}(\theta) | T_i, \delta_i, Y_i, \theta \right),
\]

\[
\frac{\partial^2 \log L}{\partial \eta^T \partial \alpha} = \sum_{i=1}^m \text{Cov} \left( \frac{\partial}{\partial \alpha} \ell_{iq}(\theta), \frac{\partial}{\partial \eta} \ell_{iq}(\theta) | T_i, \delta_i, Y_i, \theta \right),
\]

\[
\frac{\partial^2 \log L}{\partial \beta \partial \alpha} = \sum_{i=1}^m \text{Cov} \left( \frac{\partial}{\partial \alpha} \ell_{iq}(\theta), \frac{\partial}{\partial \beta} \ell_{iq}(\theta) | T_i, \delta_i, Y_i, \theta \right),
\]
\[
\frac{\partial^2 \log L}{\partial \sigma^2 \partial \alpha} = \sum_{i=1}^{m} \text{Cov} \left( \frac{\partial}{\partial \alpha} \ell_{iq}(\theta), \frac{\partial}{\partial \sigma^2} \ell_{iq}(\theta) \right) | T_i, \delta_i, Y_i, \theta ,
\]

\[
\frac{\partial^2 \log L}{\partial \zeta^T \partial \alpha} = \sum_{i=1}^{m} \text{Cov} \left( \frac{\partial}{\partial \alpha} \ell_{iq}(\theta), \frac{\partial}{\partial \zeta} \ell_{iq}(\theta) \right) | T_i, \delta_i, Y_i, \theta .
\]

**Part C: Fully exponential Laplace approximation**

Denote \( M(s) = E(\exp(sh(\alpha_i)) | T_i, \delta_i, Y_i) \), then \( \partial M(s)/\partial s = E(h(\alpha_i) \exp(sh(\alpha_i)) | T_i, \delta_i, Y_i) \) and \( \partial M(s)/\partial s|_{s=0} = E(h(\alpha_i) | T_i, \delta_i, Y_i) \). Therefore we approximate \( M(s) \) to obtain the approximation of posterior expectation of \( E(h(\alpha_i) | T_i, \delta_i, Y_i) \), and since

\[
E(\exp(sh(\alpha_i)) | T_i, \delta_i, Y_i) = \frac{\int e^{sh(\alpha_i)} Pr(T_i, \delta_i, Y_i, \alpha_i) d\alpha_i}{\int Pr(T_i, \delta_i, Y_i, \alpha_i) d\alpha_i} = \frac{\int \exp \left( sh(\alpha_i) + \log Pr(T_i, \delta_i, Y_i, \alpha_i) \right) d\alpha_i}{\int \exp \left( \log Pr(T_i, \delta_i, Y_i, \alpha_i) \right) d\alpha_i}.
\]

Denote

\[
L(\alpha_i) = \log Pr(T_i, \delta_i, Y_i, \alpha_i),
\]

\[
L^*(\alpha_i) = sh(\alpha_i) + \log Pr(T_i, \delta_i, Y_i, \alpha_i) = sh(\alpha_i) + L(\alpha_i),
\]

and \( \hat{\alpha}_i = \arg \max_{\alpha_i} \log Pr(T_i, \delta_i, Y_i, \alpha_i) \), thus we have \( L'(\hat{\alpha}_i) = 0 \) and

\[
L(\alpha_i) \approx L(\hat{\alpha}_i) + \frac{1}{2}(\alpha_i - \hat{\alpha}_i)^T L''(\hat{\alpha}_i)(\alpha_i - \hat{\alpha}_i),
\]

\[
L^*(\alpha_i) \approx sh(\hat{\alpha}_i) + L(\hat{\alpha}_i) + (sh'(\hat{\alpha}_i))^T (\alpha_i - \hat{\alpha}_i) + \frac{1}{2}(\alpha_i - \hat{\alpha}_i)^T \left( sh''(\hat{\alpha}_i) + L''(\hat{\alpha}_i) \right) (\alpha_i - \hat{\alpha}_i),
\]

where \( L'(\cdot) \) and \( L''(\cdot) \) denote the first and second derivatives with respect to \( \alpha_i \) and similarly \( h'(\cdot) \) and \( h''(\cdot) \). Therefore, we have

\[
\int e^{L(\alpha_i)} d\alpha_i \approx \int e^{L(\hat{\alpha}_i) + \frac{1}{2}(\alpha_i - \hat{\alpha}_i)^T L''(\hat{\alpha}_i)(\alpha_i - \hat{\alpha}_i)} d\alpha_i = e^{L(\hat{\alpha}_i)} \int e^{-\frac{1}{2}(\alpha_i - \hat{\alpha}_i)^T \left( -L''(\hat{\alpha}_i) \right) (\alpha_i - \hat{\alpha}_i)} d\alpha_i = e^{L(\hat{\alpha}_i)} (2\pi)^{q/2} | - L''(\hat{\alpha}_i) |^{-1/2},
\]

\[
\int e^{L^*(\alpha_i)} d\alpha_i \approx \int e^{sh(\hat{\alpha}_i) + L(\hat{\alpha}_i) + (sh'(\hat{\alpha}_i))^T (\alpha_i - \hat{\alpha}_i) + \frac{1}{2}(\alpha_i - \hat{\alpha}_i)^T \left( sh''(\hat{\alpha}_i) + L''(\hat{\alpha}_i) \right) (\alpha_i - \hat{\alpha}_i)} d\alpha_i = e^{sh(\hat{\alpha}_i)} e^{L(\hat{\alpha}_i)} \int e^{-\frac{1}{2}H_1^T \left( -sh''(\hat{\alpha}_i) - L''(\hat{\alpha}_i) \right) H_1 + \frac{1}{2}(sh'(\hat{\alpha}_i))^T \left( -sh''(\hat{\alpha}_i) - L''(\hat{\alpha}_i) \right)^{-1} (sh'(\hat{\alpha}_i))} d\alpha_i = e^{sh(\hat{\alpha}_i)} e^{L(\hat{\alpha}_i)} e^{\frac{1}{2}(h'(\hat{\alpha}_i))^T \left( -sh''(\hat{\alpha}_i) - L''(\hat{\alpha}_i) \right)^{-1} (h'(\hat{\alpha}_i))} (2\pi)^{q/2} | - sh''(\hat{\alpha}_i) - L''(\hat{\alpha}_i) |^{-1/2},
\]
where $H_1 = (\alpha_i - \hat{\alpha}_i) - \left( -sh''(\alpha_i) - L''(\alpha_i) \right)^{-1} sh'(\hat{\alpha}_i)$. Therefore, we have,

$$E(\exp(h(\alpha_i)s)|T_i, \delta_i, Y_i) \approx \frac{e^{sh(\hat{\alpha}_i) + \frac{\partial^2}{2}(h'(\hat{\alpha}_i))^T\left(-sh''(\hat{\alpha}_i) - L''(\hat{\alpha}_i)\right)^{-1}(h'(\hat{\alpha}_i))} - sh''(\hat{\alpha}_i) - L''(\hat{\alpha}_i)|^{-1/2}}{| - L''(\hat{\alpha}_i)|^{-1/2}}.$$

Thus

$$\frac{\partial E(\exp(h(\alpha_i)s)|T_i, \delta_i, Y_i)}{\partial s} \approx \frac{e^{H_2(s)H_2'(s)} - sh''(\hat{\alpha}_i) - L''(\hat{\alpha}_i)|^{-1/2}}{| - L''(\hat{\alpha}_i)|^{-1/2}} \frac{\partial H_2(s)}{\partial s}$$

$$\approx 0$$

due to that

$$\frac{\partial \det(Y)}{\partial s} = \det(Y)\text{Trace}(Y^{-1}\frac{\partial Y}{\partial s}).$$

Therefore, we have

$$E(h(\alpha_i)|T_i, \delta_i, Y_i) \approx h(\hat{\alpha}_i) - \frac{1}{2}\text{Trace}\left(\left( -L''(\hat{\alpha}_i)\right)^{-1}( -h''(\hat{\alpha}_i))\right).$$

And the Laplace approximation is a second-order approximation.
Chapter 5

Conclusions and future work

5.1 Summary of the thesis

In a longitudinal study, various types of data are often recorded. Variables collected at the time of entry include but are not limited to basic factors like sex and age. Also, a large number of variables usually have been recorded both at the time of entry and at the follow-up visits, for example, the measurements of Systolic blood pressure and diastolic blood pressure may have been taken intermittently in a years-follow-up study of cardiovascular disease, and the CD4+ cell counts in an HIV/AIDS study, together with processes of events in time, which can be recurrent events, such as revisits of the hospital, or terminated events, such as death. On one hand, it is of great interest to investigate the relationships among all these different types of variables and to try to understand the underlying mechanism. On the other hand, the repeated measurements and different-types variables taken from same subjects are correlated somewhat in their very nature, and special techniques are required to account for such correlations in the analysis. By ignoring such nature, it may lead to severe bias and erroneous inference.

Repeated measurements of multivariate longitudinal processes are often recorded routinely in a longitudinal study. Apart from the within-subject correlation of repeated measurements of a single process of a subject, it is very likely that there may also exist some association among these processes and it is often of importance to include such association into analysis to improve the efficiency and therefore provide reliable inference. In chapter 2, the study is motived by a large cohort study, UK north Staffordshire osteoarthritis project, in which the statistical modelling for triple types of outcomes
of repeated measurements is of interest. The mixed random-effects models are proposed to account for the inherent association, and the association is characterized by a covariance matrix of random effects, $\text{cov}(\alpha_i) = G$. When the covariance matrix $G$ turns out to be block diagonal, it shows that the association between some processes may be weak and we can consider them to be independent, they can be analyzed separately. However, when some of the off-diagonal elements of correlation matrix have large values and indicate a strong correlation among them, the investigation of the study in chapter 2 shows that, analyzing the data by joint modelling may reduce the bias and, more importantly, lead to more reliable inference and much improve the efficiency comparing that of by modelling the processes separately, especially for the cases when the variation and correlation of random effects are relative large or strong, and similar finds also have been stated by Ye and Pan (2006) and Daniels and Zhao (2003). The study is conducted under the Bayesian framework which can pleasantly avoid the complex integration issue with respect to random effects, typically raising in the likelihood-based approach of classical statistics framework.

Apart from the repeated measurements that are taken at entry and follow-up visits, one or two processes of events in time are often being monitored as well throughout the study. Together with some other baseline covariates information, they constitute the observed data for a longitudinal study. It is quite often that the events are terminated-type events and there is no follow-up visit beyond the time of events, that is, the potential measurements may have somehow been truncated by the events. It may lead to biased or even erroneous inference by analysing the longitudinal data without noticing the issue of potential truncation (Henderson et al., 2000, among others). On the other hand, the levels of longitudinal process is often of interest in the way of being introduced in the Cox regression models to investigate the relationship of time to event and the levels of longitudinal process. The lack of complete knowledge of trajectory information of the longitudinal process inevitably calls for the joint modelling of both types of processes to assist the implementation of statistical inference for survival data, and, in the mean time, it may also significantly improve the statistical inference for longitudinal process and correct the bias caused by the truncation issue. In the past 20 years, some well-established joint modelling frameworks have been developed and some have even been implemented in the software, such as the R packages of joineR
and JM, for the public use. However, due to the complexity of the joint models, some assumptions can not be released or make them to more general circumstances, as we know that, the assumptions are not necessary to have to be true all the time, and the violation of some assumptions may lead to problematic inference. In chapter 3, we investigate the impact on statistical inference of survival data when the assumption of mutual independence of random error has been violated. Henderson et al. (2000) introduced a stationary Gaussian process, which were independent of the random effects, to account for the possible left-over serial correlation among the random errors, and Tsiatis and Davidian (2004) also mentioned additional efforts may require to account for such correlation. The conditional score estimation approach, proposed by Tsiatis and Davidian (2001), is utilised in chapter 3 to investigate the influence caused by the violation of the assumption of mutual independence on random errors. The conditional score estimation, on one hand, by introducing the sufficient statistics for random effects, it can rule out any distributed assumption on the random effects and it provides huge computational advantage, while the traditional likelihood-based approaches are usually quite computationally intensive. However, the study in chapter 3 shows that it may lead to severe bias of inference for the survival data when random errors are somehow correlated. Generalised conditional score estimators are proposed in chapter 3 to account for the random error correlation, and the data-driven method of modified Cholesky decomposition is utilised to capture the correlation left-over in random errors after the random effects. It shows that, by doing so, it can largely reduce the bias and provide nearly unbiased estimators when the covariance matrix of random errors have been characterized well.

In chapter 4, instead of default current value of longitudinal profile, which has been being typically introduced into Cox regression model within the framework of joint modelling of longitudinal and survival data, we would like to also account for the trajectory information, such as the cumulative information of longitudinal process, in the analysis of survival data. In more recent studies, it has been recognized by some investigators that both of the past and recent levels of longitudinal profile may have instructive impact on the process of events in time. It may under estimate the importance of blood pressure and cholesterol, for example, for coronary heart disease by accounting for the current information of blood pressure only, and levels of them
in last certain periods may also have important impact on the present heart disease hazards (Clarke et al., 2002; Boshuizen et al., 2007 among others). In chapter 4, the algorithm of integration is proposed to account for the trajectory information of longitudinal process, together with a weighted curve to share the flexibility of being able to cope with the fact that levels of longitudinal profile in different time periods may have different impact on the present hazards of events, more specifically,

\[
\int_{0}^{t} \gamma(t-u)m_i(u)du \quad \text{or} \quad \int_{t_k}^{t} \gamma(t-u)m_i(u)du
\]

is proposed to account for the cumulative information of the longitudinal process and therefore the past information. In the study, the join models are investigated under the framework of classical likelihood-based approach. Expectation-maximization algorithm is utilised to obtain the maximum likelihood estimators and the fully exponential Laplace approximation approach has been proposed to approximate the expectations of functions of random effects conditional on the observed data and the current estimators of parameters. The negative Hessian matrix is proposed to estimate the Fisher information matrix and therefore the statistical inference for parameters. The piecewise function is utilised to characterize the baseline hazard function, on one hand, it can turn the non-parametric problem of statistical inference for baseline hazard function into the parametric framework, on the other hand, it can help to obtain the estimation of standard deviations for the parameters, otherwise the standard deviations are often under estimate and additional efforts, such as bootstrap algorithm, may be required to obtain the proper standard deviations for the estimators of parameters (Hsieh et al., 2006, among others). The joint models of longitudinal and survival data with the term of cumulative information have been applied to both datasets of the schizophrenia trial and liver cirrhosis trial and have good performance. Simulation study of the chapter also shows the feasibility of the model framework and the reliability of the proposed approach. In the end, we would like to say it is a really interesting topic, in which there are still a lot of open questions and aspects that are waiting for us to explore, to conquer.
5.2 Future work

5.2.1 Model selection

Model selection, including but certainly not limited to variable selection, is one of the fundamental topics in statistics. In chapter 3 and chapter 4, the model or variable selection is not the topic of our studies, hence we did not explore the aspects of model and/or variable selection, and when it goes to the demand of choosing variables for the models, we simply either based on the inference that the variables are significant or not, or the interest lied in the variables, that is, if a certain variable is of our interest, it may still be included in the final analysis models that we report in the end. However, we can sense the urgent need for proposing a systematic approach for the model and variable selection under the joint modelling framework, which, to date, still few literatures talk about. It is due to the complexity of the joint models and the fact of that special efforts are required to deal with the random effects. Although the computer technologies have make the inference for joint modelling feasible, the algorithm is still sometimes quite computationally intensive. Therefore, it would really help out if a systematic procedure can be proposed to serve efficiently as the tool of model/variable selection under the joint modelling framework, such as in the way of LASSO approach.

5.2.2 Informative visiting process

It is quite a standard framework to assume that the schedule for follow-up visits is pre-specified and the longitudinal measurements are taken as designed. It may be the case, or at least almost the circumstance, with a good experiment design and the study also has been conducted well per protocol. However, in practice, we also can have huge reachable data that have been recorded by clinics or hospitals when the patients come and visit. These data are recorded without previous design at all and they are not experimental data, but they may be valuable as well to study and improve the public health. In this circumstance, the visiting process may be no longer non-informative and the patients come to visit hospitals obviously due to some symptoms of diseases or some unpleasant responses, and the frequency of revisits of hospital is strongly influenced by the levels of some health indicators. Furthermore, even in certain types
of experiment studies, such as the study that consists of several phases, the schedule of follow-up visits may also be specified conditional on some previous levels of indicators and the visiting process may not be non-informative neither.

In literature, there have been several articles noticing such issue, for example, the analysis of longitudinal data with outcome-dependent follow-up by Lin, Scharfstein and Rosenheck (2004), Sun et al. (2005), and Liang, Lu and Ying (2009) among others. Liu, Huang and O’Quigley (2008) discussed the analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, in which the shared random-effects models are proposed, that is, linear mixed random-effects model for the longitudinal process and different intensive functions for the visiting and terminal event processes, respectively, and the models are linked by shared random effects. Obviously, there are still a lot issues that proposed models can not cope with, such as when the time-dependent covariates are introduced in the intensive functions, the implementation of statistical inference turns to more complicated. Or there may be some alternative approaches to analyse such kind of observed data.

5.2.3 Measure theory to serve the joint modelling framework

It is a hunch of the author that the measure theory may serve well the framework of joint modelling of longitudinal and survival data, including the theoretical support, model selection and some elegant ways of coping with both of the longitudinal and survival process simultaneously. And the author of the thesis would like to explore some aspects of it in the coming years of her research and career.
Bibliography


[5] T. Ando. Bayesian predictive information criterion for the evaluation of hierar-


nen, S. Nedeljkovic, A. Kafatos, and D. Kromhout. Effects of past and recent
blood pressure and cholesterol level on coronary heart disease and stroke mor-
tality, accounting for measurement error. *American journal of epidemiology*,


Chapter 6

Additional materials

1. Some distributions for the simulation scenarios in chapter 3

- Multivariate skew-normal distributions

The multivariate skew-normal distribution was discussed by Azzalini and Dalla Valle (1996). A random variable $Z$ is said to be skew-normal with parameter $\lambda$, written $Z \sim SN(\lambda)$, if its density function is

$$f(z; \lambda) = 2\phi(z)\Phi(\lambda z) \quad (z \in R)$$

where $\phi(z)$ and $\Phi(z)$ denote the $N(0,1)$ density and distribution function, respectively, the parameter $\lambda$ which regulars the skewness varies in $(-\infty, \infty)$, and $\lambda = 0$ corresponds to the $N(0,1)$ density.

If $Y$ and $W$ are independent $N(0,1)$ variates, and $Z$ is set equal to $Y$ conditional on $\lambda Y > W$, for some real $\lambda$, then $Z \sim SN(\lambda)$.

For random number generation, it is more efficient to use a variant of

$$Z = \begin{cases} Y & \text{if } \lambda Y > W; \\ -Y & \text{if } \lambda Y \leq W, \end{cases}$$

hence avoiding rejection of samples.

If $Y_0$ and $Y_1$ are independent $N(0,1)$ variables and $\delta \in (-1,1)$, then

$$Z = \delta|Y_0| + (1 - \delta^2)^{\frac{1}{2}}Y_1$$

is $SN(\lambda(\delta))$, where $\lambda(\delta) = \delta/(1 - \delta^2)^{\frac{1}{2}}$ and $\lambda(\lambda) = \lambda/(1 + \lambda^2)^{\frac{1}{2}}$. 

199
Consider a k-dimensional normal random variable $Y = (Y_1, \ldots, Y_k)^T$ with standardised marginals, independent of $Y_0 \sim N(0, 1)$, thus

$$
\begin{pmatrix}
Y_0 \\
Y
\end{pmatrix} \sim N_{k+1}
\begin{pmatrix}
0, \\
1 & \Psi
\end{pmatrix}
$$

where $\Psi$ is a $k \times k$ correlation matrix. If $(\delta_1, \ldots, \delta_k)$ are in $(-1, 1)$, define,

$$Z_j = \delta_j |Y_0| + (1 - \delta_j^2)^{1/2} Y_j \quad j = 1, \ldots, k$$

Random variable $Z = (Z_1, \ldots, Z_k)^T$ is a k-dimensional skew-normal variable, with vector $\lambda$ of shape parameters and dependent parameter $\Psi$. For brevity, we write $Z \sim SN_k(\lambda, \Psi)$. Since we have

$$E(|Y_0|) = \int_{-\infty}^{0} (-y) \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy + \int_{0}^{\infty} y \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \quad = 2 \int_{0}^{\infty} y \frac{1}{\sqrt{2\pi}} e^{-y^2} dy = 2 \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u} du = \sqrt{\frac{2}{\pi}}$$

and $E(|Y_0|^2) = E(Y_0^2) = 1$, thus $\text{var}(|Y_0|) = 1 - \frac{2}{\pi}$. And

$$E(Z) = \sqrt{\frac{2}{\pi}} \omega_1(\delta)$$

$$\text{Var}(Z) = \omega_1(\delta) \text{var}(|Y_0|) \omega_1^T(\delta) + \omega_2(\delta) \Psi \omega_2^T(\delta) = \Sigma$$

where we define

$$\omega_1(\delta) = \begin{pmatrix}
\delta_1 \\
\vdots \\
\delta_k
\end{pmatrix} \quad \text{and} \quad \omega_2(\delta) = \begin{pmatrix}
(1 - \delta_1^2)^{1/2} & 0 & \cdots & 0 \\
0 & (1 - \delta_2^2)^{1/2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & (1 - \delta_k^2)^{1/2}
\end{pmatrix}$$

we have $\text{diag}(\Sigma) = (1 - \frac{2}{\pi} \delta_1^2, \ldots, 1 - \frac{2}{\pi} \delta_k^2)^T$.

Denote $Z_0 = CZ$, and given $\text{Var}(Z_0) = \Sigma_0$, and $\text{diag}(\Sigma_0) = (\sigma_1^2, \ldots, \sigma_k^2)$, and let

$$C = \text{diag} \left( \frac{\text{diag}(\Sigma_0)^{1/2}}{\text{diag}(\Sigma)^{1/2}} \right) = \begin{pmatrix}
c_1 & 0 & \cdots & 0 \\
0 & c_2 & \cdots & 0 \\
0 & 0 & \cdots & c_k
\end{pmatrix}$$

$$\Sigma = C^{-1} \Sigma_0 C^{-1}, \text{ and } \Psi = \omega_2^{-1}(\delta) \left( \Sigma - (1 - \frac{2}{\pi}) \omega_1(\delta) \omega_1^T(\delta) \right) \omega_2^{-T}(\delta)$$
\[ E(Z_0) = CE(Z) = \sqrt{\frac{2}{\pi}} C \omega_1(\delta). \]

Therefore, given \( \omega_1(\delta), \mu \) and \( \Sigma_0 \), the skew normal distribution with mean \( \mu \) and covariance \( \Sigma_0 \) can be generated by

\[ Z_0 - \sqrt{\frac{2}{\pi}} C \omega_1(\delta) + \mu \]

- Bimodal mixture of normals described in Davidian & Gallant (1993)

Bi-modal random effects were generated by mixing two normal distributions \( N(\mu, RR^T) \) and \( N(-\mu, RR^T) \) with mixing proportion \( p \) and

\[
\mu = \{(sep/2)\sqrt{r_{11}^2 + r_{12}^2}, 0\}^T
\]

where \( R \) is the upper triangular matrix.

Denote \( R = \begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix} \), then \( RR^T = \begin{pmatrix} r_{11}^2 + r_{12}^2 & r_{12}r_{22} \\ r_{12}r_{22} & r_{22}^2 \end{pmatrix} \).

Let \( Y \sim N(\mu, RR^T) \) and \( Z = \begin{cases} Y & \text{if } u \leq p \text{ and } u \sim U(0,1) \\ -Y & \text{if } u > p \text{ and } u \sim U(0,1) \end{cases} \)

Then \( E(Z) = p\mu + (1-p)(-\mu) = 2p\mu - \mu = (2p-1)\mu \) and \( E(Z^2) = E(Y^2) \)

therefore

\[ Var(Z) = RR^T - \left( (2p-1)\mu \right)^2 \]

Special case: \( E(Z) = 0 \) if \( p = 0.5 \)

The mixing normal distribution with mean \( \mu_0 \) and variance \( \Sigma_0 \), with \( p = 0.5 \) and \( sep = 4 \) can be generated by

\[ Z_0 = Z - (2p-1)\mu + \mu_0 \]

- Multivariate t distribution

Univariate t distribution with degree of \( \nu \)

\[ T = \frac{Z}{\sqrt{V/\nu}} \]

where \( Z \sim N(0,1) \) and \( V \) is from chisq distribution with degree freedom \( \nu \), and \( Z \) and \( V \) are independent.

\[ E(T) = 0 \text{ and } Var(T) = \frac{\nu}{\nu - 2} \]
Multivariate t distribution with parameters of $\Sigma$, $\mu$ and $\nu$

Let $Y \sim \mathcal{N}(0, \Sigma)$ and $V \sim \chi^2_\nu$, then random variable $T$ which is from multivariate t distribution with mean $\mu$ and variance $\Sigma_0 = \frac{\nu}{\nu-2} \Sigma$, with degree freedom of $\nu$, can be generated by

$$T = \frac{Y}{\sqrt{V/\nu}} + \mu$$

2. Additional tables and figures for the bias analysis in chapter 4

Figure 6.1: Density functions of estimators of survival parameters of default current value model $B^0(a, -, t)$ with correct model fitting.
Table 6.1: Bias analysis. $M_0$ denote the underlying models and they are fitted by the default current model $B^0(a, -, t)$. Inference are obtained by using the JM package with argument of method="piecewise-PH-GH".

<table>
<thead>
<tr>
<th>Par</th>
<th>Truval.</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>RB(%)</th>
<th>CP(%)</th>
<th>Truval.</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>RB(%)</th>
<th>CP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$B^0(a = 1.5, -, t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$B^0(a = 1, b = 0.5, t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>4.5</td>
<td>4.4709</td>
<td>0.1887</td>
<td>0.0768</td>
<td>0.7</td>
<td>55.5</td>
<td>3.5</td>
<td>3.4757</td>
<td>0.2236</td>
<td>0.0796</td>
<td>0.7</td>
<td>51.5</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.2</td>
<td>-0.1481</td>
<td>0.1535</td>
<td>0.0396</td>
<td>25.9</td>
<td>39.4</td>
<td>-0.3</td>
<td>-0.2390</td>
<td>0.2059</td>
<td>0.0453</td>
<td>20.3</td>
<td>34.4</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.1</td>
<td>0.0741</td>
<td>0.0401</td>
<td>0.0051</td>
<td>25.9</td>
<td>17.1</td>
<td>0.1</td>
<td>0.0670</td>
<td>0.0426</td>
<td>0.0058</td>
<td>33.0</td>
<td>14.9</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.3</td>
<td>0.5129</td>
<td>0.0317</td>
<td>0.0182</td>
<td>71.0</td>
<td>0.0</td>
<td>0.3</td>
<td>0.5388</td>
<td>0.0321</td>
<td>0.0188</td>
<td>79.6</td>
<td>0.0</td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>-1.5</td>
<td>-0.8930</td>
<td>0.1266</td>
<td>0.0919</td>
<td>40.5</td>
<td>2.0</td>
<td>-1.5</td>
<td>-0.8466</td>
<td>0.1289</td>
<td>0.0932</td>
<td>43.6</td>
<td>0.8</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>-1.0</td>
<td>-0.3959</td>
<td>0.2361</td>
<td>0.2529</td>
<td>60.4</td>
<td>30.8</td>
<td>-1.0</td>
<td>-0.4036</td>
<td>0.2521</td>
<td>0.2607</td>
<td>59.6</td>
<td>35.0</td>
</tr>
<tr>
<td>$a$</td>
<td>1.5</td>
<td>1.0548</td>
<td>0.1274</td>
<td>0.0947</td>
<td>29.7</td>
<td>7.1</td>
<td>-</td>
<td>1.3247</td>
<td>0.1735</td>
<td>0.1287</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$B^0(a = 1, b = 1, t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$B^0(a = 1, b = 2, t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>3.5</td>
<td>3.4904</td>
<td>0.2425</td>
<td>0.0785</td>
<td>0.3</td>
<td>50.5</td>
<td>3.5</td>
<td>3.5101</td>
<td>0.2623</td>
<td>0.0815</td>
<td>0.3</td>
<td>44.4</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.3</td>
<td>-0.2505</td>
<td>0.2104</td>
<td>0.0444</td>
<td>16.5</td>
<td>30.3</td>
<td>-0.3</td>
<td>-0.2833</td>
<td>0.2183</td>
<td>0.0452</td>
<td>5.6</td>
<td>32.3</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.1</td>
<td>0.0773</td>
<td>0.0425</td>
<td>0.0056</td>
<td>22.7</td>
<td>15.9</td>
<td>0.1</td>
<td>0.0878</td>
<td>0.0394</td>
<td>0.0057</td>
<td>12.2</td>
<td>21.9</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.3</td>
<td>0.5518</td>
<td>0.0328</td>
<td>0.0178</td>
<td>84.0</td>
<td>0.0</td>
<td>0.3</td>
<td>0.5651</td>
<td>0.0323</td>
<td>0.0172</td>
<td>88.4</td>
<td>0.0</td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>-1.5</td>
<td>-0.9336</td>
<td>0.1321</td>
<td>0.1026</td>
<td>37.8</td>
<td>2.6</td>
<td>-1.5</td>
<td>-1.0337</td>
<td>0.1542</td>
<td>0.1142</td>
<td>31.1</td>
<td>9.6</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>-1.0</td>
<td>-0.4410</td>
<td>0.2841</td>
<td>0.2739</td>
<td>55.9</td>
<td>44.3</td>
<td>-1.0</td>
<td>-0.4934</td>
<td>0.2874</td>
<td>0.2881</td>
<td>50.7</td>
<td>56.4</td>
</tr>
<tr>
<td>$a$</td>
<td>-</td>
<td>1.2198</td>
<td>0.1540</td>
<td>0.1167</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.1069</td>
<td>0.1408</td>
<td>0.1068</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$B^0(a = 1.2, b = 3, t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$B^0(a = 0.8, b = 0.3, t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>3.5</td>
<td>3.4804</td>
<td>0.2703</td>
<td>0.0789</td>
<td>0.6</td>
<td>51.1</td>
<td>3.5</td>
<td>3.4968</td>
<td>0.2427</td>
<td>0.0806</td>
<td>0.1</td>
<td>47.8</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.3</td>
<td>-0.2662</td>
<td>0.2086</td>
<td>0.0434</td>
<td>11.3</td>
<td>28.6</td>
<td>-0.3</td>
<td>-0.2974</td>
<td>0.2031</td>
<td>0.0455</td>
<td>0.9</td>
<td>32.1</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.1</td>
<td>0.0835</td>
<td>0.0421</td>
<td>0.0055</td>
<td>16.5</td>
<td>21.3</td>
<td>0.1</td>
<td>0.0958</td>
<td>0.0436</td>
<td>0.0057</td>
<td>4.2</td>
<td>24.1</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.3</td>
<td>0.5572</td>
<td>0.0312</td>
<td>0.0174</td>
<td>85.7</td>
<td>0.0</td>
<td>0.3</td>
<td>0.5628</td>
<td>0.0306</td>
<td>0.0171</td>
<td>87.6</td>
<td>0.0</td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>-1.5</td>
<td>-1.0004</td>
<td>0.1568</td>
<td>0.1122</td>
<td>33.3</td>
<td>8.7</td>
<td>-1.5</td>
<td>-0.8626</td>
<td>0.0804</td>
<td>0.0826</td>
<td>42.5</td>
<td>0.0</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>-1.0</td>
<td>-0.4640</td>
<td>0.3060</td>
<td>0.2875</td>
<td>53.6</td>
<td>51.1</td>
<td>-1.0</td>
<td>-0.4770</td>
<td>0.2501</td>
<td>0.2540</td>
<td>52.3</td>
<td>43.6</td>
</tr>
<tr>
<td>$a$</td>
<td>-</td>
<td>1.1835</td>
<td>0.1656</td>
<td>0.1173</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.6691</td>
<td>0.0591</td>
<td>0.0549</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
### 3. Schizophrenia trial data: PANSS data

Table 6.2: PANSS data from APTS short course (blank is missing) - Table 1/4

<table>
<thead>
<tr>
<th>ID</th>
<th>Yt0</th>
<th>Yt1</th>
<th>Yt2</th>
<th>Yt4</th>
<th>Yt6</th>
<th>Yt8</th>
<th>trt</th>
<th>nobs</th>
<th>survTime</th>
<th>δi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>91</td>
<td>1</td>
<td>1</td>
<td>0.70</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>72</td>
<td></td>
<td>1</td>
<td>0.74</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>108</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>106</td>
<td>93</td>
<td>1</td>
<td>1.22</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>77</td>
<td>80</td>
<td>1</td>
<td>1.30</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>120</td>
<td>133</td>
<td>1</td>
<td>1.54</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>113</td>
<td>130</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>101</td>
<td>134</td>
<td>2</td>
<td>1.03</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>97</td>
<td>97</td>
<td>2</td>
<td>1.04</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>97</td>
<td>118</td>
<td>2</td>
<td>1.12</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>96</td>
<td>90</td>
<td>2</td>
<td>1.19</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>76</td>
<td>76</td>
<td>2</td>
<td>1.24</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>99</td>
<td>123</td>
<td>2</td>
<td>1.32</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>111</td>
<td>122</td>
<td>2</td>
<td>1.49</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>120</td>
<td>130</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>91</td>
<td>86</td>
<td>2</td>
<td>1.60</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>105</td>
<td>135</td>
<td>2</td>
<td>1.80</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>113</td>
<td>118</td>
<td>2</td>
<td>1.96</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>81</td>
<td>71</td>
<td>3</td>
<td>1.13</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>119</td>
<td>118</td>
<td>3</td>
<td>1.16</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>77</td>
<td>94</td>
<td>3</td>
<td>1.28</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>100</td>
<td>86</td>
<td>3</td>
<td>1.56</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>67</td>
<td>108</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>127</td>
<td>98</td>
<td>152</td>
<td>1</td>
<td>3</td>
<td>2.14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>161</td>
<td>116</td>
<td>85</td>
<td>1</td>
<td>3</td>
<td>3.10</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>87</td>
<td>92</td>
<td>103</td>
<td>1</td>
<td>3</td>
<td>3.22</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>73</td>
<td>87</td>
<td>91</td>
<td>1</td>
<td>3</td>
<td>3.92</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>118</td>
<td>116</td>
<td>153</td>
<td>2</td>
<td>3</td>
<td>2.06</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>92</td>
<td>72</td>
<td>78</td>
<td>2</td>
<td>3</td>
<td>2.10</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>73</td>
<td>74</td>
<td>68</td>
<td>2</td>
<td>3</td>
<td>2.20</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>112</td>
<td>112</td>
<td>117</td>
<td>2</td>
<td>3</td>
<td>2.50</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>90</td>
<td>100</td>
<td>113</td>
<td>2</td>
<td>3</td>
<td>2.55</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>60</td>
<td>93</td>
<td>116</td>
<td>2</td>
<td>3</td>
<td>2.71</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>81</td>
<td>114</td>
<td>135</td>
<td>2</td>
<td>3</td>
<td>3.07</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Yt0</td>
<td>Yt1</td>
<td>Yt2</td>
<td>Yt4</td>
<td>Yt6</td>
<td>Yt8</td>
<td>trt</td>
<td>nobs</td>
<td>survTime</td>
<td>δ_i</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>----------</td>
<td>-----</td>
</tr>
<tr>
<td>35</td>
<td>93</td>
<td>86</td>
<td>94</td>
<td>2</td>
<td>3</td>
<td>3.15</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>36</td>
<td>94</td>
<td>107</td>
<td>104</td>
<td>3</td>
<td>3</td>
<td>2.42</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>37</td>
<td>104</td>
<td>106</td>
<td>102</td>
<td>3</td>
<td>3</td>
<td>2.50</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>38</td>
<td>87</td>
<td>87</td>
<td>92</td>
<td>3</td>
<td>3</td>
<td>2.68</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>39</td>
<td>75</td>
<td>52</td>
<td>46</td>
<td>3</td>
<td>3</td>
<td>2.68</td>
<td>0</td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>65</td>
<td>57</td>
<td>62</td>
<td>3</td>
<td>3</td>
<td>3.06</td>
<td>0</td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>41</td>
<td>84</td>
<td>80</td>
<td>72</td>
<td>3</td>
<td>3</td>
<td>3.71</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>42</td>
<td>75</td>
<td>92</td>
<td>117</td>
<td>3</td>
<td>3</td>
<td>3.79</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>43</td>
<td>102</td>
<td>109</td>
<td>111</td>
<td>108</td>
<td>1</td>
<td>4</td>
<td>4.34</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>44</td>
<td>84</td>
<td>70</td>
<td>55</td>
<td>74</td>
<td>1</td>
<td>4</td>
<td>4.42</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>45</td>
<td>99</td>
<td>98</td>
<td>92</td>
<td>112</td>
<td>1</td>
<td>4</td>
<td>4.71</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>46</td>
<td>75</td>
<td>70</td>
<td>69</td>
<td>71</td>
<td>1</td>
<td>4</td>
<td>4.71</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>47</td>
<td>84</td>
<td>99</td>
<td>105</td>
<td>98</td>
<td>1</td>
<td>4</td>
<td>4.86</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>48</td>
<td>107</td>
<td>84</td>
<td>100</td>
<td>122</td>
<td>1</td>
<td>4</td>
<td>5.00</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>49</td>
<td>129</td>
<td>113</td>
<td>116</td>
<td>114</td>
<td>1</td>
<td>4</td>
<td>5.25</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>50</td>
<td>64</td>
<td>68</td>
<td>66</td>
<td>106</td>
<td>1</td>
<td>4</td>
<td>5.33</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>51</td>
<td>79</td>
<td>91</td>
<td>73</td>
<td>89</td>
<td>1</td>
<td>4</td>
<td>5.54</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>52</td>
<td>120</td>
<td>103</td>
<td>86</td>
<td>111</td>
<td>1</td>
<td>4</td>
<td>5.60</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>53</td>
<td>98</td>
<td>98</td>
<td>96</td>
<td>80</td>
<td>1</td>
<td>4</td>
<td>5.61</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>54</td>
<td>94</td>
<td>98</td>
<td>111</td>
<td>110</td>
<td>1</td>
<td>4</td>
<td>5.86</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>55</td>
<td>87</td>
<td>104</td>
<td>95</td>
<td>99</td>
<td>2</td>
<td>4</td>
<td>4.28</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>56</td>
<td>80</td>
<td>84</td>
<td>77</td>
<td>85</td>
<td>2</td>
<td>4</td>
<td>4.70</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>57</td>
<td>90</td>
<td>74</td>
<td>74</td>
<td>92</td>
<td>2</td>
<td>4</td>
<td>4.73</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>58</td>
<td>91</td>
<td>90</td>
<td>90</td>
<td>104</td>
<td>2</td>
<td>4</td>
<td>5.15</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>59</td>
<td>71</td>
<td>96</td>
<td>96</td>
<td>97</td>
<td>2</td>
<td>4</td>
<td>5.31</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>60</td>
<td>123</td>
<td>113</td>
<td>104</td>
<td>121</td>
<td>2</td>
<td>4</td>
<td>5.57</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>61</td>
<td>78</td>
<td>84</td>
<td>91</td>
<td>107</td>
<td>2</td>
<td>4</td>
<td>5.90</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>62</td>
<td>83</td>
<td>92</td>
<td>88</td>
<td>100</td>
<td>3</td>
<td>4</td>
<td>4.16</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>63</td>
<td>76</td>
<td>54</td>
<td>51</td>
<td>65</td>
<td>3</td>
<td>4</td>
<td>4.41</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>64</td>
<td>106</td>
<td>102</td>
<td>95</td>
<td>95</td>
<td>3</td>
<td>4</td>
<td>5.00</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>65</td>
<td>95</td>
<td>87</td>
<td>101</td>
<td>104</td>
<td>3</td>
<td>4</td>
<td>5.57</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>66</td>
<td>108</td>
<td>100</td>
<td>74</td>
<td>72</td>
<td>3</td>
<td>4</td>
<td>5.92</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>67</td>
<td>87</td>
<td>83</td>
<td>89</td>
<td>98</td>
<td>107</td>
<td>1</td>
<td>5</td>
<td>6.16</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>68</td>
<td>110</td>
<td>95</td>
<td>92</td>
<td>103</td>
<td>93</td>
<td>1</td>
<td>5</td>
<td>6.31</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>69</td>
<td>74</td>
<td>60</td>
<td>54</td>
<td>49</td>
<td>45</td>
<td>1</td>
<td>5</td>
<td>7.98</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>70</td>
<td>100</td>
<td>90</td>
<td>91</td>
<td>90</td>
<td>98</td>
<td>2</td>
<td>5</td>
<td>6.10</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Table 6.4: PANSS data from APTS short course (blank is missing) - Table 3/4

<table>
<thead>
<tr>
<th>ID</th>
<th>Yt0</th>
<th>Yt1</th>
<th>Yt2</th>
<th>Yt4</th>
<th>Yt6</th>
<th>Yt8</th>
<th>trt</th>
<th>nobs</th>
<th>survTime</th>
<th>δi</th>
</tr>
</thead>
<tbody>
<tr>
<td>71</td>
<td>86</td>
<td>94</td>
<td>86</td>
<td>93</td>
<td>109</td>
<td></td>
<td>2</td>
<td>5</td>
<td>6.71</td>
<td>1</td>
</tr>
<tr>
<td>72</td>
<td>64</td>
<td>72</td>
<td>65</td>
<td>64</td>
<td>80</td>
<td></td>
<td>2</td>
<td>5</td>
<td>6.79</td>
<td>0</td>
</tr>
<tr>
<td>73</td>
<td>84</td>
<td>60</td>
<td>94</td>
<td>82</td>
<td>97</td>
<td></td>
<td>2</td>
<td>5</td>
<td>7.17</td>
<td>1</td>
</tr>
<tr>
<td>74</td>
<td>96</td>
<td>85</td>
<td>88</td>
<td>114</td>
<td>100</td>
<td></td>
<td>2</td>
<td>5</td>
<td>7.18</td>
<td>1</td>
</tr>
<tr>
<td>75</td>
<td>121</td>
<td>103</td>
<td>84</td>
<td>62</td>
<td>144</td>
<td></td>
<td>2</td>
<td>5</td>
<td>7.59</td>
<td>1</td>
</tr>
<tr>
<td>76</td>
<td>107</td>
<td>111</td>
<td>95</td>
<td>97</td>
<td>109</td>
<td></td>
<td>2</td>
<td>5</td>
<td>7.79</td>
<td>1</td>
</tr>
<tr>
<td>77</td>
<td>70</td>
<td>57</td>
<td>57</td>
<td>52</td>
<td>47</td>
<td></td>
<td>3</td>
<td>5</td>
<td>6.80</td>
<td>0</td>
</tr>
<tr>
<td>78</td>
<td>113</td>
<td>102</td>
<td>93</td>
<td>111</td>
<td>106</td>
<td></td>
<td>3</td>
<td>5</td>
<td>7.02</td>
<td>0</td>
</tr>
<tr>
<td>79</td>
<td>100</td>
<td>54</td>
<td>53</td>
<td>58</td>
<td>44</td>
<td></td>
<td>3</td>
<td>5</td>
<td>7.30</td>
<td>0</td>
</tr>
<tr>
<td>80</td>
<td>124</td>
<td>107</td>
<td>115</td>
<td>135</td>
<td>152</td>
<td></td>
<td>3</td>
<td>5</td>
<td>7.30</td>
<td>1</td>
</tr>
<tr>
<td>81</td>
<td>105</td>
<td>74</td>
<td>89</td>
<td>75</td>
<td>89</td>
<td></td>
<td>3</td>
<td>5</td>
<td>7.34</td>
<td>1</td>
</tr>
<tr>
<td>82</td>
<td>140</td>
<td>109</td>
<td>76</td>
<td>63</td>
<td>54</td>
<td></td>
<td>3</td>
<td>5</td>
<td>7.64</td>
<td>0</td>
</tr>
<tr>
<td>83</td>
<td>93</td>
<td>90</td>
<td>91</td>
<td>94</td>
<td>102</td>
<td>109</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>1</td>
</tr>
<tr>
<td>84</td>
<td>60</td>
<td>79</td>
<td>55</td>
<td>58</td>
<td>65</td>
<td>51</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>85</td>
<td>98</td>
<td>78</td>
<td>64</td>
<td>55</td>
<td>57</td>
<td>62</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>86</td>
<td>90</td>
<td>96</td>
<td>86</td>
<td>78</td>
<td>78</td>
<td>86</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>87</td>
<td>123</td>
<td>126</td>
<td>104</td>
<td>101</td>
<td>94</td>
<td>93</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>88</td>
<td>102</td>
<td>119</td>
<td>112</td>
<td>98</td>
<td>90</td>
<td>80</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>89</td>
<td>83</td>
<td>73</td>
<td>82</td>
<td>80</td>
<td>76</td>
<td>86</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>90</td>
<td>60</td>
<td>52</td>
<td>49</td>
<td>45</td>
<td>46</td>
<td>50</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>91</td>
<td>75</td>
<td>79</td>
<td>76</td>
<td>79</td>
<td>86</td>
<td>91</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>92</td>
<td>78</td>
<td>74</td>
<td>62</td>
<td>60</td>
<td>73</td>
<td>82</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>93</td>
<td>71</td>
<td>73</td>
<td>71</td>
<td>61</td>
<td>56</td>
<td>58</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>94</td>
<td>110</td>
<td>106</td>
<td>115</td>
<td>131</td>
<td>128</td>
<td>125</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>95</td>
<td>98</td>
<td>102</td>
<td>85</td>
<td>86</td>
<td>101</td>
<td>94</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>96</td>
<td>83</td>
<td>80</td>
<td>88</td>
<td>85</td>
<td>85</td>
<td>79</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>97</td>
<td>105</td>
<td>53</td>
<td>53</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>98</td>
<td>84</td>
<td>67</td>
<td>80</td>
<td>67</td>
<td>77</td>
<td>74</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>99</td>
<td>104</td>
<td>75</td>
<td>73</td>
<td>76</td>
<td>75</td>
<td>71</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>105</td>
<td>100</td>
<td>109</td>
<td>91</td>
<td>95</td>
<td>69</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>101</td>
<td>64</td>
<td>62</td>
<td>82</td>
<td>83</td>
<td>74</td>
<td>65</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>102</td>
<td>93</td>
<td>97</td>
<td>104</td>
<td>107</td>
<td>94</td>
<td>98</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>103</td>
<td>73</td>
<td>75</td>
<td>65</td>
<td>64</td>
<td>66</td>
<td>62</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>104</td>
<td>88</td>
<td>60</td>
<td>49</td>
<td>65</td>
<td>46</td>
<td>47</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>77</td>
<td>82</td>
<td>73</td>
<td>63</td>
<td>52</td>
<td>56</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>106</td>
<td>107</td>
<td>48</td>
<td>73</td>
<td>75</td>
<td>54</td>
<td>56</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>107</td>
<td>77</td>
<td>82</td>
<td>70</td>
<td>66</td>
<td>60</td>
<td>55</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>108</td>
<td>106</td>
<td>76</td>
<td>84</td>
<td>88</td>
<td>88</td>
<td>88</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>109</td>
<td>72</td>
<td>64</td>
<td>53</td>
<td>60</td>
<td>98</td>
<td>98</td>
<td>3</td>
<td>5</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>110</td>
<td>113</td>
<td>105</td>
<td>116</td>
<td>95</td>
<td>113</td>
<td>111</td>
<td>2</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 6.5: PANSS data from APTS short course (blank is missing) - Table 4/4

<table>
<thead>
<tr>
<th>ID</th>
<th>Yt0</th>
<th>Yt1</th>
<th>Yt2</th>
<th>Yt4</th>
<th>Yt6</th>
<th>Yt8</th>
<th>trt</th>
<th>nobs</th>
<th>survTime</th>
<th>δi</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>89</td>
<td>90</td>
<td>66</td>
<td>76</td>
<td>101</td>
<td>113</td>
<td>2</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>112</td>
<td>72</td>
<td>57</td>
<td>57</td>
<td>53</td>
<td>55</td>
<td>48</td>
<td>2</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>113</td>
<td>95</td>
<td>110</td>
<td>107</td>
<td>86</td>
<td>95</td>
<td>108</td>
<td>2</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>114</td>
<td>77</td>
<td>57</td>
<td>64</td>
<td>52</td>
<td>52</td>
<td>47</td>
<td>2</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>115</td>
<td>84</td>
<td>93</td>
<td>75</td>
<td>79</td>
<td>79</td>
<td>73</td>
<td>2</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>116</td>
<td>132</td>
<td>138</td>
<td>110</td>
<td>119</td>
<td>110</td>
<td>106</td>
<td>2</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>117</td>
<td>94</td>
<td>81</td>
<td>86</td>
<td>84</td>
<td>81</td>
<td>84</td>
<td>2</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>118</td>
<td>63</td>
<td>66</td>
<td>39</td>
<td>58</td>
<td>73</td>
<td>95</td>
<td>2</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>119</td>
<td>96</td>
<td>82</td>
<td>73</td>
<td>85</td>
<td>81</td>
<td>89</td>
<td>2</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>120</td>
<td>67</td>
<td>77</td>
<td>85</td>
<td>113</td>
<td>122</td>
<td>106</td>
<td>2</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>130</td>
<td>67</td>
<td>74</td>
<td>70</td>
<td>68</td>
<td>66</td>
<td>99</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>131</td>
<td>110</td>
<td>87</td>
<td>95</td>
<td>61</td>
<td>56</td>
<td>59</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>132</td>
<td>137</td>
<td>85</td>
<td>56</td>
<td>60</td>
<td>49</td>
<td>59</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>133</td>
<td>101</td>
<td>65</td>
<td>44</td>
<td>36</td>
<td>33</td>
<td>37</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>134</td>
<td>98</td>
<td>95</td>
<td>91</td>
<td>91</td>
<td>89</td>
<td>93</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>135</td>
<td>95</td>
<td>88</td>
<td>90</td>
<td>61</td>
<td>70</td>
<td>70</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>136</td>
<td>78</td>
<td>60</td>
<td>61</td>
<td>49</td>
<td>55</td>
<td>52</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>137</td>
<td>82</td>
<td>79</td>
<td>73</td>
<td>68</td>
<td>64</td>
<td>66</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>138</td>
<td>92</td>
<td>96</td>
<td>81</td>
<td>75</td>
<td>92</td>
<td>95</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>139</td>
<td>70</td>
<td>66</td>
<td>63</td>
<td>55</td>
<td>46</td>
<td>57</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>140</td>
<td>62</td>
<td>60</td>
<td>56</td>
<td>54</td>
<td>53</td>
<td>54</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>141</td>
<td>60</td>
<td>52</td>
<td>55</td>
<td>46</td>
<td>39</td>
<td>42</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>142</td>
<td>93</td>
<td>96</td>
<td>84</td>
<td>85</td>
<td>98</td>
<td>78</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>143</td>
<td>79</td>
<td>86</td>
<td>91</td>
<td>66</td>
<td>57</td>
<td>106</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>144</td>
<td>113</td>
<td>99</td>
<td>98</td>
<td>84</td>
<td>98</td>
<td>95</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>145</td>
<td>74</td>
<td>60</td>
<td>55</td>
<td>69</td>
<td>58</td>
<td>50</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>146</td>
<td>96</td>
<td>78</td>
<td>77</td>
<td>77</td>
<td>117</td>
<td>114</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>147</td>
<td>74</td>
<td>62</td>
<td>70</td>
<td>58</td>
<td>59</td>
<td>57</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>148</td>
<td>81</td>
<td>84</td>
<td>78</td>
<td>54</td>
<td>54</td>
<td>53</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>149</td>
<td>96</td>
<td>75</td>
<td>67</td>
<td>74</td>
<td>74</td>
<td>93</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>150</td>
<td>89</td>
<td>87</td>
<td>82</td>
<td>76</td>
<td>61</td>
<td>64</td>
<td>3</td>
<td>6</td>
<td>8.00</td>
<td>0</td>
</tr>
</tbody>
</table>