UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística e Computação Científica

ALIAKBAR MASTANISHIRAZI

Self-Modeling regression and application

Regressão de auto-modelagem e aplicações

Campinas

2016

Aliakbar Mastanishirazi

Self-Modeling regression and application

Regressão de auto-modelagem e aplicações

Tese apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Estatística .

e

Thesis presented to the Institute of Mathematics, Statistics and Scientific Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Statistics.

Orientador: Aluisio de Souza Pinheiro

Este exemplar corresponde à versão final da Tese defendida pelo aluno Aliakbar Mastanishirazi e orientada pelo Prof. Dr. Aluisio de Souza Pinheiro.

Campinas

2016

Agência(s) de fomento e nº(s) de processo(s): CAPES

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

M391s Mastanishirazi, Aliakbar, 1982-
Self-modeling regression and application / Aliakbar Mastanishirazi. – Campinas, SP : [s.n.], 2016.

Orientador: Aluísio de Souza Pinheiro.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Estatística não paramétrica. 2. Spline, Teoria do. 3. Estatística matemática - Estudos longitudinais. 4. Modelos lineares (Estatística). 5. Estatística ordinal. I. Pinheiro, Aluísio de Souza,1967-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Regressão de auto-modelagem e aplicações
Palavras-chave em inglês:
Nonparametric statistics
Spline theory
Mathematical statistics - Longitudinal studies
Linear models (Statistics)
Order statistics
Área de concentração: Estatística
Titulação: Doutor em Estatística
Banca examinadora:
Aluísio de Souza Pinheiro [Orientador]
Júlia Maria Pavan Soler
Aldo William Medina Garay
Luiz Koodi Hotta
Mariana Rodrigues Motta
Data de defesa: 02-03-2016
Programa de Pós-Graduação: Estatística

Tese de Doutorado defendida em 02 de março de 2016 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). ALUÍSIO DE SOUZA PINHEIRO

Prof(a). Dr(a). JÚLIA MARIA PAVAN SOLER

Prof(a). Dr(a). ALDO WILLIAM MEDINA GARAY

Prof(a). Dr(a). LUIZ KOODI HOTTA

Prof(a). Dr(a). MARIANA RODRIGUES MOTTA

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no processo de vida acadêmica do aluno.

À minha família

Agradecimentos

Agradeço a Deus por mais esta realização em minha vida,

A meu orientador e amigo, Professor Dr. Aluisio de Souza Pinheiro, por sua ajuda constante e papos construtivos, não só na parte acadêmica, mas também na minha estadia longe de casa,

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro concedido, sem o qual não seria possível a realização de meus estudos de Doutorado em Estatística,

Aos professores Julia Pavan Soler, Luiz Koodi Hotta, Mariana Rodrigues Motta and Aldo Medina Garay pelas correções e sugestões,

Ao Brasil, minha gratidão eterna,

Ao meus queridos pais, meus primeiros e eternos professores.

Resumo

Essa tese discute quatro modelos para problemas cujas respostas sejam curvas contínuas. Para a modelagem utilizam-se modelos semi-paramétricos de auto-modelagem. A forma comum às curvas é obtida por *spline* enquanto as especificidades das curvas são obtidas por modelos lineares mistos. Com isso, incorporam se as diferenças de amplitude e de deformação temporal a forma geral não-paramétrico.

Todos os modelos propostos têm em comum serem modelos de regressão de auto-modelagem. Eles diferem nas características dos dados em que são aplicáveis: covariaveis invariantes em tempo; censura à direita; modelos de curva com riscos proporcionais; ou riscos acelerados.

Para cada modelos, é apresentado um estudo de simulação e uma analise de dados reais.

Palavras-chave: Não-paramétrico, *spline* penalizada, dados longitudinais, semi-paramétrico.

Abstract

We discuss in this dissertation models for problems in which responses are continues curves. We employ a semi-parametric model as follows. The common shape of the curves is modeled by spline. The specificities of the curves are modeled by linear mixed models so that different amplitude and time deformations may be incorporated to the underlying shape.

All the four proposed models are self-modeling regression models. They differ in their applications; time-invariant covariates; right censoring; proportional hazard cure; or accelerated hazards.

For each model a simulation study and an analysis of real data are provided.

Keywords: Non-parametric, penalized spline, longitudinal data, semi-parametric.

Lista de ilustrações

Figura 1.1 – Threshold concept for an ordinal response with three categories. 27
Figura 1.2 – Hazard curves from the PH model and AH model. 27
Figura 2.1 – Spirometer data. Volume of air expelled versus time for 18 individuals. 40

Lista de tabelas

Tabela 2.1 – The estimates of all parameters of the proportional odds model by MCEMNR (SE in parenthesis) 40
Tabela 2.2 – The estimates of all parameters of the Self generating ordinal model by MCEMNR (SE in parenthesis) 41
Tabela 2.2 – The estimates of all parameters of the Self generating ordinal model by MCEMNR (SE in parenthesis) 42
Tabela 2.3 – Values of the deviance for the best subset of each size 42
Tabela 2.4 – Estimates of intercepts and coefficients of the fourth model 42
Tabela 3.1 – The estimates of all parameters of the proportional odds model under right censoring (10%) by MCEMNR (SE in parenthesis) 53
Tabela 3.2 – The estimates of all parameters of the proportional odds model under right censoring (20%) by MCEMNR (SE in parenthesis) 54
Tabela 3.3 – The estimates of all parameters of the proportional odds model under right censoring (40%) by MCEMNR (SE in parenthesis) 55
Tabela 3.4 – The estimates of all parameters of the Self generating ordinal model under right censoring (10%) by MCEMNR (SE in parenthesis) 56
Tabela 3.4 – The estimates of all parameters of the Self generating ordinal model under right censoring (10%) by MCEMNR (SE in parenthesis) 57
Tabela 3.5 – The estimates of all parameters of the Self generating ordinal model under right censoring (20%) by MCEMNR (SE in parenthesis) 58
Tabela 3.5 – The estimates of all parameters of the Self generating ordinal model under right censoring (20%) by MCEMNR (SE in parenthesis) 59
Tabela 3.6 – The estimates of all parameters of the Self generating ordinal model under right censoring (40%) by MCEMNR (SE in parenthesis) 60
Tabela 3.6 – The estimates of all parameters of the Self generating ordinal model under right censoring (40%) by MCEMNR (SE in parenthesis) 61
Tabela 3.7 – Values of the deviance for the best subset of each size in two censoring rates 61
Tabela 3.8 – Values of AIC and BIC for the best subset of each size in two censoring rates 61
Tabela 3.9 – Estimates of intercepts and coefficients of the sixth model for two censoring rates 62
Tabela 4.1 – The estimates of all parameters of the PHC model with censoring rate (10%) by MCEMNR (SE in parenthesis) 74
Tabela 4.2 – The estimates of all parameters of the PHC model with censoring rate (20%) by MCEMNR (SE in parenthesis) 75

Tabela 4.3 – The estimates of all parameters of the PHC model with censoring rate (40%) by MCEMNR (SE in parenthesis) 76
Tabela 4.4 – The estimates of all parameters of the SI-PHC model with censoring rate (10%) by MCEMNR (SE in parenthesis) 77
Tabela 4.4 – The estimates of all parameters of the SI-PHC model with censoring rate (10%) by MCEMNR (SE in parenthesis) 78
Tabela 4.5 – The estimates of all parameters of the SI-PHC model with censoring rate (20%) by MCEMNR (SE in parenthesis) 79
Tabela 4.5 – The estimates of all parameters of the SI-PHC model with censoring rate (20%) by MCEMNR (SE in parenthesis) 80
Tabela 4.6 – The estimates of all parameters of the SI-PHC model with censoring rate (40%) by MCEMNR (SE in parenthesis) 81
Tabela 4.6 – The estimates of all parameters of the SI-PHC model with censoring rate (40%) by MCEMNR (SE in parenthesis) 82
Tabela 4.7 – Estimates of intercepts and coefficients of the model for two censoring rates 82
Tabela 5.1 – The estimates of all parameters of the AHMC model with censoring rate (10%) by MCEMNR (SE in parenthesis) 93
Tabela 5.2 – The estimates of all parameters of the AHMC model with censoring rate (20%) by MCEMNR (SE in parenthesis) 94
Tabela 5.3 – The estimates of all parameters of the AHMC model with censoring rate (40%) by MCEMNR (SE in parenthesis) 95
Tabela 5.4 – The estimates of all parameters of the SI-AHMC model with censoring rate (10%) by MCEMNR (SE in parenthesis) 96
Tabela 5.4 – The estimates of all parameters of the SI-AHMC model with censoring rate (10%) by MCEMNR (SE in parenthesis) 97
Tabela 5.5 – The estimates of all parameters of the SI-AHMC model with censoring rate (20%) by MCEMNR (SE in parenthesis) 98
Tabela 5.5 – The estimates of all parameters of the SI-AHMC model with censoring rate (20%) by MCEMNR (SE in parenthesis) 99
Tabela 5.6 – The estimates of all parameters of the SI-AHMC model with censoring rate (40%) by MCEMNR (SE in parenthesis) 100
Tabela 5.6 – The estimates of all parameters of the SI-AHMC model with censoring rate (40%) by MCEMNR (SE in parenthesis) 101
Tabela 5.7 – Values of the deviance for the best subset of each size in two censoring rates 101
Tabela 5.8 – Values of AIC and BIC for the best subset of each size in two censoring rates 101

Tabela 5.9 – Estimates of intercepts and coefficients of the sixth model for two censoring rates 102

Sumário

Agradecimentos **6**

1 INTRODUCTION **15**

2 SELF-MODELING ORDINAL REGRESSION WITH TIME INVARIANT COVARIATES **28**

2.1 Introduction **28**

2.2 Ordinal longitudinal logistic regression model **30**

2.3 Self-Modeling Regression **32**

2.3.1 Penalized regression spline for Nonparametric Regression 33

2.4 Likelihood function and estimation **35**

2.4.1 An Estimation algorithm 35

2.4.2 MCEMNR algorithms 36

2.5 Simulation **38**

2.6 An Application to the Prostate Cancer Data **38**

3 SELF-MODELING ORDINAL REGRESSION UNDER RIGHT CENSORING **43**

3.1 Introduction **43**

3.2 Self-Modeling Ordinal model under right censoring **44**

3.3 Estimation method **47**

3.3.1 The estimation algorithm 48

3.3.2 MCEMNR algorithms 49

3.4 Simulation **51**

3.5 An Application to Anaesthesia Recovery **51**

4 A PROPORTIONAL HAZARD CURE MODEL FOR ORDINAL RESPONSES BY SELF-MODELING REGRESSION **63**

4.1 Introduction **63**

4.2 Self-Modeling Ordinal model **65**

4.3 The model **67**

4.4 Estimation method **69**

4.4.1 MCEMNR algorithms 70

4.5 Simulation **71**

4.6 An Application to schizophrenia data **72**

5 SELF-MODELING REGRESSION IN THE ACCELERATED HAZARDS MIXTURE CURE MODEL . . . 83
5.1 Introduction . . . 83
5.2 Self-Modeling binary model . . . 85
5.3 The SI-AHMC model . . . 86
5.4 Estimation method . . . 88
5.4.1 MCEMNR algorithms . . . 89
5.5 Simulation . . . 90
5.6 An Application to respiratory illness Data . . . 91

6 CONCLUSIONS AND PROPOSALS FOR FUTURE WORK . . . 103

REFERÊNCIAS . . . 105

APÊNDICES 106

APÊNDICE A – THE DERIVATIVES FOR THE NEWTON-RAPHSON AND EM PROCEDURES . . . 107
A.1 Chapter 2 . . . 107
A.1.1 The Newton-Raphson procedure . . . 107
A.1.2 The E and M steps . . . 110
A.2 Chapter 3 . . . 112
A.2.1 The Newton-Raphson procedure . . . 112
A.2.2 The E- and M- steps . . . 115
A.3 Chapter 4 . . . 117
A.3.1 The Newton-Raphson procedure . . . 117
A.3.2 The E- and M- steps . . . 122
A.4 Chapter 5 . . . 124
A.4.1 The Newton-Raphson procedure . . . 124
A.4.2 The E- and M- steps . . . 128

1 Introduction

In many longitudinal studies such as electrocardiogram readings, growth curves or serum glucose levels following a meal, the response to be modelled is a continuous curve measured over time. The common factor in all of these examples is that the response curves share a similar shape, e.g., the same number of extrema or inflection points located relatively near some common region. The data can be represented as Y_{ij} for $i = 1, ..., N$ and $j = 1, ..., n_i$, a sample of N individuals, curves, or experimental units, with the ith individual measured at n_i times t_{ij}. In such circumstances the interest may be to estimate the shape of the response curve itself or the value of a response feature such as the slope at a particular point or the location of a maximum.

Lawton, Sylvestre and Maggio (1972) proposed the self-modeling approach to such data. Their method is based on the assumption that all individual's response curves have a common shape and that a particular individual's curve is some simple transformation of the common shape curve. The self-modeling model is

$$Y_{ij} = \upsilon_i \left\{\mu_0 \left[\kappa_i(t_{ij})\right]\right\} + e_{ij}, \tag{1.1}$$

where, Y_{ij} is the observed response on subject i at time t_{ij}. $\upsilon_i(x)$ is a monotone inverse link transforming the regression function and $\kappa_i(x)$ is a monotone transformation of the time axis. μ_0 is a shape function that is common to all the curves, and ϵ_{ij} are unobserved errors which may be correlated within subjects. In this dissertation, we will consider the Shape-Invariant model (SI model) introduced by Altman e Villarreal (2004) which is a special case of their self-modelling regression method. The model assumes a common underlying regression function μ_0 and transformations of both the time (t) and response (Y) axes. The SI model is

$$Y_{ij} = \alpha_{0i} + A_{1i}\mu_0(\beta_{0i} + B_{1i}t_{ij}) + \epsilon_{ij} \tag{1.2}$$

Here α_{0i}, A_{1i}, β_{0i} and B_{1i} are unknown parameters which may be functions of observed covariates. Often A_{1i} and B_{1i} are constrained to be positive, in which case we set $A_{1i} = \exp(\alpha_{1i})$ and $B_{1i} = \exp(\beta_{1i})$. In this paper, we parametrize in this exponential form.

When μ_0 is modelled by a parametric function, this can be fitted as a nonlinear mixed or fixed effects model. However, when a parametric form is not known, when goodness of fit to a parametric form is an issue, or for other reasons, it may be convenient to fit a non-parametric function for μ_0. Here we consider only nonparametric fits for μ_0. In this model, α_{0i} and α_{1i} (or A_{1i}) will be referred to as the response scale parameters and β_{0i} e β_{1i} (or B_{1i}) will be referred to as the time scale parameters, although in some contexts we will model them as random effects.

The SI model has several advantages over fitting a separate nonparametric curve to each subject:

1. Differences among the subjects due to time-invariant covariates are captured by the parameters. We will demonstrate that when the parameters can be modelled by a linear mixed model inference about the parameters can proceed as if μ_0 were a known regression function.

2. Curve summaries such as time until maximum, maximum response, area under the curve, time to return to baseline, and most other common summaries can be expressed as a functional of μ_0 which does not depend on the covariates and a function of the parameters which depends on the covariates. Hence, covariate effects on these curve summaries are efficiently expressed by covariate effects on the parameters.

3. The assumption that all subjects have the same response profile allows pooling of information about the profile over subjects. This produces a much better fit for subjects and better interpolation.

4. When the response is multivariate, use of the SI model, with a possibly different shape function for each response, allows a common setting for comparing the effects of covariates across responses (whether the shape is fitted parametrically or non-parametrically). For example in Altman e Villarreal (2004), in the nestling growth study described later in the paper, a dietary supplement (calcium) was given to some of the parent birds, and the effects of this supplement on several aspects of nestling growth are of interest. Using the SI model, we can meaningfully determine whether the effects of the supplement on bone growth were similar to the effects on body mass, for example. This type of comparability across responses might be particularly interesting in very high dimensional problems such as time course studies in gene expression experiments

We are interested in the case when there are time invariant covariates X and Z (possibly multivariate) which might have an effect on the parameters. In particular, letting

$$\boldsymbol{\theta}_i' = (\alpha_{0i}, \alpha_{1i}, \beta_{0i}, \beta_{1i}) \text{ or } \boldsymbol{\theta}_i' = (\alpha_{0i}, A_{1i}, \beta_{0i}, B_{1i})\,, \tag{1.3}$$

we consider a mixed model $\theta_{ik} = g(\boldsymbol{X}_i, \boldsymbol{Z}_i; \phi_k, \psi_{ik}) + \eta_{ik}$ where i refers to the subject k refers to the particular component of the vector $\boldsymbol{\theta}$, g is a parametric function, ϕ_k are fixed effects associated with $\boldsymbol{X}$ and ψ_{ik} are random effects associated with $\boldsymbol{Z}$. For example, if the subjects are measured as part of a randomized complete block (RCB) design, we would fit the randomized complete block model to each of the response and time parameter. Models 1.2 and 1.3 are extensions of the semi-parametric model of Zeger and Diggle (1994) which can be thought of as a SI model with $\alpha_{1i} = \beta_{0i} = \beta_{1i} = 0$.

Kneip and Gasser (1988) and Kneip and Engel (1995) explored the problem of fitting the parameters as fixed effects with no covariate. Kneip and Gasser (1988) also considered identifiability. Capra and Muller (1997) demonstrated that nonparametric smooth functions of the parameters can also be fitted. Lindstrom (1995) tackled the problem of estimating the response and time scale parameters as random effects. Ke and Wang (2001) consider self-modelling regression as a specific case of semi-parametric regression with mixed effects.

There are several special cases of the model that we might consider. In many cases, $t = 0$ has a special meaning, such as the onset of treatment or birth, and in those cases β_{0i} may be fixed at zero. In gene expression data, we might expect some genes to be promoted while others are repressed, so we do not exponentiate. The errors ϵ_{ij} can be modelled parametrically or non-parametrically and can be serially correlated or heteroscedastic, which is useful in the longitudinal setting

In many situations, inference focuses on the parametric effects on one or more summary statistics. Many of the commonly used summary statistics can be expressed as functions of the parameters and a functional of the shape. For example, if $\mu_0(t)$ has a critical point at t_0, then the ith curve has a critical point at $(t_0 - \beta_{0i})/\exp(\beta_{1i})$ and the value of the ith curve at the critical point is $\alpha_{0i} + \exp(\alpha_{1i})\ \mu_0(t_0)$. Hence such summaries as time until optimum, the value at the optimum and return to baseline can be expressed as functions of the parameters. Since the functional of the shape does not vary with the treatment and covariate effects, inferences about the treatment and covariate effects depend only on the parameters. Use of the SI model for inference is therefore very similar to the use of parametric nonlinear mixed models in that the parameters embody all of the information about covariate effects. This is quite different from models in which each curve is fitted non-parametrically with its own shape.

Under Penalized spline smoothing has become increasingly popular in recent years. The idea of penalized splines has led to a powerful and applicable smoothing technique (see Ruppert, Wand and Carroll, 2003). We use penalized regression spline introduced by Ruppert and Carroll (1997) for Nonparametric Regression. The main computational advance of this thesis is the use of penalized regression splines to estimate the shape function μ_0. This provides a sieve likelihood framework which can readily incorporate parametric models for the error covariance structure. As well, the equivalence of this model to a linear random effects model allows automatic selection of the smoothing parameter(s) by the generalized maximum likelihood (GML) method due to Wahba (1985). This provides considerable computational simplification over previous work fitting the SI model with kernel smoothers (e.g., Kneip and Gasser, 1988 and Kneip and Engel, 1995) or regression splines (Lindstrom, 1995), since the smoothing parameter for the non-parametric regression is chosen as part of the overall fitting algorithm.

Often, the response of interest is measured in a series of ordered categories. Such measures termed as *ordinal*, can represent a variety of graded responses such as agreement ratings (disagree, undecided, agree). In some cases, the response measure of interest may represent a count (e.g., number of heath service visits) that has large probability mass at zero (i.e., no service use), a majority of values in the one to two-visit range, and a few extreme values. In these cases, an ordinal variable can be constructed with ordered categories of 0, 1, 2, and 3 or more visits. The relative frequency with which the categories are endorsed is not a factor for the ordinal regression model, whereas quite strict requirements are imposed under the assumption of a Poisson process.

A variable with an unordered categorical scale is called nominal. Examples of nominal variables are religious affiliation (Protestant, Catholic, Jewish, Muslim, other), marital status (married, divorced, widowed, never married), favorite type of music (classical, folk, jazz, rock, other), and preferred place to shop (downtown, Internet, suburban mall). Distinct levels of such variables differ in quality, not in quantity. Therefore, the listing order of the categories of a nominal variable should not affect the statistical analysis.

Many well-known statistical methods for categorical data treat all response variables as nominal. That is, the results are invariant to permutations of the categories of those variables, so they do not utilize the ordering if there is one. Examples are the Pearson chi-squared test of independence and multinomial response modeling using baseline-category logits. Test statistics and P-values take the same values regardless of the order in which categories are listed. Some researchers routinely apply such methods to nominal and ordinal variables alike because they are both categorical.

Recognizing the discrete nature of categorical data is useful for formulating sampling models, such as in assuming that the response variable has a multinomial distribution rather than a normal distribution. However, the distinction regarding whether data are continuous or discrete is often less crucial to substantive conclusions than whether the data are qualitative (nominal) or quantitative (ordinal or interval). Since ordinal variables are inherently quantitative, many of their descriptive measures are more like those for interval variables than those for nominal variables.

Many advantages can be gained from treating an ordered categorical variable as ordinal rather than nominal. They include:

1. Ordinal data description can use measures that are similar to those used in ordinary regression and analysis of variance for quantitative variables, such as correlations, slopes, and means.

2. Ordinal analyses can use a greater variety of models, and those models are more parsimonious and have simpler interpretations than the standard models for nominal variables, such as baseline-category logit models.

3. Ordinal methods have greater power for detecting relevant trend or location alternatives to the null hypothesis of "no effect"of an explanatory variable on the response variable.

4. Interesting ordinal models apply in settings for which standard nominal models are trivial or else have too many parameters to be tested for goodness of fit.

Methods for ordinal response data analysis have been actively pursued, Harville and Mee (1984), Jansen (1990), Ezzet and Whitehead (1991), Agresti and Lang (1993), Hedeker and Gibbons (1994), Have (1996), Tutz and Hennevogl (1996), Fielding (1999) and Santos and Berridge (2000).

In particular, because the proportional odds assumption described by McCullagh (1980), which is based on the logistic regression formulation, is a common choice for analysis of ordinal data, many of the mixed models for ordinal data are generalizations of this model. The proportional odds model characterizes the ordinal responses in L categories $(l = 1, 2, \ldots, L)$ in terms of $L - 1$ cumulative category comparisons, specifically $L - 1$ cumulative logits (i.e., log odds). In the proportional odds model, the covariate effects are assumed to be the same across these cumulative logits, or proportional across the cumulative odds.

Let the L ordered response categories be coded as $l = 1, 2, ..., L$. As ordinal models often utilize cumulative comparisons of the categories, define the cumulative probabilities for the L categories of the outcome y_{ij} as

$$P_{ijl} = \Pr\left(y_{ij} \leqslant l\right) \; = \sum_{k=1}^{l} p_{ijk}, \tag{1.4}$$

where p_{ijk} represents the probability of response in category k. The mixed-effects logistic regression model for the cumulative probabilities is given in terms of the cumulative logits as

$$log\left[\frac{P_{ijl}}{1 - P_{ijl}}\right] = \tau_l - \left[x'_{ij}\beta + z'_{ij}\upsilon_i\right], \qquad l = 1, \ldots, L-1, \tag{1.5}$$

with $L - 1$ strictly increasing model thresholds τ_l $(i.e., \tau_1 < \tau_2 < \ldots < \tau_{L-1})$. x_{ij} is the $(p+1) \times 1$ covariate vector (including the intercept), and z_{ij} is the design vector for the r random effects, both vectors being for the jth timepoint nested within subject i. Also,

$\boldsymbol{\beta}$ is the $(p+1) \times 1$ vector of unknown fixed regression parameters. Let $\upsilon = T\theta$, where $TT^{'} = \Sigma_{\upsilon}$ is the Cholesky factorization of random-effect variance covariance matrix Σ_{υ}.

The relationship between the latent continuous variable y and an ordinal outcome with three categories is depicted in Figure 1.1. In this case, the observed ordinal outcome $Y_{ij} = l$ if $\tau_{l-1} \leqslant y_{ij} < \tau_l$ for the latent variable (with $\tau_0 = -\infty$ and $\tau_L = \infty$). To set the location of the latent variable, it is common to set a threshold to zero. Typically, this is done in terms of the first threshold (i.e., $\tau_1 = 0$). Figure 1.1 illustrates this concept assuming that the continuous latent variable y follows either a normal or logistic probability density function.

In Figure 1.1, setting $\tau_1 = 0$ implies that $\tau_2 = 2$. These threshold parameters, in addition to the model intercept, represent the marginal response probabilities in the L categories. For example, for this case with $L = 3$, $0 - \beta_0$ represents the log odds for a response in the first category, relative to categories 2 and 3; $\tau_2 - \beta_0$ represents the log odds for a response in the first two categories, relative to the third category. An alternative specification is to set the model intercept $\beta_0 = 0$ and to estimate $L-1$ thresholds. Denoting these $L-1$ thresholds as τ^*, we would then have the following relationship between these two parameterizations: $\tau^* = 0 - \beta_0$ e $\tau_2^* = \tau_2 - \beta_0$.

Our primary focus is to frame up a model for longitudinal data where patients suffering from a common disease, exhibits a similar shape even though each patient's response varies substantially over time. In longitudinal analysis for continuous data, the response is a continuous curve measured over time. For example, serum glucose level following a meal for individuals observed over certain time points or the value of air expelled by persons measured different time points. In such situations, plot of raw data indicates a possibly similar shape among the individual's response curve. Lawton, Sylvestre and Maggio (1972) have developed a model such that the model for every individual quite naturally exhibits a common shape function. It becomes a challenge then to develop a shape invariant model when the response variable ceases to be continuous. In fact, in medical studies, often we come across binary or ordinal outcomes observed longitudinally. Typically this scenario happens in our example on prostrate cancer study where the severity of cancer for each patient is observed over different points of time. We build up our shape invariant model based on the following three justifications.

1. One reason for using the proportional odds cumulative logit model is its connection to the idea of a continuous latent response. We assume that the categorical outcome is actually a categorized version of an unobservable continuous variable. In that case it is reasonable to think that a Likert scale is a coarsened version of a continuous variable. The continuous scale is divided into regions formed according to the category. Suppose y^* is related to x through a nonlinear regression $y^* = \eta + e$ where $e \sim$ logistic

(mean=0 and variance $=c_0$), η being a nonlinear function that involves parameters as well as the covariates x and individual specific random effects (indicating variation over individuals though shape invariant). Then the coarsened version y will be related to x by a proportional odds cumulative logit model. The motivation for self generating shape invariant model then comes up through the latent variable.

2. In disease progression or pharmacokinetic studies where disease status (ordinal) is observed longitudinally, the predictors usually appear through nonlinear relationship. In the current investigation we bring non-linearity through the logit function maintaining common shape for all individuals. Suppressing all suffixes, we write a simple version of cumulative logit model where

$$\log\left(\frac{P(y \leqslant l|x)}{P(y > l|x)}\right) = \eta = \tau_l - \omega(x),$$

τ_l being category boundary parameter and $\omega(x)$ being a nonlinear function of covariates and does not involve any category specific parameter. In our set up, $\omega(x) = \alpha_0 + \exp(\alpha_1)\mu_0(t^*)$. As has been assumed in our model, α_0 is a random intercept parameter (not involving x), hence

$$\log\left(\frac{P(y \leqslant l|x_1)/P(y > l|x_1)}{P(y \leqslant l|x_2)/P(y > l|x_2)}\right) = \omega(x_1) - \omega(x_2) = \mu_0(t^*)(e^{\alpha_1(x_1)} - e^{\alpha_1(x_2)}),$$

Thus the common shape is virtually a scale multiple of the log odds under two covariates.

3. Lindstrom and Bates (1990) in their work on nonlinear mixed effects model for longitudinal data, assume a common parametric function that relates the conditional mean of the response variable to the covariate t_{ij} and random parameter vector φ_i. An alternative way to analyze longitudinal data is to use a Self modeling nonlinear regression (SEMOR) introduced by Lawton, Sylvestre and Maggio (1972) where a common curve exists for all individuals. Later Ke and Wang (2001) extend this idea of common shape function through Semi parametric Nonlinear mixed Models. In the same light we develop our ordinal nonlinear mixed model (GLMM) where instead of the conditional mean, the conditional canonical link is expressed in terms of the intercept and common but unknown shape function μ_0.

Survival analysis is generally defined as a set of methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest. The event can be death, occurrence of a disease, marriage, divorce, etc. The time to event or survival time can be measured in days, weeks, years, etc. For example, if the event of interest is heart attack, then the survival time can be the time in years until a person develops a heart attack. In survival analysis, subjects are usually followed over a specified time period and the focus is on the time at which the event of interest occurs. Why not use linear regression to model the survival time as a function of a set of predictor variables? First, survival times are typically positive numbers; ordinary linear regression may not be the best choice unless these times are first transformed in a way that removes this restriction. Second, and more importantly, ordinary linear regression cannot effectively handle the censoring of observations.

Censoring is said to be present when information on time to outcome event is not available for all study participants. Participant is said to be censored when information on time to event is not available due to loss to follow-up or non-occurrence of outcome event before the trial end. Censoring can be classified into three types:

1. Right censoring: a data point is above a certain value but it is unknown by how much.
2. Left censoring: a data point is below a certain value but it is unknown by how much.
3. Interval censoring: a data point is somewhere on an interval between two values.

In a post-surgical recovery study, the status of recovery is assessed for patients who were given different dose of anaesthetic. The ordinal responses are recorded longitudinally along with the recovery stage of a patient. Differences among the patients due to time-invariant covariates are captured by the parameters. Since patients having a common surgery usually exhibit a similar pattern, it is natural to build up a nonlinear model that is shape invariant. In Chapter 3, we proposed the use of self-modeling ordinal longitudinal model based on right-censoring where the conditional cumulative probabilities for a category of an outcome has a relation with shape-invariant model. We focus on the question of whether the dose of anesthesia affects the post-surgical recovery. In particular, we investigate the interaction between the dose effect and time to follow-up.

Statistical models for survival data with a surviving or cure fraction, often called cure models, have received a great deal of attention in the last decade. In medicine and public health researches, survival cure models are widely used to analyse time-to-event data in which some subjects are reasonably believed to be medically cured. In general, there are two types of models for estimation of the cure fraction. The first one is the Mixture Cure Model (MCM), which was developed by Boag (1949). This type of models assumes that the whole population is composed of susceptible subjects and cured subjects. Boag (1949) proposed lognormal normal distribution to model the failure time of the susceptible group and assumed the cure probability to be constant. This model was further developed three years later by Berkson and Gage (1952) and later studied extensively by several authors, e.g., Farewell (1986), Goldman (1984), Kuk and Chen (1992), Maller and Zhou (1996), Taylor (1995), Peng and Dear (2000) and Banerjee and Carlin (2004), among many others.

The second cure model type was proposed by Yakovlev et al. (1993) based on the assumption that the treatment leaves the patient with a number of cancer cells, which may grow slowly over time and produce a detectable recurrence of cancer. It is known as the Non-Mixture Cure Model (NMCM). These two models are related and the NMCM can be transformed into the MCM, when the cure fraction specially specified. NMCM was further discussed by Chen, Ibrahim and Sinha (1999), Ibrahim and Chen (2001),Chen, Ibrahim and Lipsitz (2002) and Tsodikov (2002). This model was motivated by the underlying biological mechanism and developed based on assumption that the number of cancer cells that remain active after cancer treatment follow Poisson distribution (Yakovlev et al., 1993; Chen, Ibrahim and Sinha, 1999; Gutierrez, 2002; Uddin et al., 2006).

For decades, the MCM has been a popular method in analysing time-to-event data in which some subjects are reasonably believed to be cured. This model assumes that a proportion, π, of the subjects will be cured and that these subjects are not at risk of experiencing the re-occurrence of the event. The other proportion $(1 - \pi)$ is for the individuals who are expected to experience the event in some future time eventually. The MCM can be derived as following: Suppose that T denotes the occurrence time of a disease with population survival function $S(t)$ and that y expresses a binary random variable taking the values 1 and 0 with probability $(1 - \pi)$ (event rate) and π (cure rate), respectively, where $\pi = Pr(T = \infty)$. Furthermore, let $S_u(t)$ and $f_u(t)$ be the survival and density function for uncured group. So, the population survival function $S(t)$ can be represented as:

$$S(t|\boldsymbol{x}, \boldsymbol{z}) = \pi(\boldsymbol{z})S_u(t|\boldsymbol{x}) + [1 - \pi(\boldsymbol{z})], \tag{1.6}$$

where $\boldsymbol{x}$ and $\boldsymbol{z}$ are two sets of covariates that have effects on π and $S_u(t)$. The density function corresponding to 1.6 is $f(t) = (1 - \pi)f_u(t)$. Different parametric distributions have been used to model the density function $f_u(t)$, including Exponential distribution (Ghitany, Maller and Zhou, 1994), Weibull distribution (Farewell, 1986), Log normal (Boag, 1949; Gamel, McLean and Rosenberg, 1990). Nonparametric approaches for $f_u(t)$ have also been considered in the literature (Taylor, 1995; Kuk and Chen, 1992; Sy and Taylor, 2000; Peng and Dear, 2000). The advantage of the mixture cure model is that the proportion of cured patients and the survival distribution of uncured patients are modeled separately and the interpretation of the parameters of x and z in the model is straightforward.

Similar to the classical survival models, there are a number of methods to specify the effects of $\boldsymbol{x}$ on $S_u(t)$. Let $S_0(t)$ be an arbitrary baseline survival function.

Similar to the proportional hazards (PH) model in survival analysis, one can assume

$$S_u(t|\boldsymbol{x}) = \{S_0(t)\}^{\exp(\beta \boldsymbol{x})} \tag{1.7}$$

or equivalently

$$h_u(t|\boldsymbol{x}) = \{h_0(t)\} \exp(\beta \boldsymbol{x}) \tag{1.8}$$

where h_u and $h_0(t)$ are the corresponding hazard functions of S_u and $S_0(t)$. This model is referred to as the proportional hazards mixture cure (PHMC) model. The model can be easily estimated if the baseline survival function $S_0(t)$ is specified up to a few unknown parameters.

The most common method to specify the effects of $\boldsymbol{z}$ on π is via a logit link function:

$$\pi(\boldsymbol{z}_i) = \frac{\exp(\gamma' \boldsymbol{z}_i)}{1 + \exp(\gamma' \boldsymbol{z}_i)}, \tag{1.9}$$

where γ is a vector of unknown parameters. Other link functions may be considered, such as the complementary log–log and the probit link functions in the generalized linear models for binary data. In this paper, we will use the logit link function only because of its simplicity and popularity. Kuk and Chen (1992) considered the semi-parametric logistic proportional hazard mixture model. They focused on estimation of regression parameters using a marginal likelihood method. Sy and Taylor (2000) and Peng and Dear (2000) used the full likelihood approach and derived some EM algorithms to compute the maximum likelihood estimator.

In a cure model, the population can be a mixture of several levels of susceptible and non-susceptible (cured) individuals. In a cancer study, patients have various stages of illness. After treatment the patient will be cured or the stage of illness will change. In such situations where there is good scientific or empirical evidence of a susceptible population by several levels, we combine a Self-Modeling ordinal model for the probability of occurrence of an event with a Cox regression for the time of occurrence of an event. Unlike the univariate mixture cure model with two stages (Boag,1949, Farewell, 1982), a mixture cure model with several levels of cured and uncured patients is proposed in Chapter 4. The model allocates the probability of these stages by the ordinal regression of self-modeling precisely. We will consider schizophrenia as a mental illness that can come in various forms with different symptoms and outcomes and effects of four medications on schizophrenia patients.

To model a gradual treatment effect for data without a cure fraction, Chen and Wang (2001) and Chen (2000) proposed an accelerated hazard (AH) model

$$h_u(t|\boldsymbol{x}) = h_0\left(t\,\exp(\beta\boldsymbol{x})\right). \tag{1.10}$$

For the binary treatment covariate defined above, it is easy to see that the hazard functions of the new and the standard treatments are $h_0(t\exp(\beta))$ and $h_0(t)$ respectively, and the difference of the two hazard functions starts at 0 when $t = 0$. Thus the AH model assumes that the hazard does not change at time 0 and then change gradually with time. Unless $h_0(t) \equiv$ constant or $lim_{t\to0+}h_0(t) = 0$, the AH model provides a useful way to model the gradual effect of a treatment that other existing models cannot handle properly.

To better demonstrate the differences, we plot the hazard curves based on the two models in Figure 1.2. We consider two groups with $x = 0$ for the control (baseline) group and $x = 1$ for the treatment group. The baseline hazard function is a U-shape function, which is often employed in health research. The value of β is set to -0.8. Comparing the hazard curves from the two groups, we can see that the PH model implies that the treatment decreases the hazard rate by $e^{-0.8} = 0.45$ for the whole period. The AH model provides a simple scenario: the treatment starts at the same hazard rate as the control group, it has a higher hazard rate than the control group at the early period due to, say, the toxicity of the treatment. However, after certain time point, the positive effect of the treatment is demonstrated with a smaller hazard rate than the control group.

Chen and Wang (2001) proposed estimating equations to estimate the parameters semi-parametrically in the AH model 1.10. When there is a cure fraction in the data, the model 1.10 is clearly not appropriate. It is unclear whether the model and the semi-parametrically estimation method can be easily adapted to incorporate the cure fraction.

To allow a gradual effect of covariates on the failure time of uncured patients, Zhang and Peng (2009) proposed to model $S_u(t)$ in the mixture cure model 1.7 by the AH model proposed by Chen (2000). That is,

$$S_u(t|\boldsymbol{x}) = S_0\left(t\,\exp(\beta\boldsymbol{x})\right)^{\exp(-\beta\boldsymbol{x})}, \tag{1.11}$$

where $S_0(t)$ is the corresponding survival function. Zhang and Peng (2009) refered to the model specified by Eqs. 1.7, 1.9, and 1.11 as the AH mixture cure (AHMC) model. Their mixture cure model (AHMC) employs a AH model to model the effects of x on $S_u(t)$ in the mixture cure model 1.6. the AHMC model allows covariate effects on the failure time distribution of uncured patients to be negligible at time zero and to increase as time goes by. Such a model is particularly useful in some cancer treatments when the treat effect increases gradually from zero. If $h_0(t)$ is specified up to a few unknown parameters in the AHMC model, the parameters in the model can be estimated by the maximum likelihood approach. Zhang and Peng (2009) focused on a semi-parametric estimation approach where $h_0(t)$ is not parametrically specified. This approach is more attractive in application because it does not rely on a parametric assumption that may be difficult to verify.

A AHMCM with random effects is proposed in Chapter 5. We extend the AHMCM such that the extended model can be applied for the time of occurrence of an event when Self-Modeling binary model is used for the probability of occurrence of an event. As an application of the model, we employ the proposed model to the respiratory illness data set.

The aims of this dissertation are:

1. The shape invariant model is required to be developed through proportional odds model.

2. The use of self-modeling longitudinal ordinal model based on right censoring is developed.

3. A new cure model with several stages of cured or uncured patients combining the self-modeling ordinal model for the probability of occurrence of an event with a Cox regression analysis for the occurrence time of an event.

4. The longitudinal binary model is employed for a accelerated hazard mixture cure model such that the extended model can be applied to the occurrence time of an event.

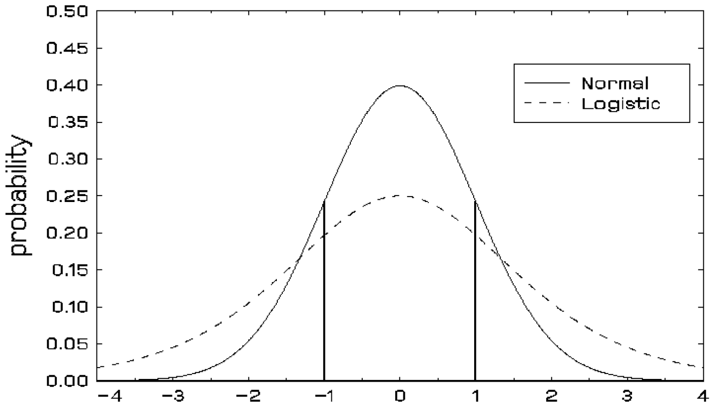

Figura 1.1 – Threshold concept for an ordinal response with three categories.

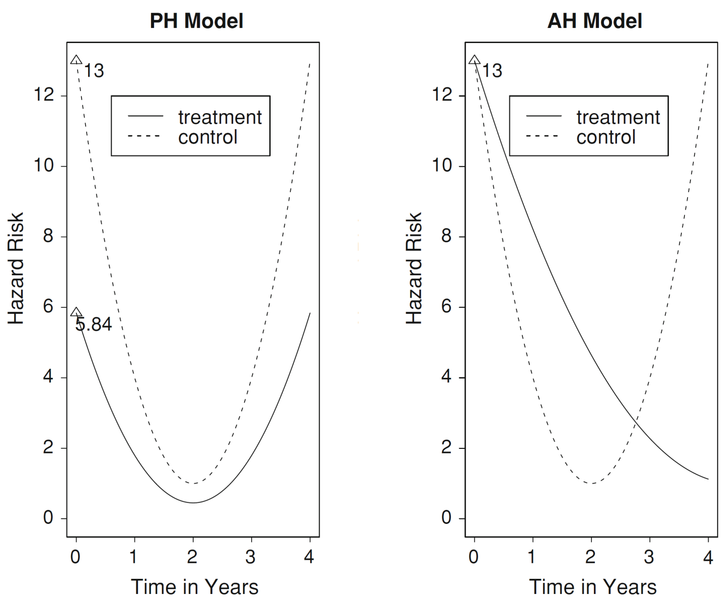

Figura 1.2 – Hazard curves from the PH model and AH model.

2 Self-Modeling Ordinal regression with Time Invariant Covariates

2.1 Introduction

Often, in longitudinal studies involving ordinal outcomes, the curve obtained by smoothing over those outcome points exhibits a specific pattern measured over time. Examples include severity/progression status of disease, Pharmacokinetic data of ordinal nature etc. where response curves share a similar shape e.g. the same number of extremes or inflection points located relatively near some common region. In such circumstances the interest may be to estimate the shape of the response curve itself or the value of a response feature such as the slope at a particular point or the location of a maximum. Often polynomials or other standard regression models are used to fit the data. However, in some cases no standard model is appropriate. Lawton, Sylvestre and Maggio (1972) proposed the self-modeling approach for data of this type. Their method is based on the assumption that all individual's response curve have a common shape and that a particular individual's curve is some simple transformation of the common shape curve. The model which implements this idea is called the shape invariant (SI) model. Estimation and testing of parameters have been discussed by Kneip and Engel (1995).

Often, in longitudinal studies involving ordinal outcomes, the curve obtained by smoothing over those outcome points exhibits a specific pattern measured over time. Examples include severity/progression status of disease, Pharmacokinetic data of ordinal nature etc. where response curves share a similar shape e.g. the same number of extremes or inflection points located relatively near some common region. In such circumstances the interest may be to estimate the shape of the response curve itself or the value of a response feature such as the slope at a particular point or the location of a maximum. Often polynomials or other standard regression models are used to fit the data. However, in some cases no standard model is appropriate. Lawton, Sylvestre and Maggio (1972) proposed proposed the self-modeling approach for data of this type. Their method is based on the assumption that all individual's response curve have a common shape and that a particular individual's curve is some simple transformation of the common shape curve. The model which implements this idea is called the shape invariant (SI) model. Estimation and testing of parameters have been discussed by Kneip and Engel (1995).

The shape-invariant model is a special case of the self-modeling regression method (Lawton, Sylvestre and Maggio, 1972, Altman and Villarreal, 2004). In the context of continuous data, the model can be expressed as

$$Y_{ij} = \alpha_{0i} + \exp(\alpha_{1i})\mu_0(\beta_{0i} + \exp(\beta_{1i})t_{ij}) + \epsilon_{ij} \tag{2.1}$$

where Y_{ij} is the observed response on subject i at time t_{ij}, for $i = 1, ..., N$ and $j = 1, ..., n_i$. Here α_{0i}, α_{1i}, β_{0i} and β_{1i} are unknown parameters which may be functions of observed covariates, ϵ_{ij} is an unobserved error which may be correlated within subject, and μ_0 is a shape function which is common to all subjects.

In this case, the responses are of ordinal nature, like severity of bladder toxicity in prostate cancer patients, the shape invariant model is required to be developed through proportional odds model (see Section 2.2).

In many situations, inference focuses on the parametric effects on one or more summary statistics. Many of the commonly used summary statistics can be expressed as functions of the parameters and functions of the shape. For example, if $\mu_0(t)$ has a critical point at t_0, then the ith curve has a critical point at $(t_0 - \beta_{0i})/\exp(\beta_{1i})$ and the value of the ith curve at the critical point is $\alpha_{0i} + \exp(\alpha_{1i})\ \mu_0(t_0)$. Since the function of the shape does not vary with the treatment and covariate effects, inference about the treatment covariate effects depend only on the parameters. Use of the SI model for inference is very similar to the use of parametric nonlinear mixed models in that the parameters embody all the information about covariate effects. This is quite different from models in which each curve is fitted non-parametrically with its own shape.

Inference for the parameters conditionally on the fitted shape will be comparatively simpler if the conditional model is an ordinary nonlinear mixed effect model. In Section 2.3 we consider the SI model defined for the conditional cumulative probabilities for a category of an outcome.

Penalized spline smoothing has become increasingly popular in recent years. A smooth unknown regression function is estimated by assuming a functional parametric shape constructed via a high dimensional basis function. The basis dimension is chosen to achieve the desired flexibility, while the basis coefficients are penalized to ensure smoothness of the resulting functional estimates. The idea of penalized splines has led to a powerful and applicable smoothing technique.

Maximum likelihood estimation in the nonlinear mixed effects model brings up a substantial challenge because the likelihood of observations cannot typically be expressed in closed form. Several different approximations to the log-likelihood have been proposed. These include the linearization approximation, the LME approximation, and the Laplace's approximation.

These likelihood approximations often perform well if the number of the intraindividual measurements is not small and the variability of random effects is not large, but when some of the individuals have sparse data or the variability of the random effects is large there are considerable errors in approximating the likelihood function via these approximations (Davidian and Giltinan, 1995, Pinheiro and D.M., 1995, Lindstrom and D.M., 1988). This has motivated the use of exact methods such as Monte Carlo methods. In particular, the Monte Carlo EM algorithm, (Wei and Tanner, 1990) in which the E step is approximated using simulated samples from the exact conditional distribution of the random effects given the observed data, has been used for estimation in mixed models. In Section 2.4, parameter estimation using the Monte Carlo method in Newton-Raphson and EM algorithms are introduced. Section 2.5 discuss simulation results of the estimation methods.

In Section 2.6, as an application, we focus on the question of whether the dose level of radiation affects the severity of genito-urinary (bladder) toxicity, which is a side effect of radiation therapy. In particular, we investigate the interaction between the dose effect and follow-up time.

2.2 Ordinal longitudinal logistic regression model

The proportional odds model (see McCullagh, 1980) is essentially based on the logistic regression formulation. It is commonly used for analysis of ordinal data. The proportional odds model characterizes the ordinal responses in L categories ($l = 1, 2, \ldots, L$) in terms of $L-1$ cumulative category comparisons, specifically, $L-1$ cumulative logits. In the proportional odds model, the covariate effects are assumed to be the same across these cumulative logits, or proportional across the cumulative odds. Let $\pi_1(x_i), \ldots, \pi_L(x_i)$ denote the response probabilities at value x_i for a set of explanatory variables, such that:

$$F_l(x_i) = \Pr(Y \leqslant l \,|x_i) = \pi_1(x_i) + \ldots + \pi_l(x_i) \quad for \;\; l = 1, \ldots, L-1,$$

so that cumulative logits are then formed as follows:

$$G_l(x_i) = logit\left[F_l(x_i)\right] = log\left[\frac{F_l(x_i)}{1-F_l(x_i)}\right],$$

where $F_l(x_i)$ is the cumulative probability up to and including category l, the Proportional Odds Model can be expressed as follows:

$$G_l(x_i) = \alpha_l + \beta' x_i.$$

The parameters $\alpha_1, \ldots, \alpha_{L-1}$ are non decreasing in l and are known as the intercepts or "cut-points". The parameter vector β contains the regression coefficients for the covariate vector x_i. Inherent in this model is the proportional odds assumption, which states that the cumulative odds ratio for any two values of the covariates is constant across response categories. Its interpretation is that the odds of being in category less than or equals l is $\exp\left[\beta'\left(x_1 - x_2\right)\right]$ times higher at $x = x_1$ than at $x = x_2$. The model $L-1$ response curves to have the same shape, and therefore we cannot estimate by fitting separate logit models for each cut-point. We must maximize the multinomial likelihood subject to this constraint. The model assumes that effects are the same for each cut-point, $l = 1, ..., L-1$.

Here, let the conditional cumulative probabilities for the L categories of the outcome y_{ij} be called $P_{ijl} = \Pr\left(y_{ij} \leqslant l \mid \upsilon_i, x_{ij}, z_{ij}\right) \ = \sum_{k=1}^{l} p_{ijk}$, where p_{ijk} represents the conditional probability in category k. The logistic Generalized Linear Mixed Model for the conditional cumulative probabilities is given in terms of the cumulative logits as

$$log\left[\frac{P_{ijl}}{1 - P_{ijl}}\right] = \eta_{ijl}, \tag{2.2}$$

where the linear predictor is given by

$$\eta_{ijl} = \tau_l + \omega_{ij}, \tag{2.3}$$

and

$$\omega_{ij} = x_{ij}^{'}\boldsymbol{\beta} + z_{ij}^{'}\upsilon_i. \tag{2.4}$$

We have $L-1$ strictly increasing model thresholds τ_l $(i.e., \tau_1 < \tau_2 < \ldots < \tau_{L-1})$. The x_{ij} is the $(p+1) \times 1$ covariate vector (including the intercept), and z_{ij} is design vector for the r random effects, both vectors being for the jth time-point nested within subject i. Also, $\boldsymbol{\beta}$ is the $(p+1) \times 1$ vector of unknown fixed regression parameters. Let $\upsilon = T\theta$, where $TT^{'} = \Sigma_\upsilon$ is the Cholesky factorization of random-effect variance covariance matrix Σ_υ.

The thresholds allow the cumulative response probabilities to differ by categories. For identification, either the first threshold τ_1 or the model intercept β_0 is typically set to zero. As the regression coefficients $\boldsymbol{\beta}$ do not carry the l subscript, the effects of the regressors do not vary across categories. McCullagh calls this assumption of identical odds ratios across the $L-1$ cutoffs the proportional odds assumption. Because the ordinal model is defined in terms of the cumulative probabilities, the conditional probability of a response in category l is obtained as the difference of two conditional cumulative probabilities:

$$p_{ijl} = \Pr\left(Y_{ij} = l \mid \upsilon_i, x_{ij}, z_{ij}\right) \ = P_{ijl} - P_{ijl-1}. \tag{2.5}$$

Here, $\tau_0 = -\infty$ and $\tau_L = \infty$, and so $P_{ij0} = 0$ and $P_{ijL} = 1$.

2.3 Self-Modeling Regression

In many longitudinal studies, the response to be modeled is a continuous curve measured over time. The volume of air expelled by 18 individuals is measured at 20 time points that vary among individuals. A plot of the raw data values with straight lines connecting the observations in each individual's response sequence (Figure 2.1) indicates a possibly similar shape among the individual's response curves. The data set, as reported, includes one or two zeroes at the beginning of each individual's data sequence.

The Self-Modeling Regression (SEMOR) Model is expressed as

$$Y_{ij} = \pi_i \left\{\mu_0 \left[\kappa_i(t_{ij})\right]\right\} + e_{ij} \tag{2.6}$$

where Y_{ij} is the response for curve i, $i = 1, \ldots, N$, measured at n_i times, t_{ij}. $\pi_i(x)$ is a monotone inverse link transforming the regression function and $\kappa_i(x)$ is a monotone transformation of the time axis. μ_0 is a shape function that is common to all the curves, and e_{ij} are the error terms. For building up the SEMOR model for ordinal data, we will focus on non-parametric modeling of μ_0 and parametric modeling of $\pi_i(x)$ and $\kappa_i(x)$.

We give special attention to Shape Invariant Model and apply the SI model (SIM) for equation (2.4), so

$$\omega_{ij} = \alpha_{0i} + \exp\left(\alpha_{1i}\right)\, \mu_0\left(t^*_{ij}\right), \tag{2.7}$$

where $t^*_{ij} = \beta_{0i} + \exp\left(\beta_{1i}\right) t_{ij}$. Therefore, equations (2.2), (2.3) and (2.7) we have

$$\eta_{ijl} = log\left[\frac{P_{ijl}}{1 - P_{ijl}}\right] = \tau_l - \left[\alpha_{0i} + \exp\left(\alpha_{1i}\right)\, \mu_0\left(t^*_{ij}\right)\right]. \tag{2.8}$$

If one has physical or theoretical justification to pre-specify $\mu_0\left(t^*_{ij}\right)$ parametrically, this is just a special case of nonlinear regression. The semi-parametric SEMOR model allows flexible modeling by estimating $\mu_0\left(t^*_{ij}\right)$ non-parametrically. Several different approaches have been studied earlier for fitting the SIM model. Denoting

$$\boldsymbol{\theta}_i = \left(\alpha_{0i}, \alpha_{1i}, \beta_{0i}, \beta_{1i}\right)', \tag{2.9}$$

we consider a mixed model

$$\boldsymbol{\theta}_{ij} = \ x'_{ij}\boldsymbol{\phi} + \ z'_{ij}\boldsymbol{\psi}_i + \boldsymbol{\varepsilon}_{ij}, \tag{2.10}$$

where z'_{ij} is the design (or covariate) vector associated with the random effect vector $\boldsymbol{\psi}_i$. If μ_0 is a known parametric function, and if we assume that $\boldsymbol{\psi}_i$ and $\boldsymbol{\varepsilon}_i$ are normally distributed, then equation (2.7) is a parametric nonlinear mixed model and the model can be fitted using maximum likelihood using any standard software.

Here, we fit μ_0 using the penalized spline model. Use of the penalized spline method with penalty chosen by generalized maximum likelihood is equivalent to fitting the model

$$\mu_0(t^*_{ij}) = \mathbf{U}\boldsymbol{\gamma} + \mathbf{V}\boldsymbol{\zeta}, \tag{2.11}$$

where $\mu_0(t^*_{ij})$ is the vector of means at the transformed times, $\mathbf{U}$ is a design matrix for a cubic polynomial in t^*_{ij}, $\mathbf{V}$ is a design matrix for cubic in t^*_{ij} which are left-truncated at the knot, $\boldsymbol{\gamma}$ is a vector of unknown parameters and $\boldsymbol{\zeta}$ is normally distributed with zero mean and covariance matrix $\boldsymbol{\Sigma}_\zeta$.

Note that for the analysis of ordinal longitudinal data, one can think of proportional odds model given in equation 2.3 where ω_{ij} is straightway taken as $\boldsymbol{\theta}_{ij}$ considered in equation 2.10. In our simulation section, we compare the performance of estimators for our model and for the proportional odds model even though the models are a bit different. In fact our model involves a larger number of parameters for prediction of individual's response curve which is a transformation of the common shape curve. This situation is automatic in proportional odds model where the predictor is linked through a linear mixed model.

2.3.1 Penalized regression spline for Nonparametric Regression

Consider the following non-parametric regression analysis is a curve fitted to the data set $(\mathbf{t}^*_i, \mathbf{w}_i)$

$$\mathbf{w}_i = \mu_0(\mathbf{t}^*_i) + \boldsymbol{\varepsilon}_i, \tag{2.12}$$

where μ_0 is a smooth function giving the conditional mean of $\mathbf{w}_i$ given $\mathbf{t}^*_i$ and the $\boldsymbol{\varepsilon}_i$ are mutually independent with $N(0, \sigma^2_\varepsilon)$ for $i = 1, 2, \ldots, n$. The smooth function μ_0 can be found as a result of minimization of the residual sum of squares plus a roughness penalty,

$$\sum_{i=1}^{n} (\mathbf{w}_i - \mu_0(\mathbf{t}^*_i))^2 + \lambda \int ({\mu_0}^p(\mathbf{t}^*_i))^2 d\mathbf{t}^*_i, \tag{2.13}$$

where ${\mu_0}^p(\mathbf{t}^*_i)$ is the pth derivative of the function $\mu_0(\mathbf{t}^*_i)$. The fitted curve is a piecewise polynomial of degree $2p - 1$, so that the smoothing parameter λ governs the trade-off between smoothness and goodness of fit. This parameter is often unknown in practice and needs to be estimated from the data. A classical data-driven approach to selecting the smoothing parameter is cross-validation.

A random coefficient linear regression spline model of order p for $\mu_0(t^*_{ij})$ is

$$\mu_0\left(t^*_{ij}\right) = \beta_0 + \beta_1 t^*_{ij} + \cdots + \beta_p {t^*_{ij}}^p + \beta_{p+1}\left(t^*_{ij} - \xi_1\right)^p_+ + \cdots + \beta_{p+K}\left(t^*_{ij} - \xi_K\right)^p_+, \quad (2.14)$$

where the parameters to be estimated are $\boldsymbol{\beta}=(\beta_0, \ldots, \beta_{p+k})$ and $\{\xi_1, \ldots, \xi_k\}$ are K fixed knots with $a \leqslant \xi_1 < \ldots < \xi_k \leqslant b$ and $(x)^p_+ = x^p I_{\{x\geqslant 0\}}$. We can write the equation (2.11) with $p = 3$ as

$$\mu_0\left(t^*_{ij}\right) = \sum_{m=1}^{4} {t^*_{ij}}^{m-1}\gamma_m + \sum_{k=1}^{K}\left(t^*_{ij} - \xi_k\right)^3_+ \zeta_k, \quad (2.15)$$

where $\boldsymbol{\zeta} = [\zeta_1, \ldots, \zeta_k]' \sim N(\mathbf{0}, \sigma^2_\zeta I)$ is independent of $\boldsymbol{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_k]'$. We can combine equations (2.12) and (2.15) in one model

$$\mathbf{w} = \mathbf{U}\boldsymbol{\gamma} + \mathbf{V}\boldsymbol{\zeta} + \boldsymbol{\varepsilon}, \quad (2.16)$$

with $\boldsymbol{\varepsilon} \sim N(\mathbf{0},\ \sigma^2_\varepsilon I)$ and

$$\mathbf{U}=\begin{bmatrix} 1 & t^*_{i1} & t^{*2}_{i1} & t^{*3}_{i1} \\ 1 & t^*_{i2} & t^{*2}_{i2} & t^{*3}_{i2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t^*_{in_i} & t^{*2}_{in_i} & t^{*3}_{in_i} \end{bmatrix}, \mathbf{V}=\begin{bmatrix} \left(t^*_{i1} - \xi_1\right)^3 & \cdots & \left(t^*_{i1} - \xi_k\right)^3 \\ \vdots & \vdots & \vdots \\ \left(t^*_{in_i} - \xi_1\right)^3 & \cdots & \left(t^*_{in_i} - \xi_k\right)^3 \end{bmatrix}. \quad (2.17)$$

Equation (2.16) is nothing but a normal linear mixed model and, for any given σ^2_ζ and σ^2_ε, the estimated best linear unbiased predictor (EBLUP) of $\mathbf{w}$

$$\widehat{\mathbf{w}} = \widehat{\mu_0} = \mathbf{U}\widehat{\boldsymbol{\gamma}} + \mathbf{V}\widehat{\boldsymbol{\zeta}}. \quad (2.18)$$

$E\left(\widehat{\mu_0}\right) = E(\mu_0)$, and that equation (2.18) can be rewritten as

$$\widehat{\mu_0}^* = \mathbf{C}(\mathbf{C}'\mathbf{C} + \lambda^p\mathbf{D})^{-1}\mathbf{C}'\mathbf{w}, \quad (2.19)$$

where $\mathbf{C} = [\mathbf{U}\ \mathbf{V}]$, $\mathbf{D}=Diag(0_{p+1}, I_k)$ and $\lambda^p = \sigma^2_\varepsilon/\sigma^2_\zeta$ for the pth degree of the penalized spline model. The EBLUP estimates (2.19) evaluated at design points are the same as the penalized regression spline solution to equation (2.13). Thus it turns out that the nonparametric smoothing spline regression is equivalent to a mixed-effects model (2.16). We see that the smoothing parameter λ^p is the ratio of the variance components $\sigma^2_\varepsilon/\sigma^2_\zeta$, and that fitting can be done using standard linear mixed effects software. This is the GML method of Wahba (1985).

2.4 Likelihood function and estimation

In view of (2.5) the conditional likelihood can be expressed as

$$\ell(\boldsymbol{\theta} \mid Y_i) = \left\{ \prod_{j=1}^{n_i} \prod_{l=1}^{L} (P_{ijl} - P_{ijl-1})^{I_{y_{ij}}(l)} \right\}, \tag{2.20}$$

where $\boldsymbol{\theta} = (\boldsymbol{\phi}',\ \sigma_\psi^2,\ \sigma_\varepsilon^2,\ \boldsymbol{\tau}')'$ and

$$I_{y_{ij}}(l) = \begin{cases} 1 & if\ y_{ij}=l \\ 0 & if\ y_{ij} \neq l \end{cases}. \tag{2.21}$$

Then the marginal likelihood function is

$$L(\boldsymbol{\theta}, \mathbf{Y}) = \prod_{i=1}^{N} \int \left\{ \prod_{j=1}^{n_i} \prod_{l=1}^{L} (P_{ijl} - P_{ijl-1})^{I_{y_{ij}}(l)} \right\} f(\boldsymbol{\theta}_i \mid \boldsymbol{\psi}_i)\ f\left(\boldsymbol{\psi}_i \mid \sigma_{\psi_i}^2\right)\ d\boldsymbol{\theta}_i\ d\boldsymbol{\psi}_i. \tag{2.22}$$

Since $\boldsymbol{\psi}_i$ and $\boldsymbol{\varepsilon}_i$ have $N_q\left(\mathbf{0}, \boldsymbol{\Sigma}\left(\sigma_\psi^2\right)\right)$ and $N_4\left(\mathbf{0},\ \sigma_\varepsilon^2 I_4\right)$, respectively, in equation (2.10), we write the joint likelihood of conditional cumulative probabilities of ordered outcomes and the random effects $\boldsymbol{\theta}_i$ and ψ_i simplicity in the cumulative logit function as

$$\begin{aligned} l_{ic} = \sum_{j=1}^{n_i} \sum_{l=1}^{L} I_{y_{ij}}(l) \log(P_{ijl} - P_{ijl-1}) + log\Big(\boldsymbol{\phi}\left(\boldsymbol{\theta}_i \mid \mathbf{X}_i\boldsymbol{\phi} + \mathbf{Z}_i\boldsymbol{\psi}_i,\ \sigma_\varepsilon^2 I_4\right)\Big) \\ + log\Big(\boldsymbol{\phi}\left(\boldsymbol{\psi}_i \mid 0,\ \sigma_{\psi_i}^2\right)\Big), \end{aligned} \tag{2.23}$$

where Φ is a probability density function. Therefore, the complete log-likelihood is given by

$$l_c = \sum_{i=1}^{N} l_{ic}. \tag{2.24}$$

2.4.1 An Estimation algorithm

In this subsection we propose an algorithm for parameter estimation in the SI ordinal model. Essentially steps 0-2 are considered for predicting the shape function μ_0 and step 3 provides estimate of the parameters for the cumulative logit model through MCNR and MCEM approaches.

Step 0: Choose initial estimates of $\boldsymbol{\gamma}^{(s)}$, variances $\sigma_\varepsilon^{2(s)}, \sigma_\zeta^{2(s)}$ and $\sigma_\psi^{2(s)}$, random effects $\boldsymbol{\psi}_i^{(s)} \sim N_q\left(\mathbf{0}, \boldsymbol{\Sigma}\left(\sigma_\psi^{2(s)}\right)\right)$, errors $\boldsymbol{\varepsilon}_i^{(s)} \sim N_4\left(\mathbf{0}, \sigma_\varepsilon^{2(s)} I_4\right)$. Set $s = 0$.

Step 1: Compute $\boldsymbol{\theta}_i^{(s)} = \mathbf{X}_i\boldsymbol{\phi}^{(s)} + \mathbf{Z}_i\boldsymbol{\psi}_i^{(s)} + \boldsymbol{\varepsilon}_i^{(s)}$ and extract $\alpha_{0i}^{(s)},\ \alpha_{1i}^{(s)}, \beta_{0i}^{(s)}$ and $\beta_{1i}^{(s)}$

Step 1.1: $t_{ij}^{*(s)} = \beta_{0i}^{(s)} + \exp\left(\beta_{1i}^{(s)}\right) t_{ij}$,

$$\text{Step 1.2: } \mathbf{U}^{(s)} = \begin{bmatrix} 1 & t_{i1}^{*(s)} & t_{i1}^{*2(s)} & t_{i1}^{*3(s)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_{in_i}^{*(s)} & t_{in_i}^{*2(s)} & t_{in_i}^{*3(s)} \end{bmatrix}, \mathbf{V}^{(s)} = \begin{bmatrix} \left(t_{i1}^{*(s)} - \xi_1\right)^3 & \dots & \left(t_{i1}^{*(s)} - \xi_k\right)^3 \\ \vdots & \vdots & \vdots \\ \left(t_{in_i}^{*(s)} - \xi_1\right)^3 & \dots & \left(t_{in_i}^{*(s)} - \xi_k\right)^3 \end{bmatrix},$$

$$\mathbf{C}^{(s)} = \left[\mathbf{U}^{(s)} \quad \mathbf{V}^{(s)}\right].$$

Step 1.3: The pth degree of penalized spline model $\lambda^{p(s)} = \sigma_{\varepsilon}^{2(s)}/\sigma_{\zeta}^{2(s)}$ (See equation (A.8) for more details). $\mathbf{w}^{(s)} = \mathbf{U}^{(s)}\hat{\boldsymbol{\gamma}}^{(s)} + \mathbf{V}^{(s)}\hat{\boldsymbol{\zeta}}^{(s)}$.

Step 1.4: $\hat{\mu_0}^{*(s)} = \mathbf{C}^{(s)}\left(\mathbf{C}^{'(s)}\mathbf{C}^{(s)} + \lambda^{p(s)}\mathbf{D}\right)^{-1}\mathbf{C}^{'(s)}\mathbf{w}^{(s)}$, $D = Diag(0_{p+1}; I_k)$.

Step 2: Using linear mixed model estimation, estimate $\boldsymbol{\gamma}^{(s+1)}$ and $\boldsymbol{\zeta}^{(s+1)}\left(\sigma_{\zeta}^{2(s+1)}\right)$ by fitting

$$\hat{\mu_0}^{*(s)} = \mathbf{U}^{(s)}\boldsymbol{\gamma} + \mathbf{V}^{(s)}\boldsymbol{\zeta}.$$

Step 3: Using nonlinear mixed model estimation, estimate $\boldsymbol{\phi}^{(s+1)}$, $\sigma_{\psi}^{2(s+1)}$, $\sigma_{\varepsilon}^{2(s+1)}$, and $\boldsymbol{\tau}^{(s+1)}$ by fitting the model

$$\begin{aligned} P_{ijl} = \Pr\left(Y_{ij} \leqslant l|\tau_l, \boldsymbol{\theta}_i\right) &= \frac{exp(\eta_{ijl})}{1 + exp(\eta_{ijl})}, \\ \eta_{ijl} = \tau_l - \Big[\alpha_{0i} + \exp(\alpha_{1i})\ \mu_0\Big(\beta_{0i} &+ \exp(\beta_{1i})t_{ij}\Big)\Big], \\ \boldsymbol{\theta}_i = \mathbf{X}_i\boldsymbol{\phi} &+ \mathbf{Z}_i\boldsymbol{\psi}_i + \boldsymbol{\varepsilon}_i, \end{aligned}$$

conditional on $\hat{\mu_0}^{*(s)}$.

Step 4: Check for convergence. If the algorithm has converged, then stop. Otherwise, increase the iteration counter s by one, and return to step 1.

2.4.2 MCEMNR algorithms

In the current setup, the E-M steps are as follows

E-Step

The $(s+1)th$ step computes the conditional expectation of l_{ic}

$$Q\left(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\right) = \sum_{i=1}^{N} E_{\boldsymbol{\theta}_i}\left[l_{ic}|D_i, \boldsymbol{\theta}^{(s)}\right], \tag{2.25}$$

where $D_i = (y_i, \mathbf{Z}_i)$. An alternative is to replace the E step with Monte Carlo approximations

constructed using a sample $l_{ic}^{1}, \ldots, l_{ic}^{(M_i)}$

$$Q\big(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\big) \approx \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} l_{ic}^{(m)}\left(\boldsymbol{\theta} \mid D_i, \boldsymbol{\theta}^{(s)}\right). \tag{2.26}$$

We assume without loss of generality that the number of iterations $M_i = M\ \forall\ i$.

M-Step

The $(s+1)th$ step then finds $\boldsymbol{\theta}^{(s+1)}$ as the maximizer of $Q\big(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\big)$:

$$Q\left(\boldsymbol{\theta}^{(s+1)} \mid \boldsymbol{\theta}^{(s)}\right) \geqslant Q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(s)}\right), \tag{2.27}$$

for all $\boldsymbol{\theta}$ in the parameter space. In principle, the M step is carried out by solving the score equations

$$\frac{\partial}{\partial \boldsymbol{\theta}} Q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(s)}\right) = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \frac{\partial}{\partial \boldsymbol{\theta}} l_{ic}^{(m)}\left(\boldsymbol{\theta} \mid D_i, \boldsymbol{\theta}^{(s)}\right) = 0, \quad for\ \boldsymbol{\theta}. \tag{2.28}$$

The essence of the EM algorithm is that increasing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)})$ forces an increase in the log-likelihood of the observed data. To obtain a satisfactory accuracy, the MC sample size M needs to be a large number. For example, to obtain two decimal digits of accuracy, $M \geqslant 10\ 000$ is required. Solution for θ in (2.28) can be obtained by the NR procedure. To solve for $\boldsymbol{\theta}$, we proceed through NR steps described below.

NR-Step

Let $\boldsymbol{\theta}^{(s)} = (\boldsymbol{\phi}^{(s)},\ \sigma_{\psi}^{2(s)},\ \sigma_{\varepsilon}^{2(s)},\ \boldsymbol{\tau}^{(s)})$, the Newton Raphson procedure is applied to estimate $\boldsymbol{\theta}$ at the $(s+1)$ iteration by using equation (2.28) as follows.

$$\widehat{\boldsymbol{\theta}}^{(s+1)} = \widehat{\boldsymbol{\theta}}^{(s)} + [V^{-1}\hat{S}(\boldsymbol{\theta})]_{|\widehat{\boldsymbol{\theta}}^{(s)}}\ , \tag{2.29}$$

where

$$\hat{S}\left(\boldsymbol{\theta}\right) = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left[\frac{\partial l_{ic}^{(m)}}{\partial \boldsymbol{\theta}}\right]_{|\widehat{\boldsymbol{\theta}}^{(s)}}, \tag{2.30}$$

and

$$V = \frac{\partial \hat{S}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}_{|\widehat{\boldsymbol{\theta}}^{(s)}} = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \frac{\partial}{\partial \boldsymbol{\theta}^T} \left[\frac{\partial l_{ic}^{(m)}}{\partial \boldsymbol{\theta}} \right]_{|\widehat{\boldsymbol{\theta}}^{(s)}} . \tag{2.31}$$

M is the number of iterations in the Monte Carlo method. This process is repeated until $\left\| \widehat{\boldsymbol{\theta}}^{(s+1)} - \widehat{\boldsymbol{\theta}}^{(s)} \right\| \leqslant \epsilon \downarrow 0$. See Section A.1 for more details.

2.5 Simulation

In this section, we use simulate of data sets to compare the performances of the proportional odds model and our model by the MCNREM algorithm. We assume the number of categories $L = 5$ and for $i = 1, .., N$, $N = 50, 100, 200$ and for $j = 1, .., n$, $n = 10, 25$. The number of iterations in the Monte Carlo is 10000. The true values and the estimates of all parameters of the proportional odds model and our model by the MCNREM algorithm are displayed in Tables 2.1 and 2.2 respectively. Note that both for the proportional odds model and the proposed model the true value of the regression parameter vector $\boldsymbol{\phi}$ has been taken to be a null vector. Also the dispersion matrix of the random vector $\boldsymbol{\psi_i}$ is assumed to be $\sigma_{\psi}^2 I_4$ for both models with the true value of $\sigma_{\psi}^2 = 0.1$.

Since we do not have a closed form expression for the maximizer $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)})$, we use Newton-Raphson iterations to find the maximizer in the EM algorithm. The standard errors (SE) based on generated random samples of different sizes are also computed for each estimator. All the computations are carried out in MATLAB.

Tables 2.1 and 2.2 give us an idea about the performance of the estimators under Proportional odds model and the self generating model of interest . In both tables, we observe the variations in the estimates when $n = 10$ is fixed and the value of N is 50, 100, 200. The situation is pretty similar when $n = 25$ and N varies. From Table 2.2, it is evident that performance of the estimates is quite satisfactory. The iteration convergence figures for different parameters under both models have been appended at the end.

2.6 An Application to the Prostate Cancer Data

The primary motivation for considering the self generating model comes up while analyzing the data recorded at some hospitals in Chicago [1]. The objective is to see to what extent the different doses of radiation affects the prostate cancer patients. Two hundred and forty three patients are treated with radiation, under one of the three dose levels of radiation (D = 1,2,3 for weak, medium, strong) applied randomly to each patient. The stage of prostate cancer is associated with the spread and severity of the cancer (S = 1,2,3 for minor, medium, severe). The initial investigation has been carried out in two hospitals (H

[1] www.biometrics.tibs.org/datasets/980326.txt

= 0,1) where the patients (N = 243) have been followed up over 6 years. The time points of measurement differ from patient to patient. In each of the two hospitals, a physician assesses the severity of genito-urinary (bladder) toxicity (gu = 0 for no symptoms, gu = 1 for pain/local bleeding, does not require intervention, gu = 2 for bleeding lesion, requires minor intervention, no hospitalization, and gu = 3 for serious lesion, requires hospitalization). The detailed analysis is based on our proposed model where we consider the severity of toxicity as a response variable.

For choosing the best model in this example, we use the deviance which is the difference between the log-likelihood of the fitted model and the maximum possible log-likelihood. Table 2.3 gives the values of the deviance. We can note that increasing the number of the predictors in the model, the deviance decreases.

Table 2.4 shows the estimates of intercepts and coefficients for the fourth model in Table 2.3. The p-value of 0.0065 and 0.0120 for age and hospital respectively indicate that these factors are significant on the odds of the genito-urinary toxicity of a patient being less than or equal to a certain value versus being greater than that value. The p-values of 0.0520 and 0.6814 for dose and stage of a patient indicate that these factors are not significant.

The self generating model with respect to the fourth model is

$$log\left(\frac{P(gu_{ijl} \leqslant 0)}{P(gu_{ijl} > 0)}\right) = 3.9847 - \left[\hat{\alpha}_{0i} + \exp\left(\hat{\alpha}_{1i}\right)\,\hat{\mu}_0\left(t^*_{ij}\right)\right],$$

$$log\left(\frac{P(gu_{ijl} \leqslant 1)}{P(gu_{ijl} > 1)}\right) = 5.7957 - \left[\hat{\alpha}_{0i} + \exp\left(\hat{\alpha}_{1i}\right)\,\hat{\mu}_0\left(t^*_{ij}\right)\right],$$

$$log\left(\frac{P(gu_{ijl} \leqslant 2)}{P(gu_{ijl} > 2)}\right) = 7.0340 - \left[\hat{\alpha}_{0i} + \exp\left(\hat{\alpha}_{1i}\right)\,\hat{\mu}_0\left(t^*_{ij}\right)\right].$$

$$\hat{\boldsymbol{\theta}}_i = \left(\hat{\alpha}_{0i}, \hat{\alpha}_{1i}, \hat{\beta}_{0i}, \hat{\beta}_{1i}\right)' = -0.1682D_i' - 0.0221A_i' - 0.3421H_i' + 0.0312S_i',$$

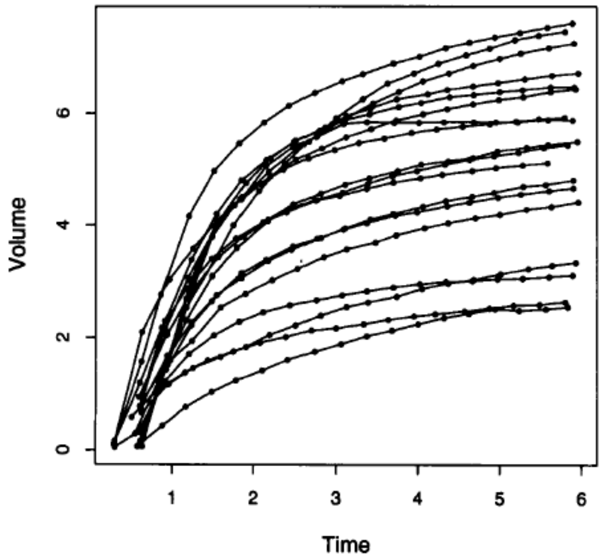

Figura 2.1 – Spirometer data. Volume of air expelled versus time for 18 individuals.

Tabela 2.1 – The estimates of all parameters of the proportional odds model by MCEMNR (SE in parenthesis)

N		50		100		200	
n		10	25	10	25	10	25
Parameter	*True value*						
τ_1	$-log3$	0.2311	0.2540	0.2476	0.2675	0.2616	0.2857
		(0.2562)	(0.2044)	(0.1993)	(0.1592)	(0.1481)	(0.1071)
τ_2	0	0.2692	0.2859	0.2810	0.2988	0.2932	0.3160
		(0.2508)	(0.2027)	(0.1960)	(0.1549)	(0.1422)	(0.1055)
τ_3	$log3$	0.2984	0.3193	0.3143	0.3345	0.3203	0.3426
		(0.2593)	(0.2133)	(0.1967)	(0.1584)	(0.1470)	(0.1037)
$\boldsymbol{\phi}$	$\mathbf{0}$	-0.1412	-0.1207	-0.1125	-0.0991	-0.0832	-0.0683
		(0.0705	(0.0596)	(0.0410)	(0.0280)	(0.0217)	(0.0121)
		-0.1448	-0.1213	-0.1111	-0.1008	-0.0864	-0.0720
		(0.0945)	(0.0760)	(0.0586)	(0.0364)	(0.0331)	(0.0205)
		-0.1406	-0.1290	-0.1181	-0.0952	-0.0811	-0.0682
		(0.0763)	(0.0588)	(0.0450)	(0.0271)	(0.0238)	(0.0110)
		-0.1486	-0.1223	-0.1144	-0.1017	-0.0848	-0.0704
		(0.0835)	(0.0696)	(0.0548)	(0.0357)	(0.0331)	(0.0209)
$\boldsymbol{\sigma}^2_\psi$	0.1	0.0573	0.0668	0.0761	0.0805	0.0841	0.0948
		(0.0130)	(0.0116)	(0.0109)	(0.0088)	(0.0080)	(0.0063)
		0.0602	0.0699	0.0780	0.0811	0.0871	0.0961
		(0.0141)	(0.0121)	(0.0111)	(0.0093)	(0.0087)	(0.0069)
		0.0611	0.0692	0.0730	0.0813	0.0825	0.0955
		(0.0145)	(0.0121)	(0.0113)	(0.0097)	(0.0086)	(0.0070)
		0.0585	0.0670	0.0758	0.0803	0.0863	0.0932
		(0.0132)	(0.0118)	(0.0101)	(0.0089)	(0.0083)	(0.0061)
σ^2_ε	0.1	0.0918	0.0930	0.0947	0.0961	0.0966	0.0980
		(0.0052)	(0.0044)	(0.0039)	(0.0031)	(0.0029)	(0.0022)

Tabela 2.2 – The estimates of all parameters of the Self generating ordinal model by MCEMNR (SE in parenthesis)

N		50		100		200	
n		10	25	10	25	10	25
Parameter	*True value*						
τ_1	$-log3$	-0.7293	-0.7280	-0.7273	-0.7258	-0.7253	-0.7239
		(0.1763)	(0.0973)	(0.1241)	(0.0618)	(0.0758)	(0.0330)
τ_2	0	0.1275	0.1269	0.1263	0.1257	0.1250	0.1244
		(0.1865)	(0.1204)	(0.1440)	(0.0853)	(0.0946)	(0.0439)
τ_3	$log3$	1.2122	1.1951	1.1863	1.1758	1.1689	1.1580
		(0.1967)	(0.1372)	(0.1678)	(0.0995)	(0.1299)	(0.0751)
$\boldsymbol{\phi}$	$\mathbf{0}$	-0.0112	-0.0097	-0.0090	-0.0074	-0.0070	-0.0054
		(0.0221)	(0.0207)	(0.0150)	(0.0136)	(0.0087)	(0.0075)
		-0.0155	-0.0134	-0.0129	-0.0105	-0.0099	-0.0080
		(0.0249)	(0.0225)	(0.0169)	(0.0157)	(0.0098)	(0.0089)
		-0.0141	-0.0122	-0.0116	-0.0098	-0.0059	-0.0051
		(0.0138)	(0.0130)	(0.0097)	(0.0089)	(0.0059)	(0.0051)
		-0.0109	-0.0095	-0.0087	-0.0075	-0.0073	-0.0060
		(0.0228)	(0.0221)	(0.0156)	(0.0139)	(0.0111)	(0.0099)
$\boldsymbol{\sigma}^2_{\psi}$	0.1	0.0986	0.0990	0.0991	0.0995	0.0997	0.0999
		(0.0059)	(0.0054)	(0.0055)	(0.0049)	(0.0048)	(0.0044)
		0.0988	0.0989	0.0992	0.0996	0.0995	0.0998
		(0.0060)	(0.0055)	(0.0053)	(0.0050)	(0.0051)	(0.0046)
		0.0983	0.0991	0.0990	0.0993	0.0995	0.0997
		(0.0062)	(0.0058)	(0.0059)	(0.0055)	(0.0053)	(0.0049)
		0.0985	0.0990	0.0994	0.0996	0.0997	0.0999
		(0.0065)	(0.0062)	(0.0059)	(0.0056)	(0.0055)	(0.0052)
σ^2_{ε}	0.1	0.1128	0.1125	0.1115	0.1112	0.1101	0.1097
		(0.0038)	(0.0036)	(0.0027)	(0.0024)	(0.0018)	(0.0015)
$\boldsymbol{\gamma}$	0	-0.0130	-0.0112	-0.0115	-0.0091	-0.0095	-0.0081
		(0.1129)	(0.0935)	(0.0942)	(0.0785)	(0.0790)	(0.0601)
		-0.0163	-0.0147	-0.0149	-0.0130	-0.0135	-0.0113
		(0.1387)	(0.1159)	(0.1163)	(0.0927)	(0.0930)	(0.0697)
		-0.0143	-0.0121	-0.0127	-0.0109	-0.0111	-0.093
		(0.1336)	(0.1181)	(0.1190)	(0.1049)	(0.1057)	(0.0864)
		-0.0150	-0.0123	-0.0128	-0.0105	-0.0107	-0.0084
		(0.1419)	(0.1204)	(0.1220)	(0.1007)	(0.1010)	(0.0801)

Tabela 2.2 – The estimates of all parameters of the Self generating ordinal model by MCEMNR (SE in parenthesis)

N		50		100		200	
n		10	25	10	25	10	25
Parameter	*True value*						
$\boldsymbol{\sigma}^2_\zeta$	0.1	0.0982	0.0984	0.0984	0.0987	0.0990	0.0993
		(0.1409)	(0.1405)	(0.1406)	(0.1400)	(0.1395)	(0.1391)
		0.0981	0.0983	0.0984	0.0986	0.0991	0.0994
		(0.1412)	(0.1409)	(0.1402)	(0.1397)	(0.1397)	(0.1392)
		0.0980	0.0982	0.0981	0.0985	0.0995	0.0998
		(01410)	(0.1404)	(0.1405)	(0.1400)	(0.1398)	(0.1390)
		0.0978	0.0985	0.0982	0.0990	0.0993	0.0999
		(0.1424)	(0.1408)	(0.1409)	(0.1402)	(0.1394)	(0.1387)
		0.0973	0.0977	0.0975	0.0983	0.0996	0.0999
		(0.1421)	(0.1412)	(0.1414)	(0.1407)	(0.1390)	(0.1386)
$\sigma^2_{\varepsilon_p}$	0.1	0.0930	0.0941	0.0937	0.0950	0.0946	0.0960
		(0.0299)	(0.0284)	(0.0286)	(0.0270)	(0.0275)	(0.0259)

Tabela 2.3 – Values of the deviance for the best subset of each size

Subset size	Predictors	Dev
1	*Dose*	2210.1
2	*Dose, Age*	2203.0
3	*Dose, Age, Hosp*	2195.8
4	*Dose, Age, Hosp, Stage*	2195.6

Tabela 2.4 – Estimates of intercepts and coefficients of the fourth model

Variable	Estimates	Std. Error	P-value
τ_1	3.9847	0.6714	0.0000
τ_2	5.7957	0.6820	0.0000
τ_3	7.0340	0.7147	0.0000
Dose	−0.1682	0.0847	0.0520
Age	−0.0221	0.0085	0.0065
Hosp	−0.3421	0.1468	0.0120
Stage	0.0312	0.1147	0.6814

3 Self-Modeling ordinal regression under right censoring

3.1 Introduction

We discuses the Self-Modeling ordinal regression under right censoring. Self-Modeling regressions form a class of models for functional data observed for many individuals. We consider the Self-Modeling Regression (SEMOR) Model introduced by Lawton, Sylvestre and Maggio (1972)

$$y_{ij} = \pi_i \left\{ \mu_0 \left[\kappa_i(t_{ij}) \right] \right\} + e_{ij}, \tag{3.1}$$

where y_{ij} is the response for curve i, $i = 1, \ldots, N$, measured at n_i times, t_{ij}. $\pi_i(x)$ is a monotone inverse link transforming the regression function and $\kappa_i(x)$ is a monotone transformation of the time axis. μ_0 is a shape function that is common to all the curves, and e_{ij} are errors. This chapter will focus on non-parametric modeling of μ_0 and parametric modeling of $\pi_i(x)$ and $\kappa_i(x)$ with known correlation structure for e_{ij}.

Since the proportional odds assumption described by McCullagh (1980) is a common choice for the analysis of ordinal data, many of the mixed models for ordinal data are generalizations of this model. The proportional odds model characterizes the ordinal responses in L categories ($l = 1, 2, \ldots, L$) in terms of $L - 1$ cumulative category comparisons, specifically, $L - 1$ cumulative logits. In the proportional odds model, the covariate effects are assumed to be the same across these cumulative logits, or proportional across the cumulative odds. We denote the conditional cumulative probabilities for the L categories of the outcome y_{ij} as $P_{ijl} = \Pr(y_{ij} \leqslant l | x_{ij}, z_{ij}) \ = \sum_{k=1}^{l} p_{ijk}$, where p_{ijk} represents the conditional probability of response in category k. The logistic Generalized Linear Mixed Models for the conditional cumulative probabilities is given in terms of the cumulative logits as

$$P_{ijl} = \frac{\exp(\eta_{ijl})}{1 + \exp(\eta_{ijl})}, \qquad l = 1, \ldots, L - 1, \tag{3.2}$$

where the linear predictor is

$$\eta_{ijl} = \tau_l + \omega_{ij}, \tag{3.3}$$

and

$$\omega_{ij} = x_{ij}^{'}\boldsymbol{\beta} + z_{ij}^{'}\upsilon_i. \tag{3.4}$$

Take $L-1$ strictly increasing model thresholds τ_l $(i.e., \tau_1 < \tau_2 < \ldots < \tau_{L-1})$. $\boldsymbol{x}_{ij}$ the $(p+1)\times 1$ covariate vector (including the intercept), and $\boldsymbol{z}_{ij}$ the design vector for the r random effects, both vectors being for the jth time-point nested within subject i. Also, $\boldsymbol{\beta}$ let be $(p+1)\times 1$ vector of unknown fixed regression parameters. Let $\upsilon = T\theta$, where $TT^{'} = \Sigma_\upsilon$ is the Cholesky factorization of random-effect variance covariance matrix Σ_υ. Because the ordinal model is defined in terms of the cumulative probabilities, the conditional probability of a response in category l is obtained as the difference of two conditional cumulative probabilities:

$$p_{ijl} = \Pr\left(Y_{ij} = l | x_{ij}, z_{ij}\right) = P_{ijl} - P_{ijl-1},$$

where, $\tau_0 = -\infty$, $\tau_L = \infty$ so that $P_{ij0} = 0$ and $P_{ijL} = 1$.

In Section 3.2, we will consider the Self-Modeling defined for the conditional cumulative probabilities for a category of an outcome. Therefore, unlike the model 3.1, in our model there is no relation between the observed response and parameters directly. In fact, unlike the non-linear mixed model, it is rather difficult to find the relation between the observed response and parameters directly. In Section 3.3, parameter estimation using the Monte Carlo method in Newton-Raphson and EM algorithms are introduced. Section 3.4 discuss simulation results of the estimation methods.

In a post-surgical recovery study, the status of recovery is assessed for patients who were given different dose of anaesthetic. The ordinal responses are recorded longitudinally along with the recovery stage of a patient. Differences among the patients due to time-invariant covariates are captured by the parameters. Since patients having a common surgery usually exhibit a similar pattern, it is natural to build up a nonlinear model that is shape invariant. In Section 3.5, we focus on the question of whether the dose of anesthesia affects the post-surgical recovery. In particular, we investigate the interaction between the dose effect and follow-up time.

3.2 Self-Modeling Ordinal model under right censoring

The Shape-Invariant (SI) model is a special case of the self-modeling regression method (Altman and Villarreal, 2004),

$$Y_{ij} = \alpha_{0i} + \exp(\alpha_{1i})\mu_0(\beta_{0i} + \exp(\beta_{1i})t_{ij}) + \epsilon_{ij} \tag{3.5}$$

where Y_{ij} is the observed response on subject i at time t_{ij}. We apply the SI model for equation (3.4), so

$$\omega_{ij} = \alpha_{0i} + \exp\left(\alpha_{1i}\right)\, \mu_0\left(t^*_{ij}\right), \tag{3.6}$$

where $t^*_{ij} = \beta_{0i} + \exp\left(\beta_{1i}\right) t_{ij}$. From equations (3.2), (3.3) and (3.6) we have

$$\eta_{ijl} = log\left[\frac{P_{ijl}}{1-P_{ijl}}\right] = \tau_l + \left[\alpha_{0i} + \exp\left(\alpha_{1i}\right)\, \mu_0\left(t^*_{ij}\right)\right]. \tag{3.7}$$

Lawton, Sylvestre and Maggio (1972), Kneip and Gasser (1988) and Kneip and Engel (1995) considered

$$\boldsymbol{\theta}_i = \left(\alpha_{0i}, \alpha_{1i}, \beta_{0i}, \beta_{1i}\right)', \tag{3.8}$$

while we consider a mixed model

$$\boldsymbol{\theta}_i = \mathbf{X}_i\boldsymbol{\phi} + \mathbf{Z}_i\boldsymbol{\psi}_i + \boldsymbol{\varepsilon}_i, \tag{3.9}$$

where $\boldsymbol{Z}_i$ is the design (or covariate) matrix for the random effect vector $\boldsymbol{\psi}_i$. If μ_0 is a known parametric function, and if we assume that $\boldsymbol{\psi}_i$, $\boldsymbol{\varepsilon}_i$ and ϵ_{ij} are normally distributed, then equation (3.6) is a parametric non-linear mixed model and the model can be fitted using maximum likelihood using standard software (e.g., Lindstrom and D.M., 1988).

Here, we fit μ_0 using the penalized spline model of Ruppert and Carroll (1997). Use of the penalized spline method with penalty chosen by generalized maximum likelihood (Wahba, 1985) is equivalent to fitting the model

$$\mu_0(t^*_{ij}) = \mathbf{U}\boldsymbol{\gamma} + \mathbf{V}\boldsymbol{\zeta}, \tag{3.10}$$

where $\mu_0(t^*_{ij})$ is the vector of means at the transformed times, $\mathbf{U}$ is a design matrix for a cubic polynomial in t^*_{ij}, $\mathbf{V}$ is a design matrix for cubics in t^*_{ij} which are left-truncated at the knot, $\boldsymbol{\gamma}$ is a vector of unknown parameters and $\boldsymbol{\zeta}$ is normally distributed with zero mean and covariance matrix $\boldsymbol{\Sigma}_{\zeta}$.

The result of a non-parametric regression analysis is a curve fitted to a set of data $(\mathbf{t}^*_i, \mathbf{w}_i)$

$$\mathbf{w}_i = f_0(\mathbf{t}^*_i) + \boldsymbol{\varepsilon}_i, \tag{3.11}$$

where f_0 is a smooth function giving the conditional mean of $\mathbf{w}_i$ given $\mathbf{t}_i^*$ and the $\boldsymbol{\varepsilon}_i$ are mutually independent with $N(0, \sigma_{\varepsilon}^2)$ for $i = 1, 2, \ldots, n$. The smooth function f_0 can be found as result of minimization of the residual sum of squares plus a roughness penalty,

$$\sum_{i=1}^{n} (\mathbf{w}_i - \mu_0(\mathbf{t}_i^*))^2 + \lambda \int (f_0^p(\mathbf{t}_i^*))^2 d\mathbf{t}_i^*, \tag{3.12}$$

where $f_0^p(\mathbf{t}_i^*)$ is the pth derivative of the function $f_0(\mathbf{t}_i^*)$. The resultant curve fitted to the data is a piecewise polynomial of degree $2p-1$. The smoothing parameter λ governs the trade-off between smoothness and goodness of fit. This parameter is often unknown in practice and needs to be estimated from the data. A classical data-driven approach to selecting the smoothing parameter is cross-validation. A random coefficient linear regression spline model of order p for $\mu_0(t_{ij}^*)$ by Ruppert and Carroll (1997)., is

$$\mu_0\left(t_{ij}^*\right) = \beta_0 + \beta_1 t_{ij}^* + \cdots + \beta_p {t_{ij}^*}^p + \beta_{p+1}\left(t_{ij}^* - \xi_1\right)_+^p + \cdots + \beta_{p+K}\left(t_{ij}^* - \xi_K\right)_+^p, \tag{3.13}$$

where the parameters to be estimated are $\boldsymbol{\beta}$=$(\beta_0, \ldots, \beta_{p+k})$ and $\{\xi_1, \ldots, \xi_k\}$ are K fixed knots with $a \leqslant \xi_1 < \ldots < \xi_k \leqslant b$ and $(x)_+^p = x^p I_{\{x \geqslant 0\}}$. We can write equation (3.8) with $p = 3$ as

$$\mu_0\left(t_{ij}^*\right) = \sum_{m=1}^{4} {t_{ij}^*}^{m-1} \gamma_m + \sum_{k=1}^{K} \left(t_{ij}^* - \xi_k\right)_+^3 \zeta_k, \tag{3.14}$$

where $\boldsymbol{\zeta} = [\zeta_1, \ldots, \zeta_k]' \sim N(\mathbf{0}, \sigma_\zeta^2 I)$ is independent of $\boldsymbol{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_k]'$. We can combine equations (3.11) and (3.14) in one model

$$\mathbf{w} = \mathbf{U}\boldsymbol{\gamma} + \mathbf{V}\boldsymbol{\zeta} + \boldsymbol{\varepsilon}, \tag{3.15}$$

with $\boldsymbol{\varepsilon} \sim N(\mathbf{0},\ \sigma_\varepsilon^2 I)$ and

$$\mathbf{U}=\begin{bmatrix} 1 & t_{i1}^* & t_{i1}^{*2} & t_{i1}^{*3} \\ 1 & t_{i2}^* & t_{i2}^{*2} & t_{i2}^{*3} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_{in_i}^* & t_{in_i}^{*2} & t_{in_i}^{*3} \end{bmatrix}, \mathbf{V}=\begin{bmatrix} (t_{i1}^* - \xi_1)^3 & \ldots & (t_{i1}^* - \xi_k)^3 \\ \vdots & \vdots & \vdots \\ (t_{in_i}^* - \xi_1)^3 & \cdots & (t_{in_i}^* - \xi_k)^3 \end{bmatrix}. \tag{3.16}$$

Equation (3.15) is nothing but a normal linear mixed model and, for any given σ_ζ^2 and σ_ε^2, the estimated best linear unbiased predictor (EBLUP) of $\mathbf{w}$ by Robinson (1991) is given by

$$\widehat{\mathbf{w}} = \widehat{\mu_0} = \mathbf{U}\widehat{\boldsymbol{\gamma}} + \mathbf{V}\widehat{\boldsymbol{\zeta}}. \tag{3.17}$$

Note that $E(\widehat{\mu_0}) = E(\mu_0)$. Equation (3.17) can be rewritten by McCulloch and Searle (2001) as

$$\widehat{\mu_0}^* = \mathbf{C}(\mathbf{C}'\mathbf{C} + \lambda^p \mathbf{D})^{-1}\mathbf{C}'\mathbf{w}, \tag{3.18}$$

where $\mathbf{C} = [\mathbf{U}\ \mathbf{V}]$, $\mathbf{D}{=}Diag(0_{p+1}, I_k)$ and $\lambda^p = \sigma_\varepsilon^2/\sigma_\zeta^2$ for the pth degree of penalized spline model. It has been shown in Wang (1998) and Brumback and J.A. (1998) that the EBLUP estimates given by (3.18) evaluated at design points are the same as the penalized regression spline solution to equation (3.12). Thus it turns out that the non-parametric smoothing spline regression is equivalent to a mixed-effects model (3.15). We see that the smoothing parameter λ^p is the ratio of the variance components $\sigma_\varepsilon^2/\sigma_\zeta^2$, and that fitting can be done using standard linear mixed effects software. This is the GML method of Wahba (1985).

3.3 Estimation method

Let Y_i , denote the vector of ordinal responses from subject i (for the n_i repeated observations nested within). The conditional likelihood for the observed responses can be expressed as

$$\ell(\boldsymbol{\theta}|\mathbf{Y}_i) = \prod_{j=1}^{n_i}\prod_{l=1}^{L}(P_{ijl} - P_{ijl-1})^{I_{y_{ij}}(l)}, \tag{3.19}$$

where $\boldsymbol{\theta} = (\boldsymbol{\phi}',\ \sigma_\psi^2,\ \sigma_\varepsilon^2,\ \boldsymbol{\tau}')'$ and

$$I_{y_{ij}} = \begin{cases} 1 & \text{if } y_{ij} = l \\ 0 & \text{if } y_{ij} \neq l. \end{cases}$$

For the ordinal representation of the survival model, where right-censoring is present, the above likelihood is generalized to

$$\ell(\mathbf{Y}_i|\boldsymbol{\theta}) = \prod_{j=1}^{n_i}\prod_{l=1}^{L}\left[(P_{ijl} - P_{ijl-1})^{c_{ij}}(1 - P_{ijl})^{1-c_{ij}}\right]^{I_{y_{ij}}(l)}, \tag{3.20}$$

where

$$c_{ij} = \begin{cases} 1 & \text{if } y_{ij} : Uncensored \\ 0 & \text{if } y_{ij} : Censored. \end{cases}$$

With right-censoring, because there is essentially one additional response category (for those censored at the last category L), given by $\tau_{L+1} = \infty$ and $P_{ijL+1} = 1$. Then the marginal likelihood function under right-censoring is

$$L(\boldsymbol{\theta}, \mathbf{Y}) = \prod_{i=1}^{N} \int \left\{ \prod_{j=1}^{n_i} \prod_{l=1}^{L} \left[(P_{ijl} - P_{ijl-1})^{c_{ij}} (1 - P_{ijl})^{1-c_{ij}} \right]^{I_{y_{ij}}(l)} f(\boldsymbol{\theta}_i | \boldsymbol{\psi}_i) \, f\left(\boldsymbol{\psi}_i | \sigma^2_{\psi_i}\right) \right\} d\boldsymbol{\theta}_i \, d\boldsymbol{\psi}_i. \quad (3.21)$$

Since $\boldsymbol{\psi}_i$ and $\boldsymbol{\varepsilon}_i$ have $N_q\Big(\mathbf{0}, \boldsymbol{\Sigma}\left(\sigma^2_{\psi_i}\right)\Big)$ and $N_4\big(\mathbf{0},\ \sigma^2_{\varepsilon_i} I_4\big)$ distributions, respectively, in equation (3.9), we write the joint likelihood of the conditional cumulative probabilities of ordered outcomes and the random effects $\boldsymbol{\theta}_i$ and ψ_i in the cumulative logit function as

$$l_{ic} = \sum_{j=1}^{n_i} \sum_{l=1}^{L} I_{y_{ij}}(l) \Big[c_{ij} \log(P_{ijl} - P_{ijl-1}) + (1 - c_{ij}) \log(1 - P_{ijl}) \Big] + \log\Big(\boldsymbol{\Phi}\left(\boldsymbol{\theta}_i | \mathbf{X}_i \boldsymbol{\phi} + \mathbf{Z}_i \boldsymbol{\psi}_i,\ \sigma^2_{\varepsilon_i} I_4\right) \Big) + \log\Big(\boldsymbol{\Phi}\left(\boldsymbol{\psi}_i | 0,\ \sigma^2_{\psi_i}\right) \Big), \quad (3.22)$$

where Φ is some probability density function. Therefore, the complete log-likelihood is

$$l_c = \sum_{i=1}^{N} l_{ic}. \quad (3.23)$$

3.3.1 The estimation algorithm

In this section we propose an algorithm for the parameters estimation in the SI ordinal model. Essentially steps 0-2 are considered for predicting the shape function μ_0 and step 3 provides estimate of the parameters for the cumulative logit model through MCNR and MCEM approaches.

Step 0: Choose initial estimates of $\boldsymbol{\tau}^{(s)}$, $\boldsymbol{\gamma}^{(s)}$, variances $\sigma^{2(s)}_{\varepsilon}, \sigma^{2(s)}_{\zeta}$ and $\sigma^{2(s)}_{\psi}$, random effects $\boldsymbol{\psi}^{(s)}_i \sim N_q\left(\mathbf{0}, \boldsymbol{\Sigma}\left(\sigma^{2(s)}_{\psi}\right)\right)$, errors $\boldsymbol{\varepsilon}^{(s)}_i \sim N_4\left(\mathbf{0}, \sigma^{2(s)}_{\varepsilon} I_4\right)$. Set $s = 0$.

Step 1: Compute $\boldsymbol{\theta}^{(s)}_i = \mathbf{X}_i \boldsymbol{\phi}^{(s)} + \mathbf{Z}_i \boldsymbol{\psi}^{(s)}_i + \boldsymbol{\varepsilon}^{(s)}_i$ and extract $\alpha^{(s)}_{0i}$, $\alpha^{(s)}_{1i}$, $\beta^{(s)}_{0i}$ and $\beta^{(s)}_{1i}$

Step 1.1: $t^{*(s)}_{ij} = \beta^{(s)}_{0i} + \exp\left(\beta^{(s)}_{1i}\right) t_{ij}$,

Step 1.2: $\mathbf{U}^{(s)} = \begin{bmatrix} 1 & t^{*(s)}_{i1} & t^{*2(s)}_{i1} & t^{*3(s)}_{i1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t^{*(s)}_{in_i} & t^{*2(s)}_{in_i} & t^{*3(s)}_{in_i} \end{bmatrix}$, $\mathbf{V}^{(s)} = \begin{bmatrix} \left(t^{*(s)}_{i1} - \xi_1\right)^3 & \dots & \left(t^{*(s)}_{i1} - \xi_k\right)^3 \\ \vdots & \vdots & \vdots \\ \left(t^{*(s)}_{in_i} - \xi_1\right)^3 & \dots & \left(t^{*(s)}_{in_i} - \xi_k\right)^3 \end{bmatrix}$,

$$\mathbf{C}^{(s)} = \left[\mathbf{U}^{(s)} \quad \mathbf{V}^{(s)}\right].$$

Step 1.3: The pth degree of penalized spline model $\lambda^{p(s)} = \sigma_{\varepsilon}^{2(s)}/\sigma_{\zeta}^{2(s)}$. $\mathbf{w}^{(s)} = \mathbf{U}^{(s)}\widehat{\boldsymbol{\gamma}}^{(s)} + \mathbf{V}^{(s)}\widehat{\boldsymbol{\zeta}}^{(s)}$.

Step 1.4: $\widehat{\mu_0}^{*(s)} = \mathbf{C}^{(s)}\left(\mathbf{C}'^{(s)}\mathbf{C}^{(s)} + \lambda^{p(s)}\mathbf{D}\right)^{-1}\mathbf{C}'^{(s)}\mathbf{w}^{(s)}$, $D = diag(0_{p+1}; I_k)$.

Step 2: Using linear mixed model estimation, estimate $\boldsymbol{\gamma}^{(s+1)}$ and $\boldsymbol{\zeta}^{(s+1)}\left(\sigma_{\zeta}^{2(s+1)}\right)$ by fitting

$$\widehat{\mu_0}^{*(s)} = \mathbf{U}^{(s)}\boldsymbol{\gamma} + \mathbf{V}^{(s)}\boldsymbol{\zeta}.$$

Step 3: Using non-linear mixed model estimation, estimate $\boldsymbol{\phi}^{(s+1)}$, $\sigma_{\boldsymbol{\psi}}^{2(s+1)}$, $\sigma_{\boldsymbol{\varepsilon}}^{2(s+1)}$, and $\boldsymbol{\tau}^{(s+1)}$ by fitting the model

$$P_{ijl} = \Pr\left(Y_{ij} \leqslant l|\tau_l, \boldsymbol{\theta}_i\right) = \frac{exp(\eta_{ijl})}{1 + exp(\eta_{ijl})},$$
$$\eta_{ijl} = \tau_l + \left[\alpha_{0i} + \exp(\alpha_{1i})\ \mu_0\left(\beta_{0i} + \exp(\beta_{1i})t_{ij}\right)\right],$$
$$\boldsymbol{\theta}_i = \mathbf{X}_i\boldsymbol{\phi} + \mathbf{Z}_i\boldsymbol{\psi}_i + \boldsymbol{\varepsilon}_i,$$

conditional on $\widehat{\mu_0}^{*(s)}$.

Step 4: Check for convergence. If the algorithm has converged, then stop. Otherwise, increase the iteration counter s by one, and return to step 1.

3.3.2 MCEMNR algorithms

In the current set-up, the E-M steps are as follows

E-Step

The $(s+1)th$ step computes the conditional expectation of l_{ic}

$$Q\left(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\right) = \sum_{i=1}^{N} E_{\boldsymbol{\theta}_i}\left[l_{ic}|D_i, \boldsymbol{\theta}^{(s)}\right], \tag{3.24}$$

where $D_i = (y_i, \mathbf{Z}_i)$. An alternative is to replace the E step with Monte Carlo approximations

constructed using a sample $l_{ic}^{1}, \ldots, l_{ic}^{(M_i)}$

$$Q\left(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\right) \approx \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} l_{ic}^{(m)}\left(\boldsymbol{\theta} \mid D_i, \boldsymbol{\theta}^{(s)}\right). \tag{3.25}$$

We assume without loss of generality, that the number of iterations $M_i = M\ \forall\ i$.

M-Step

The $(s+1)$th step then finds $\boldsymbol{\theta}^{(s+1)}$ as the maximizer of $Q\big(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\big)$:

$$Q\left(\boldsymbol{\theta}^{(s+1)} \mid \boldsymbol{\theta}^{(s)}\right) \geqslant Q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(s)}\right), \tag{3.26}$$

for all $\boldsymbol{\theta}$ in the parameter space. In principle, the M step is carried out by solving the score equations

$$\frac{\partial}{\partial \boldsymbol{\theta}} Q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(s)}\right) = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \frac{\partial}{\partial \boldsymbol{\theta}} l_{ic}^{(m)} \left(\boldsymbol{\theta} \mid D_i, \boldsymbol{\theta}^{(s)}\right) = 0, \quad for\ \boldsymbol{\theta}. \tag{3.27}$$

The essence of the EM algorithm is that increasing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)})$ forces an increase in the log-likelihood of the observed data. To obtain a satisfactory accuracy, the MC sample size M needs to be a large number. For example, to obtain two decimal digits of accuracy, $M \geqslant 10\ 000$ is required. Solution for θ in (3.27) can be obtained by NR procedure.To solve for $\boldsymbol{\theta}$, we proceed through NR steps described below.

NR-Step

With $\boldsymbol{\theta}^{(s)} = (\boldsymbol{\phi}^{(s)},\ \sigma_{\boldsymbol{\psi}}^{2(s)},\ \sigma_{\boldsymbol{\varepsilon}}^{2(s)},\ \boldsymbol{\tau}^{(s)})$, in Newton Raphson procedure is applied to estimate $\boldsymbol{\theta}$ at the $(s+1)$ iteration by using equation (3.27) as follows.

$$\widehat{\boldsymbol{\theta}}^{(s+1)} = \widehat{\boldsymbol{\theta}}^{(s)} + [V^{-1}\hat{S}(\boldsymbol{\theta})]_{|\widehat{\boldsymbol{\theta}}^{(s)}}, \tag{3.28}$$

where

$$\hat{S}\left(\boldsymbol{\theta}\right) = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left[\frac{\partial l_{ic}^{(m)}}{\partial \boldsymbol{\theta}}\right]_{|\widehat{\boldsymbol{\theta}}^{(s)}}, \tag{3.29}$$

and

$$V = \frac{\partial \hat{S}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}_{|\widehat{\boldsymbol{\theta}}^{(s)}} = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \frac{\partial}{\partial \boldsymbol{\theta}^T} \left[\frac{\partial l_{ic}^{(m)}}{\partial \boldsymbol{\theta}}\right]_{|\widehat{\boldsymbol{\theta}}^{(s)}}. \tag{3.30}$$

M is the number of iterations in the Monte Carlo method. This process is repeated until $\left\|\widehat{\boldsymbol{\theta}}^{(s+1)} - \widehat{\boldsymbol{\theta}}^{(s)}\right\| \leqslant \epsilon \downarrow 0$. See Section **??** for more details.

3.4 Simulation

In this section, we use simulated data sets to compare the performances of the proportional odds model under right censoring and our model by the MCNREM algorithm. We assume the number of categories $L = 5$, $N = 50, 100, 200$ and $n = 10, 25$. The number of iterations in the Monte Carlo is 10000. The true values and the estimates of all parameters of the proportional odds model under right censoring and our model by the MCNREM algorithm are displayed in Tables 3.1-3.3 and 3.4-3.6 respectively. Note that both for the proportional odds model under right censoring and the proposed model the true value of the regression parameter vector $\boldsymbol{\phi}$ has been taken to be a null vector. Also the dispersion matrix of the random vector $\boldsymbol{\psi_i}$ is assumed to be $\sigma^2_{\psi} I_4$ for both models with the true value of $\sigma^2_{\psi} = 0.35$. The censoring time is generated from a uniform distribution in $(0, r)$ with proper values of r so that the corresponding censoring rates are 10%, 20% and 40%.

We do not have the closed form expression so we use Newton-Raphson iterations to find the maximizer in EM algorithm. The standard errors (SE) based on generated random samples of different sizes are also computed for each estimator. All the computations are carried out in MATLAB.

Tables 3.1-3.3 and 3.4-3.6 respectively give us an idea about the performance of the estimators under the proportional odds model under right censoring and the self generating model of interest. In tables, we observe the variations in the estimates when $n = 10$ is fixed and the value of N changes from 50 to 100 and then to 200. The situation is pretty similar when $n = 25$ and N increases. From tables 3.4-3.6, it is evident that performance of the estimates is quite satisfactory for our model built up for longitudinal ordinal data.

3.5 An Application to Anaesthesia Recovery

The objective of focusing on the self generating model under right censoring is to study the effects of varying dosages of an anaesthetic on post-surgical recovery. In Davis (1991), 60 young children undergoing outpatient surgery were randomized to one of four dosages (15, 20, 25 and 30 mg/kg) of the anaesthesia, with 15 children per dose group. Recovery scores on a seven-point scale (0:least favorable; 6: most favorable) were assigned upon admission to the recovery room and at minutes 5, 15 and 30 following admission. In addition to the dosage, time when the measurement was taken, age of the patient (in months) and duration of the surgery (in minutes) were considered as covariates. The detailed analysis is based on our proposed model where we consider the recovery as a response variable. The corresponding censoring rates are about 25% and 50%. For choosing the best model in this example, we use the deviance which is the difference between the log-likelihood of the fitted model and the maximum possible log-likelihood. Table 3.7 gives the values of the deviance.

Table 3.9 shows the estimates of intercepts and coeffcients for the sixth model in Table 3.7. Under censoring rate 25%, the p-value of 0.0080, 0.0023 and 0.0162 for Dose, Age and Dose×Age respectively indicate that these factors are significant on the odds of the post-surgical recovery of a patient being less than or equal to a certain value versus being greater than that value. The p-values of 0.0586 Dose×Duration of a patient indicates that the factor is not significant. In another censoring rate only factors Dose and Age with p-values of 0.0194 and 0.0480 respectively are significant. The self generating model with respect to the sixth model with censoring rate 25% is

$$\begin{aligned}\hat{\theta}_i &= \left(\hat{\alpha}_{0i}, \hat{\alpha}_{1i}, \hat{\beta}_{0i}, \hat{\beta}_{1i}\right)' \\ &= 0.0581 Dose_i' + 0.0447 Age_i' - 0.0016 (Dose \times Age)_i' + 0.0003 (Dose \times Dur.)_i',\end{aligned}$$

$$\begin{aligned}
log\left(\frac{P(r_{ijl} \leqslant 0)}{P(r_{ijl} > 0)}\right) &= -5.0521 - \left[\hat{\alpha}_{0i} + \exp\left(\hat{\alpha}_{1i}\right)\, \hat{\mu}_0\left(t_{ij}^*\right)\right],\\
log\left(\frac{P(r_{ijl} \leqslant 1)}{P(r_{ijl} > 1)}\right) &= -4.3099 - \left[\hat{\alpha}_{0i} + \exp\left(\hat{\alpha}_{1i}\right)\, \hat{\mu}_0\left(t_{ij}^*\right)\right],\\
log\left(\frac{P(r_{ijl} \leqslant 2)}{P(r_{ijl} > 2)}\right) &= -2.6073 - \left[\hat{\alpha}_{0i} + \exp\left(\hat{\alpha}_{1i}\right)\, \hat{\mu}_0\left(t_{ij}^*\right)\right],\\
log\left(\frac{P(r_{ijl} \leqslant 3)}{P(r_{ijl} > 3)}\right) &= -2.1794 - \left[\hat{\alpha}_{0i} + \exp\left(\hat{\alpha}_{1i}\right)\, \hat{\mu}_0\left(t_{ij}^*\right)\right],\\
log\left(\frac{P(r_{ijl} \leqslant 4)}{P(r_{ijl} > 4)}\right) &= -1.5271 - \left[\hat{\alpha}_{0i} + \exp\left(\hat{\alpha}_{1i}\right)\, \hat{\mu}_0\left(t_{ij}^*\right)\right],\\
log\left(\frac{P(r_{ijl} \leqslant 5)}{P(r_{ijl} > 5)}\right) &= -1.0530 - \left[\hat{\alpha}_{0i} + \exp\left(\hat{\alpha}_{1i}\right)\, \hat{\mu}_0\left(t_{ij}^*\right)\right].
\end{aligned}$$

Tabela 3.1 – The estimates of all parameters of the proportional odds model under right censoring (10%) by MCEMNR (SE in parenthesis)

N		50		100		200	
n		10	25	10	25	10	25
Parameter	*True value*						
τ_1	$-log3$	-0.8461	-0.8971	-0.8919	-0.9372	-0.9428	-0.9847
		(0.1227)	(0.0775)	(0.0757)	(0.0557)	(0.0472)	(0.0374)
τ_2	0	0.4738	0.4679	0.4082	0.3991	0.3475	0.3247
		(0.1172)	(0.0740)	(0.0728)	(0.0526)	(0.0428)	(0.0335)
τ_3	$log3$	2.6219	2.4929	2.4885	2.3320	2.2441	2.0471
		(0.2500)	(0.1986)	(0.1848)	(0.1268)	(0.1192)	(0.0984)
$\boldsymbol{\phi}$	$\mathbf{0}$	0.0976	0.0871	0.0801	0.0714	0.0641	0.0552
		(0.0813)	(0.0580)	(0.0563)	(0.0403)	(0.0356)	(0.0276)
		0.0982	0.0882	0.0811	0.0722	0.0652	0.0559
		(0.0797)	(0.0584)	(0.0557)	(0.0409)	(0.0362)	(0.0281)
		0.0967	0.0869	0.0799	0.0706	0.0638	0.0545
		(0.0802)	(0.0578)	(0.0561)	(0.0414)	(0.0356)	(0.0284)
		0.0985	0.0885	0.0809	0.0719	0.0647	0.0560
		(0.0788)	(0.0591)	(0.0574)	(0.0402)	(0.0355)	(0.0275)
$\boldsymbol{\sigma}^2_\psi$	0.35	0.3018	0.3105	0.3169	0.3251	0.3310	0.3405
		(0.0716)	(0.0594)	(0.0507)	(0.0401)	(0.0355)	(0.0264)
		0.3020	0.3108	0.3165	0.3248	0.3305	0.3399
		(0.0701)	(0.0622)	(0.0484)	(0.0389)	(0.0349)	(0.0257)
		0.3016	0.3100	0.3171	0.3256	0.3311	0.3410
		(0.0683)	(0.0625)	(0.0468)	(0.0394)	(0.0371)	(0.0291)
		0.3013	0.3099	0.3170	0.3254	0.3316	0.3408
		(0.0697)	(0.0611)	(0.0538)	(0.0441)	(0.0368)	(0.0253)
σ^2_ε	0.25	0.2429	0.2444	0.2451	0.2463	0.2467	0.2479
		(0.0153)	(0.0099)	(0.0091)	(0.0069)	(0.0067)	(0.0040)

Tabela 3.2 – The estimates of all parameters of the proportional odds model under right censoring (20%) by MCEMNR (SE in parenthesis)

N		50		100		200	
n		10	25	10	25	10	25
Parameter	*True value*						
τ_1	$-log3$	-1.3874	-1.2484	-1.3076	-1.1859	-1.2791	-1.1170
		(0.1272)	(0.0801)	(0.0790)	(0.0574)	(0.0492)	(0.0334)
τ_2	0	0.2519	0.2309	0.1907	0.1829	0.1347	0.1201
		(0.1180)	(0.0743)	(0.0729)	(0.0527)	(0.0430)	(0.0337)
τ_3	$log3$	2.5453	2.3208	2.3475	2.1071	1.9889	1.8640
		(0.2034)	(0.1558)	(0.1545)	(0.1008)	(0.0997)	(0.0561)
$\boldsymbol{\phi}$	$\mathbf{0}$	0.0980	0.0876	0.0806	0.0720	0.0639	0.0560
		(0.0812)	(0.0581)	(0.0563)	(0.0403)	(0.0357)	(0.0221)
		0.0976	0.0882	0.0815	0.0726	0.0645	0.0555
		(0.0796)	(0.0585)	(0.0547)	(0.0409)	(0.0363)	(0.0234)
		0.0985	0.0870	0.0801	0.0714	0.0631	0.0569
		(0.0803)	(0.0578)	(0.0563)	(0.0414)	(0.0357)	(0.0228)
		0.0991	0.0888	0.0820	0.0731	0.0643	0.0548
		(0.0788)	(0.0592)	(0.0575)	(0.0402)	(0.0355)	(0.0220)
$\boldsymbol{\sigma}^2_\psi$	0.35	0.3026	0.3118	0.3180	0.3264	0.3316	0.3411
		(0.0697)	(0.0560)	(0.0520)	(0.0391)	(0.0354)	(0.0231)
		0.3018	0.3129	0.3171	0.3270	0.3310	0.3406
		(0.0729)	(0.0603)	(0.0480)	(0.0366)	(0.0343)	(0.0218)
		0.3020	0.3120	0.3182	0.3275	0.3321	0.3419
		(0.0694)	(0.0614)	(0.0476)	(0.0395)	(0.0365)	(0.0261)
		0.3015	0.3117	0.3175	0.3261	0.3324	0.3404
		(0.0692)	(0.0611)	(0.0539)	(0.0434)	(0.0334)	(0.0254)
σ^2_ε	0.25	0.2428	0.2444	0.2450	0.2461	0.2478	0.2483
		(0.0153)	(0.0099)	(0.0090)	(0.0069)	(0.0068)	(0.0039)

Tabela 3.3 – The estimates of all parameters of the proportional odds model under right censoring (40%) by MCEMNR (SE in parenthesis)

N		50		100		200	
n		10	25	10	25	10	25
Parameter	*True value*						
τ_1	$-log3$	-1.5059	-1.4607	-1.4269	-1.3889	-1.3467	-1.2947
		(0.1352)	(0.0887)	(0.0843)	(0.0637)	(0.0521)	(0.0351)
τ_2	0	-0.5524	-0.5030	-0.4518	-0.3970	-0.3491	-0.3017
		(0.1191)	(0.0775)	(0.0746)	(0.0548)	(0.0436)	(0.0340)
τ_3	$log3$	1.4129	1.3658	1.3205	1.2756	1.2391	1.1964
		(0.1536)	(0.1162)	(0.1125)	(0.0765)	(0.0794)	(0.0381)
$\boldsymbol{\phi}$	$\mathbf{0}$	0.0991	0.0883	0.0821	0.0740	0.0648	0.0573
		(0.0814)	(0.0582)	(0.0564)	(0.0405)	(0.0362)	(0.0254)
		0.0979	0.0878	0.0816	0.0732	0.0639	0.0561
		(0.0798)	(0.0587)	(0.0549)	(0.0411)	(0.0368)	(0.0260)
		0.0998	0.0888	0.0830	0.0739	0.0653	0.0580
		(0.0805)	(0.0580)	(0.0566)	(0.0416)	(0.0362)	(0.0251)
		0.0986	0.0891	0.0836	0.0746	0.0660	0.0569
		(0.0798)	(0.0594)	(0.0576)	(0.0404)	(0.0360)	(0.0249)
$\boldsymbol{\sigma}^2_{\psi}$	0.35	0.3018	0.3109	0.3173	0.3258	0.3308	0.3415
		(0.0697)	(0.0560)	(0.0520)	(0.0396)	(0.0347)	(0.0224)
		0.3022	0.3116	0.3184	0.3250	0.3311	0.3406
		(0.0729)	(0.0603)	(0.0480)	(0.0354)	(0.0342)	(0.0228)
		0.3020	0.3121	0.3169	0.3261	0.3305	0.3414
		(0.0694)	(0.0614)	(0.0476)	(0.0390)	(0.0369)	(0.0270)
		0.3031	0.3111	0.3180	0.3248	0.3320	0.3410
		(0.0692)	(0.0601)	(0.0540)	(0.0451)	(0.0339)	(0.0260)
σ^2_{ε}	0.25	0.2438	0.2449	0.2453	0.2462	0.2478	0.2487
		(0.0153)	(0.0099)	(0.0091)	(0.0070)	(0.0069)	(0.0035)

Tabela 3.4 – The estimates of all parameters of the Self generating ordinal model under right censoring (10%) by MCEMNR (SE in parenthesis)

N		50		100		200	
n		10	25	10	25	10	25
Parameter	*True value*						
τ_1	$-log3$	-0.9824	-1.0039	-1.0128	-1.0352	-1.1477	-1.1632
		(0.0609)	(0.0516)	(0.0469)	(0.0388)	(0.0348)	(0.0261)
τ_2	0	0.0892	0.0803	0.0756	0.0708	0.0639	0.0569
		(0.0519)	(0.0433)	(0.0401)	(0.0321)	(0.0294)	(0.0219)
τ_3	$log3$	1.2122	1.1951	1.1863	1.1758	1.1689	1.1580
		(0.0829)	(0.0709)	(0.0661)	(0.0550)	(0.0511)	(0.0416)
$\boldsymbol{\phi}$	$\mathbf{0}$	-0.1032	-0.0906	-0.0824	-0.0699	-0.0635	-0.0503
		(0.0403)	(0.0328)	(0.0286)	(0.0226)	(0.0208)	(0.0145)
		-0.1028	-0.0904	-0.0826	-0.0701	-0.0630	-0.0501
		(0.0408)	(0.0321)	(0.0289)	(0.0233)	(0.0208)	(0.0146)
		-0.1030	-0.0908	-0.0821	-0.0703	-0.0637	-0.0505
		(0.0404)	(0.0322)	(0.0289)	(0.0231)	(0.0207)	(0.0145)
		-0.1035	-0.0901	-0.0823	-0.0702	-0.0631	-0.0504
		(0.0405)	(0.0325)	(0.0286)	(0.0231)	(0.0207)	(0.0145)
$\boldsymbol{\sigma}^2_\psi$	0.35	0.3761	0.3719	0.3683	0.3641	0.3606	0.3577
		(0.0504)	(0.0401)	(0.0367)	(0.0264)	(0.0210)	(0.0108)
		0.3765	0.3720	0.3684	0.3641	0.3606	0.3578
		(0.0504)	(0.0402)	(0.0367)	(0.0264)	(0.0211)	(0.0108)
		0.3764	0.3721	0.3684	0.3642	0.3607	0.3578
		(0.0505)	(0.0401)	(0.0368)	(0.0263)	(0.0210)	(0.0107)
		0.3761	0.3720	0.3685	0.3640	0.3607	0.3577
		(0.0503)	(0.0401)	(0.0368)	(0.0264)	(0.0210)	(0.0107)
σ^2_ε	0.25	0.2449	0.2462	0.2466	0.2478	0.2481	0.2491
		(0.0078)	(0.0055)	(0.0048)	(0.0039)	(0.0035)	(0.0024)
$\boldsymbol{\gamma}$	0	0.0908	0.0763	0.0690	0.0630	0.0571	0.0504
		(0.0405)	(0.0353)	(0.0276)	(0.0237)	(0.0168)	(0.0130)
		0.0908	0.0763	0.0689	0.0630	0.0572	0.0505
		(0.0405)	(0.0354)	(0.0277)	(0.0236)	(0.0167)	(0.0129)
		0.0905	0.0761	0.0690	0.0632	0.0572	0.0505
		(0.0407)	(0.0354)	(0.276)	(0.0237)	(0.0167)	(0.0130)
		0.0904	0.0763	0.0691	0.0633	0.0571	0.0504
		(0.0404)	(0.0352)	(0.0276)	(0.0236)	(0.0168)	(0.0130)

Tabela 3.4 – The estimates of all parameters of the Self generating ordinal model under right censoring (10%) by MCEMNR (SE in parenthesis)

N		50		100		200	
n		10	25	10	25	10	25
Parameter	*True value*						
$\boldsymbol{\sigma}^2_\zeta$	0.3	0.2871	0.2898	0.2912	0.2941	0.2959	0.2976
		(0.0762)	(0.0657)	(0.0469)	(0.0362)	(0.0147)	(0.0055)
		0.2873	0.2896	0.2912	0.2939	0.2960	0.2976
		(0.0761)	(0.0657)	(0.0471)	(0.0362)	(0.0147)	(0.0056)
		0.2871	0.2896	0.2912	0.2940	0.2960	0.2976
		(0.0761)	(0.0653)	(0.0471)	(0.0362)	(0.0148)	(0.0055)
		0.2869	0.2897	0.2911	0.2937	0.2958	0.2977
		(0.0759)	(0.0656)	(0.0471)	(0.0361)	(0.0149)	(0.0055)
		0.2869	0.2899	0.2911	0.2938	0.2959	0.2977
		(0.0761)	(0.0657)	(0.0469)	(0.0362)	(0.0149)	(0.0055)
$\sigma^2_{\varepsilon_p}$	0.2	0.1945	0.1955	0.1959	0.1967	0.1974	0.1982
		(0.0130)	(0.0086)	(0.0085)	(0.0058)	(0.0055)	(0.0039)

Tabela 3.5 – The estimates of all parameters of the Self generating ordinal model under right censoring (20%) by MCEMNR (SE in parenthesis)

N		50		100		200	
n		10	25	10	25	10	25
Parameter	*True value*						
τ_1	$-log3$	-0.9867	-1.0081	-1.0190	-1.0403	-1.1511	-1.1686
		(0.0646)	(0.0550)	(0.0501)	(0.0416)	(0.0381)	(0.0290)
τ_2	0	0.0921	0.0839	0.0798	0.0739	0.0684	0.0605
		(0.0544)	(0.0462)	(0.0420)	(0.0336)	(0.0305)	(0.0231)
τ_3	$log3$	1.2094	1.1926	1.1848	1.1733	1.1652	1.1544
		(0.0801)	(0.0698)	(0.0670)	(0.0526)	(0.0498)	(0.0386)
$\boldsymbol{\phi}$	$\mathbf{0}$	-0.1030	-0.0905	-0.0806	-0.0681	-0.0624	0.0497
		(0.0404)	(0.0328)	(0.0282)	(0.0228)	(0.0207)	(0.0145)
		-0.1032	-0.0901	-0.0809	-0.0682	-0.0625	0.0498
		(0.0404)	(0.0321)	(0.0283)	(0.0232)	(0.0207)	(0.0144)
		-0.1025	-0.0904	-0.0808	-0.0683	-0.0624	0.0496
		(0.0406)	(0.0323)	(0.0281)	(0.0231)	(0.0208)	(0.0144)
		-0.1033	-0.0907	-0.0805	-0.0684	-0.0623	0.0495
		(0.0407)	(0.0324)	(0.0283)	(0.0232)	(0.0208)	(0.0145)
$\boldsymbol{\sigma}^2_{\psi}$	0.35	0.3765	0.3720	0.3685	0.3642	0.3605	0.3579
		(0.0485)	(0.0394)	(0.0331)	(0.0235)	(0.0199)	(0.0101)
		0.3766	0.3721	0.3685	0.3643	0.3605	0.3578
		(0.0487)	(0.0393)	(0.0331)	(0.0235)	(0.0198)	(0.0102)
		0.3761	0.3721	0.3687	0.3643	0.3606	0.3578
		(0.0488)	(0.0394)	(0.0330)	(0.0234)	(0.0199)	(0.0102)
		0.3766	0.3725	0.3684	0.3645	0.3606	0.3579
		(0.0485)	(0.0393)	(0.0330)	(0.0235)	(0.0199)	(0.0101)
σ^2_{ε}	0.25	0.2450	0.2461	0.2467	0.2478	0.2482	0.2491
		(0.0076)	(0.0051)	(0.0047)	(0.0038)	(0.0034)	(0.0023)
$\boldsymbol{\gamma}$	0	0.0909	0.0760	0.0689	0.0629	0.0571	0.0504
		(0.0405)	(0.0353)	(0.0277)	(0.0238)	(0.0168)	(0.0129)
		0.0908	0.0767	0.0689	0.0630	0.0570	0.0504
		(0.0406)	(0.0353)	(0.0277)	(0.0237)	(0.0168)	(0.0129)
		0.0905	0.0760	0.0690	0.0631	0.0572	0.0506
		(0.0407)	(0.0354)	(0.275)	(0.0237)	(0.0167)	(0.0130)
		0.0904	0.0767	0.0692	0.0636	0.0574	0.0506
		(0.0401)	(0.0351)	(0.0275)	(0.0237)	(0.0166)	(0.0129)

Tabela 3.5 – The estimates of all parameters of the Self generating ordinal model under right censoring (20%) by MCEMNR (SE in parenthesis)

N		50		100		200	
n		10	25	10	25	10	25
Parameter	*True value*						
$\boldsymbol{\sigma}_{\zeta}^{2}$	0.3	0.2872	0.2898	0.2911	0.2940	0.2959	0.2975
		(0.0761)	(0.0656)	(0.0469)	(0.0361)	(0.0148)	(0.055)
		0.2873	0.2897	0.2912	0.2939	0.2960	0.2976
		(0.0761)	(0.0657)	(0.0470)	(0.0361)	(0.0148)	(0.0055)
		0.2872	0.2896	0.2912	0.2938	0.2960	0.2977
		(0.0760)	(0.0652)	(0.0471)	(0.0362)	(0.0149)	(0.0056)
		0.2868	0.2900	0.2913	0.2936	0.2957	0.2977
		(0.0759)	(0.0656)	(0.0471)	(0.0360)	(0.0149)	(0.0056)
		0.2869	0.2901	0.2911	0.2935	0.2956	0.2976
		(0.0760)	(0.0657)	(0.0469)	(0.0363)	(0.0149)	(0.0055)
$\sigma_{\varepsilon_p}^{2}$	0.2	0.1944	0.1953	0.1959	0.1968	0.1973	0.1982
		(0.0131)	(0.0086)	(0.0084)	(0.0057)	(0.0055)	(0.0038)

Tabela 3.6 – The estimates of all parameters of the Self generating ordinal model under right censoring (40%) by MCEMNR (SE in parenthesis)

N		50		100		200	
n		10	25	10	25	10	25
Parameter	*True value*						
τ_1	$-log3$	-0.9816	-1.0028	-1.0131	-1.0343	-1.0460	-1.0654
		(0.0724)	(0.0601)	(0.0544)	(0.0453)	(0.0427)	(0.0298)
τ_2	0	0.0980	0.0896	0.0851	0.0792	0.0706	0.0638
		(0.0560)	(0.0485)	(0.0439)	(0.0378)	(0.0336)	(0.0268)
τ_3	$log3$	1.2050	1.1886	1.1801	1.1692	1.1607	1.1491
		(0.0774)	(0.0662)	(0.0640)	(0.0481)	(0.0434)	(0.0277)
$\boldsymbol{\phi}$	$\mathbf{0}$	-0.1029	-0.0907	-0.0811	-0.0677	-0.0626	-0.0493
		(0.0405)	(0.0329)	(0.0286)	(0.0223)	(0.0206)	(0.0141)
		-0.1033	-0.0906	-0.0815	-0.0679	-0.0624	-0.0495
		(0.0406)	(0.0320)	(0.0290)	(0.0224)	(0.0206)	(0.0142)
		-0.1026	-0.0905	-0.0816	-0.0681	-0.0625	-0.0495
		(0.0408)	(0.0322)	(0.0289)	(0.0224)	(0.0205)	(0.0141)
		-0.1031	-0.0908	-0.0813	-0.0676	-0.0626	-0.0494
		(0.0409)	(0.0326)	(0.0286)	(0.0223)	(0.0205)	(0.0142)
$\boldsymbol{\sigma}^2_{\psi}$	0.35	0.3770	0.3724	0.3686	0.3649	0.3608	0.3582
		(0.0498)	(0.0414)	(0.0352)	(0.0248)	(0.0214)	(0.0123)
		0.3771	0.3721	0.3687	0.3647	0.3608	0.3583
		(0.0495)	(0.0414)	(0.0352)	(0.0248)	(0.0214)	(0.0123)
		0.3766	0.3721	0.3687	0.3647	0.3609	0.3583
		(0.0500)	(0.0416)	(0.0351)	(0.0248)	(0.0213)	(0.0123)
		0.3769	0.3722	0.3689	0.3646	0.3609	0.3580
		(0.0498)	(0.0416)	(0.0353)	(0.0248)	(0.0213)	(0.0123)
σ^2_{ε}	0.25	0.2451	0.2461	0.2468	0.2478	0.2483	0.2491
		(0.0077)	(0.0053)	(0.0046)	(0.0038)	(0.0035)	(0.0024)
$\boldsymbol{\gamma}$	0	0.0911	0.0764	0.0691	0.0630	0.0573	0.0506
		(0.0395)	(0.0348)	(0.0274)	(0.0231)	(0.0164)	(0.0126)
		0.0915	0.0769	0.0688	0.0631	0.0571	0.0508
		(0.0396)	(0.0351)	(0.0275)	(0.0230)	(0.0164)	(0.0128)
		0.0905	0.0760	0.0695	0.0635	0.0572	0.0507
		(0.0400)	(0.0352)	(0.274)	(0.0230)	(0.0165)	(0.0128)
		0.0901	0.0768	0.0694	0.0638	0.0575	0.0507
		(0.0387)	(0.0349)	(0.0275)	(0.0229)	(0.0165)	(0.0127)

Tabela 3.6 – The estimates of all parameters of the Self generating ordinal model under right censoring (40%) by MCEMNR (SE in parenthesis)

N		50		100		200	
n		10	25	10	25	10	25
Parameter	*True value*						
$\boldsymbol{\sigma}_\zeta^2$	0.3	0.2861	0.2891	0.2908	0.2933	0.2948	0.2972
		(0.0756)	(0.0651)	(0.0466)	(0.0358)	(0.0143)	(0.0059)
		0.2866	0.2890	0.2911	0.2931	0.2947	0.2971
		(0.0758)	(0.0656)	(0.0467)	(0.0357)	(0.0145)	(0.0059)
		0.2862	0.2895	0.2911	0.2931	0.2947	0.2972
		(0.0760)	(0.0653)	(0.0467)	(0.0358)	(0.0144)	(0.0058)
		0.2865	0.2898	0.2910	0.2932	0.2947	0.2972
		(0.0756)	(0.0653)	(0.0470)	(0.0359)	(0.0144)	(0.0058)
		0.2860	0.2897	0.2909	0.2934	0.2948	0.2973
		(0.0758)	(0.0658)	(0.0462)	(0.0359)	(0.0143)	(0.0059)
$\sigma_{\varepsilon_p}^2$	0.2	0.1945	0.1951	0.1960	0.1967	0.1974	0.1981
		(0.0132)	(0.0087)	(0.0078)	(0.0056)	(0.0045)	(0.0026)

Tabela 3.7 – Values of the deviance for the best subset of each size in two censoring rates

			Deviance	
Model	Subset size	Predictors	CR:25%	CR:50%
1	1	*Dose*	630.2133	624.0238
2	2	*Dose, Age*	625.3785	622.5002
3	2	*Dose, Dur.*	622.9848	620.9668
4	3	*Dose, Age, Dur.*	620.6077	618.3663
5	4	*Dose, Age, Dur., Age*×*Dur.*	617.2568	614.7388
6	4	*Dose, Age, Dose*×*Age, Dose*×*Dur.*	615.3196	613.7527

CR: Censoring Rate, Dur.: Duration

Tabela 3.8 – Values of AIC and BIC for the best subset of each size in two censoring rates

	AIC		BIC	
Model	CR:25%	CR:50%	CR:25%	CR:50%
1	2022.66	2041.44	2703.78	2724.84
2	2012.01	2032.56	2694.16	2716.19
3	2003.97	2024.31	2684.48	2707.25
4	1996.74	2018.92	2678.02	2700.67
5	1984.65	2007.33	2666.40	2689.08
6	1969.37	1991.53	2651.12	2673.28

Tabela 3.9 – Estimates of intercepts and coefficients of the sixth model for two censoring rates

Censoring Rate	Variable	Estimates	Std. Error	P-value
	τ_1	-5.0521	0.4013	0.0000
	τ_2	-4.3099	0.3784	0.0000
	τ_3	-2.6073	0.3202	0.0000
	τ_4	-2.1794	0.3108	0.0000
	τ_5	-1.5271	0.2965	0.0000
25%	τ_6	-1.0530	0.2867	0.0000
	$Dose$	0.0581	0.0219	0.0080
	Age	0.0447	0.0147	0.0023
	$Dose \times Age$	-0.0016	0.0006	0.0162
	$Dose \times Dur$	0.0003	0.0002	0.0586
	τ_1	-5.0472	0.3264	0.0000
	τ_2	-3.6004	0.3007	0.0000
	τ_3	-1.8698	0.2186	0.0000
	τ_4	-1.4377	0.2047	0.0000
	τ_5	-0.7807	0.1872	0.0000
50%	τ_6	-0.3033	0.1801	0.0000
	$Dose$	0.0564	0.0241	0.0194
	Age	0.0329	0.0166	0.0480
	$Dose \times Age$	-0.0013	0.0008	0.0905
	$Dose \times Dur$	0.0002	0.0002	0.2374

4 A proportional hazard cure model for ordinal responses by Self-Modeling regression

4.1 Introduction

Farewell (1982), Farewell (1986) used mixture models for the analysis of survival data with long-term survivors. On the cure models, the population is considered as a mixture of cured patients and uncured patients. Let Y be the indicator variable for an uncured patient with $Y = 1$ if the patient is uncured and 0 if cured, where T is the failure time of a patient. We denote $\boldsymbol{x}$ and $\boldsymbol{z}$ as the covariate vectors, $\pi(\boldsymbol{z})$ as the uncured probability for a subject, and $S(t|\boldsymbol{x}, \boldsymbol{z})$ as the survival function for T, respectively. The proportional hazard cure (PHC) model is given by

$$S(t|\boldsymbol{x}, \boldsymbol{z}) = \pi(\boldsymbol{z})S_u(t|\boldsymbol{x}) + [1 - \pi(\boldsymbol{z})], \tag{4.1}$$

where $S_u(t|\boldsymbol{x})$ be the survival function for uncured subjects. This formulation implies that the survivor function $S_u(t|\boldsymbol{x})$ for t is related to $S_0(t)$ as $S_u(t|\boldsymbol{x}) = \{S_0(t)\}^{\exp(\beta \boldsymbol{x})}$. To specify the effects of $\boldsymbol{z}$ on π, we have

$$\pi(\boldsymbol{z}) = \frac{\exp(\gamma' \boldsymbol{z})}{1 + \exp(\gamma' \boldsymbol{z})}, \tag{4.2}$$

where γ is a vector of unknown parameters. Kuk and Chen (1992) considered the semiparametric logistic PH mixture model. They focused on estimation of the regression parameters using a marginal likelihood method. Sy and Taylor (2000) and Peng and Dear (2000) used the full likelihood approach and derived some EM algorithms to compute the maximum likelihood estimator.

McCullagh (1980) described the proportional odds model for analysis of ordinal data. As a logit link function, the proportional odds model characterizes the ordinal responses in L categories ($l = 1, 2, \ldots, L$) in terms of $L-1$ cumulative category comparisons, specifically, $L - 1$ cumulative logits. In the proportional odds model, the covariate effects are assumed to be the same across these cumulative logits, or proportional across the cumulative odds. Here, denote the conditional cumulative probabilities for the L categories of the outcome y_{ij} as $P_{ijl} = \Pr(y_{ij} \leqslant l | x_{ij}, z_{ij}) = \sum_{k=1}^{l} p_{ijk}$, where p_{ijk} represents the conditional probability of response in category k. The logistic Generalized Linear Mixed Models for the conditional cumulative probabilities is given in terms of the cumulative logits as

$$P_{ijl} = \frac{\exp(\eta_{ijl})}{1+\exp(\eta_{ijl})}, \qquad l = 1, \ldots, L-1, \tag{4.3}$$

where the linear predictor is

$$\eta_{ijl} = \tau_l + \omega_{ij}, \tag{4.4}$$

and

$$\omega_{ij} = x'_{ij}\boldsymbol{\beta} + z'_{ij}\upsilon_i. \tag{4.5}$$

Take $L-1$ strictly increasing model thresholds τ_l $(i.e., \tau_1 < \tau_2 < \ldots < \tau_{L-1})$. x_{ij} is the $(p+1)\times 1$ covariate vector (including the intercept), and z_{ij} is design vector for the r random effects, both vectors being for the jth timepoint nested within subject i. Also, $\boldsymbol{\beta}$ is the $(p+1)\times 1$ vector of unknown fixed regression parameters. Let $\upsilon = T\theta$, where $TT' = \Sigma_\upsilon$ is the Cholesky factorization of random-effect variance covariance matrix Σ_υ.

In this work we will consider the Self-Modeling Regression (SEMOR) Model introduced by Lawton, Sylvestre and Maggio (1972)

$$y_{ij} = \upsilon_i\left\{\mu_0\left[\kappa_i(t_{ij})\right]\right\} + e_{ij}, \tag{4.6}$$

where y_{ij} is the response for curve i, $i = 1, \ldots, N$, measured at n_i times, t_{ij}. $\upsilon_i(x)$ is a monotone inverse link transforming the regression function and $\kappa_i(x)$ is a monotone transformation of the time axis. μ_0 is a shape function that is common to all the curves, and e_{ij} are errors. This paper will focus on nonparametric modeling of μ_0 and parametric modeling of $\upsilon_i(x)$ and $\kappa_i(x)$ with known correlation structure for e_{ij}. In Section 4.2, we will consider the Self-Modeling model defined for the conditional cumulative probabilities for a category of an outcome. Therefore, unlike equation (4.6), in our model there is no relation between the observed response and parameters directly.

In a medical study, patients have various stages of illness. For example the stage of cancer is associated with the severity of the cancer (S = 1,2,3 for minor, medium, severe). The patients are treated with radiation, where one of the three dose levels of radiation *(D* = 1,2,3 for weak, medium, strong) is applied to each patient. After using radiation, the patient will be cured or the stage of his cancer will change. Changing of the stage of the cancer could be an deviance of health or not, therefore the changing will effect on later decisions. To build up a suitable framework for an analysis of such data, we propose the use of self-modeling ordinal longitudinal model. Unlike the univariate mixture cure model (Boag, 1949, Farewell, 1982), a mixture cure model based on ordinal responses is proposed in Section 4.3. We outline the proportional hazard mixture cure model with random effects. In Section 4.4, parameter estimation with using the Monte Carlo method in Newton-Raphson and EM algorithms are introduced. Section 4.5 discuss simulation results of the estimation methods.

Schizophrenia is a disorder of the mind that results in disorganisation of normal thinking and feeling. Schizophrenia as a mental illness can come in various forms with different symptoms and outcomes. In Section 4.6, as an application, we focus on the question of whether four medications affects schizophrenia patients. In particular, we investigate the interaction between the medication effect and time to follow-up.

4.2 Self-Modeling Ordinal model

Consider the Shape-Invariant (SI),

$$Y_{ij} = \alpha_{0i} + \exp(\alpha_{1i})\mu_0(\beta_{0i} + \exp(\beta_{1i})t_{ij}) + \epsilon_{ij} \tag{4.7}$$

where Y_{ij} is the observed response on subject i at time t_{ij}. With

$$\omega_{ij} = \alpha_{0i} + \exp\left(\alpha_{1i}\right)\, \mu_0\left(t^*_{ij}\right), \tag{4.8}$$

where $t^*_{ij} = \beta_{0i} + \exp\left(\beta_{1i}\right) t_{ij}$ and

$$\eta_{ijl} = log\left[\frac{P_{ijl}}{1 - P_{ijl}}\right] = \tau_l + \left[\alpha_{0i} + \exp\left(\alpha_{1i}\right)\, \mu_0\left(t^*_{ij}\right)\right]. \tag{4.9}$$

Lawton, Sylvestre and Maggio (1972), Kneip and Gasser (1988) and Kneip e Engel (1995) considered

$$\boldsymbol{\theta}_i = \left(\alpha_{0i}, \alpha_{1i}, \beta_{0i}, \beta_{1i}\right)', \tag{4.10}$$

while we consider a mixed model

$$\boldsymbol{\theta}_i = \mathbf{X}_i\boldsymbol{\phi} + \mathbf{Z}_i\boldsymbol{\psi}_i + \boldsymbol{\varepsilon}_i, \tag{4.11}$$

where $\boldsymbol{Z}_i$ is the design (or covariate) matrix for the random effect vector $\boldsymbol{\psi}_i$.

As discussed before the use of the penalized spline method with penalty chosen by generalized maximum likelihood (Wahba, 1985) is equivalent to fitting the model

$$\mu_0(t^*_{ij}) = \mathbf{U}\boldsymbol{\gamma} + \mathbf{V}\boldsymbol{\zeta}, \tag{4.12}$$

where $\mu_0(t^*_{ij})$ is the vector of means at the transformed times, $\mathbf{U}$ is a design matrix for a cubic polynomial in t^*_{ij}, $\mathbf{V}$ is a design matrix for cubics in t^*_{ij} which are left-truncated at the knot, $\boldsymbol{\gamma}$ is a vector of unknown parameters and $\boldsymbol{\zeta}$ is normally distributed with zero mean and covariance matrix $\boldsymbol{\Sigma}_{\zeta}$.

The result of a nonparametric regression analysis is a curve fitted to a set of data $(\mathbf{t}^*_i, \mathbf{w}_i)$

$$\mathbf{w}_i = f_0(\mathbf{t}^*_i) + \boldsymbol{\varepsilon}_i, \tag{4.13}$$

where f_0 is a smooth function giving the conditional mean of $\mathbf{w}_i$ given $\mathbf{t}^*_i$ and the $\boldsymbol{\varepsilon}_i$ are mutually independent with $N(\mathbf{0}, \sigma^2_{\varepsilon})$ for $i = 1, 2, \ldots, n$. The smooth function μ_0 can be found as result of minimization of the residual sum of squares plus a roughness penalty,

$$\sum_{i=1}^{n} \left(\mathbf{w}_i - f_0\left(\mathbf{t}^*_i\right)\right)^2 + \lambda \int \left(f_0^p(\mathbf{t}^*_i)\right)^2 d\mathbf{t}^*_i, \tag{4.14}$$

where $f_0^p(\mathbf{t}^*_i)$ is the pth derivative of the function $f_0(\mathbf{t}^*_i)$. The resultant curve fitted to the data is a piecewise polynomial of degree $2p-1$. The smoothing parameter λ governs the trade-off between smoothness and goodness of fit. This parameter is often unknown in practice and needs to be estimated from the data. A classical data-driven approach to selecting the smoothing parameter is cross-validation, which leaves out one subject's entire data at a time.

A random coefficient linear regression spline model of order p for $\mu_0(t^*_{ij})$ by Ruppert and Carroll (1997)., is

$$\mu_0\left(t^*_{ij}\right) = \beta_0 + \beta_1 t^*_{ij} + \cdots + \beta_p {t^*_{ij}}^p + \beta_{p+1}\left(t^*_{ij} - \xi_1\right)^p_+ + \cdots + \beta_{p+K}\left(t^*_{ij} - \xi_K\right)^p_+, \tag{4.15}$$

where the parameters to be estimated are $\boldsymbol{\beta}$=$(\beta_0, \ldots, \beta_{p+k})$ and $\{\xi_1, \ldots, \xi_k\}$ are K fixed knots with $a \leqslant \xi_1 < \ldots < \xi_k \leqslant b$ and $(x)^p_+ = x^p I_{\{x \geqslant 0\}}$. We can write equation (4.12) with $p = 3$ as

$$\mu_0\left(t^*_{ij}\right) = \sum_{m=1}^{4} {t^*_{ij}}^{m-1} \gamma_m + \sum_{k=1}^{K} \left(t^*_{ij} - \xi_k\right)^3_+ \zeta_k, \tag{4.16}$$

where $\boldsymbol{\zeta} = [\zeta_1, \ldots, \zeta_k]' \sim N(0, \sigma^2_{\zeta} I)$ is independent of $\boldsymbol{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_k]'$. We can combine equations (4.13) and (4.16) in one model

$$\mathbf{w} = \mathbf{U}\boldsymbol{\gamma} + \mathbf{V}\boldsymbol{\zeta} + \boldsymbol{\varepsilon}, \tag{4.17}$$

with $\boldsymbol{\varepsilon} \sim N(0,\ \sigma_\varepsilon^2 I)$ and

$$\mathbf{U}=\begin{bmatrix} 1 & t_{i1}^* & t_{i1}^{*2} & t_{i1}^{*3} \\ 1 & t_{i2}^* & t_{i2}^{*2} & t_{i2}^{*3} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_{in_i}^* & t_{in_i}^{*2} & t_{in_i}^{*3} \end{bmatrix}, \mathbf{V}=\begin{bmatrix} (t_{i1}^* - \xi_1)^3 & \dots & (t_{i1}^* - \xi_k)^3 \\ \vdots & \vdots & \vdots \\ \left(t_{in_i}^* - \xi_1\right)^3 & \cdots & \left(t_{in_i}^* - \xi_k\right)^3 \end{bmatrix}. \tag{4.18}$$

Equation (4.17) is nothing but a normal linear mixed model and, for any given σ_ζ^2 and σ_ε^2, the estimated best linear unbiased predictor (EBLUP) of $\mathbf{w}$ by Robinson (1991)

$$\widehat{\mathbf{w}} = \widehat{\mu}_0 = \mathbf{U}\widehat{\boldsymbol{\gamma}} + \mathbf{V}\widehat{\boldsymbol{\zeta}}. \tag{4.19}$$

Unbiased refers here to the property that the average value of the estimate is equal to the average value of the quantity being estimated, that is $E\left(\widehat{\mu}_0\right) = E(\mu_0)$. Equation (4.19) can be rewritten by McCulloch and Searle (2001) as

$$\widehat{\mu}_0^* = \mathbf{C}(\mathbf{C}'\mathbf{C} + \lambda^p\mathbf{D})^{-1}\mathbf{C}'\mathbf{w}, \tag{4.20}$$

where $\mathbf{C} = [\mathbf{U}\ \mathbf{V}]$, $\mathbf{D}=Diag(0_{p+1}, I_k)$ and $\lambda^p = \sigma_\varepsilon^2/\sigma_\zeta^2$ for the pth degree of penalized spline model. It has been shown in Wang (1998) and Brumback and J.A. (1998) that the EBLUP estimates (4.20) evaluated at design points are the same as the penalized regression spline solution to equation (4.14). Thus it turns out that the nonparametric smoothing spline regression is equivalent to a mixed-effects model (4.17). We see that the smoothing parameter λ^p is the ratio of the variance components $\sigma_\varepsilon^2/\sigma_\zeta^2$, and that fitting can be done using standard linear mixed effects software. This is the GML method of Wahba (1985).

4.3 The model

Suppose an ordinal variable y_{ij} where $y_{ij} \leqslant l$ indicates that an individual will experience a particular event and $y_{ij} > l$ indicates that the individual will never experience the event. Let $O = \left\{t_{ij} = min(t_{ij}^*, c_{ij}), \delta_{ij} = I(t_{ij}^* \leqslant c_{ij}), x_{ij}^*, z_{ij}^*, i = 1, \ldots, N, j = 1, \ldots, n_i\right\}$ be the observed multivariate failure data, where (t_{ij}, δ_{ij}) denote the failure time and censoring indicator and x_{ij}^* and z_{ij}^* be the covariates of the jth observation within the ith individual that may affect the failure time distribution of uncured individuals and the cure proportion. t_{ij}^* is the value of the underlying failure time T_{ij} that may be subject to censoring with the censoring time c_{ij}. We assume that given covariates, the censoring time c_{ij} is independent of the failure time t_{ij}^* and the cure status y_{ij}. Let $f_u\left(t_{ij}|y_{ij} \leqslant l, x_{ij}^*, z_{ij}^*\right)$ and $S_u\left(t_{ij}|y_{ij} \leqslant l, x_{ij}^*, z_{ij}^*\right)$ be the probability density function (pdf) and the survival function for uncured subjects. Since the ordinal model is defined in terms of the cumulative probabilities, the conditional probability of a response in category l is obtained as the difference of two conditional cumulative probabilities:

$$p_{ijl} = \Pr\left(Y_{ij} = l|x_{ij}, z_{ij}\right) = P_{ijl} - P_{ijl-1}, \tag{4.21}$$

where P_{ijl} is given by equation (4.8). Here, $\tau_0 = -\infty$ and $\tau_L = \infty$, and so $P_{ij0} = 0$ and $P_{ijL} = 1$. The SI-PHC model can be written as a mixture model in terms of the survival function

$$S\left(t_{ij}|y_{ij} \leqslant l, x^*_{ij}, z^*_{ij}\right) = (1 - P_{ijl}) + p_{ijl} S_u\big(t_{ij}|y_{ij} \leqslant l, x^*_{ij}, z^*_{ij}\big), \tag{4.22}$$

where

$$S_u\big(t_{ij}|y_{ij} \leqslant l, x^*_{ij}, z^*_{ij}\big) = \Big[S_0\big(t_{ij}|y_{ij} \leqslant l\big)\Big]^{\exp(x^*_{ij}\phi^* + z^*_{ij}\psi^*_i)}. \tag{4.23}$$

$S_0(\cdot)$ is an arbitrary baseline survival function for uncured individuals. In this paper, we will consider the Weibull distribution. Under the Weibull baseline assumption, $S_0(t) = [-(t/b)^a]$ where $a, b > 0$ are the shape and scale parameters of the distribution, respectively. The marginal likelihood for the multivariate cure model is

$$\begin{aligned} L\left(\boldsymbol{\theta}\right) = \prod_{i=1}^{N} \int \Bigg\{ \prod_{j=1}^{n_i} \prod_{l=1}^{L} \Bigg(& \Big[p_{ijl} f_u\big(t_{ij}|y_{ij} \leqslant l, x^*_{ij}\phi^* + z^*_{ij}\psi^*_i\big)\Big]^{\delta_{ij}} \\ & \times \Big[(1 - P_{ijl}) + p_{ijl} S_u\big(t_{ij}|y_{ij} \leqslant l, x^*_{ij}\phi^* + z^*_{ij}\psi^*_i\big)\Big]^{1-\delta_{ij}} \Bigg)^{I_{y_{ij}}(l)} \\ & \times f\left(\boldsymbol{\theta}_i|\boldsymbol{\psi}_i\right) f\left(\boldsymbol{\psi}_i|\sigma^2_{\psi_i}\right) f\left(\boldsymbol{\psi}^*_i|\sigma^2_{\psi^*_i}\right) \Bigg\} d\boldsymbol{\theta}_i \; d\boldsymbol{\psi}_i \; d\boldsymbol{\theta}^*_i \; d\boldsymbol{\psi}^*_i, \end{aligned} \tag{4.24}$$

where

$$I_{y_{ij}} = \begin{cases} 1 & \text{if } y_{ij} = l \\ 0 & \text{if } y_{ij} \neq l. \end{cases}$$

and $\boldsymbol{\theta} = (\boldsymbol{\phi}', \ \sigma^2_{\psi}, \ \sigma^2_{\varepsilon}, \ \boldsymbol{\tau}', \ \boldsymbol{\phi}^{*'}, \ \sigma^2_{\psi^*})'$. Since $\boldsymbol{\psi}^*_i$ has $N\left(\mathbf{0}, \boldsymbol{\Sigma}\left(\sigma^2_{\psi^*_i}\right)\right)$, therefore the complete log-likelihood is

$$
\begin{aligned}
l_c = \sum_{i=1}^{N} l_{ic} = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{l=1}^{L} I_{y_{ij}}(l)\Bigg(\delta_{ij}\Big[\log p_{ijl} + \log f_u\big(t_{ij}|y_{ij} \leqslant l, x_{ij}^{*}\phi^{*} + z_{ij}^{*}\psi_i^{*}\big)\Big] \\
+ (1-\delta_{ij})\log\Big[(1-P_{ijl}) + p_{ijl}S_u\big(t_{ij}|y_{ij} \leqslant l, x_{ij}^{*}\phi^{*} + z_{ij}^{*}\psi_i^{*}\big)\Big]\Bigg) \\
+ \log\Big[\boldsymbol{\Phi}\left(\boldsymbol{\theta}_i|\mathbf{X}_i\boldsymbol{\phi} + \mathbf{Z}_i\boldsymbol{\psi}_i,\ \sigma^2_{\varepsilon_i}I_4\right)\Big] + \log\Big[\boldsymbol{\Phi}\left(\boldsymbol{\psi}_i|0,\ \sigma^2_{\psi_i}\right)\Big] \\
+ \log\Big[\boldsymbol{\Phi}\left(\boldsymbol{\theta}_i^{*}|\mathbf{X}_i^{*}\boldsymbol{\phi}^{*} + \mathbf{Z}_i^{*}\boldsymbol{\psi}_i^{*},\ \sigma^2_{\varepsilon_i^{*}}I_4\right)\Big] + \log\Big[\boldsymbol{\Phi}\left(\boldsymbol{\psi}_i^{*}|0,\ \sigma^2_{\psi_i^{*}}\right)\Big],
\end{aligned}
\tag{4.25}
$$

where Φ is a probability density function.

4.4 Estimation method

In this section we propose an algorithm for parameter estimation in the SI ordinal model. Essentially steps 0-2 are considered for predicting the shape function μ_0 and step 3 provides estimate of the parameters for the cummulative logit model through MCNR and MCEM approaches.

Step 0: Choose initial estimates of $\tau^{(s)}$, $\phi^{(s)}$, $\phi^{*(s)}$ and $\boldsymbol{\gamma}^{(s)}$, variances ${\sigma^2_{\psi}}^{(s)}$, $\sigma^{2(s)}_{\varepsilon}$, ${\sigma^2_{\psi^*}}^{(s)}$, ${\sigma^2_{\varepsilon^*}}^{(s)}$, $\sigma^{2(s)}_{\zeta}$ and $\sigma^2_{\varepsilon_p}$, random effects ${\boldsymbol{\psi}_i}^{(s)} \sim N_4\left(\mathbf{0}, \boldsymbol{\Sigma}\left({\sigma^2_{\psi}}^{(s)}\right)\right)$ and $\boldsymbol{\psi}_i^{*(s)} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}\left({\sigma^2_{\psi^*}}^{(s)}\right)\right)$, errors ${\boldsymbol{\varepsilon}_i}^{(s)} \sim N_4\left(\mathbf{0}, \sigma^{2(s)}_{\varepsilon}I_4\right)$ and $\boldsymbol{\varepsilon}_i^{*(s)} \sim N\left(0, {\sigma^2_{\varepsilon^*}}^{(s)}\right)$. Set $s = 0$.

Step 1: Compute $\boldsymbol{\theta}_i^{*(s)} = \mathbf{X}_i^{*}\boldsymbol{\phi}^{*(s)} + \mathbf{Z}_i^{*}\boldsymbol{\psi}_i^{*(s)} + \boldsymbol{\varepsilon}_i^{*(s)}$ and $\boldsymbol{\theta}_i^{(s)} = \mathbf{X}_i\boldsymbol{\phi}^{(s)} + \mathbf{Z}_i\boldsymbol{\psi}_i^{(s)} + \boldsymbol{\varepsilon}_i^{(s)}$ and extract $\alpha_{0i}^{(s)}$, $\alpha_{1i}^{(s)}$, $\beta_{0i}^{(s)}$ and $\beta_{1i}^{(s)}$

Step 1.1: ${t_{ij}^{*}}^{(s)} = \beta_{0i}^{(s)} + \exp\left(\beta_{1i}^{(s)}\right)t_{ij}$,

Step 1.2: $\mathbf{U}^{(s)} = \begin{bmatrix} 1 & {t_{i1}^{*}}^{(s)} & t_{i1}^{*2(s)} & t_{i1}^{*3(s)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & {t_{in_i}^{*}}^{(s)} & {t_{in_i}^{*2}}^{(s)} & {t_{in_i}^{*3}}^{(s)} \end{bmatrix}$, $\mathbf{V}^{(s)} = \begin{bmatrix} \left({t_{i1}^{*}}^{(s)} - \xi_1\right)^3 & \cdots & \left({t_{i1}^{*}}^{(s)} - \xi_k\right)^3 \\ \vdots & \vdots & \vdots \\ \left({t_{in_i}^{*}}^{(s)} - \xi_1\right)^3 & \cdots & \left({t_{in_i}^{*}}^{(s)} - \xi_k\right)^3 \end{bmatrix}$,

$$\mathbf{C}^{(s)} = \left[\mathbf{U}^{(s)} \quad \mathbf{V}^{(s)}\right].$$

Step 1.3: The pth degree of penalized spline model $\lambda^{p(s)} = {\sigma^2_{\varepsilon_p}}^{(s)}/\sigma^{2(s)}_{\zeta}$. $\mathbf{w}^{(s)} = \mathbf{U}^{(s)}\widehat{\boldsymbol{\gamma}}^{(s)} + \mathbf{V}^{(s)}\widehat{\boldsymbol{\zeta}}^{(s)} + \widehat{\boldsymbol{\varepsilon}}_p^{(s)}$.

Step 1.4: $\widehat{\mu}_0^{*(s)} = \mathbf{C}^{(s)}\left(\mathbf{C}'^{(s)}\mathbf{C}^{(s)} + \lambda^{p(s)}\mathbf{D}\right)^{-1}\mathbf{C}'^{(s)}\mathbf{w}^{(s)}$, $D = diag(0_{p+1}; I_k)$.

Step 2: Using linear mixed model estimation, estimate $\boldsymbol{\gamma}^{(s+1)}$ and $\boldsymbol{\zeta}^{(s+1)}\left(\sigma^{2(s+1)}_{\zeta}\right)$ by fitting

$$\widehat{\mu}_0^{*(s)} = \mathbf{U}^{(s)}\boldsymbol{\gamma} + \mathbf{V}^{(s)}\boldsymbol{\zeta} + \boldsymbol{\varepsilon}_p.$$

Step 3: Using nonlinear mixed model estimation, estimate $\tau^{(s+1)}$, $\boldsymbol{\phi}^{(s+1)}$, $\sigma_{\psi}^{2\,(s+1)}$, $\sigma_{\varepsilon}^{2(s+1)}$, $\boldsymbol{\phi}^{*(s+1)}$, $\sigma_{\psi*}^{2\;(s+1)}$ and $\sigma_{\varepsilon*}^{2\;(s+1)}$ by fitting the model

$$\eta_{ijl} = \tau_l + \Big[\alpha_{0i} + \exp(\alpha_{1i})\; \mu_0\Big(\beta_{0i} + \exp(\beta_{1i})t_{ij}\Big)\Big],$$

$$P_{ijl} = \frac{exp(\eta_{ijl})}{1 + exp(\eta_{ijl})},\; p_{ijl} = P_{ijl} - P_{ijl-1}$$

$$\boldsymbol{\theta}_i = \mathbf{X}_i\boldsymbol{\phi} + \mathbf{Z}_i\boldsymbol{\psi}_i + \boldsymbol{\varepsilon}_i,$$

$$\boldsymbol{\theta}_i^* = \mathbf{X}_i^*\boldsymbol{\phi}^* + \mathbf{Z}_i^*\boldsymbol{\psi}_i^* + \boldsymbol{\varepsilon}_i^*$$

,

conditional on $\widehat{\mu}_0^{*(s)}$.

Step 4: Check for convergence. If the algorithm has converged, then stop. Otherwise, increase the iteration counter s by one, and return to step 1.

4.4.1 MCEMNR algorithms

In the current setup, the E-M steps are as follows

E-Step

The $(s+1)$th step computes the conditional expectation of l_{ic}

$$Q\big(\boldsymbol{\theta}\,\big|\boldsymbol{\theta}^{(s)}\big) = \sum_{i=1}^{N} E_{\boldsymbol{\theta}_i}\left[l_{ic}|D_i, \boldsymbol{\theta}^{(s)}\right], \tag{4.26}$$

where $D_i = (y_i,\ \mathbf{X}_i,\ \mathbf{Z}_i,\ \mathbf{X}_i^*,\ \mathbf{Z}_i^*)$. An alternative is to replace the E step with Monte Carlo approximations constructed using a sample $l_{ic}^{(1)}, \ldots, l_{ic}^{(M_i)}$

$$Q\big(\boldsymbol{\theta}\,\big|\boldsymbol{\theta}^{(s)}\big) \approx \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} l_{ic}^{(m)}\left(\boldsymbol{\theta} \mid D_i, \boldsymbol{\theta}^{(s)}\right). \tag{4.27}$$

M_i is the number of iterations in the Monte Carlo method. We assume without loss of generality, the number of iterations $M_i = M\ \forall\ i$.

M-Step

The $(s+1)$th step then finds $\boldsymbol{\theta}^{(s+1)}$ as the maximizer of $Q\big(\boldsymbol{\theta}\,\big|\boldsymbol{\theta}^{(s)}\big)$:

$$Q\left(\boldsymbol{\theta}^{(s+1)} \mid \boldsymbol{\theta}^{(s)}\right) \geqslant Q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(s)}\right), \tag{4.28}$$

for all $\boldsymbol{\theta}$ in the parameter space. In principle, the M step is carried out by solving the score equations

$$\frac{\partial}{\partial \boldsymbol{\theta}} Q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(s)}\right) = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \frac{\partial}{\partial \boldsymbol{\theta}} l_{ic}^{(m)}\left(\boldsymbol{\theta} \mid D_i, \boldsymbol{\theta}^{(s)}\right) = 0, \quad for\ \boldsymbol{\theta}. \tag{4.29}$$

The essence of the EM algorithm is that increasing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)})$ forces an increase in the log-likelihood of the observed data. To obtain a satisfactory accuracy, the MC sample size M needs to be a large number. For example, to obtain two decimal digits of accuracy, $M \geqslant 10\ 000$ is required. Solution for θ in (4.29) can be obtained by NR procedure.To solve for $\boldsymbol{\theta}$, we proceed through NR steps described below.

NR-Step

With $\boldsymbol{\theta}^{(s)} = (\boldsymbol{\phi}^{(s)},\ \sigma_{\boldsymbol{\psi}}^{2\,(s)},\ \sigma_{\varepsilon}^{2(s)},\ \tau^{(s)},\ \boldsymbol{\phi}^{*(s)},\ \sigma_{\boldsymbol{\psi}*}^{2\ \ (s)},\ \sigma_{\varepsilon*}^{2\ \ (s)})$, Newton Raphson procedure is applied to estimate $\boldsymbol{\theta}$ at the $(s+1)$ iteration by using equation (20) as follows.

$$\widehat{\boldsymbol{\theta}}^{(s+1)} = \widehat{\boldsymbol{\theta}}^{(s)} + [V^{-1}\hat{S}(\boldsymbol{\theta})]_{|\widehat{\boldsymbol{\theta}}^{(s)}}\ , \tag{4.30}$$

where

$$\hat{S}\left(\boldsymbol{\theta}\right) = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left[\frac{\partial l_{ic}^{(m)}}{\partial \boldsymbol{\theta}}\right]_{|\widehat{\boldsymbol{\theta}}^{(s)}}, \tag{4.31}$$

and

$$V = \frac{\partial \hat{S}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}_{|\widehat{\boldsymbol{\theta}}^{(s)}} = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \frac{\partial}{\partial \boldsymbol{\theta}^T} \left[\frac{\partial l_{ic}^{(m)}}{\partial \boldsymbol{\theta}}\right]_{|\widehat{\boldsymbol{\theta}}^{(s)}}. \tag{4.32}$$

This process is repeated until $\left\|\widehat{\boldsymbol{\theta}}^{(s+1)} - \widehat{\boldsymbol{\theta}}^{(s)}\right\| \leqslant \epsilon \downarrow 0$. See Section **??** for more details

4.5 Simulation

In this section, we use simulated data sets to compare the performances of the PHC model with random effect and the SI-PHC model by the MCNREM algorithm. The sample size is assumed to be $N = 100, 250, 500$ for $i = 1, .., N$ and $n = 10, 25$ for $j = 1, .., n$. The number of iterations in the Monte Carlo is 10000. The true values and the estimates of all parameters of the PHC model and the SI-PHC by the MCNREM algorithm are displayed in Tables 4.1-4.3 and 4.4-4.6 respectively. Note that both for the PHC model and the SI-PHC model, the true values of the regression parameter vector $\boldsymbol{\phi}$

and $\boldsymbol{\phi}^*$ respectively have been taken to be a null vector and $log(0.5)$. Also the dispersion matrices of the random vectors $\boldsymbol{\psi}_i$ and $\boldsymbol{\psi}_i^*$ are assumed to be $\sigma^2_{\psi} I_4$ and $\sigma^2_{\psi*}$ for both models with the true value of $\sigma^2_{\psi} = 0.35$ and $\sigma^2_{\psi*} = 0.3$ respectively. The baseline hazard function $h_0(t)$ is assumed to be Weibull distribution with $a = 2$ and $b = 1$. The censoring time is generated from a uniform distribution in $(0, r)$ with proper values of r so that the corresponding censoring rates are about 10%, 20% and 40%. The standard errors (SE) based on generated random samples of different sizes are also computed for each estimator. All the computations are carried out in MATLAB.

Tables 4.1-4.3 and 4.4-4.6 respectively give us an idea about the performance of the estimators under the PHC model and the SI-PHC model of interest. The tables show that the estimates from the proposed model tend to have smaller variances than those from the PHC model. In the tables, the results also show that a larger sample size improves the standard error estimation. From Table 4.4-4.6, it is evident that performance of the estimates is quite satisfactory for the SI-PHC model built up for longitudinal data.

4.6 An Application to schizophrenia data

Schizophrenia is a mental disorder often characterized by abnormal social behaviour and failure to recognize what is real. Common symptoms include false beliefs, unclear or confused thinking, auditory hallucinations, reduced social engagement and emotional expression, and lack of motivation. Schizophrenia affects around 0.3-0.7% of people at some point in their life. It occurs 1.4 times more frequently in males than females and typically appears earlier in men. The peak ages of onset are 25 years for males and 27 years for females. Onset in childhood is much rarer, as is onset in middle or old age.

In the U.S. National Institute of Mental Health (NIMH) Schizophrenia Collaborative Study [1], 329 patients were randomly assigned to receive one of four medications: placebo, chlorpromazine, fluphenazine, or thioridazine; the latter three medications are anti-psychotic drugs. The study protocol called for measurements to be made at weeks 0, 1, 3, and 6. The outcome variable of interest is a 4-level ordinal scale measuring "severity of illness", derived from item 79 of the Inpatient Multidimensional Psychiatric Scale (Lorr and Klett, 1966). The four categories of the ordinal response correspond to: 1 = normal or borderline mentally ill, 2 = mildly or moderately ill, 3 = markedly ill, and 4 = severely or among the most extremely ill. To illustrate application the SI-PHC model, we consider the schizophrenia data with censoring rates 25% and 50% that we defined for this data.

Table 4.7 shows the estimates of intercepts and coeffcients for the model. Under censoring rates 25% and 50%, the p-value of 0.0971 and 0.1016 for Treatment indicate

[1] www.hsph.harvard.edu/fitzmaur/ala2e/schizophrenia.txt

that this factor is not significant on the odds of schizophrenia of a patient being less than or equal to a certain value versus being greater than that value.

The self generating model with respect to the model with censoring rate 25% is

$$\hat{\boldsymbol{\theta}}_i = \left(\hat{\alpha}_{0i}, \hat{\alpha}_{1i}, \hat{\beta}_{0i}, \hat{\beta}_{1i}\right)' = 0.1843 Treatment'_i,$$

$$log\left(\frac{P(r_{ijl} \leqslant 1)}{P(r_{ijl} > 1)}\right) = 1.4812 + \left[\hat{\alpha}_{0i} + \exp\left(\hat{\alpha}_{1i}\right)\,\hat{\mu}_0\left(t^*_{ij}\right)\right],$$
$$log\left(\frac{P(r_{ijl} \leqslant 2)}{P(r_{ijl} > 2)}\right) = 2.8267 + \left[\hat{\alpha}_{0i} + \exp\left(\hat{\alpha}_{1i}\right)\,\hat{\mu}_0\left(t^*_{ij}\right)\right],$$
$$log\left(\frac{P(r_{ijl} \leqslant 3)}{P(r_{ijl} > 3)}\right) = 3.2670 + \left[\hat{\alpha}_{0i} + \exp\left(\hat{\alpha}_{1i}\right)\,\hat{\mu}_0\left(t^*_{ij}\right)\right],$$

Tabela 4.1 – The estimates of all parameters of the PHC model with censoring rate (10%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
$\boldsymbol{\phi}$	**0**	0.1168	0.0923	0.0876	0.0638	0.569	0.0326
		(0.0679)	(0.0434)	(0.0422)	(0.0301)	(0.0269)	(0.0189)
		0.1179	0.0944	0.0890	0.0647	0.0575	0.0333
		(0.0670)	(0.439)	(0.0429)	(0.0304)	(0.0271)	(0.0191)
		0.1196	0.0966	0.0901	0.0653	0.0583	0.0340
		(0.0677)	(0.0441)	(0.0426)	(0.0301)	(0.0268)	(0.0192)
		0.1154	0.0911	0.0861	0.0630	0.0561	0.0321
		(0.0672)	(0.0437)	(0.0423)	(0.0300)	(0.0267)	(0.0191)
$\boldsymbol{\sigma}^2_\psi$	0.35	0.3221	0.3277	0.3289	0.3343	0.3354	0.3389
		(0.0421)	(0.0348)	(0.0329)	(0.0239)	(0.0211)	(0.0146)
		0.3219	0.3276	0.3291	0.3339	0.3350	0.3392
		(0.0419)	(0.0353)	(0.0334)	(0.0237)	(0.0216)	(0.0149)
		0.3225	0.3280	0.3287	0.3345	0.3352	0.3394
		(0.0425)	(0.0350)	(0.0331)	(0.0241)	(0.0210)	(0.0145)
		0.3220	0.3275	0.3292	0.3340	0.3351	0.3388
		(0.0423)	(0.0346)	(0.0327)	(0.0243)	(0.0215)	(0.0151)
σ^2_ε	0.25	0.2416	0.2434	0.2441	0.2458	0.2462	0.2475
		(0.0186)	(0.0143)	(0.0136)	(0.0097)	(0.0089)	(0.0056)
ϕ^*	$log(0.5)$	-0.6778	-0.6796	-0.6801	-0.6818	-0.6823	-0.6847
		(0.0348)	(0.0274)	(0.0251)	(0.0189)	(0.0165)	(0.0111)
$\sigma^2_{\psi*}$	0.3	0.2910	0.2926	0.2932	0.2945	0.2953	0.2967
		(0.0358)	(0.0334)	(0.0279)	(0.0256)	(0.0203)	(0.0181)
$\sigma^2_{\varepsilon*}$	0.2	0.1959	0.1968	0.1971	0.1980	0.1983	0.1990
		(0.0091)	(0.0073)	(0.0070)	(0.0054)	(0.0052)	(0.0034)

Tabela 4.2 – The estimates of all parameters of the PHC model with censoring rate (20%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
$\boldsymbol{\phi}$	**0**	0.1185	0.0947	0.0894	0.0655	0.591	0.0343
		(0.0698)	(0.0459)	(0.0439)	(0.0316)	(0.0288)	(0.0211)
		0.1196	0.0959	0.0908	0.0669	0.0603	0.0351
		(0.0690)	(0.0463)	(0.0451)	(0.0323)	(0.0294)	(0.0208)
		0.1208	0.0981	0.0913	0.0670	0.0599	0.0357
		(0.0699)	(0.0469)	(0.0448)	(0.0327)	(0.0285)	(0.0215)
		0.1178	0.0939	0.0883	0.0648	0.0583	0.0346
		(0.0689)	(0.0453)	(0.0443)	(0.0318)	(0.0291)	(0.0206)
$\boldsymbol{\sigma}^2_{\psi}$	0.35	0.3245	0.3296	0.3311	0.3360	0.3373	0.3414
		(0.0429)	(0.0356)	(0.0334)	(0.0247)	(0.0218)	(0.0152)
		0.3236	0.3294	0.3309	0.3366	0.3369	0.3410
		(0.0425)	(0.0360)	(0.0340)	(0.0243)	(0.0224)	(0.0155)
		0.3243	0.3299	0.3306	0.3363	0.3370	0.3409
		(0.0431)	(0.0358)	(0.0339)	(0.0250)	(0.0216)	(0.0153)
		0.3240	0.3291	0.3314	0.3361	0.3367	0.3412
		(0.0430)	(0.0351)	(0.0334)	(0.0249)	(0.0222)	(0.0157)
σ^2_{ε}	0.25	0.2408	0.2427	0.2432	0.2449	0.2453	0.2466
		(0.0191)	(0.0148)	(0.0142)	(0.0102)	(0.0094)	(0.0062)
ϕ^*	$log(0.5)$	-0.6769	-0.6788	-0.6792	-0.6810	-0.6815	-0.6836
		(0.0353)	(0.0279)	(0.0255)	(0.0195)	(0.0169)	(0.0116)
$\sigma^2_{\psi*}$	0.3	0.2917	0.2933	0.2938	0.2951	0.2959	0.2972
		(0.0363)	(0.0340)	(0.0284)	(0.0262)	(0.0207)	(0.0187)
$\sigma^2_{\varepsilon*}$	0.2	0.1951	0.1959	0.1963	0.1973	0.1975	0.1984
		(0.0099)	(0.0081)	(0.0077)	(0.0061)	(0.0059)	(0.0041)

Tabela 4.3 – The estimates of all parameters of the PHC model with censoring rate (40%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
$\boldsymbol{\phi}$	0	0.1203	0.0971	0.0913	0.0682	0.0619	0.0368
		(0.0707)	(0.0468)	(0.0447)	(0.0323)	(0.0296)	(0.0219)
		0.1211	0.0983	0.0924	0.0691	0.0628	0.0379
		(0.0701)	(0.0470)	(0.0458)	(0.0330)	(0.0300)	(0.0216)
		0.1223	0.0994	0.0936	0.0697	0.0615	0.0374
		(0.0706)	(0.0475)	(0.0455)	(0.0333)	(0.0293)	(0.0224)
		0.1199	0.0964	0.0906	0.0675	0.0609	0.0363
		(0.0797)	(0.0462)	(0.0450)	(0.0324)	(0.0297)	(0.0213)
$\boldsymbol{\sigma}^2_{\psi}$	0.35	0.3267	0.3319	0.3331	0.3384	0.3394	0.3431
		(0.0436)	(0.0363)	(0.0342)	(0.0253)	(0.0226)	(0.0161)
		0.3258	0.3315	0.3328	0.3390	0.3390	0.3435
		(0.0431)	(0.0367)	(0.0348)	(0.0250)	(0.0231)	(0.0163)
		0.3261	0.3321	0.3335	0.3386	0.3396	0.3438
		(0.0438)	(0.0364)	(0.0345)	(0.0258)	(0.0224)	(0.0160)
		0.3263	0.3314	0.3334	0.3381	0.3393	0.3433
		(0.0440)	(0.0359)	(0.0340)	(0.0254)	(0.0229)	(0.0163)
σ^2_{ε}	0.25	0.2400	0.2419	0.2423	0.2441	0.2445	0.2458
		(0.0196)	(0.0154)	(0.0146)	(0.0105)	(0.0096)	(0.0065)
ϕ^*	$log(0.5)$	-0.6759	-0.6778	-0.6783	-0.6801	-0.6807	-0.6827
		(0.0359)	(0.0285)	(0.0261)	(0.0200)	(0.0174)	(0.0122)
$\sigma^2_{\psi*}$	0.3	0.2925	0.2940	0.2944	0.2958	0.2964	0.2978
		(0.0368)	(0.0347)	(0.0291)	(0.0270)	(0.0215)	(0.0194)
$\sigma^2_{\varepsilon*}$	0.2	0.1942	0.1950	0.1953	0.1965	0.1967	0.1978
		(0.0105)	(0.0090)	(0.0086)	(0.0071)	(0.0068)	(0.0049)

Tabela 4.4 – The estimates of all parameters of the SI-PHC model with censoring rate (10%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
τ_1	$-log3$	-1.1680	-1.1606	-1.1589	-1.1512	-1.1496	-1.1415
		(0.0556)	(0.0483)	(0.0467)	(0.0398)	(0.0382)	(0.0311)
τ_2	0	0.0513	0.0456	0.0453	0.0396	0.0384	0.0321
		(0.0448)	(0.0389)	(0.0372)	(0.0311)	(0.0296)	(0.0241)
τ_3	$log3$	1.1649	1.1572	1.1553	1.1460	1.1441	1.1364
		(0.0561)	(0.0473)	(0.0456)	(0.0376)	(0.0360)	(0.0289)
$\boldsymbol{\phi}$	$\mathbf{0}$	0.0446	0.0369	0.0352	0.0273	0.0259	0.0186
		(0.0351)	(0.0284)	(0.0216)	(0.0161)	(0.0152)	(0.0098)
		0.0451	0.0373	0.0357	0.0280	0.0264	0.0190
		(0.0356)	(0.0287)	(0.0218)	(0.0166)	(0.0154)	(0.0094)
		0.0442	0.0364	0.0349	0.0278	0.0261	0.0181
		(0.0359)	(0.0280)	(0.0213)	(0.0159)	(0.0150)	(0.0101)
		0.0455	0.0370	0.0353	0.0275	0.0266	0.0184
		(0.0360)	(0.0291)	(0.0223)	(0.0168)	(0.0157)	(0.0103)
$\boldsymbol{\sigma}^2_{\psi}$	0.35	0.3360	0.3385	0.3391	0.3414	0.3420	0.3446
		(0.0261)	(0.0216)	(0.0201)	(0.0152)	(0.0139)	(0.0087)
		0.3358	0.3381	0.3393	0.3416	0.3421	0.3445
		(0.0259)	(0.0213)	(0.0197)	(0.0149)	(0.0136)	(0.0089)
		0.3363	0.3384	0.3396	0.3410	0.3418	0.3448
		(0.0263)	(0.0214)	(0.0199)	(0.0151)	(0.0141)	(0.0080)
		0.3361	0.3380	0.3390	0.3411	0.3420	0.3442
		(0.0260)	(0.0219)	(0.0203)	(0.0155)	(0.0137)	(0.0083)
σ^2_{ε}	0.25	0.2453	0.2464	0.2467	0.2476	0.2479	0.2488
		(0.0060)	(0.0047)	(0.0044)	(0.0032)	(0.0030)	(0.0019)
ϕ^*	$log(0.5)$	-0.6805	-0.6827	-0.6832	-0.6851	-0.6855	-0.6873
		(0.0149)	(0.0126)	(0.0117)	(0.0100)	(0.0094)	(0.0078)
$\sigma^2_{\psi*}$	0.3	0.2932	0.2943	0.2947	0.2959	0.2961	0.2972
		(0.0195)	(0.0174)	(0.0178)	(0.0160)	(0.0163)	(0.0145)
$\sigma^2_{\varepsilon*}$	0.2	0.1980	0.1985	0.1983	0.1988	0.1989	0.1993
		(0.0052)	(0.0043)	(0.0040)	(0.0031)	(0.0029)	(0.0021)

Tabela 4.4 – The estimates of all parameters of the SI-PHC model with censoring rate (10%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
$\boldsymbol{\gamma}$	0	0.0549	0.0503	0.0486	0.0439	0.0423	0.0379
		(0.0435)	(0.0414)	(0.0410)	(0.0389)	(0.0383)	(0.0359)
		0.0538	0.0496	0.0490	0.0442	0.0419	0.0370
		(0.0429)	(0.0410)	(0.0407)	(0.0385)	(0.0382)	(0.0361)
		0.0534	0.0508	0.0492	0.0435	0.0412	0.0366
		(0.0431)	(0.0412)	(0.408)	(0.0384)	(0.0380)	(0.0362)
		0.0552	0.0500	0.0483	0.0431	0.0415	0.0375
		(0.0432)	(0.0409)	(0.0406)	(0.0387)	(0.0381)	(0.0358)
$\boldsymbol{\sigma}^2_\zeta$	0.3	0.2875	0.2893	0.2899	0.2921	0.2928	0.2944
		(0.0389)	(0.0346)	(0.0338)	(0.0295)	(0.0286)	(0.0239)
		0.2879	0.2889	0.2895	0.2919	0.2924	0.2946
		(0.0384)	(0.0347)	(0.0339)	(0.0301)	(0.0284)	(0.0236)
		0.2881	0.2891	0.2897	0.2925	0.2929	0.2940
		(0.0380)	(0.0341)	(0.0335)	(0.0297)	(0.0288)	(0.0235)
		0.2876	0.2890	0.2895	0.2921	0.2926	0.2941
		(0.0383)	(0.0339)	(0.0330)	(0.0298)	(0.0286)	(0.0237)
		0.2875	0.2894	0.2898	0.2926	0.2929	0.2947
		(0.0390)	(0.0344)	(0.0337)	(0.0296)	(0.0289)	(0.0241)
$\sigma^2_{\varepsilon_p}$	0.2	0.1934	0.1945	0.1949	0.1960	0.1963	0.1975
		(0.0071)	(0.0069)	(0.0066)	(0.0054)	(0.0050)	(0.0039)

Tabela 4.5 – The estimates of all parameters of the SI-PHC model with censoring rate (20%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
τ_1	$-log3$	-1.1689	-1.1613	-1.1596	-1.1518	-1.1503	-1.1421
		(0.0562)	(0.0489)	(0.0472)	(0.0403)	(0.0387)	(0.0315)
τ_2	0	0.0520	0.0463	0.0461	0.0404	0.0390	0.0326
		(0.0452)	(0.0393)	(0.0376)	(0.0314)	(0.0300)	(0.0244)
τ_3	$log3$	1.1655	1.1576	1.1561	1.1467	1.1446	1.1372
		(0.0566)	(0.0476)	(0.461)	(0.0380)	(0.0365)	(0.0296)
$\boldsymbol{\phi}$	$\mathbf{0}$	0.0459	0.0378	0.0363	0.0284	0.0270	0.0195
		(0.0356)	(0.0290)	(0.0221)	(0.0167)	(0.0158)	(0.0103)
		0.0460	0.0384	0.0366	0.0292	0.0276	0.0201
		(0.0363)	(0.0294)	(0.0224)	(0.0172)	(0.0161)	(0.0099)
		0.0451	0.0375	0.0361	0.0289	0.0272	0.0192
		(0.0364)	(0.0286)	(0.0219)	(0.0166)	(0.0157)	(0.0106)
		0.0463	0.0381	0.0365	0.0287	0.0279	0.0196
		(0.0366)	(0.0297)	(0.0228)	(0.0173)	(0.0162)	(0.0107)
$\boldsymbol{\sigma}^2_{\psi}$	0.35	0.3351	0.3376	0.3382	0.3405	0.3413	0.3437
		(0.0264)	(0.0218)	(0.0204)	(0.0154)	(0.0141)	(0.0090)
		0.3349	0.3372	0.3385	0.3407	0.3412	0.3439
		(0.0260)	(0.0216)	(0.0200)	(0.0153)	(0.0139)	(0.0091)
		0.3354	0.3374	0.3487	0.3402	0.3410	0.3438
		(0.0265)	(0.0217)	(0.0201)	(0.0155)	(0.0143)	(0.0083)
		0.3352	0.3373	0.3381	0.3401	0.3412	0.3436
		(0.0263)	(0.0222)	(0.0205)	(0.0158)	(0.0140)	(0.0086)
σ^2_{ε}	0.25	0.2447	0.2458	0.2462	0.2472	0.2475	0.2485
		(0.0063)	(0.0050)	(0.0047)	(0.0035)	(0.0032)	(0.0021)
ϕ^*	$log(0.5)$	-0.6799	-0.6822	-0.6826	-0.6846	-0.6851	-0.6869
		(0.0153)	(0.0129)	(0.0120)	(0.0102)	(0.0096)	(0.0080)
$\sigma^2_{\psi *}$	0.3	0.2928	0.2939	0.2944	0.2957	0.2959	0.2970
		(0.0198)	(0.0177)	(0.0180)	(0.0162)	(0.0165)	(0.0146)
$\sigma^2_{\varepsilon *}$	0.2	0.1978	0.1984	0.1981	0.1987	0.1988	0.1992
		(0.0054)	(0.0044)	(0.0042)	(0.0032)	(0.0031)	(0.0023)

Tabela 4.5 – The estimates of all parameters of the SI-PHC model with censoring rate (20%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
$\boldsymbol{\gamma}$	0	0.0553	0.0508	0.0491	0.0444	0.0426	0.0383
		(0.0437)	(0.0416)	(0.0412)	(0.0391)	(0.0385)	(0.0361)
		0.0544	0.0501	0.0494	0.0446	0.0425	0.0375
		(0.0431)	(0.0411)	(0.0409)	(0.0386)	(0.0383)	(0.0363)
		0.0540	0.0511	0.0495	0.0440	0.0419	0.0373
		(0.0433)	(0.0414)	(0.409)	(0.0385)	(0.0382)	(0.0363)
		0.0556	0.0504	0.0488	0.0438	0.0420	0.0379
		(0.0435)	(0.0412)	(0.0407)	(0.0389)	(0.0383)	(0.0361)
$\boldsymbol{\sigma}^2_{\zeta}$	0.3	0.2869	0.2885	0.2890	0.2909	0.2915	0.2936
		(0.0382)	(0.0342)	(0.0333)	(0.0291)	(0.0282)	(0.0235)
		0.2871	0.2883	0.2887	0.2906	0.2911	0.2933
		(0.0380)	(0.0344)	(0.0335)	(0.0295)	(0.0281)	(0.0234)
		0.2874	0.2886	0.2890	0.2912	0.2917	0.2935
		(0.0377)	(0.0338)	(0.0332)	(0.0293)	(0.0285)	(0.0231)
		0.2869	0.2884	0.2889	0.2908	0.2912	0.2930
		(0.0379)	(0.0337)	(0.0326)	(0.0296)	(0.0283)	(0.0234)
		0.2870	0.2889	0.2902	0.2913	0.2917	0.2935
		(0.0385)	(0.0341)	(0.0334)	(0.0293)	(0.0286)	(0.0238)
$\sigma^2_{\varepsilon_p}$	0.2	0.1929	0.1941	0.1944	0.1956	0.1958	0.1971
		(0.0073)	(0.0071)	(0.0068)	(0.0057)	(0.0053)	(0.0042)

Tabela 4.6 – The estimates of all parameters of the SI-PHC model with censoring rate (40%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
τ_1	$-log3$	-1.1700	-1.1622	-1.1605	-1.1527	-1.1510	-1.1428
		(0.0568)	(0.0496)	(0.0479)	(0.0410)	(0.0393)	(0.0321)
τ_2	0	0.0528	0.0470	0.0469	0.0411	0.0397	0.0331
		(0.0455)	(0.0397)	(0.0378)	(0.0316)	(0.0303)	(0.0246)
τ_3	$log3$	1.1663	1.1583	1.1568	1.1474	1.1453	1.1376
		(0.0572)	(0.0484)	(0.0465)	(0.0388)	(0.0372)	(0.0302)
$\boldsymbol{\phi}$	**0**	0.0472	0.0390	0.0372	0.0296	0.0281	0.0207
		(0.0363)	(0.0295)	(0.0227)	(0.0173)	(0.0164)	(0.0108)
		0.0469	0.0393	0.0377	0.0301	0.0285	0.0210
		(0.0369)	(0.0300)	(0.0231)	(0.0177)	(0.0168)	(0.0105)
		0.0462	0.0386	0.0370	0.0300	0.0283	0.0203
		(0.0371)	(0.0302)	(0.0224)	(0.0171)	(0.0164)	(0.0110)
		0.0474	0.0392	0.0374	0.0298	0.0288	0.0205
		(0.0372)	(0.0303)	(0.0233)	(0.0179)	(0.0169)	(0.0113)
$\boldsymbol{\sigma}^2_\psi$	0.35	0.3342	0.3368	0.3374	0.3396	0.3407	0.3428
		(0.0270)	(0.0225)	(0.0210)	(0.0160)	(0.0147)	(0.0096)
		0.3340	0.3361	0.3376	0.3398	0.3405	0.3430
		(0.0268)	(0.0222)	(0.0207)	(0.0158)	(0.0145)	(0.0098)
		0.3345	0.3365	0.3379	0.3394	0.3403	0.3429
		(0.0272)	(0.0221)	(0.0206)	(0.0161)	(0.0149)	(0.0093)
		0.3341	0.3362	0.3375	0.3392	0.3402	0.3428
		(0.0269)	(0.0228)	(0.0211)	(0.0163)	(0.0148)	(0.0092)
σ^2_ε	0.25	0.2441	0.2452	0.2456	0.2467	0.2470	0.2481
		(0.0068)	(0.0054)	(0.0051)	(0.0039)	(0.0035)	(0.0023)
ϕ^*	$log(0.5)$	-0.6791	-0.6813	-0.6716	-0.6837	-0.6843	-0.6861
		(0.0159)	(0.0135)	(0.0127)	(0.0106)	(0.0100)	(0.084)
$\sigma^2_{\psi*}$	0.3	0.2923	0.2934	0.2941	0.2953	0.2955	0.2964
		(0.0202)	(0.0181)	(0.0184)	(0.0165)	(0.0168)	(0.0149)
$\sigma^2_{\varepsilon*}$	0.2	0.1975	0.1981	0.1978	0.1984	0.1986	0.1990
		(0.0056)	(0.0047)	(0.0045)	(0.0034)	(0.0033)	(0.0025)

Tabela 4.6 – The estimates of all parameters of the SI-PHC model with censoring rate (40%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
$\boldsymbol{\gamma}$	0	0.0567	0.0521	0.0506	0.0456	0.0438	0.0397
		(0.0448)	(0.0425)	(0.0422)	(0.0400)	(0.0397)	(0.0372)
		0.0560	0.0515	0.0507	0.0457	0.0430	0.0391
		(0.0443)	(0.0422)	(0.0420)	(0.0398)	(0.0394)	(0.0375)
		0.0558	0.0520	0.0509	0.0455	0.0436	0.0393
		(0.0442)	(0.0425)	(0.0419)	(0.0395)	(0.0392)	(0.0374)
		0.0565	0.0519	0.0504	0.0453	0.0439	0.0396
		(0.0446)	(0.0421)	(0.0420)	(0.0401)	(0.0395)	(0.0372)
$\boldsymbol{\sigma}^2_\zeta$	0.3	0.2856	0.2870	0.2876	0.2893	0.2899	0.2918
		(0.0387)	(0.0346)	(0.0338)	(0.0294)	(0.0286)	(0.0240)
		0.2857	0.2868	0.2873	0.2891	0.2896	0.2920
		(0.0385)	(0.0349)	(0.0339)	(0.0299)	(0.0287)	(0.0239)
		0.2855	0.2869	0.2873	0.2890	0.2896	0.2923
		(0.0382)	(0.0342)	(0.0335)	(0.0296)	(0.0289)	(0.0236)
		0.2850	0.2865	0.2870	0.2891	0.2895	0.2919
		(0.0384)	(0.0341)	(0.0331)	(0.0299)	(0.0287)	(0.0237)
		0.2853	0.2867	0.2871	0.2894	0.2900	0.2924
		(0.0389)	(0.0346)	(0.0337)	(0.0297)	(0.0289)	(0.0240)
$\sigma^2_{\varepsilon_p}$	0.2	0.1919	0.1930	0.1935	0.1947	0.1949	0.1962
		(0.0080)	(0.0077)	(0.0074)	(0.0062)	(0.0058)	(0.0047)

Tabela 4.7 – Estimates of intercepts and coefficients of the model for two censoring rates

Censoring Rate	Variable	Estimates	Std. Error	P-value
25%	τ_1	1.4812	0.0941	0.0000
	τ_2	2.8267	0.0976	0.0000
	τ_3	3.2670	0.0990	0.0000
	Treatment	0.1843	0.0896	0.0971
50%	τ_1	1.6695	0.1067	0.0000
	τ_2	2.9870	0.1134	0.0000
	τ_3	3.4937	0.1186	0.0000
	Treatment	0.2163	0.0936	0.1016

5 Self-Modeling regression in the accelerated hazards mixture cure model

5.1 Introduction

Binary outcomes are common in biomedical research, where success may indicate that the patient is alive after treatment, develops no particular disease after exposure, or develops no complication after a surgical operation. In health services research a common binary outcome is the use or non-use of services. Logistic regression is a widely accepted method for describing the relationship between a binary or dichotomous outcome and a set of explanatory variables. It is used in many areas such as health care research and biomedical studies (Kramer et al., 1983, Tsutakawa, 1988, Cleary and Angel, 1984, Khuri et al., 1997).

Suppose the probability of a response for a subject i at a time-point j is conditional on the random (subject) effect, and so $p_{ij} = \Pr(y_{ij} = 1 | x_{ij}, z_{ij}),\ i = 1, \ldots, N,\ j = 1, \ldots, n_i$. The mixed-effects logistic regression model is

$$\log\left[\frac{p_{ij}}{1 - p_{ij}}\right] = \eta_{ij} \tag{5.1}$$

where

$$\eta_{ij} = x_{ij}^{'}\boldsymbol{\beta} + z_{ij}^{'}\upsilon_i. \tag{5.2}$$

x_{ij} is the $(p+1) \times 1$ covariate vector (including the intercept), and z_{ij} is design vector for the r random effects, both vectors being for the jth time-point nested within subject i. Also, $\boldsymbol{\beta}$ is the $(p+1) \times 1$ vector of unknown fixed regression parameters. Let $\upsilon = T\theta$, where $TT^{'} = \Sigma_\upsilon$ is the Cholesky factorization of random effect variance covariance matrix Σ_υ.

Survival models incorporating a cure fraction, often called cure models, are often used in analyzing data from cancer clinical trials. In these clinical trials, a group of patients respond favorably to the treatment. They usually have long-term censored survival times and tend to be regarded as cured. Since the mixture nature of patients in the trials, the most popular type of cure models is the mixture model. Farewell (1982), Farewell (1986) used mixture models for the analysis of survival data with Long-Term Survivors. Kuk and Chen (1992) considered the semiparametric logistic proportional hazard mixture model.

The accelerated hazards (AH) model was first considered by Chen (2000) and Chen and Wang (2001) for survival data without a fraction of cured subjects. The AH model is useful to model the situation when the effect of treatments or other covariates is gradually released on the failure time distribution. Let Y be the indicator variable for an uncured patient with $Y = 1$ if the patient is uncured and 0 if cured, T be the failure time of a patient. Define $\pi = P(Y = 1), S(t) = P(T > t)$ and $S_u(t) = P(T > t|Y = 1)$. That is, π is the probability of being uncured, and $S(t)$ and $S_u(t)$ are the survival functions of the failure time of a patient and the failure time of an uncured patient respectively. The AH model is given by

$$S(t|\boldsymbol{x}, \boldsymbol{z}) = \pi(\boldsymbol{z}) S_u(t|\boldsymbol{x}) + [1 - \pi(\boldsymbol{z})], \tag{5.3}$$

where $\boldsymbol{x}$ and $\boldsymbol{z}$ are two sets of covariates that have effects on π and $S_u(t)$. The advantage of the mixture cure model is that the proportion of cured patients and the survival distribution of uncured patients are modeled separately and the interpretation of the parameters of $\boldsymbol{x}$ and $\boldsymbol{z}$ in the model is straightforward. To specify the effects of $\boldsymbol{z}$ on π, the most common method is the logit link function

$$\pi(z) = \frac{\exp(\gamma' z)}{1 + \exp(\gamma' z)}, \tag{5.4}$$

where γ is a vector of unknown parameters. In fact, equation 5.4 is a simple form of equation 5.1 without random effects. To allow a gradual effect of covariates on the failure time of uncured patients, Zhang and Peng (2009)proposed to model $S_u(t)$ in the mixture cure model 5.3 by the AH model proposed by Chen and Wang (2001). That is

$$S_u(t|x) = [S_0(t \exp(\beta^T x))]^{\exp(-\beta^T x)}. \tag{5.5}$$

where $S_0(t)$ is an arbitrary baseline survival function. We refer to the model specified by Eqs. 5.3–5.5 as the AH mixture cure (AHMC) model. The AHMCM allows covariate effects on the failure time distribution of uncured patients to be negligible at time zero and to increase as time goes by. In some cancer treatments when the treat effect increases gradually from zero, the AHMCM is particularly useful.

In this work, we propose a model for longitudinal data where patients suffering from a common disease exhibits a similar shape even though each patient's response varies substantially over time. In longitudinal analysis for continuous data, the response is a continuous curve measured over time. We will consider the Self-Modeling Regression (SEMOR) model introduced by Lawton, Sylvestre and Maggio (1972)

$$y_{ij} = \pi_i \left\{ \mu_0 \left[\kappa_i(t_{ij}) \right] \right\} + e_{ij}, \tag{5.6}$$

where y_{ij} is the response for curve i, $i = 1, \ldots, N$, measured at n_i times, t_{ij}. $\pi_i(x)$ is a monotone inverse link transforming the regression function and $\kappa_i(x)$ is a monotone transformation of the time axis. μ_0 is a shape function that is common to all the curves, and e_{ij} are errors. This paper will focus on nonparametric modeling of μ_0 and parametric modeling of $\pi_i(x)$ and $\kappa_i(x)$ with known correlation structure for e_{ij}. The Lawton model exhibits a common shape function for every individual quite naturally. It becomes a challenge then to develop a shape invariant model when the response variable ceases to be continuous. In fact, in medical studies, often we come across binary outcome observed longitudinally. Typically this scenario happens in our example on respiratory illness study where the stage of illness for each patient is observed over different points of time. In section 5.2, we will consider the Self-Modeling model defined for the conditional probabilities for a category of an outcome. Therefore, unlike equation 5.6, in our model there is no relation between the observed response and parameters directly.

A AHMCM with random effects is proposed in Section 5.3. We extend the AHMCM such that the extended model can be applied for the time of occurrence of an event when Self-Modeling binary model is used for the probability of occurrence of an event. In section 5.4, parameter estimation with using the Monte Carlo method in Newton-Raphson and EM algorithms are introduced. Section 5.5 discuss simulation results of the estimation methods. As an application of the model, we apply the model to the respiratory illness data set in Section 5.6.

5.2 Self-Modeling binary model

We give special attention to shape invariant (SI) Model (Lawton, Sylvestre and Maggio, 1972, Altman and Villarreal, 2004) and apply the SI model for equation (5.2), so

$$\eta_{ij} = \alpha_{0i} + \exp\left(\alpha_{1i}\right)\,\mu_0\left(t^*_{ij}\right), \tag{5.7}$$

where $t^*_{ij} = \beta_{0i} + \exp\left(\beta_{1i}\right) t_{ij}$. Here α_{0i}, α_{1i}, β_{0i} and β_{1i} are unknown parameters which may be functions of observed covariates and μ_0 is a shape function which is common to all subjects. Therefore, with equations (5.1), (5.2) and (5.7) we have

$$\eta_{ij} = log\left[\frac{p_{ij}}{1-p_{ij}}\right] = \alpha_{0i} + \exp\left(\alpha_{1i}\right)\,\mu_0\left(t^*_{ij}\right). \tag{5.8}$$

If one has physical or theoretical justification to pre-specify $\mu_0\left(t^*_{ij}\right)$ parametrically, this is just a special case of nonlinear regression. The semi-parametric SEMOR model allows flexible modeling by estimating $\mu_0\left(t^*_{ij}\right)$ non-parametrically. Several different approaches have been studied in fitting the SI model. Lawton, Sylvestre and Maggio (1972), Kneip and Gasser (1988), Kneip and Engel (1995) considered

$$\boldsymbol{\theta}_i = (\alpha_{0i}, \alpha_{1i}, \beta_{0i}, \beta_{1i})', \tag{5.9}$$

we consider a mixed model

$$\boldsymbol{\theta}_i = \mathbf{X}_i \boldsymbol{\phi} + \mathbf{Z}_i \boldsymbol{\psi}_i + \boldsymbol{\varepsilon}_i, \tag{5.10}$$

where $\boldsymbol{Z}_i$ is the design (or covariate) matrix for the random effect vector $\boldsymbol{\psi}_i$. If μ_0 is a known parametric function, and if we assume that $\boldsymbol{\psi}_i$, $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{e}_{ij}$ are normally distributed, then equation (5.7) is a parametric nonlinear mixed model and the model can be fitted using maximum likelihood using standard software (e.g., Lindstrom and D.M., 1988).

Here, we fit μ_0 using the penalized spline model of Ruppert and Carroll (1997). Use of the penalized spline method with penalty chosen by generalized maximum likelihood (Wahba, 1985) is equivalent to fitting the model

$$\mu_0(t^*_{ij}) = \mathbf{U}\boldsymbol{\gamma} + \mathbf{V}\boldsymbol{\zeta}, \tag{5.11}$$

where $\mu_0(t^*_{ij})$ is the vector of means at the transformed times, $\mathbf{U}$ is a design matrix for a cubic polynomial in t^*_{ij}, $\mathbf{V}$ is a design matrix for cubics in t^*_{ij} which are left-truncated at the knot, $\boldsymbol{\gamma}$ is a vector of unknown parameters and $\boldsymbol{\zeta}$ is normally distributed with zero mean and covariance matrix $\boldsymbol{\Sigma}_\zeta$.

5.3 The SI-AHMC model

Suppose a binary variable y_{ij} where $y_{ij} = 1$ indicates that an individual will experience a particular event and $y_{ij} = 0$ indicates that the individual will never experience the event. let $O = \{t_{ij},\ \delta_{ij},\ x^*_{ij}, z^*_{ij}, x_{ij}, z_{ij}\}$ denote the observed data for the ith individual, $i = 1, \ldots, N$ on time-point j, $j = 1, \ldots, n_i$, where (t_{ij}, δ_{ij}) denote the observed survival time and censoring indicator with 1 if t_{ij} is uncensored (i.e., observed) and 0 if censored. x^*_{ij} and z^*_{ij} be the covariates of the jth observation within the ith individual that may affect the survival time distribution of uncured individuals and the cure proportion. x_{ij} and z_{ij} are covariate vector and design vector for the random effect ψ_i in the logistic model for p_{ij}. Let $h_u\big(t_{ij}|y_{ij} = 1, x^*_{ij}, z^*_{ij}\big)$ and $S_u\big(t_{ij}|y_{ij} = 1, x^*_{ij}, z^*_{ij}\big)$ be the hazard function and the survival function for uncured subjects. The SI-AHMC model can be written as

$$h_u\big(t_{ij}|y_{ij} = 1, x^*_{ij}, z^*_{ij}\big) = h_0\big(t_{ij}\exp(x^*_{ij}\phi^* + z^*_{ij}\psi^*_i)\big) \tag{5.12}$$

or

$$S_u\big(t_{ij}|y_{ij}=1,x_{ij}^*,z_{ij}^*\big)=\Big[S_0\big(t_{ij}\exp(x_{ij}^*\phi^*+z_{ij}^*\psi_i^*)\big)\Big]^{\exp\left[-(x_{ij}^*\phi^*+z_{ij}^*\psi_i^*)\right]}. \tag{5.13}$$

$h_0(\cdot)$ and $S_0(\cdot)$ are respectively arbitrary baseline hazard and survival functions for uncured individuals. Therefore we can write the Cox PH cure model as a mixture model in terms of the survival function

$$S\left(t_{ij}|y_{ij}=1,x_{ij}^*,z_{ij}^*\right)=(1-p_{ij})+p_{ij}S_u\big(t_{ij}|y_{ij}=1,x_{ij}^*,z_{ij}^*\big). \tag{5.14}$$

where p_{ij} is given by equation (5.8) as

$$p_{ij}=\frac{\exp\left(\eta_{ij}\right)}{1+\exp\left(\eta_{ij}\right)}. \tag{5.15}$$

Note for creating the AHMC model, we use the logit link function defining in equations (5.1) and (5.2). In this paper, we will consider the Weibull distribution. Under the Weibull baseline assumption, $h_0(t)=(a/b^a)\,t^{(a-1)}$ and $S_0(t)=\exp\left[-(t/b)^a\right]$ where $a,b>0$ are the shape and scale parameters of the distribution, respectively. The marginal likelihood for the SI-AHMC model is

$$\begin{aligned}L\left(\boldsymbol{\theta}\right)=\prod_{i=1}^{N}\int\Bigg\{\prod_{j=1}^{n_i}&\Big[p_{ij}f_u\big(t_{ij}|y_{ij}=1,x_{ij}^*,z_{ij}^*\big)\Big]^{\delta_{ij}}\\&\times\Big[(1-p_{ij})+p_{ij}S_u\big(t_{ij}|y_{ij}=1,x_{ij}^*,z_{ij}^*\big)\Big]^{(1-\delta_{ij})}\\&\times f\left(\boldsymbol{\theta}_i^*|\boldsymbol{\psi}_i^*\right)f\left(\boldsymbol{\psi}_i^*|\sigma^2_{\psi_i^*}\right)f\left(\boldsymbol{\theta}_i|\boldsymbol{\psi}_i\right)f\left(\boldsymbol{\psi}_i|\sigma^2_{\psi_i}\right)\Bigg\}d\boldsymbol{\theta}_i\;d\boldsymbol{\psi}_i\;d\boldsymbol{\theta}_i^*\;d\boldsymbol{\psi}_i^*,\end{aligned} \tag{5.16}$$

where $\boldsymbol{\theta}=(\boldsymbol{\phi}',\,\sigma^2_{\boldsymbol{\psi}},\,\sigma^2_{\boldsymbol{\varepsilon}},\,\boldsymbol{\phi}^{*'},\,\sigma^2_{\boldsymbol{\psi}^*},\,\sigma^2_{\boldsymbol{\varepsilon}^*})'$ and $f_u\big(t_{ij}|y_{ij}=1,x_{ij}^*,z_{ij}^*\big)$ is the probability density function for uncured subjects. Since $\boldsymbol{\psi}_i^*$ and $\boldsymbol{\varepsilon}^*$ have $N\left(\mathbf{0},\boldsymbol{\Sigma}\left(\sigma^2_{\psi_i^*}\right)\right)$ and $N\left(0,\sigma^2_{\varepsilon_i^*}\right)$ respectively, therefore the complete log-likelihood is

$$\begin{aligned}l_c=\sum_{i=1}^{N}l_{ic}=\sum_{i=1}^{N}\sum_{j=1}^{n_i}\delta_{ij}&\Big[\log p_{ij}+\log f_u\big(t_{ij}|y_{ij}=1,x_{ij}^*\phi^*+z_{ij}^*\psi_i^*\big)\Big]\\&+(1-\delta_{ij})\log\Big[(1-p_{ij})+p_{ij}S_u\big(t_{ij}|y_{ij}=1,x_{ij}^*\phi^*+z_{ij}^*\psi_i^*\big)\Big]\\&+\log\Big[\boldsymbol{\Phi}\left(\boldsymbol{\theta}_i^*|\mathbf{X}_i^*\boldsymbol{\phi}^*+\mathbf{Z}_i^*\boldsymbol{\psi}_i^*,\;\sigma^2_{\varepsilon_i^*}I\right)\Big]+\log\Big[\boldsymbol{\Phi}\left(\boldsymbol{\psi}_i^*|0,\;\sigma^2_{\psi_i^*}\right)\Big]\\&+\log\Big[\boldsymbol{\Phi}\left(\boldsymbol{\theta}_i|\mathbf{X}_i\boldsymbol{\phi}+\mathbf{Z}_i\boldsymbol{\psi}_i,\;\sigma^2_{\varepsilon_i}I_4\right)\Big]+\log\Big[\boldsymbol{\Phi}\left(\boldsymbol{\psi}_i|0,\;\sigma^2_{\psi_i}\right)\Big)\Big],\end{aligned} \tag{5.17}$$

where Φ is a probability density function.

5.4 Estimation method

In this section we propose an algorithm for parameter estimation in the SI-AHMC model. Essentially steps 0-2 are considered for predicting the shape function μ_0 and step 3 provides estimate of the parameters for the SI-AHMC model through MCNR and MCEM approaches.

Step 0: Choose initial estimates of $\phi^{(s)}, \phi^{*(s)}$ and $\boldsymbol{\gamma}^{(s)}$, variances $\sigma_{\boldsymbol{\psi}}^{2\,(s)}, \sigma_{\varepsilon}^{2(s)}, \sigma_{\boldsymbol{\psi}*}^{2\;\;(s)}, \sigma_{\boldsymbol{\varepsilon}*}^{2\;\;(s)}$, $\sigma_{\boldsymbol{\zeta}}^{2(s)}$ and $\sigma_{\varepsilon_p}^2$, random effects $\boldsymbol{\psi}_i{}^{(s)} \sim N_4\left(\mathbf{0}, \boldsymbol{\Sigma}\left(\sigma_{\boldsymbol{\psi}}^{2\,(s)}\right)\right)$ and $\boldsymbol{\psi}_i^{*(s)} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}\left(\sigma_{\boldsymbol{\psi}*}^{2\;\;(s)}\right)\right)$, errors $\boldsymbol{\varepsilon}_i{}^{(s)} \sim N_4\left(\mathbf{0}, \sigma_{\varepsilon}^{2(s)} I_4\right)$ and $\boldsymbol{\varepsilon}_i^{*(s)} \sim N\left(0, \sigma_{\boldsymbol{\varepsilon}*}^{2\;\;(s)}\right)$. Set $s = 0$.

Step 1: Compute $\boldsymbol{\theta}_i^{*(s)} = \mathbf{X}_i^* \boldsymbol{\phi}^{*(s)} + \mathbf{Z}_i^* \boldsymbol{\psi}_i^{*(s)} + \boldsymbol{\varepsilon}_i^{*(s)}$ and $\boldsymbol{\theta}_i^{(s)} = \mathbf{X}_i \boldsymbol{\phi}^{(s)} + \mathbf{Z}_i \boldsymbol{\psi}_i^{(s)} + \boldsymbol{\varepsilon}_i^{(s)}$ and extract $\alpha_{0i}^{(s)}$, $\alpha_{1i}^{(s)}, \beta_{0i}^{(s)}$ and $\beta_{1i}^{(s)}$

Step 1.1: $t_{ij}^{*\,(s)} = \beta_{0i}^{(s)} + \exp\left(\beta_{1i}^{(s)}\right) t_{ij}$,

Step 1.2: $\mathbf{U}^{(s)} = \begin{bmatrix} 1 & t_{i1}^{*\,(s)} & t_{i1}^{*2(s)} & t_{i1}^{*3(s)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_{in_i}^{*\,(s)} & t_{in_i}^{*2\,(s)} & t_{in_i}^{*3\,(s)} \end{bmatrix}$, $\mathbf{V}^{(s)} = \begin{bmatrix} \left(t_{i1}^{*\,(s)} - \xi_1\right)^3 & \dots & \left(t_{i1}^{*\,(s)} - \xi_k\right)^3 \\ \vdots & \vdots & \vdots \\ \left(t_{in_i}^{*\,(s)} - \xi_1\right)^3 & \dots & \left(t_{in_i}^{*\,(s)} - \xi_k\right)^3 \end{bmatrix}$,

$$\mathbf{C}^{(s)} = \begin{bmatrix}\mathbf{U}^{(s)} & \mathbf{V}^{(s)}\end{bmatrix}.$$

Step 1.3: The pth degree of penalized spline model $\lambda^{p(s)} = \sigma_{\varepsilon_p}^{2\;\;(s)} / \sigma_{\boldsymbol{\zeta}}^{2(s)}$. $\mathbf{w}^{(s)} = \mathbf{U}^{(s)} \widehat{\boldsymbol{\gamma}}^{(s)} + \mathbf{V}^{(s)} \widehat{\boldsymbol{\zeta}}^{(s)} + \widehat{\boldsymbol{\varepsilon}}_p^{(s)}$.

Step 1.4: $\widehat{\mu}_0^{*(s)} = \mathbf{C}^{(s)} \left(\mathbf{C}'^{(s)} \mathbf{C}^{(s)} + \lambda^{p(s)} \mathbf{D}\right)^{-1} \mathbf{C}'^{(s)} \mathbf{w}^{(s)}$, $D = diag(0_{p+1}; I_k)$.

Step 2: Using linear mixed model estimation, estimate $\boldsymbol{\gamma}^{(s+1)}$ and $\boldsymbol{\zeta}^{(s+1)}\left(\sigma^{2(s+1)}_{\zeta}\right)$ by fitting

$$\widehat{\mu}_0^{*(s)} = \mathbf{U}^{(s)} \boldsymbol{\gamma} + \mathbf{V}^{(s)} \boldsymbol{\zeta} + \boldsymbol{\varepsilon}_p.$$

Step 3: Using nonlinear mixed model estimation, estimate $\boldsymbol{\phi}^{(s+1)}$, $\sigma_{\boldsymbol{\psi}}^{2\,(s+1)}$, $\sigma_{\varepsilon}^{2(s+1)}$, $\boldsymbol{\phi}^{*(s+1)}$, $\sigma_{\boldsymbol{\psi}*}^{2\;\;(s+1)}$ and $\sigma_{\boldsymbol{\varepsilon}*}^{2\;\;(s+1)}$ by fitting the model

$$\eta_{ij} = \alpha_{0i} + \exp(\alpha_{1i})\; \mu_0\Big(\beta_{0i} + \exp(\beta_{1i}) t_{ij}\Big)$$

,

$$p_{ij} = \frac{exp(\eta_{ij})}{1 + exp(\eta_{ij})}$$

,

$$\boldsymbol{\theta}_i = \mathbf{X}_i \boldsymbol{\phi} + \mathbf{Z}_i \boldsymbol{\psi}_i + \boldsymbol{\varepsilon}_i$$

,

$$\boldsymbol{\theta}_i^* = \mathbf{X}_i^* \boldsymbol{\phi}^* + \mathbf{Z}_i^* \boldsymbol{\psi}_i^* + \boldsymbol{\varepsilon}_i^*$$

Conditional on $\widehat{\mu}_0^{*(s)}$.

Step 4: Check for convergence. If the algorithm has converged, then stop. Otherwise, increase the iteration counter s by one, and return to step 1.

5.4.1 MCEMNR algorithms

In the current setup, the E-M steps are as follows

E-Step

The $(s+1)$th step computes the conditional expectation of l_{ic}

$$Q\big(\boldsymbol{\theta}\,\big|\boldsymbol{\theta}^{(s)}\big) = \sum_{i=1}^{N} E_{\boldsymbol{\theta}_i}\big[l_{ic}|D_i, \boldsymbol{\theta}^{(s)}\big], \tag{5.18}$$

where $D_i = (y_i,\ \mathbf{X}_i,\ \mathbf{Z}_i,\ \mathbf{X}_i^*,\ \mathbf{Z}_i^*)$. An alternative is to replace the E step with Monte Carlo approximations constructed using a sample $l_{ic}^{(1)}, \ldots, l_{ic}^{(M_i)}$

$$Q\big(\boldsymbol{\theta}\,\big|\boldsymbol{\theta}^{(s)}\big) \approx \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} l_{ic}^{(m)}\left(\boldsymbol{\theta} \mid D_i, \boldsymbol{\theta}^{(s)}\right). \tag{5.19}$$

M_i is the number of iterations in the Monte Carlo method. We assume without loss of generality, the number of iterations $M_i = M\ \forall\ i$.

M-Step

The $(s+1)$th step then finds $\boldsymbol{\theta}^{(s+1)}$ as the maximizer of $Q\big(\boldsymbol{\theta}\,\big|\boldsymbol{\theta}^{(s)}\big)$:

$$Q\left(\boldsymbol{\theta}^{(s+1)} \mid \boldsymbol{\theta}^{(s)}\right) \geqslant Q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(s)}\right), \tag{5.20}$$

for all $\boldsymbol{\theta}$ in the parameter space. In principle, the M step is carried out by solving the score equations

$$\frac{\partial}{\partial \boldsymbol{\theta}} Q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(s)}\right) = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \frac{\partial}{\partial \boldsymbol{\theta}} l_{ic}^{(m)}\left(\boldsymbol{\theta} \mid D_i, \boldsymbol{\theta}^{(s)}\right) = 0, \quad for\ \boldsymbol{\theta}. \tag{5.21}$$

The essence of the EM algorithm is that increasing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)})$ forces an increase in the log-likelihood of the observed data. To obtain a satisfactory accuracy, the MC sample size

M needs to be a large number. For example, to obtain two decimal digits of accuracy, $M \geqslant 10\ 000$ is required. Solution for θ in (5.21) can be obtained by NR procedure.To solve for $\boldsymbol{\theta}$, we proceed through NR steps described below.

NR-Step

With $\boldsymbol{\theta}^{(s)} = (\boldsymbol{\phi}^{(s)},\ \sigma_{\psi}^{2\,(s)},\ \sigma_{\varepsilon}^{2(s)},\ \boldsymbol{\phi}^{*(s)},\ {\sigma_{\psi*}^{2}}^{(s)},\ {\sigma_{\varepsilon*}^{2}}^{(s)})$, Newton Raphson procedure is applied to estimate $\boldsymbol{\theta}$ at the $(s+1)$ iteration by using equation (20) as follows.

$$\widehat{\boldsymbol{\theta}}^{(s+1)} = \widehat{\boldsymbol{\theta}}^{(s)} + [V^{-1}\hat{S}(\boldsymbol{\theta})]_{|\widehat{\boldsymbol{\theta}}^{(s)}}\ , \tag{5.22}$$

where

$$\hat{S}\left(\boldsymbol{\theta}\right) = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left[\frac{\partial l_{ic}^{(m)}}{\partial \boldsymbol{\theta}}\right]_{|\widehat{\boldsymbol{\theta}}^{(s)}}, \tag{5.23}$$

and

$$V = \frac{\partial \hat{S}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}_{|\widehat{\boldsymbol{\theta}}^{(s)}} = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \frac{\partial}{\partial \boldsymbol{\theta}^T} \left[\frac{\partial l_{ic}^{(m)}}{\partial \boldsymbol{\theta}}\right]_{|\widehat{\boldsymbol{\theta}}^{(s)}}. \tag{5.24}$$

This process is repeated until $\left\|\widehat{\boldsymbol{\theta}}^{(s+1)} - \widehat{\boldsymbol{\theta}}^{(s)}\right\| \leqslant \epsilon \downarrow 0$. See Apendix **??** for more details **??**

5.5 Simulation

In this section, we use simulated data sets to compare the performances of the AHMC model with random effect and the SI-AHMC model by the MCNREM algorithm. The sample size is assumed to be $N = 100, 250, 500$ for $i = 1, .., N$ and $n = 10, 25$ for $j = 1, .., n$. The number of iterations in the Monte Carlo is 10000. The true values and the estimates of all parameters of the AHMCM and the SI-AHMCM by the MCNREM algorithm are displayed in Tables 5.1-5.3 and 5.4-5.6 respectively. Note that both for the AHMCM and the SI-AHMCM model, the true values of the regression parameter vector $\boldsymbol{\phi}$ and $\boldsymbol{\phi}^*$ respectively have been taken to be a null vector and $log(0.5)$. Also the dispersion matrices of the random vectors $\boldsymbol{\psi}_i$ and $\boldsymbol{\psi}_i^*$ are assumed to be $\sigma_{\psi}^2 I_4$ and $\sigma_{\psi*}^2$ for both models with the true value of $\sigma_{\psi}^2 = 0.35$ and $\sigma_{\psi*}^2 = 0.3$ respectively. The baseline hazard function $h_0(t)$ is assumed to be Weibull distribution with $a = 0.5$ and $b = 1.5$. The censoring time is generated from a uniform distribution in $(0, r)$ with proper values of r so that the corresponding censoring rates are about 10%, 20% and 40%. The standard errors (SE) based on generated random samples of different sizes are also computed for each estimator. All the computations are carried out in MATLAB.

Tables 5.1-5.3 and 5.4-5.6 respectively give us an idea about the performance of the estimators under the AHMC model and the SI-AHMC model of interest. The tables show that the estimates from the proposed model tend to have smaller variances than those from the AHMC model. In the tables, the results also show that a larger sample size improves the standard error estimation. From Table 5.4-5.6, it is evident that performance of the estimates is quite satisfactory for the SI-AHMC model built up for longitudinal data.

5.6 An Application to respiratory illness Data

Respiratory disease encompasses pathological conditions affecting the organs and tissues that make gas exchange possible in higher organisms, and includes conditions of the upper respiratory tract, trachea, bronchi, bronchioles, alveoli, pleura and pleural cavity, and the nerves and muscles of breathing. Respiratory diseases range from mild and self-limiting, such as the common cold, to life-threatening entities like bacterial pneumonia, pulmonary embolism, and lung cancer. Respiratory disease is a common and significant cause of illness and death around the world. In 2012, respiratory conditions were the most frequent reasons for hospital stays among children. Therefore, the data respiratory illness are important for researchers, clinicians, policy makers, and citizens in understanding this disease.

Davis (1991) displayed the raw data from a clinical trial comparing two treatments for a respiratory illness. In each of two centres, eligible patients were randomly assigned to active treatment or placebo. During treatment, respiratory status (categorized here as 0 = poor, 1 = good) was determined at four visits. Potential covariates were centre and sex (all dichotomous), as well as age (in years) at the time of study entry. There were 111 patients (54 active, 57 placebo) with no missing data for responses or covariates. The detailed analysis is based on SI-AHMC model where we consider the respiratory status as a response variable. We defined censoring rates 25% and 40% for the data. For choosing the best model in this example, we use the deviance which is the difference between the log-likelihood of the fitted model and the maximum possible log-likelihood. Table 5.7 gives the values of the deviance.

In view of the deviance values observed in table 5.7, we carry out our analysis. Table 5.9 shows the estimates of intercepts and coeffcients for the sixth model in table 5.7. Under censoring rate 20%, the p-value of 0.0017, 0.0001 and 0.0004 for Treatment, Age and Center respectively indicate that these factors are significant on the odds of the post-surgical recovery of a patient being less than or equal to a certain value versus being greater than that value. The p-values of 0.5129 Sex of a patient indicates that the factor is not significant. In another censoring rate we have same results in significance of estimators.

The self generating model with respect to the sixth model with censoring rate 20% is

$$\begin{aligned}\hat{\boldsymbol{\theta}}_i &= \left(\hat{\alpha}_{0i}, \hat{\alpha}_{1i}, \hat{\beta}_{0i}, \hat{\beta}_{1i}\right)' \\ &= 0.7094Treatment_i' - 0.0229Age_i' - 0.1269Sex_i' + 0.6302Center_i',\end{aligned}$$

$$log\left[\frac{p_{ij}}{1-p_{ij}}\right] = \alpha_{0i} + \exp\left(\alpha_{1i}\right)\, \mu_0\left(t_{ij}^*\right),$$

Tabela 5.1 – The estimates of all parameters of the AHMC model with censoring rate (10%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
$\boldsymbol{\phi}$	**0**	0.0633	0.0591	0.576	0.0528	0.0495	0.0447
		(0.0661)	(0.0418)	(0.0410)	(0.0300)	(0.0267)	(0.0189)
		0.0641	0.0599	0.0571	0.0526	0.0491	0.0441
		(0.0645)	(0.0417)	(0.0406)	(0.0303)	(0.0269)	(0.0191)
		0.0639	0.0587	0.0569	0.0524	0.0488	0.0439
		(0.0639)	(0.0421)	(0.0411)	(0.0300)	(0.0266)	(0.0191)
		0.0635	0.0595	0.0577	0.0531	0.0499	0.0451
		(0.0641)	(0.0421)	(0.0419)	(0.0299)	(0.0264)	(0.0190)
$\boldsymbol{\sigma}^2_\psi$	0.35	0.3124	0.3186	0.3203	0.3279	0.3298	0.3358
		(0.0405)	(0.0371)	(0.0359)	(0.0310)	(0.0298)	(0.0251)
		0.3129	0.3179	0.3210	0.3283	0.3302	0.3364
		(0.0411)	(0.0372)	(0.0356)	(0.0311)	(0.0294)	(0.0253)
		0.3119	0.3191	0.3204	0.3281	0.3303	0.3368
		(0.0409)	(0.0378)	(0.0355)	(0.0316)	(0.0297)	(0.0254)
		0.3123	0.3183	0.3209	0.3288	0.3307	0.3371
		(0.0415)	(0.0375)	(0.0360)	(0.0314)	(0.0296)	(0.0255)
σ^2_ε	0.25	0.2461	0.2470	0.2472	0.2478	0.2480	0.2486
		(0.0101)	(0.0066)	(0.0065)	(0.0049)	(0.0043)	(0.0031)
ϕ^*	$log(0.5)$	-0.6823	-0.6844	-0.6850	-0.6862	-0.6867	-0.6880
		(0.0323)	(0.0211)	(0.0209)	(0.0158)	(0.0142)	(0.0099)
$\sigma^2_{\psi*}$	0.3	0.2890	0.2913	0.2921	0.2939	0.2946	0.2957
		(0.0366)	(0.0329)	(0.0255)	(0.0240)	(0.0193)	(0.0174)
$\sigma^2_{\varepsilon*}$	0.2	0.1978	0.1981	0.1982	0.1985	0.1987	0.1989
		(0.0086)	(0.0055)	(0.0054)	(0.0039)	(0.0035)	(0.0025)

Tabela 5.2 – The estimates of all parameters of the AHMC model with censoring rate (20%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
$\boldsymbol{\phi}$	**0**	0.0638	0.0594	0.0580	0.0531	0.0499	0.455
		(0.0659)	(0.0420)	(0.0412)	(0.0301)	(0.0268)	(0.0192)
		0.0647	0.0603	0.0575	0.0529	0.0502	0.0450
		(0.0644)	(0.0419)	(0.0407)	(0.0305)	(0.0271)	(0.0193)
		0.0644	0.0590	0.0571	0.0525	0.0501	0.0446
		(0.0638)	(0.0422)	(0.0412)	(0.0301)	(0.0268)	(0.0192)
		0.0640	0.0609	0.0581	0.0534	0.0505	0.0456
		(0.0639)	(0.0410)	(0.0422)	(0.0300)	(0.0269)	(0.0191)
$\boldsymbol{\sigma}^2_{\psi}$	0.35	0.3108	0.3167	0.3184	0.3249	0.3269	0.3330
		(0.0419)	(0.0392)	(0.0376)	(0.0331)	(0.0314)	(0.0262)
		0.3111	0.3164	0.3187	0.3243	0.3275	0.3334
		(0.0421)	(0.0393)	(0.0377)	(0.0329)	(0.0316)	(0.0263)
		0.3109	0.3172	0.3188	0.3240	0.3264	0.3339
		(0.0423)	(0.0398)	(0.0379)	(0.0333)	(0.0315)	(0.0261)
		0.3107	0.3166	0.3193	0.3247	0.3273	0.3336
		(0.0418)	(0.0399)	(0.0381)	(0.0332)	(0.0314)	(0.0264)
σ^2_{ε}	0.25	0.2459	0.2467	0.2469	0.2477	0.2479	0.2485
		(0.0103)	(0.0068)	(0.0067)	(0.0050)	(0.0044)	(0.0032)
ϕ^*	$log(0.5)$	-0.6820	-0.6842	-0.6859	-0.6860	-0.6864	-0.6876
		(0.0325)	(0.0213)	(0.0210)	(0.0163)	(0.0148)	(0.0101)
$\sigma^2_{\psi*}$	0.3	0.2886	0.2905	0.2914	0.2933	0.2938	0.2952
		(0.0371)	(0.0338)	(0.0263)	(0.0249)	(0.0198)	(0.0181)
$\sigma^2_{\varepsilon*}$	0.2	0.1976	0.1979	0.1980	0.1982	0.1983	0.1986
		(0.0087)	(0.0057)	(0.0056)	(0.0040)	(0.0036)	(0.0027)

Tabela 5.3 – The estimates of all parameters of the AHMC model with censoring rate (40%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
$\boldsymbol{\phi}$	$\mathbf{0}$	0.0646	0.0604	0.0583	0.0536	0.0505	0.0461
		(0.0661)	(0.0425)	(0.0413)	(0.0305)	(0.0270)	(0.0197)
		0.0653	0.0609	0.0579	0.0533	0.0507	0.0452
		(0.0650)	(0.0421)	(0.0410)	(0.0307)	(0.0274)	(0.0193)
		0.0651	0.0597	0.0577	0.0529	0.0506	0.0453
		(0.0642)	(0.0423)	(0.0414)	(0.0302)	(0.0270)	(0.0195)
		0.0648	0.0616	0.0585	0.0536	0.0508	0.0460
		(0.0643)	(0.0417)	(0.0421)	(0.0301)	(0.0271)	(0.0193)
$\boldsymbol{\sigma}^2_{\psi}$	0.35	0.3090	0.3155	0.3169	0.3231	0.3247	0.3316
		(0.0431)	(0.0403)	(0.0382)	(0.0339)	(0.0323)	(0.0270)
		0.3093	0.3151	0.3163	0.3227	0.3245	0.3311
		(0.0434)	(0.0405)	(0.0384)	(0.0337)	(0.0324)	(0.0269)
		0.3086	0.3153	0.3161	0.3230	0.3250	0.3214
		(0.0437)	(0.0409)	(0.0386)	(0.0340)	(0.0321)	(0.0268)
		0.3099	0.3160	0.3168	0.3236	0.3249	0.3219
		(0.0430)	(0.0411)	(0.0389)	(0.0338)	(0.0322)	(0.0271)
σ^2_{ε}	0.25	0.2456	0.2465	0.2466	0.2474	0.2476	0.2482
		(0.0106)	(0.0071)	(0.0070)	(0.0053)	(0.0047)	(0.0035)
ϕ^*	$log(0.5)$	-0.6817	-0.6836	-0.6843	-0.6857	-0.6861	-0.6874
		(0.0332)	(0.0221)	(0.0216)	(0.0170)	(0.0153)	(0.0105)
$\sigma^2_{\psi*}$	0.3	0.2879	0.2897	0.2906	0.2924	0.2931	0.2947
		(0.0378)	(0.0344)	(0.0270)	(0.0255)	(0.0199)	(0.0188)
$\sigma^2_{\varepsilon*}$	0.2	0.1973	0.1977	0.1978	0.1980	0.1982	0.1985
		(0.0089)	(0.0059)	(0.0058)	(0.0042)	(0.0038)	(0.0028)

Tabela 5.4 – The estimates of all parameters of the SI-AHMC model with censoring rate (10%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
$\boldsymbol{\phi}$	$\mathbf{0}$	0.0486	0.0437	0.0428	0.0371	0.0351	0.0296
		(0.0322)	(0.0205)	(0.0202)	(0.0142)	(0.0126)	(0.0090)
		0.0479	0.0429	0.0417	0.368	0.0348	0.0293
		(0.0325)	(0.0206)	(0.0201)	(0.0143)	(0.0127)	(0.0090)
		0.0475	0.0431	0.0423	0.0361	0.0346	0.0289
		(0.0326)	(0.0202)	(0.0197)	(0.0142)	(0.0127)	(0.0090)
		0.0483	0.0435	0.0429	0.0366	0.0350	0.0294
		(0.0320)	(0.0200)	(0.0198)	(0.0142)	(0.0127)	(0.0090)
$\boldsymbol{\sigma}^2_\psi$	0.35	0.3276	0.3327	0.3336	0.3381	0.3397	0.3439
		(0.0285)	(0.0258)	(0.0252)	(0.0228)	(0.0221)	(0.0202)
		0.3283	0.3321	0.3334	0.3387	0.3396	0.3442
		(0.0282)	(0.0259)	(0.0250)	(0.0228)	(0.0222)	(0.0196)
		0.3286	0.3324	0.3333	0.3289	0.3401	0.3440
		(0.0281)	(0.0259)	(0.0250)	(0.0229)	(0.0219)	(0.0196)
		0.3280	0.3320	0.3331	0.3384	0.3400	0.3443
		(0.0289)	(0.0260)	(0.0253)	(0.0231)	(0.0220)	(0.0198)
σ^2_ε	0.25	0.2473	0.2478	0.2480	0.2487	0.2489	0.2495
		(0.0054)	(0.0035)	(0.0034)	(0.0024)	(0.0022)	(0.0015)
ϕ^*	$log(0.5)$	-0.6801	-0.6823	-0.6830	-0.6852	-0.6860	-0.6875
		(0.0311)	(0.0198)	(0.0189)	(0.0141)	(0.0136)	(0.0087)
$\sigma^2_{\psi*}$	0.3	0.2924	0.2939	0.2948	0.2954	0.2961	0.2975
		(0.0342)	(0.0311)	(0.0239)	(0.0221)	(0.0167)	(0.0133)
$\sigma^2_{\varepsilon*}$	0.2	0.1983	0.1986	0.1987	0.1989	0.1990	0.1992
		(0.0077)	(0.0048)	(0.0047)	(0.0029)	(0.0028)	(0.0019)
$\boldsymbol{\gamma}$	0	0.0572	0.0511	0.0498	0.0435	0.0416	0.0359
		(0.0617)	(0.0498)	(0.0478)	(0.0209)	(0.0184)	(0.0089)
		0.0585	0.0519	0.0502	0.0438	0.0419	0.0356
		(0.0259)	(0.0225)	(0.0207)	(0.0181)	(0.0165)	(0.0080)
		0.0579	0.0513	0.0496	0.0435	0.0411	0.0351
		(0.0106)	(0.0081)	(0.0070)	(0.0051)	(0.0049)	(0.0034)
		0.0570	0.0508	0.0495	0.0431	0.0421	0.0357
		(0.0011)	(0.0008)	(0.0007)	(0.0005)	(0.0005)	(0.0004)

Tabela 5.4 – The estimates of all parameters of the SI-AHMC model with censoring rate (10%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
σ^2_ζ	0.25	0.2431	0.2444	0.2452	0.2468	0.2475	0.2486
		(0.0341)	(0.0322)	(0.0225)	(0.0208)	(0.0165)	(0.0138)
		0.2436	0.2447	0.2451	0.2469	0.2473	0.2487
		(0.0357)	(0.0331)	(0.0217)	(0.0207)	(0.0163)	(0.0139)
		0.2433	0.2448	0.2456	0.2466	0.2474	0.2487
		(0.0346)	(0.0330)	(0.0229)	(0.0209)	(0.0159)	(0.0134)
		0.2431	0.2450	0.2453	0.2469	0.2477	0.2489
		(0.0344)	(0.0328)	(0.0234)	(0.0201)	(0.0145)	(0.0133)
		0.2429	0.2446	0.2452	0.2465	0.2480	0.2491
		(0.0359)	(0.0329)	(0.0216)	(0.0202)	(0.0166)	(0.0138)
$\sigma^2_{\varepsilon_p}$	0.15	0.1485	0.1488	0.1489	0.1491	0.1493	0.1495
		(0.0064)	(0.0043)	(0.0041)	(0.0032)	(0.0026)	(0.0018)

Tabela 5.5 – The estimates of all parameters of the SI-AHMC model with censoring rate (20%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
$\boldsymbol{\phi}$	**0**	0.0496	0.0443	0.0439	0.0386	0.0363	0.0306
		(0.0327)	(0.0212)	(0.0204)	(0.0148)	(0.0129)	(0.0094)
		0.0485	0.0438	0.0429	0.0381	0.0362	0.0300
		(0.0329)	(0.0213)	(0.0207)	(0.0146)	(0.0131)	(0.0092)
		0.0482	0.0440	0.0430	0.0376	0.0367	0.0297
		(0.0328)	(0.0216)	(0.0201)	(0.0147)	(0.0133)	(0.0093)
		0.0494	0.0445	0.0437	0.0379	0.0359	0.0299
		(0.0322)	(0.0215)	(0.0203)	(0.0146)	(0.0130)	(0.0091)
$\boldsymbol{\sigma}^2_\psi$	0.35	0.3261	0.3318	0.3327	0.3369	0.3381	0.3428
		(0.0297)	(0.0269)	(0.0262)	(0.0242)	(0.0232)	(0.0211)
		0.3264	0.3319	0.3330	0.3366	0.3379	0.3430
		(0.0294)	(0.0273)	(0.0260)	(0.0245)	(0.0232)	(0.0215)
		0.3260	0.3313	0.3326	0.3367	0.3384	0.3431
		(0.0293)	(0.0276)	(0.0259)	(0.0246)	(0.0223)	(0.0210)
		0.3267	0.3322	0.3331	0.3370	0.3375	0.3426
		(0.0300)	(0.0271)	(0.0261)	(0.0240)	(0.0228)	(0.0209)
σ^2_ε	0.25	0.2474	0.2477	0.2479	0.2486	0.2488	0.2494
		(0.0054)	(0.0036)	(0.0035)	(0.0024)	(0.0023)	(0.0016)
ϕ^*	$log(0.5)$	-0.6792	-0.6814	-0.6821	-0.6843	-0.6851	-0.6869
		(0.0322)	(0.0213)	(0.0208)	(0.0148)	(0.0138)	(0.0093)
$\sigma^2_{\psi*}$	0.3	0.2916	0.2933	0.2941	0.2950	0.2957	0.2969
		(0.0352)	(0.0319)	(0.0248)	(0.0230)	(0.0175)	(0.0139)
$\sigma^2_{\varepsilon*}$	0.2	0.1981	0.1984	0.1985	0.1988	0.1989	0.1991
		(0.0079)	(0.0050)	(0.0048)	(0.0031)	(0.0029)	(0.0020)
$\boldsymbol{\gamma}$	0	0.0589	0.0521	0.0505	0.0449	0.0425	0.0366
		(0.0629)	(0.0507)	(0.0492)	(0.0215)	(0.0187)	(0.0089)
		0.0593	0.0529	0.0503	0.0446	0.0429	0.0360
		(0.0270)	(0.0233)	(0.0211)	(0.0188)	(0.0172)	(0.0085)
		0.0591	0.0525	0.0508	0.0451	0.0418	0.0369
		(0.0113)	(0.0086)	(0.0072)	(0.0052)	(0.0052)	(0.0041)
		0.0581	0.0519	0.0511	0.0453	0.0426	0.0365
		(0.0012)	(0.0008)	(0.0008)	(0.0006)	(0.0006)	(0.0005)

Tabela 5.5 – The estimates of all parameters of the SI-AHMC model with censoring rate (20%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
σ^2_ζ	0.25	0.2417	0.2430	0.2437	0.2451	0.2459	0.2473
		(0.0352)	(0.0333)	(0.0230)	(0.0211)	(0.0167)	(0.0142)
		0.2413	0.2428	0.2433	0.2455	0.2463	0.2478
		(0.0361)	(0.0340)	(0.0234)	(0.0210)	(0.0165)	(0.0143)
		0.2421	0.2431	0.2436	0.2450	0.2458	0.2480
		(0.0355)	(0.0339)	(0.0238)	(0.0213)	(0.0161)	(0.0143)
		0.2418	0.2426	0.2435	0.2454	0.2463	0.2479
		(0.0351)	(0.0337)	(0.0240)	(0.0211)	(0.0152)	(0.0145)
		0.2420	0.2429	0.2438	0.2451	0.2458	0.2477
		(0.0363)	(0.0334)	(0.0229)	(0.0210)	(0.0168)	(0.0141)
$\sigma^2_{\varepsilon_p}$	0.15	0.1479	0.1483	0.1484	0.1488	0.1490	0.1494
		(0.0068)	(0.0046)	(0.0043)	(0.0032)	(0.0027)	(0.0019)

Tabela 5.6 – The estimates of all parameters of the SI-AHMC model with censoring rate (40%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
$\boldsymbol{\phi}$	$\mathbf{0}$	0.0501	0.0456	0.0443	0.0399	0.0380	0.0320
		(0.0329)	(0.0214)	(0.0206)	(0.0156)	(0.0133)	(0.0097)
		0.0493	0.0447	0.0441	0.0396	0.0375	0.0321
		(0.0331)	(0.0215)	(0.0209)	(0.0154)	(0.0137)	(0.0094)
		0.0490	0.0451	0.0439	0.0393	0.0381	0.0317
		(0.0330)	(0.0223)	(0.0204)	(0.0155)	(0.0140)	(0.0096)
		0.0502	0.0460	0.0446	0.0395	0.0376	0.0313
		(0.0326)	(0.0224)	(0.0207)	(0.0157)	(0.0139)	(0.0093)
$\boldsymbol{\sigma}^2_\psi$	0.35	0.3251	0.3306	0.3316	0.3357	0.3369	0.3418
		(0.0307)	(0.0284)	(0.0273)	(0.0251)	(0.0239)	(0.0214)
		0.3020	0.3308	0.3314	0.3355	0.3367	0.3420
		(0.0303)	(0.0281)	(0.0275)	(0.0256)	(0.0240)	(0.0218)
		0.3016	0.3305	0.3317	0.3359	0.3371	0.3416
		(0.0304)	(0.0284)	(0.0271)	(0.0257)	(0.0234)	(0.0215)
		0.3013	0.3310	0.3321	0.3360	0.3366	0.3417
		(0.0311)	(0.0283)	(0.0277)	(0.0251)	(0.0237)	(0.0213)
σ^2_ε	0.25	0.2473	0.2476	0.2478	0.2484	0.2487	0.2491
		(0.0055)	(0.0037)	(0.0036)	(0.0026)	(0.0023)	(0.0017)
ϕ^*	$log(0.5)$	-0.6787	-0.6809	-0.6813	-0.6836	-0.6843	-0.6862
		(0.0327)	(0.0216)	(0.0211)	(0.0151)	(0.0140)	(0.0096)
$\sigma^2_{\psi*}$	0.3	0.2912	0.2929	0.2937	0.2947	0.2953	0.2966
		(0.0355)	(0.0321)	(0.0251)	(0.0234)	(0.0178)	(0.0141)
$\sigma^2_{\varepsilon*}$	0.2	0.1978	0.1981	0.1982	0.1985	0.1986	0.1989
		(0.0081)	(0.0052)	(0.0050)	(0.0033)	(0.0031)	(0.0022)
$\boldsymbol{\gamma}$	0	0.0602	0.0533	0.0516	0.0463	0.0439	0.0383
		(0.0636)	(0.0515)	(0.0501)	(0.0228)	(0.0195)	(0.0091)
		0.0596	0.0536	0.0520	0.0467	0.0441	0.0386
		(0.0276)	(0.0240)	(0.0218)	(0.0193)	(0.0178)	(0.0089)
		0.0599	0.0531	0.0514	0.0461	0.0438	0.0387
		(0.0120)	(0.0092)	(0.0077)	(0.0060)	(0.0059)	(0.0046)
		0.0597	0.0539	0.0516	0.0459	0.0435	0.0390
		(0.0013)	(0.0010)	(0.0009)	(0.0008)	(0.0007)	(0.0005)

Tabela 5.6 – The estimates of all parameters of the SI-AHMC model with censoring rate (40%) by MCEMNR (SE in parenthesis)

N		100		250		500	
n		10	25	10	25	10	25
Parameter	*True value*						
σ^2_ζ	0.25	0.2399	0.2416	0.2422	0.2439	0.2445	0.2460
		(0.0371)	(0.0347)	(0.0236)	(0.0216)	(0.0169)	(0.0146)
		0.2401	0.2418	0.2424	0.2441	0.2447	0.2464
		(0.0370)	(0.0348)	(0.0239)	(0.0218)	(0.0170)	(0.0148)
		0.2397	0.2420	0.2426	0.2443	0.2449	0.2463
		(0.0378)	(0.0350)	(0.0240)	(0.0219)	(0.0168)	(0.0147)
		0.2494	0.2419	0.2427	0.2444	0.2450	0.2467
		(0.0367)	(0.0346)	(0.0242)	(0.0217)	(0.0158)	(0.0148)
		0.2403	0.2424	0.2429	0.2447	0.2453	0.2470
		(0.0375)	(0.0352)	(0.0234)	(0.0217)	(0.0171)	(0.0146)
$\sigma^2_{\varepsilon_p}$	0.15	0.1474	0.1478	0.1480	0.1485	0.1487	0.1491
		(0.0073)	(0.0045)	(0.0042)	(0.0033)	(0.0027)	(0.0019)

Tabela 5.7 – Values of the deviance for the best subset of each size in two censoring rates

			Deviance	
Model	Subset size	Predictors	CR:20%	CR:40%
1	1	*Tre.*	751.4400	761.2625
2	2	*Tre., Age*	747.4016	756.5138
3	2	*Tre., Tre.× Cen.*	737.6777	749.6382
4	3	*Tre., Tre.× Age, Tre.× Cen.*	737.3458	749.5815
5	3	*Tre., Age, Cen.*	732.6792	748.4719
6	4	*Tre., Age, Sex, Cen.*	732.2774	747.1268

Tre.: Treatment, Cen.: Center, CR: Censoring rate

Tabela 5.8 – Values of AIC and BIC for the best subset of each size in two censoring rates

	AIC		BIC	
Model	CR:20%	CR:40%	CR:20%	CR:40%
1	1601.77	1652.91	1951.46	1993.65
2	1591.39	1640.31	1942.91	1976.19
3	1578.64	1615.44	1931.39	1957.83
4	1564.17	1592.73	1918.51	1945.48
5	1557.59	1585.27	1909.43	1934.09
6	1548.23	1576.18	1896.74	1923.11

Tabela 5.9 – Estimates of intercepts and coefficients of the sixth model for two censoring rates

Censoring Rate	Variable	Estimates	Std. Error	P-value
20%	*Treatment*	0.7094	0.1984	0.0017
	Age	-0.0229	0.0054	0.0001
	Sex	-0.1269	0.2009	0.5129
	Center	0.6302	0.1563	0.0004
40%	*Treatment*	0.5817	0.2218	0.0101
	Age	-0.0236	0.0065	0.0007
	Sex	-0.2648	0.2279	0.2316
	Center	0.5537	0.1761	0.0020

6 Conclusions and proposals for future work

Chapter 2 proposes the use of self-modeling regression to analysis ordinal longitudinal data when the conditional cumulative probabilities for a category of an outcome is linked with a shape-invariant model. The proposed model ensures that individual's response curves have a common shape and that a particular individual's response curve is some simple transformation of the common shape curve. The model is essentially semi-parametric where the population time curve is modeled with a penalized regression spline. Besides simulation studies, we have also analyzed a Prostate cancer data based on our model to see whether the dose level of radiation and other factors affect the severity of genito-urinary (bladder) toxicity.

The model we consider though a bit complex, has the capability of covering the nonlinear as well as linear proportional odds model with simpler assumptions. Data obeying this framework are common and hence are quite useful to the medical statisticians for taking appropriate decisions. Some more issues like missing outcomes and measurement errors are also of interest to medical people.

Survival analysis is plagued by problem of censoring in design of clinical trials which renders routine methods of determination of central tendency redundant in computation of average survival time. Since patients having a common surgery usually exhibit a similar pattern, it is natural to build up a nonlinear model that is shape invariant. In Chapter 3, we extended the self-modeling ordinal model for censored data. As an application for anesthesia recovery data based on our model, we focus on the question of whether the dose of anesthesia affects the post-surgical recovery. In particular, we investigate the interaction between the dose effect and time to follow-up

Unlike the uni-variate mixture cure model, a cure model has been presented in Chapter 4 that models survival data in the case of ordinal response. The model is based on a proportional hazards assumption with random effects. We used the self-modeling ordinal model and the Cox regression respectively for the probability and time of occurrence of an event. A schizophrenia illness data is considered based on our model to verify effects of the treatment on the schizophrenia illness.

In Chapter 5 we used of self-modeling regression to the accelerated hazards mixture cure model. The extended model can be applied for the time of occurrence of an event when self-modeling binary model is used for the probability of occurrence of an event. As an application of the model, we apply the model to the respiratory illness data set to see whether the treatment affects the respiratory illness.

Monte Carlo Expectation Maximization (MCEM) technique is used to estimate the parameters of the models. Simulation studies are also carried out to justify the methodologies used. The results in the simulations and applications show the advantages of the models with respect to the old models very well.

The future researches emphasize the following topics:

1. Since in many researches observed responses are binary or nominal, we can apply the presented models for these kinds of data. Also we can use the provided model in chapter 5 for ordinal and nominal responses.

2. Using the interval censoring instead of right censoring in the self-modeling regression model.

3. Applying the self-modeling regression model in the definition of the survival function for uncured individuals and comparing the models in presence of the correlation between covariates of the model.

4. Considering to another distributions such as log-normal and Gompertz-Makeham for the baseline survivor function in the cure models and comparing the models with the accelerated failure time mixture cure model.

5. Extension of a the new cure models when cure information is partially known, especially for the accelerated hazard form.

Referências

BOAG, J. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *The Royal Statistical Society*, B, n. 11, p. 15–44, 1949.

CHEN, Y. Accelerated hazards regression model and its adequacy for censored survival data. *Biometrics*, v. 57, p. 853–860, 2000.

DAVIS, C. Semi-parametric and non-parametric methods for the analysis of repeated measurement with applications to clinical trials. *Statistics in Medicine*, n. 10, p. 1959–1980, 1991.

FAREWELL, V. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, v. 38, p. 1041–1046, 1982.

______. Mixture models in survival analysis: Are they worth the risk? *Canadian Journal of Statistics*, v. 14, p. 257–262, 1986.

MCCULLAGH, P. Regression models for ordinal data. *The Royal Statistical Society*, B, n. 42, p. 109–142, 1980.

ROBINSON, G. That blup is a good thing: the estimation of random effects. *Statistical Science*, v. 6, p. 15–51, 1991.

WAHBA, G. A comparison of gcv and gml for choosing the smoothing parameter in the generalized spline smoothing problem. *The Annals of Statistics*, v. 13, p. 1378–1402, 1985.

WANG, Y. Smoothing spline models with correlated random errors. *The American Statistical Association*, v. 93, p. 341–348, 1998.

YAKOVLEV, A.; ASSELAIN, B.; BARDOU, V.; FOURQUET, A.; HOANG, T.; ROCHEFEDIERE, A.; TSODIKOV, A. A simple stochastic model of tumor recurrence and its applications to data on pre-menopausal breast cancer. *French Society of Biometrics*, p. 66–82, 1993.

Apêndices

APÊNDICE A – The derivatives for the Newton-Raphson and EM procedures

A.1 Chapter 2

A.1.1 The Newton-Raphson procedure

The first derivative of l_{ic} respect to thresholds τ_l, for l= 1,...,L is given by:

$$\frac{\partial P_{ijl}}{\partial \tau_l} = P_{ijl}(1 - P_{ijl}),$$

$$\frac{\partial \eta_{ijl}}{\partial \tau_l} = 1, \ \frac{\partial \eta_{ij(l-1)}}{\partial \tau_l} = 0,$$

$$\frac{\partial P_{ij(l-1)}}{\partial \tau_l} = \frac{\partial P_{ij(l-1)}}{\partial \eta_{ij(l-1)}} \frac{\partial \eta_{ij(l-1)}}{\partial \tau_l} = 0,$$

$$\frac{\partial l_{ic}}{\partial \tau_l} = \sum_{j=1}^{n_i} I_{y_{ijl}}(l) \ \frac{P_{ijl}(1 - P_{ijl})}{P_{ijl} - P_{ijl-1}}.$$

The first derivative of l_{ic} respect to ϕ is given by:

$$\frac{\partial l_{ic}}{\partial \phi} = \sum_{j=1}^{n_i} \sum_{l=1}^{L} I_{y_{ij}}(l) \frac{\frac{\partial P_{ijl}}{\partial \phi} - \frac{\partial P_{ijl-1}}{\partial \phi}}{P_{ijl} - P_{ijl-1}} + \frac{(\theta_i - X_i\phi - Z_i\psi_i)' X_i}{2\sigma_\varepsilon^2},$$

$$\frac{\partial P_{ijl}}{\partial \phi} = P_{ijl}(1 - P_{ijl}) \frac{\partial \eta_{ijl}}{\partial \phi},$$

$$\frac{\partial \alpha_{0i}}{\partial \phi} = X_i \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \ \frac{\partial \alpha_{1i}}{\partial \phi} = X_i \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \ \frac{\partial \beta_{0i}}{\partial \phi} = X_i \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \ \frac{\partial \beta_{1i}}{\partial \phi} = X_i \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

$$\frac{\partial \eta_{ijl}}{\partial \phi} = - \left[X_i \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + X_i \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \exp(\alpha_{1i}) \ \mu_0(t_{ij}^*) + \exp(\alpha_{1i}) \ \frac{\partial \mu_0(t_{ij}^*)}{\partial \phi} \right],$$

$$\frac{\partial \mu_0(t^*_{ij})}{\partial \phi} = \frac{\partial \mu_0(t^*_{ij})}{\partial t^*_{ij}} \frac{\partial t^*_{ij}}{\partial \phi},$$

$$\frac{\partial \mu_0(t^*_{ij})}{\partial t^*_{ij}} = \sum_{m=1}^{4} (m-1){t^*_{ij}}^{m-2}\gamma_m + \sum_{k=1}^{K} 3\Big(t^*_{ij} - \xi_k\Big)^2 \zeta_k.$$

$$\frac{\partial t^*_{ij}}{\partial \phi} = X_i \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + X_i \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \exp(\beta_{1i}) t_{ij}.$$

The first derivative of l_{ic} respect to σ^2_ε is given by:

$$\frac{\partial l_{ic}}{\partial \sigma^2_\varepsilon} = -\frac{1}{2\sigma^2_\varepsilon} + \frac{(\theta_i - X_i\phi - Z_i\psi_i)'(\theta_i - X_i\phi - Z_i\psi_i)}{2\Big(\sigma^2_\varepsilon\Big)^2},$$

The first derivative of l_{ic} respect to σ^2_ψ is given by:

$$\frac{\partial l_{ic}}{\partial \sigma^2_\psi} = -\frac{1}{2\sigma^2_\psi} + \frac{\psi^2_i}{2\Big(\sigma^2_\psi\Big)^2}.$$

The second derivative of l_{ic} respect to thresholds τ_l is given by:

$$\frac{\partial^2 P_{ijl}}{\partial {\tau_l}^2} = P_{ijl}(1-P_{ijl})(1-2P_{ijl}),$$

$$\frac{\partial^2 P_{ijl}}{\partial \tau_l \partial {\tau_l}'} = 0, \qquad l \neq l'.$$

$$\frac{\partial^2 l_{ic}}{\partial {\tau_l}^2} = \sum_{j=1}^{n_i} I_{y_{ij}}(l) \frac{P^2_{ijl}(1-P_{ijl})P_{ijl-1} - P^3_{ijl}(1-P_{ijl}) - P_{ijl}(1-P_{ijl})^2 P_{ijl-1}}{(P_{ijl} - P_{ijl-1})^2},$$

$$\frac{\partial^2 l_{ic}}{\partial \tau_l \partial \tau_{l'}} = 0, \qquad l \neq l'.$$

The second derivative of l_{ic} respect to ϕ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \phi' \partial \phi} = \sum_{j=1}^{n_i}\sum_{l=1}^{L} I_{y_{ij}}(l) \left(\frac{\frac{\partial^2 P_{ijl}}{\partial \phi' \partial \phi} - \frac{\partial^2 P_{ijl-1}}{\partial \phi' \partial \phi}}{P_{ijl} - P_{ijl-1}} - \frac{\Big(\frac{\partial P_{ijl}}{\partial \phi'} - \frac{\partial P_{ijl-1}}{\partial \phi'}\Big)\Big(\frac{\partial P_{ijl}}{\partial \phi} - \frac{\partial P_{ijl-1}}{\partial \phi}\Big)}{(P_{ijl} - P_{ijl-1})^2} \right) - \frac{X'_i X_i}{2\sigma^2_\varepsilon},$$

$$\frac{\partial^2 P_{ijl}}{\partial \phi' \partial \phi} = P_{ijl}(1-P_{ijl})^2 \frac{\partial \eta_{ijl}}{\partial \phi'} \frac{\partial \eta_{ijl}}{\partial \phi} - P_{ijl}^2 (1-P_{ijl}) \frac{\partial \eta_{ijl}}{\partial \phi'} \frac{\partial \eta_{ijl}}{\partial \phi} + P_{ijl}(1-P_{ijl}) \frac{\partial^2 \eta_{ijl}}{\partial \phi' \partial \phi}.$$

$$\begin{aligned} \frac{\partial^2 \eta_{ijl}}{\partial \phi' \partial \phi} &= -\Bigg[X_i \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix} X_i' \ \exp(\alpha_{1i}) \ \mu_0(t^*_{ij}) + X_i \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \exp(\alpha_{1i}) \frac{\partial \mu_0(t^*_{ij})}{\partial \phi'} \\ & + \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix} X_i' \exp(\alpha_{1i}) \frac{\partial \mu_0(t^*_{ij})}{\partial \phi} + \exp(\alpha_{1i}) \frac{\partial^2 \mu_0(t^*_{ij})}{\partial \phi' \partial \phi} \Bigg]. \end{aligned}$$

$$\frac{\partial^2 \mu_0(t^*_{ij})}{\partial \phi' \ \partial \phi} = \frac{\partial^2 \mu_0(t^*_{ij})}{\partial \phi' \ \partial t^*_{ij}} \ \frac{\partial t^*_{ij}}{\partial \phi} + \frac{\partial \mu_0(t^*_{ij})}{\partial t^*_{ij}} \ \frac{\partial^2 t^*_{ij}}{\partial \phi' \ \partial \phi},$$

$$\begin{aligned} \frac{\partial^2 \mu_0(t^*_{ij})}{\partial \phi' \ \partial t^*_{ij}} &= \sum_{m=1}^{4} (m-1)(m-2) \Big[\begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix} X_i' + \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} X_i' \ \exp(\beta_{1i}) t_{ij} \Big] {t^*_{ij}}^{m-3} \gamma_m \\ & + \sum_{k=1}^{K} 6 \Big[\begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix} X_i' + \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} X_i' \ \exp(\beta_{1i}) t_{ij} \Big] \Big(t^*_{ij} - \xi_k \Big) \zeta_k, \end{aligned}$$

$$\frac{\partial^2 t^*_{ij}}{\partial \phi' \partial \phi} = X_i \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} X_i' \ \exp(\beta_{1i}) t_{ij}.$$

The second derivative of l_{ic} respect to thresholds τ_l and ϕ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \tau_l \partial \phi} = \sum_{j=1}^{n_i} \left(\frac{\frac{\partial^2 P_{ijl}}{\partial \tau_l \partial \phi} - \frac{\partial^2 P_{ijl-1}}{\partial \tau_l \partial \phi}}{P_{ijl} - P_{ijl-1}} - \frac{\left(\frac{\partial P_{ijl}}{\partial \tau_l} - \frac{\partial P_{ijl-1}}{\partial \tau_l} \right) \left(\frac{\partial P_{ijl}}{\partial \phi} - \frac{\partial P_{ijl-1}}{\partial \phi} \right)}{(P_{ijl} - P_{ijl-1})^2} \right).$$

$$\frac{\partial^2 P_{ijl}}{\partial \tau_l \partial \phi} = P_{ijl}(1-P_{ijl})(1-2P_{ijl}) \frac{\partial \eta_{ijl}}{\partial \phi}.$$

The second derivative of l_{ic} respect to ϕ and σ^2_ε is given by:

$$\frac{\partial^2 l_{ic}}{\partial \sigma^2_\varepsilon \partial \phi} = -\frac{(\theta_i - X_i \phi - Z_i \psi_i)' X_i}{2 \Big(\sigma^2_\varepsilon \Big)^2}.$$

$$\frac{\partial^2 P_{ijl}}{\partial \sigma^2_\varepsilon \partial \phi} = \frac{\partial}{\partial \sigma^2_\varepsilon} \left(P_{ijl}(1-P_{ijl}) \frac{\partial \eta_{ijl}}{\partial \phi} \right) = 0.$$

The second derivative of l_{ic} respect to σ^2_ε is given by:

$$\frac{\partial^2 l_{ic}}{\partial\left(\sigma^2_\varepsilon\right)^2} = \frac{1}{2\left(\sigma^2_\varepsilon\right)^2} - \frac{\left(\theta_i - X_i\phi - Z_i\psi_i\right)'\left(\theta_i - X_i\phi - Z_i\psi_i\right)}{\left(\sigma^2_\varepsilon\right)^3}.$$

The second derivative of l_{ic} respect to σ^2_ψ is given by:

$$\frac{\partial^2 l_{ic}}{\partial\left(\sigma^2_\psi\right)^2} = \frac{1}{2\left(\sigma^2_\psi\right)^2} - \frac{\psi_i^2}{\left(\sigma^2_\psi\right)^3}.$$

A.1.2 The E and M steps

E- step:

$$Q\left(\boldsymbol{\theta}\,\middle|\boldsymbol{\theta}^{(s)}\right) \approx \sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\left(\sum_{j=1}^{n_i}\sum_{l=1}^{L} I_{y_{ij}}(l)\ log\left(P_{ijl} - P_{ijl-1}\right) - \frac{1}{2}log\left(\sigma^{2(s)}_\varepsilon\right)\right.$$
$$\left. - \frac{\left(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\right)'\left(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\right)}{2\sigma^{2(s)}_\varepsilon} - \frac{1}{2}log\left(\sigma^{2(s)}_{\psi_i}\right) - \frac{\psi_i^{2(s)}}{2\sigma^{2(s)}_{\psi_i}}\right).$$

M- step:

For $l = 1,\ldots,L$

$$\frac{\partial Q\left(\boldsymbol{\theta}\,\middle|\boldsymbol{\theta}^{(s)}\right)}{\partial\tau_l} = \sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\sum_{j=1}^{n_i} I_{y_{ij}}(l)\frac{P^{(s)}_{ijl}\left(1 - P^{(s)}_{ijl}\right)}{P^{(s)}_{ijl} - P^{(s)}_{ijl-1}} = 0,$$

$$\frac{P_{ijl}(1 - P_{ijl})}{P_{ijl} - P_{ijl-1}} = \frac{\exp(\eta_{ijl})\Big(1 + \exp(\eta_{ijl-1})\Big)}{\Big(1 + \exp(\eta_{ijl})\Big)\Big(\exp(\eta_{ijl}) - \exp(\eta_{ijl-1})\Big)}$$

$$With\ \exp(\eta_{ijl}) = \exp(\tau_l - \omega_{ij})$$
$$= \frac{\exp(\tau_l)}{\exp(\tau_l) - \exp(\tau_{l-1})}\ \frac{\exp(\omega_{ij}) + \exp(\tau_{l-1})}{\exp(\omega_{ij}) + \exp(\tau_l)}.$$

$$\frac{\partial Q\left(\boldsymbol{\theta}\,\middle|\boldsymbol{\theta}^{(s)}\right)}{\partial\tau_l} = \sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\sum_{j=1}^{n_i} I_{y_{ij}}(l)\frac{\exp\left(\omega^{(s)}_{ij}\right) + \exp\left(\tau_{l-1}\right)}{\exp\left(\omega^{(s)}_{ij}\right) + \exp\left(\tau_l\right)} = 0,$$

$$\sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\sum_{j=1}^{n_i} 1 + \frac{\exp\left(\tau_{l-1}\right) - \exp\left(\tau_l\right)}{\exp\left(\omega^{(s)}_{ij}\right) + \exp\left(\tau_l\right)} = NM_in_i + \sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\sum_{j=1}^{n_i}\frac{\exp\left(\tau_{l-1}\right) - \exp\left(\tau_l\right)}{\exp\left(\omega^{(s)}_{ij}\right) + \exp\left(\tau_l\right)} = 0,$$

$$\sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\sum_{j=1}^{n_i}\frac{1}{\exp\left(\omega_{ij}^{(s)}\right)+\exp\left(\tau_l\right)}=\frac{NM_in_i}{\exp\left(\tau_l\right)-\exp\left(\tau_{l-1}\right)}$$

For the parameter τ_l, we can't get the closed form expression for the maximizer $Q\left(\boldsymbol{\theta}\mid\boldsymbol{\theta}^{(s)}\right)$. We use the Newton-Raphson iterations as:

$$\begin{aligned}\frac{\partial Q\left(\boldsymbol{\theta}\left|\boldsymbol{\theta}^{(s)}\right.\right)}{\partial\phi} &= \sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\left(\sum_{j=1}^{n_i}\sum_{l=1}^{L}I_{y_{ij}}\left(l\right)\frac{\frac{\partial\eta_{ijl}}{\partial\phi}\Big(P_{ijl}(1-P_{ijl})-P_{ijl-1}(1-P_{ijl-1})\Big)}{P_{ijl}-P_{ijl-1}}\right.\\ &\left.+\frac{\left(\theta_i-X_i\phi^{(s)}-Z_i\psi_i^{(s)}\right)'X_i}{2\sigma_\varepsilon^{2(s)}}\right)\\ &= 0,\end{aligned}$$

$$\begin{aligned}\phi^{(s+1)} &= \sum_{i=1}^{N}\frac{1}{M_iX_i'X_i}\sum_{m=1}^{M_i}\left[2\sigma_\varepsilon^{2(s)}\left(\sum_{j=1}^{n_i}\sum_{l=1}^{L}I_{y_{ij}}(l)\frac{\frac{\partial\eta_{ijl}}{\partial\phi}\Big(P_{ijl}(1-P_{ijl})-P_{ijl-1}(1-P_{ijl-1})\Big)}{P_{ijl}-P_{ijl-1}}\right)\right.\\ &\left.+\theta_i'X_i-\psi_i^{'(s)}Z_i'X_i\right].\end{aligned}$$

$$\frac{\partial Q\left(\boldsymbol{\theta}\left|\boldsymbol{\theta}^{(s)}\right.\right)}{\partial\sigma_\varepsilon^2}=\sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\left[-\frac{1}{2\sigma_\varepsilon^{2(s)}}+\frac{\left(\theta_i-X_i\phi^{(s)}-Z_i\psi_i^{(s)}\right)'\left(\theta_i-X_i\phi^{(s)}-Z_i\psi_i^{(s)}\right)}{2\left(\sigma_\varepsilon^{2(s)}\right)^2}\right]=0,$$

$$\sigma_\varepsilon^{2(s+1)}=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\left(\theta_i-X_i\phi^{(s)}-Z_i\psi_i^{(s)}\right)'\left(\theta_i-X_i\phi^{(s)}-Z_i\psi_i^{(s)}\right).$$

$$\frac{\partial Q\left(\boldsymbol{\theta}\left|\boldsymbol{\theta}^{(s)}\right.\right)}{\partial\sigma_\psi^2}=\sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\left[-\frac{1}{2\sigma_\psi^{2(s)}}+\frac{\psi_i^{2(s)}}{2\left(\sigma_\psi^{2(s)}\right)^2}\right]=0,$$

$$\sigma_\psi^{2(s+1)}=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\psi_i^{2(s)}.$$

A.2 Chapter 3

A.2.1 The Newton-Raphson procedure

The first derivative of l_{ic} respect to thresholds τ_l, for $l = 1, \ldots, L$ is given by:

$$\frac{\partial l_{ic}}{\partial \tau_l} = \sum_{j=1}^{n_i} I_{y_{ijl}}(l) \left[c_{ij} \frac{P_{ijl}(1-P_{ijl})}{P_{ijl} - P_{ijl-1}} + (1-c_{ij}) \frac{-P_{ijl}(1-P_{ijl})}{1-P_{ijl}} \right],$$

$$\frac{\partial \eta_{ijl}}{\partial \tau_l} = 1, \; \frac{\partial P_{ijl}}{\partial \tau_l} = P_{ijl}(1-P_{ijl}), \; \frac{\partial \eta_{ij(l-1)}}{\partial \tau_l} = 0, \; \frac{\partial P_{ij(l-1)}}{\partial \tau_l} = 0.$$

The first derivative of l_{ic} respect to ϕ is given by:

$$\frac{\partial l_{ic}}{\partial \phi} = \sum_{j=1}^{n_i} \sum_{l=1}^{L} I_{y_{ij}}(l) \left[c_{ij} \frac{\frac{\partial P_{ijl}}{\partial \phi} - \frac{\partial P_{ijl-1}}{\partial \phi}}{P_{ijl} - P_{ijl-1}} + (1-c_{ij}) \frac{-\frac{\partial P_{ijl}}{\partial \phi}}{1-P_{ijl}} \right] + \frac{(\theta_i - X_i\phi - Z_i\psi_i)' X_i}{2\sigma_\varepsilon^2},$$

$$\frac{\partial P_{ijl}}{\partial \phi} = P_{ijl}(1-P_{ijl}) \frac{\partial \eta_{ijl}}{\partial \phi},$$

$$\frac{\partial \alpha_{0i}}{\partial \phi} = X_i \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \; \frac{\partial \alpha_{1i}}{\partial \phi} = X_i \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \; \frac{\partial \beta_{0i}}{\partial \phi} = X_i \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \; \frac{\partial \beta_{1i}}{\partial \phi} = X_i \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

$$\frac{\partial \eta_{ijl}}{\partial \phi} = X_i \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + X_i \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \exp(\alpha_{1i}) \; \mu_0(t_{ij}^*) + \exp(\alpha_{1i}) \; \frac{\partial \mu_0(t_{ij}^*)}{\partial \phi},$$

$$\frac{\partial \mu_0(t_{ij}^*)}{\partial \phi} = \frac{\partial \mu_0(t_{ij}^*)}{\partial t_{ij}^*} \; \frac{\partial t_{ij}^*}{\partial \phi},$$

$$\frac{\partial \mu_0(t_{ij}^*)}{\partial t_{ij}^*} = \sum_{m=1}^{4} (m-1) {t_{ij}^*}^{m-2} \gamma_m \; + \sum_{k=1}^{K} 3 \left(t_{ij}^* - \xi_k \right)^2 \zeta_k.$$

$$\frac{\partial t_{ij}^*}{\partial \phi} = X_i \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + X_i \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \exp(\beta_{1i}) t_{ij}.$$

The first derivative of l_{ic} respect to $\sigma^2_{\varepsilon_i}$ is given by:

$$\frac{\partial l_{ic}}{\partial \sigma^2_{\varepsilon_i}} = -\frac{1}{2\sigma^2_{\varepsilon_i}} + \frac{(\theta_i - X_i\phi - Z_i\psi_i)' (\theta_i - X_i\phi - Z_i\psi_i)}{2\left(\sigma^2_{\varepsilon_i}\right)^2}.$$

The first derivative of l_{ic} respect to $\sigma^2_{\psi_i^2}$ is given by:

$$\frac{\partial l_{ic}}{\partial \sigma^2_{\psi_i}} = -\frac{1}{2\sigma^2_{\psi_i}} + \frac{\psi_i^2}{2\left(\sigma^2_{\psi_i}\right)^2}.$$

The second derivative of l_{ic} respect to thresholds τ_l, for $l = 1, \ldots, L$ is given by:

$$\frac{\partial^2 P_{ijl}}{\partial {\tau_l}^2} = P_{ijl}(1 - P_{ijl})(1 - 2P_{ijl}),$$

$$\frac{\partial^2 P_{ijl}}{\partial \tau_l \partial {\tau_l}'} = 0, \qquad l \neq l'.$$

$$\frac{\partial^2 l_{ic}}{\partial {\tau_l}^2} = \sum_{j=1}^{n_i} I_{y_{ij}}(l)\Bigg[c_{ij}\frac{P^2_{ijl}\left(1 - P_{ijl}\right)P_{ijl-1} - P^3_{ijl}\left(1 - P_{ijl}\right) - P_{ijl}(1 - P_{ijl})^2 P_{ijl-1}}{(P_{ijl} - P_{ijl-1})^2}$$
$$- (1 - c_{ij})\frac{P_{ijl} - 3P^2_{ijl} - 7P^3_{ijl} - P^4_{ijl}}{(1 - P_{ijl})^2}\Bigg],$$

$$\frac{\partial^2 l_{ic}}{\partial \tau_l \partial \tau_{l'}} = 0, \qquad l \neq l'.$$

The second derivative of l_{ic} respect to ϕ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \phi' \partial \phi} = \sum_{j=1}^{n_i}\sum_{l=1}^{L} I_{y_{ijl}}(l)\Bigg[c_{ij}\left(\frac{\frac{\partial^2 P_{ijl}}{\partial\phi'\partial\phi} - \frac{\partial^2 P_{ijl-1}}{\partial\phi'\partial\phi}}{P_{ijl} - P_{ijl-1}} - \frac{\left(\frac{\partial P_{ijl}}{\partial\phi'} - \frac{\partial P_{ijl-1}}{\partial\phi'}\right)\left(\frac{\partial P_{ijl}}{\partial\phi} - \frac{\partial P_{ijl-1}}{\partial\phi}\right)}{(P_{ijl} - P_{ijl-1})^2}\right)$$
$$- (1 - c_{ij})\left(\frac{\frac{\partial^2 P_{ijl}}{\partial\phi'\partial\phi}(1 - P_{ijl}) + \frac{\partial P_{ijl}}{\partial\phi'}\frac{\partial P_{ijl}}{\partial\phi}}{(1 - P_{ijl})^2}\right)\Bigg] - \frac{X_i' X_i}{2\sigma^2_{\varepsilon_i}}.$$

$$\frac{\partial^2 P_{ijl}}{\partial\phi'\partial\phi} = P_{ijl}(1 - P_{ijl})^2\frac{\partial\eta_{ijl}}{\partial\phi'}\frac{\partial\eta_{ijl}}{\partial\phi} - P^2_{ijl}(1 - P_{ijl})\frac{\partial\eta_{ijl}}{\partial\phi'}\frac{\partial\eta_{ijl}}{\partial\phi} + P_{ijl}(1 - P_{ijl})\frac{\partial^2\eta_{ijl}}{\partial\phi'\partial\phi}.$$

$$\frac{\partial^2 \eta_{ijl}}{\partial \phi' \partial \phi} = X_i \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix} X_i' \ \exp(\alpha_{1i}) \ \mu_0(t^*_{ij}) + X_i \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \exp(\alpha_{1i}) \frac{\partial \mu_0(t^*_{ij})}{\partial \phi'}$$
$$+ \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix} X_i' \exp(\alpha_{1i}) \frac{\partial \mu_0(t^*_{ij})}{\partial \phi} + \exp(\alpha_{1i}) \frac{\partial^2 \mu_0(t^*_{ij})}{\partial \phi' \partial \phi}.$$

$$\frac{\partial^2 \mu_0(t^*_{ij})}{\partial \phi' \ \partial \phi} = \frac{\partial^2 \mu_0(t^*_{ij})}{\partial \phi' \ \partial t^*_{ij}} \ \frac{\partial t^*_{ij}}{\partial \phi} + \frac{\partial \mu_0(t^*_{ij})}{\partial t^*_{ij}} \ \frac{\partial^2 t^*_{ij}}{\partial \phi' \ \partial \phi},$$

$$\frac{\partial^2 \mu_0(t^*_{ij})}{\partial \phi' \ \partial t^*_{ij}} = \sum_{m=1}^{4} (m-1)(m-2) \frac{\partial t^*_{ij}}{\partial \phi'} {t^*_{ij}}^{m-3} \gamma_m + \sum_{k=1}^{K} 6 \frac{\partial t^*_{ij}}{\partial \phi'} \Big(t^*_{ij} - \xi_k \Big) \zeta_k,$$

$$\frac{\partial t^*_{ij}}{\partial \phi'} = \begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix} X_i' + \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} X_i' \ \exp(\beta_{1i}) t_{ij},$$

$$\frac{\partial^2 t^*_{ij}}{\partial \phi' \partial \phi} = X_i \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} X_i' \ \exp(\beta_{1i}) t_{ij}.$$

The second derivative of l_{ic} respect to thresholds τ_l and ϕ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \tau_l \partial \phi} = \sum_{j=1}^{n_i} I_{y_{ijl}}(l) \Bigg[c_{ij} \left(\frac{\frac{\partial^2 P_{ijl}}{\partial \tau_l \partial \phi} - \frac{\partial^2 P_{ijl-1}}{\partial \tau_l \partial \phi}}{P_{ijl} - P_{ijl-1}} - \frac{\left(\frac{\partial P_{ijl}}{\partial \tau_l} - \frac{\partial P_{ijl-1}}{\partial \tau_l} \right) \left(\frac{\partial P_{ijl}}{\partial \phi} - \frac{\partial P_{ijl-1}}{\partial \phi} \right)}{(P_{ijl} - P_{ijl-1})^2} \right)$$
$$- (1 - c_{ij}) \frac{\frac{\partial^2 P_{ijl}}{\partial \tau_l \partial \phi} (1 - P_{ijl}) + \frac{\partial P_{ijl}}{\partial \tau_l} \frac{\partial P_{ijl}}{\partial \phi}}{(1 - P_{ijl})^2} \Bigg].$$

$$\frac{\partial^2 P_{ijl}}{\partial \tau_l \partial \phi} = P_{ijl}(1 - P_{ijl})(1 - 2P_{ijl}) \frac{\partial \eta_{ijl}}{\partial \phi}.$$

The second derivative of l_{ic} respect to ϕ and $\sigma^2_{\varepsilon_i}$ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \sigma^2_{\varepsilon_i} \partial \phi} = - \frac{(\theta_i - X_i \phi - Z_i \psi_i)' X_i}{2 \Big(\sigma^2_{\varepsilon_i} \Big)^2}.$$

$$\frac{\partial^2 P_{ijl}}{\partial \sigma^2_{\varepsilon_i} \partial \phi} = \frac{\partial}{\partial \sigma^2_{\varepsilon_i}} \left(P_{ijl}(1 - P_{ijl}) \frac{\partial \eta_{ijl}}{\partial \phi} \right) = 0.$$

The second derivative of l_{ic} respect to $\sigma^2_{\varepsilon_i}$ is given by:

$$\frac{\partial^2 l_{ic}}{\partial\left(\sigma^2_{\varepsilon_i}\right)^2} = \frac{1}{2\left(\sigma^2_{\varepsilon_i}\right)^2} - \frac{(\theta_i - X_i\phi - Z_i\psi_i)'\,(\theta_i - X_i\phi - Z_i\psi_i)}{\left(\sigma^2_{\varepsilon_i}\right)^3}.$$

The second derivative of l_{ic} respect to $\sigma^2_{\psi_i}$ is given by:

$$\frac{\partial^2 l_{ic}}{\partial\left(\sigma^2_{\psi_i}\right)^2} = \frac{\partial}{\partial\sigma^2_{\psi_i}}\left(-\frac{1}{2\sigma^2_{\psi_i}} + \frac{\psi_i^2}{2\sigma^2_{\psi_i}}\right) = \frac{1}{2\left(\sigma^2_{\psi_i}\right)^2} - \frac{\psi_i^2}{\left(\sigma^2_{\psi_i}\right)^3}.$$

A.2.2 The E- and M- steps

E- step:

$$Q\big(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\big) \approx \sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\left(\sum_{j=1}^{n_i}\sum_{l=1}^{L} I_{y_{ij}}(l)\Big[c_{ij}\log(P_{ijl} - P_{ijl-1}) + (1-c_{ij})\log(1-P_{ijl})\Big]\right.$$
$$\left. + \log\Big(\boldsymbol{\Phi}\left(\boldsymbol{\theta}_i|\mathbf{X}_i\boldsymbol{\phi} + \mathbf{Z}_i\boldsymbol{\psi}_i,\ \sigma^2_{\varepsilon_i}I_4\right)\Big) + \log\Big(\boldsymbol{\Phi}\left(\boldsymbol{\psi}_i|0,\ \sigma^2_{\psi_i}\right)\Big)\right).$$

M- step:

For l= $1,\ldots,L$

$$\frac{\partial Q\left(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\right)}{\partial\tau_l} = \sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\sum_{j=1}^{n_i} I_{y_{ijl}}(l)\left[c_{ij}\frac{P_{ijl}(1-P_{ijl})}{P_{ijl}-P_{ijl-1}} + (1-c_{ij})\frac{-P_{ijl}(1-P_{ijl})}{1-P_{ijl}}\right] = 0,$$

$$\frac{P_{ijl}(1-P_{ijl})}{P_{ijl}-P_{ijl-1}} = \frac{\exp(\eta_{ijl})\Big(1+\exp(\eta_{ijl-1})\Big)}{\Big(1+\exp(\eta_{ijl})\Big)\Big(\exp(\eta_{ijl}) - \exp(\eta_{ijl-1})\Big)}$$

$$With\ \exp(\eta_{ijl}) = \exp(\tau_l + \omega_{ij})$$
$$= \frac{\exp(\tau_l)}{\exp(\tau_l) - \exp(\tau_{l-1})}\ \frac{\frac{1}{\exp(\omega_{ij})} + \exp(\tau_{l-1})}{\frac{1}{\exp(\omega_{ij})} + \exp(\tau_l)}.$$

$$\frac{\partial Q\left(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\right)}{\partial\tau_l} = \sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\sum_{j=1}^{n_i} I_{y_{ij}}(l)\left[c_{ij}\frac{1}{\exp(\tau_l)-\exp(\tau_{l-1})}\,\frac{\frac{1}{\exp\left(\omega^{(s)}_{ij}\right)} + \exp\left(\tau_{l-1}\right)}{\frac{1}{\exp\left(\omega^{(s)}_{ij}\right)} + \exp\left(\tau_l\right)}\right.$$
$$\left. - (1-c_{ij})\exp(\omega^{(s)}_{ij})\right] = 0,$$

For the parameter τ_l, we can't get the closed form expression for the maximizer $Q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(s)}\right)$. We use the Newton-Raphson iterations as:

$$\frac{\partial Q\left(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\right)}{\partial \phi} = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left(\sum_{j=1}^{n_i} \sum_{l=1}^{L} I_{y_{ij}}(l) \left[c_{ij} \frac{\frac{\partial \eta_{ijl}}{\partial \phi}\Big(P_{ijl}(1-P_{ijl}) - P_{ijl-1}(1-P_{ijl-1})\Big)}{P_{ijl} - P_{ijl-1}} - (1-c_{ij})\frac{\partial \eta_{ijl}}{\partial \phi} P_{ijl} \right] + \frac{\left(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\right)' X_i}{2\sigma_{\varepsilon}^{2(s)}} \right) = 0,$$

$$\phi^{(s+1)} = \sum_{i=1}^{N} \frac{1}{M_i X_i' X_i} \sum_{m=1}^{M_i} \left[2\sigma_{\varepsilon}^{2(s)} \left(\sum_{j=1}^{n_i} \sum_{l=1}^{L} I_{y_{ij}}(l) \left[c_{ij} \frac{\frac{\partial \eta_{ijl}}{\partial \phi}\Big(P_{ijl}(1-P_{ijl}) - P_{ijl-1}(1-P_{ijl-1})\Big)}{P_{ijl} - P_{ijl-1}} \right) - (1-c_{ij})\frac{\partial \eta_{ijl}}{\partial \phi} P_{ijl} \right] + \theta_i' X_i - \psi_i'^{(s)} Z_i' X_i \right].$$

$$\frac{\partial Q\left(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\right)}{\partial \sigma_{\varepsilon}^2} = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left[-\frac{1}{2\sigma_{\varepsilon}^{2(s)}} + \frac{\left(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\right)' \left(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\right)}{2\left(\sigma_{\varepsilon}^{2(s)}\right)^2} \right] = 0,$$

$$\sigma_{\varepsilon}^{2(s+1)} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\right)' \left(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\right).$$

$$\frac{\partial Q\left(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\right)}{\partial \sigma_{\psi}^2} = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left[-\frac{1}{2\sigma_{\psi}^{2(s)}} + \frac{\psi_i^{2(s)}}{2\left(\sigma_{\psi}^{2(s)}\right)^2} \right] = 0,\ \ \sigma_{\psi}^{2(s+1)} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \psi_i^{2(s)}.$$

A.3 Chapter 4

A.3.1 The Newton-Raphson procedure

$$C = 1 - P_{ijl} + p_{ijl} S_u\big(t_{ij}|y_{ij} = 1, x^*_{ij}, z^*_{ij}\big),$$

The first derivative of l_{ic} respect to thresholds τ_l, for $l = 1, \dots, L$ is given by:

$$\frac{\partial l_{ic}}{\partial \tau_l} = \sum_{j=1}^{n_i} I_{y_{ij}}(l) \left[\delta_{ij} \frac{P_{ijl}(1 - P_{ijl})}{P_{ijl} - P_{ijl-1}} - (1 - \delta_{ij}) \frac{\frac{\partial C}{\partial \tau_l}}{C} \right].$$

$$\frac{\partial \eta_{ijl}}{\partial \tau_l} = 1, \;\; \frac{\partial \eta_{ij(l-1)}}{\partial \tau_l} = 0, \;\; \frac{\partial P_{ijl}}{\partial \tau_l} = P_{ijl}(1 - P_{ijl}), \;\; \frac{\partial P_{ij(l-1)}}{\partial \tau_l} = 0,$$

$$\frac{\partial C}{\partial \tau_l} = -\frac{\partial P_{ijl}}{\partial \tau_l} + S_u\big(t_{ij}|y_{ij} = 1, x^*_{ij}, z^*_{ij}\big) \frac{\partial p_{ijl}}{\partial \tau_l}$$

The first derivative of l_{ic} respect to ϕ is given by:

$$\frac{\partial l_{ic}}{\partial \phi} = \sum_{j=1}^{n_i} I_{y_{ij}}(l) \left[\delta_{ij} \frac{\frac{\partial p_{ijl}}{\partial \phi}}{p_{ijl}} + (1 - \delta_{ij}) \frac{\frac{\partial C}{\partial \phi}}{C} \right] + \frac{(\theta_i - X_i\phi - Z_i\psi_i)' X_i}{2\sigma^2_{\varepsilon_i}},$$

$$\frac{\partial P_{ijl}}{\partial \phi} = P_{ijl}(1 - P_{ijl}) \frac{\partial \eta_{ijl}}{\partial \phi}, \;\; \frac{\partial C}{\partial \phi} = -\frac{\partial P_{ijl}}{\partial \phi} + S_u\big(t_{ij}|y_{ij} = 1, x^*_{ij}, z^*_{ij}\big) \frac{\partial p_{ijl}}{\partial \phi},$$

$$\frac{\partial \alpha_{0i}}{\partial \phi} = X_i \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \;\; \frac{\partial \alpha_{1i}}{\partial \phi} = X_i \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \;\; \frac{\partial \beta_{0i}}{\partial \phi} = X_i \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \;\; \frac{\partial \beta_{1i}}{\partial \phi} = X_i \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

$$\frac{\partial \eta_{ijl}}{\partial \phi} = X_i \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + X_i \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \exp(\alpha_{1i}) \; \mu_0(t^*_{ij}) + \exp(\alpha_{1i}) \; \frac{\partial \mu_0(t^*_{ij})}{\partial \phi},$$

$$\frac{\partial \mu_0(t^*_{ij})}{\partial \phi} = \frac{\partial \mu_0(t^*_{ij})}{\partial t^*_{ij}} \; \frac{\partial t^*_{ij}}{\partial \phi},$$

$$\frac{\partial \mu_0(t^*_{ij})}{\partial t^*_{ij}} = \sum_{m=1}^{4} (m-1){t^*_{ij}}^{m-2}\gamma_m \ + \sum_{k=1}^{K} 3\Big(t^*_{ij} - \xi_k\Big)^2 \zeta_k,$$

$$\frac{\partial t^*_{ij}}{\partial \phi} = X_i \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + X_i \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \exp(\beta_{1i}) t_{ij}.$$

The first derivative of l_{ic} respect to $\sigma^2_{\psi^2_i}$:

$$\frac{\partial l_{ic}}{\partial \sigma^2_{\psi_i}} = -\frac{1}{2\sigma^2_{\psi_i}} + \frac{\psi_i^2}{2\Big(\sigma^2_{\psi_i}\Big)^2}.$$

The first derivative of l_{ic} respect to $\sigma^2_{\varepsilon_i}$ is given by:

$$\frac{\partial l_{ic}}{\partial \sigma^2_{\varepsilon_i}} = -\frac{1}{2\sigma^2_{\varepsilon_i}} + \frac{(\theta_i - X_i\phi - Z_i\psi_i)'\,(\theta_i - X_i\phi - Z_i\psi_i)}{2\Big(\sigma^2_{\varepsilon_i}\Big)^2}.$$

The first derivative of l_{ic} respect to ϕ^* is given by:

$$A = -(\frac{t_{ij}}{b})^a,$$

$$\frac{\partial l_{ic}}{\partial \phi^*} = \sum_{j=1}^{n_i} I_{y_{ij}}(l) \left[\delta_{ij} \frac{\partial \log f_u\big(t_{ij}|y_{ij}=1, x^*_{ij}, z^*_{ij}\big)}{\partial \phi^*} + (1-\delta_{ij}) \frac{\frac{\partial C}{\partial \phi^*}}{C} \right] + \frac{(\theta^*_i - X^*_i\phi^* - Z^*_i\psi^*_i)' X^*_i}{2\sigma^2_{\varepsilon^*_i}},$$

$$\frac{\partial \log f_u\big(t_{ij}|y_{ij}=1, x^*_{ij}, z^*_{ij}\big)}{\partial \phi^*} = x^*_{ij}(1 + exp(\theta^*)\ A),\ \ \frac{\partial C}{\partial \phi^*} = p_{ijl} \frac{\partial S_u\big(t_{ij}|y_{ij}=1, x^*_{ij}, z^*_{ij}\big)}{\partial \phi^*},$$

$$\frac{\partial S_u\big(t_{ij}|y_{ij}=1, x^*_{ij}, z^*_{ij}\big)}{\partial \phi^*} = A\ x^*_{ij}\ exp(\theta^*)\ S_u\big(t_{ij}|y_{ij}=1, x^*_{ij}, z^*_{ij}\big).$$

The first derivative of l_{ic} respect to $\sigma^2_{\psi^*_i}$ is given by:

$$\frac{\partial l_{ic}}{\partial \sigma^2_{\psi^*_i}} = -\frac{1}{2\sigma^2_{\psi^*_i}} + \frac{\psi_i^{*2}}{2\Big(\sigma^2_{\psi^*_i}\Big)^2}.$$

The first derivative of l_{ic} respect to $\sigma^2_{\varepsilon^*_i}$ is given by:

$$\frac{\partial l_{ic}}{\partial \sigma^2_{\varepsilon_i^*}} = -\frac{1}{2\sigma^2_{\varepsilon_i^*}} + \frac{(\theta_i^* - X_i^*\phi^* - Z_i^*\psi_i^*)'\,(\theta_i^* - X_i^*\phi^* - Z_i^*\psi_i^*)}{2\left(\sigma^2_{\varepsilon_i^*}\right)^2}.$$

The second derivative of l_{ic} respect to thresholds τ_l, for $l = 1, \ldots, L$ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \tau_l' \partial \tau_l} = \sum_{j=1}^{n_i} I_{y_{ij}}(l)\left[\delta_{ij}\frac{\frac{\partial^2 p_{ijl}}{\partial \tau_l' \partial \tau_l}p_{ijl} - \frac{\partial p_{ijl}}{\partial \tau_l'}\frac{\partial p_{ijl}}{\partial \tau_l}}{p_{ijl}^2} + (1-\delta_{ij})\frac{\frac{\partial^2 C}{\partial \tau_l' \partial \tau_l}C - \frac{\partial C}{\partial \tau_l'}\frac{\partial C}{\partial \tau_l}}{C^2}\right],$$

$$\frac{\partial^2 l_{ic}}{\partial \tau_l \partial \tau_{l'}} = 0, \qquad l \neq l'.$$

$$\frac{\partial^2 P_{ijl}}{\partial {\tau_l}^2} = P_{ijl}(1-P_{ijl})(1-2P_{ijl}),$$

$$\frac{\partial^2 P_{ijl}}{\partial \tau_l \partial {\tau_l}'} = 0, \qquad l \neq l'.$$

$$\frac{\partial^2 C}{\partial \tau_l' \partial \tau_l} = -\frac{\partial^2 P_{ijl}}{\partial \tau_l' \partial \tau_l} + S_u\big(t_{ij}|y_{ij}=1, x_{ij}^*, z_{ij}^*\big)\frac{\partial^2 p_{ijl}}{\partial \tau_l' \partial \tau_l}.$$

The second derivative of l_{ic} respect to ϕ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \phi' \partial \phi} = \sum_{j=1}^{n_i} I_{y_{ij}}(l)\left[\delta_{ij}\frac{\frac{\partial^2 p_{ijl}}{\partial \phi' \partial \phi}p_{ijl} - \frac{\partial p_{ijl}}{\partial \phi'}\frac{\partial p_{ijl}}{\partial \phi}}{p_{ijl}^2} + (1-\delta_{ij})\frac{\frac{\partial^2 C}{\partial \phi' \partial \phi}C - \frac{\partial C}{\partial \phi'}\frac{\partial C}{\partial \phi}}{C^2}\right] - \frac{X_i' X_i}{2\sigma^2_{\varepsilon_i}},$$

$$\frac{\partial^2 P_{ijl}}{\partial \phi' \partial \phi} = P_{ijl}(1-P_{ijl})^2\frac{\partial \eta_{ijl}}{\partial \phi'}\frac{\partial \eta_{ijl}}{\partial \phi} - P_{ijl}^2(1-P_{ijl})\frac{\partial \eta_{ijl}}{\partial \phi'}\frac{\partial \eta_{ijl}}{\partial \phi} + P_{ijl}(1-P_{ijl})\frac{\partial^2 \eta_{ijl}}{\partial \phi' \partial \phi},$$

$$\frac{\partial^2 C}{\partial \phi' \partial \phi} = -\frac{\partial^2 P_{ijl}}{\partial \phi' \partial \phi} + S_u\big(t_{ij}|y_{ij}=1, x_{ij}^*, z_{ij}^*\big)\frac{\partial^2 p_{ijl}}{\partial \phi' \partial \phi},$$

$$\begin{aligned}\frac{\partial^2 \eta_{ijl}}{\partial \phi' \partial \phi} = {} & X_i\begin{pmatrix}0\\1\\0\\0\end{pmatrix}\begin{pmatrix}0 & 1 & 0 & 0\end{pmatrix}X_i'\ \exp(\alpha_{1i})\ \mu_0(t_{ij}^*) + X_i\begin{pmatrix}0\\1\\0\\0\end{pmatrix}\exp(\alpha_{1i})\frac{\partial \mu_0(t_{ij}^*)}{\partial \phi'} \\ & + \begin{pmatrix}0 & 1 & 0 & 0\end{pmatrix}X_i' \exp(\alpha_{1i})\frac{\partial \mu_0(t_{ij}^*)}{\partial \phi} + \exp(\alpha_{1i})\frac{\partial^2 \mu_0(t_{ij}^*)}{\partial \phi' \partial \phi},\end{aligned}$$

$$\frac{\partial^2 \mu_0(t^*_{ij})}{\partial \phi' \, \partial \phi} = \frac{\partial^2 \mu_0(t^*_{ij})}{\partial \phi' \, \partial t^*_{ij}} \, \frac{\partial t^*_{ij}}{\partial \phi} + \frac{\partial \mu_0(t^*_{ij})}{\partial t^*_{ij}} \, \frac{\partial^2 t^*_{ij}}{\partial \phi' \, \partial \phi},$$

$$\frac{\partial^2 \mu_0(t^*_{ij})}{\partial \phi' \, \partial t^*_{ij}} = \sum_{m=1}^{4} (m-1)(m-2) \frac{\partial t^*_{ij}}{\partial \phi'} {t^*_{ij}}^{m-3} \gamma_m + \sum_{k=1}^{K} 6 \frac{\partial t^*_{ij}}{\partial \phi'} \Big(t^*_{ij} - \xi_k \Big) \zeta_k,$$

$$\frac{\partial t^*_{ij}}{\partial \phi'} = \begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix} X_i' + \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} X_i' \exp(\beta_{1i}) t_{ij},$$

$$\frac{\partial^2 t^*_{ij}}{\partial \phi' \partial \phi} = X_i \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} X_i' \, \exp(\beta_{1i}) t_{ij}.$$

The second derivative of l_{ic} respect to ϕ and $\sigma^2_{\varepsilon_i}$ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \sigma^2_{\varepsilon_i} \partial \phi} = -\frac{(\theta_i - X_i \phi - Z_i \psi_i)' X_i}{2 \Big(\sigma^2_{\varepsilon_i} \Big)^2}.$$

The second derivative of l_{ic} respect to $\sigma^2_{\psi_i}$ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \big(\sigma^2_{\psi_i} \big)^2} = \frac{1}{2 \big(\sigma^2_{\psi_i} \big)^2} - \frac{\psi_i^2}{\big(\sigma^2_{\psi_i} \big)^3}.$$

The second derivative of l_{ic} respect to $\sigma^2_{\varepsilon_i}$ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \Big(\sigma^2_{\varepsilon_i} \Big)^2} = \frac{1}{2 \left(\sigma^2_{\varepsilon_i} \right)^2} - \frac{(\theta_i - X_i \phi - Z_i \psi_i)' \, (\theta_i - X_i \phi - Z_i \psi_i)}{\Big(\sigma^2_{\varepsilon_i} \Big)^3}.$$

The second derivative of l_{ic} respect to thresholds τ_l and ϕ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \phi \partial \tau_l} = \sum_{j=1}^{n_i} I_{y_{ij}}(l) \left[\delta_{ij} \frac{\frac{\partial^2 p_{ijl}}{\partial \phi \partial \tau_l} p_{ijl} - \frac{\partial p_{ijl}}{\partial \phi} \frac{\partial p_{ijl}}{\partial \tau_l}}{p^2_{ijl}} + (1 - \delta_{ij}) \frac{\frac{\partial^2 C}{\partial \phi \partial \tau_l} C - \frac{\partial C}{\partial \phi} \frac{\partial C}{\partial \tau_l}}{C^2} \right],$$

$$\frac{\partial^2 P_{ijl}}{\partial \phi \partial \tau_l} = P_{ijl}(1 - P_{ijl})(1 - 2P_{ijl}) \frac{\partial \eta_{ijl}}{\partial \phi}.$$

$$\frac{\partial^2 C}{\partial \phi \partial \tau_l} = -\frac{\partial^2 P_{ijl}}{\partial \phi \partial \tau_l} + S_u \big(t_{ij} | y_{ij} = 1, x^*_{ij}, z^*_{ij} \big) \frac{\partial^2 p_{ijl}}{\partial \phi \partial \tau_l}.$$

The second derivative of l_{ic} respect to ϕ^* is given by:

$$\frac{\partial^2 l_{ic}}{\partial \phi^{*'} \partial \phi^*} = \sum_{j=1}^{n_i} I_{y_{ij}}(l) \left[\delta_{ij} \frac{\partial^2 \log f_u\big(t_{ij}|y_{ij}=1, x_{ij}^*, z_{ij}^*\big)}{\partial \phi^{*'} \partial \phi^*} + (1-\delta_{ij}) \frac{\frac{\partial^2 C}{\partial \phi^{*'} \partial \phi^*} C - \frac{\partial C}{\partial \phi^{*'}} \frac{\partial C}{\partial \phi^*}}{C^2} \right] - \frac{X_i^{*'} X_i^*}{2\sigma^2_{\varepsilon_i^*}},$$

$$\frac{\partial^2 \log f_u\big(t_{ij}|y_{ij}=1, x_{ij}^*, z_{ij}^*\big)}{\partial \phi^{*'} \partial \phi^*} = x_{ij}^{*'}\, x_{ij}^*\, exp(-\theta^*)\, A, \;\; \frac{\partial^2 C}{\partial \phi^{*'} \partial \phi^*} = p_{ijl} \frac{\partial^2 S_u\big(t_{ij}|y_{ij}=1, x_{ij}^*, z_{ij}^*\big)}{\partial \phi^{*'} \partial \phi^*},$$

$$\frac{\partial^2 S_u\big(t_{ij}|y_{ij}=1, x_{ij}^*, z_{ij}^*\big)}{\partial \phi^{*'} \partial \phi^*} = x_{ij}^*\; exp(-\theta^*)\; A \left[x_{ij}^{*'} S_u + \frac{\partial S_u}{\partial \phi^{*'}} \right].$$

The second derivative of l_{ic} respect to ϕ^* and $\sigma^2_{\varepsilon_i^*}$ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \sigma^2_{\varepsilon_i^*} \partial \phi^*} = - \frac{(\theta_i^* - X_i^* \phi^* - Z_i^* \psi_i^*)' X_i^*}{2\left(\sigma^2_{\varepsilon_i^*}\right)^2}.$$

The second derivative of l_{ic} respect to $\sigma^2_{\psi_i^*}$ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \left(\sigma^2_{\psi_i^*}\right)^2} = \frac{1}{2\left(\sigma^2_{\psi_i^*}\right)^2} - \frac{\psi_i^{*2}}{\left(\sigma^2_{\psi_i^*}\right)^3}.$$

The second derivative of l_{ic} respect to $\sigma^2_{\varepsilon_i^*}$ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \left(\sigma^2_{\varepsilon_i^*}\right)^2} = \frac{1}{2\left(\sigma^2_{\varepsilon_i^*}\right)^2} - \frac{(\theta_i^* - X_i^* \phi^* - Z_i^* \psi_i^*)' (\theta_i^* - X_i^* \phi^* - Z_i^* \psi_i^*)}{\left(\sigma^2_{\varepsilon_i^*}\right)^3}.$$

The second derivative of l_{ic} respect to ϕ and ϕ^* is given by:

$$\frac{\partial^2 l_{ic}}{\partial \phi\; \partial \phi^*} = \sum_{j=1}^{n_i} (1-\delta_{ij}) \frac{\frac{\partial^2 C}{\partial \phi\; \partial \phi^*} C - \frac{\partial C}{\partial \phi} \frac{\partial C}{\partial \phi^*}}{C^2}, \;\; \frac{\partial^2 C}{\partial \phi\; \partial \phi^*} = \frac{\partial p_{ijl}}{\partial \phi} \frac{\partial S_u\big(t_{ij}|y_{ij}=1, x_{ij}^*, z_{ij}^*\big)}{\partial \phi^*}.$$

A.3.2 The E- and M- steps

E- step:

$$
\begin{aligned}
Q\left(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\right) \approx & \sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\left(\sum_{j=1}^{n_i}\sum_{l=1}^{L} I_{y_{ij}}(l)\Big[\delta_{ij}\log(p_{ijl})+\delta_{ij}\log f_u\big(t_{ij}|y_{ij}=1,x_{ij}^{*},z_{ij}^{*}\big)\right.\\
& +(1-\delta_{ij})\log\Big(1-P_{ijl}+p_{ijl}S_u\big(t_{ij}|y_{ij}=1,x_{ij}^{*},z_{ij}^{*}\big)\Big)\Big]\\
& -\frac{1}{2}log\left(\sigma_{\varepsilon_i}^{2(s)}\right)-\frac{\left(\theta_i-X_i\phi^{(s)}-Z_i\psi_i^{(s)}\right)'\left(\theta_i-X_i\phi^{(s)}-Z_i\psi_i^{(s)}\right)}{2\sigma_{\varepsilon_i}^{2(s)}}\\
& -\frac{1}{2}log\left(\sigma_{\varepsilon_i^{*}}^{2(s)}\right)-\frac{\left(\theta_i^{*}-X_i^{*}\phi^{*(s)}-Z_i^{*}\psi_i^{*(s)}\right)'\left(\theta_i^{*}-X_i^{*}\phi^{*(s)}-Z_i^{*}\psi_i^{*(s)}\right)}{2\sigma_{\varepsilon_i^{*}}^{2(s)}}\\
& \left.-\frac{1}{2}log\left(\sigma_{\psi_i}^{2(s)}\right)-\frac{\psi_i^{2(s)}}{2\sigma_{\psi_i}^{2(s)}}-\frac{1}{2}log\left(\sigma_{\psi_i^{*}}^{2(s)}\right)-\frac{\psi^{*}{}_i^{2(s)}}{2\sigma_{\psi_i^{*}}^{2(s)}}\right).
\end{aligned}
$$

M- step: For $l = 1,\ldots,L$

$$
\frac{\partial Q\left(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\right)}{\partial\tau_l}=\sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\sum_{j=1}^{n_i} I_{y_{ijl}}(l)\left[\delta_{ij}\frac{\frac{\partial p_{ijl}}{\partial\tau_l}}{p_{ijl}}+(1-\delta_{ij})\frac{\frac{\partial C}{\partial\tau_l}}{C}\right]=0,
$$

$$
\frac{P_{ijl}(1-P_{ijl})}{P_{ijl}-P_{ijl-1}}=\frac{\exp(\tau_l)}{\exp(\tau_l)-\exp(\tau_{l-1})}\,\frac{1+\exp(\tau_{l-1})\exp(\omega_{ij})}{1+\exp(\tau_l)\exp(\omega_{ij})},
$$

$$
\begin{aligned}
\frac{\partial Q\left(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\right)}{\partial\tau_l} & =\sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\sum_{j=1}^{n_i} I_{y_{ij}}(l)\left[\delta_{ij}\frac{1}{\exp(\tau_l)-\exp(\tau_{l-1})}\,\frac{1+\exp(\tau_{l-1})\exp(\omega_{ij}^{(s)})}{1+\exp(\tau_l)\exp(\omega_{ij}^{(s)})}\right.\\
& \left.-(1-\delta_{ij})\frac{P_{ijl}(1-P_{ijl})(S_u-1)}{C}\right]=0.
\end{aligned}
$$

For the parameter τ_l, we can't get the closed form expression for the maximizer $Q\left(\boldsymbol{\theta}\mid\boldsymbol{\theta}^{(s)}\right)$. We use the Newton–Raphson iterations as:

$$
\begin{aligned}
\frac{\partial Q\left(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\right)}{\partial\phi} & =\sum_{i=1}^{N}\frac{1}{M_i}\sum_{m=1}^{M_i}\left(\sum_{j=1}^{n_i}\sum_{l=1}^{L} I_{y_{ij}}(l)\left[\delta_{ij}\frac{\frac{\partial p_{ijl}}{\partial\phi}}{p_{ijl}}+(1-\delta_{ij})\frac{\frac{\partial C}{\partial\phi}}{C}\right]\right.\\
& \left.+\frac{\left(\theta_i-X_i\phi^{(s)}-Z_i\psi_i^{(s)}\right)'X_i}{2\sigma_{\varepsilon_i}^{2(s)}}\right)=0,
\end{aligned}
$$

$$\phi^{(s+1)} = \sum_{i=1}^{N} \frac{1}{M_i X_i' X_i} \sum_{m=1}^{M_i} \left(2\sigma_{\varepsilon_i}^{2(s)} \sum_{j=1}^{n_i} \sum_{l=1}^{L} I_{y_{ij}}(l) \left[\delta_{ij} \frac{\frac{\partial p_{ijl}}{\partial \phi}}{p_{ijl}} + (1-\delta_{ij}) \frac{\frac{\partial C}{\partial \phi}}{C} \right] + \theta_i' X_i - {\psi_i'}^{(s)} Z_i' X_i \right).$$

$$\frac{\partial Q\left(\boldsymbol{\theta} \,|\boldsymbol{\theta}^{(s)}\right)}{\partial \sigma_{\varepsilon_i}^2} = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left[-\frac{1}{2\sigma_{\varepsilon_i}^{2(s)}} + \frac{\left(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\right)' \left(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\right)}{2\left(\sigma_{\varepsilon_i}^{2(s)}\right)^2} \right] = 0,$$

$$\sigma_{\varepsilon_i}^{2(s+1)} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\right)' \left(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\right).$$

$$\frac{\partial Q\left(\boldsymbol{\theta} \,|\boldsymbol{\theta}^{(s)}\right)}{\partial \sigma_{\psi_i}^2} = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left[-\frac{1}{2\sigma_{\psi_i}^{2(s)}} + \frac{\psi_i^{2(s)}}{2\left(\sigma_{\psi_i}^{2(s)}\right)^2} \right] = 0, \;\; \sigma_{\psi_i}^{2(s+1)} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \psi_i^{2(s)}.$$

$$\frac{\partial Q\left(\boldsymbol{\theta} \,|\boldsymbol{\theta}^{(s)}\right)}{\partial \phi^*} = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left(\sum_{j=1}^{n_i} \left[\delta_{ij} \frac{\partial \log f_u\left(t_{ij}|y_{ij}=1, x_{ij}^*, z_{ij}^*\right)}{\partial \phi^*} + (1-\delta_{ij}) \frac{\frac{\partial C}{\partial \phi^*}}{C} \right] \right.$$
$$\left. + \frac{(\theta_i^* - X_i^*\phi^* - Z_i^*\psi_i^*)' X_i^*}{2\sigma_{\varepsilon_i^*}^2}, \right) = 0,$$

$$\phi^{*(s+1)} = \sum_{i=1}^{N} \frac{1}{M_i {X_i^*}' X_i^*} \sum_{m=1}^{M_i} \left(2\sigma_{\varepsilon_i}^{2(s)} \sum_{j=1}^{n_i} \left[\delta_{ij} \frac{\partial \log f_u\left(t_{ij}|y_{ij}=1, x_{ij}^*, z_{ij}^*\right)}{\partial \phi^*} + (1-\delta_{ij}) \frac{\frac{\partial C}{\partial \phi^*}}{C} \right] \right.$$
$$\left. + {\theta_i^*}' X_i^* - {\psi_i^{*}}'^{(s)} {Z_i^*}' X_i^* \right).$$

$$\frac{\partial Q\left(\boldsymbol{\theta} \,|\boldsymbol{\theta}^{(s)}\right)}{\partial \sigma_{\varepsilon_i^*}^2} = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left[-\frac{1}{2\sigma_{\varepsilon_i^*}^{2(s)}} \right.$$
$$\left. + \frac{\left(\theta_i^* - X_i^*\phi^{*(s)} - Z_i^*\psi_i^{*(s)}\right)' \left(\theta_i^* - X_i^*\phi^{*(s)} - Z_i^*\psi_i^{*(s)}\right)}{2\left(\sigma_{\varepsilon_i^*}^{2(s)}\right)^2} \right] = 0,$$

$$\sigma_{\varepsilon_i^*}^{2(s+1)} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left(\theta_i^* - X_i^*\phi^{*(s)} - Z_i^*\psi_i^{*(s)}\right)' \left(\theta_i^* - X_i^*\phi^{*(s)} - Z_i^*\psi_i^{*(s)}\right).$$

$$\frac{\partial Q\left(\boldsymbol{\theta} \,|\boldsymbol{\theta}^{(s)}\right)}{\partial \sigma_{\psi_i^*}^2} = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left[-\frac{1}{2\sigma_{\psi_i^*}^{2(s)}} + \frac{\psi_i^{*2(s)}}{2\left(\sigma_{\psi_i^*}^{2(s)}\right)^2} \right] = 0, \;\; \sigma_{\psi_i^*}^{2(s+1)} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \psi_i^{*2(s)}.$$

A.4 Chapter 5

A.4.1 The Newton-Raphson procedure

The first derivative of l_{ic} respect to ϕ is given by:

$$C = 1 - p_{ij} + p_{ij} S_u\big(t_{ij}|y_{ij} = 1, x^*_{ij}, z^*_{ij}\big),$$

$$\frac{\partial l_{ic}}{\partial \phi} = \sum_{j=1}^{n_i} \left[\delta_{ij} \frac{\frac{\partial p_{ij}}{\partial \phi}}{p_{ij}} + (1-\delta_{ij}) \frac{\frac{\partial C}{\partial \phi}}{C} \right] + \frac{(\theta_i - X_i\phi - Z_i\psi_i)' X_i}{2\sigma^2_{\varepsilon_i}},$$

$$\frac{\partial p_{ij}}{\partial \phi} = p_{ij}(1-p_{ij}) \frac{\partial \eta_{ij}}{\partial \phi}, \;\; \frac{\partial C}{\partial \phi} = S_u\big(t_{ij}|y_{ij} = 1, x^*_{ij}, z^*_{ij}\big) \frac{\partial p_{ij}}{\partial \phi},$$

$$\frac{\partial \alpha_{0i}}{\partial \phi} = X_i \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \;\; \frac{\partial \alpha_{1i}}{\partial \phi} = X_i \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \;\; \frac{\partial \beta_{0i}}{\partial \phi} = X_i \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \;\; \frac{\partial \beta_{1i}}{\partial \phi} = X_i \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

$$\frac{\partial \eta_{ijl}}{\partial \phi} = X_i \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + X_i \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \exp(\alpha_{1i}) \; \mu_0(t^*_{ij}) + \exp(\alpha_{1i}) \; \frac{\partial \mu_0(t^*_{ij})}{\partial \phi},$$

$$\frac{\partial \mu_0(t^*_{ij})}{\partial \phi} = \frac{\partial \mu_0(t^*_{ij})}{\partial t^*_{ij}} \; \frac{\partial t^*_{ij}}{\partial \phi},$$

$$\frac{\partial \mu_0(t^*_{ij})}{\partial t^*_{ij}} = \sum_{m=1}^{4} (m-1) {t^*_{ij}}^{m-2} \gamma_m \; + \sum_{k=1}^{K} 3\Big(t^*_{ij} - \xi_k\Big)^2 \zeta_k,$$

$$\frac{\partial t^*_{ij}}{\partial \phi} = X_i \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + X_i \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \exp(\beta_{1i}) t_{ij}.$$

The first derivative of l_{ic} respect to $\sigma^2_{\psi^2_i}$ is given by:

$$\frac{\partial l_{ic}}{\partial \sigma^2_{\psi_i}} = -\frac{1}{2\sigma^2_{\psi_i}} + \frac{\psi_i^2}{2\Big(\sigma^2_{\psi_i}\Big)^2}.$$

The first derivative of l_{ic} respect to $\sigma^2_{\varepsilon_i}$ is given by:

$$\frac{\partial l_{ic}}{\partial \sigma^2_{\varepsilon_i}} = -\frac{1}{2\sigma^2_{\varepsilon_i}} + \frac{(\theta_i - X_i\phi - Z_i\psi_i)'\,(\theta_i - X_i\phi - Z_i\psi_i)}{2\left(\sigma^2_{\varepsilon_i}\right)^2}.$$

The first derivative of l_{ic} respect to ϕ^* is given by:

$$A = -(b\ t_{ij}\ exp(\theta^*))^a,$$

$$\frac{\partial l_{ic}}{\partial \phi^*} = \sum_{j=1}^{n_i}\left[\delta_{ij}\frac{\partial \log f_u\big(t_{ij}|y_{ij}=1, x^*_{ij}, z^*_{ij}\big)}{\partial \phi^*} + (1-\delta_{ij})\frac{\frac{\partial C}{\partial \phi^*}}{C}\right] + \frac{(\theta^*_i - X^*_i\phi^* - Z^*_i\psi^*_i)'X^*_i}{2\sigma^2_{\varepsilon^*_i}},$$

$$\frac{\partial \log f_u\big(t_{ij}|y_{ij}=1, x^*_{ij}, z^*_{ij}\big)}{\partial \phi^*} = (a-1)x^*_{ij}(1-A\,exp(-\theta^*)),\ \frac{\partial C}{\partial \phi^*} = p_{ij}\frac{\partial S_u\big(t_{ij}|y_{ij}=1, x^*_{ij}, z^*_{ij}\big)}{\partial \phi^*},$$

$$\frac{\partial S_u\big(t_{ij}|y_{ij}=1, x^*_{ij}, z^*_{ij}\big)}{\partial \phi^*} = exp(-\theta^*)\Big[-x^*_{ij}A + \frac{\partial A}{\partial \phi^*}\Big]S_u\big(t_{ij}|y_{ij}=1, x^*_{ij}, z^*_{ij}\big),\ \frac{\partial A}{\partial \phi^*} = -ax^*_{ij}A.$$

The first derivative of l_{ic} respect to $\sigma^2_{\psi^*_i}$ is given by:

$$\frac{\partial l_{ic}}{\partial \sigma^2_{\psi^*_i}} = -\frac{1}{2\sigma^2_{\psi^*_i}} + \frac{\psi^{*2}_i}{2\left(\sigma^2_{\psi^*_i}\right)^2}.$$

The first derivative of l_{ic} respect to $\sigma^2_{\varepsilon^*_i}$ is given by:

$$\frac{\partial l_{ic}}{\partial \sigma^2_{\varepsilon^*_i}} = -\frac{1}{2\sigma^2_{\varepsilon^*_i}} + \frac{(\theta^*_i - X^*_i\phi^* - Z^*_i\psi^*_i)'\,(\theta^*_i - X^*_i\phi^* - Z^*_i\psi^*_i)}{2\left(\sigma^2_{\varepsilon^*_i}\right)^2}.$$

The second derivative of l_{ic} respect to ϕ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \phi' \partial \phi} = \sum_{j=1}^{n_i}\left[\delta_{ij}\frac{\frac{\partial^2 p_{ij}}{\partial \phi' \partial \phi}p_{ij} - \frac{\partial p_{ij}}{\partial \phi'}\frac{\partial p_{ij}}{\partial \phi}}{p^2_{ij}} + (1-\delta_{ij})\frac{\frac{\partial^2 C}{\partial \phi' \partial \phi}C - \frac{\partial C}{\partial \phi'}\frac{\partial C}{\partial \phi}}{C^2}\right] - \frac{X'_iX_i}{2\sigma^2_{\varepsilon_i}},$$

$$\frac{\partial^2 P_{ijl}}{\partial \phi' \partial \phi} = P_{ijl}(1-P_{ijl})^2\frac{\partial \eta_{ijl}}{\partial \phi'}\frac{\partial \eta_{ijl}}{\partial \phi} - P^2_{ijl}(1-P_{ijl})\frac{\partial \eta_{ijl}}{\partial \phi'}\frac{\partial \eta_{ijl}}{\partial \phi} + P_{ijl}(1-P_{ijl})\frac{\partial^2 \eta_{ijl}}{\partial \phi' \partial \phi},$$

$$\frac{\partial^2 C}{\partial \phi' \partial \phi} = \frac{\partial^2 P_{ijl}}{\partial \phi' \partial \phi} S_u\big(t_{ij}|y_{ij} = 1, x^*_{ij}, z^*_{ij}\big),$$

$$\frac{\partial^2 \eta_{ijl}}{\partial \phi' \partial \phi} = X_i \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix} X_i' \ \exp(\alpha_{1i}) \ \mu_0(t^*_{ij}) + X_i \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \exp(\alpha_{1i}) \frac{\partial \mu_0(t^*_{ij})}{\partial \phi'}$$
$$+ \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix} X_i' \exp(\alpha_{1i}) \frac{\partial \mu_0(t^*_{ij})}{\partial \phi} + \exp(\alpha_{1i}) \frac{\partial^2 \mu_0(t^*_{ij})}{\partial \phi' \partial \phi},$$

$$\frac{\partial^2 \mu_0(t^*_{ij})}{\partial \phi' \ \partial \phi} = \frac{\partial^2 \mu_0(t^*_{ij})}{\partial \phi' \ \partial t^*_{ij}} \ \frac{\partial t^*_{ij}}{\partial \phi} + \frac{\partial \mu_0(t^*_{ij})}{\partial t^*_{ij}} \ \frac{\partial^2 t^*_{ij}}{\partial \phi' \ \partial \phi},$$

$$\frac{\partial^2 \mu_0(t^*_{ij})}{\partial \phi' \ \partial t^*_{ij}} = \sum_{m=1}^{4} (m-1)(m-2) \frac{\partial t^*_{ij}}{\partial \phi'} {t^*_{ij}}^{m-3} \gamma_m + \sum_{k=1}^{K} 6 \frac{\partial t^*_{ij}}{\partial \phi'} \Big(t^*_{ij} - \xi_k\Big) \zeta_k,$$

$$\frac{\partial t^*_{ij}}{\partial \phi'} = \begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix} X_i' + \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} X_i' \ \exp(\beta_{1i}) t_{ij},$$

$$\frac{\partial^2 t^*_{ij}}{\partial \phi' \partial \phi} = X_i \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} X_i' \ \exp(\beta_{1i}) t_{ij}.$$

The second derivative of l_{ic} respect to ϕ and $\sigma^2_{\varepsilon_i}$ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \sigma^2_{\varepsilon_i} \partial \phi} = -\frac{(\theta_i - X_i \phi - Z_i \psi_i)' X_i}{2 \Big(\sigma^2_{\varepsilon_i}\Big)^2}.$$

The second derivative of l_{ic} respect to $\sigma^2_{\psi_i}$ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \big(\sigma^2_{\psi_i}\big)^2} = \frac{1}{2 \big(\sigma^2_{\psi_i}\big)^2} - \frac{\psi_i^2}{\big(\sigma^2_{\psi_i}\big)^3}.$$

The second derivative of l_{ic} respect to $\sigma^2_{\varepsilon_i}$ is given by:

$$\frac{\partial^2 l_{ic}}{\partial \Big(\sigma^2_{\varepsilon_i}\Big)^2} = \frac{1}{2 \left(\sigma^2_{\varepsilon_i}\right)^2} - \frac{(\theta_i - X_i \phi - Z_i \psi_i)' \left(\theta_i - X_i \phi - Z_i \psi_i\right)}{\Big(\sigma^2_{\varepsilon_i}\Big)^3}.$$

The second derivative of l_{ic} respect to ϕ^* is given by:

$$\frac{\partial^2 l_{ic}}{\partial\phi^{*\prime}\partial\phi^*} = \sum_{j=1}^{n_i}\left[\delta_{ij}\frac{\partial^2 \log f_u\big(t_{ij}|y_{ij}=1,x_{ij}^*,z_{ij}^*\big)}{\partial\phi^{*\prime}\partial\phi^*} + (1-\delta_{ij})\frac{\frac{\partial^2 C}{\partial\phi^{*\prime}\partial\phi^*}C - \frac{\partial C}{\partial\phi^{*\prime}}\frac{\partial C}{\partial\phi^*}}{C^2}\right] - \frac{X_i^{*\prime}X_i^*}{2\sigma^2_{\varepsilon_i^*}},$$

$$\frac{\partial^2 \log f_u\big(t_{ij}|y_{ij}=1,x_{ij}^*,z_{ij}^*\big)}{\partial\phi^{*\prime}\partial\phi^*} = -(a-1)x_{ij}^* exp(-\theta^*)\Big(\frac{\partial A}{\partial\phi^{*\prime}} - Ax_{ij}^{*\prime}\Big),$$

$$\frac{\partial^2 C}{\partial\phi^{*\prime}\partial\phi^*} = p_{ij}\frac{\partial^2 S_u\big(t_{ij}|y_{ij}=1,x_{ij}^*,z_{ij}^*\big)}{\partial\phi^{*\prime}\partial\phi^*}, \quad \frac{\partial^2 A}{\partial\phi^{*\prime}\partial\phi^*} = -ax_{ij}^*\frac{\partial A}{\partial\phi^{*\prime}}.$$

$$\begin{aligned}\frac{\partial^2 S_u\big(t_{ij}|y_{ij}=1,x_{ij}^*,z_{ij}^*\big)}{\partial\phi^{*\prime}\partial\phi^*} &= exp(-\theta^*)\Big[x_{ij}^*x_{ij}^{*\prime}A - x_{ij}^*\frac{\partial A}{\partial\phi^{*\prime}} - x_{ij}^{*\prime}\frac{\partial A}{\partial\phi^*} + \frac{\partial^2 A}{\partial\phi^{*\prime}\partial\phi^*}\Big]S_u \\ &+ exp(-\theta^*)\Big[-x_{ij}^*A + \frac{\partial A}{\partial\phi^*}\Big]\frac{\partial S_u\big(t_{ij}|y_{ij}=1,x_{ij}^*,z_{ij}^*\big)}{\partial\phi^{*\prime}}.\end{aligned}$$

The second derivative of l_{ic} respect to ϕ^* and $\sigma^2_{\varepsilon_i^*}$ is given by:

$$\frac{\partial^2 l_{ic}}{\partial\sigma^2_{\varepsilon_i^*}\partial\phi^*} = -\frac{(\theta_i^* - X_i^*\phi^* - Z_i^*\psi_i^*)'X_i^*}{2\left(\sigma^2_{\varepsilon_i^*}\right)^2}.$$

The second derivative of l_{ic} respect to $\sigma^2_{\psi_i^*}$ is given by:

$$\frac{\partial^2 l_{ic}}{\partial\left(\sigma^2_{\psi_i^*}\right)^2} = \frac{\partial}{\partial\sigma^2_{\psi_i^*}}\left(-\frac{1}{2\sigma^2_{\psi_i^*}} + \frac{\psi_i^*}{2\sigma^2_{\psi_i^*}}\right) = \frac{1}{2\left(\sigma^2_{\psi_i^*}\right)^2} - \frac{\psi_i^{*2}}{\left(\sigma^2_{\psi_i^*}\right)^3}.$$

The second derivative of l_{ic} respect to $\sigma^2_{\varepsilon_i^*}$ is given by:

$$\frac{\partial^2 l_{ic}}{\partial\left(\sigma^2_{\varepsilon_i^*}\right)^2} = \frac{1}{2\left(\sigma^2_{\varepsilon_i^*}\right)^2} - \frac{(\theta_i^* - X_i^*\phi^* - Z_i^*\psi_i^*)'(\theta_i^* - X_i^*\phi^* - Z_i^*\psi_i^*)}{\left(\sigma^2_{\varepsilon_i^*}\right)^3}.$$

The second derivative of l_{ic} respect to ϕ and ϕ^* is given by:

$$\frac{\partial^2 l_{ic}}{\partial\phi\,\partial\phi^*} = \sum_{j=1}^{n_i}(1-\delta_{ij})\frac{\frac{\partial^2 C}{\partial\phi\,\partial\phi^*}C - \frac{\partial C}{\partial\phi}\frac{\partial C}{\partial\phi^*}}{C^2}, \quad \frac{\partial^2 C}{\partial\phi\,\partial\phi^*} = \frac{\partial p_{ij}}{\partial\phi}\frac{\partial S_u\big(t_{ij}|y_{ij}=1,x_{ij}^*,z_{ij}^*\big)}{\partial\phi^*}.$$

A.4.2 The E- and M- steps

E- step:

$$
\begin{aligned}
Q\big(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\big) \approx & \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left(\sum_{j=1}^{n_i} \Big[\delta_{ij} \log(p_{ij}) + \delta_{ij} \log f_u\big(t_{ij}|y_{ij}=1, x_{ij}^*, z_{ij}^*\big) \right. \\
& + (1-\delta_{ij}) \log \Big(1 - p_{ij} + p_{ij} S_u\big(t_{ij}|y_{ij}=1, x_{ij}^*, z_{ij}^*\big)\Big)\Big] \\
& - \frac{1}{2} log\left(\sigma_{\varepsilon_i}^{2(s)}\right) - \frac{\Big(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\Big)' \Big(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\Big)}{2\sigma_{\varepsilon_i}^{2(s)}} \\
& - \frac{1}{2} log\left(\sigma_{\varepsilon_i^*}^{2(s)}\right) - \frac{\Big(\theta_i^* - X_i^*\phi^{*(s)} - Z_i^*\psi_i^{*(s)}\Big)' \Big(\theta_i^* - X_i^*\phi^{*(s)} - Z_i^*\psi_i^{*(s)}\Big)}{2\sigma_{\varepsilon_i^*}^{2(s)}} \\
& \left. - \frac{1}{2} log\left(\sigma_{\psi_i}^{2(s)}\right) - \frac{\psi_i^{2(s)}}{2\sigma_{\psi_i}^{2(s)}} - \frac{1}{2} log\left(\sigma_{\psi_i^*}^{2(s)}\right) - \frac{{\psi^*}_i^{2(s)}}{2\sigma_{\psi_i^*}^{2(s)}} \right).
\end{aligned}
$$

M- step:

$$
\begin{aligned}
\frac{\partial Q\left(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\right)}{\partial \phi} &= \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left(\sum_{j=1}^{n_i} \left[\delta_{ij} \frac{\frac{\partial p_{ij}}{\partial \phi}}{p_{ij}} + (1-\delta_{ij}) \frac{\frac{\partial C}{\partial \phi}}{C} \right] + \frac{\Big(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\Big)' X_i}{2\sigma_{\varepsilon_i}^{2(s)}} \right) \\
&= 0,
\end{aligned}
$$

$$
\phi^{(s+1)} = \sum_{i=1}^{N} \frac{1}{M_i X_i' X_i} \sum_{m=1}^{M_i} \left(2\sigma_{\varepsilon_i}^{2(s)} \sum_{j=1}^{n_i} \left[\delta_{ij} \frac{\frac{\partial p_{ij}}{\partial \phi}}{p_{ij}} + (1-\delta_{ij}) \frac{\frac{\partial C}{\partial \phi}}{C} \right] + \theta_i' X_i - {\psi_i'}^{(s)} Z_i' X_i \right).
$$

$$
\begin{aligned}
\frac{\partial Q\left(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\right)}{\partial \sigma_{\varepsilon_i}^2} &= \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left[-\frac{1}{2\sigma_{\varepsilon_i}^{2(s)}} + \frac{\Big(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\Big)' \Big(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\Big)}{2\Big(\sigma_{\varepsilon_i}^{2(s)}\Big)^2} \right] \\
&= 0,
\end{aligned}
$$

$$
\sigma_{\varepsilon_i}^{2(s+1)} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \Big(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\Big)' \Big(\theta_i - X_i\phi^{(s)} - Z_i\psi_i^{(s)}\Big).
$$

$$
\frac{\partial Q\left(\boldsymbol{\theta}\,|\boldsymbol{\theta}^{(s)}\right)}{\partial \sigma_{\psi_i}^2} = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left[-\frac{1}{2\sigma_{\psi_i}^{2(s)}} + \frac{\psi_i^{2(s)}}{2\Big(\sigma_{\psi_i}^{2(s)}\Big)^2} \right] = 0,\ \ \sigma_{\psi_i}^{2(s+1)} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \psi_i^{2(s)}.
$$

$$\frac{\partial Q\left(\boldsymbol{\theta}\left|\boldsymbol{\theta}^{(s)}\right.\right)}{\partial \phi^*} = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left(\sum_{j=1}^{n_i} \left[\delta_{ij} \frac{\partial \log f_u\big(t_{ij}|y_{ij}=1, x_{ij}^*, z_{ij}^*\big)}{\partial \phi^*} + (1-\delta_{ij}) \frac{\frac{\partial C}{\partial \phi^*}}{C} \right] \right.$$
$$\left. + \frac{(\theta_i^* - X_i^* \phi^* - Z_i^* \psi_i^*)' X_i^*}{2\sigma^2_{\varepsilon_i^*}}, \right) = 0,$$

$$\phi^{*(s+1)} = \sum_{i=1}^{N} \frac{1}{M_i X_i^{*'} X_i^*} \sum_{m=1}^{M_i} \left(2\sigma^{2(s)}_{\varepsilon_i} \sum_{j=1}^{n_i} \left[\delta_{ij} \frac{\partial \log f_u\big(t_{ij}|y_{ij}=1, x_{ij}^*, z_{ij}^*\big)}{\partial \phi^*} + (1-\delta_{ij}) \frac{\frac{\partial C}{\partial \phi^*}}{C} \right] \right.$$
$$\left. + \theta_i^{*'} X_i^* - \psi_i^{*'(s)} Z_i^{*'} X_i^* \right).$$

$$\frac{\partial Q\left(\boldsymbol{\theta}\left|\boldsymbol{\theta}^{(s)}\right.\right)}{\partial \sigma^2_{\varepsilon_i^*}} = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left[-\frac{1}{2\sigma^{2(s)}_{\varepsilon_i^*}} \right.$$
$$\left. + \frac{\left(\theta_i^* - X_i^* \phi^{*(s)} - Z_i^* \psi_i^{*(s)}\right)' \left(\theta_i^* - X_i^* \phi^{*(s)} - Z_i^* \psi_i^{*(s)}\right)}{2\left(\sigma^{2(s)}_{\varepsilon_i^*}\right)^2} \right] = 0,$$

$$\sigma^{2(s+1)}_{\varepsilon_i^*} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left(\theta_i^* - X_i^* \phi^{*(s)} - Z_i^* \psi_i^{*(s)}\right)' \left(\theta_i^* - X_i^* \phi^{*(s)} - Z_i^* \psi_i^{*(s)}\right).$$

$$\frac{\partial Q\left(\boldsymbol{\theta}\left|\boldsymbol{\theta}^{(s)}\right.\right)}{\partial \sigma^2_{\psi_i^*}} = \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \left[-\frac{1}{2\sigma^{2(s)}_{\psi_i^*}} + \frac{\psi_i^{*2(s)}}{2\left(\sigma^{2(s)}_{\psi_i^*}\right)^2} \right] = 0, \ \sigma^{2(s+1)}_{\psi_i^*} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{M_i} \sum_{m=1}^{M_i} \psi_i^{*2(s)}.$$