Multi-state Markov models with fixed and random effects

Wenyu Wang

A thesis submitted for the degree
of Doctor of Philosophy

Department of Statistical Science
University College London

2022

Declaration

I, Wenyu Wang confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Multi-state models describe a process where individuals move among a series of states over time. They are increasingly popular in a wide range of applications in biostatistics. For instance, breast cancer, HIV and ageing problems. There are two types of effects when describing the hazards for change of status: fixed effects and random effects. For the fixed-effects multi-state model, the characteristics of individuals are usually considered as covariates, such as age and gender. However, there is still some unobserved heterogeneity, which can be taken into account as random effects. Models with both fixed effects and random effects in survival analysis are called frailty models. A large number of papers discusses parametric univariate frailties in multi-state models. This study presents both parametric and non-parametric frailty models. For the parametric frailty model, both univariate and bivariate frailties in multi-state models are discussed, in which frailties follow several common distributions. In particular, the contribution of this study is to apply a bivariate gamma-distributed frailty in the multi-state model for the interval-censored data, in order to describe the unobserved heterogeneity and investigate the correlation between two transition hazards. Model validation and prediction are discussed as well. In the application, we illustrate both fixed-effect models and frailty models for a cardiac allograft vasculopathy study and a cognitive impairment process.

Impact statement

This thesis offers a number of models and applications which have a potential impact on both inside and outside academia. Firstly, this thesis developed a comprehensive overview of multi-state models. The multi-state model is a very useful method in a wide range of applications in biostatistics. The parametric distributions we have discussed in the model would provide multiple options for different applications. Next, we have explored the frailty model, which can describe the unobserved heterogeneity comparing with the fixed-effect multi-state model. Also, frailty models may counterbalance the violation of Markov assumption. Therefore, the frailty model is a good choice to check whether it will improve the model performance. Moreover, this thesis discussed both univariate and bivariate frailty models. The multi-state bivariate frailty model is quite useful in the application, since it can not only describe the unobserved heterogeneity, but also investigate the correlation between two transition hazards. In particular, this thesis presented a bivariate gamma frailty model, which has the lowest AIC value among fitted models in the study. Therefore, this thesis provides more options for researchers who would like to use frailty models.

Acknowledgements

First, I would like to express my deepest gratitude to my supervisor Dr. Ardo van den Hout. I am thankful for his support and encouragement over the last four years. All the opportunities he provided and the weekly meetings we had are helping me and motivating me a lot. Also, I would like to thank the China Scholarship Council for funding this research. Special thanks to Dr. Boo Johansson and Dr. Graciela Muniz-Terrera for permission to use the OCTO-Twin data. Finally, I am grateful to my mom and friends for their supports in my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The multi-state model, regarded as an extension of a survival model, describes a process where individuals move among a series of states over time. It can be called a multi-state survival model if there is an absorbing state. Usually, the absorbing state is death. Multi-state models are broadly applied to analyse longitudinal data, where the change of state over time is of interest. The start time in this kind of multi-state survival process should be specific and clearly defined. The exact times when the status of individuals change are sometimes impossible or unnecessary to measure and record, especially for long-term studies. Consider an example of a dementia study, in which the process can be described by a four-state survival model. State 1 is healthy, state 2 is mild dementia, state 3 is severe dementia and state 4 is death. An example to demonstrate transitions between states is: A patient was firstly observed at state 1 at age 70. The second observation of this patient was at state 2 and age 72. After a set of observations, this patient was observed to die at age 80. Here we know the state of this patient changed from state 1 to state 2 during these two years, but we do not know the exact point in time of change.

Usually not all times in survival analysis are observed exactly, and censoring is common. We assume T_i as the exact time that the event of interest occurs for individual i, consider the situation where T_i is unknown. Say t_1 and t_2 are two

known time points and $t_1 < t_2$. Left-censoring indicates that the event of interest happens before a point in time t_1, where $T_i < t_1$. Right-censoring indicates the event of interest happens after a point in time t_2, where $T_i > t_2$. Interval-censoring indicates the event of interest happens between two exact time points ($t_1 < T_i < t_2$), see more details in Collett (2015).

The multi-state model can have back and forth transitions, resulting in a more complex censoring of time. An example of this situation for the above dementia study: for an individual who is observed at state 2 at time t_1, all other states (state 1, state 3 and state 4) can be considered as the next potential state. If this individual is next observed at state 1 at time $t_2 \neq t_1$, the exact time of transition t_e is interval-censored, where $t_1 < t_e < t_2$. If state 4 is observed at time $t_3 \neq t_1$, examples of possible trajectories are $2 \to 4$, $2 \to 3 \to 4$ or $2 \to 1 \to 4$. The censoring is severe if potential trajectories include $2 \to 1$, because there are potentially multiple times of transitions between state 1 and state 2.

The death time is usually observed. Therefore, in many applications, the transition times between living states are interval-censored, unlike the dead time which is right-censored or exact.

Another important concept for a multi-state model is the distinction of continuous time and discrete time. Discrete models describe the processes step by step, the time unit in this model is fixed on a uniform grid. For the continuous-time model, the transition process can take place at any time. It is more practical and realistic in a wide range of applications. This thesis will consider the continuous-time models with data in which times of transitions between living states are interval-censored but times of death (transitions to the absorbing state) are observed exactly.

1.1 Overview literature for the multi-state model

The multi-state survival model has widespread use in many areas of application. We would like to give some examples in biostatistics for both exact-time data

and interval-censored data. Details of modelling interval-censored data will be introduced in Chapter 3. Although we will mainly focus on the interval-censored data in our study, some previous studies for exact time data are good references for the basic idea of the multi-state model. For exact time data, Anderson and Keiding (2002) discuss several multi-state models to deal with event history analysis. They investigate the mortality and bleeding episodes in a liver cirrhosis trial. Both parametric and non-parametric methods are discussed. Putter et al. (2006) implement a multi-state model for analysing data of thousands of early breast cancer patients. They restricted the attention to the non-parametric hazards in the framework of the Cox model. Escolano et al. (2000) build a multi-state model and emphasize the prediction of the occurrence of hospital-acquired infections in ICU. Similar approaches are also used in many other kinds of papers, see Kay (1986), Marshall and Jones (1995) and Rickayzen and Duncan (2002). The non-parametric method is typically used for the exact time data, such as the Kaplan-Meier estimate (Kaplan and Meier (1958)). For more details of non-parametric methods, please see Collett (2015), Chapter 2. The parametric method is more common to use for interval-censored data, since the time of transition is unknown and the parametric models are better at dealing with the uncertainty with respect to the exact time.

Commenges (2002) give a comprehensive inference for multi-state models from interval-censored data. Transition probabilities and likelihood are discussed. Foucher et al. (2010) fit a four-state model to analyse the deterioration process of kidney transplant recipients, a goodness-of-fit for the model is also discussed. Buter et al. (2008) analyse dementia and survival in Parkinson disease by fitting a three-state model, and also show the difference in survival time between women and men. Machado et al. (2009) review several modelling methods for multi-state models and illustrate them by using two interval-censored datasets. The multi-state model is also useful in other areas, such as social statistics. For instance, Uhlendorff (2006) uses working in low paid jobs, higher paid jobs, and not working as the three states to apply a multi-state model. See Wu et al. (2009),

Nicolas (1995), Li (2012) for other applications.

Multi-state models are often based on the Markov assumption. For a multi-state process, define the variable Y_t as the state of the process at time t, the state space is the set S of all possible values of Y_t. S usually is a finite discrete space $\{1, 2, 3..., D\}$, where D is an integer. First, we assume that the continuous-time stochastic process $\{Y_t | t \in (0, \infty)\}$ is a Markov chain, for which we have

$$P(Y_{u+t} = s | Y_u = r, Y_v, 0 \leq v \leq u) = P(Y_{u+t} = s | Y_u = r), \tag{1.1.1}$$

for all states $r, s \in S$ and $u, t \geq 0$ in the multi-state process. This indicates that only the current state determines the future.

If the movements between states in the future are affected by the duration which the process has stayed in a state, the process is called a semi-Markov model. There are plenty of studies that discuss this model, such as Foucher et al. (2007) and Ramezankhani et al. (2020). Moreover, if the future state depends on two or more preceding states in the past, it is called the second or higher order Markov model. Note that the "higher order" is only used for the discrete time model, and not for the continuous time Markov model. Please see Cox and Miller (1965) and Kulkarni (2011) for more information on the Markov process.

The Markov chain is time homogeneous if

$$P(Y_{u+t} = s | Y_u = r) = P(Y_t = s | Y_0 = r), \tag{1.1.2}$$

for $t, u \geq 0$.

As will be shown, the time-homogeneous Markov chain can be defined by the constant transition hazard. Please see Section 3.4.1 for details. The assumption of the continuous-time Markov model was discussed in Kalbfleisch and Lawless (1985), in order to facilitate the computation process of the multi-state model when maximising the likelihood function. Gentleman et al. (1994) illustrate the usefulness of this Markov model for disease history data. Both the advantages

and disadvantages are well discussed. One of the advantages is that it can make more efficient use of incomplete information when only fairly short portions of individual's disease histories are available. However, the model is quite restrictive under the Markov assumption, since only the current state is related to the next state.

For the computation of multi-state Markov models for panel data, Jackson (2011) presents the `msm` package in R, introducing the usages and examples of fitting a wide range of multi-state models and the hidden Markov models. Extensions and limitations of `msm` are also well discussed. Hidden Markov models are used for multi-state processes which are subject to misclassification of states or where states are not observed. The hidden Markov model is not discussed in this thesis, please see Satten and Longini (1996) and Bureau et al. (2003) for details. There is also a tutorial about statistical methods for multi-state models provided by Putter et al. (2007). It details both concepts and practices from data preparation to estimation. It also shows examples with the R software. Please note that the event times of data in this tutorial are observed exactly or right-censored. Therefore, some concepts cannot be used for interval-censored data, such as the likelihood function of transition probabilities.

1.2 Overview literature for the frailty model

A multi-state model can be specified through transition hazards, which are instantaneous risks of transitioning from one state to another. The details of the hazard and hazard models will be introduced in Chapter 3. This study discusses two types of effects when describing hazards for moving across states: fixed effects and random effects. For the fixed-effect multi-state model, the variety of hazards for different individuals are explained by the observed explanatory variables, such as gender, ages or educational levels. However, there might be some unobserved heterogeneity that affects the hazard, even if all the explanatory variables are considered (Collett (2015)). For example, some factors do affect the sur-

vival, but were not measured in the data. Therefore, under the same value of covariates, some individuals are more likely to move to severe disease states or death than others, and some individuals are more likely to remain in a healthier state. In survival analysis, if such unobserved effects are modelled by a random effect, such a random effect in survival analysis is called frailty.

The main reason for adding a frailty term to the fixed-effect multi-state model is to describe the unobserved susceptibility for illness and death. Frailty can be defined either at the individual or group level. The former implies individual-specific transition hazards, the latter implies that groups of individuals share the group-specific frailty. Adding a group-specific frailty can investigate people from different groups without adding the dummy covariates. For example, people from different hospitals or countries can share a hospital- or country- specific frailty. If the number of hospitals or countries are large, the model will be overly complex by adding many dummy variables, which is not a good choice, especially if researchers do not aim to analyse the effects of individual hospitals or countries (Van Den Hout (2016)).

Except for the main advantage of adding a frailty term (to describe the unobserved heterogeneity), another advantage is that the frailty can model the association between transitions. As discussed in Section 1.1, it is assumed that the continuous-time multi-state model is a Markov process, which means only the current state determines the future. However, the Markov assumption could fail to hold because of the association between transition times. For example, individuals who have been longer in a specific state, are more likely to move to death. This kind of effect might be addressed by fitting a frailty model, where the frailty can be used to model such associations.

There is a large volume of published studies describing the role of frailty in survival analysis. The book *The Frailty Model* (Duchateau and Janssen (2007) introduced various types of frailty models and copula models to analyse clustered survival data. Many studies started with adding frailty as an extension for survival models, for example, see Anderson (1995) and Hougaard (1995) for details.

Some discussed multi-state frailty models with different distributions. For example, Bijwaard (2014) considered a Cox model in continuous time with frailties in univariate survival models and semi-Markov multi-state models. This study also gave an application for the labour market and migration dynamics of recent immigrants to the Netherlands. Putter and van Houwelingen (2015) discussed the way to identify frailties in multi-state models. They suggested considering whether we need frailties because they come with some disadvantages: firstly the effect of covariates might be more difficult to quantify since the computation is more complex due to the frailty parameters; secondly, frailty models are more difficult to fit; thirdly, sometimes frailties are not easy to understand and may lead to an unexpected result. In our study, we admit that the frailty model is not easy to fit because of the complexity of modelling and computation, but it is a good method for describing the unobserved heterogeneity and investigating the association between transitions. Therefore, we would like to fit both the multi-state fixed-effect model and the frailty model and compare their results. The question of whether we need frailty can be considered by comparing the parameter estimation, the assessment of goodness-of-fit and the AIC value of these two models.

Furthermore, a number of studies presented frailty models for multivariate survival data. For instance, Xue et al.(1996) fit a bivariate frailty model to analyse the bivariate survival data, to overcome the limitations of the univariate frailty model when analysing the multivariate survival data. Hougaard (2012) present four approaches to handle multivariate survival data. Many other papers discuss multivariate frailty models as well, such as Pickles et al.(1995) and O'Keeffe et al. (2018).

Lognormal and gamma distributions are widely used for frailties in the model. For example, the bivariate lognormal-distributed frailty is used in Lindeboom and Van Den Berg (1994). Joly et al. (2012) and Pak et al. (2017) both used the multivariate lognormal-distributed frailty in the multi-state model for interval-censored data. For the gamma frailty model, various types of multivariate gamma

distributions were defined in previous studies. For example, Mckay's bivariate gamma, Kibble and Moran's bivariate gamma and Sarmanov's bivariate gamma, and so on. These were introduced in Chapter 48 of Kotz et al. (2004). Some studies illustrate multivariate gamma frailty models, please see Wassell and Zelen (1993), Yashin (1995), Hens et al. (2009) and Martins et al. (2019). The gamma distribution is one of the most widely used frailty distributions in the standard survival model. The reason is that it is a very convenient distribution both from the analytical and computational view (Bijwaard (2014)). In the standard survival model, it has a closed-form expression for the hazard. But the situation of multi-state models is more complex, and we cannot get a closed-form of the likelihood for the gamma frailty. Details of the hazard and likelihood will be introduced in Chapters 3-5. An advantage for the gamma distribution for the multi-state model is that the mean of the gamma distribution refers to the mean of the frailty. However, when using a loglinear model, the mean of the normal distribution is zero do not imply the mean of the lognormal-distributed frailty is one, because there is a non-linear exponential transformation in the model. Details of modelling and this advantage will be discussed in Chapter 4.

In the multi-state process, the multivariate frailty model is a good approach to investigate the correlation of hazards between multiple transitions. In this study, we will mainly focus on bivariate frailty models. For the bivariate frailty model, the lognormal distribution and gamma distribution will both be discussed. Since there are many types of the bivariate gamma distribution, we aim to explore a bivariate gamma frailty model, which is relatively easy to fit and allows simple computation, in order to describe the unobserved heterogeneity and analyse the correlation between different transition hazards. Therefore, the Cheriyan and Ramabhadran's bivariate gamma distribution in Kotz (2004) is selected to fit frailty models. The correlated frailty models with this distribution was first proposed by Yashin et al.(1995) for univariate survival data on related event times. For example, the event times for twins or family members. The novelty of this thesis is that we use this distribution to define a bivariate gamma multi-state

frailty model for the interval-censored data. The correlation is defined in different transitions rather than paired-individuals.

1.3 Overview thesis

Chapters 2-6 provide an overview of fixed-effect multi-state models and frailty models. Chapter 2 introduces the Cardiac allograft vasculopathy (CAV) data and the Origins of Variance in the Oldest-Old (OCTO-Twin) data. Details of these datasets are shown in tables and graphs. Chapter 3 gives an overview and statistical modelling of multi-state fixed-effect models. In particular, we contribute the analytic expression of the transition probability for the four-state progressive model, the details are illustrated in Section 3.2. The comparison of this method and the piecewise constant approximation is investigated in Section 3.6. In Chapter 4 and 5, univariate multi-state frailty models and bivariate multi-state frailty models are discussed, respectively. The main contribution in these two chapters is proposing the bivariate frailty multi-state model for interval-censored data. Chapter 6 describes non-parametric frailty models. In this chapter, we also extend the general non-parametric frailty model by adding a covariate-specific frailty term. Chapter 7 analyses the data introduced in Chapter 2 and presents the results. Comparison of different models and the assessment of goodness of fit are discussed as well. Moreover, there might be some issues for identifiability and sensitivity of the proposed model regarding misspecifications These will be investigated in a simulation study, which is shown at the end of this chapter. In Chapter 8, we discuss both the contributions and limitations of this thesis.

Chapter 2

Data

2.1 Cardiac allograft vasculopathy data

To illustrate the approaches to fixed-effect multi-state models and frailty models mentioned in Chapter 1, the longitudinal data about cardiac allograft vasculopathy are used in this thesis. Cardiac allograft vasculopathy (CAV) commonly happens after heart transplantation. The CAV progress of patients is monitored by coronary angiography. Sharples et al. (2003) defined the process by three living states and one death state, the living states are distinguished by the grades of CAV. The states and transitions of this multi-state process are shown in Figure 2.1: State 1 to 3 are defined by no CAV, moderate CAV, severe CAV, respectively. State 4 is an absorbing state representing the dead. The process can transit between the living states. The CAV data were collected at Papworth Hospital in the United Kingdom and are also included in `msm` package in the R software (Jackson (2011)). Note that the transition times here for CAV data are interval-censored for three living states and are exact or right-censored for death. The time is given in unites of years, all the individuals are observed from time $t = 0$, which is time of transplantation.

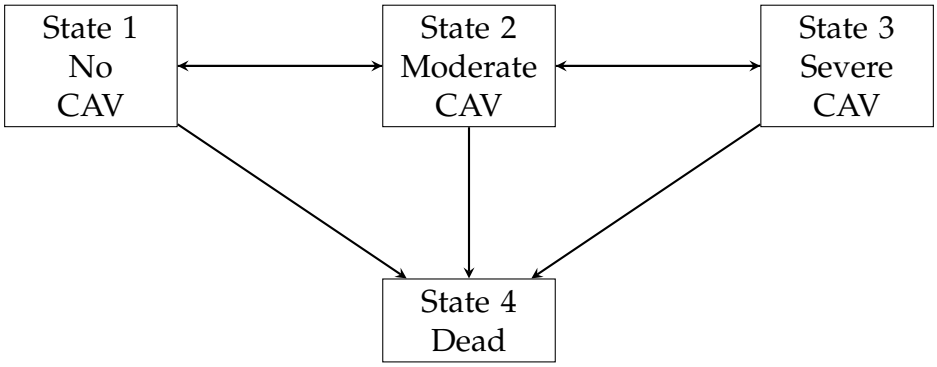

Figure 2.1: Transitions in the four-state model for cardiac allograft vasculopathy (CAV) data.

Table 2.1 depicts the data information: The data frame contains 2846 records which are derived from 622 individuals. There are 87 females and 535 males. Information on the reason for transplantation is also available. 313 individuals have the ischaemic heart disease(IHD) and 309 individuals have the idiopathic dilated cardiomyopathy(IDC).

The minimum and maximum ages (years) of observations in this study are 6.30 and 74.33. The frequencies of age at baseline and during the follow-up are more specifically shown on the top left and top right of Figure 2.2: The highest frequency of age interval of recipients for their first observations is 50yrs-55yrs. The majority of people have their first records during 40yrs-60yrs. Only a few people join this study when they are young. The histograms of age at baseline and age during follow-up show that there are data throughout the whole age range. The baseline age denotes the age of individuals at the first observation. For the time interval of observations, the minimum, median and maximum lengths (years) are 0.003, 1.32 and 16.48 years, respectively. The individual who has the time interval 0.003 years is a male, he only has two observations at time 0 in state 1 and 0.002739 (round to 0.003 in Table 2.1) in state 4. The time 0.003 years is approximately one day, which indicates that this individual died in a very short time after the transplant. The frequencies are shown on the middle left of Figure 2.2. Length as 2-3 years is the most common time interval between two observations. Although it is

Table 2.1: The summary of CAV data

Number of records	2846
Sample size	622 individuals
Gender	87 females and 535 males
Reason for transplant	313 for IHD and 309 for IDC
Description of age and age interval (years)	Minimum age =6.30 Maximum age = 74.33 Min length of intervals = 0.003 Median length of intervals = 1.32 Max length of intervals = 16.48
Frequencies number of states	1 2 3 4 2039 351 205 251
State table	*to* *from* 1 2 3 4 1 1367 204 44 148 2 46 134 54 48 3 4 13 107 55

a yearly examination for this study, some recipients skipped one or more of this examination. Commonly, the time intervals between two observations are during 0-5 years. Also, there are a few numbers of individuals whose time intervals are more than 10 years between two records. There might be some informative observation in the data. Informative observation means that the fact that an individual is observed potentially informative with respect to his or her health condition (Sisk et al. (2021)). If informative observation is included in the data, there might be some biased likelihood inference if it is not taken into account. For further discussion, please see Barrett (2011). The middle right of Figure 2.2 shows the years of follow-up for each individual until death or right-censoring. As we can see, the most frequent follow-up time is around 3-6 years. The longest follow-up time is around 18 years. The bottom left of Figure 2.2 shows the number of times of each individual. The most frequent number of times is two, there is only a few people are observed more than ten times in this study.

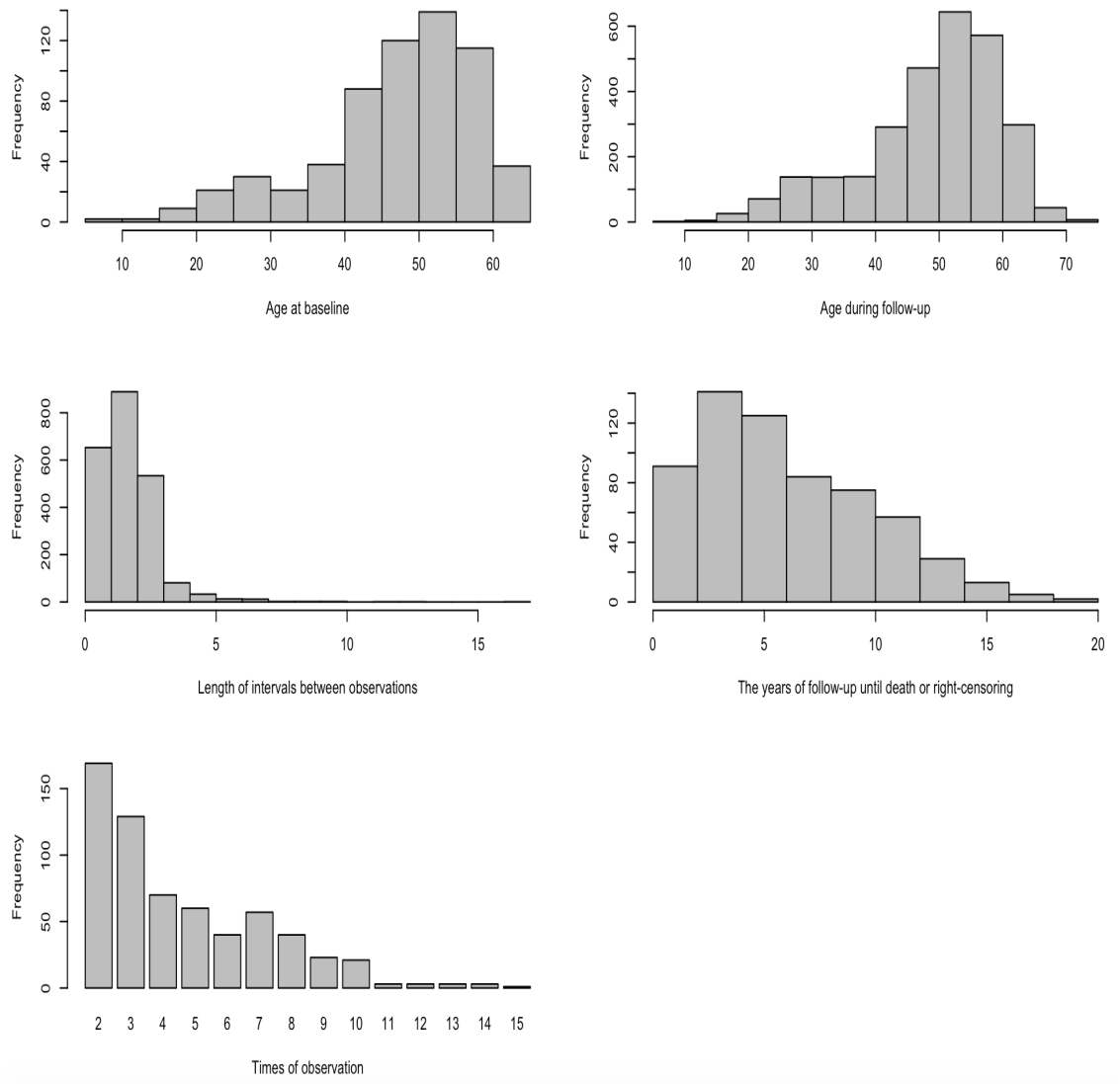

Figure 2.2: The histograms of age at baseline, age during follow-up, length of the interval between observations, follow-up years until death or right-censoring and times of observation for each individual for cardiac allograft vasculopathy (CAV) data on the top left, top right, middle left, middle right and bottom left, respectively. The units of ages and time intervals are years.

Table 2.1 lists how often the different states were observed to be occupied by individuals in the sample during the complete follow-up. The frequencies number of states 1, 2, 3, 4 are 2039, 351, 205 and 251, respectively. It illustrates that 251 individuals have died by the end of this study. The state table is shown in the last row of this table. It is obvious that the majority of frequencies are in

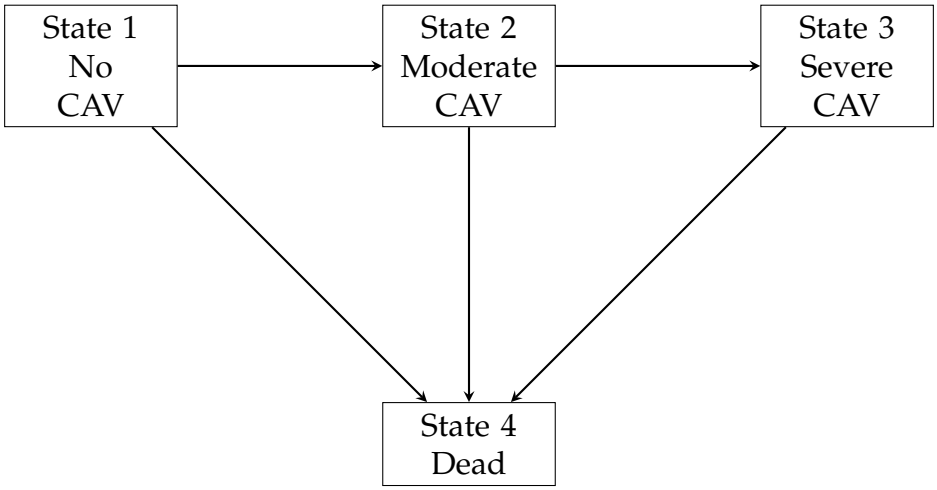

Figure 2.3: Progressive transitions in the four-state model for cardiac allograft vasculopathy (CAV) data.

the diagonal, which indicates that given the time between observations, most of people tend to stay in their current states between two observations. Also, there are only a few observations that go backwards, such as 4 observed in transition $(3, 1)$ and 13 observed in transition $(3, 2)$. This demonstrates the CAV is more likely a progressive process.

In the paper by Sharples et al. (2003), it is assumed that CAV is a deteriorating process with no recovery, which means there is no backward transitions in the model and all the backward transitions (2,1) (3,2) (3,1) are seen as misclassifications. Here for our study, we also assume that CAV is a progressive process with no backward transitions, and set individuals who have backward transitions (2,1) (3,2) and (3,1) to stay at states 2, 3 and 3, respectively. See Figure 2.3 for the progressive CAV process, the transitions of the process are (1,2), (1,4), (2,3), (2,4), (3,4). The frequencies number of states and state table are given by Table 2.2.

Table 2.2: The frequencies number of states and state table of CAV observed history data

Frequencies number of states	1	2	3	4
	1958	405	232	251

State table		*to*			
	from	1	2	3	4
	1	1336	185	40	139
	2	0	220	52	49
	3	0	0	140	63

2.2 Origins of Variance in the Oldest-Old data

The Origins of Variance in the Oldest-Old (OCTO-Twin) data includes the twin pairs, dizygotic or monozygotic, who are 79 years old or older. The sample was selected from older adults in the population-based Swedish Twin Registry (Robitaille et al. (2018)). Five cycles of longitudinal data were collected at 2-year intervals. The initial sample consisted of 702 participants (351 same-sex pairs), the final analysis included 694 participants. The permission for using this data was given by Dr. Boo Johansson.

In the study, participants were tested by nurses in their residence at two years time intervals for five cycles. The process of interest of OCTO-Twin data is cognitive function. Mini-mental state examination (MMSE) is a known clinical method for grading the cognitive state of patients, see Folstein et al. (1975). In this paper, three living states are defined by cutoff scores of MMSE: state 1 is defined as no cognitive impairment ($27 \leq$ MMSE ≤ 30), state 2 is a mild cognitive impairment ($23 \leq$ MMSE ≤ 26), state 3 is a severe cognitive impairment (MMSE≤ 22). This scoring method was discussed in Robitaille et al.(2018).

In this study, we assume that severe cognitive impairment (state 3) is an irreversible process. It means if individuals have been recorded in state 3, they are not allowed to move back to state 2 or state 1. The process is shown in 2.4. Additionally, we dismiss some individuals with missing values. The OCTO-Twin

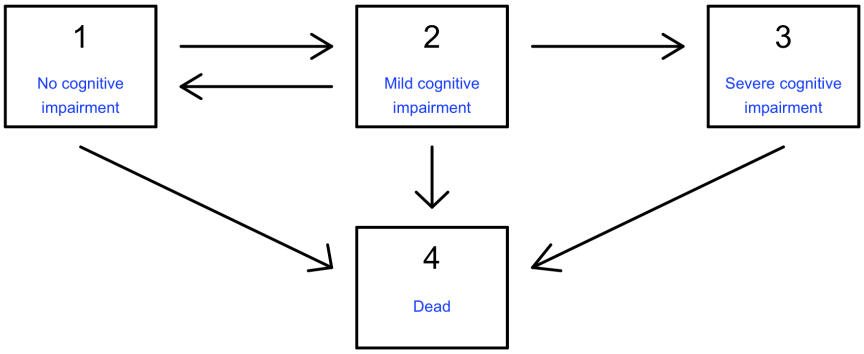

Figure 2.4: Transitions in the four-state model for the Origins of Variance in the Oldest-Old (OCTO-Twin) data.

data frame using in this study is shown in Table 2.3. There are totally 692 individuals with 461 females and 231 males. As we introduced at the beginning of this section, there are 694 individuals in the start of the analysis. We remove two individuals, since both of them only have two observations (state 1 and censoring). The minimum age (years) of participants is 79.37 and the maximum age is 104.205. 662 participants were recorded as dead at the end of the study. The state table shows the details of the transitions of individuals. Notice that the minimum length of intervals (years) is 0.003, which is the same as the one for CAV data. The individual who has the time interval 0.003 in OCTO-Twin data is a male, he was observed in state 3 at age 95.32329, and died at 95.32603. This time interval is 0.002739 year (round to 0.003 in the table), which is one day.

Figure 2.5 indicates that the baseline age of majority people is between 80 and 85. According to the length of time intervals, it demonstrates that participants are usually tested by two years. More specifically, 75.4% of time intervals between observations are between 1.5 and 2.5 years, which is consistent with the study design. The histogram of the years of follow up shows that most individuals are

Table 2.3: The Summary of OCTO-Twin data. The state table is a frequency table counting the number of times each pair of states were observed in successive observation times

Number of records	2876
Sample size	692 individuals
Gender	461 females and 231 males
Description of age and age interval (years)	Minimum age = 79.37 Maximum age = 104.21 Min length of intervals = 0.003 Median length of intervals = 1.99 Max length of intervals = 13.86
Frequencies number of states	1 2 3 4 1170 374 670 662
State table	*to* *from* 1 2 3 4 1 710 130 79 229 2 66 98 100 104 3 0 0 339 329

observed in 2-10 years, some participants are recorded over 15 years. The times of observation shows that all the people are seen two to six times during this study.

The process time of this study is age. People enter the study at different ages, which is shown in the first picture of Figure 2.5. There are 18 people entering this study between ages 79-80 (first column), and 3 people entering this study between ages 97-98 (last column), nobody entered the study between ages 94-96. Therefore, there is left truncation in OCTO-Twin data. As we discussed in Section 1.1, the multi-state model in this study is based on the Markov assumption, so we would not consider the left truncation for the fixed-effect multi-state model. However, for the frailty model, the frailty distributions may vary for people entering the study at different ages. For example, an individual who enters the study at age 85 in a healthy state may have a lower frailty than people who enter at age 80 in an illness state. The left truncation of frailty will be discussed in the simulation study and Conclusion in Section 7.4 and Chapter 8.

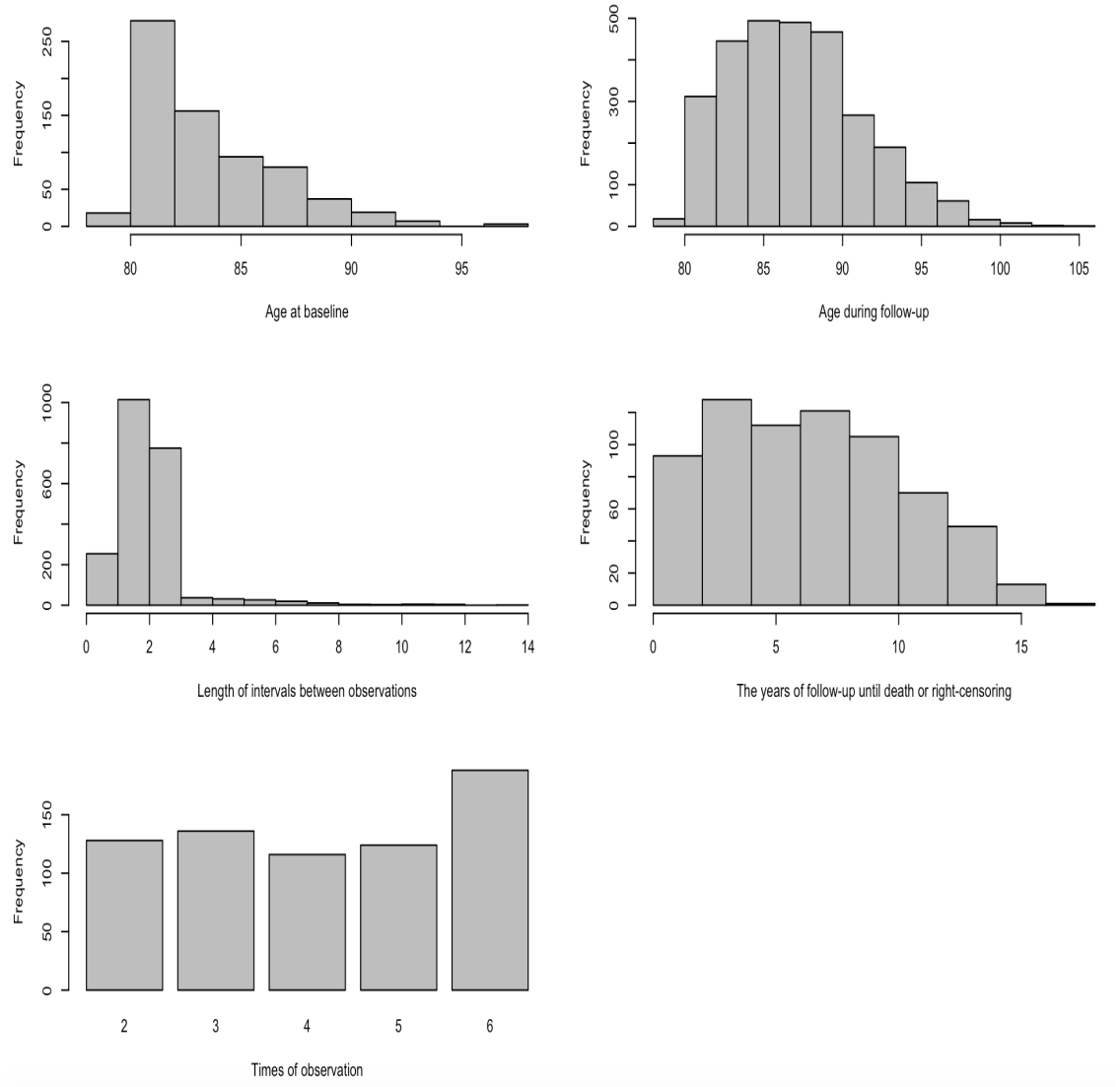

Figure 2.5: The histograms of age at baseline, age during follow-up, length of the interval between observations, follow-up years until death or right-censoring for OCTO-Twin data and times of observation of each individual respectively. The units of ages and time intervals are years.

Chapter 3

Fixed-effect multi-state model

3.1 Introduction of survival analysis

The fixed-effect multi-state model can be seen as an extension of the survival model. Therefore, we would like to introduce some basic terminology and information through a survival model. As it is shown in Figure 3.1, the survival model has only two states. It describes the process that people transit from state 1 (alive) to state 2 (death). In medical research, the time origin often corresponds to the recruitment of individuals into the study or the beginning of the treatment. The transition time from state 1 to state 2 is called the time to event, other common names are event time, survival time, failure time, and so on. Materials of the standard survival analysis mentioned in this section are partly based on Collett (2015).

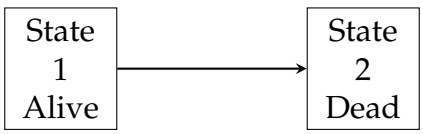

Figure 3.1: The process of univariate survival model.

Let T be the time to event, the cumulative distribution function of T is given

by

$$F(t) = P(T \leq t) = \int_0^t f(u)du, \tag{3.1.1}$$

where T is a continuous variable with time space $[0, \infty)$, f is the probability density function.

The survival function is given by

$$S(t) = P(T > t) = 1 - F(t). \tag{3.1.2}$$

The hazard function is defined as the infinitesimal risk of death occurring at some time t, conditional on an individual's survival time T beyond t.

$$\begin{aligned} h(t) &= \lim_{\triangle \to 0} \frac{P(t \leq T \leq t + \triangle | T > t)}{\triangle} \\ &= \lim_{\triangle \to 0} \frac{P(t \leq T \leq t + \triangle)}{P(T > t)\triangle} \\ &= \lim_{\triangle \to 0} \frac{F(t + \triangle) - F(t)}{S(t)\triangle} \\ &= \lim_{\triangle \to 0} \{\frac{F(t + \triangle) - F(t)}{\triangle}\} \frac{1}{S(t)} \\ &= \frac{f(t)}{S(t)}, \end{aligned} \tag{3.1.3}$$

where $h(t) \geq 0$ for all t.

According to Equations (3.1.1), (3.1.2) and (3.1.3), the hazard can be deduced as

$$h(t) = -\frac{d}{dt} \log(S(t)). \tag{3.1.4}$$

Therefore, the survival function can be derived from the hazard

$$S(t) = \exp(-H(t)), \tag{3.1.5}$$

where $H(t) = \int_0^t h(u)du$ is the cumulative function of $h(t)$.

The probability of time T in time interval $(t_1, t_2]$ is given by

$$P(t_1 \leq T \leq t_2) = F(t_2) - F(t_1) = \int_{t_1}^{t_2} f(u)du = \int_{t_1}^{t_2} h(u)S(u)du. \tag{3.1.6}$$

More details about the survival function, hazard function and the cumulative hazard function can be found in Section 1.3 of Collett (2015).

The hazard as defined above for standard survival analysis can be used for multi-state models as well. Several transitions are possible in a multi-state model and each of them should have its own transition hazard. We present an example of the four-state progressive CAV model introduced in Chapter 2. Let t denotes the time in this continuous-time model, the random variable $Y_t \in \{1, 2, 3\}$ denotes the state occupied at time t. For each transition (r, s), the hazard at time t is defined as $h_{rs}(t)$. As a consequence, the cumulative hazard functions for leaving state 1 (no CAV), state 2 (moderate CAV), and state 3 (severe CAV) during the time interval $(t_1, t_2]$ are followed, respectively.

$$H_1(t_1, t_2) = \int_{t_1}^{t_2} h_{12}(u) + h_{14}(u)du, \tag{3.1.7}$$

$$H_2(t_1, t_2) = \int_{t_1}^{t_2} h_{23}(u) + h_{24}(u)du, \tag{3.1.8}$$

and

$$H_3(t_1, t_2) = \int_{t_1}^{t_2} h_{34}(u)du. \tag{3.1.9}$$

Van Den Hout (2016) discussed the transition probabilities for a three-state model. In our study, we will discuss the transition probabilities for a progressive

four-state model according to his approach. To the best of our knowledge, this has not been discussed in the literature before. Based on the Equations 3.1.5 and 3.1.6, the transition probabilities $P_{rs}(t_1, t_2) = P(Y_{t_2} = s | Y_{t_1} = r)$ are given by

$$
\begin{aligned}
p_{11}(t_1, t_2) &= \exp(-H_1(t_1, t_2)) \\
p_{12}(t_1, t_2) &= \int_{t_1}^{t_2} \exp(-H_1(t_1, u)) h_{12}(u) \exp(-H_2(u, t_2)) du \\
p_{13}(t_1, t_2) &= \int_{t_1}^{t_2} \int_{u_1}^{t_2} \exp(-H_1(t_1, u_1)) h_{12}(u_1) \\
&\qquad \times \exp(-H_2(u_1, u_2)) h_{23}(u_2) \exp(-H_3(u_2, t_2)) du_2 du_1 \\
p_{14}(t_1, t_2) &= 1 - p_{11}(t_1, t_2) - p_{12}(t_1, t_2) - p_{13}(t_1, t_2) \\
p_{21}(t_1, t_2) &= 0 \\
p_{22}(t_1, t_2) &= \exp(-H_2(t_1, t_2)) \\
p_{23}(t_1, t_2) &= \int_{t_1}^{t_2} \exp(-H_2(t_1, u)) h_{23}(u) \exp(-H_3(u, t_2)) du \\
p_{24}(t_1, t_2) &= 1 - p_{22}(t_1, t_2) - p_{23}(t_1, t_2) \\
p_{31}(t_1, t_2) &= 0 \\
p_{32}(t_1, t_2) &= 0 \\
p_{33}(t_1, t_2) &= \exp(-H_3(t_1, t_2)) \\
p_{34}(t_1, t_2) &= 1 - p_{33}(t_1, t_2).
\end{aligned}
\tag{3.1.10}
$$

The transition probabilities are used to calculate the likelihood function. This will be introduced in Section 3.4.

3.2 Parametric distribution of time

The models in this chapter are parametric. We have briefly mentioned non-parametric methods in Section 1.1. Please see Akritas (2004) and Collett (2015) for details. In many studies for survival data, supplementary information will also be investigated for individuals, such as age, gender and dietary habits. These may all affect the time an individual survives. Therefore, we would like to fit

parametric models instead of non-parametric ones, since the former can take into account the effect of covariates.

We will introduce three different distributions in this section: Exponential, Weibull and Gompertz distribution. For each of them, we contribute an analytic expression of transition probabilities during the time interval (t_1, t_2) for a four-state progressive model. The transition probabilities are derived by using the method discussed in Section 3.1. A major advantage of this method is that it can simplify the computation of transition probabilities. The result of this method can be directly used in the progressive four-state model. However, there are certain drawbacks if there are backward transitions or a large number of states in multi-state models. For these kinds of models, it is hard to get the analytic expression of transition probabilities, so a method called piecewise-constant approximation is usually used, which will be discussed in Section 3.4. Although this analytic expression will not be used for the application of CAV and OCTO-Twin data, it is still good to show the expression. On the one hand, it is a direct solution for the progressive four-state models, which typically leads to a fast computation time, on the other hand, such a closed-form expression is helpful for readers to understand the transition probability for each transition in the multi-state model. Moreover, we compare the results and speed of computations of these two methods, see Section 3.6 for the example.

3.2.1 Exponential distribution

A continuous random variable T follows an exponential distribution (i.e. $T \sim Exp(\lambda)$), if the probability density function is

$$f(t) = \begin{cases} \lambda \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{3.2.1}$$

where the parameter $\lambda > 0$.

The cumulative distribution function is given by

$$F(t) = P(T \leq t) = 1 - \exp(-\lambda t) \quad for \quad t \geq 0, \tag{3.2.2}$$

the hazard function is

$$h(t) = \lambda \quad for \quad \lambda > 0, \tag{3.2.3}$$

the survival function is

$$S(t) = \exp(-\lambda t) \quad for \quad t \geq 0. \tag{3.2.4}$$

The expectation is $E(T) = \frac{1}{\lambda}$, the variance is $Var(T) = \frac{1}{\lambda^2}$.

Equation 3.2.3 shows that the transition hazards are constants in the exponential model. Let the transition hazards $h_{rs}(t)$ for each transition (r, s) are λ_{rs} at time t, the transition probabilities for four-state progressive model are given by

$$\begin{aligned}
p_{11}(t_1, t_2) &= \exp(-H_1(t_1, t_2)) = \exp(-(\lambda_{12} + \lambda_{14})(t_2 - t_1)) \\
p_{22}(t_1, t_2) &= \exp(-H_2(t_1, t_2)) = \exp(-(\lambda_{23} + \lambda_{24})(t_2 - t_1)) \\
p_{33}(t_1, t_2) &= \exp(-H_3(t_1, t_2)) = \exp(-(\lambda_{34})(t_2 - t_1)) \\
p_{12}(t_1, t_2) &= \int_{t_1}^{t_2} \exp(-H_1(t_1, u)) h_{12}(u) \exp(-H_2(u, t_2)) du \\
&= \int_{t_1}^{t_2} \exp(-H_1(t_1, u)) \lambda_{12} \exp(-H_2(u, t_2)) du \\
&= \exp((\lambda_{12} + \lambda_{14})t_1 - (\lambda_{23} + \lambda_{24})t_2) \\
&\quad \times \int_{t_1}^{t_2} \lambda_{12} \exp(-(\lambda_{12} + \lambda_{14} - \lambda_{23} - \lambda_{24})u) du \\
&= \frac{\lambda_{12}}{\lambda_{23} + \lambda_{24} - \lambda_{12} - \lambda_{14}} \\
&\quad \times (\exp((\lambda_{12} + \lambda_{14})(t_1 - t_2)) - \exp((\lambda_{23} + \lambda_{24})(t_1 - t_2))) \\
p_{13}(t_1, t_2) &= \int_{t_1}^{t_2} \int_{u_1}^{t_2} \exp(-H_1(t_1, u_1)) h_{12}(u_1) \\
&\quad \times \exp(-H_2(u_1, u_2)) h_{23}(u_2) \exp(-H_3(u_2, t_2)) du_2 du_1
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\lambda_{12}\lambda_{23}}{(\lambda_{34}-\lambda_{23}-\lambda_{24})(\lambda_{23}+\lambda_{24}-\lambda_{12}-\lambda_{14})} \\
&\quad \times (\exp((\lambda_{12}+\lambda_{14})(t_1-t_2)) - \exp((\lambda_{23}+\lambda_{24})(t_1-t_2))) \\
&\quad - \frac{\lambda_{12}\lambda_{23}}{(\lambda_{34}-\lambda_{23}-\lambda_{24})(\lambda_{34}-\lambda_{12}-\lambda_{14})} \\
&\quad \times (\exp((\lambda_{12}+\lambda_{14})(t_1-t_2)) - \exp(\lambda_{34}(t_1-t_2))) \\
p_{23}(t_1,t_2) &= \int_{t_1}^{t_2} \exp(-H_2(t_1,u))h_{23}(u)\exp(-H_3(u,t_2))du \\
&= \frac{\lambda_{23}}{\lambda_{34}-\lambda_{23}-\lambda_{24}}(\exp((\lambda_{23}+\lambda_{24})(t_1-t_2)) - \exp(\lambda_{34}(t_1-t_2)))
\end{aligned}
$$

$p_{21}(t_1,t_2) = p_{31}(t_1,t_2) = p_{32}(t_1,t_2) = 0$

$p_{14}(t_1,t_2), p_{24}(t_1,t_2), p_{34}(t_1,t_2)$ can be derived from other probabilities by Equation (3.1.10)

3.2.2 Weibull distribution

A continuous random variable T follows a Weibull distribution (i.e. $T \sim Weibull(\lambda,\tau)$), if the probability density function is defined as

$$
f(t) = \begin{cases} \lambda\tau t^{\tau-1}\exp(-\lambda t^{\tau}) & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{3.2.5}
$$

where the scale parameter $\lambda > 0$ and shape parameter $\tau > 0$.

The cumulative distribution function is given by

$$
F(t) = P(T \leq t) = 1 - \exp(-\lambda t^{\tau}) \quad for \quad t \geq 0, \tag{3.2.6}
$$

the hazard function is

$$
h(t) = \lambda\tau t^{\tau-1} \quad for \quad t \geq 0, \tag{3.2.7}
$$

the survival function is

$$S(t) = \exp(-\lambda t^{\tau}) \quad for \quad t \geq 0. \tag{3.2.8}$$

The expected value and variance are

$$E(T) = \lambda^{\frac{-1}{\tau}} \Gamma(1 + \tau^{-1}) \tag{3.2.9}$$

and

$$Var(T) = \lambda^{\frac{-2}{\tau}} (\Gamma(1 + \frac{2}{\tau}) - \Gamma(1 + \frac{1}{\tau})^2), \tag{3.2.10}$$

respectively, where Γ is the gamma function.

The transition hazards are time-dependent in the Weibull model. Let the transition hazards $h_{rs}(t)$ for each transition (r, s) are $h_{rs}(t) = \lambda_{rs}\tau_{rs}t^{\tau_{rs}-1}$ at time t, the transition probabilities for the four-state progressive model are given by

$$p_{11}(t_1, t_2) = \exp(-H_1(t_1, t_2)) = \exp(-\lambda_{12}(t_2^{\tau_{12}} - t_1^{\tau_{12}}) - \lambda_{14}(t_2^{\tau_{14}} - t_1^{\tau_{14}}))$$
$$p_{22}(t_1, t_2) = \exp(-H_2(t_1, t_2)) = \exp(-\lambda_{23}(t_2^{\tau_{23}} - t_1^{\tau_{23}}) - \lambda_{24}(t_2^{\tau_{24}} - t_1^{\tau_{24}}))$$
$$p_{33}(t_1, t_2) = \exp(-H_3(t_1, t_2)) = \exp(-\lambda_{34}(t_2^{\tau_{34}} - t_1^{\tau_{23}}))$$

For the transition probabilities in transition $(1,2)$, $(1,3)$, $(2,3)$, they follow that

$$\begin{aligned}
p_{12}(t_1,t_2) &= \int_{t_1}^{t_2} \exp(-H_1(t_1,u))h_{12}(u)\exp(-H_2(u,t_2))du \\
&= \int_{t_1}^{t_2} \exp(-H_1(t_1,u))\lambda_{12}\tau_{12}u^{\tau_{12}-1}\exp(-H_2(u,t_2))du \\
p_{13}(t_1,t_2) &= \int_{t_1}^{t_2}\int_{u_1}^{t_2} \exp(-H_1(t_1,u_1))h_{12}(u_1) \\
&\quad \times \exp(-H_2(u_1,u_2))h_{23}(u_2)\exp(-H_3(u_2,t_2))du_2du_1 \\
&= \int_{t_1}^{t_2}\int_{u_1}^{t_2} \exp(-H_1(t_1,u_1))\lambda_{12}\tau_{12}u_1^{\tau_{12}-1} \\
&\quad \times \exp(-H_2(u_1,u_2))\lambda_{23}\tau_{23}u_2^{\tau_{23}-1}\exp(-H_3(u_2,t_2))du_2du_1 \\
p_{23}(t_1,t_2) &= \int_{t_1}^{t_2} \exp(-H_2(t_1,u))h_{12}(u)\exp(-H_3(u,t_2))du \\
&= \int_{t_1}^{t_2} \exp(-H_2(t_1,u))\lambda_{23}\tau_{23}u^{\tau_{23}-1}\exp(-H_3(u,t_2))du
\end{aligned}$$

For other transition probabilities,

$p_{21}(t_1,t_2) = p_{31}(t_1,t_2) = p_{32}(t_1,t_2) = 0$

$p_{14}(t_1,t_2), p_{24}(t_1,t_2), p_{34}(t_1,t_2)$ can be derived from other probabilities by Equation (3.1.10)

Generally, there are no numerical-specific expressions for the integrations of $p_{12}(t_1,t_2)$, $p_{13}(t_1,t_2)$ and $p_{23}(t_1,t_2)$. Omar et al. (1995) discussed closed-form expressions for models with specific restrictions.

3.2.3 Gompertz distribution

A continuous random variable T follows a Gompertz distribution (i.e. $T \sim (\lambda,\xi)$), if the probability density function is

$$f(t) = \begin{cases} \lambda\exp(-\xi t)\exp(-\lambda\xi^{-1}(\exp(\xi t)-1)) & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{3.2.11}$$

where $\lambda > 0, \xi > 0$. The distribution function is therefore given by

$$F(t) = P(T \leq t) = 1 - \exp(-\lambda\xi^{-1}(\exp(\xi t) - 1)), \tag{3.2.12}$$

the hazard function is given by

$$h(t) = \lambda \exp(\xi t) \quad for \quad t \geq 0, \tag{3.2.13}$$

the survival function is given by

$$S(t) = \exp(-\lambda\xi^{-1}(\exp(\xi t) - 1)) \quad for \quad t \geq 0. \tag{3.2.14}$$

There is no closed form expression of expectation and variance for Gompertz distribution, for more details about Gompertz distribution, please also see Porllard et al. (1992).

The transition hazards are time-dependent in the Weibull model. Let the transition hazards $h_{rs}(t)$ for each transition (r, s) are $h_{rs}(t) = \lambda_{rs} \exp(\xi_{rs} t)$ at time t, the transition probabilities for four-state progressive model are given by

$$\begin{aligned}
p_{11}(t_1, t_2) &= \exp(-H_1(t_1, t_2)) = \exp(-\int_{t_1}^{t_2} \lambda_{12} \exp(\xi_{12} u) + \lambda_{13} \exp(\xi_{13} u)) \\
&= \exp(-(\frac{\lambda_{12}}{\xi_{12}}(\exp(\xi_{12} t_2) - \exp(\xi_{12} t_1)) \\
&\quad + \frac{\lambda_{13}}{\xi_{13}}(\exp(\xi_{13} t_2) - \exp(\xi_{13} t_1)))) \\
p_{22}(t_1, t_2) &= \exp(-H_2(t_1, t_2)) = \exp(-\int_{t_1}^{t_2} \lambda_{23} \exp(\xi_{23} u) + \lambda_{24} \exp(\xi_{24} u)) \\
&= \exp(-(\frac{\lambda_{23}}{\xi_{23}}(\exp(\xi_{23} t_2) - \exp(\xi_{23} t_1)) \\
&\quad + \frac{\lambda_{24}}{\xi_{24}}(\exp(\xi_{24} t_2) - \exp(\xi_{24} t_1)))) \\
p_{33}(t_1, t_2) &= \exp(-H_3(t_1, t_2)) = \exp(-\int_{t_1}^{t_2} \lambda_{34} \exp(\xi_{34} u)) \\
&= \exp(-\frac{\lambda_{34}}{\xi_{34}}(\exp(\xi_{34} t_2) - \exp(\xi_{34} t_1)))
\end{aligned}$$

For the transition probabilities in transition $(1,2)$, $(1,3)$, $(2,3)$, they follow that

$$\begin{aligned}
p_{12}(t_1,t_2) &= \int_{t_1}^{t_2} \exp(-H_1(t_1,u))h_{12}(u)\exp(-H_2(u,t_2))du \\
&= \int_{t_1}^{t_2} \exp(-H_1(t_1,u))\lambda_{12}\exp(\xi_{12}u)\exp(-H_2(u,t_2))du \\
p_{13}(t_1,t_2) &= \int_{t_1}^{t_2}\int_{u_1}^{t_2} \exp(-H_1(t_1,u_1))h_{12}(u_1) \\
&\quad \times \exp(-H_2(u_1,u_2))h_{23}(u_2)\exp(-H_3(u_2,t_2))du_2du_1 \\
&= \int_{t_1}^{t_2}\int_{u_1}^{t_2} \exp(-H_1(t_1,u_1)))\lambda_{12}\exp(\xi_{12}u_1) \\
&\quad \times \exp(-H_2(u_1,u_2)))\lambda_{23}\exp(\xi_{23}u_2)\exp(-H_3(u_2,t_2))du_2du_1 \\
p_{23}(t_1,t_2) &= \int_{t_1}^{t_2} \exp(-H_2(t_1,u))h_{12}(u)\exp(-H_3(u,t_2))du \\
&= \int_{t_1}^{t_2} \exp(-H_2(t_1,u))\lambda_{23}\exp(\xi_{23}u)\exp(-H_3(u,t_2))du
\end{aligned}$$

For other transition probabilities,

$p_{21}(t_1,t_2) = p_{31}(t_1,t_2) = p_{32}(t_1,t_2) = 0$

$p_{14}(t_1,t_2), p_{24}(t_1,t_2), p_{34}(t_1,t_2)$ can be derived from other probabilities by Equation (3.1.10)

3.3 Hazard function

In general, additional to the time of event, other information on the individuals is available. This information can be included as covariates in the model. Therefore, we can define a regression model to combine the effect of time and covariates. The general expression for transition (r,s) is given by

$$h_{rs}(t|\boldsymbol{x}) = h_{rs.0}(t)\exp(\boldsymbol{\beta}_{rs}^T\boldsymbol{x}), \tag{3.3.1}$$

where $\boldsymbol{\beta}_{rs} = (\beta_{rs.1}, \beta_{rs.2,\ldots,\beta_{rs.n}})^\top$ is a parameter vector, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top$ is the vector of covariates, $h_{rs.0}(t)$ is called baseline hazard, which illustrates how the

hazard changes with time t. See Equation (3.2.3), (3.2.7) and (3.2.13) for the expressions of hazard $h_{rs.0}(t)$ when T follows an exponential distribution, Weibull distribution and Gompertz distribution.

The model in Equation (3.3.1) was introduced by Cox (1972) and is called the proportional hazards model. This model is widely used in modelling survival data. The assumption of proportionality indicates that the hazard of the transition $r \rightarrow s$ at any given time for an individual is proportional to the hazard at that time for another individual (Collett (2015)). Alternative models for time-to-event data that do not require this assumption are the accelerated failure time model, the proportional odds model, the Cox regression model that includes a time-dependent covariate and non-proportional hazards. These models are not discussed in this thesis, please see Chapters 6, 8 and 11 in Collett (2015) for details.

3.4 Likelihood function

3.4.1 Markov assumption

Mostly, multi-state models are based on the Markov assumption. This idea was proposed to facilitate the computation process of the multi-state model when maximising the likelihood function, as in Kalbfleisch and Lawless (1985).

The likelihood function for the Markov multi-state model for interval-censored data is calculated by using the transition probability matrix $P(t)$. We have introduced the transition probabilities for the time interval (t_1, t_2) in section 3.2. For the continuous-time process, we describe a transition probability matrix $P_{rs}(u, t+u)$, which represents the probability of being in state s at time $t+u$ conditionally on being in state r at time u. The Kolmogorov differential equations is a method to derive P matrix by using a generator matrix Q, see Cox and Miller (1965) for details.

A generator matrix Q collects the transition hazards in a matrix, where the

off-diagonal elements are $q_{rs} = h_{rs}(t)$ and the diagonal elements are defined as $q_{rr}(t) = -\Sigma_{r \neq s} q_{rs}(t)$. For example, the Q matrix for the OCTO-Twin data is

$$Q(t) = \begin{pmatrix} -(h_{12}(t) + h_{14}(t)) & h_{12}(t) & 0 & h_{14}(t) \\ h_{21}(t) & -(h_{21}(t) + h_{23}(t) + h_{24}(t)) & h_{23}(t) & h_{24}(t) \\ 0 & 0 & -h_{34}(t) & h_{34}(t) \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{3.4.1}$$

The Q matrix for the CAV study is

$$Q = \begin{pmatrix} -(h_{12}(t) + h_{14}(t)) & h_{12}(t) & 0 & h_{14}(t) \\ 0 & -(h_{23}(t) + h_{24}(t)) & h_{23}(t) & h_{24}(t) \\ 0 & 0 & -h_{34}(t) & h_{34}(t) \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{3.4.2}$$

If the generator matrix Q is constant over the time interval $(u, t+u)$, as in a time-homogeneous process (discussed in Section 1.1), then the transition probability of this interval can be derived by the matrix exponential of Q scaled by the time interval, see Jackson (2011).

$$P(u, t+u) = P(t) = \sum_{k=1}^{\inf} \frac{(tQ)^k}{k!} = \exp(tQ), \tag{3.4.3}$$

where the computation of the matrix exponential function $\exp()$ is introduced by Moler and Van Loan (2003). This function is also used in the package `msm` in R, please see Jackson (2011) for details. Alternatively, it can be calculated by using the eigenvalue decomposition of Q. The matrix Q can be deduced by

$$Q = ABA^{-1}, \tag{3.4.4}$$

where A is a $D \times D$ matrix consist of D eigenvectors in columns of Q, B is a

diagonal matrix with D eigenvalues of matrix Q. It follows that

$$P(t) = A\text{diag}(\exp(b_1 t), \exp(b_2 t), \ldots, \exp(b_D t))A^{-1}, \tag{3.4.5}$$

where $b_1, b_2, \ldots, b_D$ are the eigenvalues of Q.

3.4.2 Piecewise-constant approximation

In the time-homogeneous case discussed in the previous section, hazards are assumed to be constant over time. However, for the time-inhomogeneous process, the transition probability matrix cannot be calculated, because the generator matrix Q is changing over the time interval $(u, t+u)$. This problem can be dealt with by using a piecewise-constant approximation. This method also can be found in Van Den Hout (2016). He discussed this method for the survival model, the progressive three-state model and the general multi-state model. For this method, the hazard is constant during a time interval but can change between the intervals. Therefore, a constant Q can be used during a time interval. Here for the `msm` package in R software, it defaults that $Q = Q(u_1)$ during a time interval (u_1, u_2), thus the transition probability matrix $P(u_1, u_2) \approx \exp((u_1 - u_2)Q(u_1))$, see Jackson (2011) for details. There are several other ways to do the piecewise-constant approximation. For instance, during the time interval (u_1, u_2), we can define the time of transition hazards in midway intervals.

$$Q = Q(\frac{u_1 + u_2}{2}),$$

therefore, the transition probability matrix

$$P(u_1, u_2) \approx \exp((u_1 - u_2)Q(\frac{u_1 + u_2}{2})).$$

Figure 3.2 illustrates these two options of the piecewise-constant approxima-

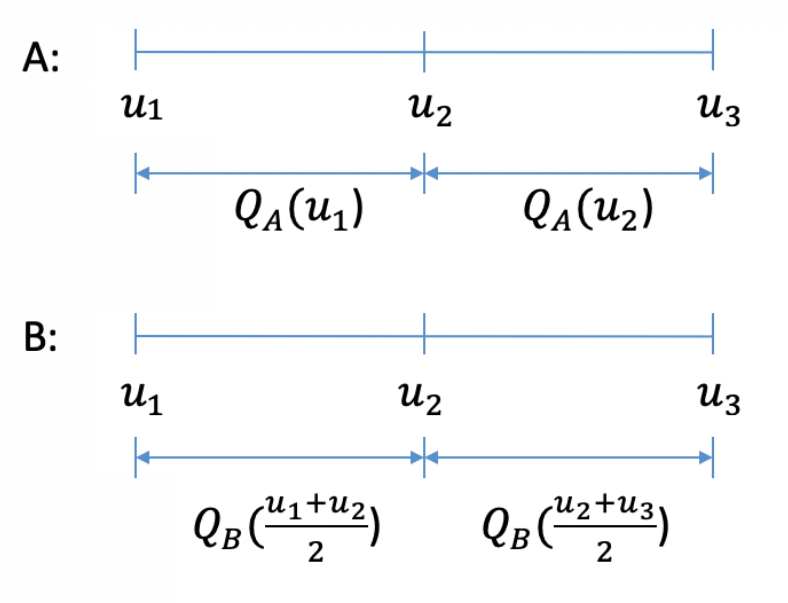

Figure 3.2: Two options of the piecewise-constant approximation

tion. If Q matrix is defined as

$$Q = Q(u_2),$$

then the transition probability matrix

$$P(u_1, u_2) \approx \exp((u_1 - u_2)Q(u_2)).$$

In practice, if we want to model a slow changing process like dementia, it might be sufficient to observe individuals every six months, perhaps even at longer intervals. Then different types of piecewise-constant approximation do not have a large effect on the result. We follow the choice made in `msm`. On the other hand, if we want to model a rapidly evolving disease like common cold, the piecewise-constant approximation can help with it. The sub-interval should be relative to how quickly the process changes, because we might need more time intervals if it changes quickly.

3.4.3 Maximising the likelihood function

The likelihood contribution for an individual i with the observed time $t_1, t_2, \ldots, t_J$, conditional on the first observed state is given by

$$L_i(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}) = P(Y_2 = y_2, \ldots, Y_J = y_J | Y_1 = y_1, \boldsymbol{\theta}, \boldsymbol{x}), \tag{3.4.6}$$

where $\boldsymbol{\theta}$ represents the vector of all the model parameters (see Section 3.3 for the parameters), $\boldsymbol{y} = (y_1, y_2, \ldots, y_J)$ represents the state at times $t_1, t_2, \ldots, t_J$, and $\boldsymbol{x} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_J})$ represents the vector of the covariate values. Each element of the $\boldsymbol{x}$ is a vector represents the covariate value at observed time $t_1, t_2, \ldots, t_J$.

Under the Markov assumption, the likelihood contribution is

$$\begin{aligned} L_i(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}) &= \prod_{j=2}^{J} P(Y_j = y_j | Y_{j-1} = y_{j-1}, \boldsymbol{\theta}, \boldsymbol{x}) \\ &= \prod_{j=2}^{J-1} P(Y_j = y_j | Y_{j-1} = y_{j-1}, \boldsymbol{\theta}, \boldsymbol{x})\, S(y_J | y_{J-1}, \boldsymbol{\theta}, \boldsymbol{x}), \end{aligned}$$

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_J)$, $S(y_J|y_{J-1}, \boldsymbol{x})$ represents the contribution of time interval (t_{J-1}, t_J) with observations y_{J-1} and y_J, respectively.

There is an absorbing state for some multi-state models, such as the state 4 (dead) in CAV data and OCTO-Twin data. The time of transition to the absorbing state Y_J denoted by t_J can be either observed exactly or be right-censored.

For a multi-state model with or without an absorbing state, let A denotes the largest state number. For example, $A = 4$ for a four-state multi-state process, such as CAV data and OCTO-Twin data. The probability $S(y_J|y_{J-1}, \boldsymbol{\theta}, \boldsymbol{x})$ for transition (Y_{J-1}, Y_J) can be classified into three types:

If the time t_J at the transition to state Y_J is right-censored, it means that the individual is alive at time t_J, but it is unknown that which state the individual is

in.

$$S(y_J|y_{J-1},\boldsymbol{\theta},\boldsymbol{x}) = \sum_{s=1}^{A-1} P(Y_J = s|Y_{J-1} = y_{J-1},\boldsymbol{\theta},\boldsymbol{x}), \tag{3.4.7}$$

here s represents the observed states except for the number of the state$= A$. For example, $s = 1, 2, 3$ for CAV and OCTO-Twin data. This expression can be refer to Equation (3.1.3), where $S(t)h(t) = f(t)$.

If the state Y_J at time t_J is an observed absorbing state,

$$S(y_J|y_{J-1},\boldsymbol{\theta},\boldsymbol{x}) = \sum_{s=1}^{A-1} P(Y_J = s|Y_{J-1} = y_{J-1},\boldsymbol{\theta},\boldsymbol{x})\, h_{sA}(t_{J-1}|\boldsymbol{\theta},\boldsymbol{x}), \tag{3.4.8}$$

where h_{sA} denotes the transition intensity for the transition (s, A).

For the state Y_J at time t_J is a living state,

$$S(y_J|y_{J-1},\boldsymbol{\theta},\boldsymbol{x}) = P(Y_J = y_J|Y_{J-1} = y_{J-1},\boldsymbol{\theta},\boldsymbol{x}). \tag{3.4.9}$$

Assuming independence between individuals, the likelihood function for all N individuals is

$$L(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{x}) = \prod_{i=1}^{N} L_i(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{x}) \tag{3.4.10}$$

The expressions discussed in this section also can be found in Van Den Hout (2016) and Jackson (2011).

As an example consider the likelihood function for the CAV data. There are 622 individuals, three living states and one dead state (state 4), the transitions into the dead state are exactly observed. Assuming the time in the hazard function is Gompertz-distributed, the parameter $\boldsymbol{\theta}$ consists of $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$, $\boldsymbol{y}$ is the sequence of observed states of individuals. Therefore, the contribution for individual i who

died is given by

$$
\begin{aligned}
L_i(\boldsymbol{\beta}, \boldsymbol{\xi} | \boldsymbol{state}_i, \boldsymbol{x}_i) = & \Big(\prod_{j=2}^{J-1} P(state_{ij} = state_{ij} | state_{ij-1} = state_{ij-1}, \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{x}_i)\Big) \\
& \times \sum_{s=1}^{3} P(Y_{iJ} = s | Y_{iJ-1} = y_{iJ-1}, \boldsymbol{\theta}, \boldsymbol{x}_i)\, q_{s4}(t_{iJ-1} | \boldsymbol{\theta}, \boldsymbol{x}_i),
\end{aligned} \tag{3.4.11}
$$

the contribution for individual i who is alive at the end of the study is given by

$$
\begin{aligned}
L_i(\boldsymbol{\beta}, \boldsymbol{\xi} | \boldsymbol{state}_i, \boldsymbol{x}_i) = & \Big(\prod_{j=2}^{J-1} P(state_{ij} = state_{ij} | state_{ij-1} = state_{ij-1}, \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{x}_i)\Big) \\
& \times \sum_{s=1}^{3} P(Y_{iJ} = s | Y_{iJ-1} = y_{iJ-1}, \boldsymbol{\theta}, \boldsymbol{x}_i).
\end{aligned} \tag{3.4.12}
$$

Therefore, the likelihood function is given by

$$
L(\boldsymbol{\beta}, \boldsymbol{\xi} | state, \boldsymbol{x}) = \prod_{i=1}^{622} L_i(\boldsymbol{\beta}, \boldsymbol{\xi} | state_i, \boldsymbol{x}_i), \tag{3.4.13}
$$

where $\boldsymbol{x}$ denotes the covariates in the model.

3.4.4 Maximum likelihood estimation in R software

The logarithm of likelihood is maximised in the R software by using the function `optim`. General-purpose Optimisation (`optim`) is based on Nelder-Mead, quasi-Newton and conjugate-gradient algorithms. The default method is `"Nelder-Mead"`, which was published by Nelder and Mead (1965). This method is robust but relatively slow for computing. There is another quicker but not robust common method called `"BFGS"`, which is a quasi-Newton method. It was proposed by Broyden (1970), Fletcher (1970), Goldfarb (1970) and Shanno and Kettler (1970), simultaneously. See the `stats` package in R software for details and other methods such as `"CG"` and `"L-BFGS-B"`.

Other commonly used methods are scoring algorithm, EM algorithm (Baum

et al. (1970); Dempster et al. (1977)) and Markov chain Monte Carlo (MCMC) (Ibrahim et al. (2001)). The scoring algorithm needs a function of the first-order derivative. It is more complex for computation and coding but might be easier to understand. The EM algorithm only can be used in a very specific situation, such as data with missing data. The benefits of the EM algorithm are that it is easier to use in much standard software and robust. Therefore, it is also friendly for researchers who do not have a strong statistics background. However, there are certain problems: this method is quite slow to get results, and standard errors of parameters cannot be calculated directly, it need more steps of computation. The major advantage of MCMC is that it can be used in multi-dimension frailty models. The frailty model will be discussed in Chapter 4.

3.5 Model selection

In the analysis of multi-state models, we can define different covariates in different transition hazards. Therefore, it is necessary to compare different models and select the proper one. This section below will introduce two useful methods of model comparison.

Akaike's information criterion (AIC) is a useful method to compare and select models, see Akaike (1974). This criterion is given by

$$AIC = -2log(\hat{L}) + 2k, \tag{3.5.1}$$

where, for the model under consideration, $\hat{L}$ is the maximum of the likelihood function and k is the number of independent regression parameters in the model. AIC is reported in the software R in some functions, like `AIC` in the package `stats`. Since a good model has a high likelihood and hopefully not many parameters, we look for models with the smallest AIC. The AIC can be seen as a combination of the likelihood and a penalty term, where the term $2k$ in the equation (3.5.1) can be interpreted as a penalty for too large models. The penalty $2k$ can avoid mod-

els containing too many parameters. More details about Akaike's information criterion are specified by Akaike (1974).

The Bayesian information criterion (BIC) is an alternative method to compare the models, see Schwarz (1978). It is given by

$$BIC = -2log(\hat{L}) + log(n)\, k, \tag{3.5.2}$$

where, for the model under consideration, n is the sample size of the data, $\hat{L}$ is the maximum of the likelihood function and k is the number of independent regression parameters in the model (including the intercept).

BIC has a stronger penalty than AIC for the sample size $n \geq 8$, but here for longitudinal data, it is hard to decide whether the n is the number of individuals or the number of observations. n as the number of individuals is the common choice (see Van Den Hout (2016); Muthén and Asparouhov (2008)). However, n is defined as the total number of independent observations, so neither of the above methods is optimal. There is a discussion of this problem in Carlin and Louis (2009). Therefore, we prefer to use AIC to do the model comparison in this study. Note that Böhnstedt and Gampe (2019) discussed that the standard AIC is biased in some specific frailty models due to the boundary parameter. The frailty model in our thesis will be discussed in the next two chapters, and the AIC of our frailty models are not biased, since we use the one parameter gamma distribution. Please see Section 4.2.2 for details.

3.6 Comparison of analytic expression and piecewise-constant approximation

As we discussed in Section 3.2 and 3.4, there are two methods to calculate transition probabilities. Although we will not use the analytic expression in data analysis for CAV and OCTO-Twin data, it is a good method for simple three or four state models because it does not need to use statistical software for the com-

putation. In this section, we would like to apply these two methods for a progressive four-state model, and compare their results.

For example, there is a progressive four-state model with five transitions: $1 \to 2$, $1 \to 4$, $2 \to 3$, $2 \to 4$ and $3 \to 4$. Assuming the time t follows the exponential distribution and transition hazards are

$$\begin{aligned} h_{12} &= \lambda_{12} = 0.20 \\ h_{14} &= \lambda_{14} = 0.40 \\ h_{23} &= \lambda_{23} = 0.30 \\ h_{24} &= \lambda_{24} = 0.04 \\ h_{34} &= \lambda_{34} = 0.27. \end{aligned}$$

Therefore, the Q matrix is

$$Q = \begin{pmatrix} 0.60 & 0.20 & 0 & 0.40 \\ 0 & 0.34 & 0.30 & 0.04 \\ 0 & 0 & 0.27 & 0.27 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

According to analytic expressions of transition probabilities in Section 3.2.1, if we assume $t_1 = 0$ and $t_2 = 1$, then

$$\begin{aligned} p_{11}(t_1, t_2) &= \exp(-(\lambda_{12} + \lambda_{14})(t_2 - t_1)) = 0.549 \\ p_{22}(t_1, t_2) &= \exp(-(\lambda_{23} + \lambda_{24})(t_2 - t_1)) = 0.712 \\ p_{33}(t_1, t_2) &= \exp(-(\lambda_{34})(t_2 - t_1)) = 0.763 \\ p_{12}(t_1, t_2) &= \frac{\lambda_{12}}{\lambda_{23} + \lambda_{24} - \lambda_{12} - \lambda_{14}} \\ &\quad \times (\exp((\lambda_{12} + \lambda_{14})(t_1 - t_2)) - \exp((\lambda_{23} + \lambda_{24})(t_1 - t_2))) \\ &= 0.125 \\ p_{13}(t_1, t_2) &= \frac{\lambda_{12}\lambda_{23}}{(\lambda_{34} - \lambda_{23} - \lambda_{24})(\lambda_{23} + \lambda_{24} - \lambda_{12} - \lambda_{14})} \end{aligned}$$

$$
\begin{aligned}
&\times (\exp((\lambda_{12}+\lambda_{14})(t_1-t_2)) - \exp((\lambda_{23}+\lambda_{24})(t_1-t_2))) \\
&- \frac{\lambda_{12}\lambda_{23}}{(\lambda_{34}-\lambda_{23}-\lambda_{24})(\lambda_{34}-\lambda_{12}-\lambda_{14})} \\
&\times (\exp((\lambda_{12}+\lambda_{14})(t_1-t_2)) - \exp(\lambda_{34}(t_1-t_2))) \\
&= 0.020
\end{aligned}
$$

$$
\begin{aligned}
p_{23}(t_1,t_2) &= \frac{\lambda_{23}}{\lambda_{34}-\lambda_{23}-\lambda_{24}}(\exp((\lambda_{23}+\lambda_{24})(t_1-t_2)) - \exp(\lambda_{34}(t_1-t_2))) \\
&= 0.221
\end{aligned}
$$

$$p_{21}(t_1,t_2) = p_{31}(t_1,t_2) = p_{32}(t_1,t_2) = 0$$

$$p_{14}(t_1,t_2) = 0.305$$

$$p_{24}(t_1,t_2) = 0.067$$

$$p_{34}(t_1,t_2) = 0.237$$

Alternatively, we use the function *MatrixExp* in the package *msm* in the R software to calculate the transition probability matrix, and we get the same results.

Next, Table 3.1 shows the comparison of the speed of these two methods. Since it is too fast to get a result for this simple example, we set a loop to execute the computation 10000 times. The time needed for the computation can be investigated by the function `system.time` in the R software. The user time is the CPU time charged for the execution of user instructions of the calling process. The system time is the CPU time charged for execution by the system on behalf of the calling process, see R Core Team et al. (2013) for details. As we can see, the analytic expression is much faster than the piecewise constant approximation. It is reasonable since the analytic expression can get the results directly. Therefore, analytic expression is a good choice when we model a simple multi-state process, like the three-state model and the progressive four-state model.

Table 3.1: The speed of computation in the R software of analytic expression and piecewise-constant approximation

Method	Speed(in seconds)		
	User	System	Elapsed
Piecewise-constant approximation	3.729	0.036	3.852
Analytic expression	0.097	0.001	0.098

Chapter 4

Univariate frailty model

4.1 Introduction to frailty

In the previous chapter, we discussed the fixed-effect parametric hazard regression model. However, variations in transition times and survival times are often found between individuals in the study, even beyond the variation that can be explained by the covariates. Under the same value of covariates in the model, some individuals are more likely to move to severe disease states or death than others, and some individuals are more likely to remain in a healthier state. In other words, some people may take a shorter time to death than others. These kinds of individuals can be described as frailer than others. These differences might be because of many reasons, such as some variables that affect the survival are not included in the data. In survival analysis, these kinds of unobserved effects on the hazard are called frailty (Vaupel (1979)). Including a frailty in the model can help us to describe the unobserved heterogeneity, the estimated spread of the frailty indicates the amount of unobserved heterogeneity. Frailty is usually defined as a multiplicative random effect in the hazard function.

There are univariate or multivariate frailties in the multi-state model. The univariate frailty model can be used to explain the lack of fit (Balan and Putter (2020)), such as the heterogeneity due to the missing covariates. The multivariate

frailty model can explain the dependency between different transition hazards, as we discussed above. This chapter only discusses univariate frailty models, bivariate frailty models will be discussed in the next chapter. Further multivariate frailty models (more than two frailties) are outside our scope, but is discussed in Chapter 8 as a topic for future study.

Furthermore, including frailty may help to explain and quantify the association of transition hazards in the multi-state model. This type of association of transition hazards indicates a violation of the Markov assumption. In this study, we assume that the continuous-time multi-state model is a Markov process conditional on covariates. In other words, given the covariate values, we assume that only the current state determines the future. In reality, future transitions between states are probably affected by other factors, such as the duration of time in the current state, or the history of states. For example, individuals who have been in a disease state for a long time, may be more likely to move to the death state than individuals who just entered the disease state. This type of deviation from a Markov process might be addressed with fitting a frailty model. For more detailed information, we recommend the paper by Putter and Houwelingen (2011).

Frailties can be defined either at the individual or group level. The former implies individual-specific transition hazards, the latter implies groups of individuals share a group-specific frailty. The group-specific frailty can be applied where the survival time for a group of individuals is not independent. For example, individuals in the same hospital may have a common impact factor, their transition hazard and transition times will therefore be correlated. In this situation, it is good to assume that all the individuals in this hospital share the same frailty. This is a major advantage of using frailty instead of the fixed-effect parameter to represent different groups. For individuals from a large number of different hospitals, using an indicator variable in the fixed-effect model will increase the number of parameters. It tends to produce a more complicated model and more complex for estimation.

In our study of multi-state models, we will mainly focus on adding the uni-

variate and bivariate individual-specific frailty in the hazard of some transitions to describe the unobserved heterogeneity and investigate the correlation between transitions. The dependency of individuals is not considered in this study, and the shared frailty is thereby beyond the scope of our study. In the application, there is no proper group information in the CAV data for fitting shared frailty. For the OCTO-Twin data, the participants are identical and same-sex fraternal twins, but we do not take this genetic dependency into account in the present study, it is a topic for further work, see more discussion in Chapter 8.

4.2 Parametric frailty model

For the multi-state survival model, the variation of hazards between different individuals can be defined as individual-specific frailties, which means each individual has his or her own frailty. This frailty parameter can be multiplied by the regression hazard function mentioned in Chapter 3.

For the parametric frailty model, the hazard function for individual i in transition (r, s) is given by

$$h_{rs.i}(t|\boldsymbol{x}, i) = h_{rs.0}(t) \exp(\boldsymbol{\beta}_{rs}^{\top}\boldsymbol{x}) b_{rs.i}\,, \tag{4.2.1}$$

where $\boldsymbol{x}$ is the vector of fixed-effect covariates, $\boldsymbol{\beta}_{rs}$ is a parameter vector, $h_{rs.0}(t)$ is the baseline hazard. $B_{rs.i}$ is the frailty which is assumed to be positive, since the hazard function Equation (4.2.1) must be positive. Note that $B_{rs.i}$ can be changed to $B_{rs.g}$ for a group shared random effect.

Several parametric frailty models are discussed in Munda et al. (2012) In this chapter, we discuss when the frailty $B_{rs.i}$ in Equation (4.2.1) follows lognormal distribution or gamma distribution.

4.2.1 Lognormal distribution

The frailty variable $B_{rs.i}$ follows a lognormal distribution:

$$Brs.i = \exp(Vrs.i) \sim lognormal(\sigma_{rs}), \tag{4.2.2}$$

where $V_{rs.i} \sim N(0, \sigma_{rs}^2)$.

The expected value of $B_{rs.i}$ is given by

$$E(B_{rs.i}) = \exp(\sigma_{rs}^2/2),$$

and variance of $B_{rs.i}$ is given by

$$Var(B_{rs.i}) = \exp(\sigma_{rs}^2)(\exp(\sigma_{rs}^2) - 1).$$

It can be derived from Equation (4.2.1) that there is no frailty when $B_{rs.i} = 1$, which equals to $V_{rs.i} = 0$ from Equation (4.2.2).

4.2.2 Gamma distribution

The gamma distribution is another widely used distribution for frailty. The probability density function of the gamma distributed frailty B is given by

$$f(b) = \frac{\gamma^\alpha}{\Gamma(\alpha)} b^{\alpha-1} e^{-\gamma b}, \quad \text{for} \quad b > 0$$

where $\alpha > 0$ is the shape parameter, $\gamma > 0$ is the rate parameter, and the gamma function $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}/\exp(t)dt$. The expectation and variance of B in gamma distribution are $E(B) = \alpha/\gamma$, $Var(B) = \alpha/\gamma^2$.

The gamma distributed frailty is popular used for the standard survival model, since it leads to a closed form of the likelihood function (Abbring and Van Den Berg (2007)), but the advantage does not hold in this study because of the com-

plexity of the likelihood function of multi-state models discussed in Section 3.4. The advantage of gamma distribution in multi-state model will be discussed below. For the more simple computation in this study, we would like to use the one parameter gamma distribution, i.e. $B_{rs.i} \sim Gamma(\kappa)$, where the shape parameter $\alpha = \kappa$ and the rate parameter $\gamma = \kappa$. It follows that the probability density of $B_{rs.i}$ is given by

$$f(B_{rs.i}) = \frac{\kappa^\kappa}{\Gamma(\kappa)} B_{rs.i}^{\kappa-1} e^{-\kappa B_{rs.i}}, \quad \text{for} \quad B_{rs.i} > 0, \tag{4.2.3}$$

where the expectation and variance of $B_{rs.i}$ are given by

$$E(B_{rs.i}) = 1, \quad Var(B_{rs.i}) = \frac{1}{\kappa}.$$

Except for a more simple computation, an alternative advantage if frailty follows the one parameter gamma distribution is that the expected value of frailty will always be 1. By contrast, if the frailty follows the lognormal distribution, the expected value $E(V_{rs.i}) = 0$ does not imply the expected value $E(B_{rs.i}) = 1$.

4.3 Parametric likelihood function

For fitting parametric frailty models, consider the conditional hazard function for transition $r \to s$

$$h_{rs}(t|b_{rs.i}, \boldsymbol{x}, i) = h_{rs.0}(t) \exp(\boldsymbol{\beta}_{rs}^\top \boldsymbol{x}) b_{rs.i}. \tag{4.3.1}$$

Conditionally on $b_{rs.i}$, this hazard function has the fixed-effect format with an additional intercept $b_{rs.i}$. Thus, the transition hazard matrix $\boldsymbol{Q}$ and transition probability matrix $\boldsymbol{P}(t)$ conditional on the frailty term can be calculated in a similar approach with the fixed-effect model discussed in Chapter 3.

It follows that the likelihood contribution for the frailty model for individual

i with absorbing state is given by

$$\begin{aligned} L_i(\boldsymbol{\theta}|i,\boldsymbol{y},\boldsymbol{x}) &= P(Y_2 = y_2,\ldots,Y_J = y_J|Y_1 = y_1,\boldsymbol{\theta},\boldsymbol{x}) \\ &= \int_{\Omega_{b_i}} P(Y_2 = y_2,\ldots,Y_J = y_J|Y_1 = y_1,\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{b}_i) f(b_i|\boldsymbol{\theta}) db_i \\ &= \int_{\Omega_{b_i}} \prod_{j=2}^{J-1} P(Y_j = y_j|Y_{j-1} = y_{j-1},\boldsymbol{\theta},\boldsymbol{x},b_i)\, S(y_J|y_{J-1},\boldsymbol{\theta},\boldsymbol{x},b_i) \\ &\quad \times f(b_i|\boldsymbol{\theta}) db_i, \end{aligned} \tag{4.3.2}$$

where $\boldsymbol{\theta}$ is the vector of all the parameters in the model, $\boldsymbol{y} = (y_1, y_2, \ldots, y_J)$ are the observed states corresponding to observed time $t_1, t_2, \ldots, t_J$, b_i denotes the frailty of individual i, the $\Omega_{\boldsymbol{b}_i}$ is the parameter space of frailty b_i, multiplicative term $S(y_J|y_{J-1})$ is same defined as the expression in Chapter 3. Note that the transition probability for the frailty model are conditional independent on the frailty b_i.

Assuming the independence of individuals, the likelihood function for N individuals is given by

$$L(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{x}) = \prod_{i=1}^{N} L_i(\boldsymbol{\theta}|i,\boldsymbol{y},\boldsymbol{x}). \tag{4.3.3}$$

4.3.1 Common used numerical integration methods

Integrals in the likelihood function of the frailty model in Equation (4.3.2) and (4.3.3) can be approximately calculated by several numerical integration methods. The materials in this section also can be found in Davis and Rabinowitz (2007).

Trapezoidal rule

Trapezoidal rule is a numerical method to be used to approximate an integral. It is only suitable for one-dimension integral $\int_a^b f(x)dx$.

This method is used by dividing the area under the curve of $f(x)$ into many parts of trapezoids, where the number of trapezoids is n, then approximate the

total areas. The expression is given by

$$\int_a^b f(x)dx \approx \frac{b-a}{2n}(f(x_0) + 2f(x_1) + 2f(x_2) + \ldots + 2f(x_{n-1}) + f(x_n)),$$

where $x_i = a + \frac{(b-a)i}{n}$ for $i = 1, 2, \ldots, n$.

An alternative trapezoidal method is more accurate, it follows

$$\int_a^b f(x)dx \approx \sum_{n=1}^{N} \frac{f(x_{n-1}) + f(x_n)}{2}(x_n - x_{n-1}).$$

Simpson's rule

Simpson's rule, also called Kepler's rule, is similar to the trapezoidal rule but more accurate. It is only used for one-dimension integral as well. The expression of this method is given by

$$\begin{aligned}\int_a^b f(x)dx \approx & \frac{b-a}{3n}(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \ldots \\ & + 2f(x_{n-1}) + f(x_n)),\end{aligned} \tag{4.3.4}$$

where $x_i = a + \frac{(b-a)i}{n}$ for $i = 1, 2, \ldots, n$.

Gauss-Hermite quadrature

The Gauss-Hermite quadrature is a form of Gaussian quadrature, which can be used to approximate the integrals in the form as follow:

$$\int_{-\infty}^{+\infty} e^{-x^2} f(x)dx \approx \sum_{i=1}^{n} w_i f(x_i),$$

where n is the number of sample points, the nodes x_i ($i = 1, 2, \ldots, n$) is the root of the physicists' version of the Hermite polynomial $H_n(x)$, and the associated

weight w_k is

$$w_i = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 (H^{n-1} x_i)^2}.$$

See Shao et al. (1964) and Oliver et al. (2010) about the details and examples of Hermite polynomial and Gauss-Hermite formula. The Gauss-Hermite quadrature is a good option when the frailty follows a lognormal distribution.

4.3.2 Other numerical methods

One common approach for approximating the multivariate integral is phrasing the multiple integrals as repeated one-dimensional integrals. Alternatively, there are some methods that are easy to apply for multivariate integrals.

For example, see Press and Farrar (1990), and Weinzierl (2000) for Monte Carlo integration; See Hilgenfeldt et al. (1995) and Garcke (2012) for sparse grids.

There are also other approaches that can avoid quadrature and are simpler to evaluate the maximum likelihood function of the frailty model. For instance, the Laplace approximation and EM algorithm. For the EM algorithm, frailty can be seen as a missing value.

4.4 Transformation of parameters

When we do the computation in the R software, the fixed-effect model can be fitted by using `msm` package in R, the estimation of parameters and the standard errors can be derived directly. However, for the frailty model, we need to fit it by using user-written code, since there is no proper package for these models. Several no restricted parameters are defined when we maximise the likelihood function for frailty models by using general-purpose optimisation, see the line 3-5 in Appendix A.2. Thus, the frailty parameters $\sigma \geq 0$ and $\kappa \geq 0$ for the log-normal distribution and one parameter gamma distribution are supposed to be transformed to $\sigma^* \in \mathbf{R}$ and $\kappa^* \in \mathbf{R}$, where $\sigma = \exp(\sigma^*)$ and $\kappa = \exp \kappa^*$. The corresponding standard error for these parameters can be derived by the delta

method. Please see Appendix B for the details of the delta method. The materials of this method is partly based on the book *Delta Method* (Cox (2005)).

Chapter 5

Bivariate frailty model

Chapter 4 discussed the general concept of frailty models, and presented the hazard function and likelihood of multi-state models which include univariate frailties. In this chapter, we will explore the bivariate frailty model.

For bivariate survival analysis, the shared frailty model and the correlated frailty model are two important approaches, which are widely used in previous publications, such as Hougaard (2012), Duchateau and Jansen (2007). In a shared frailty model, the frailty is common for individuals in the same group, and it can help to describe the dependence between individuals. The correlated frailty model is usually seen as an extension of the shared frailty model (Hens et al. (2009)). In the correlated frailty model, frailties are correlated instead of shared for individuals in the same group. There is an additional correlation parameter, and all the correlations between group members are equal. For example, Yashin et al. (1995) introduced statistical modelling for data on twin pairs. They discussed a shared-frailty model at first, where frailty is defined as a measure of the relative risk, and both twins in a pair shared the same frailty value. Next, they defined two different frailties for each individual in a twin pair, but these values are correlated. This is called a bivariate correlated frailty model. The association was described by a correlation coefficient between the two frailties of the twin pair.

Both methods mentioned above are used for survival models. In this chapter, we would like to investigate the correlation between different transition hazards in a multi-state model. Therefore, the correlated frailty models are discussed. The main contribution in this chapter is that we will use Cheriyan and Ramabhadran's bivariate gamma distribution (Kotz (2004)) to fit a frailty model. The correlated frailty model with this distribution was first proposed by Yashin et al. (1995) for univariate survival data on related event times. Thereafter this approach was used in a number of papers, such as Hens et al. (2009) and Martins et al.(2018). The novelty of our study is that we consider a bivariate gamma multi-state frailty model for the interval-censored data based on this distribution. The correlation is defined for two different transitions rather than paired individuals.

5.1 Parametric frailty model

In order to investigate the correlation between two transitions in the multi-state model, we are able to consider an extension of the frailty model in Equation (4.2.1): use the bivariate frailty distribution to describe frailties in two transition hazards for an individual. For example, for an individual i who makes transitions from $r \to s$ and $p \to q$:

$$\begin{aligned} h_{rs.i}(t|\boldsymbol{x}, B_{rs.i}) &= h_{rs.0}(t)\exp(\boldsymbol{\beta}_{rs}^{\top}\boldsymbol{x})B_{rs.i} \\ h_{pq.i}(t|\boldsymbol{x}, B_{pq.i}) &= h_{pq.0}(t)\exp(\boldsymbol{\beta}_{pq}^{\top}\boldsymbol{x})B_{pq.i} \end{aligned} \tag{5.1.1}$$

where $B_{rs.i}$ and $B_{pq.i}$ follow a bivariate distribution which can be correlated or independent.

5.1.1 Lognormal distribution

The bivariate lognormal distribution that may be defined by two univariate lognormal variables B_1 and B_2. The materials in this section also can be found in Patel and Read (1996).

Define $B_1 = \exp(V_1)$, $B_2 = \exp(V_2)$, where $V_1 \sim N(\mu_1, \sigma_1^2)$ and $V_2 \sim N(\mu_2, \sigma_2^2)$, the joint probability density function for V_1, V_2 is

$$f(v_1, v_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(\frac{\frac{(v_1-\mu_1)^2)}{\sigma_1^2} - \frac{2\rho(v_1-\mu_1)(v_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(v_2-\mu_2)^2}{\sigma_2^2}}{-2(1-\rho^2)} \right).$$

Thus, the conditional probability density function for $(V_2|V_1)$ is

$$\begin{aligned} f(v_2|v_1) =& \frac{f(v_1, v_2)}{f(v_1)} \\ =& \frac{1}{\sqrt{2\pi\sigma_1\sigma_2(1-\rho^2)}} \exp\left(\frac{\frac{(v_1-\mu_1)^2)\rho^2\sigma_2^2}{\sigma_1^2} - \frac{2\rho\sigma_2(v_1-\mu_1)(v_2-\mu_2)}{\sigma_1} + (v_2-\mu_2)^2}{-2\sigma_2^2(1-\rho^2)} \right). \end{aligned}$$

Therefore, the probability density function for (V_1, V_2) is

$$f(v_1, v_2) = f_{V_2|V_1}(v_2|V_1 = v_1) f_{V_1}(v_1),$$

where $V_2|V_1 \sim N(\mu_2 + \rho(\frac{\sigma_2}{\sigma_1})(v_1 - \mu_1), \sigma_2^2(1-\rho^2))$, $\rho = Cor(v_1, v_2) = \frac{Var(V_1,V_2)}{\sigma_1\sigma_2}$.

Thus, the correlation coefficient of B_1 and B_2 (Mostafa and Mahmoud (1964)) is given by

$$\rho_{B_1B_2} = \frac{Cov(B_1, B_2)}{\sqrt{(\exp(\sigma_1^2 - 1))(\exp(\sigma_2^2 - 1))}}.$$

5.1.2 Gamma distribution

The bivariate gamma distributions are regarded as a generalization of univariate gamma distributions. There are various forms of bivariate gamma distributions, see Kotz (2004). In this study, we use Cheriyan and Ramabhadran's bivariate gamma distribution, the advantages are: (1) there are only two parameters, and (2) the bivariate density function is constructed by three univariate gamma variables which is convenient when using a marginal likelihood function. Both advantages reduce the computational challenge of fitting frailty models. Alternative bivariate gamma-distributed frailties for survival models as well as their

advantages and disadvantages can be found in Martins et al. (2019).

For the Cheriyan and Ramabhadran's bivariate gamma distribution, let Z_0, Z_1 and Z_2 be three independent gamma random variables with probability density functions

$$f(z_k) = \frac{\gamma_k^{\alpha_k}}{\Gamma(\alpha_k)} z^{\alpha_k - 1} e^{-\gamma_k z}, \quad \text{for} \quad z_k > 0,$$

where shape $\alpha_k > 0$, rate $\gamma_0 = \gamma_1 = \gamma_2 = \gamma > 0$, $k = 0, 1, 2$.

We obtain the joint density function of $(Z_0, Z_1, Z_2)^\top$ as

$$f(z_0, z_1, z_2) = \frac{e^{-\gamma(z_0+z_1+z_2)}}{\Gamma(\alpha_0)\Gamma(\alpha_1)\Gamma(\alpha_2)} z_0^{\alpha_0-1} z_1^{\alpha_1-1} z_2^{\alpha_2-1} \gamma^{\alpha_0+\alpha_1+\alpha_2},$$

where $z_k > 0$, $\alpha_k > 0$ $(k = 0, 1, 2)$.

Let B_1, B_2 be the correlated gamma frailties,

$$B_1 = Z_0 + Z_1, \quad B_2 = Z_0 + Z_2,$$

then the joint density function of $(Z_0, B_1, B_2)^\top$ is given by

$$f(z_0, b_1, b_2) = \frac{e^{-\gamma(b_1+b_2-z_0)}}{\Gamma(\alpha_0)\Gamma(\alpha_1)\Gamma(\alpha_2)} z_0^{\alpha_0-1} (b_1 - z_0)^{\alpha_1-1} (b_2 - z_0)^{\alpha_2-1} \gamma^{\alpha_0+\alpha_1+\alpha_2}, \quad (5.1.2)$$

where $b_1, b_2 \geq z_0 \geq 0$.

In order to obtain the joint density function of $(B_1, B_2)^\top$, it is necessary to integrate out Z_0 in Equation (5.1.2)

$$\int_0^{min(b_1,b_2)} z_0^{\alpha_0-1} (b_1 - z_0)^{\alpha_1-1} (b_2 - z_0)^{\alpha_2-1} e^{\gamma z_0} dz_0$$

Therefore, the bivariate gamma density function is given by

$$f(b_1, b_2) = \frac{\gamma^{\alpha_0+\alpha_1+\alpha_2} e^{-\gamma(b_1+b_2)}}{\Gamma(\alpha_0)\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^{min(b_1,b_2)} z_0^{\alpha_0-1}(b_1 - z_0)^{\alpha_1-1}(b_2 - z_0)^{\alpha_2-1} e^{\gamma z_0} dz_0. \tag{5.1.3}$$

Since $E(B_1) = \frac{\alpha_0+\alpha_1}{\gamma} = 1$, $E(B_2) = \frac{\alpha_0+\alpha_2}{\gamma} = 1$, we obtain that $\alpha_1 = \alpha_2$, $\gamma = \alpha_0 + \alpha_1$. Therefore, frailties B_1 and B_2 follow the same marginal gamma distribution:

$$B_1 \sim Gamma(\alpha_0 + \alpha_1, \alpha_0 + \alpha_1), \;\; B_2 \sim Gamma(\alpha_0 + \alpha_1, \alpha_0 + \alpha_1), \tag{5.1.4}$$

The correlation coefficient is

$$\rho_B = \frac{Cov(B_1, B_2)}{\sqrt{Var(B_1)Var(B_2)}} = \frac{\alpha_0}{\alpha_0 + \alpha_1} > 0. \tag{5.1.5}$$

Since $\rho > 0$, this method is only suitable for adding frailties when transition hazards are positively correlated transition hazards. For example, we can not define frailties in transition $2 \to 3$ and $2 \to 1$ for OCTO-Twin data in this study, see Figure 2.4 in Chapter 2 for the transition information.

5.2 Parametric likelihood function

The likelihood contribution for frailty model for individual i is given by

$$\begin{aligned} L_i(\boldsymbol{\theta}|i, \boldsymbol{y}, \boldsymbol{x}) &= P(Y_2 = y_2, \ldots, Y_J = y_J | Y_1 = y_1, \boldsymbol{\theta}, \boldsymbol{x}) \\ &= \int_{\Omega_{b_i}} P(Y_2 = y_2, \ldots, Y_J = y_J | Y_1 = y_1, \boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{b}_i) f(\boldsymbol{b}_i|\boldsymbol{\theta}) d\boldsymbol{b}_i \end{aligned} \tag{5.2.1}$$

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_J)$ are the observed states corresponding to observed time $t_1, t_2, \ldots, t_J$, $\boldsymbol{b}_i = (b_{1i}, b_{2i}, \ldots, b_{ni})$ denotes a vector of random effects for individual i for n-dimensional frailty models. In this thesis, we only discuss the bivariate frailty model, so $\boldsymbol{b}_i = (b_{1i}, b_{2i})$. $\int_{\Omega_{b_i}}$ denotes the integration for 2-dimensional frailty models.

Similar to the fixed-effect model in Chapter 4, the likelihood for an individual with no observed death time is

$$L_i(\boldsymbol{\theta}|i,\boldsymbol{y},\boldsymbol{x}) = \int_{\Omega_{b_i}} \prod_{j=2}^{J} P(Y_j = y_j|Y_{j-1} = y_{j-1},\boldsymbol{\theta},\boldsymbol{x}) f(\boldsymbol{b}_i|\boldsymbol{\theta}) d\boldsymbol{b}_i, \tag{5.2.2}$$

and the likelihood for an individual with observed death time is

$$\begin{aligned} L_i(\boldsymbol{\theta}|i,\boldsymbol{y},\boldsymbol{x}) = \int_{\Omega_{b_i}} & \prod_{j=2}^{J-1} P(Y_j = y_j|Y_{j-1} = y_{j-1},\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{b}_i) \\ & \times \sum_{s=1}^{A-1} P(Y_J = s|Y_{J-1} = y_{J-1},\boldsymbol{\theta},\boldsymbol{x})\, h_{sA}(t_{J-1}|\boldsymbol{\theta},\boldsymbol{x}) f(\boldsymbol{b}_i|\boldsymbol{\theta}) d\boldsymbol{b}_i, \end{aligned} \tag{5.2.3}$$

where s and A are defined the same as for the univariate likelihood function in Chapter 4. Note that the transition probability for the frailty model are conditional independent on frailty B_i.

Assuming the independence of individuals, the likelihood function for N individuals is given by

$$L(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{x}) = \prod_{i=1}^{N} L_i(\boldsymbol{\theta}|i,\boldsymbol{y},\boldsymbol{x}). \tag{5.2.4}$$

As an example, the likelihood contribution of individual i for the bivariate

gamma frailty model in Equation (5.1.3) is given by

$$
\begin{aligned}
& L_i(\boldsymbol{\theta}|i,\boldsymbol{y},\boldsymbol{x}) \\
=& P(Y_2=y_2,\ldots,Y_J=y_J|Y_1=y_1,\boldsymbol{\theta},\boldsymbol{x}) \\
=& \int_{b_1}\int_{b_2} P(Y_2=y_2,\ldots,Y_J=y_J|Y_1=y_1,\boldsymbol{\theta},\boldsymbol{x},b_1,b_2)f(b_1,b_2)db_1db_2 \\
=& \int_{b_1}\int_{b_2}\int_{z_0} P(Y_2=y_2,\ldots,Y_J=y_J|Y_1=y_1,\boldsymbol{\theta},\boldsymbol{x},b_1,b_2) \\
& \times \frac{\gamma^{\alpha_0+\alpha_1+\alpha_2}e^{-\gamma(b_1+b_2)}}{\Gamma(\alpha_0)\Gamma(\alpha_1)\Gamma(\alpha_2)} z_0^{\alpha_0-1}(b_1-z_0)^{\alpha_1-1}(b_2-z_0)^{\alpha_2-1}e^{\gamma z_0}db_1db_2dz_0 \\
=& \int_{z_1}\int_{z_2}\int_{z_0} P(Y_2=y_2,\ldots,Y_J=y_J|Y_1=y_1,\boldsymbol{\theta},\boldsymbol{x},z_0,z_1,z_2) \\
& \times \frac{\gamma^{\alpha_0+\alpha_1+\alpha_2}e^{-\gamma(z_0+z_1+z_2)}}{\Gamma(\alpha_0)\Gamma(\alpha_1)\Gamma(\alpha_2)} z_0^{\alpha_0-1}(z_1)^{\alpha_1-1}(z_2)^{\alpha_2-1}dz_1dz_2dz_0 \\
=& \int_{z_0}\int_{z_1}\int_{z_2} P(Y_2=y_2,\ldots,Y_J=y_J|Y_1=y_1,\boldsymbol{\theta},\boldsymbol{x},z_0,z_1,z_2)f(z_0)f(z_1)f(z_2)dz_0dz_1dz_2.
\end{aligned}
\tag{5.2.5}
$$

Cook and Lawless (2018) also discussed the marginal likelihood function for multi-state frailty models, see Chapter 6 in that book for details.

Some parameters in the likelihood function are defined on restricted domains. Such as $\alpha_i > 0$, $\gamma_i > 0$ for bivariate gamma frailty and $-1 < \rho < 1$ for bivariate lognormal frailty. For the likelihood maximization in the R software, all these parameters with restricted domain can be replaced by unrestricted parameters. For example, $\alpha_i > 0$, $\gamma_i > 0$ and $-1 < \rho < 1$ are estimated by maximizing $\alpha_i^* \in \mathbb{R}$ $\gamma_i^* \in \mathbb{R}$ and $\rho^* \in \mathbb{R}$, where

$$
\alpha_i = \exp(\alpha_i^*), \quad \gamma_i = \exp(\gamma_i^*), \quad \rho = 2\frac{\exp(\rho^*)}{1+\exp(\rho^*)} - 1. \tag{5.2.6}
$$

The corresponding standard errors for these parameters are computed by using the Hessian matrix and the delta method.

Chapter 6

Non-parametric frailty model

6.1 Introduction to non-parametric frailty model

The frailty models discussed in previous chapters assumes that the frailty is parametrically distributed, such as the gamma and lognormal. Alternatively, the non-parametric frailty model can be desirable, because it can improve the flexibility of the model by allowing a less restricted form of distributions. The non-parametric model fits data well no matter whether the frailty has a parametric distribution or not.

Non-parametric frailty models were discussed in some applications before. For instance, dos Santos et al. (1995) discuss the non-parametric hazard versus the non-parametric frailty distribution. Li et al. (1998) use a Cox proportional hazards model with a non-parametric frailty with a two-points support. Caroni et al. (2010) use a Weibull baseline hazard function and an individual-specific non-parametric frailty. There are other papers that discuss the non-parametric frailty models in Bayesian approaches. For example, using the Dirichlet Process prior (Manda (2011)) and the Polya tree prior (Walker and Mallick (1997)) for the frailty term.

In this study, we will only look at the univariate non-parametric frailty model. For the univariate non-parametric model, define that there are a number of frailty

parameters B_k according to the classes C_k, where $k = 1, 2, \ldots, K$. For each individual ($i \in C_k$), the probability distribution π_k of frailties B_k is unknown.

Except for the less restricted form of the distribution, another advantage of the non-parametric frailty model is that there is no need for integration since the number of B_k is countable. Therefore, it has a simplified computation when maximising the loglikelihood. The details of the loglikelihood estimation will be introduced in Section 6.2.

However, there is no proper way to determine the optimal K when we fit the model. A good approach to solve this problem is fitting several models with different K and using model selection criteria, such as Akaike information criterion (AIC) and Bayesian information criterion (BIC) that we introduced in Chapter 3. With regard to this approach, it may be hard to interpret the reason for a certain classification after we select the proper K by comparing different models. However, the aim of the non-parametric maximum likelihood estimation is to describe the underlying heterogeneity, which means it is not necessary to interpret the classes.

6.2 Non-parametric maximum likelihood

For the non-parametric frailty model with K classes C_k, where $k = 1, 2, \ldots, K$, the likelihood contribution for individual i under the Markov assumption is given by

$$\begin{aligned} L_i(\boldsymbol{\theta}|i, \boldsymbol{y}, \boldsymbol{x}) &= P(Y_J = y_J, \ldots, Y_2 = y_2 | Y_1 = y_1, i, \boldsymbol{\theta}, \boldsymbol{x}) \\ &= \sum_{k=1}^{K} P(Y_J = y_J, \ldots, Y_2 = y_2 | Y_1 = y_1, i \in C_k, \boldsymbol{\theta}, \boldsymbol{x}) \pi_k, \\ &= \sum_{k=1}^{K} \prod_{j=2}^{J} P(Y_j = y_j | Y_{j-1} = y_{j-1}, i \in C_k, \boldsymbol{\theta}, \boldsymbol{x}) \pi_k, \end{aligned} \tag{6.2.1}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta_0}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K)$ is a vector which combines the fixed-effects parameters and class-specific parameters. $\boldsymbol{\theta}_0$ denotes the fixed-effect parameters, $\boldsymbol{\theta}_1 =$

$\{B_1, \pi_1\}, \boldsymbol{\theta}_2 = \{B_2, \pi_2\}, \ldots, \boldsymbol{\theta}_K = \{B_K, \pi_K\}$. $\pi_k = P(i/ \in C_k)$ for each individual i. Here the class-specific parameters $\boldsymbol{B}_k$ are called the mass points, and the probability π_k are called the masses.

Therefore, the likelihood for N individuals is given by

$$L(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}) = \prod_{i=1}^{N} L_i(\boldsymbol{\theta}|i, \boldsymbol{y}, \boldsymbol{x}).$$

Computationally, it is easier to optimise over an unrestricted parameter space when maximising the loglikelihood function. Given the probabilities $\sum_{k=1}^{K} \pi_k = 1$, we use the logit link for probability π_k for class k.

For example, for a model with $K = 2$, the two probabilities π_k can be represented by a parameter η, where $\eta \in \mathbf{R}$.

$$\begin{aligned} \pi_1 &= \frac{1}{1 + \exp(\eta)} \\ \pi_2 &= 1 - \pi_1. \end{aligned} \tag{6.2.2}$$

For a model with class $K = 3$, the three probabilities π_k can be represented by two independent parameter π_1^* and π_2^*, where $\pi_1^*, \pi_2^* \in \mathbf{R}$,

$$\begin{aligned} \pi_1 &= \frac{1}{1 + \exp(\pi_1^*) + \exp(\pi_2^*)} \\ \pi_2 &= \frac{\exp(\pi_1^*)}{1 + \exp(\pi_1^*) + \exp(\pi_2^*)} \\ \pi_3 &= \frac{\exp(\pi_2^*)}{1 + \exp(\pi_1^*) + \exp(\pi_2^*)}. \end{aligned} \tag{6.2.3}$$

There are also other methods to maximise the likelihood function, such as the EM algorithm. See Aitkin (1999) for details.

6.3 Extensions of the non-parametric frailty model

We have introduced the univariate non-parametric frailty model and its likelihood function. It is possible to fit the non-parametric frailty model for clustered survival data, such as the healthcare data in which individuals are grouped by their healthcare provider. The most common method for this kind of data is the parametric shared-frailty model, which was briefly introduced in Section 4.1 and 5.1. Fitting the non-parametric frailty model for this kind of data is an alternative. The non-parametric model is more flexible in the choice of the distribution of the frailty in order to account for arbitrary multimodality and unpredictable skewness types (Walker and Mallick (1997)). For example, Manda (2011) investigates the non-parametric frailty model in a Bayesian framework to analyse community-clustered child survival in sub-Saharan Africa. Gasperoni et. al (2020) present a survival model with both a non-parametric discrete shared frailty, and a non-parametric baseline hazard. In these studies, individuals share the same non-parametric frailty in the same group, and it can explore the heterogeneity between groups.

In this thesis, we do not fit a non-parametric shared frailty model for clustered data, but explore covariate-specific frailties for people with different values of covariates. We have not seen this method used in the multi-state frailty model in the literature.

As we can see, the probabilities defined in Equation (6.2.2) and (6.2.3) do not define different probabilities π_k for different people. The model will be more useful if the probabilities of each frailty are related to the characteristic of people. Bartolucci and Farcomeni (2015) proposed an approach to consider parameterizing η with a linear predictor,

$$\eta = \boldsymbol{\delta}^T \boldsymbol{x}, \tag{6.3.1}$$

where $\boldsymbol{\delta}$ is a vector of parameters, $\boldsymbol{x}$ is a vector of covariates including an intercept.

By using this method, the probabilities π_k are related to the value of covari-

ates. For example, if we add gender as the covariates in Equation (6.3.1), then the mass probabilities for females and males with frailty B_k are not the same, they are $\pi_{female.k}$ and $\pi_{male.k}$, respectively. Moreover, we would like to explore using same covariates both for the fixed-effect part of hazard and for the frailty. This is illustrated in data analysis in the next chapter, where we plan to fit three models and compare their results: (1) adding gender as the covariate for the fixed-effect part of the hazard. (2) Adding gender as the covariate for the non-parametric frailty. (3) Adding gender as the covariate both for the fixed-effect part of hazard and for the non-parametric frailty. Please see Chapter 7.1.3 for details.

Chapter 7

Data analysis

In this chapter, we would like to explore the datasets to extend the fixed-effect analysis with frailty models. We will investigate our ideas of frailty models, and see whether we can get better fitting models and get a better insight into the process of cardiac allograft vasculopathy (CAV data) and cognitive impairment (OCTO-Twin data).

7.1 Cardiac allograft vasculopathy data

7.1.1 Fixed-effect model

We want to model a disease progression in our study. We fit five different fixed-effect models for CAV data with different covariates and distributions. As we discussed in Chapter 2, there are totally five transitions for progressive CAV data: $1 \to 2$, $1 \to 4$, $2 \to 3$, $2 \to 4$, $3 \to 4$. We define r and s as two states which describe a transition $r \to s$. The first model we fit is an intercepts-only model with no covariates. Next, we fit four Gompertz models with various covariates for different transitions. The Gompertz hazard increases exponentially with time and has been widely used to model adult mortality. The Weibull hazard is an alternative, which increases as a power function of time. In our study, we would like to use Gompertz distribution rather than Weibull distribution, since the latter

has a more complex computation w.r.t the power function of time, which sometimes lead to numerical problems in estimation. And these two distributions are quite similar, shapes of them are not very different. Since we are not focusing on the parametric time-dependent distributions but more focusing on the frailties, we would like to use Gompertz distribution for the data analysis. There are several other choices for the distribution of hazards which are not discussed in this thesis, please see Thatcher et al. (1998) and Rodrıguez (2010) for details. Specifically, Model 2 defines the effect of years of follow-up for all transitions. Model 3 defines the effect of donor's age and years of follow-up for all five transitions. For Model 4, the donor's age are defined for all the transitions as well, but years of follow-up are only defined for transition $1 \rightarrow 2$. For Model 5, individuals' baseline age are considered as a covariate for transition $1 \rightarrow 2$ and $1 \rightarrow 4$, since we would like to investigate the effects of individuals' baseline age for people from the healthy state to others. Details of hazard functions are showed below. Models with restricted covariates are able to reduce the number of parameters, resulting in a simpler computation when maximising the likelihood function.

Model 1. Intercepts-only model

This model assumes that the years of follow-up t follows the exponential distribution (i.e. $T \sim \exp(\lambda)$), where the transition-specific hazards can be represented as a constant, i.e. $h(t) = \lambda$, where $\lambda > 0$. The hazard models for all the five transitions $r \rightarrow s$ are given by

$$h_{rs}(t) = \exp(\beta_{rs.0}), \tag{7.1.1}$$

where t is the time scale represents the years of individuals during follow-up.

Model 2. Consider the effect of years of follow-up for all transitions

This model assumes that the years of follow-up t follows the Gompertz distribution, where the transition-specific hazards can be represented as $h(t) = \lambda \exp(\xi t)$ for $t \geq 0$, where $t \geq 0, \lambda > 0$. The parameters of the effect of time are defined for all the five transitions. Hazard models for these transitions are:

$$h_{rs}(t) = \exp(\beta_{rs.0} + \xi_{rs.1} t), \tag{7.1.2}$$

Model 3. Consider donor's age and years of follow-up as covariates for all transitions

This model defines the donor's age as the covariate for all transitions, since it is a reasonable variable that may affect the transition hazards. The years of follow-up t follows the Gompertz distribution as well. Hazard models for all the transitions $r \to s$ are given by

$$h_{rs}(t) = \exp(\beta_{rs.0} + \xi_{rs.1} t + \beta_{rs.2} dage), \tag{7.1.3}$$

where *dage* represents the age of heart donor for the heart transplant.

Model 4. Consider the effect of years of follow-up for specific transitions

Since the AIC value is quite large of Model 3 (please see Table 7.1), we would like to fit a model that still contains the effect of years of follow-up and donor' age as the covariate, but with a lower AIC value. Therefore, we define the effect of the years of follow-up in transition $1 \to 2$, in order to investigate more specifically when people move from the healthy state to the moderate CAV state. Hazard model for transition $1 \to 2$ is given by

$$h_{12}(t) = \exp(\beta_{12.0} + \xi_{12.1} t + \beta_{12.2} dage), \tag{7.1.4}$$

hazard models for transition $1 \to 4$, $2 \to 3$, $2 \to 4$, $3 \to 4$ are given by

$$h_{rs}(t) = \exp(\beta_{rs.0} + \beta_{rs.2} dage). \tag{7.1.5}$$

Model 5. Consider donor's age and participants' baseline age as covariates for restricted transitions

This model is based on Model 4, and newly include the participants' baseline age as the covariate for transition $1 \to 2$ and $1 \to 4$. This restriction is defined in order to investigate how the participants' baseline age affect their movement from a healthy state. The hazard model for the transition $1 \to 2$ is given by

$$h_{12}(t) = \exp(\beta_{12.0} + \xi_{12.1} t + \beta_{12.2} bage + \beta_{12.3} dage), \tag{7.1.6}$$

the hazard model for the transition $1 \to 4$ is given by

$$h_{23}(t) = \exp(\beta_{14.0} + \beta_{14.2} bage + \beta_{14.3} dage), \tag{7.1.7}$$

hazard models for transition $2 \to 3$, $2 \to 4$, $3 \to 4$ are given by

$$h_{rs}(t) = \exp(\beta_{rs.0} + \beta_{rs.3} dage), \tag{7.1.8}$$

where *bage* is the baseline age for individuals.

Table 7.1: The model covariates, distributions of baseline hazard, number of parameters, $-2\times$ maximum loglikelihood function and Akaike's information criterion of four fixed-effect multi-state models for the CAV data

Model	Baseline hazard	Number of parameters	$-2log(\hat{L})$	AIC
1	exponential	5	3519.4	3529.4
2	Gompertz	10	3450.5	3480.5
3	Gompertz	15	3490.0	3510.0
4	Gompertz	11	3463.6	3485.6
5	Gompertz	13	3446.7	3472.7

Table 7.1 shows the results of these four models. As we discussed in Chapter 3, AIC is used to select the best one among these four models. It is clear that Model 5 performs best according to the AIC value. Table 7.2 shows the parameters estimation and standard errors for Model 5. The covariate effect $\hat{\xi}_{12.1} = 0.12$ illustrates that the hazard for the transition from state 1 to state 2 increases when time goes. $\hat{\beta}_{12.2} = 0.002$ and $\hat{\beta}_{14.2} = 0.051$ indicates that people with higher baseline ages would have higher hazards for transition $1 \to 2$ and $1 \to 4$. Furthermore, conditional on the same years of follow-up, hazards for people moving to death from state 1 would increase more rapidly than them moving to state 2, since $\hat{\beta}_{14.2} > \hat{\beta}_{12.2}$. Effects for donor's age are positive for transitions $1 \to 2$ and $1 \to 4$. It illustrates a positive relationship between higher hazards and higher donor's age. For the rest of the three transitions, the standard errors are either greater or approximately equal to the absolute value of estimates. The large value of standard errors implies that we do not have enough information to estimate the parameter of interest. Therefore, it is hard to assess whether the effects are positive or negative. It is one of the reasons to explore the frailty model: if we allow more flexible models (e.g. different frailty models), we can compare the results of these frailty models with fixed-effect models fitted in this section, and check whether frailty models have lower AIC values and more reasonable estimations.

Table 7.2: The estimation (standard errors) of parameters for the model 5 for CAV data

Transitions	Intercept	t	*bage*	*dage*
$1 \to 2$	$-3.484(0.331)$	$0.12(0.023)$	$0.002(0.006)$	$0.026(0.006)$
$1 \to 4$	$-6.369(0.778)$	0	$0.051(0.015)$	$0.022(0.009)$
$2 \to 3$	$-1.255(0.287)$	0	0	$-0.006(0.009)$
$2 \to 4$	$-1.848(0.946)$	0	0	$-0.043(0.037)$
$3 \to 4$	$-1.059(0.34)$	0	0	$-0.007(0.011)$

According to AIC values, Model 5 is the best one among all these four models. However, this does not indicate that this model is a good-fitting model. The best model according to likelihood-ratio tests can still be a bad-fitting model. There-

fore, the assessment of goodness-of-fit is a necessary approach to check whether the model fits well. Since the `msm` package only works for time-independent models, where the time is restricted to be $T_{rs} \sim \exp(\lambda_{rs})$. Figure 7.1 shows the comparison of observed prevalence and expected prevalence in percentage for all the states for CAV data. From the graph, we can see that the model fits well for states 1, 2 and 3. However, there is some misfit with respect to state 4, where the prevalence of death is underestimated.

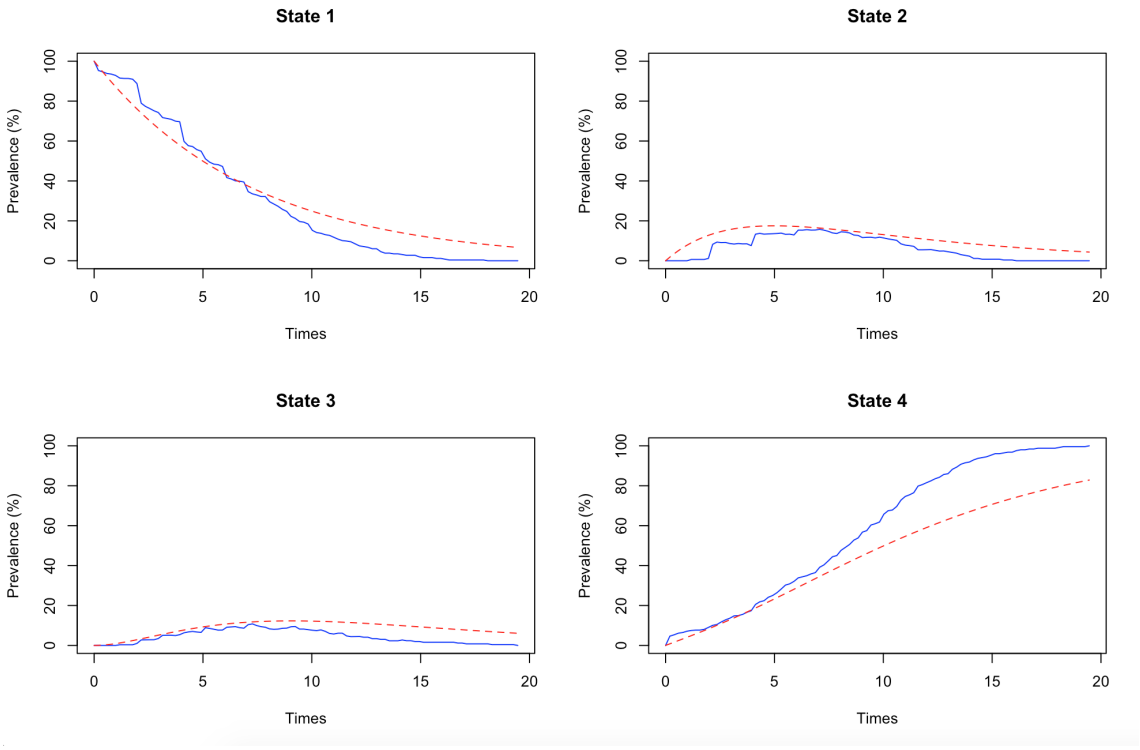

Figure 7.1: The comparison of observed prevalence (blue solid line) and expected prevalence (red dash line) for the CAV data. The observation is an approximated value due to the unknown information between two observed time points of interval-censored living states in CAV data. The expectation is based on the intercept-only model.

7.1.2 Parametric frailty model

As we mentioned in the previous section, it is good to fit frailty models to take into account the unobserved heterogeneity, and compare the results of fixed-effect models and frailty models.

In our study, we fit frailty models where the individual-specific frailty $B_{rs.i}$ follows a lognormal distribution, one parameter gamma distribution and non-parametric distribution. This section shows the result of two parametric models, results of non-parametric models and the comparison of these models will be shown in the next section.

The lognormal distribution and gamma distribution are two common choices for parametric frailty models. For the lognormal distribution, Gauss-Hermite quadrature can be used to do the integration when maximising the likelihood function. The approximation of Gauss-Hermite quadrature is more accurate compared with other numerical integration methods discussed in Chapter 4 (Van den Hout (2016)). The advantage of one parameter gamma distribution is that there is only one parameter for this distribution. It can simplify the computation, because in that case we only have one-dimensional integration when maximising the likelihood function. The frailty is defined in transition $1 \rightarrow 2$ for all the models, because we would like to investigate the latent information from healthy to not healthy. The fixed-effect parts of frailty models are derived from Model 5 discussed in the Section 7.1.1, and this model has the lowest AIC value

Table 7.3 and 7.4 shows the results for the lognormal distribution frailty model and one parameter gamma distribution frailty model, respectively. Comparing the estimation of intercept and the effects of the years of follow-up t of the lognormal frailty model and one parameter gamma frailty model, the differences are quite small. For the estimation of the effects of participants' baseline age *bage*: they are both positive in two transitions for these two models, though $\beta_{12.2} > \beta_{14.2}$ in the lognormal model but $\beta_{12.2} < \beta_{14.2}$ in the one parameter gamma model. This difference might be caused by the different frailty values. Comparing the results of these two frailty models and the best fixed-effect model (Model 5), the estimates are a little bit different. These changes are expected, since a role of the frailty can be to model the heterogeneity so that the fixed effects can be more accurately estimated. These changes of estimates might be because of some misfit of the fixed-effect model, e.g. some missing covariates.

Table 7.3: The estimation (standard errors) of parameters for the lognormal distribution frailty model for CAV data

Transitions	Intercept	t	*bage*	*dage*
$1 \to 2$	$-3.415(0.390)$	$0.127(0.036)$	$0.028(0.007)$	$-0.002(0.01)$
$1 \to 4$	$-6.326(0.786)$	0	$0.019(0.010)$	$-0.042(0.032)$
$2 \to 3$	$-1.555(0.351)$	0	0	$-0.004(0.011)$
$2 \to 4$	$-1.684(0.795)$	0	0	$-0.001(0.007)$
$3 \to 4$	$-1.208(0.347)$	0	0	$0.052(0.015)$
Frailty	$\sigma = 0.539(0.216)$			.

7.1.3 Non-parametric frailty model and model comparison

As the non-parametric frailty model described in Chapter 6, the number of classes is uncertain. A better approach is to fit several models with the different number of classes and select the best fitted one according to AIC values. Alternatively, state an explicit reason for choosing the number of classes. For example, we fit a non-parametric frailty model with $K = 2$ for CAV data, because we would like to distinguish movers (with higher frailties) from stayers (with lower frailties).

In this study, we fit three different non-parametric frailty models. (i) is a two-class non-parametric frailty model without parameterizing η. The frailty is defined in transition $1 \to 2$. (ii) is a two-class gender-specific non-parametric frailty model with the gender-specific frailty defined in transition $1 \to 2$. (iii) is a two-

Table 7.4: The estimation (standard errors) of parameters for the one parameter gamma distribution frailty model for CAV data

Transitions	Intercept	t	*bage*	*dage*
$1 \to 2$	$-3.941(0.377)$	$0.218(0.026)$	$0.011(0.007)$	$0.028(0.007)$
$1 \to 4$	$-5.646(0.670)$	0	$0.038(0.013)$	$0.021(0.010)$
$2 \to 3$	$-1.267(0.343)$	0	0	$-0.008(0.010)$
$2 \to 4$	$-2.409(0.917)$	0	0	$-0.018(0.029)$
$3 \to 4$	$-0.636(0.341)$	0	0	$-0.022(0.012)$
Frailty	$\kappa = 3.367(0.227)$			.

class non-parametric frailty model with gender as the covariate both for the fixed-effect part of hazard and the frailty. The fixed effect of gender and gender-specific frailty are both defined in transition $1 \rightarrow 2$. (iiii) is a fixed-effect model with the same covariates definition as Model (iii), in order to have a clearer comparison. Tables 7.5, 7.6, 7.7 and 7.8 shows the results for these three non-parametric frailty models and one fixed-effect model, respectively.

Regarding Model (i), the frailty parameters illustrate that the probability of a random patient being a mover is 61.8% ($b_1 > 1$) and 38.2% ($b_2 < 1$) chance to be a stayer during transition $1 \rightarrow 2$. In Model (ii), female patients are more likely to be stayers than movers during transition $1 \rightarrow 2$, since $\pi_1(35.9\%) < \pi_2(64.1\%)$. In contrast, males have a higher probability to be a mover rather than a stayer ($\pi_1(61.3\%) > \pi_2(38.7\%)$).

For the results of Model (iii) in Table 7.6, the effect of gender in transition $1 \rightarrow 2$ is -0.818, which means females (gender=1) have a lower hazard than males (gender=0). However, the probability of female patients being a mover ($b_1 = 3.177$) is 76.5% in the transition $1 \rightarrow 2$, which is higher than males (69.0%). This result is contrary to the effect of gender in the fixed-effect part of the estimation ($\hat{\beta}_{12.gender} = -0.818 < 0$). Moreover, the standard errors of both the estimation $\hat{\beta}_{12.gender}$ (1.14) and estimation of the probability of females (0.757) are quite large. This indicates that there might be a identifiability problem. To compare with Model (iii), Model (iiii) has a better result, where the estimation $\hat{\beta}_{12.gender} = -0.548$ has a small standard error (0.255). Therefore, we do not recommend including the same covariate both in the fixed-effect part and the non-parametric frailty part of the model.

In the previous section, we have discussed the fixed-effect model Model 5, which is the one that has the lowest AIC value. It is a fixed-effect Gompertz-distributed model with donor's age and participants' baseline age as covariates. We restricted the effect of time only in transition $1 \rightarrow 2$ and the effect of participants' baseline age in transition $1 \rightarrow 2$ and $1 \rightarrow 4$.

Table 7.9 compares the best fitted fixed-effect model (Model 5), the fixed-effect

Table 7.5: The estimation (standard errors) of parameters for two-class non-parametric frailty model (i) without parameterizing η for CAV data

Transitions	Intercept	t	$bage$	$dage$
$1 \to 2$	$-3.876(0.532)$	$0.177(0.047)$	$0.002(0.007)$	$0.031(0.008)$
$1 \to 4$	$-6.190(0.751)$	0	$0.047(0.014)$	$0.023(0.009)$
$2 \to 3$	$-1.445(0.501)$	0	0	$-0.017(0.011)$
$2 \to 4$	$-1.97(0.941)$	0	0	$-0.037(0.038)$
$3 \to 4$	$-1.042(0.342)$	0	0	$-0.009(0.011)$
Frailty	$b_1 = 2.237(0.539)$	$b_2 = 0.447(0.539)$		.
	$\pi_1 = 0.618(0.108)$	$\pi_2 = 0.382(0.175)$		.

Table 7.6: The estimation (standard errors) of parameters for two-class non-parametric frailty model (ii) with the gender-specific frailty defined in transition $1 \to 2$ for CAV data

Transitions	Intercept	t	$bage$	$dage$
$1 \to 2$	$-4.618(0.618)$	$0.268(0.046)$	$0.001(0.009)$	$0.041(0.009)$
$1 \to 4$	$-6.312(0.774)$	0	$0.053(0.015)$	$0.018(0.010)$
$2 \to 3$	$-1.390(0.280)$	0	0	$-0.002(0.008)$
$2 \to 4$	$-1.315(1.061)$	0	0	$-0.068(0.052)$
$3 \to 4$	$-1.153(0.334)$	0	0	$-0.005(0.011)$
Frailty	$b_1 = 3.411(0.798)$	$b_2 = 0.293(0.069)$		.
	For female:			.
	$\pi_1 = 0.359(0.023)$	$\pi_2 = 0.641(0.023)$		.
	For male:			.
	$\pi_1 = 0.613(0.084)$	$\pi_2 = 0.387(0.084)$		.

model (iiii) (adding gender as the covariate in transition $1 \to 2$), two parametric frailty models and three non-parametric frailty models. In this table, the fixed-effect parts of the models 2-5 are the same as the 1st model (fixed-effect Model 5), the fixed-effect part of the 6th model (Model (iii)) is the same as the 7th model (Model (iiii)). The AIC value for the lognormal frailty model, one parameter gamma frailty model, two-class non-parametric model (i), (ii) and (iii) and fixed-effect model (iiii) are 3470.5, 3472.1, 3468.5, 3464.8, 3468.8 and 3469.3, respectively. The AIC value for the one parameter gamma model is larger than the lognormal model. This might be caused by using different methods of integration when

Table 7.7: The estimation (standard errors) of parameters for two-class non-parametric frailty model (iii) with the gender as a covariate and a gender-specific frailty defined in transition $1 \rightarrow 2$ for CAV data

Transitions	Intercept	t	*bage*	*dage*	*gender*
$1 \rightarrow 2$	$-4.466(0.705)$	$0.225(0.082)$	$0.002(0.008)$	$0.032(0.010)$	$-0.818(1.14)$
$1 \rightarrow 4$	$-5.512(0.621)$	0	$0.035(0.012)$	$0.024(0.009)$	0
$2 \rightarrow 3$	$-1.159(0.281)$	0	0	$-0.009(0.009)$	0
$2 \rightarrow 4$	$-1.887(1.195)$	0	0	$-0.051(0.050)$	0
$3 \rightarrow 4$	$-1.005(0.337)$	0	0	$-0.009(0.011)$	0
Frailty	$b_1 = 3.177(1.166)$	$b_2 = 0.315(0.116)$		.	
	For female:			.	
	$\pi_1 = 0.765(0.757)$	$\pi_2 = 0.235(0.757)$		.	
	For male:			.	
	$\pi_1 = 0.690(0.125)$	$\pi_2 = 0.310(0.125)$		.	

Table 7.8: The estimation (standard errors) of parameters for the fixed-effect model (iiii) with gender as a covariate in transition $1 \rightarrow 2$ for CAV data

Transitions	Intercept	t	*bage*	*dage*	*gender*
$1 \rightarrow 2$	$-3.317(0.338)$	$0.118(0.023)$	$0.001(0.007)$	$0.026(0.006)$	$-0.548(0.255)$
$1 \rightarrow 4$	$-6.238(0.752)$	0	$0.049(0.014)$	$0.023(0.009)$	0
$2 \rightarrow 3$	$-1.254(0.286)$	0	0	$-0.006(0.009)$	0
$2 \rightarrow 4$	$-1.847(0.977)$	0	0	$-0.044(0.038)$	0
$3 \rightarrow 4$	$-1.058(0.338)$	0	0	$-0.007(0.011)$	0

maximising the likelihood functions: we use Gauss-Hermite quadrature for log-normal model and use the trapezoidal rule for one parameter gamma model. Different methods of integration and their usages of frailty models have been discussed in Section 4.3.1. The AIC value for both two-class non-parametric models are better than the fixed-effect model, and the two-class non-parametric model (ii) is the best among these four frailty models. Model (iii) has a AIC value close to that of Model (i), but estimation of the effect of gender and frailty in Model (iii) is problematic. Model (iiii) is fitted as a fixed-effect model, in order to compare with Model (ii) and Model (iii), we can see that the AIC value of Model (iiii) is larger than all of the non-parametric models, but lower than the fixed-effect

Table 7.9: The $-2log(\hat{L})$ and AIC values for fixed-effect Model 5 (t in $1 \rightarrow 2$ *bage* in $1 \rightarrow 2, 1 \rightarrow 4$ *dage* in all transitions), lognormal frailty model, one parameter gamma frailty model, two classes non-parametric frailty model (i), (ii), (iii) and fixed-effect model (iiii) with gender as a covariate

Model	Number of parameters	$-2log(\hat{L})$	AIC
1. Fixed-effect Model 5	13	3446.7	3472.7
2. Lognormal frailty model	14	3442.5	3470.5
3. One parameter gamma frailty model	14	3444.1	3472.1
4. Two classes non-parametric frailty model (i)	15	3438.5	3468.5
5. Two classes non-parametric frailty model (ii)	16	3432.8	3464.8
6. Two classes non-parametric frailty model (iii)	17	3434.8	3468.8
7. Fixed-effect model (iiii) with gender as a covariate	14	3441.3	3469.3

Model 5. Therefore, it is worthwhile to consider gender as the covariate for the fixed-effect model. Likewise it is also good to define gender as the covariate in the non-parametric parameterizing frailty model, if gender is not included as the fixed-effect covariate and we would like to explore the difference of frailties for different genders.

More comparison for movers ($b > 1$) and stayers ($b < 1$) for non-parametric frailty models can be illustrated by transition probabilities. They are the probabilities for each transition during a certain time interval. In the application, transition probabilities can be presented in a 4×4 matrix, where rows represent current states and columns represent the next states. Conditional on the mean of baseline age 47.1 and donor's age 30.6, transition probabilities for movers ($b_1 = 3.411$) and stayers ($b_2 = 0.293$) in Model (ii) in 2 years after transplant are

$$P(t|b_1) = \begin{pmatrix} 0.688 & 0.188 & 0.043 & 0.081 \\ 0 & 0.586 & 0.272 & 0.142 \\ 0 & 0 & 0.578 & 0.422 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$P(t|b_2) = \begin{pmatrix} 0.902 & 0.019 & 0.004 & 0.075 \\ 0 & 0.586 & 0.272 & 0.142 \\ 0 & 0 & 0.578 & 0.422 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where $t = 2$, hazards are fixed midway the interval. It is easy to see the difference in transition probabilities from the matrix above. For example, for individuals who get the transplant at age 47.1 with the donor at age 30.6, the probabilities of staying in state 1 (healthy) are 68.8% (movers) versus 90.2% (stayers).

7.2 Origins of Variance in the Oldest-Old data

7.2.1 Fitting models

In this study, we fit seven models to the OCTO-Twin data using different effects and different frailty distributions. The time-scale of OCTO-Twin data is age, and individuals enter this study at different ages. We assume that individuals are independent in the OCTO-Twin data and do not take the genetic dependency of twins into account in the present thesis.

In the CAV data, we have investigated the difference of fixed-effect models with different effects of covariates and compare them with the univariate frailty models. In this section about OCTO-Twin data, we would like to pay more attention to frailty models, especially to the bivariate frailty models. Therefore, we only fit two fixed-effect models as the standard for comparison. We start with the most simple model: intercepts-only model (Model I). Next, we add age as a covariate to fit hazard models: Model II - VII. We would like to summarize models first, and then specify them.

- Model I: Intercepts-only model.

- Model II: Fixed-effect model with age in restricted transitions as a covariate.
- Model III: Univariate lognormal distributed frailty model. Frailty is defined for transition $1 \rightarrow 2$
- Model IV: Univariate one parameter gamma distributed frailty model. Frailty is defined for transition $1 \rightarrow 2$
- Model V: Bivariate lognormal distributed frailty model. Two frailties are correlated defined for transition $1 \rightarrow 2$ and $1 \rightarrow 4$.
- Model VI: Bivariate gamma distributed frailty model. Two frailties are independently defined for transition $1 \rightarrow 2$ and $1 \rightarrow 4$.
- Model VII: Bivariate gamma distributed frailty model. Two frailties are correlated defined for transition $1 \rightarrow 2$ and $1 \rightarrow 4$.

Model II - VII share the same fixed-effect terms: define age as the covariate with restricted parameters $\beta_{age.rs}$ for transition $r \rightarrow s$: $\beta_{age.14} = \beta_{age.24} = \beta_{age.34} = \beta_{age.r}$, $\beta_{age.21} = 0$. Specifically, Model II is a fixed-effect model, which is a standard for the comparison of fixed-effect models and frailty models. Model III and IV are univariate frailty models with two common frailty distributions that we discussed in Chapter 4. The frailty is defined for transition $1 \rightarrow 2$, because we focus the investigation on people moving from the healthy state (state 1) to the mild cognitive impairment (state 2). Model V is a bivariate lognormal-distributed frailty model, this is one of the bivariate frailty model that introduced in Chapter 5. Both Model VI and Model VII are bivariate gamma frailty models. The difference is that two frailties are independent in the former model, but correlated in the latter. Fitting these two models is good to investigate whether the correlated gamma frailty model will be better (lower AIC value) than the independent one. We add frailties in transitions $1 \rightarrow 2$ and $1 \rightarrow 4$ for all these three bivariate models (Model V-Model VII), because we are interested in the correlation between hazards of people moving from the healthy state (state 1) to the mild cognitive impairment (state 2) and from the healthy state (state 1) to death (state 4).

Next, we specify the model equations. The event time t of Model I follows an exponential distribution (i.e. $T \sim \exp(\lambda)$), where the transition-specific hazards can be represented as a constant, i.e. $h(t) = \lambda$, where $\lambda > 0$. This hazard model during the transition $r \to s$ is given by

$$h_{rs}(t) = \exp(\beta_{rs.0}), \tag{7.2.1}$$

where t is the time during follow-up.

The event time t of Model II - Model VII follows a Gompertz distribution (i.e. $T \sim (\lambda, \xi)$), where the transition-specific hazards can be represented as $h(t) = \lambda \exp(\xi t)$ for $t \geq 0$, where $t \geq 0, \lambda > 0$. Table 7.10 shows the hazard functions for Model VII as an example.

Similar to the CAV data, we use Akaike's information criterion (AIC) to compare model performance. In Table 7.11, it is clear that all the models that include restricted age as covariates have lower AIC than the intercepts-only model. Univariate and bivariate lognormal distributed frailty models have similar AIC values, that are a bit higher than the fixed-effect model (Model II). For gamma-distributed frailty models, Model VII has the lowest AIC value, which is 5333.8. Although the difference between the two lowest AIC values (Model IV and Model VII) is quite minimal, there is a small improvement for the bivariate correlated gamma frailty model (Model VII).

Given the AIC values shown in Table 7.11, Model VII is selected as the best-

Table 7.10: For OCTO-Twin data: The hazard functions for Model VII where frailties follow a bivariate gamma distribution.

Transition	Hazard Function
$1 \to 2$	$h_{12}(t\|b_{12.i}, \boldsymbol{x}) = \exp(\beta_{0.12} + \beta_{t.12}t)b_{12.i}$
$1 \to 4$	$h_{14}(t\|b_{14.i}, \boldsymbol{x}) = \exp(\beta_{0.14} + \beta_{t.r}t)b_{14.i}$
$2 \to 1$	$h_{21}(t\|\boldsymbol{x}) = \exp(\beta_{0.21})$
$2 \to 3$	$h_{23}(t\|\boldsymbol{x}) = \exp(\beta_{0.23} + \beta_{t.23}t)$
$2 \to 4$	$h_{24}(t\|\boldsymbol{x}) = \exp(\beta_{0.24} + \beta_{t.r}t)$
$3 \to 4$	$h_{34}(t\|\boldsymbol{x}) = \exp(\beta_{0.34} + \beta_{t.r}t)$

Table 7.11: The model covariates, distributions of time in hazard, number of parameters, $-2\times$ maximum loglikelihood function and Akaike's information criterion of seven models for the OCTO-Twin data

Model	Distribution of time in hazard	Distribution of the frailty	Number of parameters	$-2log(\hat{L})$	AIC
I	Exponential	No frailty	6	5456.7	5468.7
II	Gompertz	No frailty	9	5349.0	5367.0
III	Gompertz	Univariate lognormal frailty	10	5352.0	5372.0
IV	Gompertz	Univariate gamma frailty	10	5316.5	5336.5
V	Gompertz	Bivariate correlated lognormal frailty	12	5348.3	5372.3
VI	Gompertz	Bivariate independent gamma frailty	10	5356.8	5376.8
VII	Gompertz	Bivariate correlated gamma frailty	11	5311.8	5333.8

fitting model for the OCTO-Twin data. Therefore, we will mainly focus on this model in the following sections.

7.2.2 Interpretation

Table 7.12 shows the parameter estimation and standard errors for Model VII. The unrestricted parameters $\hat{\alpha}_0^*$ and $\hat{\alpha}_1^*$ are mentioned in Equation (5.2.6), $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are the parameters for bivariate gamma distribution in Equation (5.1.3). Standard errors of $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are derived by using the delta method. Bivariate frailties B_1 and $B_2 \sim Gamma(10.362, 10.362)$, and the correlation coefficient is $\hat{\rho} = 0.499$. Figure 7.2 displays the density of bivariate gamma frailties B_1 and B_2. The graphs illustrate the density from two directions. The left one shows that the mean of

Table 7.12: The estimation (standard errors) of parameters for Model VII for OCTO-Twin data

transitions	intercept	t
$1 \to 2$	-2.369 (0.146)	0.137 (0.021)
$1 \to 4$	-3.031 (0.127)	0.070 (0.010)
$2 \to 1$	-1.432 (0.138)	0
$2 \to 3$	-1.259 (0.142)	0.072 (0.020)
$2 \to 4$	-3.210 (0.351)	0.070 (0.010)
$3 \to 4$	-1.781 (0.105)	0.070 (0.010)
$\hat{\alpha}_0^* = 1.642$ (0.055)	$\hat{\alpha}_1^* = 1.648$ (0.039)	
$\hat{\alpha}_0 = 5.165$ (0.284)	$\hat{\alpha}_1 = 5.197$ (0.203)	

frailty equals 1, which is assumed in Equation (5.1.3). The right one shows the symmetry of the density because B_1 and B_2 follow the same gamma distribution.

Most of the estimates are as expected. For instance, the hazard of deterioration of cognitive impairment increases with increasing age, i.e. $\hat{\beta}_{t.12}$ and $\hat{\beta}_{t.23} > 0$. Concerning the onset of cognitive impairment, the hazard ratio of people at age $t+1$ versus people at age t is $B_1 \exp(\hat{\beta}_{t.12})$. This hazard ratio is 1.147 when assuming $B_1 = E(B_1) = 1$. Furthermore, all the standard errors of estimates are relatively small. The estimation of the fixed-effect model (Model II) is slightly different from the frailty model (Models III-VII). Details of these estimations are not shown, see Section 7.1.2 for similar comparison and discussion of this aspect for CAV data.

Figure 7.3 shows the transition hazards for individuals. The graphs are helpful to understand the differences between hazards for the estimated model. It clearly illustrates that the hazards for moving forward between living state ($1 \to 2$ and $2 \to 3$) are higher than those moving to death ($1 \to 4$, $2 \to 4$ and $3 \to 4$). Given a simple example: assume there is an individual who is observed in state 1, then she or he is more likely to move to state 2 rather than death directly. For different frailties in the top two graphs, it is obvious that a larger frailty is associated with a higher risk of moving. For example, there are two individuals j and k both observed in state 1 but with different frailties: $b_{1j} = 1.5$, $b_{1k} = 0.5$. Then,

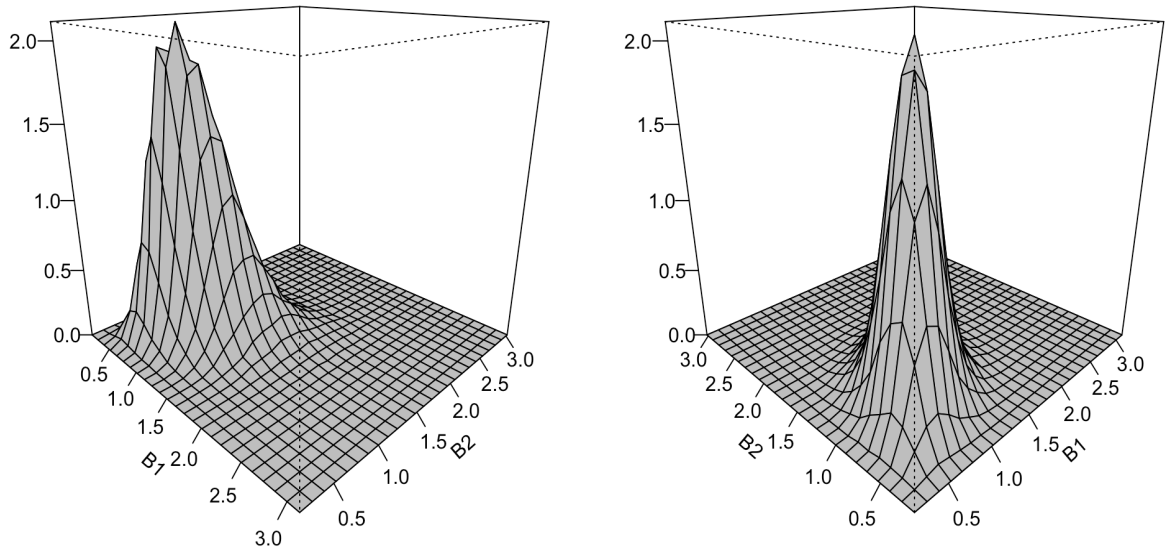

Figure 7.2: The fitted density of bivariate gamma frailty in Model VII for OCTO-Twin data in two viewing directions.

individual j will be more likely to move to state 2, compared with individual k who is more likely to stay in state 1. Therefore, we can call people like j and k movers and stayers, respectively.

More detailed interpretation and comparison can be illustrated by the transition probabilities. These are the probabilities for each transition during a certain time interval. In the application for OCTO-Twin data, transition probabilities can be presented in a 4×4 matrix, where rows represent current states and columns represent the next states. Consider individuals who are at age 85 and 90, two-year transition probabilities for them as movers ($b_1 = b_2 = 1.5$) and stayers ($b_1 = b_2 = 0.5$) in Model VII are given by

$$P(\Delta t = 2 | b_1 = b_2 = 1.5, age = 85) = \begin{pmatrix} 0.518 & 0.200 & 0.097 & 0.184 \\ 0.172 & 0.287 & 0.342 & 0.199 \\ 0 & 0 & 0.620 & 0.380 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

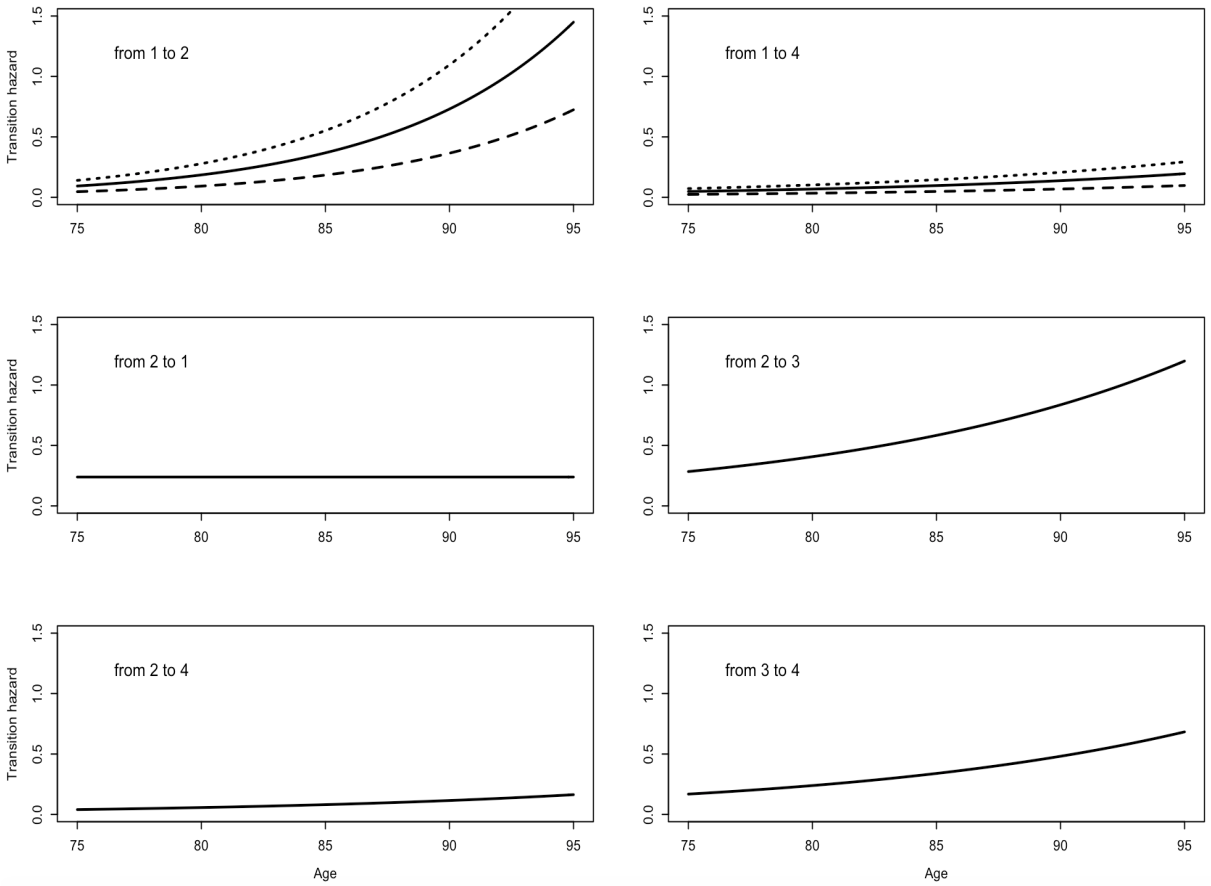

Figure 7.3: Six transition hazards in Model VII for OCTO-Twin data. Fixed effect hazard (solid line) for all transitions, frailty b=0.5 (dashed line) and b=1.5 (dotted line) for transition 1$\rightarrow$ 2 and 1 $\rightarrow$ 4.

$$P(\Delta t = 2 | b_1 = b_2 = 0.5, age = 85) = \begin{pmatrix} 0.800 & 0.087 & 0.038 & 0.075 \\ 0.223 & 0.262 & 0.334 & 0.181 \\ 0 & 0 & 0.620 & 0.380 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$P(\Delta t = 2 | b_1 = b_2 = 1.5, age = 95) = \begin{pmatrix} 0.307 & 0.244 & 0.189 & 0.260 \\ 0.106 & 0.216 & 0.381 & 0.297 \\ 0 & 0 & 0.507 & 0.493 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$P(\Delta t = 2 | b_1 = b_2 = 0.5, age = 95) = \begin{pmatrix} 0.665 & 0.131 & 0.085 & 0.119 \\ 0.170 & 0.188 & 0.366 & 0.276 \\ 0 & 0 & 0.507 & 0.493 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

We define the time of hazard as the start time of the interval, and the P matrices are calculated by using the piecewise constant approximation introduced in Section 3. The matrices above clearly show the transition probabilities for people with different frailties and ages. For example, probabilities of people at age 85 staying in state 1 are 51.8% (movers) versus 80.0% (stayers). Probabilities for people as movers staying in state 2 are 28.7% (age 85) and 21.6% (age 95). Overall, people with lower frailties are more likely to stay rather than move, older people are more likely to move than younger.

For visualization of transition probabilities, Figure 7.4 and 7.5 shows the change of two-year transition probabilities for people during age 80 to 110 for transition $1 \to 2$ and $1 \to 4$, respectively. From Figure 7.4, we can see that both movers and stayers have marked rises in the first 10-20 years, and then have dropped. Although they have similar trends, movers (peaked at 0.25 at age 90) tend to have higher transition probabilities than stayers (peaked at 0.17 at age 97) during age 80 to 105. In contrast to transition $1 \to 2$, the graph of $1 \to 4$ shows very different trends. Both movers and stayers in Figure 7.5 have steady increases over age 80 to 110.

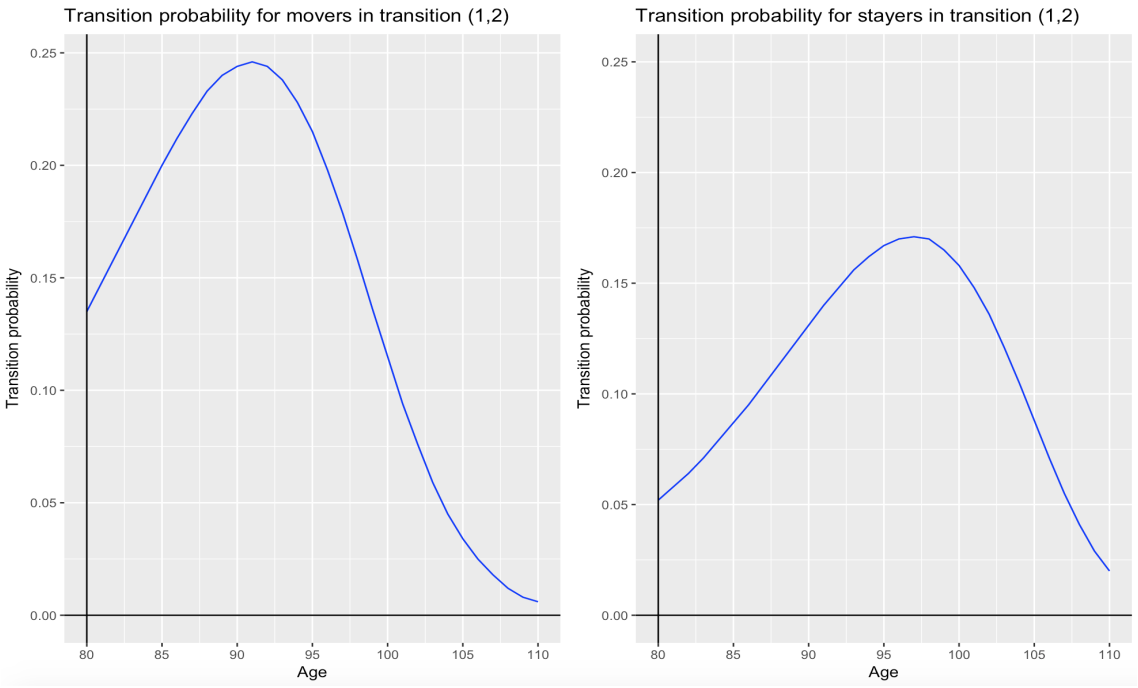

Figure 7.4: The two-year transition probabilities for transition $1 \to 2$ for movers ($b_1 = b_2 = 1.5$) and stayers ($b_1 = b_2 = 0.5$) during age 80 to 110, respectively.

7.2.3 Goodness of fit

For assessing the goodness of fit for frailty models, the estimation of individual frailties need to be derived from estimation results. The estimates of frailties are sometimes called predictors in survival analysis, see Laird (1982). In this study, we apply an empirical Bayes approach for this estimation, which is discussed in O'Keeffe et al. (2018) as well. It follows that

$$\begin{aligned}
\hat{\boldsymbol{b}}_{\boldsymbol{i}} &= E(\boldsymbol{b}_{\boldsymbol{i}}|\hat{\boldsymbol{\theta}}, i, \boldsymbol{y}, \boldsymbol{x}) \\
&= \int_{\Omega_{b_i}} \boldsymbol{b}_{\boldsymbol{i}} f(\boldsymbol{b}_{\boldsymbol{i}}|\hat{\boldsymbol{\theta}}, i, \boldsymbol{y}, \boldsymbol{x}) d\boldsymbol{b}_{\boldsymbol{i}} \\
&= \frac{\int_{\Omega_{b_i}} \boldsymbol{b}_{\boldsymbol{i}} L_i(\hat{\boldsymbol{\theta}}, i, \boldsymbol{y}, \boldsymbol{x}|\boldsymbol{b}_{\boldsymbol{i}}) f(\boldsymbol{b}_{\boldsymbol{i}}) d\boldsymbol{b}_{\boldsymbol{i}}}{L_i(\hat{\boldsymbol{\theta}}, i, \boldsymbol{y}, \boldsymbol{x})} \\
&= \frac{\int_{\Omega_{b_i}} \boldsymbol{b}_{\boldsymbol{i}} L_i(\hat{\boldsymbol{\theta}}, i, \boldsymbol{y}, \boldsymbol{x}|\boldsymbol{b}_{\boldsymbol{i}}) f(\boldsymbol{b}_{\boldsymbol{i}}) d\boldsymbol{b}_{\boldsymbol{i}}}{\int_{\Omega_{b_i}} L_i(\hat{\boldsymbol{\theta}}, i, \boldsymbol{y}, \boldsymbol{x}|\boldsymbol{b}_{\boldsymbol{i}}) f(\boldsymbol{b}_{\boldsymbol{i}}) d\boldsymbol{b}_{\boldsymbol{i}}}.
\end{aligned} \tag{7.2.2}$$

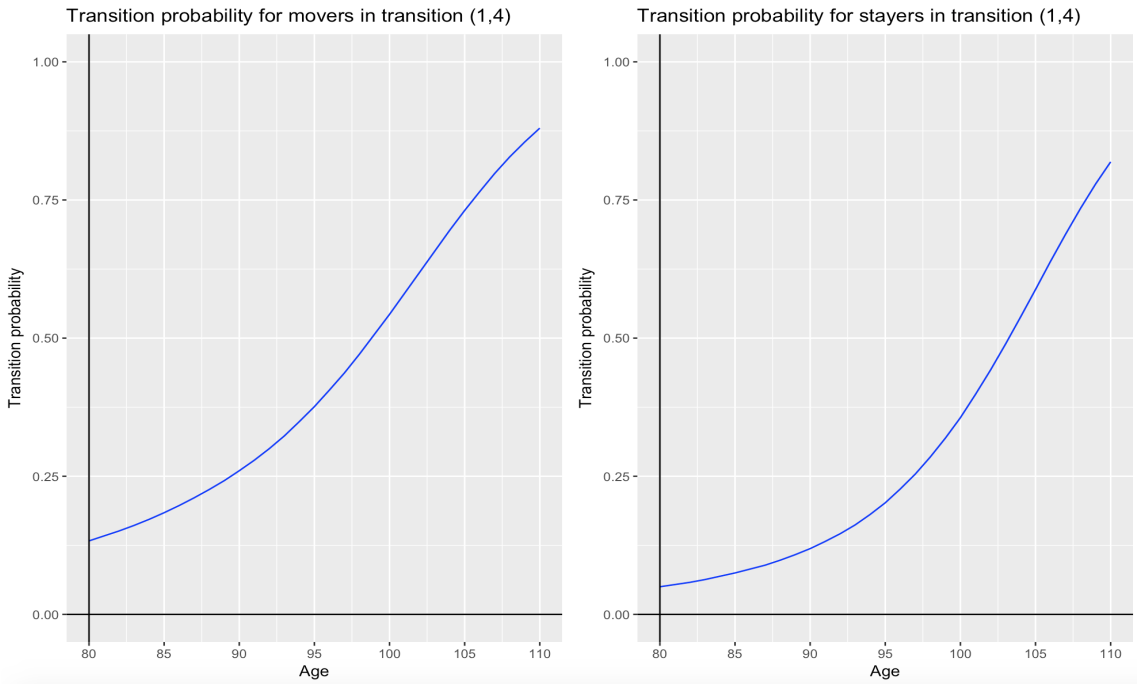

Figure 7.5: The two-year transition probabilities for transition $1 \rightarrow 4$ for movers ($b_1 = b_2 = 1.5$) and stayers ($b_1 = b_2 = 0.5$) during age 80 to 110, respectively.

Here $\boldsymbol{b_i}$ is the n-dimensional frailty, where $n = 2$ for Model VII. $\hat{\boldsymbol{\theta}}$ is the fixed-effect estimates for the corresponding parameter estimations of the model. $L_i(\hat{\boldsymbol{\theta}}, \boldsymbol{y}, \boldsymbol{x})$ represents the contribution for the model likelihood function of individual i.

The assessment of the goodness of fit for models in OCTO-Twin data is restricted by the interval censoring of the transitions between living states. However, as the death state is known at an exact time, it is good to compare estimated model-based survival with Kaplan-Meier curves. For Model VII, a baseline-state specific comparison between model-based survival and Kaplan-Meier estimates of survivor functions over time is shown in Figure 7.6. Individual survival curves start from years of baseline because each individual has different age at baseline. Sample sizes (N) are different for each state at baseline. There is some misfit for survival from state 3 and the Kaplan-Meier curve, but overall, the fit of Model VII is quite good. Possible reasons for this misfit: firstly, individuals in state 3 at baseline may be less frail when compare to individuals observed in state 3 during

the follow-up. This can be checked by adding covariate information to the model for transition-specific hazard $h_{34}(t)$ only, this was done in Van den Hout (2016) Section 4.10. Secondly, we define frailties in transition $1 \rightarrow 2$ and $1 \rightarrow 4$, but no frailty is related to state 3. This can be investigated by adding a frailty that is related to state 3 specifically. Yet another option, would be to consider different baseline hazard functions, such as the Weibull distribution. These options would be of interest in thorough data analysis, but within the current context, we focus on the methodology of the frailty model, and would not try these options since it does not give us an extra scope for our interests.

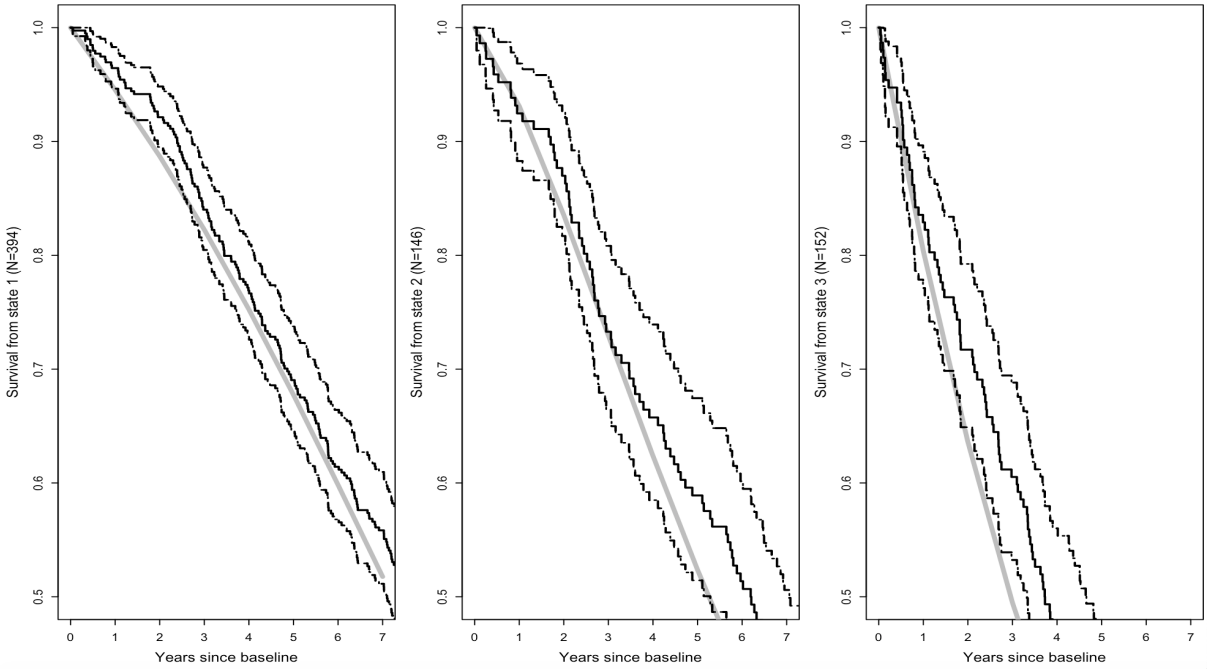

Figure 7.6: Comparison between mean of model-based survival (grey line) and Kaplan-Meier curves (black line) for Model VII for OCTO-Twin data. State-specific graphs for the three living states at baseline (with sample size 394, 146 and 152 for state 1, 2 and 3, respectively). Black dashed lines denotes the 95% confidence boundaries.

The empirical Bayes estimates can also be applied in predictions. For observed individuals in OCTO-Twin data and new individuals with observed state trajectories, estimates of individual-frailties can be illustrated by Equation (7.2.2). Therefore, predicted transition probabilities can be derived. Another way for pre-

diction is to calculate transition probabilities conditional on a fixed frailty, such as $b_1 = b_2 = 1$.

7.3 Computation in the R software

All the results including tables and graphs we have discussed in this chapter are produced using the R software. Here we would introduce some usages of packages and functions when we analysed CAV and OCTO-Twin data. For more details of the R code, please see Appendix A.

For all the fixed-effect models, we can use the package `msm` directly. It is quite easy and fast. For example, we have fitted four different fixed-effect models for CAV data. For fitting these models by using `msm` package, we choose the method of `"BFGS"` introduced in Chapter 3, and the max iteration *"maxit"* is 1000 for avoiding the failure of convergence.

However, `msm` can not be used for frailty models, since we need to integrate out the frailty B when we maximising the likelihood function, see Chapter 5 for computation details. Therefore, frailty models in this study are implemented by user-written code. We use `"Nelder-Mead"` in the function *optim* when maximising the likelihood function for all the frailty models, and use Gauss-Hermite and trapezoidal rule for quadrature for lognormal frailty models and gamma frailty models, respectively. For the non-parametric frailty model, there is no integration in the likelihood function.

Details of likelihood functions and quadrature for user-written code are shown in Appendix A.4, A.5 and A.6.

7.4 Simulation

In the previous section, we have fitted several fixed-effect models and frailty models for the OCTO-Twin data. The bivariate gamma-distributed frailty model is the best one among them, and has an overall good assessment of the goodness-

of-fit. This model is our main contribution. Therefore, we would like to perform a simulation study to investigate the performance of the bivariate gamma-distributed frailty multi-state model (BGF model in what follows).

We try to mimic the process of the OCTO-Twin data: a four-state process with backward transition $2 \rightarrow 1$. The parameter estimates for Model VII (shown in Table 7.12) are set as the true values for the simulation. Note that all the true values are round to two decimal places. The number of replications is 20 in the simulation for each scenario. The model computation is quite time-consuming, so we only do a small simulation study.

7.4.1 Sample size and interval width

First, we would like to investigate the BGF model performance for data with different sample sizes and follow-up times (censoring). We use scenarios with respect to the follow-up times that are similar to the studies in the applications for CAV and OCTO-Twin data, where the follow-up times are usually 1-3 years. Simulations are designed as follows:

- We consider four different scenarios:
  (1) Sample size =100, time length=15 years, people are observed every 3 years.
  (2) Sample size =200, time length=15 years, people are observed every 3 years.
  (3) Sample size =100, time length=9 years, people are observed every 1.5 years.
  (4) Sample size =200, time length=9 years, people are observed every 1.5 years.

- All the individuals enter the study at the same age (80 years old) in state 1. The delayed entry will be investigated in Section 7.4.2.

- The model we fit for each scenario is the same as Model VII in Section 7.2,

where age is the covariate with restricted parameters $\beta_{age.rs}$ for transition $r \to s$: $\beta_{age.14} = \beta_{age.24} = \beta_{age.34} = \beta_{age.r}$, $\beta_{age.21} = 0$ and frailties are defined for transitions $1 \to 2$ and $1 \to 4$.

- Frailty terms are generated from the bivariate gamma distribution and randomly allocated to each individual.

Sample sizes in all the scenarios are relatively small because of computational demands. Also, once it works on the small sample size, we believe that it will work on the bigger sample size as well. Table 7.13 shows the simulation results for the four scenarios. The SD and R.bias are the empirical stand deviation and the relative bias of parameter estimates. Considering the number of replications is 20, the bias of parameter estimates are relatively small in all scenarios. Comparing data with different sample sizes, we can see that some of the estimates in Scenarios (2) and (4) are a little bit better than (1) and (3), but some of them are worse. Comparing data with different time intervals, we can see that almost all the estimates in Scenarios (3) and (4) are better than (1) and (2). In particular, the means of frailty estimates α_0^* and α_1^* are all around 1.53 and 1.54 for all these four scenarios.

The result indicates that the estimation of frailty parameters might be biased in our model. The estimates are exactly same numbers because we round them, but there are small differences across the estimation in each scenarios. Moreover, if we use the mean estimates of α_0^* and α_1^* to compute the correlation coefficient of the two frailties, we get $\rho_e = \exp(1.53)/(\exp(1.53) + \exp(1.54)) \approx 0.498$, which is the same as the true value of the correlation coefficient, see Section 3.3. Figure 7.7 shows the probability density of the frailty B_1 or B_2 with true values of parameters and mean of estimates of parameters in Scenario (1). These two curves are similar. Therefore, there seems to be slight underestimations of two frailties, but they do not affect the value of the correlation coefficient of two transitions, and the picture shows that the difference is negligible. We will see the similar pattern in other simulations in the following subsections.

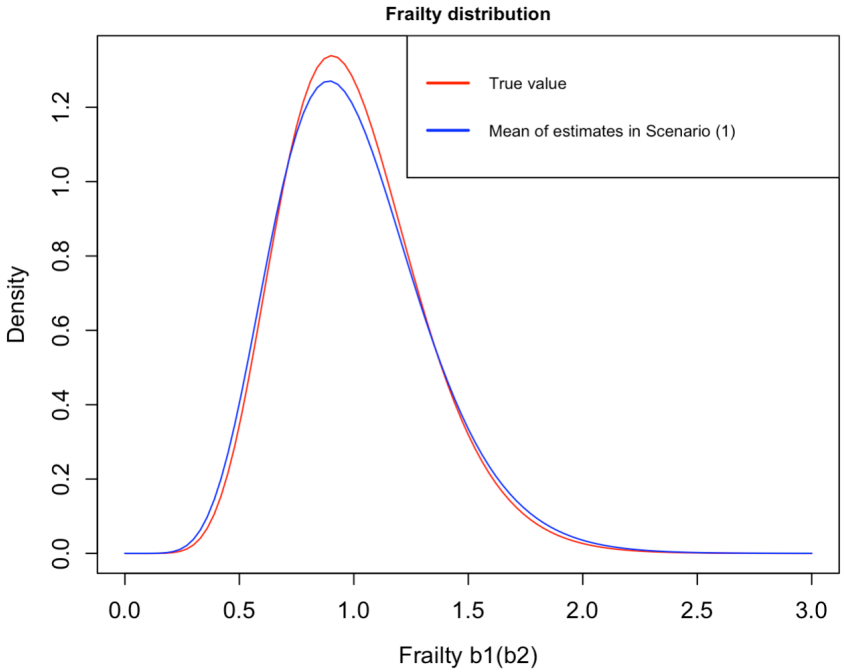

Figure 7.7: The probability density function of frailty $B_1(B_2)$ with different values of parameters. The red curve is the density of frailty $B_1(B_2) \sim Gamma(3.29, 3.29)$, where $\alpha_0^* = 1.64$ and $\alpha_0^* = 1.65$. The blue curve is the density of frailty $B_1(B_2) \sim Gamma(3.06, 3.06)$, where $\alpha_0^* = 1.53$, $\alpha_0^* = 1.53$. Frailties B_1 and B_2 has the same distribution.

In conclusion, the results are better when the time intervals are smaller, i.e. people are observed more often. The sample size does not affect the results substantially in this simulation.

7.4.2 Left truncation of frailty

Left truncation of individuals in the data is not explicitly modelled in our approach, but is accounted for by using the Markov assumption. Because of the assumption, only transition probabilities $P(Y_j = y_j | Y_{j-1} = Y_{j-1}, \boldsymbol{\theta}, \boldsymbol{x})$ are modelled. Nevertheless, the combination of left-truncation of frailty terms is an interesting aspect to be explored. Individuals entering the study at different ages may have different frailties. For example, an individual entering the study at age 80 in state 1 may be less frail than an individual entering the study at age 70 in the same state. Therefore, we would like to investigate the BGF model performance in this

Table 7.13: Parameter estimation in the Scenario (1)-(4)

Parameter	True	Scenario (1)			Scenario (2)			Scenario (3)			Scenario (4)		
	value	Mean (SD)	Bias	R.bias	Mean (SD)	Bias	R.bias	Mean (SD)	Bias	R.bias	Mean (SD)	Bias	R.bias
$\beta_{int.12}$	-2.37	-2.17(0.21)	0.20	0.08	-2.01(0.17)	0.36	0.15	-2.28(0.25)	0.09	0.04	-2.18(0.14)	0.19	0.08
$\beta_{int.14}$	-3.03	-2.90(0.19)	0.13	0.04	-2.92(0.15)	0.11	0.04	-2.93(0.19)	0.10	0.03	-2.95(0.26)	0.08	0.03
$\beta_{int.21}$	-1.43	-1.60(0.31)	-0.17	0.12	-1.36(0.3)	0.07	0.05	-1.70(0.17)	-0.27	0.19	-1.44(0.26)	-0.01	0.01
$\beta_{int.23}$	-1.26	-1.18(0.31)	0.08	0.06	-1.21(0.18)	0.05	0.04	-1.18(0.34)	0.08	0.06	-1.16(0.23)	0.10	0.08
$\beta_{int.24}$	-3.21	-3.44(0.57)	-0.23	0.07	-3.41(0.50)	-0.20	0.06	-3.35(0.41)	-0.14	0.04	3.36(0.42)	-0.15	0.05
$\beta_{int.34}$	-1.78	-1.68(0.30)	0.10	0.06	-1.53(0.25)	0.25	0.14	-1.76(0.37)	0.02	0.01	1.71(0.31)	0.07	0.04
$\beta_{age.12}$	0.14	0.14(0.04)	0.00	0	0.13(0.03)	0.01	0.07	0.14(0.04)	0.00	0	0.14(0.04)	0.00	0
$\beta_{age.r}$	0.07	0.07(0.02)	0.00	0	0.06(0.02)	0.01	0.14	0.06(0.02)	0.01	0.14	0.06(0.01)	0.01	0.14
$\beta_{age.23}$	0.07	0.07(0.03)	0.00	0	0.08(0.02)	-0.01	0.14	0.07(0.02)	0.00	0	0.07(0.02)	0.00	0
α_0^*	1.64	1.53(0.02)	0.11	0.07	1.53(0.03)	0.11	0.07	1.53(0.01)	0.11	0.07	1.53(0.01)	0.11	0.07
α_1^*	1.65	1.53(0.02)	0.11	0.06	1.54(0.01)	0.12	0.07	1.53(0.01)	0.11	0.06	1.54(0.01)	0.12	0.07

situation. Simulations are designed as follows:

- For all the scenarios: Sample size =100, time length=9 years, people are observed every 1.5 years.
- People are allocated in five groups equally, individuals in groups 1-5 enter the study at age 80, 82, 84, 86, 88, respectively.
- We consider two different scenarios:
  (5) People enter the study at different ages. Frailties are generated from the bivariate gamma distribution and randomly allocated to each individual.
  (6) People enter the study at different ages. Frailties are generated from the bivariate gamma distribution and randomly allocated to people in groups 1 and 2. For people in groups $3-5$, only frailties with value less than 0.5 are randomly allocated. The choice of 0.5 is a little bit arbitrary to some extent, but it means that people in groups $3-5$ are defined with low frailties, given the mean of the frailty distribution is 1.
- The model we fit for each scenario is the same as Model VII in Section 7.2.

Table 7.14 shows the simulation results for these two scenarios. We do not consider the left-truncation of frailties when generating the data in Scenario (5), and define lower frailties for individuals who enter the study at greater ages for data in Scenario (6). This design is in order to define a baseline age-specific frailty. The motivation is to investigate whether there is a left truncation due to the combination of age dependency and frailty. Thus, the comparison of these two scenarios can indicate whether the left-truncation of the frailty affects the parameter estimation.

The first two rows in Table 7.14 are for the intercepts for transitions $1 \rightarrow 2$ and $1 \rightarrow 4$. As we can see, the biases of estimates of $\beta_{int.12}$ and $\beta_{int.14}$ are -0.18 and -0.13 in Scenario (5), but they are -0.32 and 0.29 in Scenario (6), respectively. We can see the increases in biases for these two estimates from Scenario (5) to Scenario (6), and the differences for these estimates in Scenarios (5) and (6) are

Table 7.14: Parameter estimation in the Scenario (5) and (6)

Parameter	True	Scenario (5)			Scenario (6)		
	value	Mean(SD)	Bias	R.bias	Mean(SD)	Bias	R.bias
$\beta_{int.12}$	-2.37	-2.55(0.45)	-0.18	0.08	-2.69(0.35)	-0.32	0.13
$\beta_{int.14}$	-3.03	-3.16(0.26)	-0.13	0.04	-3.32(0.32)	0.29	0.09
$\beta_{int.21}$	-1.43	-1.43(0.36)	0.00	0	-1.50(0.34)	0.07	0.05
$\beta_{int.23}$	-1.26	-1.34(0.36)	-0.08	0.06	-1.16(0.47)	0.10	0.08
$\beta_{int.24}$	-3.21	-3.05(0.61)	0.16	0.05	-3.35(0.66)	-0.14	0.04
$\beta_{int.34}$	-1.78	-1.59(0.35)	0.19	0.11	-1.64(0.46)	0.14	0.08
$\beta_{age.12}$	0.14	0.12(0.04)	0.02	0.14	0.13(0.03)	0.01	0.07
$\beta_{age.r}$	0.07	0.06(0.01)	0.01	0.14	0.05(0.02)	0.02	0.29
$\beta_{age.23}$	0.07	0.07(0.01)	0.00	0	0.06(0.02)	0.01	0.14
α_0^*	1.64	1.53(0.01)	0.11	0.07	1.68(0.11)	-0.04	0.02
α_1^*	1.65	1.53(0.02)	0.12	0.07	1.54(0.14)	0.11	0.07

quite substantial. Since the bivariate frailties are defined in transitions $1 \to 2$ and $1 \to 4$ as well, our model may not be good at dealing with left truncation when the frailty depends on the age of entry. Entry time-related frailty as a topic for future study will be further discussed in Chapter 8.

Moreover, we would like to compare the estimates of Scenario (5) with Scenario (3). The only difference between these two scenarios is that people enter the study at the same ages in Scenario (3) and at different ages in Scenario (5). As we can see, some biases are smaller in Scenario (3), but some are smaller in Scenario (5). Considering that there are only 20 replications in this simulation, we infer that parameter estimation in Scenario (3) and Scenario (5) is similar. It indicates that the identifiability of the BGF model is not affected by the variation of observed entry times, i.e. people enter the study at different times. This is based on the assumption that there is not left truncation in the frailty distribution. However, if different times of observation at baseline induce the left truncation for frailty distributions, then the model may not be adequate.

7.4.3 Degree of heterogeneity and correlation of frailties

Next, we would like to investigate parameter estimation for data with a different degree of heterogeneity (i.e. different variance of frailties B_1 and B_2) and different correlation between frailties. These differences can be generated by different values of frailty parameters α_0^* and α_1^*. Note that $Var(B_1) = Var(B_2) = \frac{(\alpha_0+\alpha_1)}{(\alpha_0+\alpha_1)^2}$, and $Corr = \rho_B = \frac{\alpha_0}{\alpha_0+\alpha_1} > 0$, where $\alpha_0 = \exp(\alpha_0^*)$, $\alpha_1 = \exp(\alpha_1^*)$. Simulations are designed as follows:

- We consider four additional different scenarios (7)-(10). In Scenarios (7) and (8), we have the same variance (0.04), this small variance is similar to Scenarios (1)-(6). In Scenarios (9) and (10), we have the same correlation coefficient (0.5), and larger variances :
  (7) $\alpha_0^* = 1$, $\alpha_1^* = 3$, then $Var \approx 0.04$, $Corr \approx 0.12$.
  (8) $\alpha_0^* = 3$, $\alpha_1^* = 1$, then $Var \approx 0.04$, $Corr \approx 0.88$.
  (9) $\alpha_0^* = \alpha_1^* = 0.5$, then $Var \approx 0.3$, $Corr = 0.5$.
  (10) $\alpha_0^* = \alpha_1^* = 0$, then $Var = 0.5$, $Corr = 0.5$.

- For all the scenarios: Sample size =100, time length=9 years, people are observed every 1.5 years.

- All the individuals enter the study at the same age (80 years old) in state 1.

- The model we fit for each scenario is the same as Model VII in Section 7.2.

- Frailty terms are generated from the bivariate gamma distribution and randomly allocated to each individual.

Figure 7.8 shows the probability density of Scenarios (1)-(6), (7)-(8), (9) and (10), respectively. In Scenarios (1)-(6), $\alpha_0^* = 1.64$, $\alpha_1^* = 1.65$, $Var \approx 0.1$, $Corr \approx 0.498$. These values are similar to the results of Model VII of the OCTO-Twin data, so it can be seen as a good standard to compare the variance in Scenarios (7)-(10). As we can see, The top two graphs in Figure 7.8 have similar shapes: there is no

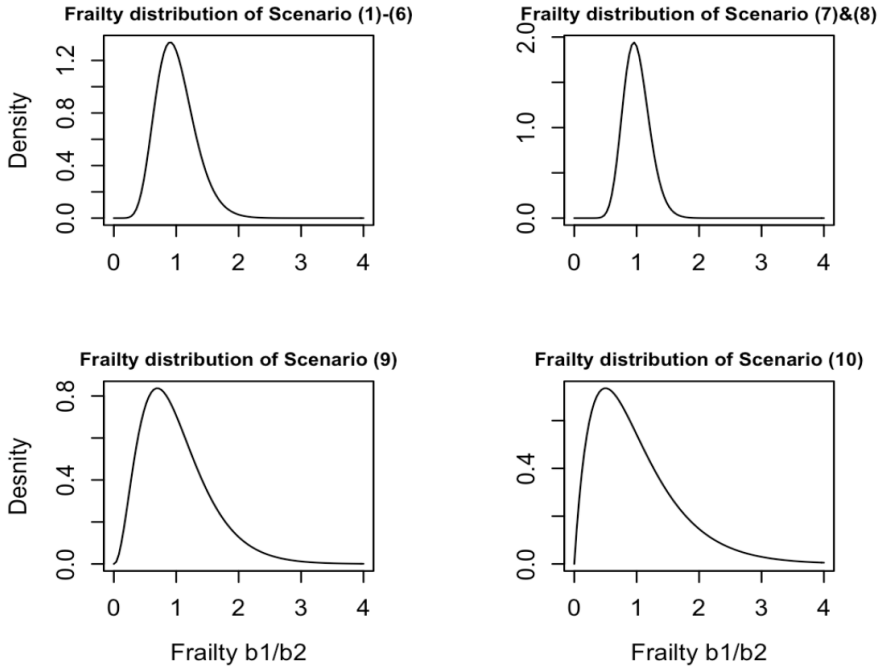

Figure 7.8: The probability density function of frailty $B_1(B_2)$ with true value in Scenarios (1)-(6), (7)-(8), (9) and (10). The distribution of B_1 is the same as it of B_2, the frailty distribution in Scenario (7) is the same it in Scenario (8).

obvious skewness, and they peak around 1. The most obvious difference between them is that the top right one is narrower, since it has a smaller variance. The shapes of the bottom two graphs are quite different from the first two graphs. They are both right skewed and the peak at values smaller than 1. The shapes of them are obviously wider than the top two graphs. Compared with variances in Scenarios (1)-(8), variances in Scenarios (9) and (10) are quite large, which should be enough to investigate the heterogeneity of the data.

Table 7.15 shows the simulation results for these four scenarios. Firstly, we would like to examine the β-related parameters, which are the fixed-effect parameters. Considering the small number of replications in our study, the biases of these parameter estimates are relatively small, and there are no big differences for each scenario. However, the estimates of frailty parameters α_0^* and α_1^* are not as good as for the fixed-effect parameters. There biases are quite large, especially for Scenarios (9) and (10). The variances of estimated frailties in Scenarios

Table 7.15: Parameter estimation in the Scenario (7)-(10)

Parameter	True value	Scenario (7) Mean(SD)	Bias	R.bias	Scenario (8) Mean(SD)	Bias	R.bias	Scenario (9) Mean(SD)	Bias	R.bias	Scenario (10) Mean(SD)	Bias	R.bias
$\beta_{int.12}$	-2.37	-2.20(0.19)	0.17	0.07	-2.50(0.24)	0.13	0.05	-2.14(0.22)	0.23	0.10	-2.25(0.15)	0.12	0.05
$\beta_{int.14}$	-3.03	-3.00(0.20)	0.03	0.01	-2.94(0.33)	0.09	0.03	-2.96(0.32)	0.07	0.02	-3.02(0.25)	0.01	0
$\beta_{int.21}$	-1.43	-1.45(0.28)	-0.02	0.01	-1.40(0.24)	0.03	0.02	-1.56(0.31)	-0.13	0.09	-1.59(0.31)	-0.16	0.11
$\beta_{int.23}$	-1.26	-1.20(0.35)	0.06	0.05	-1.30(0.23)	-0.04	0.03	-1.23(0.28)	0.03	0.02	-1.51(0.45)	-0.25	0.19
$\beta_{int.24}$	-3.21	-3.02(0.56)	0.19	0.06	-3.34(0.62)	-0.13	0.04	-3.41(0.54)	-0.20	0.06	3.03(0.52)	0.18	0.06
$\beta_{int.34}$	-1.78	-1.67(0.39)	0.11	0.06	-1.73(0.37)	0.05	0.03	-1.68(0.36)	0.10	0.06	1.79(0.24)	-0.01	0.01
$\beta_{age.12}$	0.14	0.16(0.04)	-0.02	0.14	0.13(0.06)	0.01	0.07	0.11(0.04)	0.03	0.21	0.11(0.05)	0.03	0.21
$\beta_{age.r}$	0.07	0.05(0.02)	0.02	0.29	0.06(0.01)	0.01	0.14	0.07(0.02)	0.00	0	0.06(0.03)	0.01	0.14
$\beta_{age.23}$	0.07	0.08(0.02)	-0.01	0.14	0.07(0.03)	0.00	0	0.06(0.01)	0.01	0.14	0.09(0.03)	-0.02	0.29
α_0^*	1/3/0.5/0	1.13(0.03)	-0.13	0.13	3.21(0.12)	-0.21	0.07	1.03(0.43)	-0.53	1.06	1.53(0.28)	-1.53	-
α_1^*	3/1/0.5/0	3.27(0.08)	-0.27	0.09	1.13(0.03)	-0.13	0.13	1.04(0.51)	-0.54	1.08	1.55(0.25)	-1.55	-

(7)-(10) are 0.03, 0.04, 0.20, 0.11, respectively. The estimated correlation coefficient between frailties in Scenarios (7)-(10) are 0.11, 0.89, 0.50, 0.50, respectively. Clearly, the frailty parameter estimation in Scenarios (7) and (8) are better than Scenarios (9) and (10). Moreover, it is interesting that although there are biases in the estimated frailty parameters, the estimated correlation coefficients are close to the true values. Meanwhile, the estimates of the fixed-effect parameters (β-related parameters) are relatively good as well. Therefore, it can be concluded that data with a smaller degree of heterogeneity (i.e. small variance of frailty) will lead better estimates, and data with a large degree of heterogeneity will still lead good estimates of fixed-effect parameters. Moreover, the biases of the estimates of frailty parameters would not affect the value of correlations between the hazards of two transitions, only the estimated frailty variance will be biased.

7.4.4 Incorrect model specification

This section investigates model misspecification. We would like to explore the BGF model with respect to missing covariates. There are many other situations of misspecification that can be explored as well, such as misspecifying the distribution of the hazard function, the frailty distribution. We did not investigate all the possible misspecifications, but we could use the similar approach to investigate them. For example, Gompertz versus Weibull distributions for the hazard function, lognormal versus gamma distributions for the frailty. These are good topics for future study. For investigating missing covariates, we add gender effect in the simulated process, where $\beta_{gender.14} = \beta_{gender.24} = \beta_{gender.34} \neq 0$, $\beta_{gender.12} = \beta_{gender.21} = \beta_{gender.23} = 0$. Simulations are designed as follows:

- We consider two different scenarios:
  (11) Add a gender effect in the simulated process, and fit the model without gender as the covariate.
  (12) Add a gender effect in the simulated process, and fit the model with gender as the covariate.

Table 7.16: Parameter estimation in the Scenario (11) and (12)

Parameter	True	Scenario (11)			Scenario (12)		
	value	Mean(SD)	Bias	R.bias	Mean(SD)	Bias	R.bias
$\beta_{int.12}$	-2.37	-2.31(0.17)	0.06	0.03	-2.20(0.14)	0.17	0.07
$\beta_{int.14}$	-3.03	-3.21(0.40)	-0.18	0.06	-3.10(0.14)	-0.07	0.02
$\beta_{int.21}$	-1.43	-1.51(0.24)	-0.08	0.06	-1.44(0.30)	-0.01	0.01
$\beta_{int.23}$	-1.26	-1.05(0.33)	0.21	0.17	-1.31(0.22)	-0.05	0.04
$\beta_{int.24}$	-3.21	-3.69(0.57)	-0.48	0.15	-3.43(0.42)	-0.22	0.07
$\beta_{int.34}$	-1.78	-2.00(0.41)	-0.22	0.12	-1.77(0.17)	0.01	0.01
$\beta_{age.12}$	0.14	0.16(0.04)	-0.02	0.14	0.13(0.03)	0.01	0.07
$\beta_{age.r}$	0.07	0.08(0.05)	-0.01	0.14	0.08(0.03)	-0.01	0.14
$\beta_{age.23}$	0.07	0.05(0.01)	0.02	0.29	0.08(0.01)	-0.01	0.14
β_{gender}	-0.35	-	-	-	-0.33(0.09)	0.02	0.06
α_0^*	1.64	1.53(0.03)	0.11	0.07	1.53(0.01)	0.11	0.07
α_1^*	1.65	1.53(0.04)	0.12	0.07	1.53(0.02)	0.12	0.07

- For all the scenarios: Sample size =100, time length=9 years, people are observed every 1.5 years.
- All the individuals enter the study at the same age (80 years old) in state 1.
- Frailty terms are generated from the bivariate gamma distribution and randomly allocated to each individual.

Table 7.16 shows the simulation results for these two scenarios. For Scenario (11), we can see that biases of estimates are relatively small, except for parameters $\beta_{int.14}$, $\beta_{int.24}$ and $\beta_{int.34}$. The biases of these three estimates are -0.18, -0.48 and -0.22, respectively. In contrast, results in Scenario (12) are good, all the biases are relatively small. The difference between Scenarios (11) and (12) is that gender is not defined as a covariate in the transitions $1 \to 4$, $2 \to 4$ and $3 \to 4$ in Scenario (11) but is defined in Scenario (12). The results indicate that the misspecification of the model by omitting covariates does affect the model performance and induces bias. We hope the frailty term in transition $1 \to 4$ would account for the misspecification to some extents. We see some hopefully evidence, since the bias in $\beta_{int.14}$ seems a little better than $\beta_{int.24}$ and $\beta_{int.34}$, but it is not conclusive. For the estimates of α_0^* and α_1^*, there are slight differences across these two frailty values

in Scenarios (11) and (12), where $\alpha_0^*(11) = 1.529$, $\alpha_1^*(11) = 1.530$, $\alpha_0^*(12) = 1.534$ and $\alpha_1^*(12) = 1.533$, respectively. It is not shown in the table since parameter estimates are round to two decimal places.

Chapter 8

Conclusion

In this thesis, we have discussed both fixed-effect multi-state models and frailty models, and their application to two datasets. For the fixed-effect model, we introduced the parametric hazard function, the likelihood function and model selection. In particular, we have contributed an analytic expression of the progressive four-state model. This closed-form expression is helpful for readers to understand the transition probability for each transition in the multi-state model. Moreover, we have compared the speed of calculation when using this analytic expression with the piecewise-constant approximation, which is a commonly-used method. The result shows that the former is a direct solution, which leads to a fast computation time.

For the frailty model, both the univariate frailty model and the bivariate frailty model are discussed. For the parametric univariate frailty model, we introduced two common distributions: the lognormal distribution and the gamma distribution. A novel method is presented when we introduced the univariate non-parametric frailty model. This method is used to define covariate-specific masses π_k, in order to investigate the probabilities of different individuals being movers (larger frailty values) or stayers (smaller frailty values).

For the bivariate frailty model, we discussed frailties that follow either a bivariate lognormal or a bivariate gamma distribution. The latter distribution can

be defined in more than one way, and we use Cheriyan and Ramabhadran's definition presented by Yashin et al. (1995) in the context of the correlated-frailty model. This is also the main contribution of this thesis, i.e. present an approach for including bivariate gamma-distributed frailties in the multi-state model for interval-censored data. Data analysis is challenging in this approach because of interval censoring, the multiplicity of healthy and unhealthy states and computational complexities.

The advantages of Cheriyan and Ramabhadran's definition of the bivariate gamma distribution: (1) there are only two parameters, and (2) the bivariate density function is constructed by three univariate gamma variables which is convenient when using a marginal likelihood function. Both advantages reduce the computational challenge of fitting frailty models. An important disadvantage of this approach is that the correlation of two frailties for an individual has to be positive due to the restricted range of parameters. Therefore, it is not possible to include frailties in transitions that are correlated negatively. For example, we did not define frailties in transition $2 \rightarrow 3$ and $2 \rightarrow 1$. Negatively correlated frailties are of course possible when using lognormal frailty distributions. In Section 7.2, we fit two bivariate gamma frailty models: one with independent gamma frailties (Model VI) and one with correlated gamma frailties (Model VII). Model VII has the lower AIC value. This approach is helpful for cases where it is hard to judge whether the correlation is positive and negative.

Left truncation of data for multi-state survival models is briefly discussed in our study, but not explicitly modelled in our approach, since it is accounted for by using the Markov assumption. Because of the assumption, only transition probabilities $P(Y_j = y_j | Y_{j-1} = Y_{j-1}, \boldsymbol{\theta}, \boldsymbol{x})$ are modelled. An alternative way to deal with the left truncation is to extend the model by modelling the prevalence of individuals being in the first state, this method has been discussed in previous literature, e.g. Van Den Hout (2016). When left truncation is considered, it is important to determine the time scale. For example, for the CAV data introduced in Chapter 2, all the individuals are observed after the transplant. The time scale

in our study is years since transplant, so the left truncation can be ignored in the data analysis. However, if the time scale is chosen to be the age of individuals, then the left truncation should be considered. The context we discussed above is about left truncation for the fixed-effect multi-state model. There are many papers that discussed the left truncation for multi-state models, see Commenges (1999), Andersen and Keiding (2002) and Niessl et al. (2020).

In our study of the frailty model, left truncation of frailty could be considered as well. If individuals enter the study at different times, their frailty distributions may not be the same. For example, if the time scale of the study is age, an individual entering the study at age 80 in state 1 is probably less frail than an individual entering the study at age 70 in state 1. For the frailty model, there are some publications that discussed the left truncation for the univariate survival data and clustered survival data, such as Van Den Berg and Drepper (2016) and Eriksson et al. (2015). However, few publications discuss the left truncation of frailties in the multi-state model. Therefore, the combination of left truncation and frailty terms is an interesting aspect to be explored further. An idea of this future work is to define frailty values or distributions conditional on the left truncation times. For example, define a age-dependent frailty, where the frailty is constant for an individual over time but different with individuals having different baseline ages. Individuals who enter the study at younger ages have larger frailties and individuals who enter the study older have smaller frailties.

Another aspect for frailties that can be explored further is to relax the time-constant heterogeneity assumption and consider frailty models with a time-varying frailty term. The frailty in the current study is assumed to be time-constant, but the unobserved heterogeneity might change over time. There are several approaches that discussed the time-varying frailty model for survival data. Yau and McGilchrist (1998) and Munda et al. (2016) discussed the time-dependent frailty following lognormal distribution and gamma distribution, respectively. Manton and Yashin (1997) modelled the frailty with diffusion processes. Paik et al. (1994) and Wintrebert et al. (2004) discussed the piecewise constant frailty models.

A limitation of our study is with respect to using the OCTO-Twin data. Participants in this dataset are identical and same-sex fraternal twins, but we do not take this genetic dependency into account in the present study and assume they are independent. A reason for this assumption is that the motivation of fitting our bivariate frailty models is to investigate the correlation between two transition hazards, we do not aim for describing the dependence of individuals in the model, e.g. the twin structure. We admit that our model might not be appropriately optimised for the OCTO-Twin data. Unfortunately, we do not have the access to the complete data, which allows us to identify the twins. So we cannot fit any additional models to investigate the dependence of individuals at this moment. There are publications that are in a similar situation. For example, Robitaille et al. (2018) and Machado (2018) fit multi-state models for OCTO-Twin data, but both of them do not take into account the twin structure in the model. Although it is a shortcoming that not taking this twin structure into account, we think the effect of this on our data analysis is rather minor.

The simulation study illustrates data with different sample sizes, follow-up times, degree of heterogeneity and correlation of frailties. We found that the parameters in most of the designed scenarios are identified. The biases of estimates are relatively small. Specifically, results are better with shorter follow-up times and smaller degrees of heterogeneity. Moreover, we have checked the model performance for left-truncated data, and found that our bivariate gamma-distributed model may not be good at dealing with it. Therefore, we suggest using our model for data without left truncation.

For further study, there are some aspects that we have discussed above: firstly, we would like to take into account the left truncation of frailty in the model; secondly, defining a time-varying frailty in the model would also be good to explore; thirdly, we should get access to the complete OCTO-Twin data to investigate the twin structure. Next, we can explore the approach to describe the dependence of individuals in the univariate and bivariate frailty models.

Moreover, some other further aspects can be considered as well. Firstly, we

could extend the number of covariates in the model for the data analysis. In the current study, our aim is not to provide a comprehensive analysis of the data, but to use the data to illustrate how our frailty models work. So only several covariates in some transitions are defined in the data analysis, especially for the OCTO-Twin data. In further work, it is better to consider other covariates in the model, such as gender.

Secondly, fitting the proposed bivariate gamma-distributed frailty model in the R software is very time-consuming. It may take hours to days to run depending on the sample size and complexity of models. This is a problem, which will limit the scope of applications. In further study, it will be good to explore more time-saving approaches. For example, we could simplify the code, explore different functions and packages in the R software, or try to use other software. In the simulation study, there are only 20 replications for each of the scenarios because of the time-consuming computation as well. We should take more iterations in the further study, to see whether the model results will be more stable.

Thirdly, there is a restriction on the bivariate gamma-distributed frailty model we proposed: the correlation of two frailties must be positive. Therefore, it is good to explore different bivariate gamma distributions. A good source to start is Martins et al. (2019), who compared several bivariate frailty models for survival data. Fourthly, the multivariate frailty model is a good option to be explored as well. Frailties can be defined for more than two transitions, and then the resulting model can investigate the correlation between them.

In conclusion, this thesis discussed the fixed-effect model and frailty model for particular multi-state survival model and interval-censored data. The applications show that the proposed bivariate gamma distribution is an interesting and useful addition to the current multi-state model research. It is often reasonable to assume that the frailties for the two transitions out of state 1 are positively correlated, and the data analysis and simulation study shows that using a gamma distribution can lead to a good model fit.

Appendix A

Some code for the R software

The R software is a free environment for statistical computing and graphics (Team, R Core and others, 2013). All the computation, analysis and graphs in this study are produced by the R software. We will show some basic R code for the analysis mentioned in previous chapters for people who are interested in it.

A.1 Code for Chapter 3

Fixed-effect multi-state models can be implemented in the R software by using the package `msm`. Here is an example of Model II for OCTO-Twin data. Other fixed-effect models can be illustrated using similar code.

```
# Generator matrix Q:
q <- 0.1
Q <- rbind(c(0,q,0,q), c(q,0,q,q),c(0,0,0,q),c(0,0,0,0))
qnames <- c("q12","q14","q21","q23","q24","q34")

# Model formulation:
# Covariates:
covariates <- as.formula("~age")
constraint <- list(age=c(1,2,3,4,2,2))
```

```
fixedpars  <- c(9)
# Control:
method <- "BFGS"

# Modelling:
model <- msm(state~age, subject=case, data=dta, center=FALSE,
    qmatrix=Q, death=TRUE, covariates=covariates, constraint=
   constraint, fixedpars=fixedpars, method=method, control=
   list(fnscale=50000, maxit=100000))
```

For the code above, `Q` is the starting value, `dta` is the name of OCTO-Twin data and `case` is the subject id of individuals for OCTO-Twin data.

This `msm` function is involved in the package of same name. Please see Jackson (2011) for more usages and examples.

The output information can be derived by

```
cat("\nModel with covariates: ")
print(model$covariates)
cat("and constraints:\n")
print(model$constraint)
cat("and fixedpars:\n")
print(model$fixedpars)
cat("\n-2loglik =", model$minus2loglik,"\n")
conv <- model$opt$convergence
cat("Convergence code =", conv,"\n")
# paramter estimation and standard errors:
p    <- model$estimates
p.se <- sqrt(diag(model$covmat))
```

A.2 Code for Chapter 4

The likelihood function for frailty models can not be implemented using `msm` package, since the frailty parameters need to be integrated out. The following R code is the likelihood function of univariate lognormal distributed frailty model (Model III) for OCTO-Twin data.

```
loglikelihood<-function(p){
 # Model parameters:
 beta0<-p[1:6]
 beta1<-c(p[7],p[8],0,p[9],p[8],p[8])
 theta<-p[10]
 sigma<-exp(theta)
 loglik<-0
 # Data for individual i:
 for(i in 1:N){
      data.i<-dta.split[[i]]
      o<-data.i$state
      t<-data.i$age
      # Integrate out frailty parameter vi:
      integrand<-function(vi){
        contri<-1
        # Hazard generator matrix Q:
        for(j in 2:length(o)){
        Q<-matrix(0,4,4)
        Q[1,2]<-exp(beta0[1]+beta1[1]*t[j-1]+vi)
        Q[1,4]<-exp(beta0[2]+beta1[2]*t[j-1])
        Q[1,1]<--sum(Q[1,])
        Q[2,1]<-exp(beta0[3]+beta1[3]*t[j-1])
        Q[2,3]<-exp(beta0[4]+beta1[4]*t[j-1])
        Q[2,4]<-exp(beta0[5]+beta1[5]*t[j-1])
        Q[2,2]<--sum(Q[2,])
        Q[3,4]<-exp(beta0[6]+beta1[6]*t[j-1])
```

```
27          Q[3,3]<--sum(Q[3,])
28          P <- MatrixExp(mat=Q,t=t[j]-t[j-1])
29          if(o[j]!=dead){
30            contri<-contri*P[o[j-1],o[j]]
31           }
32          else{
33              contri<-contri*(P[o[j-1],1:3]%*%Q[1:3,dead])
34           }
35        }
36  pdfvi<-dnorm(vi,mean=0,sd=sigma)
37  contri*pdfvi
38 }
39 # Likelihood contribution:
40 contri<-myintegrate(integrand,mu=0, sigma=sigma)
41 loglik<-loglik+log(contri)
42 }
43 cat("-2*Loglik = ", -2*loglik,"\n") cat("sigma",round(sigma
     ,3),"\n")
44 -loglik
45 }
```

Piecewise-constant approximation is applied in this code. t is the time scale, and the generator matrix Q is constant for each time interval `t[j]-t[j-1]`, see line 28. The function `MatrixExp` is a method to calculate the transition probability matrix, see Section 3.4 for details. This function is included in `msm`. `max` is a R object includes the parameter estimation when the maximum of likelihood function is reached.

`myintegrate` in line 40 is a user-written function, it can be any proper integration methods as we discussed in Section 4.3. The following code shows two common used methods: Trapezoidal rule and Gauss-Hermite quadrature.

```r
# Trapezoidal rule:
myintegrate <- function(integrand,lower,upper,l){
nnodes <- (l-1)
grid <- seq(lower,upper,length.out=l)
int <- 0
h <- (upper-lower) / nnodes
int <- rep(NA,l)
 for ( i in 1:l ){
  int[i] <- integrand(grid[i])
 }
return(0.5*h*(int[1]+2*sum(int[2:(l-1)])+int[l]))
}

# Gauss-Hermite quadrature:
nnodes <- 5
cat("\nNumber of nodes in Gauss-Hermite quadrature = ",
   nnodes,"\n")
quad <- gauss.quad(nnodes,"hermite")
nodes <- quad$nodes
weights <- quad$weights
myintegrate <- function(integrand,mu,sigma){
  x <- sqrt(2)*sigma*nodes+mu
  f <- rep(NA,nnodes)
for(i in 1:nnodes){f[i] <- integrand(x[i])}
      (1/sqrt(pi))*weights%*%f
}
```

To derive the parameter estimation, we need to maximising the likelihood function. It can be implemented using the function `optim`. The default algorithm method of `optim` is Nelder-Mead, which is robust and often used in this study.

```r
# Starting value:
```

```
p0 <- c(rep(-1,6),rep(0.01,6),0)

# Maximaise:
max <- optim(par=p0, fn=loglikelihood, method="Nelder-Mead",
    control=list(maxit=5000), hessian=TRUE)

# Results:
p <- max$par
p.se <- sqrt(diag(solve(max$hessian)))
```

A.3 Code for Chapter 5

The likelihood function is more complicated for bivariate frailty model, since there are multiple frailty parameters need to be integrated out. The following shows the code for bivariate correlated gamma distributed frailty model (Model VII).

```
loglikelihood<-function(p){
# Model parameters:
beta0<-p[1:6]
beta1<-c(p[7],p[8],0,p[9],p[8],p[8])
# frailty parameters:
theta1<-p[10]
theta2<-p[11]}
alpha0<-exp(theta1)
alpha1<-exp(theta2)
# Boundary of frailty parameters when integrating out
UB0<-qgamma(0.99,shape=alpha0,rate=alpha0+alpha1)
UB1<-qgamma(0.99,shape=alpha1,rate=alpha0+alpha1)
# Contribution for each individuals:
loglik<-0
```

```
for(i in 1:N){
      # Data for individual i:
      data.i<-dta.split[[i]]
      o<-data.i$state
      t<-data.i$age
      # Integral funcion
      integrand<-function(x0,x1,x2){
            contri<-1
            # Generator matrix Q
            for(j in 2:length(o)){
            Q<-matrix(0,4,4)
            Q[1,2]<-exp(beta0[1]+beta1[1]*t[j-1])*(x0+x1)
            Q[1,4]<-exp(beta0[2]+beta1[2]*t[j-1])*(x0+x2)
            Q[1,1]<--sum(Q[1,])
            Q[2,1]<-exp(beta0[3]+beta1[3]*t[j-1])
            Q[2,3]<-exp(beta0[4]+beta1[4]*t[j-1])
            Q[2,4]<-exp(beta0[5]+beta1[5]*t[j-1])
            Q[2,2]<--sum(Q[2,])
            Q[3,4]<-exp(beta0[6]+beta1[6]*t[j-1])
            Q[3,3]<--sum(Q[3,])
            P <- MatrixExp(mat=Q,t=t[j]-t[j-1])
            if(o[j]!=dead){
                  contri<-contri*P[o[j-1],o[j]]
            }
            else{
                contri<-contri*(P[o[j-1],1:3]%*%Q[1:3,dead])
            }
      }
      # probability density function for each variable
      pdfx0<-dgamma(x0,shape=alpha0,rate=alpha0+alpha1)
      pdfx1<-dgamma(x1,shape=alpha1,rate=alpha0+alpha1)
```

```
46        pdfx2<-dgamma(x2,shape=alpha1,rate=alpha0+alpha1)
47        contri*pdfx0*pdfx1*pdfx2
48 }
49 contri<- myintegrate(function(x0){
           myintegrate(function(x1) {
50    myintegrate(function(x2) {
51        integrand(x0,x1,x2)},0,UB1,6)},0,UB1,6)},0,UB0,6)
52 # Update likeihood function
53        loglik<-loglik+log(contri)
54 }
55 cat("-2*Loglik = ", -2*loglik,"\n")
56 cat("alpha0",round(alpha0,3),"\n")
57 cat("alpha1",round(alpha1,3),"\n")
58 cat("p",round(beta0,3),"\n")
59 cat("rou",round(rou,3),"\n")
60 -loglik
61 }
```

Note that `alpha0` and `alpha1` in line 8 and 9 are the parameters for the bivariate gamma distribution, see Section 5.1.2 for details.

A.4 Code for Chapter 6

The following code is the non-parametric frailty model with gender-specific frailty defined in transition $1 \to 2$, which we have discussed in Section 7.1.2.

```
1 loglikelihood<-function(p){
2 # Model parameters:
3 beta0<-p[1:5]
4 beta1<-p[6]
5 beta2<-p[7:8]; beta3<-p[9:13]
6 b1<-p[14]
```

```
b2<- -b1
b<-c(b1,b2)
# Contribution for each individual:
loglik<-0
for(i in 1:N){
      data.i<-dta.split[[i]]
      o<-data.i$state
      t<-data.i$years
      g<-data.i$sex[1]
      # Gender-specific parameters:
      pi1.i<-exp(p[15]+p[16]*g)/(1+exp(p[15]+p[16]*g))
      pi2.i<-(1-pi1.i)
      bage<-data.i$bage[1]
      dage<-data.i$dage[1]
      contri<-c(1,1)
      for(k in 1:2){
            # Generator matrix Q
            for(j in 2:length(o)){
            Q<-matrix(0,4,4)
            Q[1,2]<-exp(beta0[1]+beta1[1]*t[j-1]+beta2[1]*
               bage+beta3[1]*dage+b[k])
            Q[1,4]<-exp(beta0[2]+beta2[2]*bage+beta3[2]*dage)
            Q[1,1]<--sum(Q[1,])
            Q[2,3]<-exp(beta0[3]+beta3[3]*dage)
            Q[2,4]<-exp(beta0[4]+beta3[4]*dage)
            Q[2,2]<--sum(Q[2,])
            Q[3,4]<-exp(beta0[5]+beta3[5]*dage)
            Q[3,3]<--sum(Q[3,])
            P <- MatrixExp(mat=Q,t=t[j]-t[j-1])
            if(o[j]!=dead){
                  contri[k]<-contri[k]*P[o[j-1],o[j]]
```

```
            }
            if(o[j]==dead){
                  contri[k]<-contri[k]*(P[o[j-1],1]*Q[1,dead
                     ]+P[o[j-1],2]*Q[2,dead]+P[o[j-1],3]*Q[3,
                     dead])
            }
      }
      }
loglik<-loglik+log(contri[1]*pi1.i+contri[2]*pi2.i)
}
cat("-2*Loglik = ", -2*loglik,"\n")
cat("b1",round(b1,3),"\n")
cat("pi1", round(pi1,3),"\n"))
-loglik
}
```

A.5 Code for Chapter 7

The following code shows how to draw the bivariate frailty distribution. The graph of the fitted density of bivariate gamma frailty in Model VII for OCTO-Twin data in two viewing directions is shown in Figure 7.2 in Section 7.2.2.

```
den3dp <- kde2d(B1, B2)
persp(den3dp, box=TRUE, theta=-15,phi=0,
xlab="B1", ylab="B2",zlab="density of bivarite gamma",
d=1,col="grey",border="blue" ,axes=TRUE,ticktype="detailed")
```

Note that `B1` and `B2` in the above code are the estimated frailties for each individual in OCTO-Twin data.

Appendix B

Delta method

The delta method is generally used to calculate the approximate probability distribution of a function of the asymptotically normal distributed estimator with a known limiting variance. Here the univariate delta method can be used to calculate the standard error of frailty parameters σ, κ through σ^*, κ^* as mentioned in Section 4.2, as well as the parameter π through π^*, which will be introduced in Chapter 6. Define a consistent estimator X_n converges in probability to its true value μ when $n \to \infty$, then the uncertainty of X_n is asymptotic normality,

$$\sqrt{n}(X_n - \mu) \xrightarrow{D} N(0, \sigma^2),$$

where n is the number of observations and σ is the variance, $\xrightarrow{D}$ means the convergence in distribution. For the estimation of the variance of a function g of the estimator,

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{D} N(0, \sigma^2 \cdot (g'(\mu))^2),$$

where $g'(\mu)$ is the differentiation of $g(\mu)$. Therefore,

$$Var(g(X_n)) \approx \sigma^2 \cdot (g'(\mu))^2 / n$$

Bibliography

[1] Jaap H Abbring and Gerard J Van Den Berg. The unobserved heterogeneity distribution in duration analysis. *Biometrika*, 94(1):87–99, 2007.

[2] Murray Aitkin. A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, 55(1):117–128, 1999.

[3] Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.

[4] Michael G Akritas. Nonparametric survival analysis. *Statistical Science*, pages 615–623, 2004.

[5] Per Kragh Andersen and Niels Keiding. Multi-state models for event history analysis. *Statistical methods in medical research*, 11(2):91–115, 2002.

[6] Jon E Anderson and Thomas A Louis. Survival analysis using a scale change random effects model. *Journal of the American Statistical Association*, 90(430):669–679, 1995.

[7] Theodor A Balan and Hein Putter. A tutorial on frailty models. *Statistical methods in medical research*, 29(11):3424–3454, 2020.

[8] Jessica K Barrett, Fotios Siannis, and Vern T Farewell. A semi-competing risks model for data with interval-censoring and informative observation: an application to the mrc cognitive function and ageing study. *Statistics in medicine*, 30(1):1–10, 2011.

[9] Francesco Bartolucci and Alessio Farcomeni. A discrete time event-history approach to informative drop-out in mixed latent markov models with covariates. *Biometrics*, 71(1):80–89, 2015.

[10] Leonard E Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, 41(1):164–171, 1970.

[11] Govert E Bijwaard. Multistate event history analysis with frailty. *Demographic Research*, 30:1591–1620, 2014.

[12] Marie Böhnstedt and Jutta Gampe. Detecting mortality deceleration: Likelihood inference and model selection in the gamma-gompertz model. *Statistics & Probability Letters*, 150:68–73, 2019.

[13] Charles George Broyden. The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90, 1970.

[14] Alexandre Bureau, Stephen Shiboski, and James P Hughes. Applications of continuous time hidden markov models to the study of misclassified disease outcomes. *Statistics in medicine*, 22(3):441–462, 2003.

[15] TC Buter, A Van Den Hout, FE Matthews, JP Larsen, C Brayne, and D Aarsland. Dementia and survival in parkinson disease: a 12-year population study. *Neurology*, 70(13):1017–1022, 2008.

[16] Bradley P Carlin and Thomas A Louis. *Bayesian methods for data analysis*. CRC Press, 2008.

[17] Chrys Caroni, Martin Crowder, and Alan Kimber. Proportional hazards models with discrete frailty. *Lifetime data analysis*, 16(3):374–384, 2010.

[18] A Clifford Cohen. Maximum likelihood estimation in the weibull distribution based on complete and on censored samples. *Technometrics*, 7(4):579–588, 1965.

[19] David Collett. *Modelling survival data in medical research*. CRC press, 2015.

[20] D Commenges. Multi-state models in epidemiology. *Lifetime data analysis*, 5(4):315–327, 1999.

[21] Daniel Commenges. Inference for multi-state models from interval-censored data. *Statistical methods in medical research*, 11(2):167–182, 2002.

[22] Richard J Cook and Jerald F Lawless. *Multistate models for the analysis of life history data*. Chapman and Hall/CRC, 2018.

[23] Christopher Cox. Delta method. *Encyclopedia of biostatistics*, 2, 2005.

[24] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

[25] David Roxbee Cox and David Victor Hinkley. *Theoretical statistics*. CRC Press, 1979.

[26] David Roxbee Cox and Hilton David Miller. *The theory of stochastic processes*. Routledge, 2017.

[27] Philip J Davis and Philip Rabinowitz. *Methods of numerical integration*. Courier Corporation, 2007.

[28] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

[29] Dirley M dos Santos, Richard B Davies, and Brian Francis. Nonparametric hazard versus nonparametric frailty distribution in modelling recurrence

of breast cancer. *Journal of statistical planning and inference*, 47(1-2):111–127, 1995.

[30] Luc Duchateau and Paul Janssen. *The frailty model*. Springer Science & Business Media, 2007.

[31] Frank Eriksson, Torben Martinussen, and Thomas H Scheike. Clustered survival data with left-truncation. *Scandinavian Journal of Statistics*, 42(4):1149–1166, 2015.

[32] Sylvie Escolano, Jean-Louis Golmard, Anne-Marie Korinek, and Alain Mallet. A multi-state model for evolution of intensive care unit patients: prediction of nosocomial infections and deaths. *Statistics in medicine*, 19(24):3465–3482, 2000.

[33] Luis Fernandez. Non-parametric maximum likelihood estimation of censored regression models. *Journal of Econometrics*, 32(1):35–57, 1986.

[34] Roger Fletcher. A new approach to variable metric algorithms. *The computer journal*, 13(3):317–322, 1970.

[35] Marshal F Folstein, Susan E Folstein, and Paul R McHugh. "mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189–198, 1975.

[36] Yohann Foucher, M Giral, JP Soulillou, and JP Daures. A flexible semi-markov model for interval-censored data and goodness-of-fit testing. *Statistical methods in medical research*, 19(2):127–145, 2010.

[37] Yohann Foucher, Magali Giral, Jean-Paul Soulillou, and Jean-Pierre Daures. A semi-markov model for multistate and interval-censored data with multiple terminal events. application in renal transplantation. *Statistics in medicine*, 26(30):5381–5393, 2007.

[38] Yohann Foucher, Eve Mathieu, Philippe Saint-Pierre, Jean-François Durand, and Jean-Pierre Daurès. A semi-markov model based on generalized weibull distribution with an illustration for hiv disease. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 47(6):825–833, 2005.

[39] John E Freund. A bivariate extension of the exponential distribution. *Journal of the American Statistical Association*, 56(296):971–977, 1961.

[40] Jochen Garcke. Sparse grids in a nutshell. In *Sparse grids and applications*, pages 57–80. Springer, 2012.

[41] Francesca Gasperoni, Francesca Ieva, Anna Maria Paganoni, Christopher H Jackson, and Linda Sharples. Non-parametric frailty cox models for hierarchical time-to-event data. *Biostatistics*, 21(3):531–544, 2020.

[42] RC Gentleman, JF Lawless, JC Lindsey, and P Yan. Multi-state markov models for analysing incomplete disease history data with illustrations for hiv disease. *Statistics in medicine*, 13(8):805–821, 1994.

[43] Donald Goldfarb. A family of variable-metric methods derived by variational means. *Mathematics of computation*, 24(109):23–26, 1970.

[44] Tommi Härkänen, Hannu Hausen, Jorma I Virtanen, and Elja Arjas. A non-parametric frailty model for temporally clustered multivariate failure times. *Scandinavian journal of statistics*, 30(3):523–533, 2003.

[45] N Hens, A Wienke, M Aerts, and Geert Molenberghs. The correlated and shared gamma frailty model for bivariate current status data: an illustration for cross-sectional serological data. *Statistics in Medicine*, 28(22):2785–2800, 2009.

[46] Sascha Hilgenfeldt, Robert Balder, and Christoph Zenger. Sparse grids: applications to multi-dimensional schrödinger problems. 1995.

[47] Philip Hougaard. Frailty models for survival data. *Lifetime Data Analysis*, 1(3):255–273, 1995.

[48] Philip Hougaard. *Analysis of multivariate survival data*. Springer Science & Business Media, 2012.

[49] Joseph G Ibrahim, Ming-Hui Chen, Debajyoti Sinha, JG Ibrahim, and MH Chen. *Bayesian survival analysis*, volume 2. Springer, 2001.

[50] Christopher Jackson. Multi-state models for panel data: the msm package for r. *Journal of statistical software*, 38(1):1–28, 2011.

[51] Helena Jeličić, Erin Phelps, and Richard M Lerner. Use of missing data methods in longitudinal studies: the persistence of bad practices in developmental psychology. *Developmental psychology*, 45(4):1195, 2009.

[52] JD Kalbfleisch and Jerald Franklin Lawless. The analysis of panel data under a markov assumption. *Journal of the american statistical association*, 80(392):863–871, 1985.

[53] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[54] Richard Kay. A markov model for analysing cancer markers and disease states in survival studies. *Biometrics*, pages 855–865, 1986.

[55] Niels Keiding and Richard D Gill. Random truncation models and markov processes. *The Annals of Statistics*, pages 582–602, 1990.

[56] John F Kenney. *Mathematics of Statistics*. D. Van Nostrand, 1939.

[57] Samuel Kotz, Narayanaswamy Balakrishnan, and Norman L Johnson. *Continuous multivariate distributions, Volume 1: Models and applications*. John Wiley & Sons, 2004.

[58] Vidyadhar G Kulkarni. *Introduction to modeling and analysis of stochastic systems*. Springer, 2011.

[59] Tze Leung Lai and Zhiliang Ying. Rank regression methods for left-truncated and right-censored data. *The Annals of Statistics*, pages 531–556, 1991.

[60] Nan Laird. Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73(364):805–811, 1978.

[61] Nan M Laird. Missing data in longitudinal studies. *Statistics in medicine*, 7(1-2):305–315, 1988.

[62] Hongzhe Li, Elizabeth A Thompson, and Ellen M Wijsman. Semiparametric estimation of major gene effects for age of onset. *Genetic Epidemiology*, 15(3):279–298, 1998.

[63] Yan-Fu Li and Enrico Zio. A multi-state model for the reliability assessment of a distributed generation system via universal generating function. *Reliability Engineering & System Safety*, 106:28–36, 2012.

[64] Maarten Lindeboom and Gerard J Van den Berg. Heterogeneity in models for bivariate survival: the importance of the mixing distribution. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(1):49–60, 1994.

[65] Bruce G Lindsay. The geometry of mixture likelihoods: a general theory. *The annals of statistics*, pages 86–94, 1983.

[66] Samuel OM Manda. A nonparametric frailty model for clustered survival data. *Communications in Statistics—Theory and Methods*, 40(5):863–875, 2011.

[67] Kenneth G Manton and Anatoli I Yashin. Effects of unobserved and partially observed covariate processes on system failure: a review of models and estimation strategies. *Statistical Science*, 12(1):20–34, 1997.

[68] Robson José Mariano Machado. *Penalised maximum likelihood estimation for multi-state models*. PhD thesis, UCL (University College London), 2018.

[69] Guillermo Marshall and Richard H Jones. Multi-state models and diabetic retinopathy. *Statistics in medicine*, 14(18):1975–1983, 1995.

[70] Adelino Martins, Marc Aerts, Niel Hens, Andreas Wienke, and Steven Abrams. Correlated gamma frailty models for bivariate survival time data. *Statistical Methods in Medical Research*, 28(10-11):3437–3450, 2019.

[71] Luís Meira-Machado, Jacobo de Uña-Álvarez, Carmen Cadarso-Suárez, and Per K Andersen. Multi-state models for the analysis of time-to-event data. *Statistical methods in medical research*, 18(2):195–222, 2009.

[72] Cleve Moler and Charles Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM review*, 45(1):3–49, 2003.

[73] MD Mostafa and MW Mahmoud. On the problem of estimation for the bivariate lognormal distribution. *Biometrika*, 51(3/4):522–527, 1964.

[74] Marco Munda, Catherine Legrand, Luc Duchateau, and Paul Janssen. Testing for decreasing heterogeneity in a new time-varying frailty model. *Test*, 25(4):591–606, 2016.

[75] Marco Munda, Federico Rotolo, Catherine Legrand, et al. Parfm: parametric frailty models in r. *Journal of Statistical Software*, 51(11):1–20, 2012.

[76] Bengt Muthén, Tihomir Asparouhov, et al. Growth mixture modeling: Analysis with non-gaussian random effects. *Longitudinal data analysis*, 143165, 2008.

[77] John A Nelder and Roger Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

[78] James D Nichols and William L Kendall. The use of multi-state capture-recapture models to address questions in evolutionary ecology. *Journal of Applied Statistics*, 22(5-6):835–846, 1995.

[79] Alexandra Niessl, Arthur Allignol, Jan Beyersmann, et al. Estimating state occupation and transition probabilities in non-markov multi-state models subject to both random left-truncation and right-censoring. *arXiv preprint arXiv:2004.06514*, 2020.

[80] Frank WJ Olver, Daniel W Lozier, Ronald F Boisvert, and Charles W Clark. *NIST handbook of mathematical functions hardback and CD-ROM*. Cambridge university press, 2010.

[81] Rumana Z Omar, Nigel Stallard, and John Whitehead. A parametric multistate model for the analysis of carcinogenicity experiments. *Lifetime data analysis*, 1(4):327–346, 1995.

[82] Aidan G O'Keeffe, Li Su, and Vernon T Farewell. Correlated multistate models for multiple processes: an application to renal disease progression in systemic lupus erythematosus. *Journal of the Royal Statistical Society. Series C, Applied statistics*, 67(4):841, 2018.

[83] Myunghee Cho Paik, Wei-Yann Tsai, and Ruth Ottman. Multivariate survival analysis using piecewise gamma frailty. *Biometrics*, pages 975–988, 1994.

[84] Daewoo Pak, Chenxi Li, David Todem, and Woosung Sohn. A multistate model for correlated interval-censored life history data in caries research. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(2):413–423, 2017.

[85] Jagdish K Patel and Campbell B Read. *Handbook of the normal distribution*, volume 150. CRC Press, 1996.

[86] Andrew Pickles and Robert Crouchley. A comparison of frailty models for multivariate survival data. *Statistics in Medicine*, 14(13):1447–1461, 1995.

[87] John H Pollard and Emil J Valkovics. The gompertz distribution and its applications. *Genus*, pages 15–28, 1992.

[88] William H Press and Glennys R Farrar. Recursive stratified sampling for multidimensional monte carlo integration. *Computers in Physics*, 4(2):190–195, 1990.

[89] Hein Putter, Marta Fiocco, and Ronald B Geskus. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine*, 26(11):2389–2430, 2007.

[90] Hein Putter, Jos van der Hage, Geertruida H de Bock, Rachid Elgalta, and Cornelis JH van de Velde. Estimation and prediction in a multi-state model for breast cancer. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 48(3):366–380, 2006.

[91] Hein Putter and Hans C Van Houwelingen. Dynamic frailty models based on compound birth–death processes. *Biostatistics*, 16(3):550–564, 2015.

[92] Hein Putter and Hans C van Houwelingen. Frailties in multi-state models: Are they identifiable? do we need them? *Statistical methods in medical research*, 24(6):675–692, 2015.

[93] Azra Ramezankhani, Michael J Blaha, Mohammad hassan Mirbolouk, Fereidoun Azizi, and Farzad Hadaegh. Multi-state analysis of hypertension and mortality: application of semi-markov model in a longitudinal cohort study. *BMC Cardiovascular Disorders*, 20(1):1–13, 2020.

[94] Ben D Rickayzen and Duncan EP Walsh. A multi-state model of disability for the united kingdom: implications for future need for long-term care for the elderly. *British Actuarial Journal*, 8(2):341–393, 2002.

[95] Annie Robitaille, Ardo van den Hout, Robson JM Machado, David A Bennett, Iva Čukić, Ian J Deary, Scott M Hofer, Emiel O Hoogendijk, Martijn Huisman, Boo Johansson, et al. Transitions across cognitive states and death among older adults in relation to education: A multistate survival model using data from six longitudinal studies. *Alzheimer's & Dementia*, 14(4):462–472, 2018.

[96] Germán Rodrıguez. Parametric survival models. *Rapport technique, Princeton: Princeton University*, 2010.

[97] Glen A Satten and Ira M Longini Jr. Markov chains with measurement error: Estimating the 'true'course of a marker of the progression of human immunodeficiency virus disease. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 45(3):275–295, 1996.

[98] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.

[99] David F Shanno. Conditioning of quasi-newton methods for function minimization. *Mathematics of computation*, 24(111):647–656, 1970.

[100] TS Shao, Tien Chi Chen, and RM Frank. Tables of zeros and gaussian weights of certain associated laguerre polynomials and the related generalized hermite polynomials. *Mathematics of Computation*, 18(88):598–616, 1964.

[101] Linda D Sharples, Christopher H Jackson, Jayan Parameshwar, John Wallwork, and Stephen R Large. Diagnostic accuracy of coronary angiography and risk factors for post–heart-transplant cardiac allograft vasculopathy. *Transplantation*, 76(4):679–682, 2003.

[102] Leopold Simar. Maximum likelihood estimation of a compound poisson process. *The Annals of Statistics*, 4(6):1200–1209, 1976.

[103] Rose Sisk, Lijing Lin, Matthew Sperrin, Jessica K Barrett, Brian Tom, Karla Diaz-Ordaz, Niels Peek, and Glen P Martin. Informative presence and observation in routine health data: A review of methodology for clinical risk prediction. *Journal of the American Medical Informatics Association*, 28(1):155–166, 2021.

[104] R Core Team et al. R: A language and environment for statistical computing. 2013.

[105] A Roger Thatcher, Väinö Kannisto, and James W Vaupel. The force of mortality at ages 80 to 120. 1998.

[106] Jos Twisk and Wieke de Vente. Attrition in longitudinal studies: how to deal with missing data. *Journal of clinical epidemiology*, 55(4):329–337, 2002.

[107] Arne Uhlendorff. From no pay to low pay and back again? a multi-state model of low pay dynamics. 2006.

[108] Gerard J Van den Berg and Bettina Drepper. Inference for shared-frailty survival models with left-truncated data. *Econometric Reviews*, 35(6):1075–1098, 2016.

[109] Ardo Van Den Hout. *Multi-state survival models for interval-censored data*. CRC Press, 2016.

[110] Ardo Van Den Hout and Graciela Muniz-Terrera. Joint models for discrete longitudinal outcomes in aging research. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 65(1):167–186, 2016.

[111] James W Vaupel, Kenneth G Manton, and Eric Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454, 1979.

[112] Stephen G Walker and Bani K Mallick. Hierarchical generalized linear models and frailty models with bayesian nonparametric mixing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(4):845–860, 1997.

[113] CY Wang. Non-parametric maximum likelihood estimation for cox regression with subject-specific measurement error. *Scandinavian Journal of Statistics*, 35(4):613–628, 2008.

[114] James T Wassell and ML Moeschberger. A bivariate survival model with modified gamma frailty for assessing the impact of interventions. *Statistics in Medicine*, 12(3-4):241–248, 1993.

[115] Stefan Weinzierl. Introduction to monte carlo methods. *arXiv preprint hep-ph/0006269*, 2000.

[116] George H Weiss and Marvin Zelen. A semi-markov model for clinical trials. *Journal of Applied Probability*, 2(2):269–285, 1965.

[117] Claire MA Wintrebert, Hein Putter, Aeilko H Zwinderman, and JC Van Houwelingen. Centre-effect on survival after bone marrow transplantation: application of time-dependent frailty models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 46(5):512–525, 2004.

[118] Simon N Wood. *Core statistics*, volume 6. Cambridge University Press, 2015.

[119] Liang Wu, Jeongje Park, Jaeseok Choi, AA El-Keib, Mohammad Shahidehpour, and Roy Billinton. Probabilistic reliability evaluation of power systems including wind turbine generators using a simplified multi-state model: A case study. In *2009 IEEE Power & Energy Society General Meeting*, pages 1–6. IEEE, 2009.

[120] Xiaonan Xue and Ron Brookmeyer. Bivariate frailty model for the analysis of multivariate survival time. *Lifetime Data Analysis*, 2(3):277–289, 1996.

[121] Anatoli I Yashin, James W Vaupel, and Ivan A Iachine. Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. *Mathematical Population Studies*, 5(2):145–159, 1995.

[122] KKW Yau and CA McGilchrist. Ml and reml estimation in survival analysis with time dependent correlated frailty. *Statistics in Medicine*, 17(11):1201–1213, 1998.

[123] Shi-Tao Yeh et al. Using trapezoidal rule for the area under a curve calculation. *Proceedings of the 27th Annual SAS® User Group International (SUGI'02)*, 2002.