ANALYSIS OF DISCRETE TIME COMPETING RISKS DATA WITH MISSING FAILURE CAUSES AND CURED SUBJECTS

UNIVERSITY OF KWAZULU - NATAL

INYUVESI YAKWAZULU-NATALI

Bonginkosi Ducan Ndlovu

October, 2022

Analysis of discrete time competing risks data with missing failure causes and cured subjects

by

Bonginkosi Duncan Ndlovu

A thesis submitted to the
University of KwaZulu-Natal
in fulfilment of the requirements for the degree
of
DOCTOR OF PHILOSOPHY
in
APPLIED STATISTICS

Thesis supervisor: Prof Temesgen Zewotir

UNIVERSITY OF KWAZULU - NATAL
INYUVESI YAKWAZULU-NATALI

UNIVERSITY OF KWAZULU-NATAL
SCHOOL OF MATHEMATICS, STATISTICS AND COMPUTER SCIENCE
SCOTTSVILLE CAMPUS, PIETERMARITZBURG, SOUTH AFRICA

Declaration

I certify that this thesis, and the research to which it refers, are the product of my own work and that any ideas or quotations from the work of other people, published or unpublished, are fully acknowledged.

[redacted] 21/07/2023

Bonginkosi Duncan Ndlovu(student) Date

[redacted] 21/07/2023

Prof Sileshi Fanta Melesse(co-supervisor) Date

[redacted] 20/07/2023

Prof Temesgen Zewotir(supervisor) Date

Disclaimer

This document describes work undertaken as a PhD programme of study at the University of KwaZulu-Natal (UKZN). All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institution.

Abstract

This thesis is motivated by the limitations of the existing discrete time competing risks models vis-a-vis the treatment of data that comes with missing failure causes or a sizable proportions of cured subjects. The discrete time models that have been suggested to date (Davis and Lawrance, 1989; Tutz and Schmid, 2016; Ambrogi et al., 2009; Lee et al., 2018) are cause-specific-hazard denominated. Clearly, this fact summarily disqualifies these models from consideration if data comes with missing failure causes. It is also a well documented fact that naive application of the cause-specific-hazards to data that has a sizable proportion of cured subjects may produce downward biased estimates for these quantities. The existing models can be considered within the multiple imputation framework (Rubin, 1987) for handling missing failure causes, but the prospects of scaling them up for handling cured subjects are minimal, if not nil. In this thesis we address these issues concerning the treatment of missing failure causes and cured subjects in discete time settings. Towards that end, we focus on the mixture model (Larson and Dinse, 1985) and the vertical model (Nicolaie et al., 2010) because these models possess certain properties which dovetail with the objectives of this thesis. The mixture model has been upgraded into a model that can handle cured subjects. Nicolaie et al. (2015) have demonstrated that the vertical model can also handle missing failure causes as is. Nicolaie et al. (2018) have also extended the vertical model to deal with cured subjects. Our strategy in this thesis is to exploit both the mixture model and the vertical model as a launching pad to advance discrete time models for handling data that comes with missing failure causes or cured subjects.

Acknowledgements

I would like to express my deepest gratitude towards my supervisor, Professor T Zewotir, and co-supervisor, Professor SF Melesse, for their support and invaluable guidance.

I would also like to thank my dear friend, Professor PE Green, without whom I would not have embarked on this journey.

Contents

List of Figures

List of Tables

List of Publications

1. Ndlovu, B.D., Melesse, S.F, and Zewotir, T. (*Unpublished*). A mixture approach to the analysis of discrete time competing risks data with missing failure causes. *Iranian Journal of Statistics*

2. Ndlovu, B.D., Melesse, S.F, and Zewotir, T. (2023). A nonparametric analysis of discrete time competing risks data: A comparison of the cause-specific-hazards approach and the vertical approach. *Statistics in Transition*, 24:61-76

3. Ndlovu, B.D., Melesse, S.F, and Zewotir, T. (2022). An EM model for analysis of discrete time competing risks data with missing failure causes. *Model Assisted Statistics and Applications*, 17:167-174

4. Ndlovu, B.D., Melesse, S.F, and Zewotir, T. (2022). A regression analysis of discrete time competing risks data using a vertical model approach *South African Statistical Journal*, 1:21-36

5. Ndlovu, B.D., Melesse, S.F, and Zewotir, T. (2020). A nonparametric vertical model: an application to discrete time competing risks data with missing failure causes. *South African Statistical Journal*, 2:231-241

6. Ndlovu, B.D., Melesse, S.F, and Zewotir, T. (2019). A mixture model with application to discrete competing risks data. *South African Statistical Journal*, 2:73-86

CHAPTER 1

Introduction

Data that arises from discrete time competing risks experiments may also come with the usual complications such as cured subjects, missing failure causes, correlated failure times and so on. By discrete time we refer to experiments where time evolves in discrete units, for example, time to graduation is observed in semesters or years. Most of these issues have received attention in the continuous time realm, and, a number of purpose built models to address these unique data structures, have been advanced. This has, however, not been the case in the discrete time domain, in fact, the development of discrete time models in general has been pedestrian in comparison to advances in continuous time.

To date, the discrete time models that have been advanced are only for analysis of ordinary competing risks data (Davis and Lawrance, 1989; Ambrogi et al., 2009; Tutz and Schmid, 2016; Lee et al., 2018). With ever rising popularity of discrete times competing risks models with applied researchers, it is inevitable that these complications may arise in practice against a background of limited options to deal with them. Both the multinomial model (Ambrogi et al., 2009; Tutz and Schmid, 2016) and the binomial model (Lee et al., 2018) are *cause-specific-hazard* (CSH) denominated discrete time regression models, that is, both models advance CSHs for modelling discrete time competing risks data. This includes the nonparametric CSHs model that was suggested by Davis and Lawrance (1989). Therefore, the shortcomings of the existing discrete time competing risks models are in actual fact the limitations that are associated with modelling data with CSHs.

The limitations of modelling competing risks data with CSHs are well documented. Some of these issues have been addressed in the discrete time realm. For example, Berger et al. (2020) have advanced a discrete time subhazard regression model for the *cumulative incidence function* (CIF) which has a one-to-one relationship with data. This model was advanced in response to the regression model for the CIF that is derived from CSHs which has been notorious for complicating the evaluation of covariate effects. Steele et al. (2004), for example, have advanced a multinomial model that accounts for correlated failure times.

The focus of this thesis is on the analysis of discrete time competing risks data that comes with missing failure causes or cured subjects. These are some of the topics that have not received attention in discrete time. Both the multinomial model and the binomial model cannot be applied to data that comes with missing failure causes because the estimation of CSHs requires full information in respect of failure causes for all failures. Also, naive application of CSHs to data that has a significant proportion of cured subjects may lead to downward biased CSH estimates Anderson et al. (1996). Analysis of data that comes with missing failure causes is a topic that has received extensive coverage in continuous time, see (Dinse, 1982; Dewanji, 1992; Goetghebeur and Ryan, 1990, 1995; Lu and Tsiatis, 2001; Bakoyannis et al., 2010; Nicolaie et al., 2015) for examples of models that have been advanced to deal with this data complication. The treatment of competing risks data that has a non-ignorable poroportion of cured subjects has also received attention in continuous time (Escarela et al., 2000; Choi, 2002; Maller and Zhou, 2002; Zhiping, 2011). None of these models can be applied in discrete time. Continuous time models are premised on factorisation of the full likelihood function into cause-specific likelihood functions. This is not possible in the presence of a sizable proportion of tied failure times, a characteristic feature of discrete time data.

The mixture model (Larson and Dinse, 1985) and the vertical model (Nicolaie et al., 2010) are continuous time competing risks models that have been advanced after the CSHs model (Prentice et al., 1978). The mixture model has been upgraded into a model that can handle cured subjects. In actual fact, the continuous time competing risks models for handling cured subjects (Escarela et al., 2000; Choi, 2002; Maller and Zhou, 2002; Zhiping, 2011) are derived from the mixture model. Furthermore, the authors of the model have suggested that their model can be extended to deal with the presence of missing failure causes in data. The same authors of the vertical model, that is, Nicolaie et al. (2015), have shown that the ordinary vertical model can handle missing failure causes as is. Nicolaie et al. (2018) have further upgraded the model to deal with cured subjects. We focus on these two models in this thesis with a view to re-cast them as discrete time model as the first step and then upgrade the re-formulated models to address the issues of missing failure causes and cured subjects.

In more formal terms, the objectives of this thesis are;

1) To re-formulate the mixture model and the vertical model as discrete time competing risks models.

2) To upgrade the discrete time models that were proposed in 1) into models that can handle discrete time competing risks data that comes with missing failure causes

3) To upgrade the vertical model into a nonparametric model that can deal with the presence of cured subjects in discrete time competing risks data

The remainder of this work is organised as follows; in the next chapter, Chapter 2, we discuss the discrete time competing risks literature. This includes some of the developments that have been suggested to deal with some of the complications that often come with data. In Chapter 3 we commence with the first part of the thesis where we discuss the adjustments that must be effected such that the mixture model, as originally proposed by its authors, can be applied to discrete time competing risks data. We also upgrade this model into a model that can handle missing failure causes. This is followed by the development of a truly discrete time version of the mixture model in Chapter 4. This model is also upscaled into a model that can deal with missing failure causes. In Chapter 5 we advance a discrete time version of the vertical model and also show that the same model can equally handle missing failure causes as is. In Chapter 6 we advance a nonparametric vertical model for dealing with data that comes with a sizable proportion of cured subjects. In Chapter 7 we apply all these models to a new data set for validation purposes. Conclusions and future directions are discussed in Chapter 8, the final chapter of this thesis.

CHAPTER 2

Discrete Time Competing Risks Models

Competing risks has come to refer to survival analysis studies where subjects are exposed to multiple modes of failure. When time to failure is observed in discrete units, these studies are then categorized as discrete time competing risks as distinct from continuous time competing risks. This differentiation is important because the models that are developed in one time domain are not approprite for application in the other domain. The history of survival analysis and competing risks in particular is essentially embedded in the continuous time domain with applications typically found in medical research, clinical trials and so on, where time to failure is patently continuous. Discrete time competing risks models were developed much later specifically for analysis of discrete time competing risks data because continuous time models are not appropriate for application in discrete time. The application of competing risks models in discrete time can be traced back to authors such as (Lancaster, 1979; Nickell, 1979; Narendranathan et al., 1985) in the analysis of duration data. Then, discussions centered around the topic of modelling unemployment duration or time to re-employment for unemployed individuals. Why do some unemployed individuals take longer to find employment when other individuals are re-employed almost immediately after losing their jobs and, more importantly, what factors explain these variations in the length of unemployment spells? These are some of the issues that pre-occupied researchers of that era. Since employment can take the form of casual, part-time, full-time and so on, these studies were often framed as competing risks experiments with the length of the unemployment spell as the variable of interest. Naturally, continuous time competing risks models were considered in these studies when in fact duration data usually comes in the form of discrete time data. Even though time to re-employment may evolve continuously (especially casual employment which may occur daily) in some instances, events are often recorded periodically, that is, bi-weekly, monthly and so on. In citing challenges that were encountered when the Cox (1972) regression model, in particular, is applied to duration data, Han and Hausman (1990), in their own words, said:

1) It is a continuous time specification (in reference to the Cox (1972) regression model) while most of duration data in econometrics are discrete where the discreteness may well be important.

2) While various ad hoc procedures have been developed to treat tied failure times within the partial likelihood framework, they become cumbersome in the presence of many ties.

The difficulty with discrete time competing risks data is an excessive number of ties due to time to event that is measured in discrete units. Discrete time arises in two ways. Time to failure $\tilde{T}$ is said to be inherently discrete if it is truly observed in discrete units. Alternatively, when continuous survival times have been grouped into intervals of the form; $[a_0, a_1); [a_1, a_2), \ldots, [a_{q\text{-}1}, a_q)$, with $a_0 = 0$ and $a_q = \infty$, such that $\bigcup_{s=1}^{q} [a_{s\text{-}1}, a_s)$ exhausts *follow up* or observation period, with q as some positive integer, then time to failure is also said to be observed in discrete units. As such, $\tilde{T}$ only assumes discrete values, that is, $\tilde{T} \in \{1, 2 \ldots, q\}$. When $\tilde{T} = \tilde{t}$ and $\tilde{T}$ is inherently discrete, it means that the event actually occured at time $\tilde{t}$. When continuous survival times have been grouped into intervals, then $\tilde{T} = \tilde{t}$ implies that the event took place in the interval: $[a_{\tilde{t}\text{-}1}, a_{\tilde{t}})$, i.e.; $a_{\tilde{t}\text{-}1} \leq \tilde{t} < a_{\tilde{t}}$. Measuring time to failure in this manner gives rise to a large number of tied events. Consider, for example, a typical study in education where researchers are often interested in identifying the drivers of dropouts and graduations as well as quantifying their effect on the risks of these events (Scott and Kennedy, 2005; Cleric et al., 2014; Vallejos and Steel, 2017). A simple formulation of a study in this context frames dropouts and graduation as competing events where time to failure (graduation or dropout) is time spent continuously from registration until failure and the various forms of withdrawals such as academic exclusion, voluntary withdrawal, withdrawal for financial reasons and so on, are collapsed into one category of failure, that is, dropouts. Suppose that an academic year has been partitioned into semesters, that is, students can graduate at the end of a semester, furthermore, they are required to register at the beginning of each semester. Graduation times are inherently discrete because a graduation is only observed at the end of a semester. On the other hand, students may withdraw from an institution at any time during the semester, but the authorities are only able to determine that status at the beginning of the following semester when the student fails to re-register. Thus, dropout times are continuous survival times that have been grouped into semesters. At the end of any given semester, a substantial number of students will graduate or drop out, hence the excessive number of ties.

Continuous time competing risks models are premised on the assumption that the full likelihood function factorizes into likelihood functions according to each failure cause (Prentice et al., 1978). Therefore, the failure times due to a given failure cause can be modeled seperately by regarding the failure times due to other failure causes as censoring times. This is not possible in discrete time owing to the excessive number of tied events. Therefore, continuous time models are not appropriate for application in discrete time for theoretical reasons. In any event, continuous time models tend to produce biased estimates when they are coerced upon discrete time data (Han and Hausman, 1990). The fact that continuous time models are not appropriate for application in discrete time has provided the motivation for the development of discrete time specific models, this applies to both the single mode of failure and competing risks settings. This means that all the models that were advanced in the continuous time realm, that is, the CSHs model (Prentice et al., 1978), the mixture model (Larson and Dinse, 1985) and the vertical model (Nicolaie et al., 2010) cannot be naively applied in discrete time. Incidentally, re-formulating the last two models as discrete time models forms the main thrust of this thesis.

The first truly discrete time model was advanced by Davis and Lawrance (1989) for nonparametric analysis. This model proposes discrete time CSHs for modelling data. To fix some ideas, suppose that C denotes time to censoring and $D \in \{1, 2, \dots, J\}$, where J is the number of failure causes. In general, observed competing risks data can be represented by $\boldsymbol{y} = ((t_1, \Delta_1, \boldsymbol{x}_1), (t_2, \Delta_2, \boldsymbol{x}_2) \ldots, (t_n, \Delta_n, \boldsymbol{x}_n))$, where $T_i = \min\{\tilde{T}_i, C_i\}$, and $\Delta_i = D_i I(\tilde{T}_i < C_i)$. Clearly, $\Delta_i = 0$ when t_i is a censoring time or $\Delta_i = j$ when t_i is a failure time due to failure cause j. Subject i covariates are captured in a p-dimensional vector $\boldsymbol{x}_i$. To distinguish continuous time data from discrete time data, $\tilde{T}$ and C are assumed to be observed in discrete units, that is, $\tilde{T}, C \in \{1, 2 \ldots, q\}$, for some positive integer q. For the purpose of nonparametric analysis, we extract data that can be represented by $\boldsymbol{y} = ((t_1, \Delta_1), (t_2, \Delta_2) \ldots, (t_n, \Delta_n))$. When Prentice et al. (1978) proposed the bivariate distribution of $(\tilde{T}, D)$ framework, the CSHs were advanced as the quantities that characterize the joint distribution of failure time and failure type. The nonparametric definition of continuous time CSHs is given by

$$h_j(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt, D = j | T \geq t)}{dt}.$$

The continuous time CSH $h_j(t)$ expresses the instanteous rate of failure by cause j at time t amongst other failure causes given survival up to time t. Davis and Lawrance (1989) proposed the following nonparametric definition for discrete time CSHs;

$$h_j(t) = P(T = t, D = j | T \geq t)$$

for $t = 1, \ldots, q$, and $j = 1, 2 \ldots, J$. Davis and Lawrance (1989) also advanced their model within this framework and the definition of nonparametric discrete time CSHs is the discrete time version of continuous time CSHs. Discrete time CSHs have a more relatable meaning in contrast to their continuous time counterparts because they are true probabilities, albeit, conditional. The discrete time CSH $h_j(t)$ is the probability of failure at time t by failure cause j amongst other failure causes given survival up to time t. In the context of the graduation/dropout example, the CSH of a dropout in semester t, for example, is the probability of staying registered continuously until the beginning of semester t and then drop out during the course of semester t instead of graduating. On the other hand, the CSH of graduating in semester t is the probability of continuous registration until the end of semester t and then graduate in place of dropping out during the course of the semester. Even though dropouts occur continuously during the course of semester t, they are aggregated towards the end of semester t to allow for a meaningful contrast between the conditional probabilities of graduating and dropping out. Note that forthwith, when we make mention of the term CSH it will be in reference to the discrete time CSH, the time domain will only be included if there is a possibility of confusion.

Returning to the nonparametric CSHs (Davis and Lawrance, 1989), the authors went on to determine $\hat{h}_j(t)$. We will proceed to demonstrate the estimation of the nonparametric CSHs because the same likelihood function leads to the discrete time regression model, that is, the multinomial model, which was proposed by (Ambrogi et al., 2009; Tutz and Schmid, 2016). We will follow up the discussion of the current nonparamteric model with the discussion of the multinomial model. Naturally, the MLE for $h_j(t)$ $(t = 1, 2 \ldots, q; j = 1, 2 \ldots, J)$, is determined via the maximization of the observed data likelihood function w.r.t. $h_j(t)$ $(t = 1, 2 \ldots, q; j = 1, 2 \ldots, J)$. The observed data log-likelihood function can be written as;

$$\mathcal{L} = \sum_{j=1}^{J} \sum_{i=1}^{n} d_{ij} \log P(T_i = t_i, D_i = j) + (1 - d_i) \log P(T_i > t_i)$$

where $d_{ij} = I(D_i = j)$ and $d_i = \sum_{j=1}^{J} d_{ij}$. We need to express $\mathcal{L}$ in terms of the CSHs. Note that the expression for the survival function $S(t)$ is given by

$$S(t) = P(T > t) = \prod_{s=1}^{t} (1 - h(s))$$

where $h(t) = \sum_{j=1}^{J} h_j(t)$ and $S(0) = 1$. The expression for cause j failure time density function $f_j(t)$ is given by

$$f_j(t) = P(T = t, D = j) = h_j(t) \prod_{s=1}^{t\text{-}1} (1 - h(s)) = \frac{h_j(t)}{1 - h(t)} \prod_{s=1}^{t} (1 - h(s)).$$

After appropriate substitutions and some minimal algebra, $\mathcal{L}$ can be re-written as;

$$\mathcal{L} = \sum_{i=1}^{n} \left[\sum_{j=1}^{J} d_{ij} \log \frac{h_j(t_i)}{1 - h(t_i)} + \sum_{s=1}^{t_i} \log(1 - h(s)) \right].$$

At this point we introduce a time dependent indicator variable d_{ijs} such that $d_{ijs} = 0$ for

$s = 1, 2 \ldots, t_i$-1, and $d_{ijt_i} = d_{ij}$. We now write $\mathcal{L}$ as;

$$\mathcal{L} = \sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{s=1}^{t_i} d_{ijs} \log \frac{h_j(s)}{1-h(s)} + \log(1-h(s))$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{s=1}^{t_i} d_{ijs} \log h_j(s) + (1 - \sum_{j=1}^{J} d_{ijs}) \log(1-h(s)).$$

Note that $\mathcal{L}$ can also be written as

$$\mathcal{L} = \sum_{l=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{s=l}^{t_i} d_{ijs} \log h_j(s) + (1-d_{is}) \log(1-h(s))$$
$$= \sum_{l=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{J} d_{ijl} \log h_j(l) + (1-d_{il}) \log(1-h(l))$$

where $d_{il} = \sum_{j=1}^{J} d_{ijl}$. Suppose that $\displaystyle\sum_{r=l}^{q}\sum_{j=1}^{J} d_{(jr)} + c_{(r)} = n_{(l)}$ is the size of the risk set at time l where $d_{(jr)}$ and $c_{(r)}$ are the number of failures due to failure cause j and the number of censored subjects at time r, respectively. Taking these definitions into account, we can then re-write $\mathcal{L}$ as;

$$\mathcal{L} = \sum_{l=1}^{q}\sum_{j=1}^{J} \log h_j(l) \sum_{i=1}^{n} d_{ijl} + \log(1-h(l)) \sum_{i=1}^{n} (1-d_{il})$$
$$= \sum_{l=1}^{q}\sum_{j=1}^{J} d_{(jl)} \log h_j(l) + (n_{(l)} - d_{(l)}) \log(1-h(l))$$

where $d_{(l)} = \displaystyle\sum_{j=1}^{J} d_{(jl)}$. Thus $\mathcal{L}$ is a kernel of a multinomial log-likelihood function, and as such, $\dfrac{\partial \mathcal{L}}{\partial h_j(l)} = 0$ yields an MLE of $h_j(l)$ that is given by

$$\hat{h}_j(l) = \frac{d_{(jl)}}{n_{(l)}}.$$

When Davis and Lawrance (1989) advanced their model they did not discuss the CIF, but from the CSH estimates we can also derive the estimates for CIFs;

$$\hat{F}_j(t) = \sum_{s=1}^{t} \hat{S}(s\text{-}1)\hat{h}_j(s)$$

for $t = 1, 2 \ldots, q$, and $j = 1, 2 \ldots, J$, where $F_j(t)$ is cause j CIF. Gaynor et al. (1993) later discussed the nonparametric analysis of continuous time competing risks data. To derive the estimates for continuous time CSHs, the authors arrived at the same multinomial log-likelihood function as Davis and Lawrance (1989) because continuous time CSHs were approximated with their discrete time counterparts. Gaynor et al. (1993) went on to demonstrate that the standard errors for the CIF estimate is given by

$$\operatorname{Var}(\hat{F}_j(t)) = \sum_{s=1}^{t} \operatorname{Var}\big(\hat{S}(s\text{-}1)\hat{h}_j(s)\big) + 2\sum_{s=1}^{t\text{-}1}\sum_{k=s+1}^{t} \operatorname{Cov}\big(\hat{S}(s\text{-}1)\hat{h}_j(s), \hat{S}(k\text{-}1)\hat{h}_j(k)\big)$$

where;

$$\begin{aligned}\operatorname{Var}\big(\hat{S}(s\text{-}1)\hat{h}_j(s)\big) &= \operatorname{Var}\big(\hat{S}(s\text{-}1)\hat{h}_j(s)\big)\\ &= \big(\hat{h}_j(s)\hat{S}(s\text{-}1)\big)^2\left(\frac{n_{(s)} - d_{(js)}}{d_{(js)}n_{(s)}} + \sum_{l=1}^{s\text{-}1}\frac{d_{(l)}}{n_{(l)}(n_{(l)} - d_{(l)})}\right)\end{aligned}$$

and,

$$\begin{aligned}\operatorname{Cov}\big(\hat{S}(s\text{-}1)\hat{h}_j(s), \hat{S}(k\text{-}1)\hat{h}_j(k)\big) &= \operatorname{Cov}\big(\hat{S}(s\text{-}1)\hat{h}_j(s), \hat{S}(k\text{-}1)\hat{h}_j(k)\big)\\ &= \big(\hat{h}_j(s)\hat{S}(s\text{-}1)\hat{h}_j(k)\hat{S}(k\text{-}1)\big)\\ &\times\left(-\frac{1}{n_{(s)}} + \sum_{l=1}^{s\text{-}1}\frac{d_{(l)}}{n_{(n)}(n_{(l)} - d_{(l)})}\right).\end{aligned}$$

Note that we have adapted the original expression for standard errors of the CIF estimate as suggested by Gaynor et al. (1993) to reflect that time is observed in discrete units. We contend that this expression for standard errors equally applies for the discrete time CIF estimate. In instances when the focus on analysis shifts towards assessing covariate effects on risks of failure then regression methods must be considered. Ambrogi et al. (2009) are credited with suggesting the very first discrete time regression model. This regression model has come to be known as the multinomial model (Ambrogi et al., 2009). The model also proposes CSHs for characterising data. With covariates, the definition of CSHs is now given by

$$h_j(t|\boldsymbol{x}) = P(T = t, D = j|T \geq t, \boldsymbol{x}).$$

The model for these quantities is given by

$$h_j(t|\boldsymbol{x};\boldsymbol{\alpha}) = \frac{\exp\left(\alpha_{0jt} + \boldsymbol{x}^T\boldsymbol{\alpha}_{1j}\right)}{1 + \sum_{l=1}^{J}\exp\left(\alpha_{0lt} + \boldsymbol{x}^T\boldsymbol{\alpha}_{1l}\right)}$$

for $j = 1, \ldots, J$, and $t = 1, \ldots, q$, where for cause j, α_{0lt} is the duration coefficient at time t and $\boldsymbol{\alpha}_{1j}$ is the corresponding vector of regression coefficients. Note that at time t and conditional on $\boldsymbol{x}$, a subject may fail by failure type j with probability $h_j(t|\boldsymbol{x};\boldsymbol{\alpha})$ or survive beyond time t with probability $h_0(t|\boldsymbol{x};\boldsymbol{\alpha}) = 1 - \sum_{j=1}^{J} h_j(t|\boldsymbol{x};\boldsymbol{\alpha})$. The CSH parameters $\boldsymbol{\alpha}_j$ $(j = 1, \ldots, J)$, where $\boldsymbol{\alpha}_j = (\alpha_{0j1} \ldots \alpha_{0jq}, \boldsymbol{\alpha}_{1j}^T)^T$, are estimated simultaneously by fitting a multinomial distribution to observed data. To see this, with covariates, the full log-likelihood function can be written as

$$\mathcal{L}(\boldsymbol{\alpha}) = \sum_{j=1}^{J}\sum_{i=1}^{n} d_{ij} \log P(T_i = t_i, D_i = j; \boldsymbol{x}_i, \boldsymbol{\alpha}) + (1 - d_i) \log P(T_i > t_i; \boldsymbol{x}_i, \boldsymbol{\alpha}).$$

By following similar steps that were taken to determine the nonparametric estimates for CSHs, we can write $\mathcal{L}(\boldsymbol{\alpha})$ as;

$$\begin{aligned}\mathcal{L}(\boldsymbol{\alpha}) &= \sum_{i=1}^{n}\left[\sum_{j=1}^{J}\sum_{s=1}^{t_i} d_{ijs} \log \frac{h_j(s|\boldsymbol{x}_i,\boldsymbol{\alpha})}{1 - h(s|\boldsymbol{x}_i,\boldsymbol{\alpha})} + \log(1 - h(s|\boldsymbol{x}_i,\boldsymbol{\alpha}))\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{s=1}^{t_i} d_{ijs} \log h_j(s|\boldsymbol{x}_i,\boldsymbol{\alpha}) + (1 - \sum_{j=1}^{J} d_{ijs}) \log(1 - h(s|\boldsymbol{x}_i,\boldsymbol{\alpha})).\end{aligned}$$

The log-likelihood function $\mathcal{L}(\boldsymbol{\alpha})$ is a kernel of a multinomial log-likelihood function. Therefore, $\boldsymbol{\alpha}$ can be estimated by fitting a multinomial distribution to data in long format or person-period format. Concerns have been raised about this model regarding the large number of CSH parameters that may have to be estimated simulteneously which may lead to instability of parameter estimates. In fact a maximum number of $(J \times (p + q))$ parameters may have to be estimated simultaneously. Since the full likelihood function splits into cause specific likelihood functions in continuous time, with a p-dimensional vector $\boldsymbol{x}$, the maximum number of parameters that are estimated simulteneously is p. In discrete time, for each failure cause, there are p regression coefficients plus q duration coefficients. With a larger number q, that is, with more time points, it so happens that some intervals become thinly populated with events. Naturally, this may ultimately have a bearing on the stability of estimates for the affected duration coefficients.

Consider a typical situation where failure times are negatively skewed. Obviously, here, the intervals towards the end of follow up will have a smaller number of events in relative terms. Some authors get around this problem by collapsing the affected intervals into fewer intervals to improve the count of events per interval. Möst et al. (2016) have introduced a penalized estimation method which controls the size of duration coefficient estimates by imposing restrictions on the differences between adjacent duration coefficients to minimize larger variation amongst the coefficients. This method also performs a variable selection procedure where the less influential variables are shrunk to zero. Lee et al. (2018) have advanced a binomial model where the CSHs are estimated individually within the GEE framework. Obviously, the number of parameters that are estimated simulteneously is reduced. The model proposes the following model for CSHs;

$$h_j(t|\boldsymbol{x};\boldsymbol{\alpha}_j) = \frac{\exp\left(\alpha_{0jt} + \boldsymbol{x}^T\boldsymbol{\alpha}_{1j}\right)}{1+\exp\left(\alpha_{0jt} + \boldsymbol{x}^T\boldsymbol{\alpha}_{1j}\right)}.$$

Here, the full log-likelihood function $\mathcal{L}(\boldsymbol{\alpha})$ is collapsed such that when a target CSH is estimated the non-target failure times are censored akin to estimation of these quantities in continuous time (Prentice et al., 1978). If cause j is the cause of interest, the collapsed log-likelihood function can be written as;

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\alpha}_j) &= \sum_{i=1}^{n} d_{ij}\log P(T_i = t_i;\boldsymbol{x}_i,\boldsymbol{\alpha}_j) + (1-d_{ij})\log P(T_i > t_i;\boldsymbol{x}_i,\boldsymbol{\alpha}_j)\\
&= \sum_{i=1}^{n} d_{ij}\log\frac{h_j(t_i|\boldsymbol{x}_i;\boldsymbol{\alpha}_j)}{1-h_j(t_i|\boldsymbol{x}_i;\boldsymbol{\alpha}_j)} + \sum_{s=1}^{t_i} d_{ij}\log(1-h_j(s|\boldsymbol{x}_i;\boldsymbol{\alpha}_j))\\
&+ \sum_{s=1}^{t_i}(1-d_{ij})\log(1-h_j(s|\boldsymbol{x}_i;\boldsymbol{\alpha}_j))\\
&= \sum_{i=1}^{n} d_{ij}\log\frac{h_j(t_i|\boldsymbol{x}_i;\boldsymbol{\alpha}_j)}{1-h_j(t_i|\boldsymbol{x}_i;\boldsymbol{\alpha}_j)} + \sum_{s=1}^{t_i}\log(1-h_j(s|\boldsymbol{x}_i;\boldsymbol{\alpha}_j)).
\end{aligned}$$

Let d_{ijs} retain the same meaning as previously, then $\mathcal{L}(\boldsymbol{\alpha}_j)$ can be written as;

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\alpha}_j) &= \sum_{i=1}^{n}\sum_{s=1}^{t_i} d_{ijs}\frac{h_j(s|\boldsymbol{x}_i;\boldsymbol{\alpha}_j)}{1-h_j(s|\boldsymbol{x}_i;\boldsymbol{\alpha}_j)} + \sum_{s=1}^{t_i}\log(1-h_j(s|\boldsymbol{x}_i;\boldsymbol{\alpha}_j))\\
&= \sum_{i=1}^{n}\sum_{s=1}^{t_i} d_{ijs}\log h_j(s|\boldsymbol{x}_i;\boldsymbol{\alpha}_j) + (1-d_{ijs})\log(1-h_j(s|\boldsymbol{x}_i;\boldsymbol{\alpha}_j)).
\end{aligned}$$

Clearly, $d_{ijs} \sim \mathcal{B}(1, h_j(s|\boldsymbol{x}_i;\boldsymbol{\alpha}_j))$, and $\mathcal{L}(\boldsymbol{\alpha}_j)$ is a sequence of Bernoulli log-likelihood functions. In Table 2.1, we have created fictitious data to introduce the data structure that is used by the binomial model (Lee et al., 2018) in contrast to the multinomial model.

Table 2.1: Examples of data structures

Original Data

ID	$(\Delta = 1)$	$(\Delta = 2)$	$(\Delta = 0)$	**T**	**X**
1	0	1	0	4	$\boldsymbol{x}_1$
2	0	0	1	3	$\boldsymbol{x}_2$
3	1	0	0	1	$\boldsymbol{x}_3$
4	1	0	0	2	$\boldsymbol{x}_4$
5	0	0	1	1	$\boldsymbol{x}_5$

Table 2.2: Examples of data structures

Multinomial Data

ID	$(\Delta = 1)$	$(\Delta = 2)$	$(\Delta = 0)$	**T**	**X**
	d_{i1s}	d_{i2s}	1-d_{i1s} -d_{i2s}		
1	0	0	1	1	$\boldsymbol{x}_1$
1	0	0	1	2	$\boldsymbol{x}_1$
1	0	0	1	3	$\boldsymbol{x}_1$
1	0	1	0	4	$\boldsymbol{x}_1$
2	0	0	1	1	$\boldsymbol{x}_2$
2	0	0	1	2	$\boldsymbol{x}_2$
2	0	0	1	3	$\boldsymbol{x}_2$
3	1	0	0	1	$\boldsymbol{x}_3$
4	0	0	1	1	$\boldsymbol{x}_4$
4	1	0	0	2	$\boldsymbol{x}_4$
5	0	0	1	1	$\boldsymbol{x}_5$

This table is based on $J = 2$ and 5 subjects where subject 1 failed due to cause 2 at $T = 4$, subject 2 censored at $T = 3$, subject 3 that failed at $T = 1$ due to cause 1, subject 4 that failed at $T = 2$ due to cause 1 and subject 5 that was censored at $T = 1$. Beginning with the familiar multinomial model, in the estimation of the CSHs simultaneously, a multinomial distribution is fitted to data set in Table 2.2 with $(d_{i1s}), d_{i2s}, (1 - d_{i1s} - d_{i2s})$ as the response vector, where $\boldsymbol{T}$ as a factor, and $\boldsymbol{X}$ are explanatory variables. Usually, $(1 - d_{i1s} - d_{i2s})$ is regarded as a reference category. For the binomial model (Lee et al., 2018), the multinomial data set in Table 2.2 is also used for estimating the CSHs. For example, to estimate the CSHs according to failure type 1, a binomial distribution is fitted with d_{i1s} as the responses within the GEE framework to account for correlation between d_{i1s} and d_{i2s} with $\boldsymbol{T}$, as a factor, and $\boldsymbol{X}$ are explanatory variables.

When the CSHs are estimated via the binomial model the number of parameters that are estimated simulteneously drops by a factor of J to $(p+q)$. The additional advantage of modelling data with the binomial model is that a covariate effect on a target CSH can be read directly from the corresponding regression coefficient estimate. This is difficult to achieve when data is modeled with the multinomial model. Suppose one is interested in evaluating the effect of covariate x_l, a component of the covariate vector $\boldsymbol{x}$, on the risk of failure by cause j. One has to take into account not only α_{1jl}, but other J-1 coefficients α_{1kl}, $k \neq j$ under the multinomial model. One need only consider α_{1jl} when the binomial model is assumed to assess the effect of covariate x_l on the risk of failure by cause j.

The CIF estimates are derived from the CSH estimates;

$$\hat{F}_j(t|\boldsymbol{x};\hat{\boldsymbol{\alpha}}) = \sum_{s=1}^{t} \hat{S}(s\text{-}1|\boldsymbol{x};\hat{\boldsymbol{\alpha}})\hat{h}_j(s|\boldsymbol{x};\hat{\boldsymbol{\alpha}})$$

for $j = 1, \ldots, J$, and $t = 1, \ldots, q$, where $\hat{S}(t|\boldsymbol{x};\hat{\boldsymbol{\alpha}}) = \sum_{s=1}^{t}(1 - \hat{h}(s|\boldsymbol{x};\hat{\boldsymbol{\alpha}}))$ and $h(s|\boldsymbol{x};\boldsymbol{\alpha}) = \sum_{j=1}^{J} h_j(s|\boldsymbol{x};\boldsymbol{\alpha})$. This regression model for the CIF has become notorious for complicating the assessment of covariate effects. In continuous time, this has lead to the development of other regression models for the CIF, such as the *transformation models* (Fine and Gray, 1999; Scheike and Gerds, 2008; Klein and Anderson, 2005), where the CIF is modeled directly on covariates. The CIF regression model that was proposed by Berger et al. (2020) is an extension of the model that was advanced by Fine and Gray (1999) for application in continuous time to a discrete time model. The model proposes the following expression for CIFs;

$$F_j(t|\boldsymbol{x}) = 1 - \prod_{s=1}^{t}(1 - \acute{h}_j(s|\boldsymbol{x})) = 1 - \acute{S}_j(t|\boldsymbol{x})$$

where $\acute{h}_j(t|\boldsymbol{x})$ is referred to as the subdistribution hazard. Note, here, that there is a direct relationship between $\boldsymbol{x}$ and $F_j(t|\boldsymbol{x}))$, for example, if $\boldsymbol{x}$ engenders an increase in $\acute{h}_j(t|\boldsymbol{x})$, it leads to an increase in $1 - \acute{S}_j(t|\boldsymbol{x}) = F_j(t|\boldsymbol{x})$ as well. The difference between the CSHs and subdistribution hazards lies in their respective risks sets. The risk set for the CSH at time t, consists of all the subjects that have survived to time t. On the other hand, for the subdistribution hazard, $\acute{h}_j(t|\boldsymbol{x})$, the risk set at time t consists of subjects that have not failed due to failure cause j and those that have failed due to other failure causes before time t, that is,

$$\acute{h}_j(t|\boldsymbol{x}) = P(T = t, D = j|(T \geq t) \cup (T \leq t\text{-}1, D \neq j), \boldsymbol{x}).$$

Clearly, the risk set for the subdistribution hazard $\acute{h}_j(t|\boldsymbol{x})$ will be larger than the risk set for the CSH $h_j(t|\boldsymbol{x})$. Fine and Gray (1999), the authors who introduced the notion of a subdistribution hazard or a subhazard, are the first to admit that the idea of this quantity is somewhat unnatural. In discrete time the subdistribution hazards are also modeled on covariates within the GLM framework:

$$g(\acute{h}_j(t|\boldsymbol{x}; \boldsymbol{\xi}_j)) = \xi_{0jt} + \boldsymbol{x}^T \boldsymbol{\xi}_{1j}$$

where ξ_{0jt} is a duration coefficient and $\boldsymbol{\xi}_{1j}$ is a vector of regression coefficients. The authors proceed to argue that $\boldsymbol{\xi}_j = (\xi_{0jt}, \boldsymbol{\xi}_{1j}^T)^T$ for $j = 1, 2 \ldots, J$, can be estimated via an application of a binomial distribution with weights, that is; the solution of $\mathcal{L}(\boldsymbol{\xi}_j)$ can be found by fitting a binomial distribution with δ_{is}'s as responses and w_{is}'s as weights such that;

$$\mathcal{L}(\boldsymbol{\xi}_j) = \sum_{i=1}^{n}\sum_{s=1}^{q\text{-}1} w_{is}\{\delta_{is}\log \acute{h}_j(s|\boldsymbol{x}_i, \boldsymbol{\xi}_j)) + (1 - \delta_{is})\log(1 - \acute{h}_j(s|\boldsymbol{x}_i, \boldsymbol{\xi}_j))\}$$

where $\delta_{is} = 0$ for $s = 1, 2 \ldots, t_i\text{-}1$, $\delta_{it_i} = 1$, $\delta_{is} = 0$ for $t_i < s \leq q\text{-}1$ when subject i failed at time t_i at first due to cause j. If the subject was censored at t_i, then $\delta_{is} = 0$ for $s = 1, 2 \ldots, q\text{-}1$. For both subjects, $w_{is} = 1$ for $s \leq t_i$ and $w_{is} = 0$ for $s > t_i$. For a subject that experienced first failure by other cause other than cause j; $w_{is} = 1$ for $s \leq t_i$ and

$w_{is} = \dfrac{G(s\text{-}1)}{G(t_i\text{-}1)}$ for $t_i < s < q$-1, where $G(s)$ is an estimate of the censoring distribution. The preparation of data for the estimation of the subdistribution hazards can easily be conducted via the R package `discSurv` (Welchowski and Schmid, 2019).

While Berger et al. (2020) have managed to address the constraints of the CSH denominated regression model for the CIF, the shortcomings of modelling data with CSHs are not limited to the regression model for the CIF. In fact, this thesis is motivated by a wish to address some of the limitations of these quantities, that is, the CSHs. As already highlighted, the discrete time regression models that have been advanced to date are one dimensional, meaning that both the multinomial model and the binomial model advance CSHs for modelling data. This poses a very serious risk because if data comes with a complication that precludes the application of CSHs, then both models are summarily disqualified from consideration without an additional option to fall back on. This includes the regression model for the CIF that was suggested by Berger et al. (2020) because subhazards are merely an extension of CSHs. A case in point is when data has missing failure causes for some subjects. In such instances, neither the multinomial nor the binomial can be applied because the estimation of CSHs requires complete information regarding failure causes.

In this thesis we address this issue, that is, handling of missing failure causes, as well as the treatment of data that has a sizable proportion of cured subjects. The main thrust of this thesis is to manipulate the mixture model and the vertical model to advance models for dealing with these complications in the discrete time realm. We initiate this task by re-casting these models as discrete time competing risks models and then follow that up with upgrading these models into missing failure causes models. The first model that we focus on with a view to re-purpose as a discrete time competing risks model is the mixture model. We undertake this work in the next chapter, Chapter 3.

CHAPTER 3

The Continuous Time Mixture Model

3.1 The Ordinary Competing risks Model

3.1.1 Introduction

The main objective of this thesis is to address some of the limitations that are associated with the modelling of competing risks data with CSHs as proposed by the existing discrete time models. We focus on the analysis of data that comes with misssing failure causes and data that has a significant proportion of cured subjects. This task has been broken down into three parts or objectives as outlined in the introduction. The first part entails advancing additional models for analysis of ordinary discrete time competing risks data. This is achieved by re-casting the mixture model (Larson and Dinse, 1985) and the vertical model (Nicolaie et al., 2010) as discrete time competing risks models. In this chapter we focus on the mixture model as suggested by its authors with a view to demonstrate the adjustments that are required for the model to be applied in discrete time. We also demonstrate how the model can be upscaled into a model that can handle missing failure causes. Chapter 4 is dedicated to advancing a truly discrete time version of the mixture model. In Chapter 5 we focus on the vertical model.

When Prentice et al. (1978) introduced the joint distribution of failure time $\tilde{T}$ and failure type D framework, the CSHs were advanced as quantities to characterize this joint distribution in modelling competing risks data. Essentially, the CSHs quantify the instanteneous risk of failure at a specified time by a cause of interest given survival to that time in the presence of other risks (causes) of failure as already explained in Chapter

2. Often, interest centers upon evaluation of covariate effects on the risk of failure by a specific cause interest. Towards that end, a model that connects the CSHs to data has to be specified. By far, the Cox (1972) regression model has been the most popular model for the CSHs. To re-state the representation of compenting risks data that was in introduced in Chapter 2, suppose that C denotes censoring time, then observed data can be represented as; $\boldsymbol{y} = ((t_1, \Delta_1, \boldsymbol{x}_1), (t_2, \Delta_2, \boldsymbol{x}_2), \ldots, (t_n, \Delta_n, \boldsymbol{x}_n)$, where $T_i = \min\{\tilde{T}_i C_i\}$, and $\Delta_i = D_i I(\tilde{T}_i < C_i)$. Thus, $\Delta_i = j$ or 0 depending on whether subject i failed by cause j or is censored at time t_i. The Cox (1972) regression model for CSHs is given by

$$h_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j) = h_{0j}(t) \exp(\boldsymbol{x}^T \boldsymbol{\beta}_j).$$

Arguably, this regression model owes its popularity to the ease with which it allows for evaluation of covariate effects on risks of failure. The covariate effects on the risk of failure by any cause can be inferred directly from the corresponding regression coefficient estimates. To determine $\hat{\boldsymbol{\beta}}_j$, the regression coefficients corresponding to cause j, a partial likelihood function is maximized w.r.t. $\boldsymbol{\beta}_j$, where failure times due to failure types other than cause j are censored. With J failure causes, the CSHs; $h_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j)\,(j = 1, \ldots, J)$,completely describe competing risks data, that is, all the usual estimands are derived from these quantities, including the CIF;

$$F_j(t|\boldsymbol{x}, \boldsymbol{\beta}) = \int_0^t S(s|\boldsymbol{x}, \boldsymbol{\beta}) h_j(s|\boldsymbol{x}, \boldsymbol{\beta}_j) ds$$

for $j = 1, 2, \ldots J$. The CIF $F_j(t|\boldsymbol{x}, \boldsymbol{\beta})$ is the proportion of failures due to failure type j by time t conditional on $\boldsymbol{x}$.

The mixture model (Larson and Dinse, 1985) was introduced as an alternative to the CSHs model (Prentice et al., 1978). This model proposes component hazards and failure type probabilities for modelling data. Characterization of competing risks data with these quantities follows from the assumption that the bivariate distribution of $(\tilde{T}, D)$ can be decomposed into a marginal distribution for failure type and a distribution for failure time conditional on failure type as summarized via the failure type probabilities $\pi_j(t)\,(j = 1, \ldots, J)$, and component hazards $\lambda_j(t)\,(j = 1, \ldots, J)$, respectively. The model views the population under study as a mixture of J subpopulations that arise according to each failure cause. It is this formulation that has been extensively exploited to advance the cure model, both in single mode of failure settings, see (Peng and Taylor, 2014) for review of this model and (Choi, 2002; Maller and Zhou, 2002; Zhiping, 2011; Nicolaie et al., 2018) for examples of the cure model in competing risks settings. It is assumed that there is a stochastic process which assigns the subjects to fail according to their respective failure causes at the outset. This includes censored subjects who are also assigned to fail by their respective failure causes which are, however, unobserved due to an observation period that is cut too short for these failure causes to be realized. The model, therefore, views observed data as a mixture that consists of realizations from J failure time random variables $\tilde{T}_1, \tilde{T}_2 \ldots, \tilde{T}_J$, each with its own unique failure time distribution as summarized by $\lambda_j(t)\,(j = 1, 2 \ldots, J)$, where $\pi_j\;(j = 1, 2, \ldots, J)$, are the mixing weights.

Naturally, in the presence of covariates, models are required to link both the component hazards and failure type probabilities to data. Since $D \in \{1, 2, \ldots, J\}$, the most natural model for the failure type distribution is the multinomial distribution, that is, the model for failure type probabilities is given by

$$\pi_j(\boldsymbol{x}, \boldsymbol{\gamma}) = \exp(\gamma_{0j} + \boldsymbol{x}^T\boldsymbol{\gamma}_{1j}) \Big/ (1 + \sum_{l=1}^{J-1} \exp(\gamma_{0l} + \boldsymbol{x}^T\boldsymbol{\gamma}_{1l})$$

for $j = 1, \ldots J$-1, where $\pi_J(\boldsymbol{x}, \boldsymbol{\gamma}) = 1 - \sum_{j=1}^{J-1} \pi_j(\boldsymbol{x}, \boldsymbol{\gamma})$. Let $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^T, \ldots, \boldsymbol{\gamma}_{J\text{-}1}^T)^T$, and $\boldsymbol{\gamma}_j = (\gamma_{0j}, \boldsymbol{\gamma}_{1j}^T)^T$, where γ_{0j} is a scalar and $\boldsymbol{\gamma}_{1j}$ is a vector of regression coefficients. The model allows for flexibility in the choice of a model for component hazards. In fact, the various permutations of the model differ in terms of the models for component hazards that have been proposed over the years. When Larson and Dinse (1985) introduced the model, they assumed proportional hazards for component hazards with piece-wise constant baseline component hazards. Most of the authors who have considered this model have also assumed proportional hazards for component hazards. Ng and McLachan (2003); Escarala and Bowater (2008) have assumed a semi-parametric formulation for component hazards, Haller (2014) modeled the component hazards to follow the proportional hazards assumption where the baseline component hazards were modeled with splines, Lau et al. (2008, 2011) have assumed a parametric model where they have specified a generalized gamma distribution for each component failure time distribution.

The model for component hazards that was assumed by Larson and Dinse (1985) implies that;

$$\lambda_{0j}(t) = \exp(\beta_{0jt}) \qquad t \in [a_{t\text{-}1}, a_t),$$

so that,

$$\lambda_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j) = \exp(\beta_{0jt}) \exp(\boldsymbol{x}^T \boldsymbol{\beta}_{1j})$$

for $j = 1 \ldots, J$. Here, follow up is coarsed into mutually exclusive intervals of the form; $[a_{s\text{-}1}, a_s)$ for $s = 1 \ldots, q$, with the last interval as; $[a_{q\text{-}1}, a_q)$ where $q = \infty$ i.e., the intervals completely cover or exhaust the total length of follow up. Let $\boldsymbol{\beta}_{0j}$ capture the baseline component hazard coefficients for sub-population j, i.e. $\boldsymbol{\beta}_{0j} = (\beta_{0j1} \ldots, \beta_{0jq})^T$, so that $\boldsymbol{\beta}_j = (\boldsymbol{\beta}_{0j}^T, \boldsymbol{\beta}_{1j}^T)^T$, summarizes all cause j component hazard parameters where $\boldsymbol{\beta}_{1j}$ is a vector of regression coefficients. This is the specification of the mixture model that was advanced by Larson and Dinse (1985) for analysis of continuous time competing risks data. The objective here is to demonstrate how this model can be applied to discrete time competing risks data. Secondly, given this model for component hazards, that is, a proportional hazards assumption for component hazards with piece-wise constant baseline component hazards, it will be demonstrated that the parameter vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J)$, can be estimated via a certain Poisson regression model. This was demonstrated by Holford (1980) and Laird and Oliver (1981) independently of each other in single mode of failure settings when the hazard function is modeled to follow the proportional hazards assumption with a piece-wise constant baseline hazard. Note that partitioning follow up in this manner coincides with grouped survival times in discrete time. We exploit this connection to apply a patently continuous time model in discrete time. We will return to this point when we discuss "exposure" in Subection 3.1.3

The subpopulations are regarded as populations in their own right where each subpopulation has its own distinct set of functionals, that is, component survival functions, component density functions and so on. The regression models for these functionals can be expressed in terms of the regression model for component hazards, in particular, the definition of the component survival function conditional on covariates is given by

$$\begin{aligned} S_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j) &= P(T > t|D = j; \boldsymbol{x}, \boldsymbol{\beta}_j) \\ &= \exp\Big(- \int_0^t \lambda_j(s|\boldsymbol{x}, \boldsymbol{\beta}_j) ds \Big) \end{aligned} \tag{3.1.1}$$

for $j = 1, \ldots, J$, where $S_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j)\,(j = 1, \ldots, J)$, is the probability that a subject that eventually failed due to failure type j failed after time t conditional on $\boldsymbol{x}$. It is important to note that the emphasis is on subjects *that eventually failed* as opposed to subjects *that failed*. Observed data consists of subjects that failed by cause j, the subjects that failed by other causes and censored subjects. Recall that the failure cause for each subject is determined at the outset. This means that amongst censored subjects there are subjects that are destined to fail due to failure cause j and others that will ultimately fail by other causes, albeit, unobserved. Thus, $\lambda_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j)\,(j = 1, \ldots, J)$, is the component hazard for observed failures due to failure type j and the censored subjects that will eventually fail by cause j conditional on $\boldsymbol{x}$. The failure type probability $\pi_j(\boldsymbol{x}, \boldsymbol{\gamma})$ is, therefore, the proportion of subjects that were assigned to fail due to failure type j conditional on $\boldsymbol{x}$. This is the proportion of subjects that are observed to have failed due to cause j and censored subjects that were destined by failure cause j, but unobserved conditional on $\boldsymbol{x}$.

As hinted in Chapter 2, the model has proved to be one of the most versatile models in the literature. For example, the mixture cure model is premised on this model. One of the notable features of the mixture competing risks model is a mixed population survival function, that is, a mixture of component survival functions $S_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j)\,(j = 1, \ldots, J)$, where the failure type probabilities $\pi_j(\boldsymbol{x}, \boldsymbol{\gamma})\,(j = 1, \ldots, J)$, are mixing weights;

$$S(t|\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{j=1}^{J} \pi_j(\boldsymbol{x}, \boldsymbol{\gamma}) S_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j) \tag{3.1.2}$$

with $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$. The univariate cure mixture model assumes that the population is split into two subpopulations, one for cured subjects and another one for uncured subjects where this split is characterized via the population survival function as given by

$$S(t|\boldsymbol{x}, \boldsymbol{\theta}) = \pi(\boldsymbol{x}, \boldsymbol{\gamma}) S_u(t|\boldsymbol{x}, \boldsymbol{\beta}) + (1 - \pi(\boldsymbol{x}, \boldsymbol{\gamma})). \tag{3.1.3}$$

where $S_u(t|\boldsymbol{x}, \boldsymbol{\beta})$, is the survival function for uncured subjects. The proportion of uncured subjects conditional on $\boldsymbol{x}$, is $\pi(\boldsymbol{x}, \boldsymbol{\gamma})$ and $1 - \pi(\boldsymbol{x}, \boldsymbol{\gamma})$, is the proportion of cured subjects conditional on covariates. As alluded to in Chapter 2, cured subjects are not limited to univariate settings, there may be instances when some subjects may not fail or may exhibit an over extended survival experience in situations where there are multiple risks of failure. The survival function expression given in (3.1.2) assumes that all J subpopulations are uncured. To reflect the presence of cured subjects, it is assumed that there exist an additional subpopulation, the $(J{+}1)^{\text{th}}$ subpopulation, which consists of cured subjects. In the presence of cured subjects in competing risks settings, the unconditional survival function is now given by

$$S(t|\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{j=1}^{J} \pi_j(\boldsymbol{x}, \boldsymbol{\gamma}) S_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j) + 1 - \sum_{j=1}^{J} \pi_j(\boldsymbol{x}, \boldsymbol{\gamma}). \tag{3.1.4}$$

Note that (3.1.4) is the generalization of the mixture cure model for J failure causes, when $J = 2$, then (3.1.4) reduces to (3.1.3), the expression for the survival function which characterizes the univariate mixture cure model. We revisit this conversation in Chapter 7 where we discuss the analysis of competing risks data that has a sizable proportion of cured subjects in discrete time.

Another notable feature of the mixture model is an alternate and simpler CIF regression expression;

$$F_j(t|\boldsymbol{x}, \boldsymbol{\theta}) = \pi_j(\boldsymbol{x}, \boldsymbol{\gamma})(1 - S_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j))$$

for $j = 1, 2, \ldots, J$. This result follows immediately from the underlying assumption of the mixture model in relation to the decomposition of the bivariate distribution of $(T; D)$, that is;

$$P(T; D) = P(T|D)P(D).$$

It follows from this assumption that;

$$\begin{aligned} F_j(t) &= P(T \leq t, D = j) \\ &= P(T \leq t|D = j)P(D = j) \\ &= \pi_j(1 - S_j(t)). \end{aligned}$$

This concludes the review of the mixture model as advanced by its authors. We now attend to the estimation of $\boldsymbol{\theta}$, the vector of parameters that describe the mixture model. The MLE of $\boldsymbol{\theta}$, can be determined by applying the standard estimation methods such as the Newton-Raphson method, but the full likelihood function can prove to be intractible at times. Instead, we will demonstrate that the application of an appropriately formulated Expectation Maximazation (EM) Algorithm (Dempters et al., 1977) simplifies the estimation procedure because the full likelihood function splits into J component failure time likelihood functions and a failure type likelihood function which are specified in terms of component hazards and failure type probabilities, respectively. This means that $\boldsymbol{\gamma}$, and $\boldsymbol{\beta}_j$ $(j = 1, \ldots, J)$, are estimated seperately. In the next section, Subsection 3.1.2, we provide a detailed presentation of this EM Algorithm. In the same section we discuss the adjustments that are required such that the proposed model can be applied in discrete time. Note that when observed data is modeled with the mixture model the popular CSHs are no longer estimated directly from the data, but their estimates can be recovered from;

$$\hat{h}_j(t|\boldsymbol{x}, \hat{\boldsymbol{\theta}}) = \frac{\hat{F}_j(t|\boldsymbol{x}, \hat{\boldsymbol{\theta}}) - \hat{F}_j(t\text{-}1|\boldsymbol{x}, \hat{\boldsymbol{\theta}})}{\hat{S}(t\text{-}1|\boldsymbol{x}, \hat{\boldsymbol{\theta}})}.$$

Clearly, assessing the covariate effects on CSHs from regression coefficients is no longer possible when data is modeled with the mixture model due to the commplicated expression for the CSHs. That advantage is forfeited when data is no longer modeled with CSHs, but one of the benefits lies in a simpler regression model for the CIF when data is characterized via the parameters of the mixture model, namely; the failure type probabilities and component hazards. It is possible to infer covariate effects on the CIF under certain conditions from the regression coefficents in the corresponding models for failure type probabilities and component hazards as will be discussed when we illustrate the application of the proposed model in Subection 3.1.3. We also apply the multinomial model in Section 3.1.3 for comparison purposes to the proposed model. In Section 3.2 we upscale the proposed model into missing failure causes model. We follow this up with a demonstration of this model by applying it to to discrete time data that comes with missing failure causes in Subsection 3.2.1. We conclude this chapter with a discussion in Section 3.3. The derivation of standard errors for the CIF estimates is conducted in Appendix A. In particular we attend to the adjustment of $\mathrm{V}(\hat{\boldsymbol{\theta}})$, because $\hat{\boldsymbol{\theta}}$, was determied via the application of an EM Algorithm

3.1.2 The EM Algorithm

The unknown parameter vector $\boldsymbol{\theta}$, is estimated by maximizing the observed data likelihood function w.r.t. $\boldsymbol{\theta}$. In constructing a likelihood function from observed data, a subject i that fails at time t_i from failure type j, contributes $P(D_i = j|\boldsymbol{x}_i, \boldsymbol{\gamma})P(T_i = t_i|D_i = j; \boldsymbol{x}_i, \boldsymbol{\beta}_j)$, to the likelihood function while a censored one contributes $P(T_i > t_i|\boldsymbol{x}_i; \boldsymbol{\theta})$, at time t_i. The full log-likelihood function can be written as;

$$\mathcal{L}_0(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log\big(P(D_i = j|\boldsymbol{x}_i, \boldsymbol{\gamma}_j)P(T_i = t_i|D_i = j; \boldsymbol{x}_i, \boldsymbol{\beta}_j)\big)$$
$$+ (1 - d_i)\log P(T_i > t_i|\boldsymbol{x}_i; \boldsymbol{\theta})$$

where $d_{ij} = I(D_i = j)$, and $d_i = \sum_{j=1}^{J} d_{ij}$. To justify the application of an EM algorithm, we pose this log-likelihood function as a missing information problem regarding the eventual failure status for censored subjects. For each censored subject i, we introduce a pseudo variable $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{iJ})^T$ $(i = 1, \ldots, n)$, where v_{ij}, is 1 or 0 according to whether a censored subject i eventually fails by cause j or not. Assuming that $\boldsymbol{v}_i$, was observed, the expression for the complete data log-likelihood function can then be written as;

$$\mathcal{L}_c(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log P(D_i = j|\boldsymbol{x}_i, \boldsymbol{\gamma})P(T_i = t_i|D_i = j; \boldsymbol{x}_i, \boldsymbol{\beta}_j)$$
$$+ (1 - d_i)v_{ij}\log P(D_i = j|\boldsymbol{x}_i, \boldsymbol{\gamma})P(T_i > t_i|D_i = j; \boldsymbol{x}_i, \boldsymbol{\beta}_j).$$

With appropriate subtitutions, $\mathcal{L}_c(\boldsymbol{\theta})$, can be written as;

$$
\begin{aligned}
L_c(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}\log \pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma})\lambda_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)\big)S_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)\\
&+ (1-d_i)v_{ij}\log \pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma})S_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{J}(d_{ij}+(1-d_i)v_{ij})\log \pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma}) + d_{ij}\log\lambda_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)\big)\\
&+ (d_{ij}+(1-d_i)v_{ij})\log S_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{J} g_{ij}\log \pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma}) + d_{ij}\log\lambda_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)\big) + g_{ij}\log S_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)
\end{aligned}
$$

where $g_{ij} = d_{ij} + (1-d_i)v_{ij}$. In the E-Step, we determine the expectation of the complete data log-likelihood function conditional on $\boldsymbol{y}$, and $\boldsymbol{\theta}^{(r)}$, where $\boldsymbol{\theta}^{(r)}$, is MLE of $\boldsymbol{\theta}$, in the M-Step of the $(r)^{\text{th}}$ iteration. Since g_{ij} is linear in the complete data log-likelihood function, the E-step entails replacing g_{ij} with $g_{ij}^{(r)} = d_{ij}+(1-d_i)v_{ij}^{(r)}$, where the conditional expectation $v_{ij}^{(r)}$ of v_{ij} is given by

$$
v_{ij} = \mathrm{E}(v_{ij}|\boldsymbol{\theta}^{(r)},\boldsymbol{y}) = \frac{\pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma}^{(r)})S_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j^{(r)})}{\sum_{l=1}^{J}\pi_l(\boldsymbol{x}_i,\boldsymbol{\gamma}^{(r)})S_l(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_l^{(r)})}.
$$

Introducing the Q notation, the conditional expectation of the complete data log-likelihood function can be written as;

$$
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) = \sum_{i=1}^{n}\sum_{j=1}^{J} g_{ij}^{(r)}\log \pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma}) + d_{ij}\log\lambda_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j) + g_{ij}^{(r)}\log S_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j).
$$

Let ϵ_{is} represent the exposure or the time spent alive in the interval $[a_s, a_{s\text{-}1})$ by subject i, and define it as $\epsilon_{is} = (\min(s,a_s) - a_{s\text{-}1})$. Thus, a subject i that fails or is censored at time s during this interval has $\epsilon_{is} = s - a_{s\text{-}1}$ as its exposure, otherwise $\epsilon_{is} = a_s - a_{s\text{-}1}$ if it survives the interval. With these definitions, the component survival function can then be written as;

$$
\begin{aligned}
S_j(t|\boldsymbol{x};\boldsymbol{\beta}_j) &= \exp(-\Lambda_j(t|\boldsymbol{x},\boldsymbol{\beta}_j))\\
&= \exp -\big(\sum_{s=1}^{t}\epsilon_{is}\lambda_j(s|\boldsymbol{x},\boldsymbol{\beta}_j)\big).
\end{aligned}
$$

Furthermore, if we define $d_{ijs} = 0$ for $s \leq t_i$-1 and $d_{ijt_i} = d_{ij}$, as well as; $g_{ijs} = g_{ij}$ for $s = 1, \ldots, t_i$, the E-Step can be written as a sum of;

$$Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}) = \sum_{i=1}^{n}\sum_{j=1}^{J} g_{ij}^{(r)} \log \pi_j(\boldsymbol{x}_i, \boldsymbol{\gamma})$$

and,

$$\begin{aligned} Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(r)}) &= \sum_{j=1}^{J}\Big\{\sum_{i=1}^{n}\sum_{s=1}^{t_i} d_{ijs}\log\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j) - g_{ijs}^{(r)}\epsilon_{is}\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j)\Big\} \\ &= \sum_{j=1}^{J} Q_(\boldsymbol{\beta}|\boldsymbol{\beta}^{(r)}). \end{aligned}$$

It can easily be seen that $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)})$ is a kernel of a multinomial log-likelihood function with $(g_{i1}^{(r)}, \ldots, g_{iJ}^{(r)}) \sim \mathcal{M}(1, \pi_1(\boldsymbol{x}_i, \boldsymbol{\gamma}), \ldots, \pi_J(\boldsymbol{x}_i, \boldsymbol{\gamma}))$, for subject i. Furthermore, in likelihood, $Q_j(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^{(r)})$, is equivalent to;

$$\acute{Q}_j(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^{(r)}) = \sum_{i=1}^{n}\sum_{s=1}^{t_i} d_{ijs}\log(g_{ijs}\epsilon_{is}) + d_{ijs}\log\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j) - g_{ijs}^{(r)}\epsilon_{is}\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j)$$

where $d_{ijs} \sim \mathcal{P}(g_{ijs}^{(r)}\epsilon_{is}\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j))$, because the difference is $d_{ijs}\log(g_{ijs}^{(r)}\epsilon_{is})$, a constant term which vanishes upon differentiation of $\acute{Q}_j(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^{(r)})$, w.r.t. $\boldsymbol{\beta}_j$ in the M-Step. Therefore, the unknown parameters of the mixture model that correspond to the component hazards, with piece-wise constant baseline component hazards, can be estimated via a Poisson regression model ($d_{ijs} \sim \mathcal{P}(g_{ijs}^{(r)}\epsilon_{is}\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j)))$, with $\log(g_{ijs}^{(r)}\epsilon_{is})$ as an offset. In continuous time, even though the survival times are grouped into intervals, the exposure for each subject can be computed exactly because failure/censored times are known, but this is not the case when survival times are grouped into intervals in discrete time. In the absence of information regarding failure/censoring times, we assume that failures occur halfway through the interval, while, censoring is assumed to occur at the end of the interval. If the failure time for subject i is t_i, meaning that $a_{t\text{-}1} \leq t_i < a_t$, its total exposure is:- $\sum_{s=1}^{t_i\text{-}1}(a_s - a_{s\text{-}1}) + \frac{1}{2}(a_t - a_{t\text{-}1})$, otherwise, if t_i is a censoring time, the total exposure is:- $\sum_{s=1}^{t_i}(a_s - a_{s\text{-}1})$. This is the required adjustment to enable the model to handle discrete time competing risks data while the model itself remains intact.

The M-step entails maximizing J Poisson log-likelihood functions $\tilde{Q}_j(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^{(r)})$ $(j = 1, \ldots, J)$, individually within the GLM framework and a multinomial likelihood $Q_0(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)})$. Because some of the statistical packages cannot handle a multinomial distribution with fractional responses, an alternative is to fit J-1 binomial distributions of the form, $g_{ij} \sim \mathcal{B}(1, \pi_j(\boldsymbol{x}_i, \boldsymbol{\gamma}_j))$, with the most prevalent failure type as the reference category to minimize the standard errors (Begg and Gray, 1984). To perform the M-step, the failure time data is first rearranged into a long format. The `discSurv` **R** package (Welchowski and Schmid, 2019) can be used to facilitate the conversion of data into a long format.

3.1.3 Application

To demonstrate the application of the proposed model, we consider the unemployment data that was originally analyzed by McCall (1996). This data is given as *UnempDur* in `Ecdat` (Croissant and Graves, 2020) **R** package. The data focusses on the length of the unemployment spell (two-week intervals) until transition to either full-time or part-time employment for unemployed individuals that recently lost their jobs. The covariates in the data set are Age, Unemployment Insurance, Displacement rate, Replacement rate, Wage rate and Tenure at the lost job.

Out of a sample of 3343 subjects, 676 are excluded for lack of complete information. Eventually out of the final sample of size 2667, about 40% of the subjects exit to full-time employment, 13% to part-time employment, and 47% are censored.

The time in two-week intervals is $t \in \{1, 2, 3, \ldots, 28\}$. We consider $t \in \{1, 2, 3, \ldots, 19\}$, by collapsing the event/censoring time $t \geq 19$ into one interval $[19, 29)$ to minimize instability of parameter estimates because there are few events per timepoint beyond $t = 19$. With the exception of the unemployment benefits (UI) covariate, all covariates are continuous. For the purposes of computing exposure, an interval length is taken to be 1 so that a failure time and censoring time have 0.5 and 1 exposure, respectively, and this also applies to the interval $[19, 29)$.

This data was analysed by McCall (1996) to test the proposition that increasing the Disregard Rate for the Unemployment Insurance (UI) recipients will entice these individuals towards part-time employment away from full-time employment with the net effect of a rise in overall employment. Unemployed individuals are allowed to continue to receive full unemployment benefits while employed on part-time basis provided the benefits do not exceed a certain rate, the Disregard Rate. Replacement rate is the proportion of benefits relative to last wage earned in previous employment. We also apply the proposed model to test this claim as well.

We illustrate the application of the proposed model by demonstrating an established fact in econometrics that provision of unemployment benefits tends to discourage unemployed individuals from seeking out employment opportunities and thereby fuelling unemployment. The two theories that support this view are *labour-supply* and *job search*. As we advance various discrete time models in this thesis we will be relying on this unemployment data to demonstrate the application of these models.

Furthermore, to illustrate the application of various regression models that will be suggested throughout this thesis we will consider the effect of increasing the disregard rate on re-employment prospects for unemployment benefit recipients as well as the effect of providing unemployment benefits to unemployed individuals on overall re-employment rates. Throughout this thesis, we will set all the covariates that came with the data set as explanatory variables. We will regard Unemployment Insurance recipients (`ui`=1) as a reference category, while all the continuous covariates, that is, Disregard Rate (`dr`), Replacement Rate (`rr`), age (`age`), Logwage (`wage`) and tenure (`tenure`) will be centered at their respective averages.

We have referred to full-time employment and part-time employment as cause 1 and cause 2, respectively. Full-time employment is our cause of interest.

Table 3.1: Maximum likelihood estimates for the Mixture Model and the Multinomial Model (with standard errors) (* denotes $P < 0.05$).

	Mixture Model (Latency Model)		Multinomial Model	
Coefficient	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
T1	-2.583(0.074)*	-2.277(0.124)*	-2.694(0.077)*	-3.823(0.130)*
T2	-2.796(0.085)*	-2.677(0.150)*	-2.989(0.088)*	-4.173(0.153)*
T3	-2.936(0.098)*	-2.925(0.176)*	-3.181(0.102)*	-4.381(0.179)*
T4	-3.468(0.138)*	-3.314(0.227)*	-3.762(0.141)*	-4.766(0.229)*
T5	-2.655(0.103)*	-2.655(0.180)*	-2.926(0.108)*	-4.045(0.182)*
T6	-3.608(0.179)*	-3.714(0.322)*	-3.949(0.183)*	-5.134(0.324)*
T7	-2.436(0.112)*	-2.699(0.212)*	-2.747(0.119)*	-4.029(0.215)*
T8	-3.926(0.260)*	-3.484(0.338)*	-4.328(0.263)*	-4.857(0.341)*
T9	-3.025(0.176)*	-3.643(0.383)*	-3.411(0.182)*	-4.985(0.385)*
T10	-5.267(0.578)*	-4.353(0.580)*	-5.704(0.580)*	-5.731(0.582)*
T11	-2.989(0.198)*	-3.760(0.451)*	-3.394(0.205)*	-5.082(0.454)*
T12	-4.134(0.379)*	-4.130(0.580)*	-4.580(0.383)*	-5.473(0.583)*
T13	-2.740(0.202)*	-3.062(0.359)*	-3.150(0.210)*	-4.340(0.363)*
T14	-2.318(0.184)*	-3.174(0.413)*	-2.733(0.196)*	-4.403(0.418)*
T15	-2.461(0.231)*	-3.676(0.581)*	-2.956(0.243)*	-4.850(0.585)*
T16	-2.871(0.317)*	-3.496(0.581)*	-3.421(0.329)*	-4.660(0.587)*
T17	-2.835(0.354)*	-4.405(1.002)*	-3.417(0.367)*	-5.563(1.006)*
T18	-2.695(0.379)*	-3.588(0.710)*	-3.366(0.392)*	-4.697(0.717)*
T19	-1.211(0.214)*	-1.920(0.340)*	-1.823(0.242)*	-2.804(0.357)*
ui	1.505(0.065)*	0.556(0.116)*	1.184(0.068)*	1.199(0.119)*
dr	-0.447(0.504)	-1.729(0.835)*	-1.731(0.529)*	-0.591(0.821)
rr	0.480(0.456)	-0.025(0.784)	0.900(0.464)*	-0.334(0.742)
age	-0.011(0.003)*	-0.008(0.006)	-0.015(0.004)*	-0.003(0.006)
wage	0.227(0.096)*	0.039(0.150)	0.536(0.099)*	-0.390(0.149)
tenure	-0.014(0.006)*	0.044(0.011)*	0.002(0.006)*	0.003(0.011)

Table 3.2: Maximum likelihood estimates for the Mixture Model (with standard errors) (* denotes $P< 0.05$).

Mixture Model	
(Incidence Model)	
Coefficient	$\hat{\gamma}$
Constant	1.178(0.062)*
`ui`	-0.488(0.092)*
`dr`	-2.381(0.674)*
`rr`	0.822(0.668)*
`age`	-0.009(0.005)*
`wage`	0.897(0.131)*
`tenure`	0.038(0.009)*

Recall that the model for component hazards is given by

$$\lambda_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j) = \exp(\beta_{0jt}) \exp(\boldsymbol{x}^T \boldsymbol{\beta}_{1j}) \quad a_{t\text{-}1} \leq t < a_t$$

where we have applied a Poisson regression model to estimate $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \ldots, \boldsymbol{\beta}_J^T)^T$, the component hazard parameters. Since $J = 2$, the model for the probability of failure due to cause 1 is given by

$$\pi_1(\boldsymbol{x}; \boldsymbol{\gamma}_1) = \frac{\exp\left(\gamma_{01} + \boldsymbol{x}^T \boldsymbol{\gamma}_{11}\right)}{1 + \exp\left(\gamma_{0j} + \boldsymbol{x}^T \boldsymbol{\gamma}_{11}\right)}.$$

We have displayed the analysis results in Table 3.1 and Table 3.2. In testing McCall (1996)'s proposition, for illustrative purposes we have assessed the effect of increasing the disregard rate by 50 %. The effect of increasing the disregard rate is assessed via its effect on the CIF. The CIF estimates are plotted in Figure 3.1. It is evident from the plot that increasing the disregard rate by 50% lowers the probability of exiting the unemployment state to full-time employment significantly, but the effect on part-time employment is marginal, if any. These findings are in agreement with McCall (1996) in so far as full-time re-employment prospects are concerned. While there is an increase in part-time re-employment, it is not as substantial as suggested by McCall (1996). McCall

(1996) used even more covariates as explanatory variables than those that came with the data set. This could be the explanation for the difference between our findings and what was found by McCall (1996) with regards to the effect of increasing the disregard rate on part-time re-employment. Another possible explanation is that we only used about 80 % of the data that was used by McCall (1996) because we had to discard some portion of data that has missing failure causes, that is, it is known for these subjects that they found employment, but the type of employment was not stated.

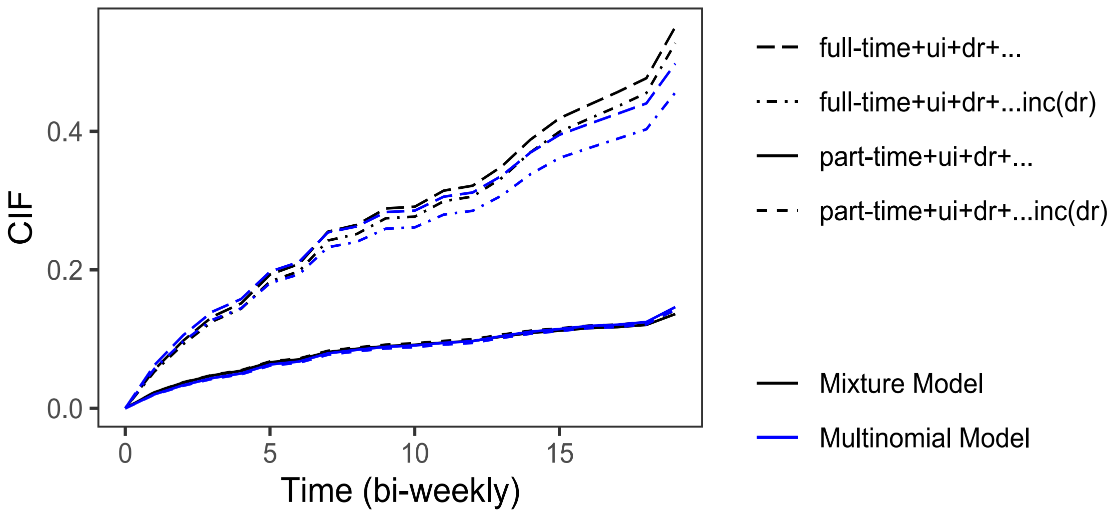

Figure 3.1: The CIF of exit to full-time and part-time employment for the `ui` recipients with the effect of increasing `dr` via the Mixture Model and the Multinomial Model.

This fact could also be the source of differences in the findings or the proposed model could be structurally flawed itself. There is very little that we can do to address the first two points, but we can, for example, fit a comparable model and contrast this model to the proposed model, particularly, in relation to the effect of increasing the disregard rate on part-time employment.

Towards that end, we fitted a multinomial model to the data. Recall that when the multinomial model is assumed the model for CSHs is given by

$$h_j(t|\boldsymbol{\alpha};\boldsymbol{x}) = \frac{\exp\left(\alpha_{0jt} + \boldsymbol{x}^T\boldsymbol{\alpha}_{1j}\right)}{1+\sum_{l=1}^{J}\exp\left(\alpha_{0lt} + \boldsymbol{x}^T\boldsymbol{\alpha}_{1l}\right)}$$

for $t = 1,\ldots,q$, and $j = 1,\ldots,J$, where $\boldsymbol{\alpha}_j = (\alpha_{0j1},\ldots,\alpha_{0jq},\boldsymbol{\alpha}_{1j}^T)^T$, and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^T,\ldots,\boldsymbol{\alpha}_J^T)^T$. For cause j, the scalar α_{0jt} represents the baseline cause-specific hazard coefficient at time t and $\boldsymbol{\alpha}_{1j}$ is a vector of regression coefficients. We have listed the results of fitting a multinomial model in Table 3.1 as well. We have plotted the CIF estimates from the multinomial model in Figure 3.1. Recall that when the multinomial model is assumed, the CIF estimates are derived from cause-specific-hazard estimates;

$$\hat{F}_j(t) = \sum_{s=1}^{t}\hat{S}(s\text{-}1|\boldsymbol{x},\hat{\boldsymbol{\alpha}})\hat{h}_j(s|\boldsymbol{x},\hat{\boldsymbol{\alpha}}).$$

It is clear from Figure 3.1 that the multinomial model is in agreement with the proposed model regarding the effect of increasing the disregard rate by 50% on full-time employment. There is a reduction in part-time re-employment, but still insignificant. In fact, it could be argued that there is no movement in part-time employment as a result of adjusting the disregard rate upwards by 50% via the proposed model or the multinomial model. According to the proposed mixture model, amongst a group of unemployed individuals who are in receipt of unemployment benefits with average values for continuous covariates, about (full-time:55.1%+ part-time:13.6%) 68.7% of these individuals found employment, but if we raise the disregard rate for the same group by 50% and hold everything else constant, about (full-time:52.7%+ part-time:14.1%) 66.8% of the individuals are re-employed. The multinomial model, on the other hand, suggests that for the reference group about (full-time:49.8%+ part-time:14.6%) 64.4% of these individuals are re-employed and raising the disregard rate, but holding everything else constant reduces re-mployment to (full-time:46.7%+ part-time:14.5%) 61.2%.

These results emphasize the fact that the proposed model and the multinomial model agree in so far as the effect of raising the disregard rate on re-employment, both models suggest that the effect of raising the disregard rate is to reduce full-time re-employment and no impact on part-time re-employment with the a net effect of a marginal reduction in total re-employment. Thus, we can safely say that the differences between our findings and McCall (1996) regarding the effect of increasing the disregard rate by 50% on part-time employment is not due to the inadequacies of the proposed model itself. That is is to say, putting aside the differences between our findings and McCall (1996), the proposed model compares favourably with an established discrete time competing risks model by the way of the multinomial model. This view is further supported by the plot of the CIF estimates via the proposed model and the multinomial model in Figure 3.1.

We did allude to the fact that it is possible to infer covariate effects on the CIF under certain conditions. Recall that the regression model for the CIF under the mixture model is given by

$$F_j(t|\boldsymbol{x}, \boldsymbol{\theta}) = \pi_j(\boldsymbol{x}, \boldsymbol{\gamma})(1 - S_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j)). \tag{3.1.5}$$

In this particular example we have two failure causes which means that the failure type probabilities were modeled via the binomial distribution. Since $\boldsymbol{\beta}_{\texttt{1dr}} = -0.447 < 0$, and $\boldsymbol{\gamma}_{\texttt{1dr}} = -2.381 < 0$, increasing the disregard rate has the effect of reducing the marginal probability of failure by cause 1 (full-time) as well as a reduction in the component hazards due to failure cause 1. This results in an increase in $S_1(t)$ and a reduction in $1-S_1(t)$ and ultimately a reduction in $\pi_1(1 - S_1(t)) = F_1(t)$, that is, a reduction in the the probability of full-time re-employment when everything else is held constant. Note that a reduction in π_1 due to an increase in the disregard rate implies an increase in π_2 because $\pi_2 = 1-\pi_1$. Since $\boldsymbol{\beta}_{\texttt{2dr}} = -1.729 < 0$, it is not possibe to predict the movement of the CIF with regard to part-time re-employment prospects engendered by an increase in the disregard rate. Clearly, it is only possible to predict the effect of a covariate on the CIF provided the regression coefficients of the covariate in the models for the failure type probability and the component hazard have the same signs. Put differently, it is only possible to predict the effect of a covariate on the CIF if the covariate induces the failure type probability and the component hazards to move in the same direction.

It was possible to predict the effect of increasing the disregard rate on the full-time CIF because there were only two failure types in the data. When $J > 2$ things become somewhat complicated. While it is always possible to assess the covariate effects on component hazards regardless of the value of J, the number of failure modes, the same cannot be said for marginal failure type probabilities. Suppose that we are interested in assessing the effect of covariate x_l on the CIF and $J > 2$. Recall that the failure type distribution is modeled via a multinomial distribution, that is, the failure type probabilities are connected to data via a multinomial model. Therefore, to assess the effect of x_l on π_j, for example, we will have to assess the effect of regression coefficients; $\gamma_{1x_l}, \gamma_{2x_l} \ldots, \gamma_{(J\text{-}1)x_l}$, simulteneously on π_j which becomes rather complicated. This is the same challenge that is encounterd when one wishes to assess the effect of a covariate on CIF from cause-specific hazard regression coefficients when the regression model for the CIF is derived from CSHs as given in (3.1.5). To assess the effect of x_l on the CIF one has to contend with J regression coefficients that correspond to x_l; $\alpha_{1x_l}, \alpha_{2x_l} \ldots, \alpha_{Jx_l}$. Recall that we did highlight the fact that existing packages may not have a multinomial routines that can deal with fractional responses. To avoid additional programming, one can fit J-1 binomial distributions to estimate failure type probabilities individually because the binomial distribution routines by most packages can handle fractional responses. Obviously, this solves our problem because we can then assess the effect of x_l on π_j from γ_{jx_l} without having to worry about the confounding effect of other regression coefficients.

We did state that we will demonstrate the application of the proposed model by testing McCall (1996)'s proposition as well as the well documented fact in econometrics that provision of unemployment benefits tends to have an unintended consequence of discouraging recipients from intensifying the search for jobs and thereby inducing a reduction in re-employment figures. In actual fact McCall (1996)'s proposition is the third way between provision of this benefit at the risk of lower rates of employment and the other extreme whereby the provision of unemployment benefits is discontinued at the expense of denying financial support for deserving unemployed individuals. We can see from Table 3.1 and Table 3.2 that $\boldsymbol{\gamma}_{1\texttt{ui}} = -0.488 < 0$ and $\boldsymbol{\beta}_{1\texttt{ui}} = 1.505 > 0$. Recall that the reference category for `ui` is benefit recipients. Moving from recipients to non-recipients induces a reduction in the failure type probability and an increase component hazards.

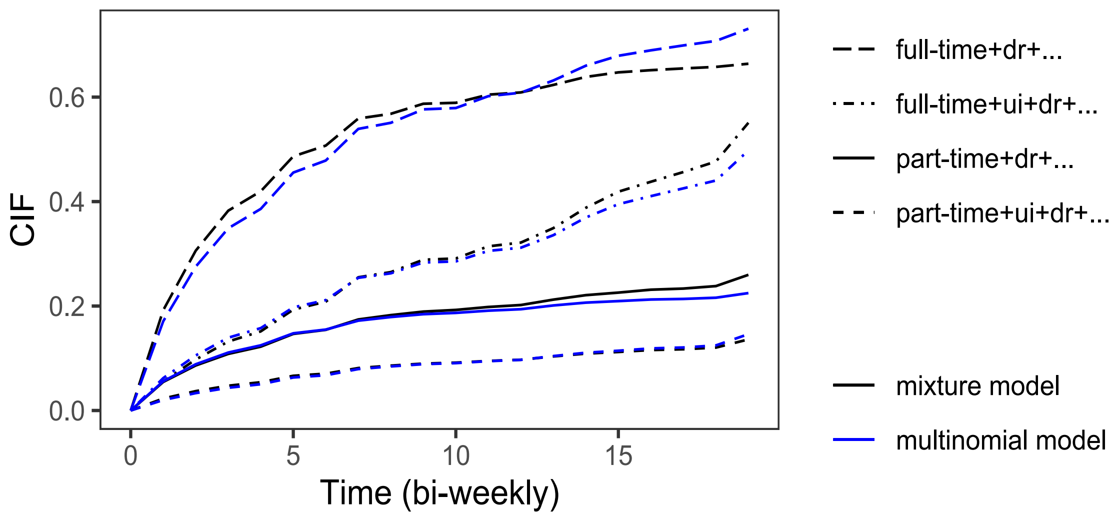

Figure 3.2: The CIF of exit to full-time and part-time employment with the effect of`ui` via the Mixture Model and the Multinomial Model.

Therefore, we cannot predict the effect of `ui` on the CIF of exit to full-time employment. Obviously, a decrease in π_1 due to `ui` induces an increase in π_2. Now, since $\boldsymbol{\beta}_{2\texttt{ui}} = 0.556 > 0$, `ui` also induces part-time component hazards to increase and the net effect is that the CIF of exit to part-time employment is larger for non-recipients of `ui` compared to recipients. An unemployed individual who does not receive the benefits is more likely to find part-time employment than a non-recipients to find part-time employment.

In Figure 3.2 we have plotted the CIF estimates from the proposed mixture model and the multinomial model. An examination of either plot leads to the same conclusion that provision of unemployment benefits have an unintended consequence of discouraging unemployed individuals from doubling up their efforts to find employment, both part-time and full-time employment, with a net effect of a rise in unemployment rates. These findings via the proposed model are consistent with the theory as it relates to the effect of unemployment benefits on employment. More importantly though, this plot serves as further evidence that the proposed mixture model compares favourably with the multinomial model because both models lead to the same conclusion regarding the impact of unemployment benefits on re-employment prospects as suggested by the theory on this subject. In the next section we will upgrade this model into a missing failure causes model. The hope is that when we apply this model to data including the subjects with missing failure causes our findings will be closer to McCall (1996)

3.2 The Missing Failure Causes Model

Recall that this thesis is divided into three parts or objectives. The first obective is to advance additional discrete time competing risks models as alternatives to the existing discrete time models, that is, the multinomial model and the binomial model. In the previous section we demonstrated how the mixture model (Larson and Dinse, 1985) as suggested by its authors can be applied in discrete time. The second part of the thesis is dedicated towards advancing discrete time models that can handle data that comes with missing failure causes. In this section we attend to the second part of the objectives by upscaling the model proposed in the previous sestion into a missing failure causes model.

Missing failure causes is a data complication that arises in competing risks settings when causes of failure for some subjects are not recorded. There are a number of reasons that may lead to missing failure causes. Anderson et al. (1996) cite examples in clinical trials where recording clerks may neglect to capture known death causes for some subjects or the cause of death may be difficult to determine due to lack of appropriate intrumentation as some of the reasons that may lead to missing failure causes. This topic has not received adequate attention in discrete time. As already highlighted in Chapter 2, the existing discrete time models are not capable of dealing with this data complication. Of course, the reason is that both the multinomial model and the binomial model are CSH denominated models. One of the options to get around this problem is to remove the affected subjects or to create an additional failure category for these subjects before applying the existing discrete time models. Obviously, this course of action is not ideal because the estimates for

CSHs may be biased downwards (Anderson et al., 1996). A more preferable alternative is to consider the existing discrete time models within the model based *Multiple Imputation Model* (MI) framework (Rubin, 1987). Bakoyannis et al. (2010) have undertaken similar work in continuous time where the authors have fitted the subhazard regression model for the CIF that was suggested by Fine and Gray (1999) within the MI framework. By existing discrete time models we also include the discrete time version of the subhazard regression model for the CIF that was suggested by Berger et al. (2020). Thus, taking our queue from Bakoyannis et al. (2010), we can fit all existing discrete time models within the MI framework to model data that comes with missing failure causes. The model suggested by Bakoyannis et al. (2010) is one of a number of models that have been advanced over the years to address the analysis of competing risks data that has missing failure causes in the continuous time domain. Dinse (1982) is often credited as one of the earliest authors to discuss this topic in continuous time. Since then, a number of authors have contributed towards this topic, see for example, (Dewanji, 1992; Goetghebeur and Ryan, 1990, 1995; Lu and Tsiatis, 2001; Nicolaie et al., 2015). For obvious reasons, these models are not appropriate for application in discrete time as these models are inherently continuous time models.

When Larson and Dinse (1985) introduced their model the authors also suggested that the model can easily be upgraded to handle missing failure causes. It is easy to see the rationale behind the authors' proposal. When we advanced this model as a model that can also be applied to discrete time competing risks data in the previous section, we demonstrated the application of the EM Algorithm in the estimation of the model parameters. Recall that it was assumed that the failure causes for all subjects were set at the outset including the failure causes for censored subjects. Amongst other things, the EM Algorithm was used to fractionate the censored subjects amongst failure causes. In the presence of subjects that have missing failure causes we can expand the same EM Algorithm to include the subjects that have missing failure causes and split these subjects amongst the failure causes as well. In fact, that is going to be our strategy as we extend the ordinary continuous time mixture model into missing failure causes model in this section. Towards that end we continue to assume the same models for component hazards and failure type probabilities. In actual fact upgrading the ordinary mixture model into a missing failure causes reduces to the estimation the component hazards and failure type probabilities in the presence of subjects with missing failure causes. In the previous section we assumed that the failure type distribution can be modeled with a multinomial distribution where the model for failure type probabilities is given by

$$\pi_j(\boldsymbol{x}, \boldsymbol{\gamma}) = \exp(\gamma_{0j} + \boldsymbol{x}^T \boldsymbol{\gamma}_{1j}) \Big/ (1 + \sum_{l=1}^{J-1} \exp(\gamma_{0l} + \boldsymbol{x}^T \boldsymbol{\gamma}_{1l})$$

The proportional hazards model for component hazards with piece-wise constant baseline component hazards, as assumed by the authors of the mixture model, is given by

$$\lambda_{0j}(t) = \exp(\beta_{0jt}) \qquad t \in [a_{t\text{-}1}, a_t),$$

so that,

$$\lambda_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j) = \exp(\beta_{0jt}) \exp(\boldsymbol{x}^T \boldsymbol{\beta}_{1j})$$

for $j = 1 \ldots, J$. The vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_1,^T, \boldsymbol{\beta}_2^T \ldots, \boldsymbol{\beta}_J^T)^T$, is a collection of all component hazard parameters and vector $\boldsymbol{\theta} = (\boldsymbol{\gamma}^T, \boldsymbol{\theta}^T)^T$, captures all the parameters of the mixture model that require estimation. Modelling data directly with CSHs has the advantage that the covariate effects on the risk of failure, as characterized via the CSHs, can be read directly from the estimates of regression coefficients. Of course, this is only possible if data is modeled with the binomial model (Lee et al., 2018) in discrete time. This advantage is forfeited when data comes with missing failure causes because, as already stated, the CSHs cannot be applied to data that has missing failure causes. Instead, the estimates for these quantities can be recovered from CIF estimates;

$$\hat{h}_j(t|\boldsymbol{x}, \hat{\boldsymbol{\theta}}) = \frac{\hat{F}_j(t|\boldsymbol{x}, \hat{\boldsymbol{\theta}}) - \hat{F}_j(t\text{-}1|\boldsymbol{x}, \hat{\boldsymbol{\theta}})}{\hat{S}(t\text{-}1|\boldsymbol{x}, \hat{\boldsymbol{\theta}})}.$$

Critically, this means that this is the first model which offers the first alternative route for estimating the CSHs in the presence of missing failure causes, albeit, without the advantage of assessing covariate effects from regression coefficients.

To determine $\hat{\boldsymbol{\theta}}$, we continue to implement an EM Algorithm in the presence of missing failure causes. As will be demonstrated in the following subsection, the full likelihood function also splits into a failure type likelihood function and J component likelihood functions for failure time, that is, a failure time likelihood function for each failure type. This means that the component hazards and failure type probabilities continue to be estimated seperately even in the presence of missing failure causes. In the next subsection, Subsection 3.2.1, we go through a detailed presentation of the EM algorithm for estimation of $\boldsymbol{\theta}$. In Subsection 3.2.1 we illustrate the application of the proposed model to real discrete time competing risks data that has missing failure causes together with complete case analysis. For comparison purposes, we also fit a multinomial model within the MI framework. We conclude with a discussion in Section 3.3. The standard errors for the CIF estimate are derived in Appendix A as well.

3.2.1 The EM Algorithm

When data comes as a mixture of subjects with known and unknown failure causes an indicator variable, R is introduced where $R_i = 1$ when subject i has failed with a known failure cause or $R_i = 0$ when the failure cause is unknown. It is assumed that $R_i = 1$ for the censored subject i because the censoring status is always known. Observed data is now represented by $(t_i, \Delta_i, \boldsymbol{x}_i, R_i = 1)$, when a subject i has failed with a known failure cause, or $(t_i, \boldsymbol{x}_i, R_i = 0)$, if the failure cause is unknown. Additional to the standard assumption that censoring is conditionally independent of the joint distribution of failure type and failure time, we also assume that missingness does not depend on failure type, i.e., we assume that the failure cause are missing at random (MAR)(Rubin, 1976). In constructing the observed data likelihood function, a subject i that failed with a known failure cause j at time t_i normally contributes $P(T_i = t_i, D_j = j)$. Assuming the mixture model, this subjects now contributes $P(D_i = j)P(T_i = t_i|D_i = j)$. If the subject is censored, then its contribution $P(T_i > t_i)$ is written as $\sum_{j=1}^{J} P(D_i = J)P(T_i > t_i)$. If subject i failed with an unknown failure cause, its contribution is $P(T_i = t_i)$ and this contribution is written as $\sum_{j=1}^{J} P(D_i = j)P(T = t_i|D_j = j)$. With covariates, the observed data log-likelihood function can be written as;

$$\mathcal{L}_0(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log P(D_i = j; \boldsymbol{x}_i, \boldsymbol{\gamma}) P(T_i = t_i | D_i = j; \boldsymbol{x}_i, \boldsymbol{\beta}_j)$$
$$+ d_{i*} \log \sum_{j=1}^{J} P(D_i = j; \boldsymbol{x}_i; \boldsymbol{\gamma}) P(T_i = t_i | D_i = j; \boldsymbol{x}_i, \boldsymbol{\beta}_j)$$
$$+ (1 - d_i) \log \sum_{j=1}^{J} P(D_i = j; \boldsymbol{x}_i, \boldsymbol{\gamma}) P(T_i > t_i | D_i = j; \boldsymbol{x}_i, \boldsymbol{\beta}_j)$$

where $d_{ij} = I(D_i = j)$, $d_{i*} = I(D_i = *)$, indicates a missing failure cause, and $d_i = \sum_{j=1}^{J} d_{ij} + d_{i*}$. When data came as normal competing risks data without missing failure causes in the previous section, the observed data log-likelihood was regarded as incomplete in respect of censored subjects to justify the implementation of an EM Algorithm. Likewise here, $\mathcal{L}_0(\boldsymbol{\theta})$ is also regarded as incomplete, vis-a-vis censoring and missing failure causes. Accordingly, data is augmented with vector $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{ij})^T$, for a censored subject i where v_{ij} assumes values 1 or 0 when the subject eventually fails by cause j or not, and vector $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{ij})^T$, for a subject i that has a missing failure cause where u_{ij} assumes values 1 or 0 if the subject has actually failed by cause j or not. If the pseudo variables were actually observed, then the complete data log-likelihood function can be written as;

$$\mathcal{L}_c(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log \pi_j(\boldsymbol{x}_i; \boldsymbol{\gamma}) \lambda_j(t_i | \boldsymbol{x}_i, \boldsymbol{\beta}_j) S_j(t_i | \boldsymbol{x}_i, \boldsymbol{\beta}_j)$$
$$+ d_{i*} u_{ij} \log \pi_j(\boldsymbol{x}_i, \boldsymbol{\gamma}) \lambda_j(t_i | \boldsymbol{x}_i, \boldsymbol{\beta}_j) S_j(t_i | \boldsymbol{x}_i; \boldsymbol{\beta}_j) + (1 - d_i) v_{ij} \log \pi_j(\boldsymbol{x}_i, \boldsymbol{\gamma}) S_j(t_i | \boldsymbol{x}_i, \boldsymbol{\beta}_j)$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{J} (d_{ij} + d_{i*} u_{ij} + (1 - d_i) v_{ij}) \log \pi_j(\boldsymbol{x}_i; \boldsymbol{\gamma}) + (d_{ij} + d_{i*} u_{ij}) \log \lambda_j(t_i | \boldsymbol{x}_i, \boldsymbol{\beta}_j)$$
$$+ (d_{ij} + d_{i*} u_{ij} + (1 - d_i) v_{ij}) \log S_j(t_i | \boldsymbol{x}_i, \boldsymbol{\beta}_j).$$

After re-arranging few terms, $\mathcal{L}_c(\boldsymbol{\theta})$, can be written as;

$$\mathcal{L}_c(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{J} g_{ij} \log \pi_j(\boldsymbol{x}_i, \boldsymbol{\gamma}) + m_{ij} \log \lambda_j(t_i | \boldsymbol{x}_i, \boldsymbol{\beta}_j) + g_{ij} \log S_j(t_i | \boldsymbol{x}_i, \boldsymbol{\beta}_j)$$

where $g_{ij} = d_{ij} + d_{i*} u_{ij} + (1 - d_i) v_{ij}$, and $m_{ij} = d_{ij} + d_{i*} u_{ij}$. As in the previous section, let $\epsilon_{is} = (\min(s, a_s) - a_{s\text{-}1})$, represent the exposure for subject i in the interval $[a_{s\text{-}1}, a_s)$, at time s such that the exposure is $\epsilon_{is} = s - a_{s\text{-}1}$. for subject i that fails or is censored at time s during the interval, otherwise $\epsilon_{is} = a_s - a_{s\text{-}1}$, if it survives the interval. Consequently, the component survival function can be written as;

$$S_j(t|\boldsymbol{x};\boldsymbol{\beta}_j) = \exp\big(-\sum_{s=1}^{t}\epsilon_{is}\lambda_j(s|\boldsymbol{x};\boldsymbol{\beta}_j)\big).$$

Furthermore, we define $d_{ijs} = 0$ for $s \leq t_i$-1 and $d_{ijt_i} = d_{ij}$ as well as $d_{i*s} = 0$ for $s \leq t_i$-1 and $d_{i*t_i} = d_{i*}$, so that $m_{ijs} = d_{ijs} + d_{i*s}u_{ij} = 0$ for $s \leq t_i$-1 and $m_{ijt_i} = m_{ij}$. The complete data log-likelihood function can now be written as a sum of;

$$\mathcal{L}_c(\boldsymbol{\gamma}) = \sum_{i=1}^{n}\sum_{j=1}^{J} g_{ij}\log\pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma})$$

and,

$$\mathcal{L}_c(\boldsymbol{\beta}) = \sum_{j=1}^{J}\sum_{i=1}^{n}\sum_{s=1}^{t_i} m_{ijs}\log\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j) - g_{ijs}\epsilon_{is}\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j)$$

where $g_{ijs} = g_{ij}$ for $s = 1,\ldots,t_i$. To complete the E-Step at the $(r{+}1)^{\text{th}}$ iteration of the algorithm, g_{ij} and m_{ij} are replaced with $g_{ij}^{(r)} = d_{ij} + d_{i*}u_{ij}^{(r)} + (1-d_i)v_{ij}^{(r)}$, and $m_{ij}^{(r)} = d_{ij} + d_{i*}u_{ij}^{(r)}$, where $u_{ij}^{(r)}$ and $v_{ij}^{(r)}$, are respective expectations for u_{ij} and v_{ij}, conditional on $\boldsymbol{\theta}^{(r)}$, and $\boldsymbol{y}$, with $\boldsymbol{\theta}^{(r)}$ as the MLE of $\boldsymbol{\theta}$ in the M-Step of the previous r^{th} iteration. The conditional expectations for pseudo variables are respectively given by

$$v_{ij}^{(r)} = \frac{\pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma}^{(r)})S_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j^{(r)})}{\sum_{l=1}^{J}\pi_l(\boldsymbol{x}_i.\boldsymbol{\gamma}^{(r)})S_l(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_l^{(r)})}$$

and,

$$u_{ij}^{(r)} = \frac{\pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma}^{(r)})\lambda_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j^{(r)})}{\sum_{l=1}^{J}\pi_l(\boldsymbol{x}_i,\boldsymbol{\gamma}^{(r)})\lambda_l(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_l^{(r)})}.$$

Introducing the Q notation, the E-Step can be written as a sum of;

$$Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}) = \sum_{i=1}^{n}\sum_{j=1}^{J} g_{ij}^{(r)}\log\pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma})$$

and,

$$\begin{aligned} Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(r)}) &= \sum_{j=1}^{J}\Big\{\sum_{i=1}^{n}\sum_{s=1}^{t_i} m_{ijs}^{(r)}\log\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j) - g_{ijs}^{(r)}\epsilon_{is}\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j)\Big\} \\ &= \sum_{j=1}^{J} Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^{(r)}). \end{aligned}$$

Again, $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)})$, is recognizable as a kernel of a multinomial log-likelihood function. Note also that $Q_j(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^{(r)})$, is equivalent to

$$\acute{Q}_j(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^{(r)}) = \sum_{i=1}^{n}\sum_{s=1}^{t_i} m_{ijs}^{(r)} \log(g_{ijs}^{(r)}\epsilon_{is}) + m_{ijs}^{(r)} \log \lambda_j(s|\boldsymbol{x}_i, \boldsymbol{\beta}_j) - g_{ijs}^{(r)}\epsilon_{is}\lambda_j(s|\boldsymbol{x}_i, \boldsymbol{\beta}_j)$$

where $m_{ijs}^{(r)} \sim \mathcal{P}(g_{ijs}^{(r)}\epsilon_{is}\lambda_j(s|\boldsymbol{x}_i, \boldsymbol{\beta}_j))$, up to a constant term; $m_{ijs}^{(r)} \log(g_{ijs}^{(r)}\epsilon_{is})$. The MLE of $\boldsymbol{\beta}_j$ for $j = 1, \dots J$, can, therefore, be determined by applying J Possion distributions to observed data in the long format with $\log(g_{ijs}^{(r)}\epsilon_{is})$ as an offset term.

Again, to apply the proposed model to discrete time data we assume that failures occurred halfway through the interval and censoring is assumed to have occured at the end of the interval.

3.2.2 Application

We apply the proposed model to the same data set *Unemployment data* that was used in the previous section, but we also include 574 individuals that have unknown failure causes such that the total size of the sample is now 3241.

Recall that when we illustrated the application of the proposed *continuous time mixture model* in the previous section we tested the assertion by McCall (1996) that increasing the disregard rate for unemployment benefit recipients will actually encourage these individuals to seek out part-time employment away from full-time employment such that overall employment improves. Recall also that we found in favor of McCall (1996) regarding full-time employment, that is, increasing the disregard rate induces a reduction in full-time employment, but our findings were in disagreement with McCall (1996) in relation to part-time employment.

We found that increasing the disregard rate almost had no impact on part-time employment. We also found that increasing the disregard rate induced a reduction in total re-employment. We highlighted the fact that we used about 80% of the data that was used by McCall (1996). Furthermore, McCall (1996) used more explanatory variables than the variables that came with the data set. These two points could be the reason that we came to different conclusions to McCall (1996) when we applied the ordinary *continuous time mixture model*. Another possibility is that the proposed model could be structurally flawed.

To eliminate the possibility that the proposed model was defective, we fitted a multinomial model for comparison purposes and found that the application of the proposed model and the multinomial model lead to the same conclusions, that is, increasing the disregard rate induces a reduction in full-time employment and almost no movement in part-time employment. There is nothing that we can do about the fact that McCall (1996) used more covariates than those that came with the data set.

In this section we have upgraded the *continuous time mixture model* into a missing failure causes model so as to be able to include the subjects with missing failure causes in the analysis.

Table 3.3: Maximum likelihood estimates for the Complete Case Analysis Mixture Model and the Missing Failure Causes Mixture Model (with standard errors) (* denotes $P<$ 0.05).

	Complete Case Analysis Mixture Model (Latency Model)		Missing Failure Causes Mixture Model (Latency Model)	
Coefficient	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
T1	-2.583(0.074)*	-2.277(0.124)*	-2.519(0.064)*	-2.170(0.109)*
T2	-2.796(0.085)*	-2.677(0.150)*	-2.583(0.0708)*	-2.376(0.123)*
T3	-2.936(0.098)*	-2.925(0.176)*	-2.593(0.077)*	-2.599(0.144)*
T4	-3.468(0.138)*	-3.314(0.227)*	-3.283(0.116)*	-3.117(0.197)*
T5	-2.655(0.103)*	-2.655(0.180)*	-2.443(0.086)*	-2.403(0.153)*
T6	-3.608(0.179)*	-3.714(0.322)*	-3.404(0.151)*	-3.461(0.274)*
T7	-2.436(0.112)*	-2.699(0.212)*	-2.241(0.095)*	-2.482(0.185)*
T8	-3.926(0.260)*	-3.484(0.338)*	-3.643(0.211)*	-3.234(0.290)*
T9	-3.025(0.176)*	-3.643(0.383)*	-2.773(0.146)*	-3.365(0.324)*
T10	-5.267(0.578)*	-4.353(0.580)*	-4.382(0.349)*	-3.757(0.420)*
T11	-2.989(0.198)*	-3.760(0.451)*	-2.875(0.175)*	-3.661(0.418)*
T12	-4.134(0.379)*	-4.130(0.580)*	-3.805(0.300)*	-3.710(0.458)*
T13	-2.740(0.202)*	-3.062(0.359)*	-2.544(0.171)*	-2.874(0.318)*
T14	-2.318(0.184)*	-3.174(0.413)*	-2.223(0.164)*	-3.059(0.379)*
T15	-2.461(0.231)*	-3.676(0.581)*	-2.539(0.220)*	-3.688(0.567)*
T16	-2.871(0.317)*	-3.496(0.581)*	-2.404(0.232)*	-3.226(0.492)*
T17	-2.835(0.354)*	-4.405(1.002)*	-2.603(0.292)*	-4.278(0.908)*
T18	-2.695(0.379)*	-3.588(0.710)*	-2.663(0.338)*	-3.566(0.675*
T19	-1.211(0.214)*	-1.920(0.340)*	-0.910(0.173)*	-1.870(0.318)*
ui	1.505(0.065)*	0.556(0.116)*	1.341(0.055)*	0.540(0.100)*
dr	-0.447(0.504)	-1.729(0.835)*	-0.215(0.413)*	-0.590(0.689)
rr	0.480(0.456)	-0.025(0.784)	0.834(0.384)*	-0.392(0.629)
age	-0.011(0.003)*	-0.008(0.006)	-0.009(0.003)*	-0.010(0.005)
wage	0.227(0.096)*	0.039(0.150)	0.332(0.080)*	-0.118(0.124)
tenure	-0.014(0.006)*	0.044(0.011)*	-0.019(0.005)*	0.033(0.010)

Table 3.4: Maximum likelihood estimates for the Complete Case Analysis Mixture Model and the Missing Failure Causes Mixture Model (with standard errors) (* denotes $P <$ 0.05).

	Complete Case Analysis Mixture Model (Incidence Model)	Missing Failure Causes Mixture Model (Incidence Model)
Coefficient	$\hat{\gamma}_1$	$\hat{\gamma}_2$
Constant	1.178(0.062)*	1.195(0.058)*
`ui`	-0.488(0.092)*	-0.267(0.085)*
`dr`	-2.381(0.674)*	-1.344(0.608)*
`rr`	0.822(0.668)*	0.504(0.579)*
`age`	-0.009(0.005)*	-0.014(0.004)*
`wage`	0.897(0.131)*	0.797(0.117)*
`tenure`	0.038(0.009)*	0.024(0.009)*

To illustrate the applicaton of the proposed model, we continue to test the same proposal by McCall (1996) to see if the inclusion of the subjects that have missing failure causes would lead us to agree with McCall (1996) or not. We continue to include all the covariates that came with the data set as explanatory variables where the unemployment benefit recipients are regarded as a reference category and all continuous covariates are centered at their respective averages. We also evaluate the impact of unemployment insurance on employment. The results of fitting the proposed missing failure causes are listed in Table 3.3 and Table 3.4. We have also fitted the proposed model excluding the subjects with missing failure causes which reduces to the ordinary *continuous time mixture model* that was advanced in the previous section. The model for failure type probabilities for the ordinary mixture model from the previous section and the proposed missing failure causes mixture model is given by

$$\pi_1(\boldsymbol{x};\boldsymbol{\gamma}_1) = \frac{\gamma_{01} + \exp(\boldsymbol{x}^T\boldsymbol{\gamma}_1)}{1 + \gamma_{01} + \exp(\boldsymbol{x}^T\boldsymbol{\gamma}_1)}$$

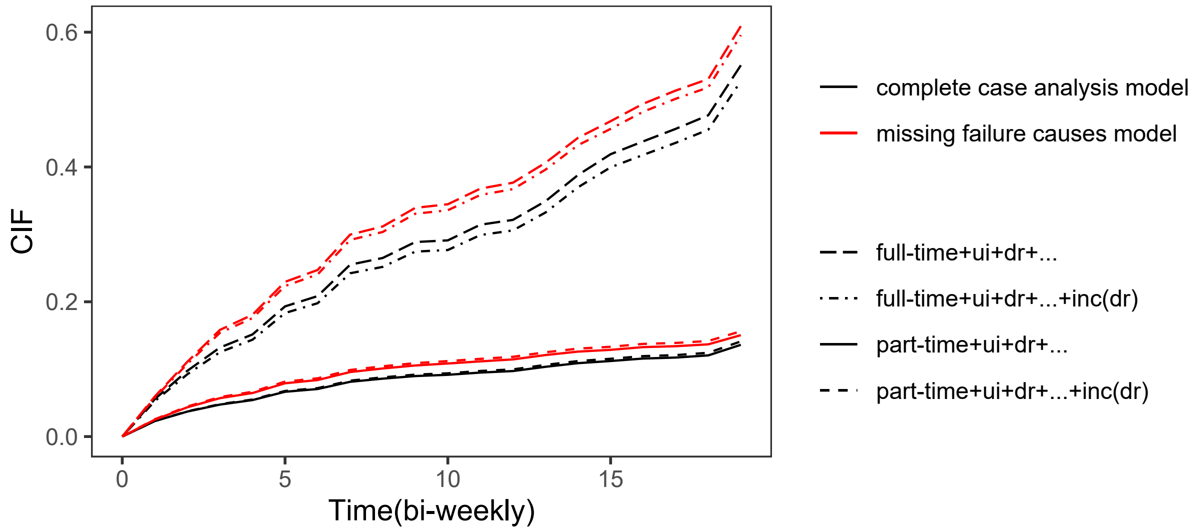

Figure 3.3: The CIF estimates of exit to full-time and part-time employment with the effect of increasing *dr* via the Complete Case Analysis Model and the Missing Failure Causes Model.

where we regard full-time employment and part-time employment as cause 1 and cause 2, respectively. The model for component hazards is given by

$$\lambda_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j) = \exp(\beta_{0jt}) \exp(\boldsymbol{x}^T \boldsymbol{\beta}_{1j}) \quad a_{t\text{-}1} \leq t < a_t$$

for both the ordinary mixture model and the missing failure causes mixture model where we have applied a Poisson regression model to estimate $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \ldots, \boldsymbol{\beta}_J^T)^T$, in the presence as well as in the absence of subjects that have missing failure causes.

Recall that according to the *continuous time mixture model* from the previous section we found that the probability of finding employment was about (full-time:47.7%+ part-time:12.0%) 59.7% for an unemployed individual who is in receipt of unemployment benefits with average values for continuous covariates, but if we raised the disregard rate for this individual by 50%, but held the other covariates at base values this probability of re-employment drops to (full-time:45.5%+ part-time:12.4%) 57.9% for this individual. These results are at odds to the claim by McCall (1996) that increasing the disregard rate for unemployment benefit recipients improves their prospects of re-employment. We have plotted the CIF estimates with the effect of raising the disregard rate by 50% including and excluding the subjects that come with missing failure causes in Figure 3.3. Clearly from Figure 3.3, we still reach the same conclusions in the presence of missing failure causes that we arrived at in their absence, that is, increasing the disregard rate induces a reduction in full-time re-employment and no effect on part-time re-employment. When we include the subjects that have missing failure causes the probability that an unemployed individual from the reference group is re-employed increases from 59.7% to 66.7%, and raising the disregard rate and holding everything else constant at base values the probability of re-employment increased from 57.9% in the absence of missing failure causes to 65.9% in the presence of missing failure causes. Thus, when the subjects that have missing failure causes are included in the analysis we have a probability of 66.7% that an unemployed individual from the reference group will be re-employment and this drops to 65.9% when we raise the disregard rate by 50% holding other covariates at their base values.

Clearly, these findings are at variance with what McCall (1996) found. The inclusion of subjects with missing failure causes does not narrow the gap between our findings and what is claimed by McCall (1996) regarding the impact of increasing the diregard rate via the missing failure causes mixture model that we have advanced in this chapter. Of course, one of the reasons could be the 102 individuals that we did not include in the anlysis or the fact that McCall (1996) used more covariates or that the proposed missing failure causes mixture model is defective.

There is nothing that can be done about the 102 subjects or the covariates that were considered by McCall (1996), but we can fit a multinomial model within the MI framework and compare this model to the proposed missing failure causes mixture model in terms of whether we would arrive at the same conclusions via the missing failure causes multinomial model as when we applied the proposed missing failure causes mixture model.

The MI method is an established procedure for handling missing data. In the present settings of missing failure causes for some subjects, this procedure entails;

1) Recreating the missing failure causes for the affected subjects following some model to create M complete samples where M is some pre-specified number

2) Applying some standard analysis method to M complete samples to produce M estimates of some parameter(s) of interest. In the present settings we apply a multinomial model to produce CSH parameter estimates; $\boldsymbol{\alpha}_1^*, \boldsymbol{\alpha}_2^*, \ldots, \boldsymbol{\alpha}_M^*$, from M complete samples.

3) Pooling these estimates to determine the final estimate;

$$\bar{\boldsymbol{\alpha}} = \frac{\sum_{m=1}^{M} \boldsymbol{\alpha}_m^*}{M}.$$

The variance of $\bar{\boldsymbol{\alpha}}$, as proposed by Little and Rubin (1987), comes from the within imputation variance $\boldsymbol{W}\text{var}(\bar{\boldsymbol{\alpha}})$ and the between imputation variance $\boldsymbol{B}\text{var}(\bar{\boldsymbol{\alpha}})$, that is;

$$\boldsymbol{W}\text{var}(\bar{\boldsymbol{\alpha}}) = \frac{1}{M} \sum_{m=1}^{M} \text{Var}(\boldsymbol{\alpha}_m^*)$$

and,

$$\boldsymbol{B}\text{var}(\bar{\boldsymbol{\alpha}}) = \sum_{m=1}^{M} \frac{(\boldsymbol{\alpha}_m^* - \bar{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_m^* - \bar{\boldsymbol{\alpha}})^T}{M-1}$$

such that,

$$\text{Var}(\bar{\boldsymbol{\alpha}}) = \boldsymbol{W}\text{var}(\bar{\boldsymbol{\alpha}}) + (1 + M^{-1})\boldsymbol{B}\text{var}(\bar{\boldsymbol{\alpha}}).$$

The asumption regarding the missingness mechanism or the cause of missingness is the first step in the exercise of recreating missing failure causes for afflicted subjects. The MI method, relies on the assumption that missingness depends on observed data only and not on the missing data. This type of missingness is referred to as missing at random (MAR) (Little and Rubin, 1987). In more formal terms, it is assumed that the distribution of missing failure causes does not depend on failure causes, but on observed data, that is;

$$P(R_i = 0 | \Delta_i, \Delta_i > 0, \boldsymbol{w}_i) = P(R_i = 0 | \Delta_i > 0, \boldsymbol{w}_i) \tag{3.2.1}$$

where $\boldsymbol{w}_i = (t_i, \boldsymbol{x}_i, \boldsymbol{z}_i)$ is a vector of observed data for subject i with $\boldsymbol{z}$ as the covariate vector that may explain missingness. Central to the recreation of complete data samples is a model $\Phi(\boldsymbol{w})$ for imputing missing failure causes. Here we assume that there are 2 causes of failure, that is, cause 1 and cause 2 where our cause of interest is cause 1 ($\Delta = 1$). The definition of $\Phi(\boldsymbol{w})$ is then given by

$$\Phi(\boldsymbol{w}_i) = P(\Delta_i = 1 | R_i = 0, \Delta_i > 0, \boldsymbol{w}_i). \tag{3.2.2}$$

The assumption given in (3.2.1) implies independence of Δ and R conditional on observed data $(I(\Delta > 0), \boldsymbol{w})$, and can be equivalently written as;

$$\Phi(\boldsymbol{w}_i) = P(\Delta_i = 1 | R_i = 0, \Delta_i > 0, \boldsymbol{w}_i) = P(\Delta_i = 1 | R_i = 1, \Delta_i > 0, \boldsymbol{w}_i).$$

This means that the failure type distribution for subjects with missing failure causes is the same as the failure type distribution for subjects with observed failure type conditional on $\boldsymbol{w}$.

This result underpins the process of imputing missing failure causes becauses it implies that we can specify a model for failure type probabilities $\Phi(\boldsymbol{w}, \boldsymbol{\tau})$, and assuming that this model is correctly specified, we can then proceed to estimate $\boldsymbol{\tau}$, from the subjects with known failure causes, whereafter, use $\hat{\Phi}(\boldsymbol{w}, \hat{\boldsymbol{\tau}})$, to re-create the missing failure causes for affected subjects.

For the purposes of our data set, the covariates that explain missingness were not collected, thus $\boldsymbol{w} = (t, \boldsymbol{x})$. With two failure causes under consideration, the natural model for failure type probabilities is a binomial model;

$$\Phi(\boldsymbol{w}, \boldsymbol{\tau}) = \frac{\exp\left(g(\boldsymbol{w})^T \boldsymbol{\tau}\right)}{1 + \exp\left(g(\boldsymbol{w})^T \boldsymbol{\tau}\right)}. \tag{3.2.3}$$

The components of $g(\boldsymbol{w})$, may include higher terms and/or interactions to improve the fit of the model as suggested by Lu and Tsiatis (2001). After the estimation of $\boldsymbol{\tau}$, the imputation procedure can be summarized as follows:

1) To generate the m^{th} complete data set, draw $\boldsymbol{\tau}_m^*$, from $\mathcal{N}(\hat{\boldsymbol{\tau}}, \hat{\mathrm{V}}(\hat{\boldsymbol{\tau}}))$.

Table 3.5: Maximum likelihood estimates for the Missing Failure Causes Multinomial Model (with standard errors) (* denotes $P < 0.05$).

Missing Failure Causes Multinomial Model		
Coefficient	$\hat{\alpha}_1$	$\hat{\alpha}_2$
`T1`	-2.622(0.068)*	-3.754(0.115)*
`T2`	-2.754(0.075)*	-3.779(0.124)*
`T3`	-2.825(0.083)*	-3.903(0.140)*
`T4`	-3.582(0.123)*	-4.472(0.190)*
`T5`	-2.677(0.093)*	-3.776(0.153)*
`T6`	-3.710(0.157)*	-3.866(0.267)*
`T7`	-2.494(0.102)*	-3.866(0.187)*
`T8`	-4.053(0.225)*	-4.757(0.282)*
`T9`	-3.117(0.154)*	-4.717(0.317)*
`T10`	-5.103(0.428)*	-5.129(0.364)*
`T11`	-3.220(0.181)*	-4.996(0.430)*
`T12`	-4.283(0.324)*	-5.061(0.439)*
`T13`	-2.930(0.184)*	-4.112(0.305)*
`T14`	-2.583(0.177)*	-4.202(0.376)*
`T15`	-2.971(0.233)*	-4.680(0.557)*
`T16`	-2.914(0.254)*	-4.055(0.465)*
`T17`	-3.246(0.355)*	-4.397(0.671)*
`T18`	-2.143(0.361)*	-3.665(0.643)*
`T19`	-1.438(0.210)*	-2.615(0.318)*
`ui`	1.136(0.059)*	-1.136(0.101)*
`dr`	-1.076(0.443)*	-0.122(0.668)
`rr`	0.840(0.395)	-0.324(0.594)
`age`	-0.015(0.003)*	-0.001(0.005)
`wage`	0.522(0.084)*	-0.495(0.122)*
`tenure`	-0.005(0.006)	-0.003(0.010)

2) For each subject i that failed at time t_i with a missing failure cause, compute;

$$\hat{\Phi}_{mi}(\boldsymbol{x}_i, \boldsymbol{\tau}^*_m)) = \frac{\exp\left(g(\boldsymbol{x}_i)^T \boldsymbol{\tau}^*\right)}{1 + \exp\left(g(\boldsymbol{x}_i)^T \boldsymbol{\tau}^*\right)}.$$

Randomly select 1 or 0 from $\mathcal{B}(1, \hat{\Phi}_{mi}(\boldsymbol{x}_i, \boldsymbol{\tau}^*_m)$, and replace the missing failure cause by cause 1 or cause 2 acording to whether 1 or 0 was drawn.

These steps are repeated until M complete data sets are obtained. In Table 3.5 we have listed the results of fitting a multinomial model within the MI framework for $M = 8$. According to the missing failure causes multinomial model, the probability of re-employment for an unemployed individual from the reference group is (full-time:49.8%+ part-time:14.9%) 64.7%. This probability drops marginally to (full-time:47.9%+ part-time:15.1%) 63.00% on account of increasing the disregard rate, but holding everything else at base level. Thus, whether we apply the proposed missing failure causes mixture

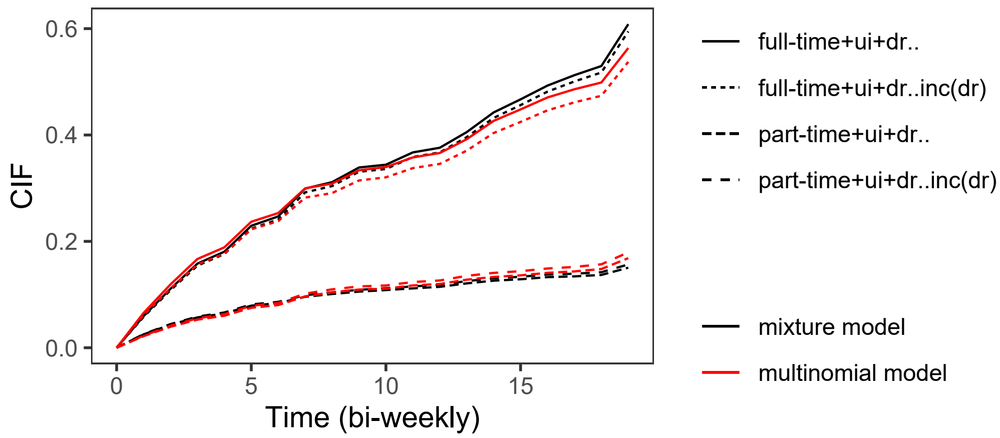

Figure 3.4: The CIF estimates of exit to full-time and part-time employment with the effect of increasing *dr* via the Missing Failure Causes Multinomial is Model and the Missing Failure Causes Mixture Model.

model or the missing failure causes multinomial model we come to the same conclusion that raising the disregard rate does not engender an increase in re-employment prospects. This holds true whether we include the subjects with missing failure causes or not. The multinomial model is an established discrete time competing risks model and so is the MI method as a model for handling missing data. Again, putting aside the fact that our findings via the proposed missing failure causes mixture model are at variance with McCall (1996) regarding the effect of increasing the disregard rate, we have managed to advance a missing failure causes model that compares favourably with an established method for handling missing failure causes. In Figure 3.4 we have plotted the CIF estimates with the effect of of increasing the disregard rate by 50% via the proposed model and the multinomial model within the MI framework.

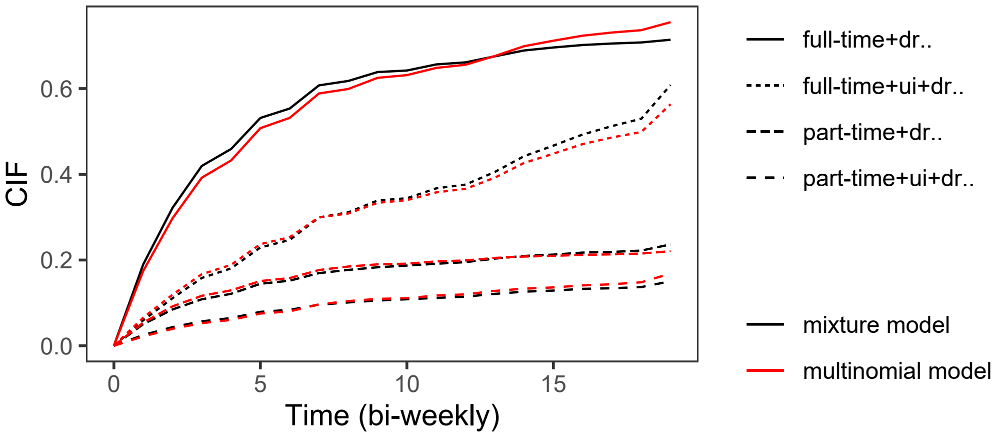

Figure 3.5: The CIF estimates of exit to full-time and part-time employment with the effect of `ui` via the Missing Failure Causes Multinomial Model and the Missing Failure Causes Mixture Model.

We have also applied both the proposed model and the missing failure causes multinomial model to assess the effect of unemployment insurance on re-employment prospects. Recall that the probalility of re-employment for the reference group (unemployed individuals who receive unemployment benefits with average values for continuous covariates) is 66.7% and 64.9% according to the proposed model and the missing failure causes multinomial model, respectively. When we consider an individual with average values for continuous covariates, but who does not receive unemployment benefits, the re-employment probability for this individual is about 93.00% according to the proposed model while the missing failure causes multinomial model suggests that this probability is 95.1%. This is further evidence that the proposed model compares favourably with the missing failure causes multinomial model, both models lead to the same conclusion that provision of unemployment benefits reduces re-employment prospects. We have plotted the CIF estimates with the effect of unemployment insurence from both models in Figure 3.5. Clearly, Figure 3.5 confirms that applying both the proposed missing failure causes mixture model and the missing failure causes multinomial model leads to the same conclusion that receipt of employment benefits reduces the prospects of re-employment. Obviously, these findings are in keeping with the theory on the effect of unemployment benefits on employment.

3.2.3 Discussion

In this chapter the main objective was to demonstrate how the mixture model as presented by its authors with associated failure time distributional assumption can be applied in discrete time for analysis of ordinary competing risks data and data that comes with missing failure causes. The secondary objective was to demonstrate that $\boldsymbol{\beta}$, the parameters of the model that are associated with the latency, can be estimated via a certain Poisson regression model. This was suggested by one of the referees to the same article by the authors of the model. This regression model relies on "exposure" of each subject or the time spent alive in a given interval for estimation of $\boldsymbol{\beta}$. The exact failure/censoring in a given time interval is known for each subject when time to failure evolves in continuous units. This is not possible in discrete time, all that is known is that a failure/censoring time occured in a given interval, but its exact value is not known. In order to apply the mixture model in discrete time we had to assume that failure times occured half way through the interval while censoring was assumed to occur at the end of the interval where it was observed to occur. Essentially, to re-purpose the continuous time mixture model for application in discrete time we left the model intact, but adjusted the structure of the data. This is a very minor adjustment. One of the advantages of the mixture model is a simpler regression model for the CIF in comparison to the cause-specific-hazard denominated regression model for the CIF that arises under the multinomial model or the binomial model. The latter model is well known for complicating the evaluation of covariate effects on the CIF. We also discussed the circumstances under which it is possible to predict the covariate effects on the CIF from regression coefficient estimates.

Essentially, in the first section of this chapter we have demonstrated how a patently continuous time mixture model can also serve as an alternate discrete time model. We have also validated the proposed model against the multinomial model and found that the application of either model led to the same conclusions when we evaluated a certain proposition by one author McCall (1996) regarding the effect of raising the disregard rate for unemployment benefit recipients on their re-employment prospects and a well documented fact in economics regarding the impact of unemployment benefits on re-employment prospects. Even though we arrived at conclusions which were contrary to what McCall (1996) proposed regarding the effect of increasing the disregard rate, but what is important for the purposes of this thesis is that we have managed to advance a discrete time model that compares favourably with an established discrete time model by the way of the multinomial model. The exclusion of the subjects with missing failure causes could be the reason for the variance in the findings or the fact that McCall (1996) used more covariates than the covariate set that came with data. In the second section of this chapter we upscaled the propose ordinary mixture into a missig failure causes model which, in essence, reduces to the estimation of the parameters of the mixture, i.e, the failure type probabilities and component hazards, but now in the presence of missing failure causes. In the application of the proposed mixture model which accounts for missing failure causes we still came to the conclusion that raising the disregard rate does induce an improvement in employment. We then fitted a multinomial model within the MI framework for comparison purposes and it was found that the proposed model compared favourably with missing failure causes multinomial model. In fact the proposed model and the missing failure causes multinomial model lead to the same conclusions when they were applied to test McCall (1996)'s proposal and the effect of unemployment benefits on re-employment prospects. The MI model is an established method for dealing with missing failure causes in competing risks literature. Clearly, if the proposed model compares favourably with an established model for dealing with missing failure causes it raises confidence in the proposed model as a reliable method for handling missing failure causes in the discrete time realm. In the next chapter, Chapter 4, we will advance a truly discrete time version of the mixture model. We will also upscale the proposed model into a missing failure causes model in the same chapter.

CHAPTER 4

The Discrete Time Mixture Model

4.1 The Discrete Time Regression Model

In the previous chapter, Chapter 3, we took a continuous time mixture competing risks model and demonstrated how the model, with a slight adjustment of failure/censoring times, can be applied to discrete time data. The best course of action though is to develop a truly discrete time mixture regression model from the outset. This section is dedicated towards that objective, that is, to reformulate the continuous time mixture model as a discrete time regression model. As in the previous chapter, we will later upgrade this model into missing failure causes regression model. Recall that observed data can be represented by $\boldsymbol{y} = ((t_1, \Delta_1, \boldsymbol{x}_1), (t_2, \Delta_2, \boldsymbol{x}_2) \ldots, (t_n, \Delta_n, \boldsymbol{x}_n))^T$, where $T_i = \min\{\tilde{T}_i, C_i\}$, $\Delta_i = I(\tilde{T} \wedge C_i)D_i$, and $\boldsymbol{x}_i$ is a p-dimensional vector of covariates. We continue to connect the failure type probabilities to data with the multinomial model;

$$\pi_j(\boldsymbol{x}, \boldsymbol{\gamma}) = \exp(\gamma_{0j} + \boldsymbol{x}^T \boldsymbol{\gamma}_{1j}) \Big/ (1 + \sum_{l=1}^{J-1} \exp(\gamma_{0l} + \boldsymbol{x}^T \boldsymbol{\gamma}_{1l})$$

for $j = 1, \ldots, J-1$, where $\pi_J(\boldsymbol{x}, \boldsymbol{\gamma}) = 1 - \sum_{j=1}^{J-1} \pi_j(\boldsymbol{x}, \boldsymbol{\gamma})$. Let $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^T, \ldots, \boldsymbol{\gamma}_{J\text{-}1}^T)^T$, and $\boldsymbol{\gamma}_j = (\gamma_{0j}, \boldsymbol{\gamma}_{1j}^T)^T$, where γ_{0j} is a scalar and $\boldsymbol{\gamma}_{1j}$ is a vector of regression coefficients.

The mixture model has been one of the most well studied models in the competing risks literature. While the multinomial distribution is the most natural model for the failure type distribution, the model allows for flexibility in the choice of a model for component hazards. In fact, the different presentations of the model differ in terms of various models that have been suggested for component hazards. When Larson and Dinse (1985) introduced the model they assumed proportional hazards for component hazards with piece-wise constant baseline component hazards. This is the *continuous time mixture model* that we worked with in Chapter 3. Ng and McLachan (2003); Escarala and Bowater (2008) have specified a semi-parametric model for component hazards. Haller (2014) have also assumed proportional hazards for component hazards and modeled the baseline component hazards with splines. These are some examples of the models for component hazards that have been suggested in the literature. To place the model in the discrete time realm we propose the following definition for component hazards;

$$\lambda_j(t) = P(T = t | T \geq t, D = j).$$

With covariates, thas is, $\lambda_j(t) = P(T = t|T \geq t, D = j; \boldsymbol{x})$, we also take advantage of this flexibility in terms modelling the component hazards as allowed by the mixture model and propose the following model for component hazards in discrete time;

$$g(\lambda_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j)) = \beta_{0jt} + \boldsymbol{x}^T \boldsymbol{\beta}_{1j}$$

for $t = 1, \ldots, q$, and $j = 1, \ldots, J$, where $g(.)$ is a link function within the GLM framework. This is in keeping with models for normal discrete time hazard functions, see for example (Allison, 1982; Singer and Willet, 2003; Tutz and Schmid, 2016). Let $\boldsymbol{\beta}_j = (\boldsymbol{\beta}_{0j}^T, \boldsymbol{\beta}_{1j}^T)^T$, where $\boldsymbol{\beta}_{0j} = (\beta_{0j1} \ldots, \beta_{0jq})^T$, and $\boldsymbol{\beta}_{1j} = (\beta_{1j1}, \ldots, \beta_{1jp})^T$, are vectors of the baseline component hazard coefficents and regression coefficients, respectively. In practice, the common link functions are the logit and the complementary log-log link functions. Some authors prefer to assume a logit link function when failure times are inherently discrete and a complementary log-log for grouped failure times. Yu et al. (2011) have advanced a model that is very similar to the proposed model, where a complementary log-log link function was assumed, however, the proposed model is more flexible as it allows for various specifications for the component hazards.

In discrete time the definition of the component density function is now given as;

$$f_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j) = \frac{\lambda_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j)}{1 - \lambda_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j)} S_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j) \tag{4.1.1}$$

where $S_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j)$ is the survival function for subjects that eventually failed due to cause j at time t conditional on $\boldsymbol{x}$ or the component survival function. The definition for the component survival function $S_j(t)$ is given by

$$S_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j) = \prod_{s=1}^{t} (1 - \lambda_j(s|\boldsymbol{x}, \boldsymbol{\beta}_j)). \tag{4.1.2}$$

Naturally, the population survival function continues to be expressed as a mixture of component survival functions with failure type probabilities as weights;

$$S(t|\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{j=1}^{J} \pi_j(\boldsymbol{x}, \boldsymbol{\gamma}) S_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j). \tag{4.1.3}$$

where, $\boldsymbol{\theta} = (\boldsymbol{\gamma}^T, \boldsymbol{\beta}^T)^T$, and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_J^T)$. These arguments locate the mixture model in the discrete time realm via the adjustment of the distributional assumptions that underly the latency component of the mixture model. We now attend to the estimation of $\boldsymbol{\theta}$, the parameter vector that describes the failure type probabilities and component hazards of the model. With the aid of the EM Algorithm, the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are also estimated seperately. We demonstrate this procedure in the next subsection, Subsection 4.1.1. In Subsection 4.1.2 we demonstrate the application of the proposed model and compare it to both the multinomial model and the *continuous time mixture model* from Chapter 3. In Section 4.2 we upgrade the proposed model into a missing failure causes model. The standard errors for the CIF are presented in Appendix B.

4.1.1 The EM Algorithm

Assuming that observed data is modeled with the proposed mixture model, the full log-likelihood function can be written as;

$$\mathcal{L}_0(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{j=1}^{J} d_{ij} \log[\pi_j(\boldsymbol{x}_i, \boldsymbol{\gamma})) f_j(t_i|\boldsymbol{x_i}, \boldsymbol{\beta}_j)] + (1 - d_i) \log \left[\sum_{j=1}^{J} \pi_j(\boldsymbol{x}_i, \boldsymbol{\gamma}) S_j(t_i|\boldsymbol{x}_i, \boldsymbol{\beta}_j) \right]$$

where $d_{ij} = I(D_i = j)$ and $d_i = 1 - \sum_{j=1}^{J} d_{ij}$ as before. We also justify the application of the EM algorithm by regarding the unknown failure causes for censored subjects as missing information. As in Chapter 3, we introduce a pseudo vector $\boldsymbol{v}_i = (v_{i1}, v_{i2} \ldots, v_{iJ})^T$, where v_{ij} is 1 or 0 according to whether subject i eventually fails due to cause j or not.

If $\boldsymbol{v}_i$ was actually observed, then, what is referred to as a complete data log-likelihood function $\mathcal{L}_c(\boldsymbol{\theta})$ can be expressed as a sum of;

$$\mathcal{L}_c(\boldsymbol{\gamma}) = \sum_{i=1}^{n}\sum_{j=1}^{J} g_{ij}\log \pi_j(\boldsymbol{x}_i, \boldsymbol{\gamma})$$

and,

$$\mathcal{L}_c(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}\log\frac{\lambda_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)}{1-\lambda_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)} + \sum_{s=1}^{t_i} g_{ij}\log(1-\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j))$$

where $g_{ij} = d_{ij} + (1-d_i)v_{ij}$. We then introduce the same time dependent indicator variable d_{ijs} such that $d_{ijs} = 0$ for $s = 1, 2\ldots, t_i$-1, and $d_{ijt_i} = d_{ij}$. Suppose that $\boldsymbol{\theta}^{(r)}$ is the MLE of $\boldsymbol{\theta}$ in the M-Step at the $(r)^{\text{th}}$ iteration of the algorithm. To complete the E-Step at the $(r{+}1)^{\text{th}}$ iteration, we replace $g_{ij} = d_{ij}+(1-d_i)v_{ij}$ with $g_{ij}^{(r)} = \mathrm{E}[g_{ij}|\boldsymbol{\theta}^{(r)}, \boldsymbol{y}] = d_{ij} + (1-d_i)v_{ij}^{(r)}$, where;

$$v_{ij}^{(r)} = \mathrm{E}[v_{ij}|\boldsymbol{\theta}^{(r)}, \boldsymbol{y}] = \frac{\pi_j(\boldsymbol{x}_i, \boldsymbol{\gamma}^{(r)})S_j(t_i|\boldsymbol{x}_i, \boldsymbol{\beta}_j^{(r)})}{\sum_{j=1}^{J}\pi_j(\boldsymbol{x}_i, \boldsymbol{\gamma}^{(r)})S_j(t_i|\boldsymbol{x}_i, \boldsymbol{\beta}_j^{(r)})}.$$

The E-Step can be written as a sum of;

$$Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}) = \sum_{i=1}^{n}\sum_{j=1}^{J} g_{ij}^{(r)}\log \pi_j(\boldsymbol{x}_i, \boldsymbol{\gamma}_j)$$

and,

$$\begin{aligned} Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(r)}) &= \sum_{j=1}^{J}\left\{\sum_{i=1}^{n}\sum_{s=1}^{t_i} d_{ijs}\log\frac{\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j)}{1-\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j)} + g_{ijs}^{(r)}\log(1-\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j))\right\} \\ &= \sum_{j=1}^{J} Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^{(r)}) \end{aligned}$$

where $g_{ijs} = g_{ij}$ for $s = 1, \ldots, t_i$. From Chapter 3, $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)})$ is the now familiar kernel of a multinomial log-likelihood function, that is, $(g_{i1}^{(r)}, \ldots, g_{iJ}^{(r)}) \sim \mathcal{M}(1, \pi_1(\boldsymbol{x}_i, \boldsymbol{\gamma}), \ldots, \pi_J(\boldsymbol{x}_i, \boldsymbol{\gamma}))$, while, $Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^{(r)})$ is recognizable as a kernel of a binomial log-likelihood function, where, $d_{ijs} \sim \mathcal{B}(g_{ijs}^{(r)}, \lambda_j(s|\boldsymbol{x}_i, \boldsymbol{\beta}_j))$. As in Chapter 3, to minimize excessive programming in the M-Step, $\boldsymbol{\gamma}_j$ $(j = 1, \ldots, J)$, can be estimated individually by fitting J-1 binomial distributions, with $g_{ij}^{(r)} \sim \mathcal{B}(1, \pi_j(\boldsymbol{x}_i, \boldsymbol{\gamma}_j))$, and the standard errors can be minimized by regarding the most prevalent failure cause as a base category (Begg and Gray, 1984). To determine the

MLE of $\boldsymbol{\beta}_j$ $(j = 1, \ldots, J)$, in the M-Step, the failure time data must be re-arranged into long format. Thus, the MLE of $\boldsymbol{\theta}$ in each iteration of the M-Step can be determined by fitting (2J-1) binomial distributions within the GLM framework.

.

4.1.2 Application

We have applied the proposed model to the same unemployment data *UnempDur* which comes with `Ecdat` Croissant and Graves (2020) **R** package. Recall that McCall (1996) argued that if the provision of unemployment benefits is accompanied by an upward adjustment of the disregard rate, unemployed individuals switch away from full-time employment towards part-time employment with a net effect of an overall improvement in the re-employment rate. As already highlighted in Chapter 3, we also test this proposition to illustrate the application of the proposed model. We also demonstrate the application of the proposed model by assessing the effect of providing financial support to unemployed individuals on re-employment prospects. Recall that the theory on unemployment benefits suggests that provision of these benefits does not improve re-employment prospects.

Again, all the covariates that came with the data are considered as explanatory variables where continuous covariates are centered at their respective averages, furthermore, the recipients of unemployment benefits are regarded as a reference category. The model for component hazards is given by

$$\lambda_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j) = 1 - \exp(-\exp(\beta_{0jt} + \boldsymbol{x}^T\boldsymbol{\beta}_{1j}))$$

for $t = 1, \ldots, 19; j = 1, 2.$

Table 4.1: Maximum likelihood estimates for component hazards from the Discrete Time Mixture Model (with standard errors) (* denotes $P< 0.05$).

Latency Model

	$\hat{\boldsymbol{\beta}}_1$	$\hat{\boldsymbol{\beta}}_2$
T1	-2.590(0.075)*	-2.253(0.125)*
T2	-2.805(0.085)*	-2.656(0.150)*
T3	-2.946(0.099)*	-2.905(0.177)*
T4	-3.479(0.138)*	-3.294(0.227)*
T5	-2.662(0.104)*	-2.634(0.180)*
T6	-3.618(0.180)*	-3.694(0.322)*
T7	-2.441(0.113)*	-2.677(0.213)*
T8	-3.936(0.260)*	-3.462(0.338)*
T9	-3.034(0.177)*	-3.621(0.383)*
T10	-5.277(0.578)*	-4.329(0.580)*
T11	-2.999(0.199)*	-3.737(0.452)*
T12	-4.145(0.380)*	-4.106(0.581)*
T13	-2.750(0.204)*	-3.037(0.360)*
T14	-2.325(0.186)*	-3.148(0.413)*
T15	-2.472(0.234)*	-3.650(0.581)*
T16	-2.883(0.318)*	-3.470(0.583)*
T17	-2.847(0.351)*	-4.377(1.004)*
T18	-2.706(0.375)*	-3.559(0.709)*
T19	-1.192(0.216)*	-1.893(0.342)*
ui	1.530(0.066)*	0.525(0.116)*
dr	-0.488(0.505)	-1.688(0.836)*
rr	0.473(0.458)	0.005(0.780)
age	-0.012(0.403)	-0.007(0.006)
wage	0.228(0.097)	0.037(0.149)
tenure	-0.014(0.006)	0.043(0.011)

Table 4.2: Maximum likelihood estimates for failure type probabilities from the Discrete Time Mixture Model (with standard errors) (* denotes $P < 0.05$).

Incidence Model	
	$\hat{\boldsymbol{\gamma}}_1$
γ_0	1.202(0.063)*
ui	-0.523(0.092)*
dr	-2.317(0.675)*
rr	0.847(0.669)*
age	-0.009(0.005)*
wage	0.898(0.131)*
tenure	0.037(0.009)*

The failure type probabilities follow a binomial model;

$$\pi_j(t|\boldsymbol{x}, \boldsymbol{\gamma}_j) = \frac{\exp\left(\alpha_{0jt} + \boldsymbol{x}^T \boldsymbol{\alpha}_{1j}\right)}{1 + \exp\left(\alpha_{0jt} + \boldsymbol{x}^T \boldsymbol{\alpha}_{1j}\right)}.$$

Full-time and part-time employment have been set as failure cause 1 and failure cause 2, respectively. The results of the analysis are displayed in Table 4.1 and Table 4.2. Naturally, the covariate effects are assessed via their effects on the CIF. The estimates for the CIF are computed from;

$$\hat{F}_j(t|\hat{\boldsymbol{\theta}}) = \hat{\pi}_j(\boldsymbol{x}, \hat{\boldsymbol{\gamma}}_j)(1 - \hat{S}_j(t|\boldsymbol{x}, \hat{\boldsymbol{\beta}}_j)).$$

There are only two failure causes for this data set as in Chapter 3. It is, therefore, possible to predict the effect of a covariate on the CIF from the regression coefficient estimates provided the covariate induces the marginal failure type probability and the component hazards to move in the same direction.

Since, for failure cause 1(full-time employment) $\beta_{1\texttt{dr}1} < 0$, and $\gamma_{1\texttt{dr}1} < 0$, increasing the disregard rate (dr) for an unemployed individual who is in reciept of unemployment benefits will diminish the likelihood that the individual will exit the state of unemployment to full-time employment. Again, increasing the disregard rate (`dr`) has the effect of reducing the marginal probability of failure due to failure type 1 and thereby increasing the marginal probability of failure due to failure type 2 because $\pi_2 = 1 - \pi_1$.

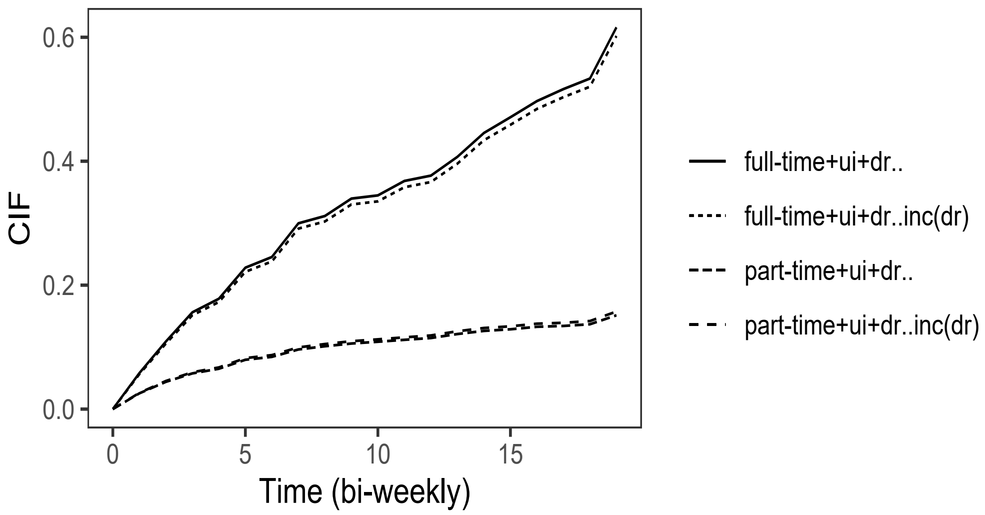

Figure 4.1: The CIF of exit to full time and part time employment for the `ui` recipients with the effect of increasing `dr` via the Discrete Time Mixture Model.

Because $\beta_{1\texttt{dr}2} < 0$, the effect of increasing disregard rate (`dr`) on part-time employment cannot be predicted as it induces the marginal failure type probability and the component hazards to move in opposite directions. Our findings are, therefore, in agreement with McCall (1996) in so far as full-time employment is concerned. To assess the effect of increasing (`dr`) on part-time re-employment prospects, we examine the CIF, and this is plotted in Figure 4.1. Again here, for illustrative purposes, we have assessed the effect of increasing the disregard rate by 50% for an average unemployed individual who receives unemployment benefits. The plot confirms our prediction regarding the effect of increasing the disregard rate on full-time re-employment. The plot also suggests that increasing the disregard rate has the effect of improving part-time re-employment prospects, albeit, marginally.

Table 4.3: Maximum likelihood estimates for the Continuous Time Mixture Model & The Multinomial Model (with standard errors) (* denotes $P < 0.05$).

	Mixture Model (Latency Model)		Multinomial Model	
Coefficient	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
T1	-2.583(0.074)*	-2.277(0.124)*	-2.694(0.077)*	-3.823(0.130)*
T2	-2.796(0.085)*	-2.677(0.150)*	-2.989(0.088)*	-4.173(0.153)*
T3	-2.936(0.098)*	-2.925(0.176)*	-3.181(0.102)*	-4.381(0.179)*
T4	-3.468(0.138)*	-3.314(0.227)*	-3.762(0.141)*	-4.766(0.229)*
T5	-2.655(0.103)*	-2.655(0.180)*	-2.926(0.108)*	-4.045(0.182)*
T6	-3.608(0.179)*	-3.714(0.322)*	-3.949(0.183)*	-5.134(0.324)*
T7	-2.436(0.112)*	-2.699(0.212)*	-2.747(0.119)*	-4.029(0.215)*
T8	-3.926(0.260)*	-3.484(0.338)*	-4.328(0.263)*	-4.857(0.341)*
T9	-3.025(0.176)*	-3.643(0.383)*	-3.411(0.182)*	-4.985(0.385)*
T10	-5.267(0.578)*	-4.353(0.580)*	-5.704(0.580)*	-5.731(0.582)*
T11	-2.989(0.198)*	-3.760(0.451)*	-3.394(0.205)*	-5.082(0.454)*
T12	-4.134(0.379)*	-4.130(0.580)*	-4.580(0.383)*	-5.473(0.583)*
T13	-2.740(0.202)*	-3.062(0.359)*	-3.150(0.210)*	-4.340(0.363)*
T14	-2.318(0.184)*	-3.174(0.413)*	-2.733(0.196)*	-4.403(0.418)*
T15	-2.461(0.231)*	-3.676(0.581)*	-2.956(0.243)*	-4.850(0.585)*
T16	-2.871(0.317)*	-3.496(0.581)*	-3.421(0.329)*	-4.660(0.587)*
T17	-2.835(0.354)*	-4.405(1.002)*	-3.417(0.367)*	-5.563(1.006)*
T18	-2.695(0.379)*	-3.588(0.710)*	-3.366(0.392)*	-4.697(0.717)*
T19	-1.211(0.214)*	-1.920(0.340)*	-1.823(0.242)*	-2.804(0.357)*
ui	1.505(0.065)*	0.556(0.116)*	1.184(0.068)*	1.199(0.119)*
dr	-0.447(0.504)	-1.729(0.835)*	-1.731(0.529)*	-0.591(0.821)
rr	0.480(0.456)	-0.025(0.784)	0.900(0.464)*	-0.334(0.742)
age	-0.011(0.003)*	-0.008(0.006)	-0.015(0.004)*	-0.003(0.006)
wage	0.227(0.096)*	0.039(0.150)	0.536(0.099)*	-0.390(0.149)
tenure	-0.014(0.006)*	0.044(0.011)*	0.002(0.006)*	0.003(0.011)

Table 4.4: Maximum likelihood estimates for the Continuous Time Mixture Model (with standard errors) (* denotes $P < 0.05$).

Mixture Model	
(Incidence Model)	
Coefficient	$\hat{\gamma}$
Constant	1.178(0.062)*
`ui`	-0.488(0.092)*
`dr`	-2.381(0.674)*
`rr`	0.822(0.668)*
`age`	-0.009(0.005)*
`wage`	0.897(0.131)*
`tenure`	0.038(0.009)*

This does not negate McCall (1996), but it is doubful if the increase in part-time re-employment is large enough to offset the reduction in full-time re-employment to the extent that overall re-employment improves according to McCall (1996).

Thus, we have come to the same conclusion that we reached from the application of the *continuous time mixture model* especially in relation to the effect of increasing (`dr`) on part-time re-employment. The results of fitting the *continuous time mixture model* and the multinomial model are reproduced in Table 4.3 and Table 4.4. Indeed, we have used about 80% of the data that was considered by McCall (1996) because the other portion of the data came with missing failure causes as before. Furthermore, we have used fewer covariates than McCall (1996). At this point we proceed to validate the proposed model by comparing it to the multinomial model as in Chapter 3 where we validated the *continuous time mixture model* by comparing it to the multinomial model. Better still, we throw the *continuous time mixture model* in the mix to see how the two mixture models compare with each other.

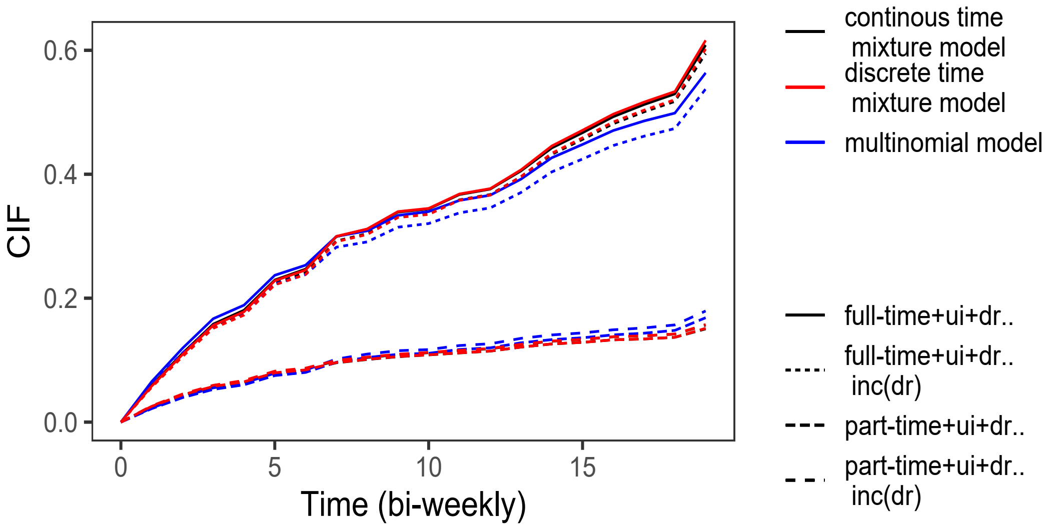

Figure 4.2: The CIF of exit to full time and part time employment for the `ui` recipients with the effect of increasing `dr` via the Discrete Time Mixture Model, the Continuous Time Mixture Model and the Multinomial Model.

Recall from Chapter 3 that according to the *continuous time mixture model*, an unemployed individual who is in receipt of unemployment benefits with average values for continuous covariates, has probability of (full-time:55.1%+ part-time:13.6%) 68.7% to be re-employed, but if the disregard rate for the same individual is raised by 50% and hold everything else constant this probaility drops to (full-time:52.7%+ part-time:14.1%) 66.8% . According to the multinomial model, an individual from the reference group has (full-time:49.8%+ part-time:14.6%) 64.4% chance of re-employment, but this probability drops to (full-time:46.7%+ part-time:14.5%) 61.2% if we raise the disregard rate by 50%, but holding everything else constant. The proposed discrete time mixture model also suggests that the probability of re-employment for an unemployed individual from the reference group is (full-time:55.3%+ part-time:13.6%) 68.9% and that this probability drops to (full-time:52.9%+ part-time:14.1%) 67.0% for the same individual if we raise the disregard rate by 50%. Evidently, the proposed *discrete time mixture model* agrees with both the multinomial model and the *continuous time mixture model* regarding the effect of raising the disregard rate on re-employment prospects.

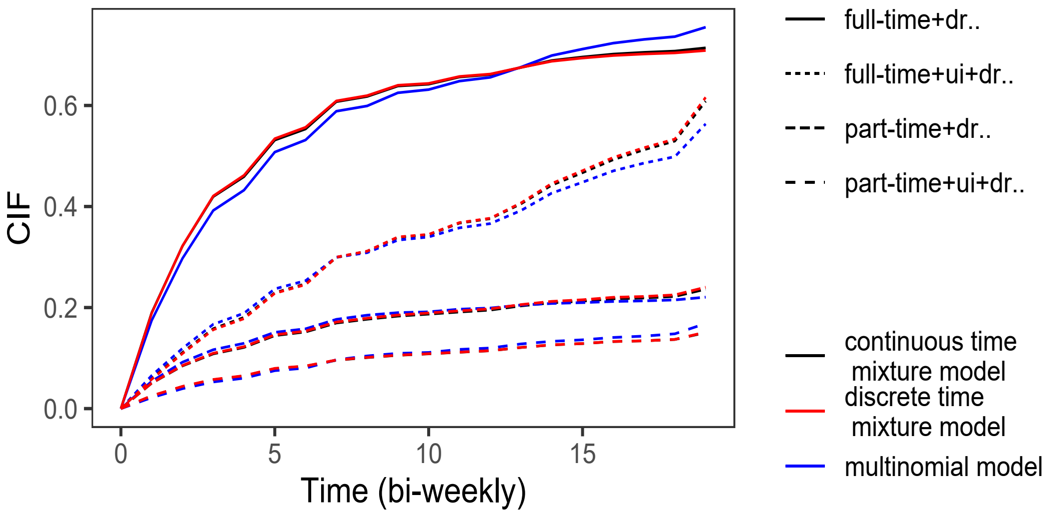

Figure 4.3: The CIF of exit to full time and part time employment with the effect `ui` via the Discrete Time Mixture Model, the Continuous Mixture Model and the Multinomial Model.

The proposed model together with the *continuous time mixture model* and the multinomial suggest that increasing the disregard rate does not induce an increase in re-employment. While there is a reduction in full-time employment there is almost no movement in part-time employment to offset the reduction in full-time employment such that there is overall improvement in employment as claimed by McCall (1996). Off course, the fact that only 80% of the total subjects were considerd in the analysis or the fact that McCall (1996) used more covariates than those that came with the data set could be some of the reasons that there are differences between ourselves (as informed by the results of fitting the two mixture models and the multinomial model) and McCall (1996) regarding the effect of raising the disregard rate on overall employment. It might be worth mentioning that the mixture models tend to produce similar CIF estimates.

We now proceed to apply all three models to test the effect of unemployment benefits on employment. Accordingly, the three CIF estimates from the three models are plotted in Figure 4.3. Evidently, Figure 4.3 suggests that the three models are all in agreement that provision of unemployment benefits does not improve the re-employment prospects, be it part-time or full-time re-employment as espoused by the theory on this subject. Note that the CIF estimates via the mixture models are almost indistinguishable. This provides further evidence that the proposed discrete time mixture model tends to compare more favorably to the *continuous time mixture model* than to the multinomial model in terms of CIF estimates. In the next section we extend the proposed discrete time mixture model into a model that can deal with missing failure causes.

4.2 The Missing Failure Causes Regression Model

Here we advance a truly discrete time regression model for analysis of data that comes with missing failure causes from the outset. We are extending the ordinary *discrete time mixture model* that we advanced in the previous section into a model that can deal with subjects that have missing failure causes. Naturally, we have to specify regression models for component hazards and failure type probabilities to connect these quantities to data. We continue to model the failure type distribution with a multinomial distribution;

$$\pi_j(\boldsymbol{x}, \boldsymbol{\gamma}) = \exp(\gamma_{0j} + \boldsymbol{x}^T \boldsymbol{\gamma}_{1j}) \Big/ (1 + \sum_{l=1}^{J-1} \exp(\gamma_{0l} + \boldsymbol{x}^T \boldsymbol{\gamma}_{1l})$$

for $j = 1, \ldots, J-1$, where $\pi_J(\boldsymbol{x}, \boldsymbol{\gamma}) = 1 - \sum_{j=1}^{J-1} \pi_j(\boldsymbol{x}, \boldsymbol{\gamma})$. Let $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^T, \ldots, \boldsymbol{\gamma}_{J\text{-}1}^T)^T$, and $\boldsymbol{\gamma}_j = (\gamma_{0j}, \boldsymbol{\gamma}_{1j}^T)^T$, where γ_{0j} is a scalar and $\boldsymbol{\gamma}_{1j}$ is a vector of regression coefficients. We also continue to model the component hazards within the GLM framework;

$$g(\lambda_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j)) = \beta_{0jt} + \boldsymbol{x}^T \boldsymbol{\beta}_{1j}$$

for $t = 1, \ldots, q$, and $j = 1, \ldots, J$. Let $\boldsymbol{\beta}_j = (\boldsymbol{\beta}_{0j}^T, \boldsymbol{\beta}_{1j}^T)^T$ where $\boldsymbol{\beta}_{0j} = (\beta_{0j1} \ldots, \beta_{0jq})^T$, and $\boldsymbol{\beta}_{1j} = (\beta_{1j1}, \ldots, \beta_{1jp})^T$, are vectors of the baseline component hazard coefficents and regression coefficients, respectively. Let $\boldsymbol{\theta} = (\boldsymbol{\gamma}^T, \boldsymbol{\beta}^T)^T$ where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \ldots, \boldsymbol{\beta}_J^T)^T$, represent the parameters that describe the component hazards and failure type probabilities.

To obtain the estimates for component hazards and failure type probabilities we have to determine $\hat{\boldsymbol{\theta}}$. Indeed, this is achieved by differentiating the observed data likelihood function w.r.t. $\boldsymbol{\theta}$. We also implement an EM Algorithm to determine $\hat{\boldsymbol{\theta}}$. This exercise is carried out in the next subsection, Subsection 4.2.1. This is followed by the application of the proposed model together with missing failure causes multinomial model to discrete time data that has missing failure causes in Subsection 4.2.2. We conclude this chapter with a discussion in Subsection 4.3. The CIF standard errors are presented in Appendix B.

4.2.1 The EM Algorithm

Following from Chapter 3, when data comes with missing failure causes, the full log-likelihood function can be written as;

$$\mathcal{L}_0(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}\log[\pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma}))f_j(t_i|\boldsymbol{x_i},\boldsymbol{\beta}_j)] + d_{i*}\log\left[\sum_{j=1}^{J}\pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma}))f_j(t_i|\boldsymbol{x_i},\boldsymbol{\beta}_j)\right](1-d_i)$$
$$+\log\left[\sum_{j=1}^{J}\pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma})S_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)\right]$$

where $d_{ij} = I(D_i = j)$ and $d_i = 1 - \sum_{j=1}^{J} d_{ij}$ as before. We also justify the application of the EM algorithm by regarding the unknown failure causes for censored subjects and the missing failure causes as missing information. As in Chapter 3, data is augmented with vector $\boldsymbol{v}_i = (v_{i1},\ldots,v_{ij})^T$, for a censored subject i where v_{ij} assumes values 1 or 0 when the subject eventually fails by cause j or not, and vector $\boldsymbol{u}_i = (u_{i1},\ldots,u_{ij})^T$, for a subject i that has a missing failure cause where u_{ij} assumes values 1 or 0 if the subject has actually failed by cause j or not.f the pseudo variables were actually observed, then the complete data log-likelihood function can be written as;

$$\mathcal{L}_c(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{2} d_{ij}\log(\pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma})f_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)) + d_{i*}u_{ij}\log(\pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma})f_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j))$$
$$+(1-d_i)v_{ij}\log(\pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma})S_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)).$$

Substituting (4.1.1) to (4.1.3), $\mathcal{L}_c(\boldsymbol{\theta})$ can be written as;

$$\mathcal{L}_c(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{2} d_{ij}\log(\pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma})\frac{\lambda_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)}{1-\lambda_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)}\prod_{s=1}^{t_i}(1-\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j)))$$
$$+ d_{i*}u_{ij}\log(\pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma})\frac{\lambda_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)}{1-\lambda_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)}\prod_{s=1}^{t_i}(1-\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j)))$$
$$+ (1-d_i)v_{ij}\log(\pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma})\prod_{s=1}^{t_i}(1-\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j)))$$

After re-arranging few terms, the complete data log-likelihood function can be expressed as a sum of;

$$\mathcal{L}_c(\boldsymbol{\pi}) = \sum_{i=1}^{n}\sum_{j=1}^{J} g_{ij}\log\pi_j(\boldsymbol{x}_i(\boldsymbol{x}_i,\boldsymbol{\gamma}),\boldsymbol{\gamma})$$

and,

$$\mathcal{L}_c(\boldsymbol{\lambda}) = \sum_{i=1}^{n}\sum_{j=1}^{2} m_{ij}\log\left\{\frac{\lambda_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)}{1-\lambda_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j)}\right\} + \sum_{s=1}^{t_i} g_{ij}\log\left(1-\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j)\right)$$

where $g_{ij} = d_{ij} + d_{i*}u_{ij} + (1-d_i)v_{ij}$ and $m_{ij} = d_{ij} + d_{i*}u_{ij}$. Already, it can be seen that $\mathcal{L}_c(\boldsymbol{\pi})$ resembles the log-likelihood of a binomial distribution. Let $m_{ijs} = d_{ijs} + d_{i*s}u_{ij}$. Assume that $d_{ijs} = 0$ and $d_{i*s} = 0$ for $s < t_i$, but $d_{ijt_i} = d_{ij}$ and $d_{i*t_i} = d_{i*}$, so that $m_{ijs} = 0$ for $s < t_i$ and $m_{ijt_i} = m_{ij}$. Taking these definitions into account, $\mathcal{L}_c(\boldsymbol{\lambda})$ can be re-written as

$$\mathcal{L}_c(\boldsymbol{\lambda}) = \sum_{i=1}^{n}\sum_{j=1}^{2}\sum_{s=1}^{t_i} m_{ijs}\log\left\{\frac{\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j)}{1-\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j)}\right\} + g_{ijs}\log\left(1-\lambda_j(s|\boldsymbol{x}_i,\boldsymbol{\beta}_j)\right)$$

Again, since $\mathcal{L}_c(\boldsymbol{\pi})$ and $\mathcal{L}_c(\boldsymbol{\lambda})$, are both linear in pseudo variables, at the $(r+1)^{\text{th}}$ iteration of the algorithm, the E-Step entails replacing g_{ij} and m_{ijs} with their conditional expectations; $g_{ij}^{(r)} = d_{ij} + d_{i*}u_{ij}^{(r)} + (1-d_i)v_{ij}^{(r)}$, and $m_{ijs}^{(r)} = d_{ijs} + d_{i*s}u_{ij}^{(r)}$, respectively. The expectation of the pseudo variables are computed conditional on $\boldsymbol{\theta}^{(r)}$ and $\boldsymbol{y}$, where $\boldsymbol{\theta}^{(r)}$ is the MLE of $\boldsymbol{\theta}$ at the M-Step of the $(r)^{\text{th}}$ iteration, and they are given by

$$v_{ij}^{(r)} = \frac{\pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma}^{(r)})S_j(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_j^{(r)})}{\sum_{l=1}^{J}\pi_l(\boldsymbol{x}_i.\boldsymbol{\gamma}^{(r)})S_l(t_i|\boldsymbol{x}_i,\boldsymbol{\beta}_l^{(r)})}$$

and,

$$u_{ij}^{(r)} = \frac{\pi_j(\boldsymbol{x}_i, \boldsymbol{\gamma}^{(r)})\lambda_j(t_i|\boldsymbol{x}_i, \boldsymbol{\beta}_j^{(r)})}{\sum_{l=1}^{J} \pi_l(\boldsymbol{x}_i, \boldsymbol{\gamma}^{(r)})\lambda_l(t_i|\boldsymbol{x}_i, \boldsymbol{\beta}_l^{(r)})}.$$

The E-Step in the $(r+1)^{\text{th}}$ iteration of the algorithm which is now expressed as a sum of;

$$Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}) = \sum_{i=1}^{n}\sum_{j=1}^{J} g_{ij}^{(r)} \log \pi_j(\boldsymbol{x}, \boldsymbol{\gamma})$$

and,

$$Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(r)}) = \sum_{j=1}^{J}\sum_{i=1}^{n}\sum_{s=1}^{t_i} m_{ijs}^{(r)} \log \lambda_j(s|\boldsymbol{x}, \boldsymbol{\beta}_j) + (g_{ijs}^{(r)} - m_{ijs}^{(r)}) \log(1 - \lambda_j(s|\boldsymbol{x}, \boldsymbol{\beta}_j))$$

with $m_{ijs}^{(r)} = d_{ijs} + u_{ij}^{(r)} d_{*js}$, $g_{ij}^{(r)} = d_{ij} + u_{ij}^{(r)} d_{*j} + (1-d_i)v_{ij}^{(r)}$ and $g_{ijs}^{(r)} = g_{ij}^{(r)}$ for $s = 1, 2 \ldots, t_i$. Clearly, failure type probabilities continue to be estimated via the application of a multinomial distribution because $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(r)})$, resembles a multinomial log-likelihood function. Of course, we can collapse the non-target failure causes into one failure cause when $J > 2$ if there only one failure cause of interest or we can fit J-1 binomial distributions to estimate $\boldsymbol{\gamma}_j$ ($j = 1, 2, \ldots, J$-1), individually where the most frequently occuring failure cause is treated as a reference category to minimize failure type probability standard errors (Begg and Gray, 1984). To determine $\boldsymbol{\beta}_j$ ($j = 1, 2 \ldots, J$), J binomial distributions are fitted to data in a long format.

4.2.2 Application

Again, we use the unemployment data to illustrate the application of the proposed model where we have 1073 full-time re-employment, 339 part-time re-employment, 1255 censored subjects and 574 subjects that were re-employed, but the type of re-employment was not specified. Time to failure is adjusted to $T \in \{1, 2, \ldots, 19\}$, as before. We also test the claim by McCall (1996) that increasing the disregard rate actually improves overall re-employment figures. We also evaluate the effect of introducing unemployment benefits on the re-employment rate as well.

Table 4.5: Maximum likelihood estimates for the Discrete Time Missing Failure Causes Mixture Model and the Continuous Time Missing Failure Causes Mixture Model (with standard errors) (* denotes $P<0.05$).

	Discrete Time Mixture Model (Latency Model)		Continuous Time Mixture Model (Latency Model)	
Coefficient	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
T1	-2.524(0.065)*	-2.163(0.109)*	-2.519(0.064)*	-2.170(0.109)*
T2	-2.589(0.071)*	-2.367(0.123)*	-2.583(0.0708)*	-2.376(0.123)*
T3	-2.599(0.078)*	-2.586(0.144)*	-2.593(0.077)*	-2.599(0.144)*
T4	-3.287(0.117)*	-3.121(0.197)*	-3.283(0.116)*	-3.117(0.197)*
T5	-2.444(0.087)*	-2.404(0.153)*	-2.443(0.086)*	-2.403(0.153)*
T6	-3.407(0.151)*	-3.464(0.274)*	-3.404(0.151)*	-3.461(0.274)*
T7	-2.239(0.096)*	-2.479(0.184)*	-2.241(0.095)*	-2.482(0.185)*
T8	-3.641(0.211)*	-3.243(0.291)*	-3.643(0.211)*	-3.234(0.290)*
T9	-2.772(0.147)*	-3.368(0.324)*	-2.773(0.146)*	-3.365(0.324)*
T10	-4.378(0.348)*	-3.769(0.422)*	-4.382(0.349)*	-3.757(0.420)*
T11	-2.874(0.176)*	-3.664(0.418)*	-2.875(0.175)*	-3.661(0.418)*
T12	-3.802(0.301)*	-3.723(0.460)*	-3.805(0.300)*	-3.710(0.458)*
T13	-2.542(0.173)*	-2.877(0.318)*	-2.544(0.171)*	-2.874(0.318)*
T14	-2.218(0.166)*	-3.060(0.379)*	-2.223(0.164)*	-3.059(0.379)*
T15	-2.535(0.223)*	-3.696(0.567)*	-2.539(0.220)*	-3.688(0.567)*
T16	-2.400(0.233)*	-3.220(0.490)*	-2.404(0.232)*	-3.226(0.492)*
T17	-2.596(0.292)*	-4.280(0.906)*	-2.603(0.292)*	-4.278(0.908)*
T18	-2.654(0.338)*	-3.573(0.675)*	-2.663(0.338)*	-3.566(0.675*
T19	-0.841(0.177)*	-1.845(0.314)*	-0.910(0.173)*	-1.870(0.318)*
ui	1.505(0.065)*	0.556(0.116)*	1.341(0.055)*	0.540(0.100)*
dr	-0.447(0.504)	-1.729(0.835)*	-0.215(0.413)*	-0.590(0.689)
rr	0.480(0.456)	-0.025(0.784)	0.834(0.384)*	-0.392(0.629)
age	-0.011(0.003)*	-0.008(0.006)	-0.009(0.003)*	-0.010(0.005)
wage	0.227(0.096)*	0.039(0.150)	0.332(0.080)*	-0.118(0.124)
tenure	-0.014(0.006)*	0.044(0.011)*	-0.019(0.005)*	0.033(0.010)

Table 4.6: Maximum likelihood estimates for the Discrete Time Missing Failure Causes Mixture Model and the Continuous Time Missing Failure Causes Mixture Model (with standard errors) (* denotes $P< 0.05$).

	Discrete Time Mixture Model (Incidence Model)	Continuous Time Causes Mixture Model (Incidence Model)
Coefficient	$\hat{\gamma}_1$	$\hat{\gamma}_1$
Constant	1.178(0.062)*	1.195(0.058)*
`ui`	-0.488(0.092)*	-0.267(0.085)*
`dr`	-2.381(0.674)*	-1.344(0.608)*
`rr`	0.822(0.668)*	0.504(0.579)*
`age`	-0.009(0.005)*	-0.014(0.004)*
`wage`	0.897(0.131)*	0.797(0.117)*
`tenure`	0.038(0.009)*	0.024(0.009)*

We also consider all the covariates that came with the data set where employment benefit recipients are set as a reference and the continuous covariates are centered at their respective averages. The model for component hazards is given by

$$g(\lambda_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j)) = 1 - \exp(-\exp(\beta_{0jt} + \boldsymbol{x}^T\boldsymbol{\beta}_j))$$

with $j = 1, 2$, and the model for failure type probabilities is given by

$$\text{Logit}(\pi_1(\boldsymbol{x}, \boldsymbol{\gamma}_1)) = \gamma_{01} + \boldsymbol{x}^T\boldsymbol{\gamma}_{11} \tag{4.2.1}$$

where we regard full-time and part-time re-employment as cause 1 and cause 2, respectively. When we advanced the *continuous time mixture model* for application to discrete time data that comes with missing failure causes we did indicate that we would wish to compare that model to this discrete time mixture model for handling missing failure causes to see if the two models will continue to compare favourably in the presence of missing as we found them to do so in the absence of these subjects. .

Table 4.7: Maximum likelihood estimates for the Missing Failure Causes Multinomial Model (with standard errors) (* denotes $P < 0.05$).

Missing Failure Causes

Multinomial Model

Coefficient	$\hat{\alpha}_1$	$\hat{\alpha}_2$
T1	-2.622(0.068)*	-3.754(0.115)*
T2	-2.754(0.075)*	-3.779(0.124)*
T3	-2.825(0.083)*	-3.903(0.140)*
T4	-3.582(0.123)*	-4.472(0.190)*
T5	-2.677(0.093)*	-3.776(0.153)*
T6	-3.710(0.157)*	-3.866(0.267)*
T7	-2.494(0.102)*	-3.866(0.187)*
T8	-4.053(0.225)*	-4.757(0.282)*
T9	-3.117(0.154)*	-4.717(0.317)*
T10	-5.103(0.428)*	-5.129(0.364)*
T11	-3.220(0.181)*	-4.996(0.430)*
T12	-4.283(0.324)*	-5.061(0.439)*
T13	-2.930(0.184)*	-4.112(0.305)*
T14	-2.583(0.177)*	-4.202(0.376)*
T15	-2.971(0.233)*	-4.680(0.557)*
T16	-2.914(0.254)*	-4.055(0.465)*
T17	-3.246(0.355)*	-4.397(0.671)*
T18	-2.143(0.361)*	-3.665(0.643)*
T19	-1.438(0.210)*	-2.615(0.318)*
ui	1.136(0.059)*	-1.136(0.101)*
dr	-1.076(0.443)*	-0.122(0.668)
rr	0.840(0.395)	-0.324(0.594)
age	-0.015(0.003)*	-0.001(0.005)
wage	0.522(0.084)*	-0.495(0.122)*
tenure	-0.005(0.006)	-0.003(0.010)

We have listed the results of fitting the two mixture models for handling missing failure causes, that is, the *continuous time mixture model* from Chapter 6 as well as the *discrete time mixture model* that we have just proposed in this chapter and the results of this exercise are displayed in Table 4.5 and Table 4.6.

We have also included the results of fitting a multinomial model for handling missing failure causes in Table 4.7. Recall from Chapter 3 that excluding missing failure causes we found that 68.7% of the individuals from the reference group found employment while raising the disregard rate for these individuals reduces the re-employment rate to 66.8% from the ordinary *continuous time mixture model* while the *discrete time mixure model* from Chapter 4 suggested that 68.9% from the reference group are re-employed and raising the disregard rate for the same group reduces this probability to 67%. When we include the subjects with missing failure causes, the *continuous time mixture model* for missing failure causes from Chapter 6 suggested that 75.9% of the individuals from the reference group were re-employed and raising the disregard rate for these individuals by 50% has no effect because 75.2% of these individuals are re-employed. The *discrete time mixture model* for handling missing failure causes that we have proposed in this chapter suggests 76.5% from the reference group were re-employed and again no real effect from increasing the disregard rate because 75.7% are re-employed as a result.

Thus, both models suggest that the effect of raising the disregard rate is to reduce the propects of re-employment, albeit, marginally if subjects with missing failure causes are excluded, but when these subjects are included then raising the disregard rate seems to have no effect. If we consider the multinomial model, this model suggests a drop in re-employment from 64.4% to 61.6% due to the increase in the disregard rate if subjects with missing failure causes are excluded, but when these subjects are included the model suggests that 73.5% are re-employed from the reference group and this drops to 72.1 % from increasing the disregard rate by 50%. Clearly, the mixture models suggest that there is marginal reduction in the prospects of re-employment if the disregard rate is increased when the subjects with missing failure causes are excluded, but if these subjects

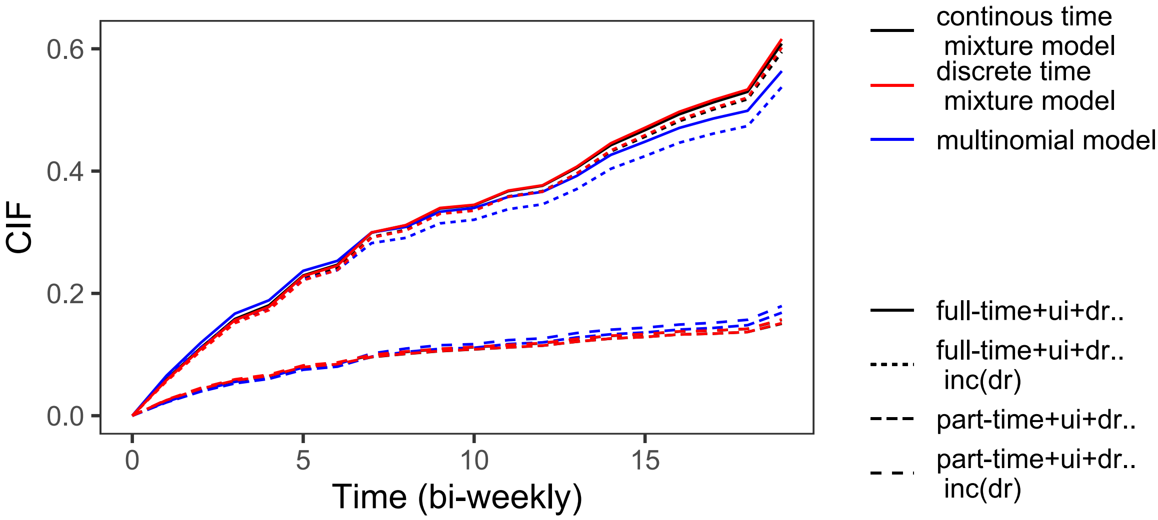

Figure 4.4: The CIF of exit to full-time and part-time employment with the effect of `dr` via the Discrete Time Mixture Model, Continuous Time Mixture Model and the Multinomial Model.

are included these models suggest that raising the disregard rate has no impact on re-employment prospects. The multinomial model, agrees with the mixture models, the model suggests a marginal reduction in the prospects of re-employment when we exclude missing failure causes and almost no effect when we include these subjects. In Figure 4.4 we have plotted the CIF estimates with effect of increasing the disregard rate by 50% from the three models. The plot confirms that the mixture models are more in agreement with each other than with with the multinomial model, but all the models suggest that raising the disregard rate for benefit recipients does not improve their prospects of re-employments. Thus, the inclusion of the subjects with missing failure causes somewhat narrows the gap between ourselves and McCall (1996) regarding the effect of increasing the disregard rate on re-employment prospects for benefit recipients. The fact that we have used fewer covariates than McCall (1996) may be the reason that there is still a gap between ourselves and McCall (1996), but despite these differences the proposed mixture models compare favourably with MI method, an established model for handling missing data (missing failure causes).

Regarding the effect of unemployment benefits on re-employment, the multinomial model suggests that an unemployed individual from the reference group, that is, an average individual who receives benefits, has a 73.5% chance of being re-employed and the same average indidual who does not recieve benefits has a 97.5% chance of being re-employed.

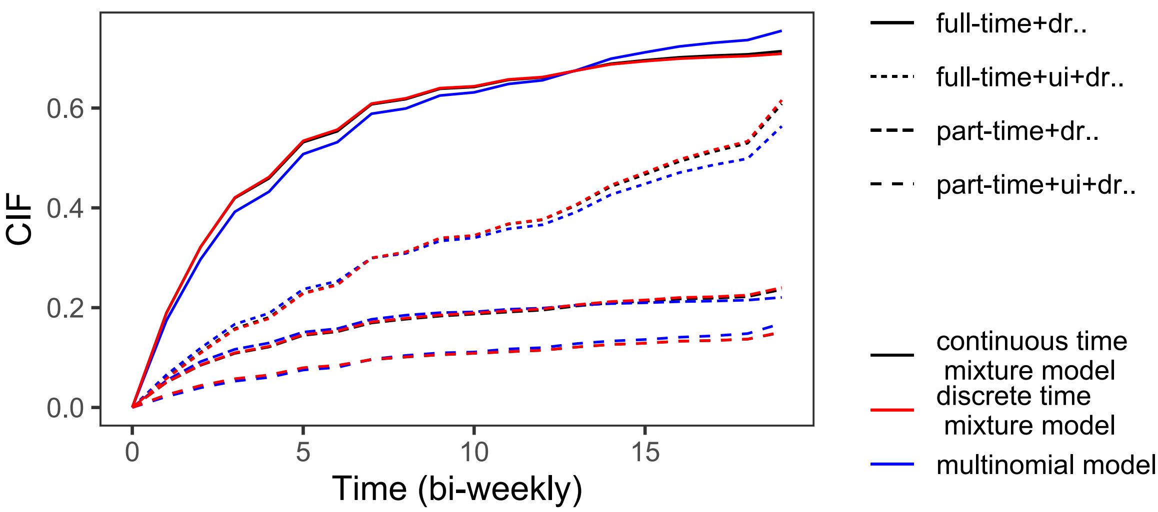

Figure 4.5: The CIF of exit to full-time and part-time employment with the effect of `ui` via the Discrete Time Mixture Model, Continuous Time Mixture Model and the Multinomial Mode.

The *discrete time mixture model* suggests 76.5% as a probability of re-employment for a benefit recepient and 94.9% for a non-recpient while the *continous time mixture model* suggests 75.9% and 95% for a recipient and a non-recipient, respectively. Again, the mixture models lead to the same conclusion as the MI model regarding the effect of unemployment benefits on re-employment prospects, that is, provision of unemployment benefits reduces employment prospects. In Figure 4.5 we have plotted the CIF estimates to evaluate the effect of unemployment benefits on re-employment prospects from the three models. Clearly, this plot provides further evidence that over and above the fact that all three models lead to the same conclusions, but the mixture models tend to agree with each other more than each model agrees with the multinomial model in terms of CIF estimates.

4.2.3 Discussion

In Chapter 3 it was demostrated how the *continuous time mixture model* can be applied to discrete time data by adjusting the failure times and censoring times. In this chapter we set out to advance a truly discrete time mixture model. This was achieved by modelling the component hazards within the GLM framework and assumed a multinomial model for failure type probabilities. As in Chapter 3, we demonstrated the application of the proposed regression model by testing McCall (1996)'s proposal regarding the effect of increasing the disregard rate for unemployed individuals who are in receipt of unemployment benefits together with the *continuous time mixture model* from Chapter 3 and the multinomial model. We found that all three models led to the conclusion that raising the disregard rate has the effect of reducing full-time re-employment and almost no effect on part-time re-employment with a net effect of a reduction in total re-employment. These findings were different from McCall (1996), but despite these differences, of paramount importance in so far as this thesis is concerned is the fact that we have advanced yet another discrete time regression model that we have managed to validate against the multinomial model. We also tested the effect of unemployment benefits on re-employment prospects via the three models and found that all three models led to the same conclusion that unemployment benefits act as a disincentive because recepients of these benefits tend not to intensify job search which ultimately results in lower re-employment rates. We also found that the two mixture models seemed to compare more favourably with one another than with the multinomial if one compared the CIF estimates that were produced by the three models. The mixture models tend to produce almost indentical CIF estimates and by implication, it means that the mixture models will also produce similar estimates for CSHs compared to the estimates for these quantities that will be derived from the multinomial model. Recall that when data is modeled with a mixture model (discrete time or continuous time) the causes-specific-hazard estimates are now derived from CIF estimates;

$$\hat{h}_j(t|\boldsymbol{x}, \hat{\boldsymbol{\theta}}) = \frac{\hat{F}_j(t|\boldsymbol{x}, \hat{\boldsymbol{\theta}}) - \hat{F}_j(t\text{-}1|\boldsymbol{x}, \hat{\boldsymbol{\theta}})}{1 - \sum_{j=1}^{J} \hat{F}_j(t\text{-}1|\boldsymbol{x}, \hat{\boldsymbol{\theta}})}.$$

We have also advanced a regression model for dealing with missing failure causes. The proposed regression model is essentially an extension of the *discrete time mixture model* into a missing failure causes regression model. We then applied the proposed model together with the multinomial model within the MI framework for comparison purposes to real discrete time competing risks data that has missing failure causes to test the same proposition by McCall (1996) regarding the effect of raising the disregard rate for employment benefit recipients and also to test the effect of unemployment benefits on re-employment prospects. We found that the two models led to the same conclusion about the effect of raising the disregard rate for benefit recipients which was, however, in contradiction to what McCall (1996) found. In the application of the two models to test the effect of unemployment benefits on re-employment prospects, we found that the two models led to the conclusion that unemployment benefits reduced the prospects of re-employment which is in line with the theory on this subject. Even though in the application of the proposed model to test the effect of increasing the disregard rate we arrived at different conclusions to McCall (1996), but for the purposes of this theis we have managed to advance a mixture model for dealing with missing failure causes which compares favourably with the MI method, an established method for handling missing data. We also compared the proposed model to the *continuous time mixture model* for handling missing failure causes from Chapter 3 and found that all three models led to the same conclusion regarding the effect of raising the disregard rate and the effect of unemployment benefits on employment prospects. We also found that the mixture models for handling missing failure causes compare more favourably with one another than each with the missing failure causes multinomial model in terms of CIF estimates. In the next chapter, Chapter 5, we focus on the vertical model (Nicolaie et al., 2010, 2015), the final model that we intend to re-frame as a discrete time model.

CHAPTER 5

The Discrete Time Vertical Model

5.1 The Ordinary Discrete Time Regression Model

The vertical competing risks model (Nicolaie et al., 2010) is the latest addition to the continuous time competing risks models. The main objective of this chapter is to carry on with the task that began in Chapter 3 and advance a discrete time version of this model. In Chapter 3 we advanced the *continuous time mixture model* for application in discrete time and followed this up with a *discrete time mixture model* in Chap4. The vertical model is the final model that we advance as a discrete time model in this thesis.

At the centre of the framework that was proposed by Prentice et al. (1978) is characterization of the bivariate distribution of (T;D) in terms of CSHs, whereafter, these quantities are estimated directly from data. The competing risks models that were suggested afterwards merely suggest an alternate means for estimating these quantities. We have already discussed the mixture competing risks model in Chapter 3 and Chapter 4. The parameters of the mixture competing risks model are component hazards and failure type probabilities and as such, competing risks data is modeled with these quantities. The CSH estimates are then derived from component hazard and failure type probability estimates. The vertical model proposes yet another decomposition of the bivariate distribution of time to failure and failure type. The model proposes a re-expression of this joint distribution in terms of a marginal distribution for failure time and a distribution for failure type conditional on time to failure. Consequently, the parameters of the model become failure type probabilities conditional on failure time (relative hazards) and total hazards to characterize the failure type distribution conditional on failure time and the time to failure distribution, respectively. The definition of continuous time total hazards is given by

$$h(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt | T \geq t)}{dt}.$$

The total hazard $h(t)$ is the instanteneous risk of failure (by any cause of failure) at time t given survival to time t. The definition of relative hazards $\Pi_j(t)$ is given by

$$\Pi_j(t) = P(D = j | T = t).$$

The relative hazard $\Pi_j(t)$ is the probability that given that a failure has occured at time t the failure is due to failure type j. The relative hazards can also be written as;

$$\Pi_j(t) = P(D = j | T = t) = \frac{P(D = j, T = t)}{P(T = t)}. \tag{5.1.1}$$

If we use the fact that $P(A = a, A \geq a) = p(A = a)$, it follows that (5.1.1) can also be written as;

$$\begin{aligned} \Pi_j(t) &= \frac{P(D = j, T = t)}{P(T = t)} \\ &= \frac{P(D = j, T = t, T \geq t)}{P(T = t, T \geq t)} \\ &= \frac{P(D = j, T = t, T \geq t)/P(T \geq t)}{P(T = t, T \geq t)/P(T \geq t)} \\ &= \frac{P(D = j, T = t | T \geq t)}{P(T = t | T \geq t)} = \frac{h_j(t)}{h(t)}. \end{aligned} \tag{5.1.2}$$

We will show later that when data is modeled with the vertical model the full likelihood function mimics the assumption that underlies the model and splits into a failure time likelihood function and a conditional failure type likelihood function as specified in terms of total hazards and relative hazards, respectively. This means that data is now modeled with total hazards and relative hazards where these quantities are estimated seperately similar to component hazards and failure type probabilities when data is modeled with the mixture model. We can re-write 5.1.2 as;

$$h_j(t) = h(t)\Pi_j(t). \tag{5.1.3}$$

The model assumes a single failure time distribution for all subjects as summarized by total hazards after which the total hazards are apportioned via the relative hazards to the

CSHs per failure type as the original quantities that characterize the joint distribution of $(T; D)$ (Prentice et al., 1978). The CSHs are now obtained indirectly from the total hazard and relative hazard estimates from (5.1.3). Arguably, this is a round about way of estimating the CSHs, but, as already pointed out previously, there are instances when the CSHs cannot be estimated directly from data. A case in point is when data comes with unknown failure causes for some subjects. In such situations, the CSHs cannot be estimated directly as their estimation requires full information regarding the failure type for all failures. Analysis of data with this complication has already been discussed in the last two chapters. We will demonstrate, amongst other things, that the vertical model can serve both as an ordinary competing risks model as well as a model that can handle data that comes with missing failure causes as is in this chapter. This model occupies a unique position amongst the competing risks models that have been advanced this far in that it is the only model that can serve as an ordinary competing risks model as well as a model that can also handle the presence of missing failure causes in data without any structural modification.

To locate the model in discrete time we propose the following definition for discrete time total hazards;

$$h(t) = P(T = t|T \geq t) = \sum_{j=1}^{J} P(T = t; D = j|T \geq t) = \sum_{j=1}^{J} h_j(t)$$

for $t = 1, 2, \ldots, q$. In discrete time the total hazards are true probabilities just as CSHs are. The total hazards $h(t)$ is the probability of failure by any cause, at time t given survival to time t. Consistent with discrete time survival analysis models, see for example, (Allison, 1982; Singer and Willet, 2003; Tutz and Schmid, 2016), the regression model for the total hazards can be expressed as;

$$g(\lambda(t|\boldsymbol{x})) = \beta_{0t} + \boldsymbol{x}^T \boldsymbol{\beta}_1 \tag{5.1.4}$$

for $t = 1, \ldots q$, where the $g(.)$ is a link function within the GLM framework. The scalar β_{0t} is the baseline total hazard coefficient at time t and $\boldsymbol{\beta}_1$ is a vector of regression coefficients.

Since $D \in \{1, \ldots, J\}$, a multinomial model is the most natural model for relative hazards;

$$\Pi_j(t|\boldsymbol{x}) = \frac{\exp\left(\phi_{0jt} + \boldsymbol{x}^T\boldsymbol{\phi}_{1j}\right)}{\left\{1 + \sum_{l=1}^{J\text{-}1} \exp\left(\phi_{0lt} + \boldsymbol{x}^T\boldsymbol{\phi}_{1l}\right)\right\}} \tag{5.1.5}$$

where $j = 1, \ldots, J$-1, and $t = 1, \ldots, q$, and $\Pi_J(t|\boldsymbol{x}, \boldsymbol{\phi}) = 1 - \sum_{j=1}^{J\text{-}1} \Pi_j(t|\boldsymbol{x})$. The scalar ϕ_{0jt} is the duration coefficient at time t, and $\boldsymbol{\phi}_{1j}$ is a vector of regression coefficients. The relative hazards convey the same meaning in discrete time as well. All the usual functionals are now expressed in terms of total hazards and relative hazards. The regression expression that connects the other functionals to data is (5.1.3) re-expressed as;

$$h_j(t|\boldsymbol{x}, \boldsymbol{\theta}) = h(t|\boldsymbol{x}, \boldsymbol{\beta})\Pi_j(t|\boldsymbol{x}, \boldsymbol{\phi})$$

where, $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T)^T$. For example, the expression for the CIF is given by

$$F_j(t|\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{j=1}^{J} S(s\text{-}1|\boldsymbol{x}, \boldsymbol{\beta})h(s|\boldsymbol{x}, \boldsymbol{\beta})\Pi_j(s|\boldsymbol{x}, \boldsymbol{\phi}) \tag{5.1.6}$$

for $t = 1, 2, \ldots, q$, and for $j = 1, 2, \ldots, J$.

Proceeding further, we need to discuss the estimation of the relative hazards and total hazards conditional on covariates by the way of estimating the corresponding parameters as collected in $\boldsymbol{\theta}$. Subsection 5.1.1., is dedicated to the estimation of $\boldsymbol{\theta}$. We follow this up with the application of the proposed model together with the multinomial model in Subsection 5.1.2. In Section 5.2 we upgrade the model into a missing failure causes model. The CIF standard errors are presented in Appendix C.

5.1.1 Estimation

To determine the MLE of $\boldsymbol{\theta}$ we maximize the log-likelihood function; $\mathcal{L}(\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$. In the construction of $\mathcal{L}(\boldsymbol{\theta})$ the contribution of a subject i that failed at time t_i is now $P(D_i = j|T_i = t_i; \boldsymbol{x}_i, \boldsymbol{\phi})P(T_i = t_i|\boldsymbol{x}_i, \boldsymbol{\beta})$, from the assumption underlying the vertical model. Let $d_{ij} = I(D_i = j)$ and $d_i = \sum_{j=1}^{J} d_{ij}$. The log-likelihood $\mathcal{L}(\boldsymbol{\theta})$, can be written as;

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}\log P(D_i=j|T_i=t_i;\boldsymbol{x}_i,\boldsymbol{\phi})P(T_i=t_i|\boldsymbol{x}_i,\boldsymbol{\beta}) + (1-d_i)\log P(T_i>t_i|\boldsymbol{x}_i,\boldsymbol{\beta})\\ &= \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}\log P(D_i=j|T_i=t_i;\boldsymbol{x}_i,\boldsymbol{\phi})\\ &+ \sum_{i=1}^{n} d_i\log P(T_i=t_i|\boldsymbol{x}_i,\boldsymbol{\beta}) + (1-d_i)\log P(T_i>t_i|\boldsymbol{x}_i,\boldsymbol{\beta})\\ &= \mathcal{L}(\boldsymbol{\phi}) + \mathcal{L}(\boldsymbol{\beta})\end{aligned}$$

The observed data log-likelihood function $\mathcal{L}(\boldsymbol{\theta})$, therefore, splits into $\mathcal{L}(\boldsymbol{\phi})$, a relative hazards log-likelihood function and $\mathcal{L}(\boldsymbol{\beta})$, a standard univariate failure time log-likelihood function. The relative hazards log-likelihood can be written as;

$$\mathcal{L}(\boldsymbol{\phi}) = \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}\log \Pi_j(t_i|\boldsymbol{x}_i;\boldsymbol{\phi})$$

The relative hazards log-likelihood function $\mathcal{L}(\boldsymbol{\phi})$, is easily recognizable as a kernel of a multinomial log-likelihood function. Thus, $\boldsymbol{\phi}$ can be estimated by fitting a multinomial distribution to the original data by including duration as factor because the failure type probailities are determined at each observed failure time, i.e., at $t \in \{1, 2 \ldots, q\}$. By definition, censored subjects are excluded from the estimation of relative hazards. The failure time log-likelihood function $\mathcal{L}(\boldsymbol{\beta})$, is a standard failure time log-likelihood function. Note that since time to failure is discrete, the definition of the density function is given by

$$P(T=t|\boldsymbol{x}) = \frac{\lambda(t|\boldsymbol{x};\boldsymbol{\beta})}{(1-\lambda(t|\boldsymbol{x};\boldsymbol{\beta}))}\prod_{s=1}^{t}(1-\lambda(s|\boldsymbol{x};\boldsymbol{\beta}))$$

Thus, the failure time log-likelihood $\mathcal{L}(\boldsymbol{\beta})$ can be written as;

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}) &= \sum_{i=1}^{n} d_i\log\frac{\lambda(t_i|\boldsymbol{x}_i;\boldsymbol{\beta})}{(1-\lambda(t_i|\boldsymbol{x}_i;\boldsymbol{\beta}))} + \sum_{s=1}^{t_i} d_i\log(1-\lambda(s|\boldsymbol{x}_i;\boldsymbol{\beta}))\\ &+ \sum_{s=1}^{t_i}(1-d_i)\log(1-\lambda(s|\boldsymbol{x}_i;\boldsymbol{\beta}))\\ &= \sum_{i=1}^{n} d_i\log\frac{\lambda(t_i|\boldsymbol{x}_i;\boldsymbol{\beta})}{(1-\lambda(t_i|\boldsymbol{x}_i;\boldsymbol{\beta}))} + \sum_{s=1}^{t_i}\log(1-\lambda(s|\boldsymbol{x}_i;\boldsymbol{\beta}))\end{aligned}$$

If we define d_{is} such that $d_{is} = 0$ for $s \leq t_i$-1 and $d_{it_i} = d_i$, then $\mathcal{L}(\boldsymbol{\beta})$, can be re-written as;

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{s=1}^{t_i} d_{is}\log\lambda(s|\boldsymbol{x}_i;\boldsymbol{\beta}) + (1-d_{is})\log(1-\lambda(s|\boldsymbol{x}_i;\boldsymbol{\beta}))$$

This means that $\boldsymbol{\beta}$ can be estimated within the GLM framework by fitting a binomial distribution, where $d_{is} \sim \mathcal{B}(1, \lambda(s|\boldsymbol{x}_i;\boldsymbol{\beta}))$. The data must be re-arranged into a long format, where subject i contributes $\boldsymbol{d}_i = (d_{i1}, \ldots, d_{it_i})^T$, as a response vector of successes out of $(\overbrace{1, \ldots, 1}^{t_i})^T$, trials with $\boldsymbol{x}_i$ repeated t_i times. Thus, both the total hazards and the relative hazards can be estimated with standard software packages for the binomial and the multinomial distribution. Naturally, this ensures consistency and asymptotic normalcy of $\hat{\boldsymbol{\theta}}$

5.1.2 Application

To illustrate the application of the proposed discrete time vertical regression model, we consider the same unemployment data that comes with `Ecdat` R package (Croissant and Graves, 2020). We continue to test the proposition by McCall (1996) that provision of unemployment benefits together with an upward adjustment of the disregard rate improves the employment rate. As before, we also apply the proposed model to assess the effect of unemployment benefits on re-employment prospects. We continue to assess the effect of increasing the disregard rate by 50% via its effect on the CIF. We also consider all the covariates that came with the data set where the continuous covariates are centered at their respective averages and the unemployment benefit recipients are set as a reference category. Since $J = 2$, we model the relative hazards with a binomial distribution;

$$\Pi_1(t|\boldsymbol{x}, \boldsymbol{\phi}) = \frac{\exp(\phi_{0t} + \boldsymbol{x}^T\boldsymbol{\phi}_1)}{1 + \exp(\phi_{0t} + \boldsymbol{x}^T\boldsymbol{\phi}_1)}$$

where full-time and part-time re-employment are regarded as failure cause 1 and 2, respectively, with failure cause 1 as our cause of interest. The total hazards are modeled with a binomial distribution as well;

$$h(t|\boldsymbol{x}, \boldsymbol{\beta}) = \frac{\exp(\beta_{0t} + \boldsymbol{x}^T\boldsymbol{\beta}_1)}{1 + \exp(\beta_{0t} + \boldsymbol{x}^T\boldsymbol{\beta}_1)}.$$

Table 5.1: Maximum likelihood estimates for the Discrete Vertical Model (with standard errors) (* denotes $P < 0.05$).

	Discrete Time Vertical Model	
Coefficient	$\hat{\boldsymbol{\phi}}$	$\hat{\boldsymbol{\beta}}$
`T1`	1.194(0.176)*	-2.399(0.067)*
`T2`	1.214(0.187)*	-2.707(0.078)*
`T3`	1.251(0.208)*	-2.902(0.090)*
`T4`	0.956(0.268)*	-3.434(0.122)*
`T5`	1.190(0.208)*	-2.627(0.095)*
`T6`	1.264(0.374)*	-3.667(0.161)*
`T7`	1.311(0.241)*	-2.487(0.106)*
`T8`	0.354(0.432)	-3.846(0.210)*
`T9`	1.527(0.423)*	-3.208(0.166)*
`T10`	0.049(0.838)	-5.001(0.412)*
`T11`	1.667(0.496)*	-3.211(0.188)*
`T12`	0.847(0.698)	-4.217(0.322)*
`T13`	1.130(0.416)*	-2.868(0.185)*
`T14`	1.682(0.460)*	-2.548(0.185)*
`T15`	1.925(0.629)*	-2.804(0.227)*
`T16`	1.084(0.678)	-3.149(0.291)*
`T17`	2.205(1.072)	-3.298(0.346)*
`T18`	1.279(0.820)	-3.115(0.349)*
`T19`	0.992(0.413)	-1.483(0.212)
`ui`	-0.062(0.150)	1.191(0.060)*
`dr`	-1.292(1.010)	-1.379(0.457)*
`rr`	1.116(0.968)*	0.478(0.406)
`age`	-0.012(0.007)*	-0.012(0.003)
`wage`	0.991(0.192)*	0.293(0.085)
`tenure`	0.002(0.013)*	-0.002(0.006)

Table 5.2: Maximum likelihood estimates for the Discrete Time Mixture Model and the Multinomial Model (with standard errors) (* denotes $P< 0.05$).

	Mixture Model (Latency Model)		Multinomial Model	
Coefficient	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
T1	-2.583(0.074)*	-2.277(0.124)*	-2.694(0.077)*	-3.823(0.130)*
T2	-2.796(0.085)*	-2.677(0.150)*	-2.989(0.088)*	-4.173(0.153)*
T3	-2.936(0.098)*	-2.925(0.176)*	-3.181(0.102)*	-4.381(0.179)*
T4	-3.468(0.138)*	-3.314(0.227)*	-3.762(0.141)*	-4.766(0.229)*
T5	-2.655(0.103)*	-2.655(0.180)*	-2.926(0.108)*	-4.045(0.182)*
T6	-3.608(0.179)*	-3.714(0.322)*	-3.949(0.183)*	-5.134(0.324)*
T7	-2.436(0.112)*	-2.699(0.212)*	-2.747(0.119)*	-4.029(0.215)*
T8	-3.926(0.260)*	-3.484(0.338)*	-4.328(0.263)*	-4.857(0.341)*
T9	-3.025(0.176)*	-3.643(0.383)*	-3.411(0.182)*	-4.985(0.385)*
T10	-5.267(0.578)*	-4.353(0.580)*	-5.704(0.580)*	-5.731(0.582)*
T11	-2.989(0.198)*	-3.760(0.451)*	-3.394(0.205)*	-5.082(0.454)*
T12	-4.134(0.379)*	-4.130(0.580)*	-4.580(0.383)*	-5.473(0.583)*
T13	-2.740(0.202)*	-3.062(0.359)*	-3.150(0.210)*	-4.340(0.363)*
T14	-2.318(0.184)*	-3.174(0.413)*	-2.733(0.196)*	-4.403(0.418)*
T15	-2.461(0.231)*	-3.676(0.581)*	-2.956(0.243)*	-4.850(0.585)*
T16	-2.871(0.317)*	-3.496(0.581)*	-3.421(0.329)*	-4.660(0.587)*
T17	-2.835(0.354)*	-4.405(1.002)*	-3.417(0.367)*	-5.563(1.006)*
T18	-2.695(0.379)*	-3.588(0.710)*	-3.366(0.392)*	-4.697(0.717)*
T19	-1.211(0.214)*	-1.920(0.340)*	-1.823(0.242)*	-2.804(0.357)*
ui	1.505(0.065)*	0.556(0.116)*	1.184(0.068)*	1.199(0.119)*
dr	-0.447(0.504)	-1.729(0.835)*	-1.731(0.529)*	-0.591(0.821)
rr	0.480(0.456)	-0.025(0.784)	0.900(0.464)*	-0.334(0.742)
age	-0.011(0.003)*	-0.008(0.006)	-0.015(0.004)*	-0.003(0.006)
wage	0.227(0.096)*	0.039(0.150)	0.536(0.099)*	-0.390(0.149)
tenure	-0.014(0.006)*	0.044(0.011)*	0.002(0.006)*	0.003(0.011)

Table 5.3: Maximum likelihood estimates for the Discrete Time Mixture Model (with standard errors) (* denotes $P < 0.05$).

Mixture Model	
(Incidence Model)	
Coefficient	$\hat{\phi}$
Constant	1.178(0.062)*
`ui`	-0.488(0.092)*
`dr`	-2.381(0.674)*
`rr`	0.822(0.668)*
`age`	-0.009(0.005)*
`wage`	0.897(0.131)*
`tenure`	0.038(0.009)*

The results of fitting the proposed model are displayed in Table 5.1. In advancing both the *continuous time mixture model* and the *discrete time mixture model* in the previous two chapters, we have validated these models against the multinomial model, an established model for regression analysis of discrete time competing risks data. Recall that we found that the mixture models compared more favourably with each other than each with the multinomial model because the mixture models tended to produce almost identical CIF estimates in comparison to the CIF estimates that were produced by the multinomial model.

We continue with this approach regarding the proposed discrete time vertical model where we compare the proposed model to the multinomial model and the *discrete time mixture model*. Accordingly, we have also included the results of fitting the *discrete time mixture model* from Chapter 4 and a multinomial model in Table 5.2 and Table 5.3. When data is modeled with the proposed vertical model the estimates for the CIF are derived from $\hat{\boldsymbol{\theta}}$;

$$\hat{F}_j(t|\boldsymbol{x}, \hat{\boldsymbol{\theta}}) = \sum_{s=1}^{t} \hat{S}(s\text{-}1|\boldsymbol{x}, \hat{\boldsymbol{\beta}})\hat{h}(s|\boldsymbol{x}, \hat{\boldsymbol{\beta}})\hat{\Pi}_j(t|\boldsymbol{x}, \hat{\boldsymbol{\phi}}). \tag{5.1.7}$$

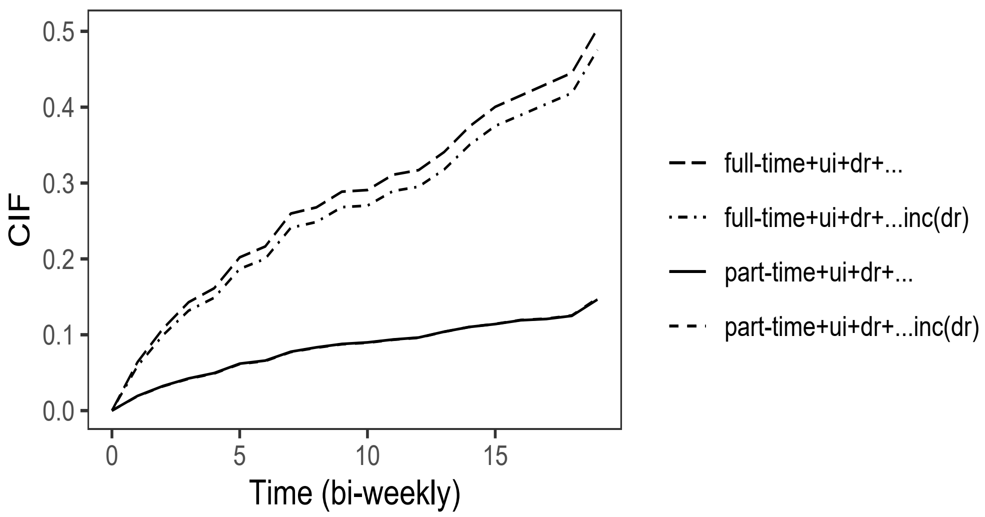

Figure 5.1: The CIF of exit to full-time and part-time employment for the `ui` recipients with the effect of increasing `dr` via the Discrete Time Vertical Model.

In Figure 5.1 we plotted the CIF estimate with the effect of increasing the disregard rate by 50%. According to Figure 5.1, increasing the disregard rate has the effect of reducing full-time re-employment and no noticeable impact on part-time employment. We reached the same conclusions when we tested McCall (1996)'s proposal regarding the effect of increasing the disregard rate via the *continuous time mixture model*, the *discrete time mixture model* and the multinomial model.

We now proceed to compare the proposed vertical model with the multinomial model and the *discrete time mixture model*. In Chapter 4 we found that according to the *discrete time mixture model* the chances of re-employment for an individual from the reference group (unemployment individuals who receive benefits with average values for the continuous covariates) are 68.9% and these chances drop to 67.1% if the disregard rate is increased by 50%, but holding the continuous covariates at their respective averages. The multinomial model suggested that these chances are 64.4% for an individual from the reference group and these chances dip to 61.6% with an increase in the disregard rate. The proposed model suggests that the probability of re-employment for an individual from the reference group is 64.9% and this drops to 62.4% with an increase in the disregard rate. Clearly these figures suggest that the multinomial model and the proposed vertical model also seem to compare more favourably with one another, in terms of producing similar CIF estimates, than each model compares to the *discrete time mixture model* or the *continuous time mixture model* by implication. In Figure 5.2 we have plotted the CIF estimates with the effect of increasing the disregard rate by 50% via the proposed vertical model, the *discrete time mixture model* and the multinomial model. The plot confirms that all the three models lead to the conclusion that raising the disregard rate does not induce an in increase in re-employment as claimed by McCall (1996), if anything, it leads to a reduction in re-employment, albeit, marginal, but certainly not an improvement in the re-employment rate. It is evident from Figure 5.2 that the movement in part-time re-employment does not offset a clear reduction in full-time employment such that there is an overall improvement in re-employment due to an upward adjustment of the disregard rate. Furthermore, Figure 5.2 confirms the observation that the proposed vertical model tends to compare more favorably with the multinomial model than with the *discrete time mixture model*. We did highlight the fact that McCall (1996) used more covariates than the covariates that came with the data set which could be the reason we have reached conclusions that are different from McCall (1996). Another reason could be that we included fewer subjects in the analysis than McCall (1996) or that the models we have used are defective. We have managed to eliminate the posibility that the discrete time models that we have advanced could be at fault. The fact that we reached the same conclusions via the multinomial model, an established discrete time model, rules out that possibility.

We also applied the three models to assess the effect of unemployment benefits on re-employment prospects.

We found that according to the *discrete time mixture model* the chances of re-employment increase from 68.9% for an individual from the reference group to 92.3% for an average individual who does not receive unemployment benefits. According to the *continuous*

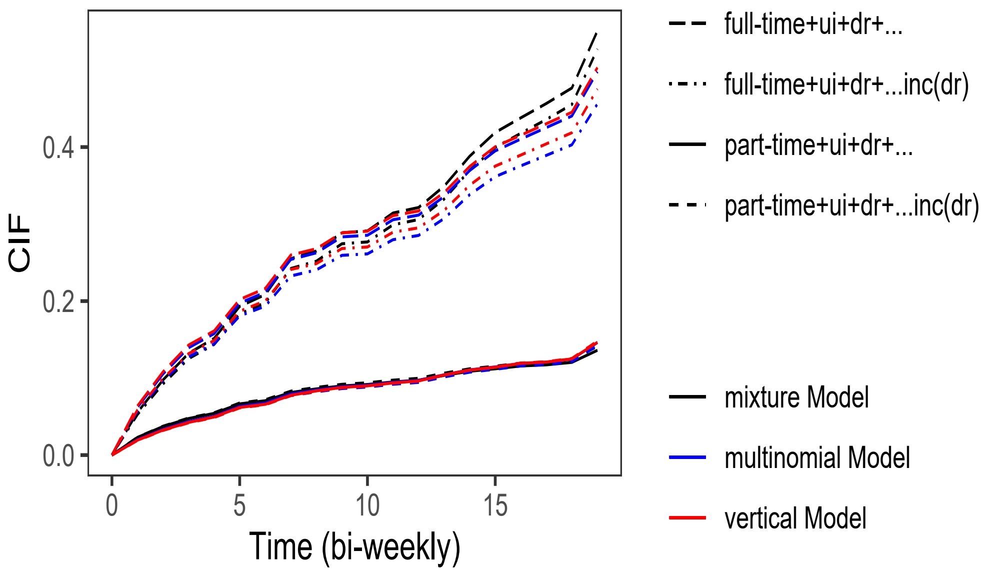

Figure 5.2: The CIF of exit to full-time and part-time employment for the `ui` recipients with the effect of increasing `dr` via the Discrete Time Vertical Model and the Multinomial Model.

time mixture model the probaility increases from 68.7% for a benefit recipient to 92.4 % for a nonrecipient. The multinomial model suggests that the chances increase from 64.4% to 95.6% while the proposed model suggests that the chances increase from 64.9% to 95.8%. Again, all three models lead us to the same conclusion regarding provision of unemployment benefits. Furthermore, these figures support the view that the mixture models tend to compare more favourablly with each other while the multinomial model and the proposed discrete time vertical model also seem to compare more favourably with one another. In Figure 5.3 we plotted the CIF estimates with the unemployment benefit effects.

Figure 5.3 provides further illustration that the proposed discrete time vertical model and the multinomial model tend to compare more favorably with each other, and by implication less favorably with the mixture models.

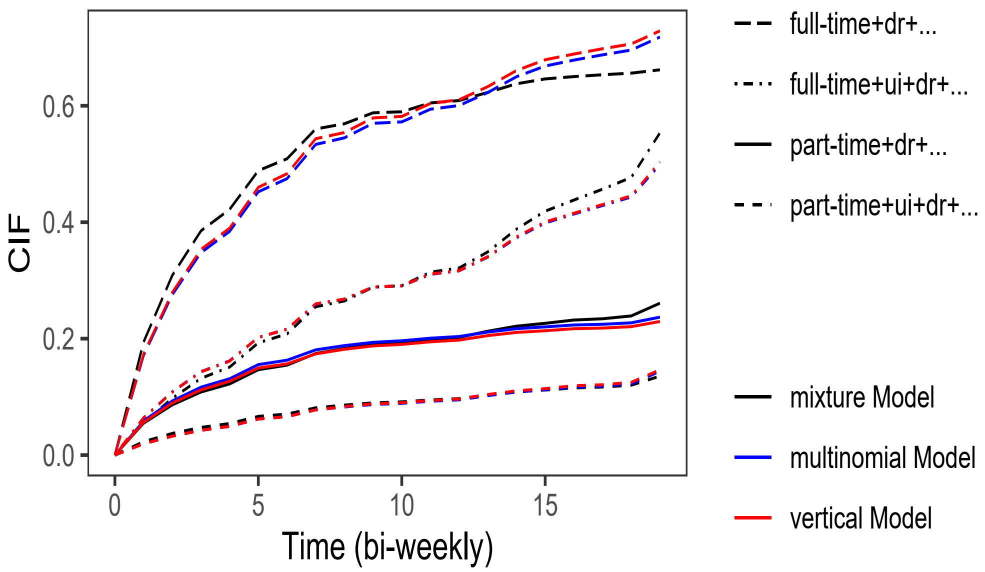

Figure 5.3: The CIF of exit to full-time and part-time employment with the effect `ui` via the Discrete Time Vertical Model, the Multinomial Model and the Discrete Time Mixture Model.

With regard to both the *discrete time mixture model* and *continuous time mixture models* we discussed the interpretation of regression coefficient estimates as it relates to their impact on the CIF. It is also instructive to discuss the interpretation of the relative hazard and total hazard parameter estimates for the proposed model. To simplify this conversation, suppose that we wished to assess the effect on re-employment that is induced by provision of unemployment benefits, that is, if provision of unemployment benefits improves the prospects of re-employment or not. Let us first consider the intepretation of $\hat{\phi}_{1\texttt{ui}}$ and $\hat{\beta}_{\texttt{ui}}$. When $J > 2$ a multinomial distribution is fitted to data excluding censored subjects to estimate $\boldsymbol{\phi}$. Here, a baseline or reference failure cause is chosen beforehand say failure cause J and then J-1 pairs of logit models;

$$\log\left(\frac{\Pi_j(t|\boldsymbol{x},\boldsymbol{\phi})}{\Pi_J(t|\boldsymbol{x},\boldsymbol{\phi}}\right) = \phi_{0jt} + \boldsymbol{x}^T\boldsymbol{\phi}_{1j} \tag{5.1.8}$$

for $j = 1, 2, \ldots, J\text{-}1$, are estimated simultaneously where $\Pi_J(t|\boldsymbol{x}, \boldsymbol{\phi}) = \sum_{j=1}^{J\text{-}1} \Pi_j(t|\boldsymbol{x}, \boldsymbol{\phi})$. Whether we write $\frac{\Pi_j(t)}{\Pi_J(t)}$ or $\frac{h_j(t)}{h_J(t)}$ really makes no difference because;

$$\begin{aligned}\frac{\Pi_j(t)}{\Pi_J(t)} &= \frac{h_j(t)/\sum_{j=1}^{J} h_j(t)}{h_J(t)/\sum_{j=1}^{J} h_j(t)}\\ &= \frac{h_j(t)}{h_J(t)}.\end{aligned}$$

The term: relative hazards could, therefore, easily have been replaced by the term: relative cause-specific-hazards. Equation (5.1.8) says, given that a subject that is described by covariate vector $\boldsymbol{x}$ may fail by cause j or cause J at time t, the log of odds that the subject will fail by cause j is: $\phi_{0jt} + \boldsymbol{x}^T\boldsymbol{\phi}_{1j}$. If we wished to compare the odds of failure by cause j relative to cause k for the same subject, then;

$$\begin{aligned}\log\left(\frac{\Pi_j(t|\boldsymbol{x}, \boldsymbol{\phi})}{\Pi_k(t|\boldsymbol{x}, \boldsymbol{\phi}}\right) &= \log\left(\frac{\Pi_j(t|\boldsymbol{x}, \boldsymbol{\phi})/\Pi_J(t|\boldsymbol{x}, \boldsymbol{\phi}}{\Pi_k(t|\boldsymbol{x}, \boldsymbol{\phi})/\Pi_J(t|\boldsymbol{x}, \boldsymbol{\phi}}\right)\\ &= \log\left(\frac{\Pi_j(t|\boldsymbol{x}, \boldsymbol{\phi})}{\Pi_J(t|\boldsymbol{x}, \boldsymbol{\phi}}\right) - \log\left(\frac{\Pi_k(t|\boldsymbol{x}, \boldsymbol{\phi})}{\Pi_J(t|\boldsymbol{x}, \boldsymbol{\phi}}\right)\\ &= \phi_{0jt} + \boldsymbol{x}^T\boldsymbol{\phi}_{1j} - \phi_{0kt} + \boldsymbol{x}^T\boldsymbol{\phi}_{1k}\\ &= (\phi_{0jt} - \phi_{0kt}) + \boldsymbol{x}^T(\boldsymbol{\phi}_{1j} - \boldsymbol{\phi}_{1k}).\end{aligned}$$

For this data there are only two failure causes and as such a binomial distribution was fitted to data to estimate the relative hazards as given in Table 5.1. Recall that the benefit recipients are regarded a the reference category while the continuous covariates are centered at their respective averages. Beginning with $\hat{\phi}_{01t}$, the following is true;

$$\frac{\hat{h}_1(t|\boldsymbol{x} = \boldsymbol{0}, \hat{\boldsymbol{\phi}})}{\hat{h}_2(t|\boldsymbol{x} = \boldsymbol{0}, \hat{\boldsymbol{\phi}})} = \exp(\hat{\phi}_{01t}). \tag{5.1.9}$$

Since $\hat{\phi}_{01t} > 0$ for all t, it means that *given* that a job has been landed, that job is more likely to be a full-time job than a part-time job for recipients of the unemployment benefits. These odds drop by $(\exp(\hat{\phi}_{ui}) = 0.939)$, about 6% in favor of a full-time job for nonrecipients. Relative hazards is a measure that is not as popular as CSHs with applied researchers. Re-consider the graduation/dropout example introduced in Chapter 2. Such studies are often undertaken to identify those factors/variables which explain graduation/dropout with the view to use this information when intervention strategies are developed to improve graduation and retention rates, and more importantly to identify when it is more opportune to implement these interventions so as to maximize their impact. Typically, a plot of the CSH estimates of graduation/dropout against time can reveal the timing as to when it is most beneficial to introduce the interventions. Suppose that from this plot of CSHs vs time we are able to identify time t_g as the time when students are least likely to graduate and t_d as the time when the students are most likely to drop out. To improve graduations, authorities must aim to intervene before time t_g, and likewise, to improve retention rates, interventions must be introduced before time t_d. Another way of looking at the same problem of identifying the best time to introduce support measures is to find the time(s) where students are most likely to drop out than to graduate. Put differently, to identify the time(s) when the least desirable (dropout) outcome/event is most likely to occur than the most desirable outcome (graduation). Now, given the present settings where the events under consideration are part-time and full-time employment let us suppose that the most desirable event is full-time employment and part-time employment is the least desirable outcome, maybe because full-time employment is more meaningful and longer lasting than part-time employment in the eyes of authorities.

We plotted the relative $RH_{F/P}$ against time in Figure 5.4 where;

$$RH_{F/P} = \frac{h_1(t)}{h_2(t)}.$$

We can deduce from the plot that at time $T = 10$, full-time re-employment is least likely to occur in comparison to or relative to part-time re-employment for both benefit recipients and nonrecipients. Since $RH_{P/F} = \frac{1}{RH_{F/P}}$, it also means that at $T = 10$ the risk of exit to part-time employment is at its largest relative to risk of exit to full-time employment where this risk is higher for nonrecipients of benefits compared to recipients as confirmed by Figure 5.5

For a moment let us assume that the full-time/part-time problem was in fact graduation /dropout example. Let us imagine that a full-time re-employment is a graduation and a part-time re-employment is a dropout while unemployment benefits recipients could be male students and female students are nonrecipients. We can imagine that the continuous variables to be age and mathematics marks and so on. Then, time $T = 10$ is when students

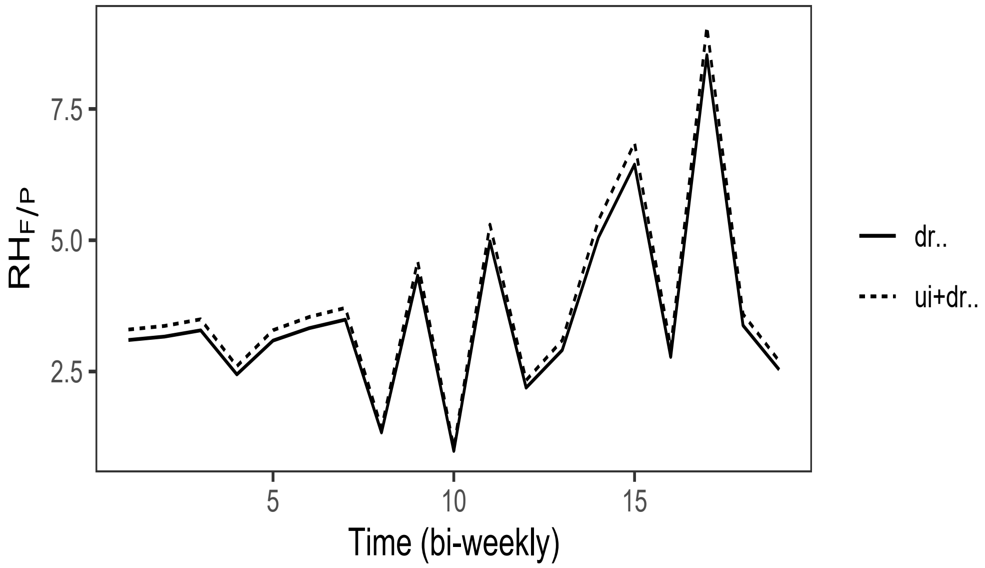

Figure 5.4: The Relative Hazards of exit to full-time employment with the effect `ui` via the Discrete Time Vertical Model.

are most likely to droupout than to graduate where male students are more likely than female students to drop out. Thus, if the authorities are planning to intervene, it would be ideal to do so before $T = 10$ taking into account that male students are at a higher risk of dropping out than female students in general including time $T = 10$. The risk of dropping out for male students is ($\frac{1}{exp(-0.0621)} = 1.06$) about 6% higher than the risk of dropping out for female students.

Going back to the full-time/part-time scenario, if we examine Table 5.1 we can tell that $RH_{[F/P]}$ is at the minimum at $T = 10$ because $\min\{\phi_{01}, \phi_{02} \ldots, \phi_{019}\}$, is $\phi_{010} = 0.049$. Furthermore, $\phi_{1\texttt{ui}} = -0.0621 < 0$. This means that the probability that an unemployed individual finds part-time employment instead of full-time employment is at its largest at time $T = 10$, and this probability is higher for unemployment benefit nonrecipients.

Note that the movement of relative hazards due to some effect of a variable of interest

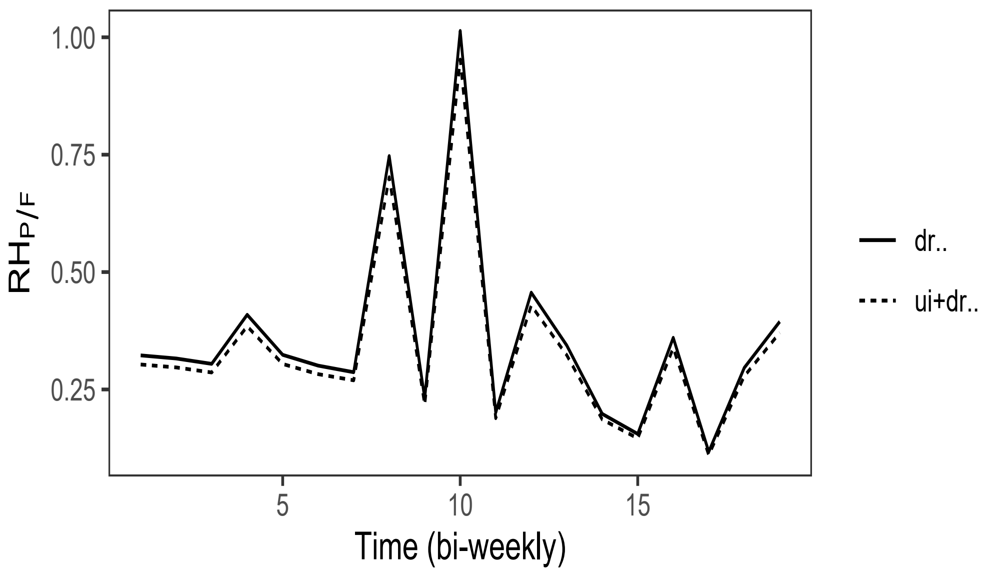

Figure 5.5: The Relative Hazards of exit to part-time employment with the effect `ui` via the Discrete Time Vertical Model.

can also be easily determined from the multinomial model parameters with minimal re-arrangement of terms. When the multinomial model is fitted to data, usually, censored subjects are regarded as a reference category such that J logit models;

$$\log \frac{h_j(t|\boldsymbol{x}, \boldsymbol{\alpha})}{h_0(t|\boldsymbol{x}, \boldsymbol{\alpha})} = \alpha_{0jt} + \boldsymbol{x}^T \boldsymbol{\alpha}_{1j} \tag{5.1.10}$$

are estimated simulteneously where $h_0(t|\boldsymbol{x}, \boldsymbol{\alpha}) = 1 - \sum_{j=1}^{J} h_j(t|\boldsymbol{x}, \boldsymbol{\alpha})$, is the probability of surviving beyond time t. Here as well, if interest is centered on comparing cause j to cause k we can write the log of odds as;

$$\log \frac{h_j(t|\boldsymbol{x}, \boldsymbol{\alpha})}{h_k(t|\boldsymbol{x}, \boldsymbol{\alpha})} = (\alpha_{0jt} - \alpha_{0kt}) + \boldsymbol{x}^T (\boldsymbol{\alpha}_{1j} - \boldsymbol{\alpha}_{1k}). \tag{5.1.11}$$

In Figure 5.6 we have plotted the probability of exit to full-time employent relative to exit to part-time employment via the proposed vertical model and the multinomial model using (5.1.11). It can be seen that the relative hazards compare favorably.

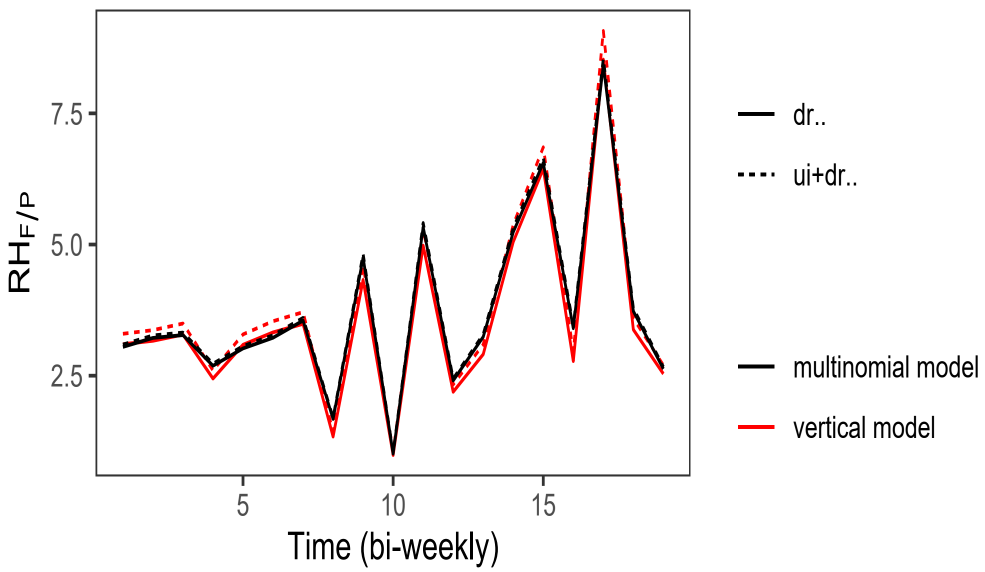

Figure 5.6: The Relative Hazards of exit to full-time employment with the effect `ui` via the Discrete Time Vertical Model and the Multinomial Model.

Obviously, if the relative hazards approach is the preferred method to compare failure causes, the advantage of the vertical model over the multinomial model is that data does not have to be re-arranged into the long format to estimate the relative hazards and the regression coefficients are readily available with no need to manipulate the coefficients as when data is modeled with the multinomial model.

Regarding $\hat{\boldsymbol{\beta}}$, since $\hat{\boldsymbol{\beta}}_{\texttt{ui}} < 0$, it means that holding the continuous covariates at their respective average values, the odds of re-employment to full-time or part-time employment are lower for `ui` recipients than for non-recipients by $\left(1 - \frac{1}{\exp(1.191)} = 0.69\right)$, i.e., about 69%. Overall, `ui` recipients tend to search for work less intensively than nonrecipients and as a result, fewer individuals land jobs (full-time or part-time) amongst `ui` recipients than amongst nonrecipients. In Figure 5.7 we have plotted the total hazards and it can be seen that both total hazards decrease until around week 20 ($T = 10$) and then pick up somewhat thereafter. In the US, most of the states provide unemployment benefits for

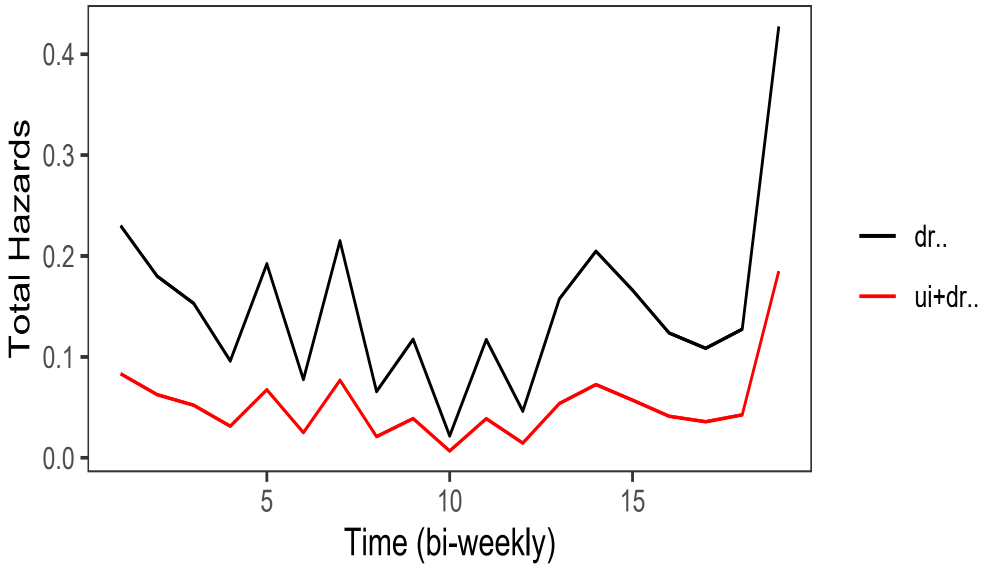

Figure 5.7: The Total Hazard of re-employment with the effect `ui` via the Discrete Time Vertical Model.

the first 26 weeks. That might be the explanation for the upward movement in the total hazard for the `ui` recipients. A possible explanation for similar movement in the total hazard curve for nonrecipients could be that the recently jobless individuals would be nearing exhausting their reserves and are, therefore, expected to double up their efforts to find employment also around that period. Thus, we were able to determine from ($\hat{\boldsymbol{\beta}}_{\texttt{ui}} < 0$), that unemployment benefits tend to reduce the prospects of re-employment across failure causes. It is difficult to predict the movement of the CSHs from the multinomial model because the regression coefficients are related to the log odds of a hazard of failure by a given cause of failure relative to the hazard of no failure if censored subjects are taken to be the reference category. It is, therefore, near impossible to predict the movement of total hazards from the regression coefficients.

In Chapter 4 we demonstrated as to when it is possible to predict the effect of a variable on the CIF from regression coefficients. Recall that the regression model for the CIF under the proposed model is given by

$$\hat{F}_j(t|\boldsymbol{x},\hat{\boldsymbol{\theta}}) = \sum_{s=1}^{t} \hat{S}(s\text{-}1|\boldsymbol{x},\hat{\boldsymbol{\beta}})\hat{h}_j(s|\boldsymbol{x},\hat{\boldsymbol{\beta}})\hat{\Pi}_j(s|\boldsymbol{x},\hat{\boldsymbol{\phi}}). \quad (5.1.12)$$

The survival function and the total hazard will always move in opposite directions regardless of what happens to the relative hazards. Thus, it is not possible to predict the effects of a variable on the CIF from the regression coefficients. In fact this model for the CIF is just as complicated as the CSH denominated regression model for the CIF that arises under the multinomial model, if not more complicated.

We can, however, predict the covariate effects on CSHs from the regression coefficients of total hazards and relative hazards in a limited way. Recall that the CSH estimates are now recovered from total hazard and relative hazard estimates;

$$\hat{h}_j(t|\boldsymbol{x},\hat{\boldsymbol{\theta}}) = \hat{h}(t|\boldsymbol{x},\hat{\boldsymbol{\beta}})\hat{\Pi}_j(t|\boldsymbol{x},\hat{\boldsymbol{\phi}}). \quad (5.1.13)$$

Here, again, if the covariate in question induces the total hazards and relative hazards to move in the same direction we can predict its effect on the CSH of interest. Say, we wished to predict the covariate effect of `ui`. Now, $\hat{\beta}_{1\mathtt{ui}} = 1.191 > 0$, and $\hat{\phi}_{1\mathtt{ui}} = -0.062 < 0$. This means that moving from recipients to nonrecipients, the total hazard increases, but the relative hazards are reduced. Effectively we cannot predict the effect of `ui` on the full-time CSHs, but if `ui` induces a reduction in $RH_{F/P}$, it then induces an increase in $RH_{P/F}$ because $RH_{P/F} = 1 - RH_{F/P}$. Thus, the part-time CSHs for benefit nonrecipients are larger than the part-time CSHs for benefit recipients. In Figure 5.8 we have plotted the CSH estimates with the effect of unemployment benefits. It is evident from the plot that receipt of unemployment benefits has the effect of dampening the risk of exit to both full-time and part-time re-employment prospects. Applied reseachers would ideally use a similar plot to identify the factors/variables that explain the risk of failure. Again, if we were to assume that full-time re-employment were graduations and part-time re-employment were dropouts where `ui`=1 represents male students and `ui`=0 are female students. To improve both graduation and retention rates, applied researchers would typically identify the time(s) when graduations were least likely to occur and the time(s) when dropouts were most likely to occur. The risk of a graduation is at its lowest at

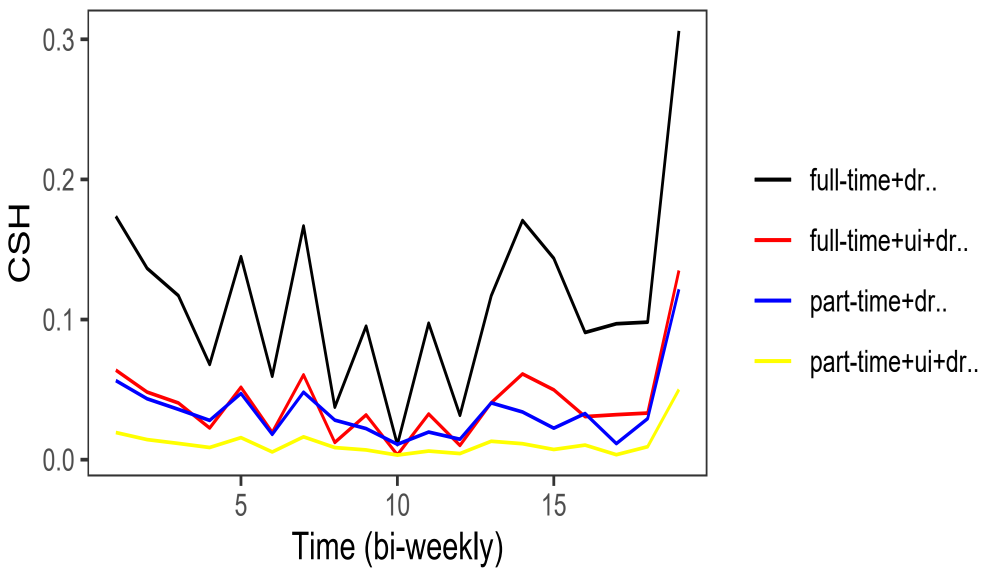

Figure 5.8: The CSHs of re-employment with the effect of `ui` via the Discrete Time Vertical Model.

time $T = 10$ and risk of a graduation is at its largest towards the end of the observation period. This is true for both male and female students, but with male students at a higher risk than their female counterparts to both graduate and drop out. This approach can be complimented with additional information provided by the relative hazards. Recall that we found that students are at the highest risk to drop out than to graduate at time $T = 10$ with male students at higher risk than female students.

It was easier to conduct this limited interpretation of covariate effects on CSHs with this data. The advantage with this data is that there are only two failure causes. When $J > 2$ things become complicated because one has to account for J regression coefficients simulteneously, one regression coefficient corresponding to the total hazard and J-1 from the model for relative hazards. If there is one cause of interest then other failure causes can be collapsed into one cause of failure such that relative hazards are modeled via a binomial model. Of course, in the end what really settles matters is to actually calculate the CSH and the CIFs estimates from (5.1.12) and (5.1.13), respectively. It is just that statisticians will always be on the lookout to simplify things where possible.

5.2 The Missing Failure Causes Vertical Model

In this section we conclude the second part of this thesis where we are addressing the analysis of data that comes with missing failure causes. In previous two chapters, that is, Chapter 3 and Chapter 4, we extended the mixture models into models that can handle missing failure causes. In this section we extend the vertical regression model from the previous section as another discrete time missing failure causes model. After introducing the vertical model (Nicolaie et al., 2010) for analysis of ordinary competing risks data, the authors of the model later demonstrated that the same model is also able to handle missing failure causes (Nicolaie et al., 2015). We also intend to follow in the footsteps of these authors and demonstrate that the regression model that was advanced in the previous chapter can also serve as missing failure causes regression model. .

To advance a discrete time vertical regression model for handling missing failure causes in this section we need to specify regression models for total hazards and relative hazards. We continue to model the failure type distribution conditional on failure time with a multinomial distribution where the model for relative hazards is given by

$$\Pi_j(t|\boldsymbol{x}) = \frac{\exp\left(\phi_{0jt} + \boldsymbol{x}^T\boldsymbol{\phi}_{1j}\right)}{1 + \sum_{l=1}^{J\text{-}1}\left(\exp\phi_{0lt} + \boldsymbol{x}^T\boldsymbol{\phi}_{1l}\right)}$$

for $j = 1, \ldots, J\text{-}1$, and $t = 1, \ldots, q$. The duration coefficient is represented by ϕ_{0jt} at time t, and $\boldsymbol{\phi}_{1j}$ is a vector of regression coefficients. We also continue to model the the total hazards within the GLM framework;

$$g(h(t|\boldsymbol{x})) = \beta_{0t} + \boldsymbol{x}^T\boldsymbol{\beta}_1$$

for $t = 1, \ldots, q$. The scalar β_{0t} is the baseline total hazard coefficient at time t and $\boldsymbol{\beta}_1$ is a vector of regression coefficients. We collect all the unknown parameters of the proposed model in $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T)^T$, where $\boldsymbol{\beta} = (\beta_{01}, \beta_{02} \ldots, \beta_{0q}, \boldsymbol{\beta}_1^T)^T$, $\boldsymbol{\phi} = (\boldsymbol{\phi}_1^T, \ldots, \boldsymbol{\phi}_{(J\text{-}1)}^T)^T$, and

$\boldsymbol{\phi}_j = (\phi_{0j1}, \ldots, \phi_{0jq}, \boldsymbol{\phi}_{1j}^T)^T$. The full likelihood function is now given by

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log P(D_i = j|T_i = t_i; \boldsymbol{x}_i, \boldsymbol{\phi}) P(T_i = t_i; \boldsymbol{x}_i, \boldsymbol{\beta}) \\
&+ d_{i*} \log P(T_i = t_i) + (1 - d_i - d_{i*}) \log P(T_i > t_i; \boldsymbol{x}_i, \boldsymbol{\beta}) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log P(D_i = j|T_i = t_i; \boldsymbol{x}_i, \boldsymbol{\phi}) + \big(\sum_{j=1}^{J} d_{ij} + d_{i*}\big) \log P(T_i = t_i; \boldsymbol{x}_i, \boldsymbol{\beta}) \\
&+ (1 - d_i - d_{i*}) \log P(T_i > t_i) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log P(D_i = j|T_i = t_i; \boldsymbol{x}_i, \boldsymbol{\phi}) + \sum_{i=1}^{n} \delta_i \log P(T_i = t_i; \boldsymbol{x}_i, \boldsymbol{\beta}) \\
&+ (1 - \delta_i) \log P(T_i > t_i; \boldsymbol{x}_i, \boldsymbol{\beta}) \\
&= \mathcal{L}(\boldsymbol{\phi}) + \mathcal{L}(\boldsymbol{\beta})
\end{aligned}$$

where $\delta_i = \sum_{j=1}^{J} d_{ij} + d_{i*}$. Clearly, the full log-likelihood function $\mathcal{L}(\boldsymbol{\theta})$, splits into $\mathcal{L}(\boldsymbol{\phi})$, and $\mathcal{L}(\boldsymbol{\beta})$, the log-likelihood functions for the conditional failure type and failure time, respectively. Thus $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$ continue to be estimated seperately. The conditional failure type log-likelihood function can be written as;

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\phi}) &= \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log P(D_i = j|T_i = t_i; \boldsymbol{x}_i, \boldsymbol{\phi}) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log \Pi_j(t_i; \boldsymbol{x}_i, \boldsymbol{\phi}).
\end{aligned}$$

This means that we continue to estimate $\boldsymbol{\phi}$ by applying a multinomial distribution where we include duration as a covariate to original data excluding censored subjects. Again, if we define δ_{is} such that $\delta_{is} = 0$ for $s \leq t_i$-1 and $d_{it_i} = d_i$, then $\mathcal{L}(\boldsymbol{\beta})$, can be re-written as;

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{s=1}^{t_i} \delta_{is} \log h(s|\boldsymbol{x}_i; \boldsymbol{\beta}) + (1 - \delta_{is}) \log(1 - h(s|\boldsymbol{x}_i; \boldsymbol{\beta}))$$

as shown in Chapter 5. Thus, $\boldsymbol{\beta}$ continues to be estimated by applying a binomial distribution to data in long format. Basically, the total hazards and relative hazards continue to be estimated exactly the same way in the presence or absence of missing failure causes where δ_{is} switches between $\delta_i = \sum_{j=1}^{J} d_{ij} + d_{i*}$, and $\delta_i = \sum_{j=1}^{J} d_{ij}$, according to the presence or absence of missing failure causes. Accordingly, we continue to implement standard packages to determine $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\phi}}$ which guarantees their consistency and asymptotic nomarlcy. In the next section we demonstrate the application of the proposed model.

5.2.1 Application

We continue to rely on unemployment data to illustrate the application of the proposed model. To illustrate the application of the proposed model we continue to test if raising the disregard rate for individuals that receive unemployment benefits will entice these individuals away from full-time employment towards part-time employment with the net effect of a rise in employment as claimed by McCall (1996). Towards that end, we consider all the covariates that came with data where unemployment benefit recipients are set as a reference category and continuous covariates are centered at their respective averages. We have re-labeled full-time and part-time employment as cause 1 and cause 2, respectively. Since $J = 2$, the relative hazards are modeled via a binomial model;

$$\Pi_1(t|\boldsymbol{x}, \boldsymbol{\phi}) = \frac{\exp(\phi_{0t} + \boldsymbol{x}^T \boldsymbol{\phi}_1)}{1 + \exp(\phi_{0t} + \boldsymbol{x}^T \boldsymbol{\phi}_1)}.$$

The total hazards are modeled with a binomial distribution as well;

$$h(t|\boldsymbol{x}, \boldsymbol{\beta}) = \frac{\exp(\beta_{0t} + \boldsymbol{x}^T \boldsymbol{\beta}_1)}{1 + \exp(\beta_{0t} + \boldsymbol{x}^T \boldsymbol{\beta}_1)}.$$

We have displayed the results of fitting the proposed vertical model for handling missing failure causes in Table 5.4. In Table 5.5 and Table 5.6 we have re-produced the results of fitting the missing failure causes *discrete time mixure model* that we advanced in Chapter 4 as well as the multinomial model for handling missing failure causes. The idea is to compare the proposed model to one of the mixture models and a multinomial model for missing failure causes. As in previous chapters, we test the effect of raising the disregard rate on re-employment prospects via its effect on the CIF estimates. We continue to assess the effect of increasing the disregard rate by 50%.

Table 5.4: Maximum likelihood estimates for the Complete Case Analysis and Missing failure Causes Discrete time Vertical Model (with standard errors) (* denotes $P< 0.05$).

Coefficient	$\hat{\boldsymbol{\phi}}$	$\hat{\boldsymbol{\beta}}_C$	$\hat{\boldsymbol{\beta}}_M$
T1	1.194(0.176)*	-2.399(0.067)*	-2.331(0.060)*
T2	1.214(0.187)*	-2.707(0.078)*	-2.450(0.073)*
T3	1.251(0.208)*	-2.902(0.073)*	-2.530(0.073)*
T4	0.956(0.268)*	-3.434(0.090)*	-3.227(0.104)*
T5	1.190(0.208)*	-2.627(0.095)*	-2.371(0.081)*
T6	1.264(0.374)*	-3.667(0.161)*	-3.425(0.136)*
T7	1.311(0.241)*	-2.487(0.106)*	-2.244(0.091)*
T8	0.354(0.432)	-3.846(0.210)*	-3.549(0.175)*
T9	1.527(0.423)*	-3.208(0.166)*	-2.904(0.139)*
T10	0.049(0.838)	-5.001(0.412)*	-4.224(0.272)*
T11	1.667(0.496)*	-3.211(0.188)*	-3.058(0.168)*
T12	0.847(0.698)	-4.217(0.322)*	-3.824(0.256)*
T13	1.130(0.416)*	-2.868(0.185)*	-2.626(0.160)*
T14	1.682(0.460)*	-2.548(0.180)*	-2.394(0.162)*
T15	1.925(0.629)*	-2.804(0.227)*	-2.815(0.217)*
T16	1.084(0.678)	-3.149(0.291)*	-2.643(0.224)*
T17	2.205(1.072)	-3.298(0.346)*	-3.000(0.291)*
T18	1.279(0.820)	-3.115(0.349)*	-3.001(0.316)*
T19	0.992(0.413)	-1.483(0.212)	-1.124(0.187)*
ui	-0.062(0.150)	1.191(0.060)*	1.138(0.052)*
dr	-1.292(1.010)	-1.379(0.457)*	-0.665(0.375)
rr	1.116(0.968)*	0.478(0.406)	0.377(0.339)
age	-0.012(0.007)*	-0.012(0.003)*	-0.012(0.003)*
wage	0.991(0.192)*	0.293(0.085)*	0.251(0.071)*
tenure	0.002(0.013)*	0.002(0.006)	-0.005(0.005)

Table 5.5: Maximum likelihood estimates for the Discrete Time Missing Failure Causes Mixture Model and the Missing Failure Causes Multinomial Model (with standard errors) (* denotes $P< 0.05$).

	Mixture Model (Latency Model)		Multinomial Model	
Coefficient	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
T1	-2.524(0.065)*	-2.163(0.109)*	-2.622(0.068)*	-3.754(0.115)*
T2	-2.589(0.071)*	-2.367(0.123)*	-2.754(0.075)*	-3.779(0.124)*
T3	-2.599(0.078)*	-2.586(0.144)*	-2.825(0.083)*	-3.903(0.140)*
T4	-3.287(0.117)*	-3.121(0.197)*	-3.582(0.123)*	-4.472(0.190)*
T5	-2.444(0.087)*	-2.404(0.153)*	-2.677(0.093)*	-3.776(0.153)*
T6	-3.407(0.151)*	-3.464(0.274)*	-3.710(0.157)*	-3.866(0.187)*
T7	-2.239(0.096)*	-2.479(0.184)*	-2.494(0.102)*	-3.866(0.187)*
T8	-3.641(0.211)*	-3.243(0.291)*	-4.053(0.225)*	-4.757(0.282)*
T9	-2.772(0.147)*	-3.368(0.324)*	-3.117(0.154)*	-4.717(0.317)*
T10	-4.378(0.348)*	-3.769(0.422)*	-5.103(0.428)*	-5.129(0.364)*
T11	-2.874(0.176)*	-3.664(0.418)*	-3.220(0.181)*	-4.996(0.430)*
T12	-3.802(0.301)*	-3.723(0.460)*	-4.283(0.324)*	-5.061(0.439)*
T13	-2.542(0.173)*	-2.877(0.318)*	-2.930(0.184)*	-4.112(0.305)*
T14	-2.218(0.166)*	-3.060(0.379)*	-2.583(0.177)*	-4.202(0.376)*
T15	-2.535(0.223)*	-3.696(0.567)*	-2.971(0.233)*	-4.680(0.557)*
T16	-2.400(0.233)*	-3.220(0.490)*	-2.914(0.254)*	-4.055(0.465)*
T17	-2.596(0.292)*	-4.280(0.906)*	-3.246(0.355)*	-4.397(0.671)*
T18	-2.654(0.338)*	-3.573(0.675)*	-2.143(0.361)*	-3.665(0.643)*
T19	-0.841(0.177)*	-1.845(0.314)*	-1.438(0.210)*	-2.615(0.318)*
ui	1.505(0.065)*	0.556(0.116)*	1.136(0.059)*	-1.136(0.101)*
dr	-0.447(0.504)	-1.729(0.835)*	-1.076(0.443)*	-0.122(0.668)
rr	0.480(0.456)	-0.025(0.784)	0.840(0.395)	-0.324(0.594)
age	-0.011(0.003)*	-0.008(0.006)	-0.015(0.003)*	-0.001(0.005)
wage	0.227(0.096)*	0.039(0.150)	0.522(0.084)*	-0.495(0.122)*
tenure	-0.014(0.006)	0.044(0.011)*	-0.005(0.006)	-0.003(0.010)

Table 5.6: Maximum likelihood estimates for the Discrete Time Missing Failure Causes Mixture Model (with standard errors) (* denotes $P< 0.05$).

Mixture Model (Incidence Model)	
Coefficient	$\hat{\phi}_1$
Constant	1.178(0.062)*
`ui`	-0.488(0.092)*
`dr`	-2.381(0.674)*
`rr`	0.822(0.668)*
`age`	-0.009(0.005)*
`wage`	0.897(0.131)*
`tenure`	0.038(0.009)*

With the previous application of the *discrete time mixture model* for handling missing failure causes we found that 76.5% of unemployed individuals from the reference were re-employed and if the disregard rate for an individual from this group is raised by 50% then the probability of re-employment is 75.7%.

The multinomial model for handling missing failure causes suggested a 73.5% re-employment probability for the reference group and a probability of 72.1% if we raised the disregard rate by 50%. So, both models suggested that raising the disregard rate has insignificant impact on re-employment. The proposed model suggests 73.8% as the probability of re-employment for an individual from the reference group and a probability of 72.6% if the disregard rate is raised by 50%. Clearly, all three models suggest that raising the disregard rate has no effect on re-employment prospects for benefit recipients. The fact these conclusions run against McCall (1996) is not as important as the fact that all three models agree with the multinomial model as fitted within the MI framework where the multinomial model is an established model for regression analysis of discrete time competing risks data and the MI is also an established framework for dealing with missing data. These results also lend credence to the suggestion that the multinomial model and the vertical model tend to produce similar results while the mixture models also tend to do the same in terms of CIFs estimates. In Figure 5.9 we have plotted the CIF estimates with the effect of increasing the disregard rate by 50% via all three models.

.

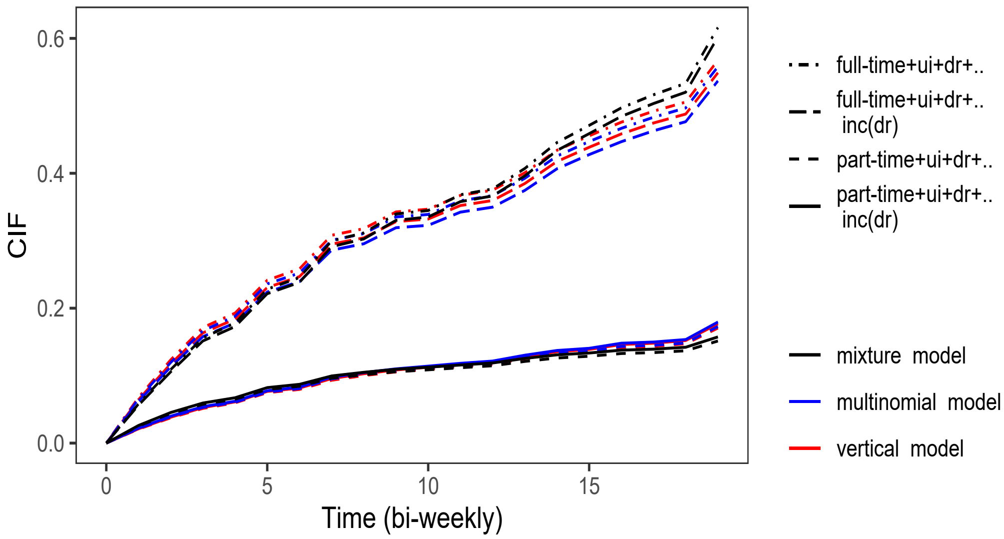

Figure 5.9: The CIF of exit to full-time and part-time employment for the `ui` recipients with the effect of increasing `dr` via the Discrete Time Vertical Model, the Multinomial Model and the Discrete Time Mixture Model.

We have also assessed the effect of unemployment benefits on re-employment prospects via their effect on CIF with all three models and plotted the results in Figure 5.10.

All three models lead to the conclusion that unemployment benefits tend to dampen the re-employment prospects as espoused by the theory on this subject. Figure 5.10 provides further evidence that the mixture models tend to compare more favorably with one another than with either the multinomial model or the vertical model, likewise, the multinomial model and the vertical model also seem to be more comparable with one another than with the mixture models. These relationships seem to hold in the presence

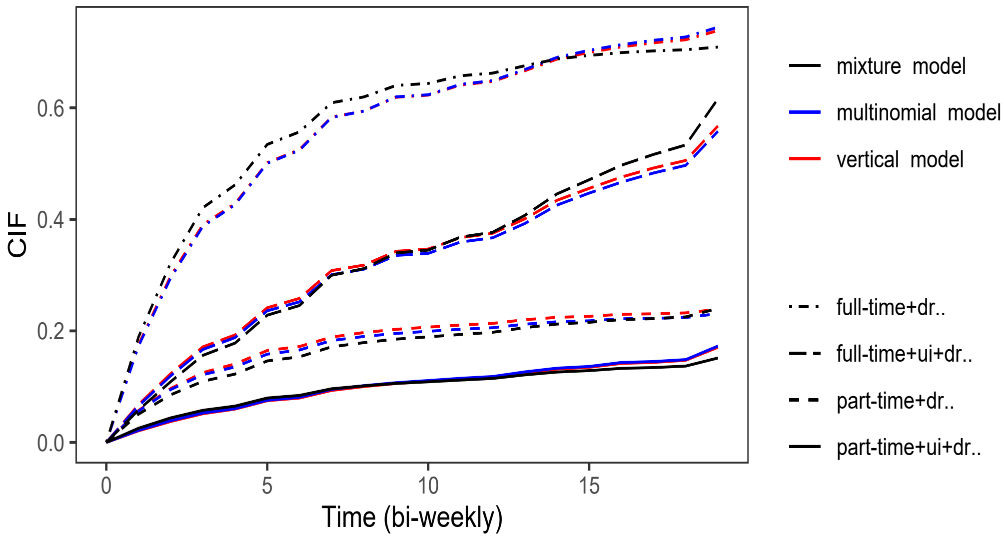

Figure 5.10: The CIF of exit to full time and part time employment with the effect of `ui` via the Discrete Time Vertical Model, the Multinomial Model and the Discrete Time Mixture Model.

as well as in the absence of missing failure causes.

We now turn our gaze on the estimates for regression coefficients. It might be instructive to compare the interpretation of the regression coefficients excluding and including the subjects with missing failure causes. Accordingly, in Table 5.4 we have re-produced the results of fitting a *discrete time vertical regression model* from the previous section.

Since the relative hazards are invariant to the presence or absence of missing failure causes, we have reprinted $\hat{\boldsymbol{\phi}}$ from the previous section in Table 5.4. Obviously, there is no need to interpret these coefficients. Recall from the prious section that we did demonstrate how the relative hazards can be recovered from the multinomial model using the following relationship;

$$\log \frac{h_j(t|\boldsymbol{x}, \boldsymbol{\alpha})}{h_k(t|\boldsymbol{x}, \boldsymbol{\alpha})} = (\alpha_{0jt} - \alpha_{0kt}) + \boldsymbol{x}^T(\boldsymbol{\alpha}_{1j} - \boldsymbol{\alpha}_{1k})$$

if we wished to compare cause j to cause k. We did establish that the relative hazards that were estimated from the multinomial model compared favourably with the estimates from the vertical model. It might also be instructive to determine if this holds true in the presence of missing failure causes, that is, if the relative hazard estimates that are now derived from the missing failure causes multinomial model are similar to relative hazard estimates that we obtained in the previous section. Evidently from Figure 5.11, the

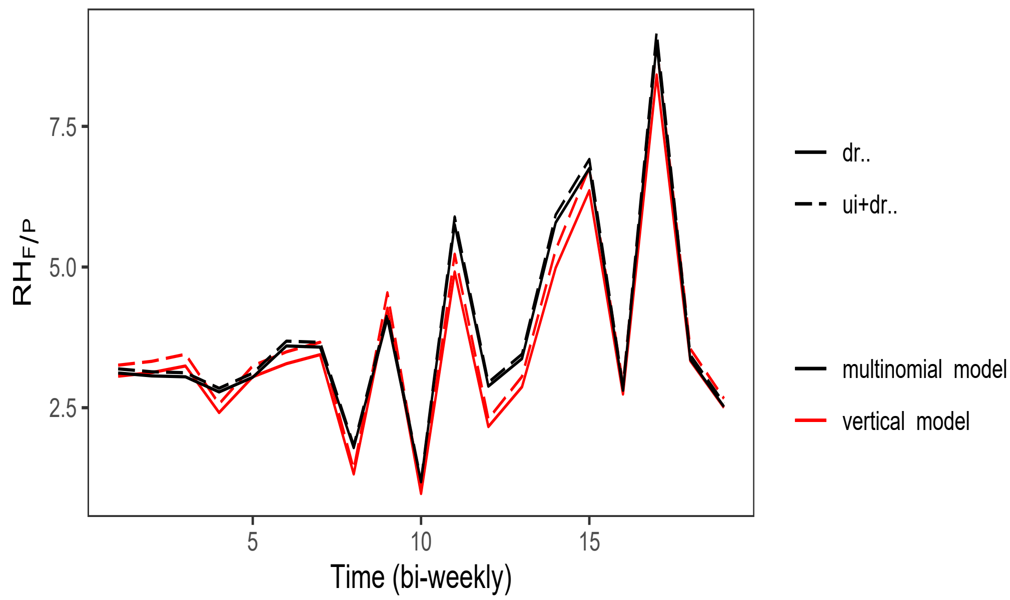

Figure 5.11: The Relative Hazards of exit to full-time employment with the effect `ui` via the Discrete Time Vertical Model and the Missing Failure Causes Multinomial Model.

relative hazard estimates from the missing failure causes multinomial model do compare favourably with estimates for the same quantities from the vetical model. When missing failure causes are recreated within the MI framework the model that is used to generate these failure causes is the model for relative hazards within the vertical model. Suppose that we generated a large number of complete samples and for each complete sample we estimated the relative hazards from ($\boldsymbol{\phi}^*$) and ultimately pooled these estimates;

$$\bar{\boldsymbol{\phi}} = \frac{\sum_{m=1}^{M} \boldsymbol{\phi}_m^*}{M}$$

we would expect that the pooled estimate of relative hazards from $\bar{\boldsymbol{\phi}}$ to approach the relative hazards that we obtained from $\hat{\boldsymbol{\phi}}$ in Chapter 5 (Nicolaie et al., 2015). In the long run, the relative hazard estimates that we obtain from the larger sample that includes subjects that had missing failure causes, but for whom failure causes have since been re-created should be the same as the relative hazards that are obtained from subjects that originally came with known failure causes. The advantage of retrieving the relative hazards from the vertical model is that we do not need to go through all the steps that are associated with the MI procedure to obtain the estimates for these quantities.

There are two estimates for $\boldsymbol{\beta}$, namely; $\hat{\boldsymbol{\beta}}_C$ and $\hat{\boldsymbol{\beta}}_M$, where $\hat{\boldsymbol{\beta}}_C$ is the estimate for $\boldsymbol{\beta}$ excluding the subjects with missing failure causes and $\hat{\boldsymbol{\beta}}_M$ is the estimate for $\boldsymbol{\beta}$ including these subjects. Now, $\hat{\beta}_{0tM} > \hat{\beta}_{0tC}$ $\forall t$, except $t = 15$. Thus, when subjects that have missing failure causes are included in the analysis the prospects of re-employment adjusts upwards except for period $t = 15$. During the fifth month ($t = 10$) the prospects of re-employment are at their lowest and highest at $t = 19$ in the absence or presence of missig failure causes. Since $\hat{\beta}_{1\text{ui}C} = 1.191$, the odds of re-employment for individuals that receive benefits are lower than the odds of re-employment for individuals that do not receive benefits by about $(1 - / \exp(1.191))$ 70% when missing failure causes are excluded and about $(1 - 1/\exp(1.138))$ 68% when these subjects are included. Clearly, provision of unemployment benefits does not improve the prospects of re-employment.

This concludes the exercise of advancing the discrete time vertical model for handling missing failure causes. We need not develop the standard errors for the CIF estimates. The expressions that were proposed in the previous section equally apply here because the models that are proposed in this chapter are the same models that were advanced in the previous section. All we did in this chapter was to demonstrate that the same models from the previous section also apply when data comes with missing failure causes.

5.2.2 Discussion

In this chapter we focussed on the vertical model with the view to advance a discrete version of this model for handling ordinary competing risks data as well as competing risks data that comes with missing failure causes. The vertical model proposes yet another decomposition of the bivariate distribution of the joint distribution of failure time and failure type which leads to characterization of data in terms of total hazards and relative hazards. We introduced this exercise by proposing a discrete time vertical regression model. We then compared the proposed regression model to the multinomial model and the mixture models by testing one proposal according to McCall (1996) that adjusting the disregard rate upwards for benefit recipients results in improved employment rates. Furtheremore, the effect of unemployment benefits on employment was also tested. It was found that all models led to the same conclusions regarding McCall (1996)'s proposal and the effect of employment benefits on re-employment prospects. It was also found that the proposed model and the multinomial model compared more favourably with each other than with the mixture models, that is, the *continuous time mixture model* and the *discrete time mixture model* in terms of CIF estimates. We then considered the regression coefficient estimates for relative hazards.

It could be argued that the popularity of competing risks with applied researchers centers on the role of CSHs and their properties. A plot of CSH estimates against time provides an indication as to when are the events most/least likely to occur. These quantities also allow for the identification of factors that explain the risks of these events. In graduation/dropout studies, for example, this analysis can be used to identify the factors that explain both graduations and dropouts so that this information can be incorporated into the development of support measures to improve both graduation and retention rates. Equally important is timing, that is, after the support measures have been designed, then when should they be implemented to maximize their impact? Additional to the use of CSHs to provide answers to that question, we illustrated how the information provided by relative hazards can be included to supplement this exercise of developing the intervention strategies. Furthermore, we discussed the interpretation of covariate effects on the total hazards from the regression coefficients. We also discussed the circumstances that allow for interpretation of covariate effects on CSHs from the regression coefficients for total hazards and relative hazards.

We then proceeded to advance a discrete time version of the vertical regression model for handling missing failure causes. When the authors of the model, that is, Nicolaie et al. (2010), proposed the vertical model for regression analysis of ordinary competing risks data in continuous time, the same authors later demonstrated that the same model can also handle missing failure causes as is (Nicolaie et al., 2015). In advancing the discrete time vertical model for handling missing failure causes in this chapter, we followed in the footsteps of these authors and showed that that ordinary discrete time vertical regression model can also deal with missing failure causes. To deal with missing failure causes the ordinary vertical model merely adjusts the total hazards accordingly while the relative hazards remain un-affected by presence or absence of these subjects. We then compared the proposed vertical regression model to the mixture regression model for handling missing failure causes which was also advanced in Chapter 4 as well as the multinomial model for handling missing failure causes by applying all these models to test McCall (1996)'s proposal. We found that all these models led to the conclusion that raising the disregard rate for unemployment benefit recipients does not improve their re-employment prospects which was, however, in contradiction to what McCall (1996) claimed. These results were adequate for the purposes of this thesis as they confirm the proposed vertical model for handling missing failure causes, the *discrete time mixture model* for handling missing failure causes, and by implication the *continuous time mixture model* for handling missing failure causes as valid models because they compare favourably with the MI method, an established method for dealing with missing data. When we applied all these models to test a well documented fact in economics that provision of unemployment benefits increases unemployment, we also found that all these models led us to the same conclusion that indeed unemployment benefits do act as a disincetive.

This chapter concludes part one and two of this thesis where we were concerned with advancing discrete time models for dealing with ordinary discrete time cometing risks data and that comes with missing failure causes. In the next chapter, Chapter 6, we attend to data that comes with a substantial proportion cured subjects.

CHAPTER 6

The Nonparametric Vertical Mixture Cure Model

6.1 Introduction

In this chapter we advance a nonparametric vertical model for handling cured subjects in discrete time. Recall that in Chapter 5 we advanced a nonparametric version of the model for analysis of ordinary discrete time competing risks data. In Chapter 8 we demonstrated that the nonparametric model that was advanced in Chapter 5 can also serve as a missing failure causes model, that is, the same model that is intended for nonparametric analysis of ordinary discrete time competing risks data can also handle missing failure causes as is. In this chapter we upgrade the model that was advanced in Chapter 5 into a model that can handle cured subjects. by cured subjects we refer to those subjects that may not fail.

The earliest attempts to analyze data that comes with a sizable proportion of cured subjects can be traced far back to (Boag, 1949; Berkson and Gage, 1952). This topic lay dormant until the contributions from Farewell (1982) and the seminal work of Kuk and Chen (1992), the authors that are often credited with laying the foundations for the modern mixture cure model as well as re-kindling interest in the topic. As acknowledged by Kuk and Chen (1992), the mixture cure model borrows heavily from the mixture competing risks model (Larson and Dinse, 1985) in terms of theoretical foundations. The mixture cure model assumes a single mode of failure, whereas, cured subjects can also arise in experiments where subjects are exposed to multiple risks of failure. Escarela et al. (2000), for example, undertook a study where the variable of interest was the time to re-conviction (for an offence that is similar or dissimilar to the previous offence) for a group of ex-convicts where the types of offence are regarded as competing risks. It is not unreasonable to expect that some of the ex-convicts may not experience re-conviction at all or may desist from committing criminal offences for an over extended period. Naturally, this latter group of convicts may be regarded as cured subjects. The discusssion of cured subjects in the context of competing risks has not been as extensive as the coverage of this topic in single mode of failure settings, see for example Peng and Taylor (2014) for review of what has come to be known as the mixture cure model. Authors such as (Choi, 2002; Maller and Zhou, 2002; Zhiping, 2011) have advanced various versions of a mixture competing risks model that can handle competing risk data that comes with cured subjects. We refer to this model as the mixture cure competing risks model. This model is also a variant of the ordinary mixture competing risks model. Recall that, with J failure causes, the ordinary mixture competing risks model proposes failure type probabilities π_j $(j = 1, 2 \ldots, J)$, and component hazards $\lambda_j(t)$ $(j = 1, 2 \ldots, J)$, for modelling competing risks data. One of the characteristic features of the ordinary mixture competing risks model is a mixed expression for the survival function;

$$S(t) = \sum_{j=1}^{J} \pi_j S_j(t). \tag{6.1.1}$$

This expression for the survival function assumes that all subjects are destined to fail by one of the J failure causes. This includes the censored subjects whose failure causes are not observed at the close of the observation period. The mixture cure competing risks model extends this argument and posits that amongst the censored subjects, there is a subgroup of subjects that may not fail by any of the J failure causes i.e., a $(J + 1)^{\text{th}}$ subpopulation of cured subjects. To characterize the presence of cured subjects in data the mixture cure competing risks model extends (6.1.1) as;

$$S(t) = \sum_{j=1}^{J} \pi_j S_j(t) + (1 - \sum_{j=1}^{J} \pi_j).$$

The mixture cure vertical model (Nicolaie et al., 2018), the subject of this chapter, is the latest regression model for handling cured subjects. In this chapter we propose a nonparametric version of this model for application to discrete time competing risks data that comes with a sizable proportion of cured subjects. An example of discrete time data that comes with cured subjects is found in loan portfolio settings where some debtors see out the term of the loan, and others repay the loan early or default (Dirick et al., 2015). This data can be regarded as a discrete time competing risks data with a possibility of cured subjects because a debtor will be observed to have defaulted if the debtor fails to honour repayments which are monthly or weekly (which will most likely give rise to a larger proportion of tied event times), and the fact that some of the debtors may default, repay the loan over the full term or earlier. Defaults or settlement of debt before the loan runs its full term are regarded as competing events while debtors who repay the loan over the original term can be viewed as cured subjects because they have survived failure due to default or early repayment.

The mixture cure vertical model upscales the ordinary vertical model (Nicolaie et al., 2010) into a model that can handle cured subjects. Recall that the ordinary vertical model proposes total hazards and relative hazards for modelling data and that characterization of data with these quantities follows from the factorization assumptions which posits that the bivariate distribution of $(\tilde{T}, D)$ can be decomposed into a marginal distribution for failure time $\tilde{T}$ and a distribution for failure type D conditional on failure time.

$$P(\tilde{T}, D) = P(D|\tilde{T})P(\tilde{T}). \tag{6.1.2}$$

Recall that observed data can be represented by $\boldsymbol{y} = \big((t_1; \Delta_1), \ldots, (t_n; \Delta_n)\big)$, where $T_i = \min\{\tilde{T}_i C_i\}$, and $\Delta_i = I(\tilde{T}_i < C_i)$. It is assumed that $T, C \in \{1, 2, \ldots, q\}$, where q is a positive integer. The definition of total hazards in discrete time is given by

$$h(t) = P(T = t|T \geq t).$$

This quantity continues to express the probability that an event of any type occurs at time t given survival up to time t. The definition of relative hazards $\Pi_j(t)$, the probability that a failure is due to failure type j given that a failure has occured at time t is given by

$$\Pi_j(t) = P(D = j|T = t).$$

When data is modeled with total hazards and relative hazards, the standard functionals are connected to data via;

$$h_j(t) = h(t)\Pi_j(t). \tag{6.1.3}$$

Recall that since data is no longer modeled with CSHs, (6.1.3) provides the means for recovering the CSH estimates from total hazard and relative hazard estimates. The estimates for CIFs are derived from;

$$F_j(t) = \sum_{s=1}^{t} S(s\text{-}1)h(s)\Pi_j(s).$$

The representation of data with or without cured subjects is identical because cured subjects are unobservable. A plot of the KM estimate of the survival function that levels off towards the close of the observation period is often taken as an indication that cured subjects may exist in data. When data comes with a sizable proportion of cured subjects the vertical model regards data as split between cured and uncured subjects, and a partially observed indicator variable Y is assumed to differentiate cured subjects from uncured subjects where Y assumes 1 or 0 according to whether a subject is uncured or cured. Note that $P(Y = 1) = p$ is the proportion of uncured subjects. The factorization of the bivariate distribution of $(\tilde{T}, D)$ which underlies the vertical model now only obtains for uncured subjects, that is;

$$P(T, D|Y = 1) = P(D|T, Y = 1)P(T|Y = 1)$$

because cured subjects are assumed not experience failure. Data is now modeled with conditional total hazards $h_u(t)$ $(t = 1, 2, \ldots, q)$, conditional relative hazards $\Pi_{uj}(t)$ $(j = 1, 2, \ldots, J; t = 1, 2, \ldots, q)$, and p where the conditional total hazards and conditional relative hazards summarize the failure time distribution for uncured subjects and $(1 - p)$ characterizes cured subjects. The definition of conditional total hazards is given by

$$h_u(t) = P(T = t|T \geq t, Y = 1).$$

This quantity conveys the same meaning, but for uncured subjects, that is, the probability of a failure at time t by any failure cause given survival to time t for uncured subjects. The definition for conditional relative hazards is given by

$$\Pi_{uj}(t) = P(D = j|T = t, Y = 1).$$

Likewise, $\Pi_{uj}(t)$ is the probability that if a failure has occured for an uncured subject at time t that failure is due to failure type j. The equation that connects all the standard functionals to data for uncured subjects is now given by

$$h_{uj}(t) = h_u(t)\Pi_{uj}(t).$$

Critically, the expression for the conditional CIF is given by

$$\begin{aligned} F_{uj}(t) &= P(T \leq t, D = j | Y = 1) \\ &= \sum_{s=1}^{t} S_u(s\text{-}1) h_u(s) \Pi_{uj}(s). \end{aligned}$$

Naturally, this expression applies to uncured subjects only. There is a special relationship between the CIF for uncured subjects and the CIF for entire sample;

$$\begin{aligned} F_j(t) &= P(T \leq t, D = j) \\ &= P(T \leq t, D = j | D = j, Y = 1) P(Y = 1) \\ &+ P(T \leq t, D = j | D = j, Y = 0) P(Y = 0) \\ &= P(T \leq t, D = j | D = j, Y = 1) P(Y = 1) + 0. \end{aligned}$$

This follows because the CIF for cured subjects is zero. Thus,

$$\begin{aligned} F_j(t) &= P(T \leq t, D = j | D = j, Y = 1) P(Y = 1) \\ &= pF_{uj}(t). \end{aligned} \tag{6.1.4}$$

Note that $S_u(t)$ is the survival function for uncured subjects, that is, $S_u(t)$ is the probability that a subject survives failure at time t given that the subject is uncured. The probability that any subject, cured or uncured, survives failure at time t or the population survival fuction is given by

$$\begin{aligned} S(t) &= P(T > t) \\ &= P(T > t | Y = 1) P(Y = 1) + P(T > t | Y = 0) P(Y = 0) \\ &= pS_u(t) + (1 - p) \end{aligned}$$

because $P(T > t | Y = 0) = 1$ as cured subjects are assumed not experience failure. This completes the presentation of the mixture vertical cure model as a nonparametric model for application to discrete time competing risks data that comes with a sizable proportion of cured subjects. This was achieved by specifying discrete time nonparametric total hazards for uncured subjects. In the next section we attend to the estimation of the model parameters. More formally, let $\boldsymbol{\theta} = (\boldsymbol{\pi}_u^T, \boldsymbol{h}_u^T, p)^T$, where $\boldsymbol{h}_u = (h_u(1), \ldots, h_u(q))^T$, $\boldsymbol{\pi}_u = (\boldsymbol{\pi}_{u1}^T, \boldsymbol{\pi}_{u2}^T, \boldsymbol{\pi}_{u(J\text{-}1)}^T)^T$, and $\boldsymbol{\pi}_{uj} = (\pi_{uj}(1), \pi_{uj}(1) \ldots, \pi_{uj}(q))^T$. In Section 9.2 we present an EM Algorithm for estimating $\boldsymbol{\theta}$. An application of the proposed model is presented in Section 9.3. The critical point to note when data comes with cured subjects is that if these subjects are not accounted for, the resulting estimates will be biased downwards. The extent of bias will vary according to the proportion of cured subjects, the larger this proportion is the larger is is the extent of bias. To demonstrate this point, we have applied the proposed model to two data sets with different sizes of cured subjects. We conclude this chapter in Section 9.4. The standard errors for the conditional CIF estimates are derived in the Appendix I.

6.2 The EM Algorithm

The MLE for $\boldsymbol{\theta}$ is obtained by differentiating the observed likelihood function w.r.t. $\boldsymbol{\theta}$. The observed data log-likelihood function in terms of conditional total hazards, conditional relative hazards and the proportion of uncured subjects can be written as;

$$
\begin{aligned}
\mathcal{L}_0(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log P(Y_i = 1)P(D_i = j|T_i = t_i, Y_i = 1)P(T_i = i|Y_i = 1) \\
&+ (1 - d_i) \log P(T_i > t_i) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij} \log \Pi_{uj}(t_i) + \sum_{i=1}^{n} d_i \log p_i P(T_i = t_i|Y_i = 1) \\
&+ (1 - d_i) \log(p_i S_u(t_i) + (1 - p_i)) \\
&= \mathcal{L}_0(\boldsymbol{\theta}_1) + \mathcal{L}_0(\boldsymbol{\theta}_2)
\end{aligned}
$$

where $d_{ij} = I(D_i = j)$, $d_i = \sum_{j=1}^{J} d_{ij}$, $\boldsymbol{\theta}_1 = \boldsymbol{\Pi}$ and $\boldsymbol{\theta}_2 = (p, \boldsymbol{h}^T)^T$. Note that;

$$
\begin{aligned}
P(T = t) &= P(T = t, Y = 1) + P(T = t, Y = 0) \\
&= P(T = t, Y = 1)
\end{aligned}
$$

since $P(T = t, Y = 0) = 0$ because cured subjects are assumed not experience failure. This implies that;

$$\begin{aligned}\Pi_{uj}(t) &= P(D = j|T = t, Y = 1)\\ &= \frac{P(D = j, T = t, Y = 1)}{P(T = t, Y = 1)} = \frac{P(D = j, T = t)}{P(T = t)} = \Pi_j(t).\end{aligned}$$

This is an important result because it means that the relative hazards are again invariant to the presence or absence of cured subjects just as they are regarding the presence or absence of missing failure causes. The conditional relative hazards can, therefore, be estimated from;

$$\hat{\Pi}_j(t) = \frac{\hat{h}_j(t)}{\hat{h}(t)} = \frac{d_{(jt)}}{d_{(t)}} = \hat{\Pi}_{uj}(t)$$

where $d_{(jt)}$ and $d_{(t)}$ denote the total number of failures due to failure cause j and the total number of failures by any failure cause at time t, respectively. We can, therefore, ignore $\mathcal{L}_0(\boldsymbol{\theta_1})$, and concentrate on $\mathcal{L}_0(\boldsymbol{\theta_2})$. To simplify the estimation of $\boldsymbol{\theta}_2$, the log-likelihood function $\mathcal{L}_0(\boldsymbol{\theta}_2)$ is regarded as "incomplete" data log-likelihood function in respect of unobserved Y vis-a-vis censored subjects so as to justify the implementation of an EM Algorithm. Suppose that Y were actually observed, then a subject i that is an observed failure contributes $p_i P(T_i = t_i|Y_i = 1)$, a censored one that eventually fails contributes $p_i P(T_i > t_i|Y_i = 1)$, while, a cured subject contributes $(1 - p_i)$ towards the complete data likelihood function. The complete data log-likelihood function can be written as;

$$\begin{aligned}\mathcal{L}_c(\boldsymbol{\theta}_2) = \sum_{i=1}^{n} d_i y_i \log p_i P(T_i = t_i|Y_i = 1) + (1 - d_i) y_i \log p_i P(T_i > t_i|Y_i = 1)\\ + (1 - y_i)(1 - d_i)\log(1 - p_i). \qquad (6.2.1)\end{aligned}$$

Note that $P(T = t|Y = 1) = \dfrac{h_u(t)}{1 - h_u(t)} \displaystyle\prod_{s=1}^{t}(1 - h_u(s))$. With appropriate substitutions and re-arrangement of few terms, $\mathcal{L}_c(\boldsymbol{\theta}_2)$, can be written as;

$$\begin{aligned}\mathcal{L}_c(\boldsymbol{\theta}_2) = \sum_{i=1}^{n} d_i y_i \log p_i + d_i y_i \log \frac{h_u(t_i)}{1 - h_u(t_i)} + \sum_{s=1}^{t_i} d_i y_i \log(1 - h_u(s))\\ + (1 - d_i) y_i \log p_i + \sum_{s=1}^{t_i}(1 - d_i) y_i \log(1 - h_u(s)) + (1 - y_i)(1 - d_i)\log(1 - p_i).\end{aligned}$$

Since $y_i = 1$ when $d_i = 1$, then $d_i y_i = 1$. When $d_i = 0$ then y_i is 0 or 1 and $d_i y_i = 0$. We can, therefore, replace $d_i y_i$ with d_i. Note also that $(1 - y_i)(1 - d_i) = 1 - (d_i + (1 - d_i)y_i)$. If we let $g_i = d_i + (1 - d_i)y_i$, we can write $\mathcal{L}_c(\boldsymbol{\theta}_2)$ as;

$$\begin{aligned}\mathcal{L}_c(\boldsymbol{\theta}_1) &= \sum_{i=1}^{n}(d_i + (1 - d_i)y_i)\log p_i + \sum_{s=1}^{t_i}(d_i + (1 - d_i)y_i)\log(1 - h_u(s)) \\ &+ (1 - (d_i + (1 - d_i)y_i)\log(1 - p_i) \\ &= \sum_{i=1}^{n} g_i \log p_i + (1 - g_i)\log(1 - p_i) + d_i \log\frac{h_u(t)}{1 - h_u(t)} + \sum_{s=1}^{t_i} g_i \log(1 - h_u(s)) \\ &= \mathcal{L}_c(p) + \mathcal{L}_c(\boldsymbol{h})\end{aligned}$$

where,

$$\mathcal{L}_c(p) = \sum_{i=1}^{n} g_i \log p_i + (1 - g_i)\log(1 - p_i)$$

and,

$$\mathcal{L}_c(\boldsymbol{h}) = \sum_{i=1}^{n} d_i \log\frac{h_u(t_i)}{1 - h_u(t_i)} + \sum_{s=1}^{t_i} g_i \log(1 - h_u(s)).$$

It can easily be seen that $\mathcal{L}_c(p)$ is a kernel of a binomial log-likelihood function. At this point, the now familiar time dependent indicator variable d_{is} is introduced, such that $d_{is} = 0$ for $s = 1, 2, \ldots, t_i$-1, and $d_{it_i} = d_i$. Now, $\mathcal{L}_c(\boldsymbol{h})$, can be written as;

$$\mathcal{L}_c(\boldsymbol{h}) = \sum_{i=1}^{n}\sum_{s=1}^{t_i} d_{is}\log h_u(s) + (g_{is} - d_{is})\log(1 - h_u(s))$$

where $g_{is} = g_i$ for $s = 1, 2 \ldots, t_i$. Since $\mathcal{L}_c(p)$ and $\mathcal{L}_c(\boldsymbol{h})$ are both linear in y_i, in the $(r{+}1)^{\text{th}}$ iteration of the EM algorithm the E-Step entails replacing y_i with its expectation conditional on $\boldsymbol{\theta}_2^{(r)}$, the estimate of $\boldsymbol{\theta}_2$ in the M-Step of the r^{th} iteration, and $\boldsymbol{y}$. The conditional expectation of y_i is given by

$$y_i^{(r)} = E[y_i|\boldsymbol{y}, \boldsymbol{\theta}_2^{(r)}] = \frac{p^{(r)}S_u^{(r)}(t_i)}{p^{(r)}S_u^{(r)}(t_i) + (1 - p^{(r)})} \tag{6.2.2}$$

as p is not subject specific. The E-Step can be written as a sum of

$$Q(p|p^{(r)}) = \sum_{i=1}^{n} g_i^{(r)}\log p + (1 - g_i^{(r)})\log(1 - p)$$

and,

$$Q(\boldsymbol{h}|\boldsymbol{h}^{(r)}) = \sum_{i=1}^{n}\sum_{s=1}^{t_i} d_{is}\log h_u(s) + (g_{is}^{(r)} - d_{is})\log(1 - h_u(s)).$$

Naturally, the MLE for p is given by

$$p^{(r+1)} = \frac{\sum_{i=1}^{n} g_i^{(r)}}{n}. \tag{6.2.3}$$

Since $y_i^{(r)}$ is a function of time only, we can re-write it as;

$$y^{(r)}(t) = \frac{p^{(r)} S_u^{(r)}(t)}{p^{(r)} S_u^{(r)}(t) + (1 - p^{(r)})}. \tag{6.2.4}$$

Suppose that $c_{(s)}$ is the number of censored subjects at time s. Define $R_{(s)}^{(r)} = \sum_{i=1}^{n} g_{is}^{(r)} = \sum_{l=s}^{q} d_{(l)} + c_{(s)} y^{(r)}(s)$, as the size of the risk set at time s in the E-Step of the $(r+1)^{\text{th}}$ iteration. We can alternatively re-write $Q(\boldsymbol{h}|\boldsymbol{h}^{(r)})$ as;

$$\begin{aligned}
Q(\boldsymbol{h}|\boldsymbol{h}^{(r)}) &= \sum_{l=1}^{q} \sum_{i=1}^{n} \sum_{s=l}^{t_i} d_{is} \log h_u(s) + (g_{is}^{(r)} - d_{is}) \log(1 - h_u(s)) \\
&= \sum_{l=1}^{q} \sum_{i=1}^{n} d_{il} \log h_u(l) + (g_{il}^{(r)} - d_{il}) \log(1 - h_u(l)) \\
&= \sum_{l=1}^{q} \log h_u(l) \sum_{i=1}^{n} d_{il} + \sum_{s=1}^{q} \log(1 - h_u(l)) \sum_{i=1}^{n} (g_{il}^{(r)} - d_{il}) \\
&= \sum_{l=1}^{q} d_{(l)} \log h_u(l) + (R_{(l)}^{(r)} - d_{(l)}) \log(1 - h_u(l))
\end{aligned}$$

where $d_{(l)} = \sum_{i=1}^{n} d_{il}$. For consistency, we swap s for l, that is, we write $Q(\boldsymbol{h}|\boldsymbol{h}^{(r)})$ as;

$$Q(\boldsymbol{h}|\boldsymbol{h}^{(r)}) = \sum_{s=1}^{q} d_{(s)} \log h_u(s) + (R_{(s)}^{(r)} - d_{(s)}) \log(1 - h_u(s)).$$

The MLE for $h_u(s)$ is, therefore, given by

$$h_u^{(r+1)}(s) = \frac{d_{(s)}}{R_{(s)}^{(r)}}. \tag{6.2.5}$$

The EM Algorithm can be summarized as follows;

1) In the $(r+1)^{\text{th}}$ iteration of the EM Algorithm, adjust the risk sets $R_{(s)}^{(r)}$ $(s = 1, 2, \ldots, q)$, from (6.2.4) in the E-Step.

2) Compute $h_u^{(r+1)}(s)$ and $p^{(r+1)}$ for use in the $(r+2)^{\text{th}}$ iteration. Repeat these steps until convergence. To initialize the algorithm we can set $y^{(0)}(t) = 0$ such that $R_{(s)}^{(0)} = \sum_{l=s}^{q} d_{(s)}$.

Then;

$$p^{(1)} = \frac{\sum_{i=1}^{n} g_{is}^{(0)}}{n} = \frac{R_{(1)}^{(0)}}{n} \quad \text{and} \quad h_u^{(1)}(s) = \frac{d_{(s)}}{R_{(s)}^{(0)}}.$$

6.3 Application

To illustrate the application of the proposed model two data sets were considered. The first data set is the same *UnempDur* data which comes with **R** package `Ecdat` (Croissant and Graves, 2020). We track 2667 unemployed individuals until they exit the state of unemployment to either full-time employment(1073), part-time employment (339) or are censored (1255). Originally there were 3343 subjects in the data, but 574 subjects with missing failure causes and the 102 subjects with incomplete information are excluded from analysis. The time in two-week interval is $T \in \{1, 2, 3, \ldots, 28\}$. Again, we collapse failure times $T \geq 19$ into a single interval $[19, 28)$ such that the adjusted failure times are $T \in \{1, 2, 3, \ldots, 19\}$. Another data set that we have considered is *melonama* data from `boot` **R** package (Canty and Ripley, 2017). This data consists of 205 subjects who underwent surgery for removal of melanoma tumor. These subjects were tracked after surgey until they either died from melanoma (57), died from other death causes (14) or were censored (134). Time to failure for this data is continuous. For our purposes we have partitioned follow up with the following cut points; $\{0.5, 1, 1.5, 2, 2.5, 3, 4, 4.5, 5, 6, 7, 8, 9, 16\}$. Both data sets have a very large proportion of censored subjects relative to failure times. This is often taken as one of the indications that data may have cured subjects.

We shall begin with the unemployment data. There are subjects that are referred to as discouraged job seekers in economics. The term refers to unemployed individuals who have given up search for employment for various reasons, but mainly because the employment opportunities on offer do not match their expectations. These individuals can be regarded as cured as they may remain unemployed for an over extended period of time or may remain permanently unemployed. In Figure 6.1 we have plotted the population KM survival function estimate for the unemployment data. It can be seen that the survival function estimate does not approach 0 fast enough towards the close of

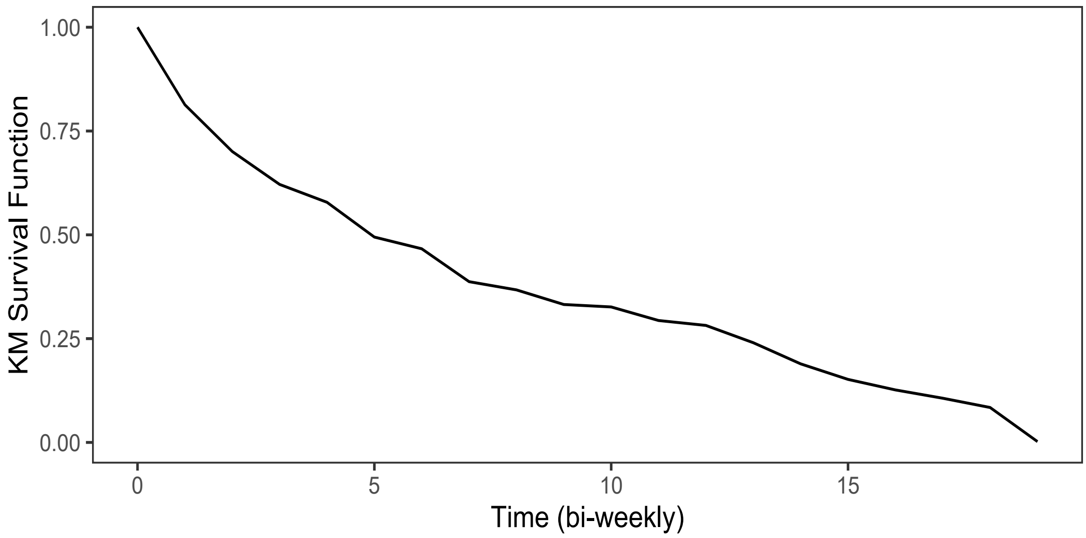

Figure 6.1: The KM Estimate for Population Survival Function for Unemployment data

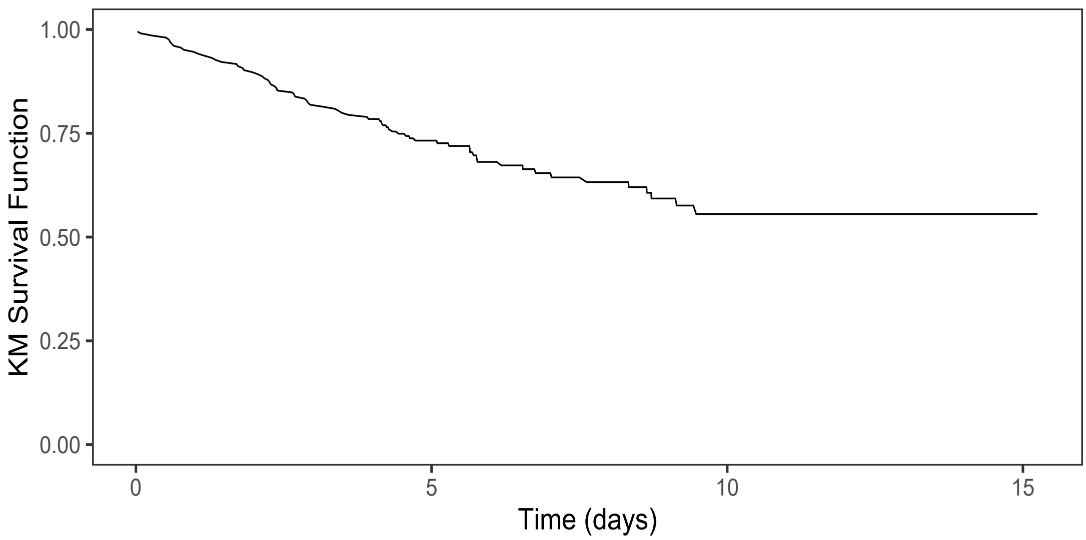

Figure 6.2: The KM Estimate for Population Survival Function for Melanoma data

the observation period. Clearly, the plot suggests a presence of cured subjects. It may be argued that follow up is not sufficiently long, that is, 1.083 years, but that is a seperate conversation. The data set suffices for the purpose of demonstrating the application of the proposed model, and besides, the proposed model is invariant to the size of cured subjects or length of follow up.

In Figure 6.2 we have plotted the KM survival function for *melanoma* data. Maller and Zhou (1995) have suggested a method for estimating the proportion of (un)cured subjects nonparametrically from the KM survival function. The method that was suggested by Maller and Zhou (1995) proposes that the proportion of uncured subjects can be read from $\hat{p} = 1 - \hat{S}(t^*)$ where t^* is the largest uncensored failure time. For *unemployment* data; $\hat{p} = 1 - \hat{S}(19) = 1 - 0.216302 = 0.783698$. Clearly, this suggests that the proportion of cured subjects for unemployment data is about 22%. For melanoma data we have $\hat{p} = 1 - \hat{S}(14) = 1 - 0.5787274 = 0.4212726$. This translates to about 42% of uncured subjects which, in turn, means 58% of the subjects are cured. When patients undergo some form of treatment against a given condition there is an expectation that some of the patients will recover from this condition. In the case of this melanoma data, the expectation is that some patient will be cured after undergoing surgery. This data suggests that 58% of the patients were cured of melanoma. We have applied the proposed model to both unemployment data and melanoma data to estimate conditional total hazards, conditional relative hazards and the proportion of uncured subjects. The results are displayed in Table 6.1 and Table 6.2. Analysis results do confirm that the proportion of cured subject for unemployment data and melanoma data is about 22% and 58%, respectively. We have also computed the estimates for conditional CIFs and also listed these estimates together with their standard errors in Table 6.1 and Table 6.2 for unemployment data and melanoma data, respectively. Recall that the conditional CIFs are computed from;

$$\hat{F}_{uj}(t) = \sum_{s=1}^{t} \hat{S}_u(s\text{-}1)\hat{h}_u(s)\hat{\Pi}_j(s). \tag{6.3.1}$$

Now, given these cure rates as suggested by unemployment data and melanoma data it is inappropriate to naively apply standard analysis methods to estimate, for example, the CSHs or the CIFs. If we were to model data with the CSHs model (Davis and Lawrance, 1989) or the ordinary nonparametric vertical model that we advanced in Chapter 5, clearly, these estimates will be understated because the proper approach is to exclude cured

Table 6.1: Maximum likelihood estimates for conditional total hazards, conditional relative hazards, p, and conditional CIFs for Unemployment data (with standard errors) (* denotes P $<$ 0.05)

	$\hat{\boldsymbol{h}}_u$	$\hat{\boldsymbol{\Pi}}_1$	$\hat{\boldsymbol{F}}_{u1}$	$\hat{\boldsymbol{F}}_{u2}$
T1	0.187(0.009)*	0.752(0.022)*	0.140(0.008)	0.046(0.005)
T2	0.138(0.009)*	0.761(0.028)*	0.226(0.009)	0.073(0.006)
T3	0.113(0.009)*	0.763(0.034)*	0.287(0.010)	0.092(0.006)
T4	0.069(0.008)*	0.727(0.051)*	0.318(0.010)	0.104(0.007)
T5	0.145(0.012)*	0.748(0.037)*	0.380(0.011)	0.125(0.008)
T6	0.057(0.009)*	0.762(0.066)*	0.402(0.011)	0.132(0.008)
T7	0.170(0.015)*	0.780(0.040)*	0.464(0.012)	0.149(0.008)
T8	0.051(0.010)*	0.625(0.099)*	0.476(0.012)	0.156(0.009)
T9	0.096(0.015)*	0.825(0.060)*	0.505(0.012)	0.163(0.009)
T10	0.018(0.007)*	0.500(0.204)*	0.508(0.013)	0.005(0.003)
T11	0.100(0.017)*	0.839(0.066)*	0.536(0.013)	0.171(0.009)
T12	0.040(0.013)*	0.700(0.145)*	0.544(0.013)	0.174(0.009)
T13	0.148(0.024)*	0.758(0.075)*	0.575(0.013)	0.184(0.010)
T14	0.212(0.033)*	0.833(0.062)*	0.618(0.014)	0.193(0.010)
T15	0.198(0.040)*	0.864(0.073)*	0.650(0.014)	0.198(0.011)
T16	0.167(0.044)*	0.769(0.117)*	0.670(0.015)	0.204(0.011)
T17	0.158(0.051)*	0.889(0.105)*	0.687(0.015)	0.206(0.011)
T18	0.210(0.068)*	0.778(0.139)*	0.705(0.015)	0.211(0.012)
T19	0.976(0.154)*	0.709(0.081)*	0.763(0.016)	0.235(0.014)

$\hat{p} = 0.785(0.0124)^*$

subjects when these estimates are computed as they relate to uncured subjects only. The extent to which the estimates are understated if cured subjects are not excluded depends

Table 6.2: Maximum likelihood estimates for conditional total hazards, conditional relative hazards, p, and conditional CIFs for Melanoma data (with standard errors)(* denotes P $<$ 0.05)

	$\hat{\boldsymbol{h}}_u$	$\hat{\boldsymbol{\Pi}}_1$	$\hat{\boldsymbol{F}}_{u1}$	$\hat{\boldsymbol{F}}_{u2}$
T1	0.035(0.020)*	0.000(0.000)*	0.000(0.000)	0.035(0.020)
T2	0.096(0.033)*	0.750(0.153)*	0.070(0.028)	0.058(0.025)
T3	0.067(0.029)*	0.800(0.179)*	0.116(0.035)	0.070(0.028)
T4	0.071(0.031)*	1.000(0.000)*	0.175(0.041)	0.070(0.028)
T5	0.139(0.044)*	0.889(0.105)*	0.268(0.048)	0.081(0.030)
T6	0.125(0.045)*	1.000(0.000)*	0.349(0.052)	0.081(0.030)
T7	0.143(0.051)*	0.857(0.132)*	0.419(0.054)	0.093(0.031)
T8	0.167(0.060)*	0.857(0.132)*	0.489(0.055)	0.105(0.033)
T9	0.093(0.052)*	1.000(0.000)*	0.527(0.055)	0.105(0.033)
T10	0.271(0.094)*	0.714(0.171)*	0.598(0.056)	0.133(0.038)
T11	0.221(0.119)*	1.000(0.000)*	0.657(0.057)	0.133(0.038)
T12	0.236(0.156)*	1.000(0.000)*	0.707(0.059)	0.133(0.038)
T13	0.523(0.252)*	0.333(0.354)*	0.735(0.058)	0.189(0.053)
T14	1.000(1.000)*	0.500(0.354)*	0.773(0.069)	0.227(0.069)

$\hat{p} = 0.421(0.043)^*$

on the relative size of cured subjects. In Figure 6.3 and Figure 6.4 we plotted conditional and population CIF estimates for unemployment and melanoma data, respectively. These plots suggest that the difference between conditional and population CIF estimates is directly proportional to the relative size of cured subjects, that is, the larger the proportion of cured subjects the larger is the difference between conditional and population CIF estimates.

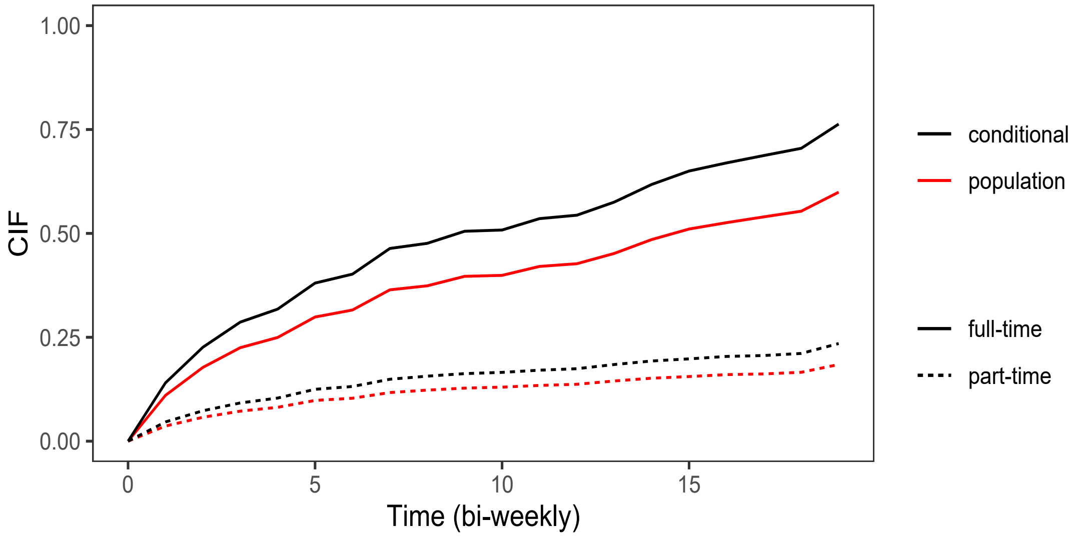

Figure 6.3: The conditional and population estimates for CIF for Unemployment data

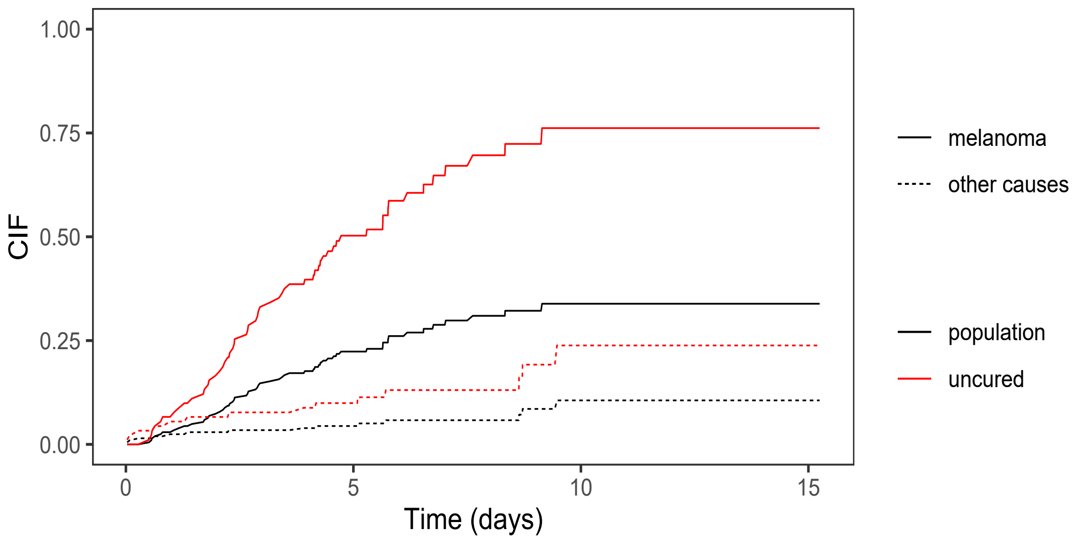

Figure 6.4: The conditional and population estimates for CIF for Melanoma data

Table 6.3: The population CIF estimates via the Mixture Cure Vertical Model and the Ordinary Vertical Model for Unemployment data (with standard errors)

	Mixture Cure Vertical Model		**Ordinary Vertical Model**	
	$\hat{\boldsymbol{F}}_1$	$\hat{\boldsymbol{F}}_2$	$\hat{\boldsymbol{F}}_1$	$\hat{\boldsymbol{F}}_2$
T1	0.110(0.006)	0.036(0.004)	0.110(0.006)	0.036 (0.004)
T2	0.177(0.008)	0.058(0.005)	0.177(0.007)	0.058 (0.005)
T3	0.225(0.009)	0.072(0.005)	0.225(0.008)	0.072 (0.005)
T4	0.250(0.009)	0.082(0.005)	0.250(0.009)	0.082 (0.005)
T5	0.299(0.010)	0.098(0.006)	0.299(0.009)	0.098 (0.006)
T6	0.316(0.010)	0.103(0.006)	0.316(0.009)	0.103 (0.006)
T7	0.364(0.011)	0.117(0.007)	0.364(0.010)	0.117 (0.007)
T8	0.374(0.011)	0.123(0.007)	0.374(0.010)	0.123 (0.007)
T9	0.397(0.012)	0.128(0.007)	0.397(0.011)	0.128 (0.007)
T10	0.399(0.012)	0.130(0.007)	0.399(0.011)	0.130 (0.007)
T11	0.421(0.012)	0.134(0.008)	0.421(0.011)	0.134 (0.007)
T12	0.427(0.012)	0.137(0.008)	0.427(0.011)	0.137 (0.008)
T13	0.452(0.013)	0.145(0.008)	0.452(0.012)	0.145 (0.008)
T14	0.485(0.013)	0.152(0.009)	0.485(0.012)	0.152 (0.008)
T15	0.511(0.014)	0.156(0.009)	0.511(0.013)	0.156 (0.009)
T16	0.526(0.014)	0.160(0.009)	0.526(0.013)	0.160 (0.009)
T17	0.540(0.014)	0.162(0.009)	0.540(0.014)	01626(0.009)
T18	0.553(0.015)	0.166(0.010)	0.553(0.014)	0.166 (0.009)
T19	0.599(0.016)	0.185(0.011)	0.599(0.015)	0.185 (0.011)

Table 6.4: The population CIF estimates via Mixture Cure Vertical Model and the Ordinary Vertical Model for Melanoma data (with standard errrs)

	Mixture Cure Vertical Model		**Ordinary Vertical Model**	
	$\hat{\boldsymbol{F}}_1$	$\hat{\boldsymbol{F}}_2$	$\hat{\boldsymbol{F}}_1$	$\hat{\boldsymbol{F}}_2$
T1	0.000(0.000)	0.015(0.008)	0.000(0.000)	0.015(0.008)
T2	0.029(0.012)	0.024(0.011)	0.029(0.012)	0.024(0.011)
T3	0.049(0.015)	0.029(0.012)	0.049(0.015)	0.029(0.012)
T4	0.074(0.019)	0.029(0.012)	0.074(0.018)	0.029(0.012)
T5	0.113(0.023)	0.034(0.013)	0.113(0.022)	0.034(0.013)
T6	0.147(0.026)	0.034(0.013)	0.147(0.025)	0.034(0.013)
T7	0.176(0.028)	0.039(0.014)	0.176(0.027)	0.039(0.014)
T8	0.206(0.030)	0.044(0.015)	0.206(0.028)	0.044(0.014)
T9	0.222(0.031)	0.044(0.015)	0.222(0.029)	0.044(0.014)
T10	0.252(0.033)	0056(0.017)	0.252(0.031)	0.056(0.017)
T11	0.277(0.036)	0.056(0.017)	0.277(0.033)	0.056(0.017)
T12	0.298(0.037)	0.056(0.017)	0.298(0.035)	0.056(0.017)
T13	0.310(0.038)	0.080(0.023)	0.310(0.037)	0.08 (0.023)
T14	0.326(0.039)	0.096(0.026)	0.326(0.039)	0.096(0.028)

While the nonparametric population CIF estimates can be estimated via the nonparametric vertical model introduced in Chapter 5, these quantities can also be estimated via the proposed nonparametric mixture cure vertical because the conditional total hazards, conditional relative hazards and the proportion of uncured subjects completely describe competing risks data with cured subjects. It may be instructive to compute the population CIF estimates via the proposed model and compare these estimates to the estimates that are obtained by applying the ordinary nonparametric vertical model that was introduced in Chapter 5. Recall that, under the nonparametric vertical model, the CIF estimates are computed from;

$$\hat{F}_j(t) = \sum_{s=}^{t} \hat{S}(s\text{-}1)\hat{h}(s)\hat{\Pi}_j(s)$$

with,

$$\begin{aligned}\mathrm{V}(\hat{F}_j(t)) = &\sum_{s=1}^{t} \mathrm{Var}\big(\hat{S}(s\text{-}1)\hat{h}(s)\hat{\Pi}_j(s)\big) \\ &+ 2\sum_{s=1}^{t\text{-}1}\sum_{k=s+1}^{t} \mathrm{Cov}\big(\hat{S}(s\text{-}1)\hat{h}(s)\hat{\Pi}_j(s), \hat{S}(k\text{-}1)\hat{h}(k)\hat{\Pi}_j(k)\big)\end{aligned}$$

where:

$$\begin{aligned}\mathrm{V}(\hat{S}(s\text{-}1)\hat{h}(s)\hat{\Pi}_j(s)) = &\big(\hat{S}(s\text{-}1)\hat{h}_j(s)\hat{\Pi}_j(s)\big)^2 \\ &\times \left(\sum_{l=1}^{s\text{-}1} \frac{d_{(l)}}{n_{(l)}(n_{(l)} - d_{(l)})} + \frac{n_{(s)} - d_{(s)}}{d_{(s)}n_{(s)}} + \frac{d_{(s)} - d_{(js)}}{d_{(s)}d_{(js)}}\right)\end{aligned}$$

and,

$$\begin{aligned}\mathrm{Cov}(\hat{S}(s\text{-}1)\hat{h}(s)\hat{\Pi}_j(s)\hat{S}(k\text{-}1)\hat{h}(k)\hat{\Pi}_j(k)) = &\big(\hat{S}(s\text{-}1)\hat{h}(s)\hat{\Pi}_j(s)\hat{S}(k\text{-}1)\hat{h}(k)\hat{\Pi}_j(k)\big) \\ &\times \left(\sum_{l=1}^{s\text{-}1} \frac{d_{(l)}}{n_{(l)}(n_{(l)} - d_{(l)})} - \frac{1}{n_{(s)}}\right).\end{aligned}$$

The population CIF estimates under the proposed model can be computed from;

$$\hat{F}_j(t) = \hat{p}\hat{F}_{uj}.$$

The expression for the variance of $\hat{F}_j(t)$, as advanced in Appendix I, is given by

$$\mathrm{V}(\hat{F}_j(t)) = \mathrm{V}(\hat{p})(\hat{F}_{uj}(t))^2 + \mathrm{V}(\hat{F}_{uj})(\hat{p})^2.$$

The estimates for the population CIF as computed via the proposed mixture cure vertical model and the ordinary vertical model as advanced in Chapter 5 are displayed in Table 6.3 and Table 6.4 for unemployment data and melanoma data, respectively. It can be seen that the population CIF estimates by the proposed mixture model and the ordinary vertical model are identical and the standard errors of the two estimation methods compare favorably. This is true for both data sets.

6.4 Discussion

In this chapter we have advanced a nonparametric vertical model for handling cured subjects in discrete time. This exercise is important if data comes with a sizable proportion of cured subjects as estimates for the standard functionals will be biased downwards if these subjects are ignored. In fact the standard functionals such as CSHs and CIFs will be underestimated if cured subjects are not accounted for and the extent of bias will depend on the size of cured subjects, the larger the proportion of cured subjects the larger is the extent of bias. We demonstrated this fact by fitting the proposed model to two data sets with different sizes of cured subjects and found that the difference between conditional and population CIF estimates is larger for data that has a larger proportion of cured subjects. We have established that the relative hazards are invariant to the presence or absence of cured subjects. Implementing an EM Algorithm is a tedious exercise and what this means is that all this could be avoided if one was only interested in relative hazards. In following chapter we apply all the regression models that were advanced in this thesis to new data set for validation purposes.

CHAPTER 7

Model Validation

7.1 The ordinary discrete time competing risks models

In the previous three chapters, Chapters 3, 4, and 5 we were concerned with advancing discrete time competing risks models as additional options to the multinomial and the binomial models. The proposed models were also upscaled as models that can handle data that comes with missing failure causes. In demonstrating the application of the proposed models we considered unemployment data that was originally analysed by McCall (1996). This data is given as '*UnempDur*' in `Ecdat` (Croissant and Graves, 2020) **R** package. In this data set, unemployed subjects are tracked from the moment they lost their jobs until they were re-employed full-time, part-time, or are still unemployed at the close of the observation period. Upon application of the proposed ordinary competing risks models to this data, we found that the models compared favourably with the multinomial model. We also established that the missing failure causes version of the proposed models compared favourably with the missing failure causes multinomial model, i.e., a multinomial model that has been upgraded to handle missing failure causes within the MI framework. One of the requirements in advancing a new model is the general applicability of the proposed model. The fact that this data set comes with only two failure causes may not be the best data set to demonstrate that the proposed models can be generalized to any discrete time competing risks data. The main thrust of this chapter is to address this very issue, that is, to demonstrate that the proposed models can be considered even in instances where the number of failure causes is more than two. Due to the limited availability of discrete time competing risks data that has more than two failure causes, we have settled for publicly available data *Clinical Randomization of an Antifibronolytic in Significant Hemorrhage 2* (CRASH 2) which is given as *crash2* from `discSurv` **R** package. In this data set patients are followed from the day of admission over a period of 28 days(4 weeks) until death or discharged alive, transfer to another hospital alive, and still alive in the same hospital when the observation was terminated. The last three states are considered as right censoring. The types of deaths are; *bleeding*, *head injury*, *vascular occlusion*,*multi-organ failure* and *other*. We have cross-classified the failure time and failure type and listed the results in Table 7.1.

Table 7.1: The distribution of failure time and failure type.

Week	censored	bleeding	head injury	vascular occlusion	multi organ failure	other
1	6026	871	901	52	259	161
2	4759	24	175	15	105	41
3	2147	2	58	8	34	25
4	1070	4	18	2	25	7
5	2245	0	1	2	0	1

Though the observation period closed in week 4, the patients listed in week 5 are those that survived week 4, but were subsequently censored (2245), died from; *head injury*(1), *vascular occlusion*(2) and *other*(1).

The explanatory variables that come with the data are; *sex (male,female)*,*age*, *injury type(blunt, penetrating, blunt and penetrating)*, *injurytime(time between injury and admission)*, *sbp(systolic blood pressure (mmHg))*,*rr(respiratory rate per minute)*, *cc(central capillary refill time in seconds)*, *hr(heart rate per minute)* and *gcs(Glascow Coma Score)*.

One of the requirements for the application of the proposed models, especially, the mixture models is that the failure times with zero events should be avoided so as to improve the

Table 7.2: The distribution of failure time and failure type.

Week	censored	bleeding	head injury	vascular occlusion	multi organ failure	other
1	6026	871	901	52	259	161
2	4759	24	175	15	105	41
3	2147	2	58	8	34	25
4	3315	4	19	4	25	8

stability of parameter estimates. Consequently, we have collapsed week 4 and week 5 into week 4 as shown in Table 7.2. We will commence this exercise of validating the proposed discrete time models by fitting them to this data set. We will first attend to the ordinary discrete time models. Unfortunately, the limitation of this data set is that it does not come with missing failure causes for the purpose of validating the proposed missing failure causes models. To overcome this hurdle, the standard procedure is to simulate these subjects from the given data set that has complete status with respect to the causes of failure. We will discuss the mechanics of this exercise in detail later when we attend to the validation of the missing failure causes models. We now proceed to fit the three proposed ordinary discrete time competing risks models, that is, the *continuous time mixture model*, the *discrete time mixture model* and the *discrete time vertical model* to the data. We commence with the *continuous time mixture model*. For illustrative purposes, we have included all the covariates that came with data except *gcs* and *sbp*. The covariates *female* and *injurytype* are factors where the latter has three levels, namely; *blunt*, *penetrating* as well as *penetrating & blunt*. For these factor variables, that is, *female* and *injurytype*, the *male* and *blunt* categories are regarded as reference categories, respectively. The continuous covariates were centered at their respective averages.

Table 7.3: Maximum likelihood estimates for the Continuous Time Mixture Model (with standard errors) (* denotes $P < 0.05$).

Continuous Time Mixture Model					
			Latency		
Coefficient	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
T1	-2.208(0.054)*	-1.738(0.042)*	-2.331(0.219)*	-4.053(0.144)*	-2.844(0.101)*
T2	-5.251(0.208)*	-2.839(0.079)*	-2.724(0.290)*	-1.660(0.113)*	-3.769(0.169)*
T3	-7.138(0.708)*	-3.358(0.134)*	-2.558(0.510)*	-2.038(0.181)*	-3.679(0.209)*
T4	-5.981(0.502)*	-3.984(0.231)*	-2.589(0.510)*	-1.699(0.208)*	-4.339(0.359)*
female	0.062(0.093)*	-0.169(0.078)*	0.852(0.337)*	-0.746(0.138)*	-0.385(0.193)*
age	0.012(0.002)*	0.016(0.002)*	0.037(0.007)*	0.004(0.003)	0.021(0.004)*
penetrate	-0.259(0.077)*	-0.965(0.109)*	2.037(0.303)	-0.202(0.119)	-0.523(0.162)*
penetrate and blunt	0.023(0.100)	0.121(0.082)*	2.085(0.346)*	0.477(0.138)*	-1.014(0.256)*
injurytime	-0.126(0.018)*	-0.023(0.014)	-0.085(0.052)	-0.018(0.194)	-0.025(0.031)
hr	0.015(0.002)*	0.009(0.001)*	-0.012(0.006)*	0.005(0.002)*	-0.025(0.031)
rr	0.027(0.044)*	0.004(0.005)*	0.006(0.014)	0.013(0.006)*	0.017(0.009)
cc	0.273(0.016)*	0.173(0.017)*	-0.306(0.074)*	0.214(0.023)*	-0.033(0.035)*

Table 7.4: Maximum likelihood estimates for the Continuous Time Mixture Model (with standard errors) (* denotes $P < 0.05$).

Continuous Time Mixture Model

Incidence

Coefficient	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$
`Constant`	-0.628(0.022)*	-0.464(0.022)*	-3.883(0.079)*	-2.827(0.045)*
`female`	-0.169(0.042)*	0.088(0.043)*	-0.853(0.210)*	0.379(0.073)*
`age`	-0.002(0.001)	-0.002(0.001)*	0.015(0.004)*	0.019(0.002)*
`penetrating`	0.747(0.034)*	-0.940(0.039)*	-1.451(0.198)*	0.113(0.069)
`penetrating & blunt`	0.199(0.048)*	-0.246(0.049)*	0.042(0.237)*	-0.128(0.098)
`injurytime`	-0.014(0.007)*	0.004(0.007)	0.042(0.019)*	0.031(0.012)*
`hr`	0.002(0.002)	-0.002(0.001)*	0.009(0.003)*	0.009(0.001)*
`rr`	0.002(0.002)	-0.009(0.003)*	0.019(0.008)*	0.003(0.004)
`cc`	-0.083(0.009)*	-0.085(0.010)*	0.228(0.025)*	0.055(0.017)*

Table 7.5: Maximum likelihood estimates for the Multinomial Model (with standard errors) (* denotes $P < 0.05$).

Multinomial Model

Coefficient	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$
`T1`	-3.185(0.056)*	-2.638(0.043)*	-5.894(0.194)*	-4.297(0.086)*	-4.463(0.101)*
`T2`	-6.305(0.209)*	-3.842(0.081)*	-6.686(0.291)*	-4.766(0.115)*	-5.383(0.169)*
`T3`	-8.186(0.713)*	-4.393(0.135)*	-6.740(0.378)*	-5.323(0.182)*	-5.291(0.210)*
`T4`	-7.008(0.503)*	-5.037(0.232)*	-6.958(0.518)*	-5.174(0.210)*	-5.947(0.360)*
`female`	-0.059(0.096)*	-0.113(0.080)	-0.351(0.320)*	-0.245(0.140)*	-0.369(0.193)*
`age`	0.013(0.002)*	0.015(0.002)*	0.040(0.007)*	0.024(0.003)*	0.019(0.004)*
`penetrating`	0.149(0.080)	-1.608(0.110)	-0.201(0.288)	-0.103(0.121)	0.459(0.162)*
`penetrating & blunt`	0.156(0.104)	-0.049(0.084)*	0.270(0.317)*	0.218(0.138)*	-0.807(0.257)*
`injurytime`	-0.127(0.019)*	-0.022(0.014)	-0.002(0.049)	0.013(0.020)*	-0.035(0.031)*
`hr`	0.014(0.002)*	0.008(0.002)*	0.004(0.006)*	0.015(0.002)*	0.011(0.003)*
`rr`	0.028(0.005)	-0.001(0.005)*	0.024(0.016)*	0.014(0.007)*	0.023(0.009)*
`cc`	0.200(0.017)*	0.113(0.017)*	0.094(0.064)*	0.206(0.022)*	0.118(0.035)*

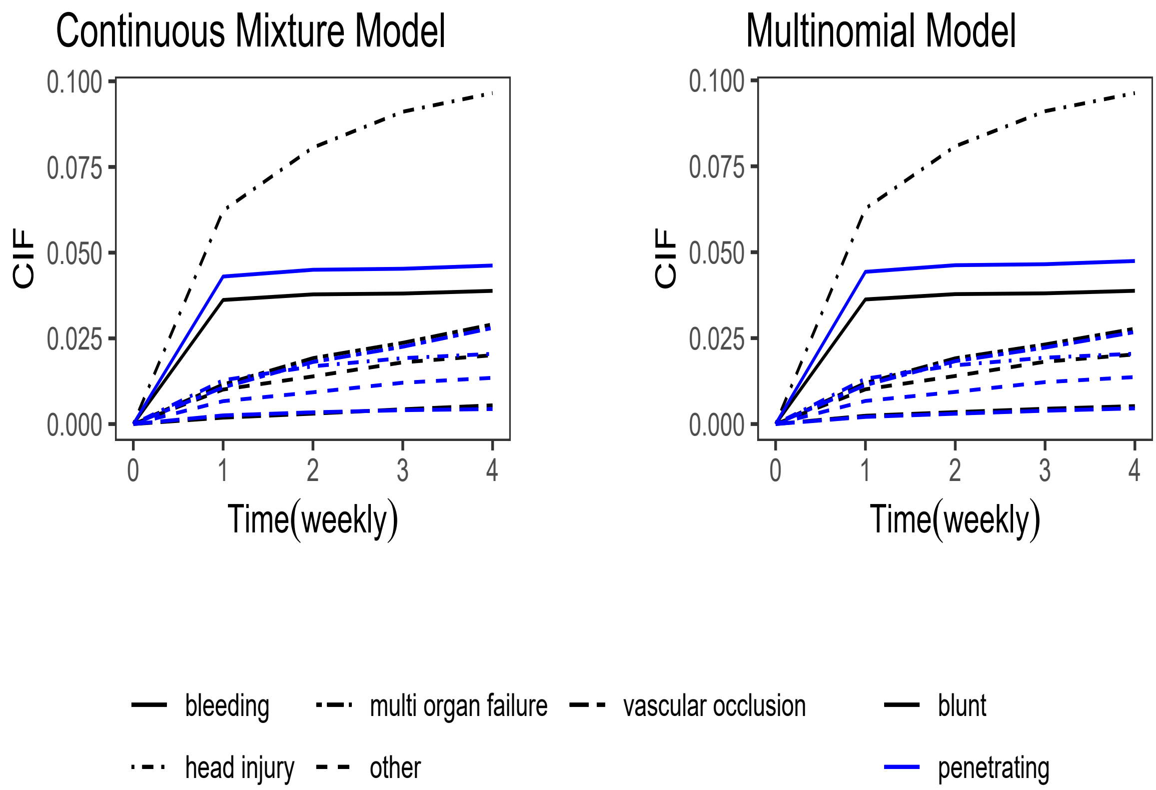

Figure 7.1: The CIF of death from bleeding, head injury, multi organ failure, vascular occlusion and other causes with the effect of injury type(*blunt* vs *penetrating*) via the Continuous Time Mixture Model and the Multinomial Model.

Naturally, we applied a Poisson regression model to estimate $\boldsymbol{\beta}$ the regression coefficients that correspond to component hazards where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\beta}_3^T, \boldsymbol{\beta}_4^T)$, such that the model for component hazards is given by

$$\lambda_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j) = \exp(\beta_{0jt}) \exp(\boldsymbol{x}^T \boldsymbol{\beta}_{1j})$$

for $j = 1 \ldots, J$. Recall that the model for failure type probabilities is given by

$$\pi_j(\boldsymbol{x}, \boldsymbol{\gamma}) = \exp(\gamma_{0j} + \boldsymbol{x}^T \boldsymbol{\gamma}_{1j}) \Big/ (1 + \sum_{l=1}^{J-1} \exp(\gamma_{0l} + \boldsymbol{x}^T \boldsymbol{\gamma}_{1l}) \tag{7.1.1}$$

where $\pi_J(\boldsymbol{x}, \boldsymbol{\gamma}) = 1 - \sum_{j=1}^{J-1} \pi_j(\boldsymbol{x}, \boldsymbol{\gamma})$ and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^T, \ldots, \boldsymbol{\gamma}_{J\text{-}1}^T)^T$.

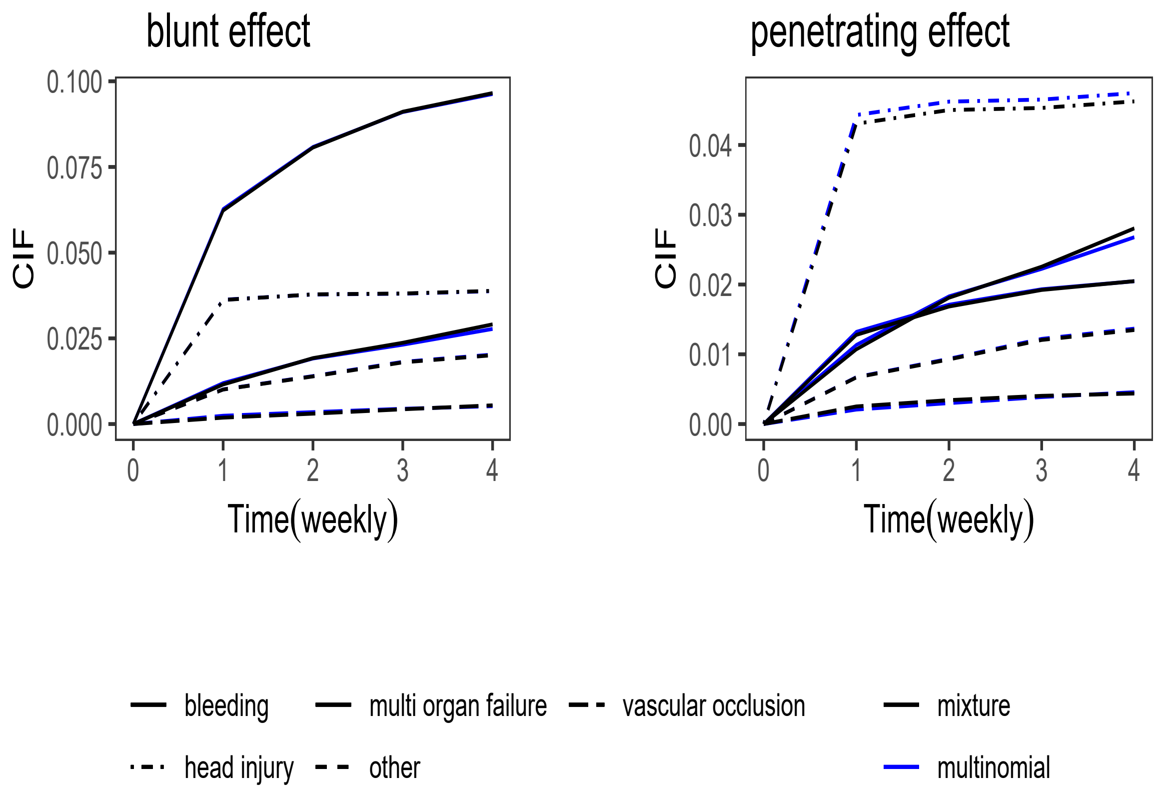

Figure 7.2: The CIF of death from bleeding, head injury, multi organ failure,vascular occlusion and other causes with the effect of injury type via the Continuous Time Mixture Model and the Multinomial Model.

Ideally, to estimate the failure type probabilities we would proceed to fit a multinomial model to estimate $\boldsymbol{\gamma}$, however, this is not possible as some of the responses may be non-integer. Instead, we fitted 4 binomial distributions to estimate $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3$ and $\boldsymbol{\gamma}_4$

The results of the analysis are displayed in Table 7.3. We also fitted a multinomial model to the data for comparison purposes where the model for CSHs is given by

$$h_j(t|\boldsymbol{x};\boldsymbol{\alpha}) = \frac{\exp\left(\alpha_{0jt} + \boldsymbol{x}^T\boldsymbol{\alpha}_{1j}\right)}{1 + \sum_{l=1}^{J}\exp\left(\alpha_{0lt} + \boldsymbol{x}^T\boldsymbol{\alpha}_{1l}\right)}$$

The results of fitting a multinomial model are displayed in Table 7.5. For illustrative purposes, throughout this chapter we will investigate the effect of injury type on the likelihood of dying by these causes of death, in particular, we will assess the effect of *penetrating* injury type relative to the reference injury type, i.e., *blunt* injury type. It must be noted that deaths are rare in this data set judging by the relative size of patients that are censored (85%) by the way of discharge or transfer. As a result, the size of the CIFs will be relatively small. We then proceeded to plot the CIFs for both the proposed mixture model and the multinomial model with the effect of *injury type*, i.e., we contrasted the effect of *blunt* and *penetrating* types of injury on failure types and plotted these results in Figure 7.1. It is evident from Figure 7.1 that the *penetrating* type of injury has the effect of dampening the likelihood of dying by all types of death except for death from *bleeding*. Put differently, a patient with a penetrating type of injury is less likely to die by all causes of death than a patient who has a blunt type of injury except for death from bleeding. Secondly, the penetrating type of injury when compared to the blunt type of injury has the largest impact on *head injury* type of death and the least effect on death by *vascular occlusion*, i.e., the gap between the penetrating and blunt effects is largest for death due to *head injury* and the smallest for death by *vascular occlusion*. If we were to examine the CIF from the multinomial model we would arrive at a similar conclusion regarding the effect of injury type. Figure 7.2 does confirm that the examination of the CIFs from the multinomial model or the continuous time mixture would lead to the same conclusion regarding the effect of injury type (*blunt* and *penetrating*) on death types. Ideally, if this data came with missing failure causes we would now proceed to fit a continuous time mixture model that can handle these subjects so as to determine the effect of their inclusion in relation to the impact of injury type.

We now attend to the *discrete time mixture model*. Recall that for this model we assumed that the failure type probabilities follow the same model given in (7.2), while, the model for component hazards is given by

$$g(\lambda_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j)) = \beta_{0jt} + \boldsymbol{x}^T \boldsymbol{\beta}_{1j}$$

We proceeded to to fit this model and the results of the analysis are displayed in Table 7.6

Table 7.6: Maximum likelihood estimates for the Discrete Time Mixture Model (with standard errors) (* denotes $P< 0.05$).

Continuous Time Mixture Model					
Latency					
Coefficient	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
`T1`	-2.350(0.058)*	-1.886(0.045)*	-1.421(0.2166)*	-0.389(0.115)*	-1.629(0.111)*
`T2`	-5.518(0.210)*	-3.079(0.082)*	-1.794(0.365)*	-0.213(0.152)*	-2.488(0.180)*
`T3`	-7.421(0.7087*	-3.628(0.136)*	-1.303(0.469)*	-0.221(0.248)*	-2.285(0.224)*
`T4`	-6.277(0.503)*	-4.278(0.233)*	-1.123(0.630)*	2.830(0.821)*	-2.890(0.373)*
`female`	-0.060(0.099)*	-0.140(0.083)*	1.527(0.520)*	-0.451(0.186)*	-0.357(0.212)*
`age`	0.014(0.002)*	0.017(0.002)*	0.032(0.010)*	-0.001(0.004)	0.014(0.005)*
`penetrate`	-0.369(0.081)*	-0.510(0.114)*	2.460(0.494)	-0.125(0.163)	-0.785(0.174)*
`penetrate and blunt`	0.020(0.108)	0.017(0.088)*	1.782(0.520)*	0.333(0.205)*	0.028(0.287)*
`injurytime`	-0.118(0.019)*	-0.039(0.015	-0.156(0.072)	-0.046(0.026)	-0.003(0.035)
`hr`	0.014(0.002)*	0.009(0.002)*	-0.004(0.008)*	0.000(0.003)*	0.012(0.004)
`rr`	0.024(0.005)*	0.006(0.005)*	-0.005(0.019)	0.0007(0.008)*	0.035(0.011)
`cc`	0.178(0.018)*	0.103(0.018)*	-0.501(0.019)*	0.227(0.042)*	0.370(0.047)*

Table 7.7: Maximum likelihood estimates for the Discrete Time Mixture Model (with standard errors) (* denotes $P < 0.05$).

Discrete Time Mixture Model

Incidence

Coefficient	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$
`Constant`	-0.359(0.022)*	-0.091(0.022)*	-4.593(0.109)*	-3.521(0.063)*
`female`	-0.333(0.039)*	0.054(0.040)*	-5.397(0.260)*	-3.604(0.073)*
`age`	-0.004(0.001)	-0.002(0.001)*	0.024(0.005)*	0.022(0.003)*
`penetrating`	1.322(0.036)*	-1.581(0.041)*	-1.023(0.230)*	-0.002(0.098)
`penetrating & blunt`	0.269(0.047)*	-0.119(0.047)*	-0.381(0.251)*	0.128(0.123)
`injurytime`	-0.031(0.007)*	0.029(0.007)	0.043(0.023)*	0.035(0.015)*
`hr`	0.000(0.001)	-0.003(0.001)*	0.004(0.004)*	0.012(0.002)*
`rr`	0.011(0.002)	-0.014(0.003)*	0.022(0.010)*	0.010(0.006)
`cc`	0.065(0.010)*	-0.083(0.010)*	0.274(0.029)*	0.132(0.020)*

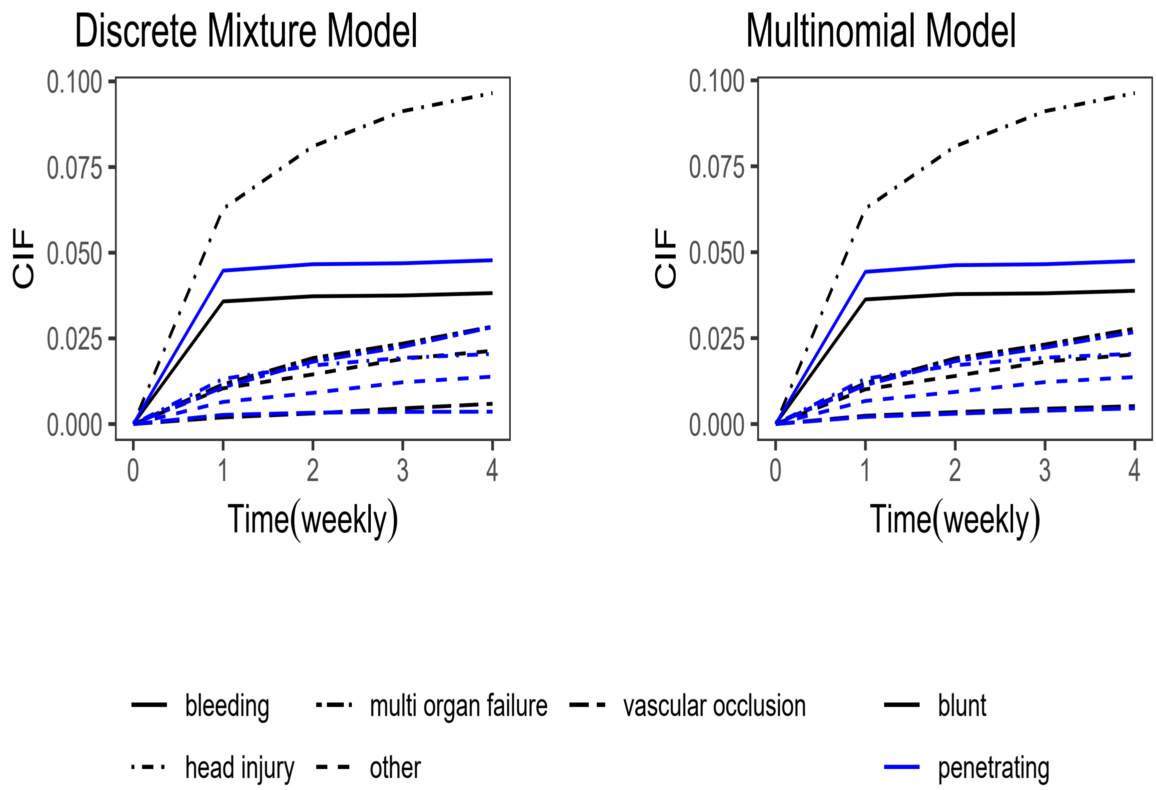

Figure 7.3: The CIF of death from bleeding, head injury, multi organ failure, vascular occlusion and other causes with the effect of injury type via the Discrete Time Mixture Model and the Multinomial Model.

Here as well we have assessed the effect of injury type, that is, *penetrating* injury relative to the *blunt* injury type. Naturally, we conducted this exercise by examining the CIFs. Clearly, from examining Figure 7.3 and Figure 7.4 we come to the same conclusion that the risk of dying by any type of death is lower for a patient that has a *penetrating* type of injury compared to a patient that has a *blunt* type of injury except for death by bleeding. Suppose we were interested in death from bleeding and the effect of injury on this type of death. The *continuous time mixture model* suggests that the chances that a patient with a *blunt* type of injury will die from *bleeding* is 3.89% and this probability increases to 4.62% for a patient with a *penetrating* injury. On the other hand, the *discrete time mixture model* indicates 3.82% and 4.77%, respectively, that is, the chances of dying from

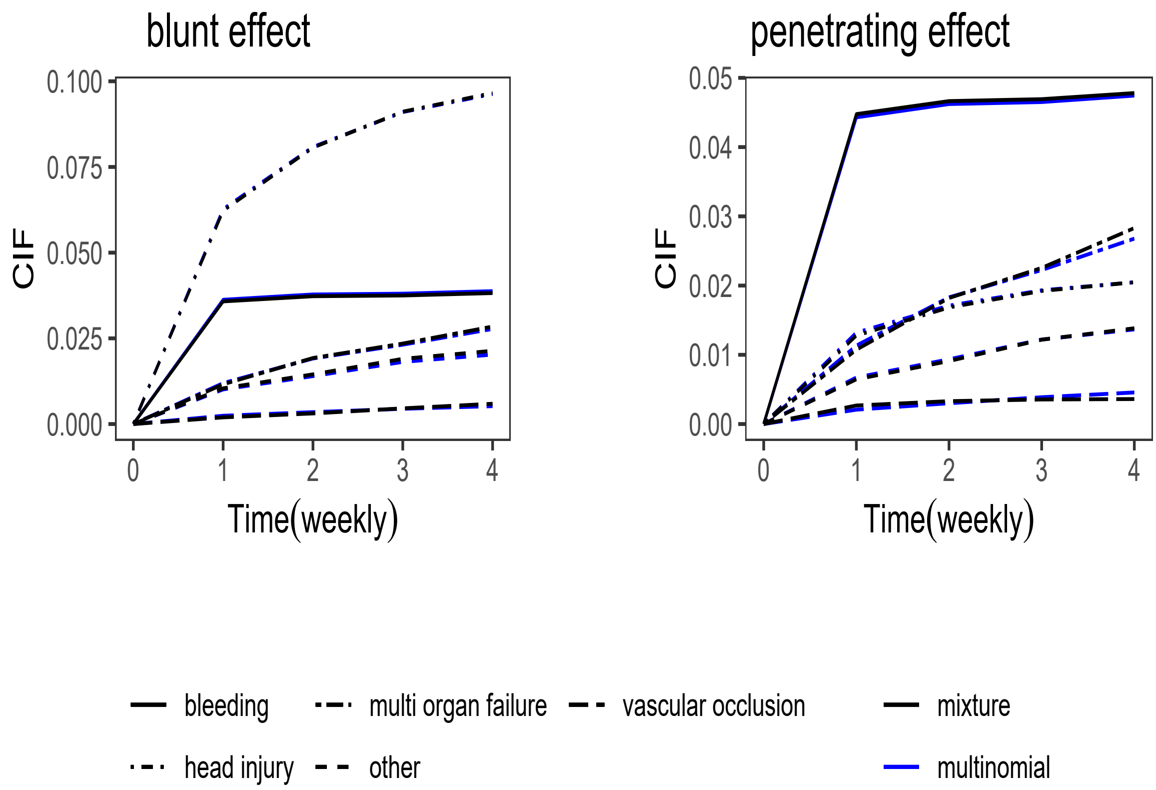

Figure 7.4: The CIF of death from bleeding, head injury, multi organ failure, vascular occlusion and other causes with the effect of injury type via the Discrete Time Mixture Model and the Multinomial Model.

bleeding is 3.82% and 4.77% for a patiens admitted with a*blunt* and a *penetrating* types of injuries, respectively. The multinomial suggests 3.88% and 4.74%, respectively for a patient with a *blunt* type of injury and another one with a *penetrating* injury to die due to *bleeding*. A patient who has suffered a *blunt* type of injury can only bleed internally while a *penetrating* injury would lead to external and visible bleeding. This evidence suggests that external loss of blood is more severe than internal bleeding.

Generally, from the plots of the CIFs, there is a strong indication that the mixture models compare favourably with a multinomial model. The next model to consider is the proposed *discrete time vertical model*.

Table 7.8: Maximum likelihood estimates for the Discrete Time Vertical Model (with standard errors) (* denotes $P < 0.05$).

Multinomial Model

Coefficient	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$\hat{\beta}$
`T1`	1.279(0.016)*	1.844(0.110)*	-1.394(0.217)*	0.159(0.133)*	-1.935(0.031)*
`T2`	-0.847(0.270)*	1.535(0.187)*	-1.290(0.335)*	0.648(0.203)*	-3.317(0.058)*
`T3`	-2.786(0.741)*	0.876(0.253)*	-1.495(0.437)*	0.014(0.279)*	-3.782(0.093)*
`T4`	-1.013(0.062)*	0.894(0.433)*	-1.023(0.636)*	0.864(0.420)*	-4.063(0.133)*
`female`	0.336(0.219)*	0.275(0.209)	0.023(0.636)*	0.087(0.239)*	-0.144(0.056)*
`age`	-0.005(0.005)*	-0.004(0.005)*	0.018(0.008)*	0.004(0.005)*	0.017(0.001)*
`penetrating`	0.616(0.183)	-1.179(0.196)	0.231(0.329)	0.305(0.202)	-0.505(0.052)*
`penetrating & blunt`	0.946(0.279)	0.735)(0.270)*	1.039(0.408)*	1.016(0.292)*	0.007(0.059)*
`injurytime`	-0.078(0.038)*	0.019(0.034)	0.027(0.057)	0.037(0.038)*	-0.044(0.010)*
`hr`	0.001(0.003)*	-0.004(0.003)*	-0.004(0.005)*	0.002(0.004)*	0.011(0.001)*
`rr`	0.005(0.009)	-0.001(0.005)*	0.001(0.015)*	-0.002(0.009)*	0.014(0.003)*
`cc`	0.070(0.044)*	-0.027(0.043)*	-0.032(0.081)*	0.119(0.046)*	0.158(0.011)*

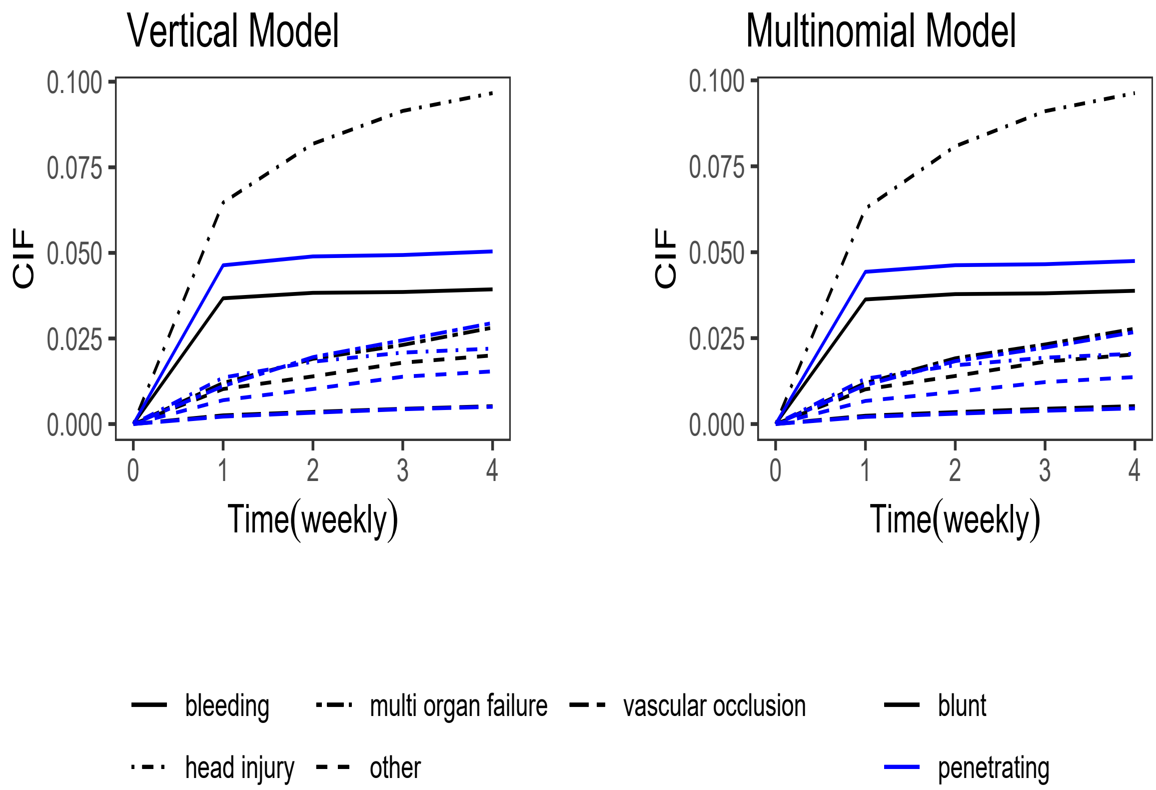

Figure 7.5: The CIF of death from bleeding, head injury, multi organ failure, vascular occlusion and other causes with the effect of injury type via the Discrete Time Vertical Model and the Multinomial Model.

Recall that, with covariates, the relative hazards model is given by

$$\Pi_j(t|\boldsymbol{x}) = \frac{\exp\left(\phi_{0jt} + \boldsymbol{x}^T\boldsymbol{\phi}_{1j}\right)}{1 + \sum_{l=1}^{J\text{-}1}\left(\exp\phi_{0lt} + \boldsymbol{x}^T\boldsymbol{\phi}_{1l}\right)}, \tag{7.1.2}$$

and we specified the following regression model for total hazards;

$$g(h(t|\boldsymbol{x})) = \beta_{0t} + \boldsymbol{x}^T\boldsymbol{\beta}_1 \tag{7.1.3}$$

The results of modelling this data with the proposed *discrete time vertical model* are listed in Table 7.8. In Figures 7.5 & 7.6 we have plotted the CIFs from the proposed *discrete time vertical model* and the multinomial model.

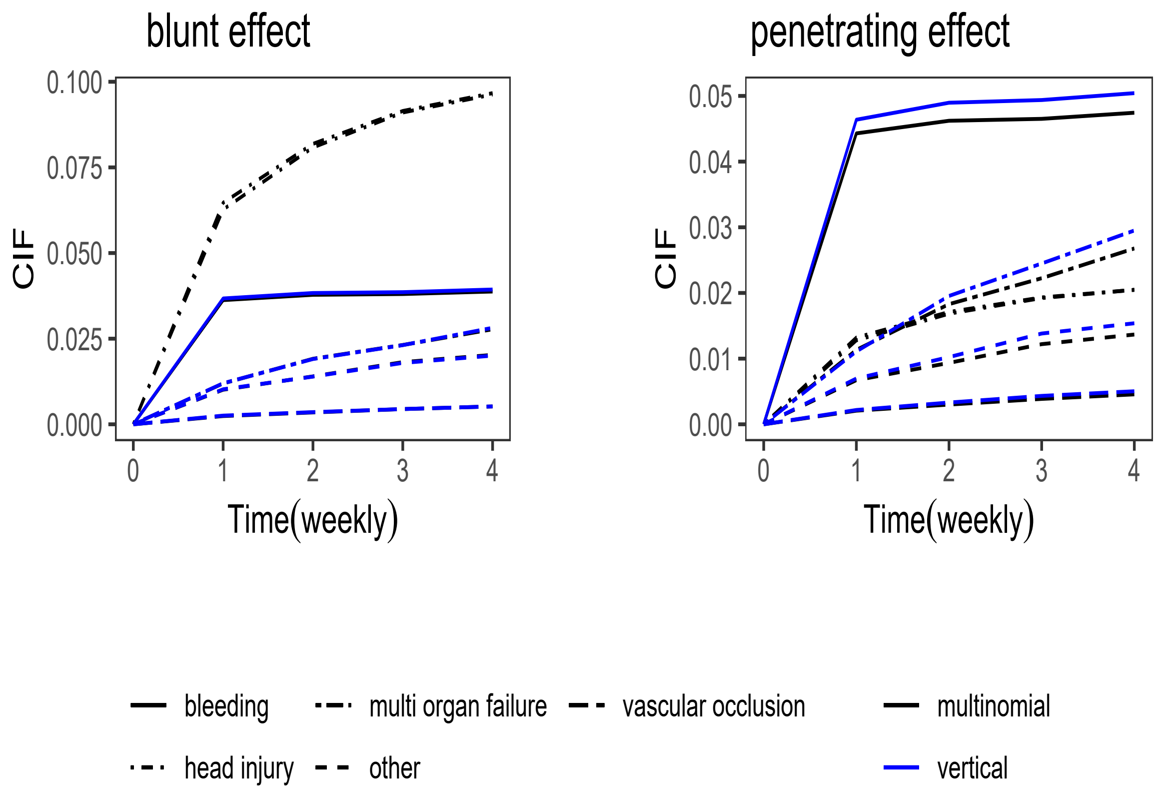

Figure 7.6: The CIF of death from bleeding, head injury, multi organ failure, vascular occlusion and other causes with the effect of injury type via the Discrete Time Vertical Model and the Multinomial Model.

If we again consider the effect of injury type on the cause of death attributable to *bleeding*, we now find that 3.93% of the patients that suffered from a *blunt* type of injury died from bleeding while 5.04% of the patients that had *penetrating* injuries died from the same cause. These figures are slightly larger than 3.87% and 4.74% that were suggested by the multinomial model and roughly by both the mixture models, but still do suggest that the chances of dying by bleeding are larger for a patient with a *penetrating* injury than a patient with a *blunt* injury.

All in all, these findings indicate that all three proposed models, i.e., the *continuous time mixture model*, the *discrete time mixture model* and the *discrete time vertical model* compare favourably with the multinomial model when applied to this data. In the next section we attend the the missing failure causes version of the we have just discussed.

7.2 The missing failure causes model

As already mentioned earlier that this data set, unlike the unemployment data, does not come with a natural set of subjects with missing failure causes. To overcome this hurdle, we will have to generate these subjects from the data so as to test the missing failure causes models that we have proposed in Chapters 3, 4 and 5. Given the fact that the number of deaths by bleeding and vascular occlusion are relatively fewer for weeks 3 and 4 we have combined these two failure causes to stabilize the sample from which we will draw the subjects with missing failure causes. We will then proceed to fit each of the proposed ordinary competing risks models to complete cases and then immediately follow that up with a missing failure causes model for comparison purposes. We will commence with the *continuous time mixture model*.

Because our data has certain failure causes with fewer events in later failure times even after combining bleeding and vascular occlussion failures, we have exercised care in generating subjects with missing failure causes such that missingness largely affects the subjects that failed earlier. Recall that when data comes with missing failure causes an indicator variable R is assumed which takes values 1 or 0 according to whether a subject has complete information regarding failure cause or not. Recall also that we assumed that the failure causes are "missing at random" (MAR) (Little and Rubin, 1987). In more formal terms we have assumed that;

$$P(R_i = 0|\Delta_i, \Delta_i > 0, \boldsymbol{w}_i) = P(R_i = 0|\Delta_i > 0, \boldsymbol{w}_i) \tag{7.2.1}$$

where $\boldsymbol{w}_i = (t_i, \boldsymbol{x}_i, \boldsymbol{z}_i)$, is a vector of observed data for subject i with $\boldsymbol{z}$ as the covariate vector that may explain missingness. As mentioned earlier, we assume that $\boldsymbol{z} = 0$, that is $\boldsymbol{w}_i = (t_i, \boldsymbol{x}_i)$.

Table 7.9: Maximum likelihood estimates for the Complete Case Multinomial Model (with standard errors) (* denotes $P < 0.05$).

Multinomial Model

Coefficient	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$
`T1`	-3.629(0.068)*	-4.483(0.063)*	-5.894(0.194)*	-4.266(0.143)*
`T2`	-6.331(0.201)*	-4.698(0.109)*	-5.670(0.157)*	-6.124(0.216)*
`T3`	-6.946(0.358)*	-5.240(0.175)*	-6.005(0.220)*	-6.304(0.297)*
`T4`	-6.481(0.359)*	-5.804(0.283)*	-6.147(0.284)*	-6.594(0.422)*
`female`	-0.252(0.111)*	-0.073(0.098)	-0.429(0.172)*	-0.257(0.222)*
`age`	0.040(0.003)*	0.043(0.002)*	0.051(0.004)*	0.048(0.005)*
`penetrating`	0.124(0.094)	-1.578(0.153)	-0.172(0.159)	-0.178(0.201)
`penetrating & blunt`	0.2386(0.116)	0.098(0.104)*	0.440(0.158)*	-0.429(0.289)*
`injurytime`	-0.152(0.022)*	-0.044(0.018)	0.006(0.023)	-0.089(0.042)*
`hr`	0.023(0.002)*	0.024(0.002)*	0.028(0.003)*	0.018(0.004)*
`rr`	-0.008(0.006)	-0.032(0.006)*	-0.021(0.009)*	-0.014(0.013)*
`cc`	0.210(0.019)*	0.132(0.021)*	0.247(0.025)*	0.154(0.042)*

We assumed a logit model for the missingness mechanism, that is;

$$\text{Logit}(P(R_i = 0|\Delta_i > 0, \boldsymbol{w}_i)) = \boldsymbol{w}^T\boldsymbol{\tau}$$

where; $\boldsymbol{\tau} = (-0.01, 2.6, -3.0, 0.5, 0.5, -0.5, -0.5, 1.5, 1.5, 1.25, 1.0)$, and $\boldsymbol{x} = (time, female$ $, age, penetrating, penetrating\ and\ blunt, injurytime, hr, rr, cc, gcs, sbp)$. This sampling scheme gave rise to 988 subjects with missing failure causes which is about 5.2% of the entire sample.

Firstly, we have proceeded to fit the proposed ordinary competing risks models together with the multinomial model to the complete case data and the results of the analysis are displayed in Table 7.9 through to Table 7.12. Secondly, we have plotted the CIFs estimates from all four models, the multinomial model, the two mixture models and the vertical model, with the impact of injury type in Figure 7.7 for comparison purposes.

Table 7.10: Maximum likelihood estimates for the Complete Case Analysis Continuous Time Mixture Model (with standard errors) (* denotes $P < 0.05$).

Continuous Time Mixture Model

Latency				
Coefficient	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
`T1`	-2.514(0.067)*	-2.309(0.062)*	-2.505(0.114)*	-4.053(0.144)*
`T2`	-5.127(0.199)*	-3.379(0.106)*	-3.079(0.154)*	-4.877(0.214)*
`T3`	-5.750(0.357)*	-3.873(0.172)*	-3.349(0.217)*	-5.350(0.420)*
`T4`	-5.294(0.357)*	-4.403(0.281)*	-3.424(0.281)*	-5.350(0.420)*
`female`	-0.282(0.107)*	0.227(0.095)*	-0.679(0.179)*	-0.348(0.220)*
`age`	0.042(0.002)*	0.033(0.002)*	0.019(0.004)*	0.063(0.005)*
`penetrating`	0.325(0.091)*	-1.770(0.152)*	-0.228(0.157)	-0.129(0.201)
`penetrating & blunt`	0.197(0.111)	0.385(0.100)*	0.331(0.156)*	-0.661(0.288)*
`injurytime`	-0.143(0.021)*	-0.030(0.018)	0.005(0.021)*	-0.097(0.042)*
`hr`	0.023(0.002)*	0.021(0.002)*	0.016(0.003)*	0.022(0.004)*
`rr`	-0.019(0.006)*	-0.019(0.006)*	-0.006(0.008)*	-0.013(0.013)*
`cc`	0.043(0.018)*	0.225(0.580)*	0.279(0.027)*	0.307(0.043)*

Incidence			
Coefficient	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
`Constant`	-0.766(0.024)*	-0.867(0.024)*	-2.431(0.040)*
`female`	0.038(0.045)	-0.305(0.047)*	0.255(0.064)*
`age`	-0.060(0.001)*	0.012(0.001)*	0.032(0.002)*
`penetrating`	-0.174(0.038)*	0.167(0.038)*	0.074(0.061)
`penetrating & blunt`	0.060(0.052)	-0.322(0.056)*	0.144(0.078)
`injurytime`	-0.010(0.008)	-0.014(0.008)	0.016(0.011)
`hr`	-0.002(0.001)*	0.004(0.001)*	-0.010(0.001)*
`rr`	0.014(0.003)*	-0.015(0.003)*	-0.012(0.004)*
`cc`	0.254(0.011)*	-0.111(0.011)*	0.008(0.016)

Table 7.11: Maximum likelihood estimates for the Complete Case Analysis Discrete Time Mixture Model (with standard errors) (* denotes $P< 0.05$).

Discrete Time Mixture Model

Latency				
Coefficient	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
`T1`	-2.939(0.069)*	-2.641(0.066)*	0.063(0.164)*	-2.193(0.155)*
`T2`	-5.678(0.202)*	-3.838(0.112)*	0.043(0.221)*	-2.993(0.2131)*
`T3`	-6.316(0.359)*	-4.381(0.178)*	0.539(0.338)*	-3.062(0.311)*
`T4`	-5.867(0.360)*	-4.954(0.287)*	13.678(164.870)*	-3.305(0.439)*
`female`	-0.290(0.114)*	-0.047(0.104)*	-0.476(0.233)*	-0.226(0.244)*
`age`	0.043(0.003)*	0.046(0.003)*	-0.017(0.006)*	0.032(0.006)*
`penetrating`	-0.159(0.096)*	-0.980(0.159)*	-0.020(0.228)	-0.491(0.215)
`penetrating & blunt`	0.149(0.120)	0.191(0.110)*	0.342(0.246)*	-0.274(0.324)*
`injurytime`	-0.149(0.022)*	-0.068(0.020)	0.082(0.037)*	-0.069(0.046)*
`hr`	0.024(0.002)*	0.026(0.002)*	0.000(0.004)*	0.069(0.046)*
`rr`	-0.017(0.006)*	-0.024(0.007)*	-0.004(0.011)*	-0.001(0.015)*
`cc`	0.137(0.020)*	0.293(0.025)*	0.146(0.050)*	0.389(0.056)*

Incidence			
Coefficient	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
`Constant`	-0.025(0.022)*	-0.276(0.023)*	-4.296(0.090)*
`female`	0.011(0.040)	-0.279(0.041)*	-4.564(0.050)*
`age`	-0.009(0.001)*	-0.001(0.001)*	0.050(0.003)*
`penetrating`	-0.825(0.036)*	-0.980(0.039)*	0.141(0.137)
`penetrating & blunt`	0.174(0.048)	-0.106(0.049)*	0.339(0.144)
`injurytime`	-0.011(0.007)	-0.029(0.007)	0.034(0.018)
`hr`	-0.002(0.001)*	0.000(0.001)*	-0.024(0.003)*
`rr`	0.018(0.003)*	-0.015(0.003)*	-0.015(0.008)*
`cc`	0.171(0.011)*	-0.186(0.011)*	0.189(0.024)

Table 7.12: Maximum likelihood estimates for the Discrete Time Vertical Model (with standard errors) (* denotes $P < 0.05$).

Discrete Time Vertical Model					
Coefficient	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\beta}_C$	$\hat{\beta}_M$
T1	0.162(0.096)*	-1.361(0.139)*	-1.628(0.163)*	-2.647(0.042)*	-1.935(0.031)*
T2	1.620(0.235)*	0.606(0.260)*	0.166(0.300)*	-4.111(0.076)*	-3.317(0.058)*
T3	1.645(0.407)*	0.855(0.427)*	0.554(0.472)*	-4.537(0.117)*	-3.782(0.093)*
T4	0.486(0.468)*	0.302(0.466)*	-0.264(0.559)*	-4.771(0.162)*	-4.063(0.133)*
female	0.212(0.150)*	-0.164(0.208)	0.003(0.248)*	-0.203(0.068)*	-0.144(0.056)*
age	0.002(0.004)*	0.010(0.005)*	0.006(0.006)*	0.044(0.002)*	0.017(0.001)*
penetrating	-1.747(0.180)	0.228(0.199)	-0.350(0.224)	-0.408(0.066)	-0.505(0.052)
penetrating & blunt	0.134(0.154)	0.228(0.199)*	0654.(0.312)*	0.169(0.071)*	0.007(0.059)*
injurytime	0.089(0.030)*	0.114(0.035)	0.030(0.048)	-0.073(0.010)*	-0.013(0.010)*
hr	0.002(0.003)*	0.006(0.003)*	-0.001(0.004)*	0.024(0.001)*	0.011(0.001)*
rr	-0.021(0.007)	-0.011(0.009)*	-0.006(0.011)*	-0.020(0.004)*	-0.014(0.013)*
cc	-0.088(0.033)*	0.088(0.037)*	-0.059(0.054)*	0.184(0.013)*	0.158(0.011)*

Recall from the previous section that we compared the effect of *penetrating* and *blunt* injuries on death by *bleeding*. We continue with this analysis in this section, but bearing in mind that we have now collapsed death by *bleeding* and *vascular occlusion* into one failure cause. Examination of Figure 7.7 suggests that all four ordinary models compare favourably.

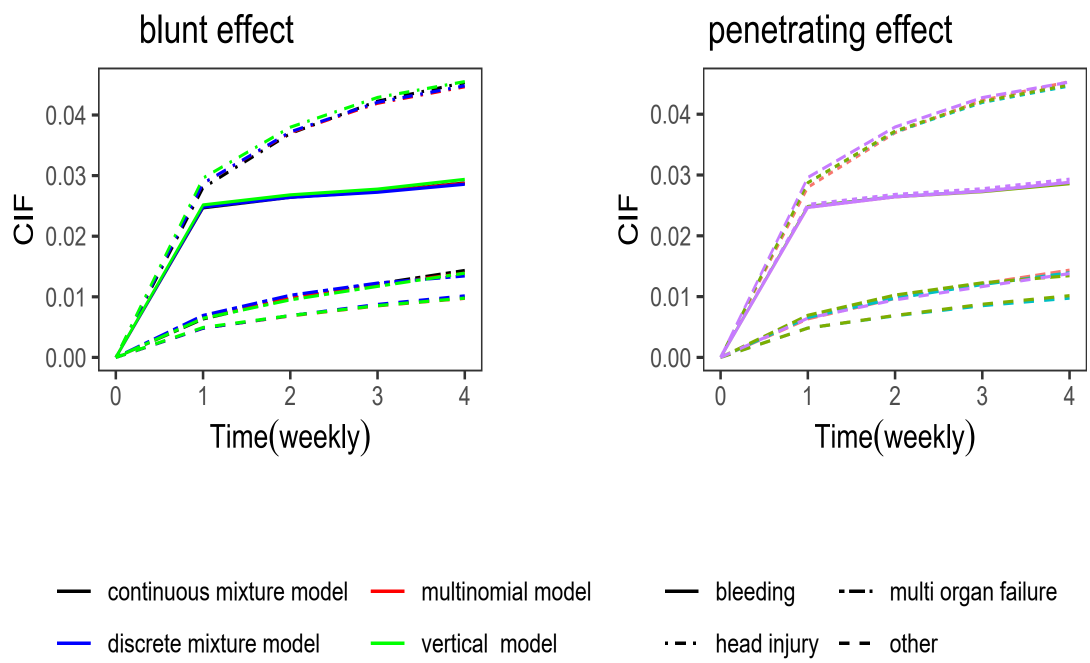

Figure 7.7: The CIF of death from bleeding, head injury, multi organ failure and other causes with the effect of injury type via the the Complete Case ordinary models.

The multinomial suggests that 2.87% of patients with *blunt* injury die from *bleeding* or *vascular occlusion* while 3.33% of the patients with *penetrating* injury die from the same causes. The *continuous time mixture model* suggests 2.88% and 3.45% of patients with *blunt* and *penetrating* injuries, respectively dying from *bleeding* or *vascular occlusion*. The *discrete time mixture model* suggests chances of 2.86% and 3.43%, respectively for patients with *blunt* and *penetrating* injuries dying from *bleeding* or *vascular occlusion*. Finally, the *vertical model* suggests slightly larger chances of 2.94% and 3.56 for patients with *blunt* and *penetrating* injuries to respectively die from *bleeding* or *vascular occlusion*. Moving to the missing failure causes models, we have fitted all three proposed models together with multinomial model to data including simulated missing failure causes, and the results are displayed in Table 7.13 to Table 7.15.

Table 7.13: Maximum likelihood estimates for the Missing Failure Causes Multinomial Model (with standard errors) (* denotes $P< 0.05$).

Multinomial Model

Coefficient	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$
T1	-2.925(0.050)*	-2.783(0.046)*	-4.235(0.084)*	-4.506(0.101)*
T2	-5.467(0.148)*	-3.892(0.082)*	-4.894(0.120)*	-5.325(0.161)*
T3	-6.547(0.335)*	-4.485(0.139)*	-5.282(0.219)*	-5.459(0.223)*
T4	-5.970(0.319)*	-5.027(0.227)*	-5.282(0.219)*	-6.119(0.385)*
female	-0.229(0.092)*	-0.005(0.081)	-0.314(0.141)*	-0.244(0.184)*
age	0.013(0.002)*	0.017(0.002)*	0.023(0.003)*	0.019(0.004)*
penetrating	0.020(0.073)	-1.788(0.126)	-0.164(0.121)	-0.283(0.155)
penetrating & blunt	0.084(0.096)	-0.058(0.088)*	0.261(0.134)*	-0.549(0.235)*
injurytime	-0.113(0.017)*	-0.013(0.014)	0.020(0.019)	-0.072(0.033)*
hr	-0.011(0.002)*	0.010(0.002)*	0.016(0.002)*	0.009(0.003)*
rr	0.027(0.004)	-0.006(0.005)*	-0.021(0.007)*	-0.023(0.009)*
cc	0.182(0.016)*	0.118(0.017)*	0.212(0.021)*	0.099(0.036)*

Table 7.14: Maximum likelihood estimates for the Missing Failure Causes Continuous Time Mixture Model (with standard errors) (* denotes $P< 0.05$).

Continuous Time Mixture Model

Latency				
Coefficient	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
`T1`	-1.945(0.047)*	-1.966(0.062)*	-2.100(0.088)*	-2.435(0.099)*
`T2`	-4.499(0.150)*	-2.940(0.081)*	-2.568(0.121)*	-3.157(0.156)*
`T3`	-5.741(0.283)*	-3.475(0.135)*	-2,858(0.179)*	-3.44(0.223)*
`T4`	-4.741(0.283)*	-4.106(0.233)*	-2.896(0.233)*	-3.745(0.342)*
`female`	-0.179(0.097)*	-0.051(0.079)*	-0.372(0.147)*	-0.156(0.179)*
`age`	0.011(0.002)*	0.017(0.002)*	0.017(0.003)*	0.025(0.004)*
`penetrating`	0.203(0.069)*	-1.408(0.122)*	-0.695(0.130)	-1.059(0.155)
`penetrating & blunt`	0.071(0.092)	0.331(0.085)*	0.423(0.135)*	-1.470(0.244)*
`injurytime`	-0.113(0.016)*	-0.014(0.014)	-0.017(0.020)*	-0.018(0.031)*
`hr`	0.001(0.002)*	0.012(0.002)*	0.015(0.003)*	0.003(0.003)*
`rr`	0.022(0.004)*	0.000(0.000)*	0.006(0.007)*	0.035(0.008)*
`cc`	0.037(0.015)*	0.225(0.580)*	0.234(0.022)*	0.246(0.035)*
Incidence				
Coefficient	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	
`Constant`	-0.553(0.023)*	-0.387(0.022)*	-2.230(0.036)*	
`female`	-0.023(0.043)	0.085(0.042)*	0.043(0.065)*	
`age`	-0.001(0.001)*	0.000(0.042)*	0.012(0.002)*	
`penetrating`	-0.166(0.036)*	-0,551(0.036)*	0.035(0.078)	
`penetrating & blunt`	-0.025(0.049)	-0.499(0.051)*	-0.035(0.078)	
`injurytime`	-0.008(0.007)	0.004(0.007)	0.041(0.010)	
`hr`	-0.001(0.001)*	0.001(0.001)*	-0.004(0.001)*	
`rr`	0.013(0.002)*	-0.012(0.002)*	-0.004(0.004)*	
`cc`	0.233(0.010)*	-0.181(0.011)*	0.027(0.014)	

Table 7.15: Maximum likelihood estimates for the Missing Failure Causes Discrete Time Mixture Model (with standard errors) (* denotes $P < 0.05$).

Continuous Time Mixture Model

Latency				
Coefficient	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
`T1`	-2.257(0.051)*	-1.903(0.050)*	-0.502(0.104)*	-1.479(0.011)*
`T2`	-5.047(0.162)*	-3.054(0.088)*	-0.391(0.140)*	-2.156(0.166)*
`T3`	-5.831(0.315)*	-3.730(0.153)*	0.076(0.216)*	-2.349(0.243)*
`T4`	-5.485(0.3343)*	-4.430(0.267)*	3.003(0.809)*	-2.851(0.391)*
`female`	-0.121(0.093)*	-0.113(0.088)*	-0.533(0.179)*	-0.444(0.193)*
`age`	0.014(0.002)*	0.021(0.002)*	0.003(0.004)*	0.125(0.005)*
`penetrating`	0.317(0.074)*	-0.917(0.129)*	-0.220(0.160)	-0.424(0.167)
`penetrating & blunt`	0.027(0.100)	0.027(0.094)*	0.368(0.185)*	0.074(0.269)*
`injurytime`	-0.109(0.017)*	-0.039(0.016)	-0.057(0.026)*	0.048(0.035)*
`hr`	0.011(0.002)*	0.014(0.002)*	0.002(0.003)*	0.004(0.003)*
`rr`	0.023(0.004)*	0.000(0.005)*	0.003(0.008)*	0.054(0.010)*
`cc`	0.128(0.016)*	0.240(0.021)*	0.223(0.039)*	0.342(0.046)*

Incidence			
Coefficient	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
`Constant`	0.013(0.022)*	-0.422(0.022)*	-3.288(0.057)*
`female`	-0.151(0.041)	0.135(0.043)*	0.129(0.108)*
`age`	-0.005(0.001)*	0.001(0.001)*	0.015(0.003)*
`penetrating`	1.002(0.036)*	-1.144(0.041)*	0.218(0.095)
`penetrating & blunt`	0.172(0.047)	-0.076(0.048)*	0.142(0.112)
`injurytime`	-0.023(0.007)	0.030(0.007)	0.045(0.013)
`hr`	-0.014(0.002)*	0.018(0.003)*	0.010(0.002)*
`rr`	0.014(0.002)*	-0.018(0.003)*	0.006(0.006)*
`cc`	0.132(0.010)*	-0.152(0.011)*	0.131(0.019)

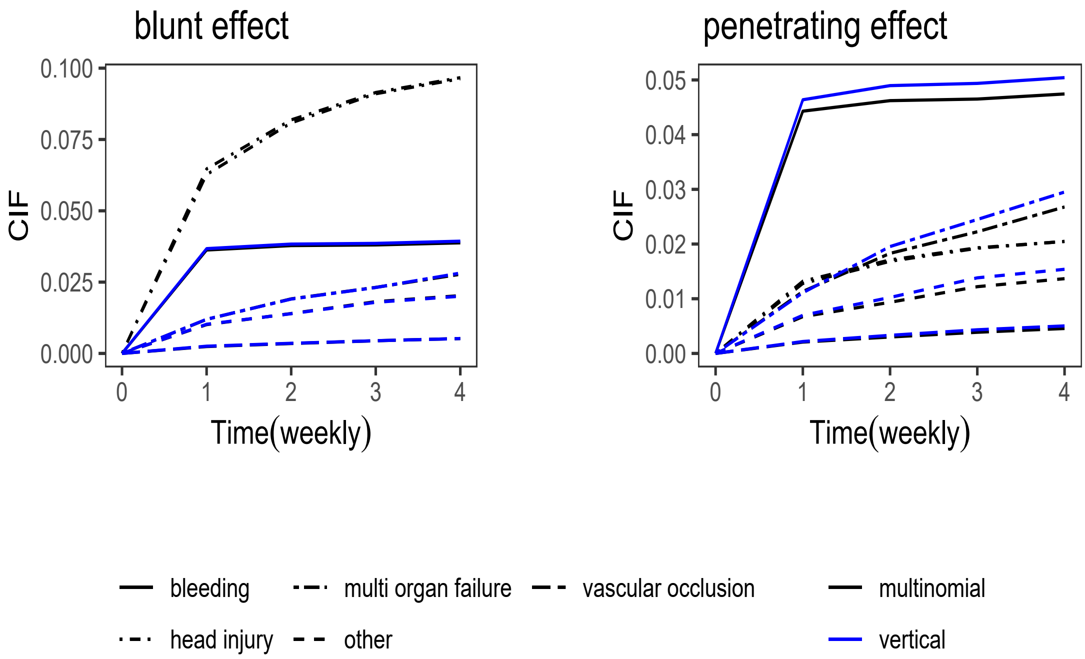

Figure 7.8: The CIF of death from bleeding, head injury, multi organ failure and other causes with the effect of injury type via the Missing Failure Causes models.

We now consider these models when data includes subjects with missing failure causes. The *continuous time mixture model* suggests that about 5.6% of the patients with a *blunt* type of injury die from *bleeding or vascular occlusion* and this figure increases to about 6.7% if a patient has a *penetrating* type of injury. The *discrete time mixture model* suggests that the probability of dying from *bleeding or vascular occlusion* if you have a *blunt* type of injury is about 5.4% and these chances increase to about 5.9%, while, the *discrete time vertical* suggests a 5.6% probability of dying from *bleeding or vascular occlusion* if you have a *blunt* type of injury and about 6.3% if you have a *penetrating type* of injury. We need to contrast all these figures to the missing failure causes multinomial model which suggests 5.3% as the chances of dying from the same death when you have a *blunt* type of injury as compared to about 5.8% if have experienced a *penetrating* type of injury. Evidently, all three proposed models have produced different figures with regard to the chances of dying from *bleeding or vascular occlusion* when a patient has *blunt* compared to a patient that has been admitted with a *penetrating* type of injury, but of critical importance is that they all agree in terms of the direction of the effect in relation to injury type even though they may differ in terms of the size of effect. It is also important to note that these figures need not necessarily compare to the figures that we obtained when we fitted these models to the entire complete data in the previous section apart from the fact that we collapsed *bleedind* and *vascular occlusion*. All four models split or share the subjects with missing failure causes, but differ in terms how this split is conducted. The mixture models employ an EM Algorithm while both the multinomial model and the vertical rely on the model for relative hazards to share these subjects. Eitherway, the splitting mechanism depends on the distribution of failure causes for the remaining subjects whose failure causes are known. For the multinomial model and the vertical model the sharing mechanism depends entirely and exclusively on the remaining subjects that do not have missing failure causes. When the missing failure causes are generated it upsets this distribution of failure causes such that the remaining subjects with known failure causes follow a different distribution with which they now split the created missing failure causes. Put diffrently, the distribution of failure causes before and after the missing failure causes are simulated are not necessarily the same.

7.3 Other Validation Methods

While competing risks methods have their origins in medical research, these methods are increasingly being applied in other field such as in economics, graduation and dropout studies in education, marriage disolution studies, etc. A case in point in the econometric field is the unemployment data that we have used extensively in this thesis to demonstrate the application of the models that have been proposed. Invariably, in almost all these settings these methods are used for the development of prediction models. In educational settings, for example, having developed a model, one can isolate or identify the student characteristics that are associated with a high or a low risk of dropping out. In fitting the proposed models to *crash2* data in Section 6.1, we were able to establish that a patient with a *penetrating* injury has a high risk of dying by *bleeding* compared to a patient who has a *blunt* injury and a smaller risk of dying by other types of death compared to a patient who has a *blunt* injury. Thus, the treatment for a given patient that is admitted with a *penetrating* type of injury should focus more on bleeding than on any other type of death. However, just like other statistical models, before these models are put to use their predictive accuracy must be assessed. The standard methods such as R^2 are not suitable for time to event data which comes with censored data. One critical issue that requires attention in the assessment of competing risks models in relation to predictive accuracy is the *discriminative* ability of the model. This is the ability of the model to distinguish between subjects with (cases) and without the outcome of interest (control). Discrimination is a measure that has its history in the assessment of classification models for binary data in cross-sectional settings. It is computed as the area under the curve (AUC) of the Reciever Operator Curve(ROC). If the model in these settings generally assigns a larger probability of the outcome interest to the subjects with the actual outcome compared to subjects without the outcome, the model is said to have a good *discriminative* ability. The idea that some subjects (cases) will experience an outcome of interest at a particular event time while others will experience it later or never at all (censored subjects) has allowed authors such as Zheng (2005) to extend *discrimination* to survival analysis. This metric is commonly referred to as the dynamic AUC because it is computed as a weighted average of the time dependent AUC_t (the AUC's that are computed at each failure time). Saha and Heagerty (2010) have extended this metric to competing risks settings and Schmid et al. (2018) are credited with the introduction of the discrete time version of this metric for data with a single mode of failure, Heyard et al. (2019) have extended this work to competing risks settings in the discrete time realm. Recall that $h_j(t)$ represents the risks of failure due to failure cause j at time t or the CSH at failure time t due to failure cause j. Heyard et al. (2019) define the discrete time version of the time dependent AUC at failure time t due to failure cause j as;

$$\text{AUC}_j(t) = P(\hat{h}_j(t|\boldsymbol{x}_i) > \hat{h}_j(t|\boldsymbol{x}_k)|d_{ijt} = 1, d_{kjt} = 0, T_i \geq t, T_k \geq t)$$

recalling that $d_{ijs} = 1$ for $s = t$ and 0 for $s < t$. If at time t we consider all subjects that failed due to cause j and pair them with other subjects that failed by other causes also at time t, then AUC_t is the proportion of concordant pairs, that is pairs for whom the predicted conditional probability of failing by cause j for a subject that *actually did* fail by cause j ($\hat{h}_{ij}(t|\boldsymbol{x}_i)$ and $d_{ijt} = 1$), is larger than the predicted conditional probability of failing by cause j for a subject that *did not* fail by cause j ($\hat{h}_{kj}(t|\boldsymbol{x})$ and $d_{kjt} = 0$). Following Li et al. (2018), the estimate for $\text{AUC}_j(t)$ that has been proposed by Heyard et al. (2019) is given by

$$\hat{\text{AUC}}_j(t) = \frac{\sum_{i=1}^{n}\sum_{k=1}^{n}\left[I(h_j(t|\boldsymbol{x}_i) > h_j(t|\boldsymbol{x}_k) + 0.5(I(h_j(t|\boldsymbol{x}_i) = h_j(t|\boldsymbol{x}_k)))\right]d_{ijt}(1-d_{kjt})}{\sum_{i=1}^{n}\sum_{k=1}^{n} d_{ijt}(1-d_{kjt})}$$

where $0.5(I(h_j(t|\boldsymbol{x}_i) = h_j(t|\boldsymbol{x}_k))$, accounts for ties which will be common in discrete time setting. Note that this estimates only takes into account the subjects that failed at time t, the censored subjects are excluded. To determine the final estimate for the cause-specific AUC_j or cause j concordence index c_j in the presence of censoring the $\text{AUC}_j(t)'s$ are weighted with $w_j(t)'s$, where $w_j(t)'s$ are inverse probability of censoring weights (IPCW) as proposed by Blanche et al. (2015) and Schmid et al. (2018);

$$\hat{c}_j = \hat{\text{AUC}}_j = \sum_{s=1}^{q} \hat{\text{AUC}}_j(t)\hat{w}_j(t)$$

where;

$$\hat{w}_j(t) = P(T = t, D = j).P(T > t)/\sum_{s=1}^{q} P(T = s, D = j).P(T > s)$$

The overall concordance index (c-index) is given by

$$\hat{c} = \sum_{j=1}^{J} \hat{c}_j e_j/e \qquad (7.3.1)$$

where e_j is the total number cause j events and $e = \sum_{j=1}^{J} e_j$. The c-index allows for comparison of various models in terms their discriminative ability. As with standard AUC in cross-sectional settings, the closer is the c-index of a given model to 1, the better is the discriminative ability of the model. Another metric that is also considered when the predictive accuracy of a given time to event model is assessed is *calibration.* This metric quantifies the extent of agreement between observations (y) and predictions ($\hat{y}$). To compute this metric, Beger and Schmid (2018) suggest partitioning data in a long format into 10 to 20 subgroups according to CSH percentiles. In each subgroup a count of $d_{ijs} = 1$ relative to total number d_{ijs} is computed which is then compared to the average $\hat{h}_j(s)$ in that subgroup. These pairs are then plotted and the closer these pairs are to $y = x$ line, the better calibrated is the model.

The next metric that we discuss is referred to as the *prediction error* (PE). It is the difference between the observed event status ($\Delta = j$) and the predicted probability of the event or the CIF ($\hat{F}_j(t)$) (Graf and Schumacher, 2008; Schoop et al., 2011a,b). These authors have adapted the Brier score (Brier, 1950) for application in time to event settings. Due to censoring, they also demonstrated how the metric can be estimated also using (IPCW) in continuous time. The discrete time version of this metric at time t is given by

$$\mathrm{PE}_j(t) = \frac{1}{n}\sum_{i=1}^{n}\left[I(T_i \leq t, \Delta_i = j) - F_j(t))\right]^2$$

The cause j *prediction error* (PE_j) is the sum of weighted $\mathrm{PE}_j(t)'s$;

$$\mathrm{PE}_j = \sum_{s=1}^{q} \mathrm{PE}_j(t) P(T = s, \Delta = j) \tag{7.3.2}$$

This metric accounts for both *discrimination* and *calibration* simultaneously. Suppose that PE_j was small. That will happen if $\mathrm{PE}_j(t)'s$ were also small, that is, if the model assigns larger values of $\hat{F}_j(t)$ at time t to subjects that actually failed due to cause j than to subjects that failed by other failure causes. Suppose $\hat{F}_j(t)$ was closer to 1 for subjects that actually failed by cause j and closer to 0 for those that did not fail by this cause, then $(I(T_i \leq t, \Delta_i = j) - F_j(t)))$, will be smaller for subjects that failed by cause j as $I(T_i \leq t, \Delta_i = j) = 1$, for these subjects, and $(I(T_i \leq t, \Delta_i = j) - F_j(t)))$, will be smaller for the other subjects as well because $I(T_i \leq t, \Delta_i = j) = 0$, for these subjects and $F_j(t)$ is closer to zero for them. It is for this reason that this metric is viewed as an alternative measure of AUC at marginal level. We considered CSHs to compute $AUC_j(t)$, the conditional probability of failure by cause J, but here these quantities are replaces by $F_j(t)$, also the probability of failure by the same cause albeit at marginal level. It is easy to see that this measure is also a proxy for *calibration* because $(I(T_i \leq t, \Delta_i = j) - F_j(t)))$, is also difference between observed failure cause and the marginal probability of failing by that failure cause.

It is imperative that both *discrimination* and *calibration* are considered to assess the predictive accuracy of a given time to event model because one metric does not necessarily imply the other. The procedure is to split available data into training and test data, where the models are developed from training data and then assessed on test data for predictive accuracy in the absence of independent and similar data set. Amongst other things, *crash2* data was collected for development of models for prognostic (future) predictions, that is to say, if a patient is admitted with these symptoms can the doctors be in a position to prescribe an appropriate treatment because the patient is at a higher risk of dying from, say *bleeding* as opposed to any other cause of death. Having developed a given model on given set of patients (training data), now given a new set of patients (test data) can we rely on our model for predictive purposes. It is largely for these reasons that a model is developed from one set of data and then tested for predictive accuracy on another independent, but similar set of data (Steyerberg, 2009).

Typically, available data is split (60%-75%):(40%-25%) as training data:test data. When we generated a set of subjects with missing failure causes in the previous section our latitude to create any number of these subjects was limited by the fact that there were fewer failures towards the close of the observation across failure causes. Recall that it was for this reason that we collapsed *bleedind* and *vascular occlusion* and even thereafter we were compelled to settle for a 5% set of subjects with missing failure causes especially for the stability of mixture models. If we follow the 75%: 25% rule we will not be able to assess any of the mixture models. Faced with the option of strictly following the split rules and thereby exclude the mixture models altogether or break with these rules and include one of the mixture models we decided to go for a 95%:5% split which allowed us to include the multinomial model, the *vertical model*, and only one of the mixture models i.e., the *discrete time mixture model*. Even if we were able to include the *continuous time mixture model* we would not be able to conduct an assessment on this model because we do not as yet possess the requisite knowledge to apply this model for prediction purpose.

We have relied heavily on the `discSurv` R Package (Welchowski and Schmid, 2019) which has routines for computing the *c-index* and to conduct *calibration*. For computing the indexes the routine requires the *risk scores* for test data that are produced by the multinomial model. Recall that the model for CSHs is given by

$$h_j(t|\boldsymbol{x};\boldsymbol{\alpha}) = \frac{\exp\left(\eta_{jt}\right)}{1+\sum_{l=1}^{J}\exp\left(\eta_{jt}\right)}$$

where $\eta_{jt} = \alpha_{0jt} + \boldsymbol{x}^T\boldsymbol{\alpha}_{1j}$, is what is referred as the risk score for a subject that is described by vector $\boldsymbol{x}$. Recall that when the multinomial is fitted to data a reference category is chosen, typically the $(J+1)^{\text{th}}$ none event category, such that;

$$\log\left[\frac{h_1(t)}{h_{J+1}(t)}\right] = \eta_{1t}, \log\left[\frac{h_2(t)}{h_{J+1}(t)}\right] = \eta_{2t}\ldots, \log\left[\frac{h_J(t)}{h_{J+1}(t)}\right] = \eta_{Jt} \tag{7.3.3}$$

The `VGAM` R package (Yee, 2010) can produce the risks scores η_{jt} for all subjects in the test data which is what is required by the `disSurv` package to compute the indexes. Most of the standard packages are only able to produce the risks only, but we can use (7.3.3) to work backwards to the risks scores because $h(\eta_{(J+1)t}) = 1 - \sum_{j=1}^{J} h_j(\eta_{jt})$. For example, for subject i that failed at time t_i the risk score of failing by failure cause j can be computed from;

$$\log\left[\frac{h_j(t_i)}{h(t_i)_{J+1}}\right] = \eta_{jt_i}$$

Things are rather straightforward when data is modelled with the multinomial, but not as such when we wish to assess the proposed *vertical model* and the *discrete time mixture model*. When data is fitted with the proposed *vertical model*, total hazards and relative hazards are produced. With most packages we can predict these quantities for each subject in the test data and easily recover the risks. Recall that, with covariates, the model for relative hazards is given by;

$$\Pi_j(t|\boldsymbol{x}) = \frac{\exp\left(\phi_{0jt} + \boldsymbol{x}^T\boldsymbol{\phi}_{1j}\right)}{1+\sum_{l=1}^{J\text{-}1}\left(\exp\phi_{0lt} + \boldsymbol{x}^T\boldsymbol{\phi}_{1l}\right)} \tag{7.3.4}$$

and we specified the following regression model for total hazards;

$$h(t|\boldsymbol{x}) = \frac{\beta_{0t} + \boldsymbol{x}^T\boldsymbol{\beta}_1}{1+\beta_{0t} + \boldsymbol{x}^T\boldsymbol{\beta}_1} \tag{7.3.5}$$

Obviously, we can retrieve the risks of failing by any cause for each subject i in the test data from;

$$h_j(t|\boldsymbol{x}_i) = \Pi_j(t|\boldsymbol{x}_i)h(t|\boldsymbol{x}_i) \tag{7.3.6}$$

Here again, the risk of none event is; $h(t|\boldsymbol{x})_{J+1} = 1 - \sum_{j=1}^{J} h_j(t|\boldsymbol{x})$. We can also exploit (7.3.3) to determine the risks score of each failure cause for all subjects in the test data. When data is modelled with the *discrete time mixture model* one can compute the CIFs of failling by any cause for all subjects in the test data. We can use the following relationship to retrive the risks;

$$h_j(t|\boldsymbol{x}) = \frac{F_j(t|\boldsymbol{x}) - F_j(t\text{-}1|\boldsymbol{x})}{1 - \sum_{j=1}^{J} F_j(t\text{-}1|\boldsymbol{x})}$$

Naturally, we can also use (7.3.3) to determine the risk scores. For conducting *calibration*, test data is expanded into standard long format and then the risks are computed for entry which is adequate for the purpose of caluculating these quantities.

Table 7.16: Maximum likelihood estimates for the Multinomial Model fitted to Training data(with standard errors) (* denotes $P < 0.05$).

Multinomial Model

Coefficient	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$
`T1`	-3.182(0.057)*	-2.653(0.044)*	-5.924(0.202)*	-4.319(0.089)*	-4.517(0.106)*
`T2`	-6.286(0.214)*	-3.815(0.082)*	-6.619(0.293)*	-4.771(0.0118)*	-5.466(0.178)*
`T3`	-8.828(1.007)*	-4.431(0.141)*	-6.963(0.431)*	-5.305(0.185)*	-5.312(0.216)*
`T4`	6.969(0.504)*	-5.046(0.238)*	-6.899(0.521)*	-5.173(0.215)*	-6.222(0.415)*
`female`	-0.089(0.099)*	-0.134(0.082)	-0.336(0.334)*	-0.238(0.143)*	-0.312(0.198)*
`age`	0.106(0.002)*	-1.632(0.002)*	0.308(0.007)*	0.025(0.003)*	0.020(0.004)*
`penetrating`	0.106(0.082)	-1.632(0.113)	-0.227(0.303)	-0.155(0.127)	0.382(0.167)*
`penetrating & blunt`	0.158(0.107)	-0.072(0.087)*	0.309(0.331)*	0.230(0.142)*	-0.684(0.259)*
`injurytime`	-0.129(0.019)*	-0.024(0.014)	-0.020(0.044)	0.021(0.019)*	-0.032(0.033)*
`hr`	0.015(0.002)*	0.008(0.002)*	0.002(0.006)*	0.014(0.003)*	0.011(0.003)*
`rr`	0.031(0.005)	0.001(0.005)*	0.022(0.016)*	0.015(0.007)*	0.021(0.010)*
`cc`	0.193(0.017)*	0.120(0.017)*	0.077(0.069)*	0.207(0.022)*	0.116(0.036)*

Table 7.17: Maximum likelihood estimates for the Discrete Time Mixture Model (with standard errors) (* denotes $P < 0.05$).

Continuous Time Mixture Model

Latency					
Coefficient	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
T1	-2.352(0.059)*	-1.878(0.046)*	-0.161(0.310)*	-0.536(0.115)*	-1.677(0.116)*
T2	-5.503(0.214)*	-3.045(0.084)*	-0.231(0.432)*	-0.396(0.152)*	-2.574(0.189)*
T3	-8.060(1.001)*	-3.659(0.143)*	0.421(0.666)*	-0.477(0.240)*	-2.312(0.229)*
T4	-6.229(0.503)*	-4.279(0.240)*	15.977(0.656)*	1.064(0.410)*	-3.185(0.426)*
female	-0.098(0.102)*	-0.137(0.086)*	0.066(0.517)*	-0.466(0.184)*	-0.223(0.216)*
age	0.014(0.003)*	0.018(0.002)*	-0.007(0.011)*	-0.001(0.004)	0.015(0.005)*
penetrate	-0.308(0.084)*	-0.743(0.117)*	1.032(0.528)	-0.081(0.166)	-0.978(0.178)*
penetrate and blunt	0.059(0.110)	-0.024(0.091)*	0.158(0.522)*	0.403(0.203)*	0.084(0.285)*
injurytime	-0.118(0.020)*	-0.042(0.015	-0.120(0.081)	-0.042(0.025)	-0.006(0.036)
hr	0.015(0.002)*	0.009(0.002)*	0.015(0.009)*	-0.001(0.003)*	0.013(0.004)
rr	0.026(0.005)*	0.008(0.005)*	0.009(0.022)	0.008(0.008)*	0.033(0.011)
cc	0.128(0.018)*	0.218(0.019)*	-0.071(0.107)*	0.278(0.043)*	0.368(0.048)*

Table 7.18: Maximum likelihood estimates for the Discrete Time Mixture Model fitted to Training Data(with standard errors) (* denotes $P < 0.05$).

Discrete Time Mixture Model

Incidence

Coefficient	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$
`Constant`	-0.349(0.022)*	-0.082(0.022)*	-5.289(0.153)*	-3.459(0.063)*
`female`	-0.303(0.040)*	0.082(0.040)*	-5.607(0.283)*	-3.504(0.108)*
`age`	-0.003(0.001)	-0.003(0.001)*	0.036(0.006)*	0.023(0.003)*
`penetrating`	1.050(0.036)*	-1.358(0.040)*	-0.409(0.269)*	-0.063(0.099)
`penetrating & blunt`	0.203(0.048)*	-0.099(0.048)*	0.248(0.289)*	0.103(0.124)
`injurytime`	-0.031(0.007)*	0.027(0.007)	0.045(0.028)*	0.042(0.014)*
`hr`	0.001(0.001)	-0.002(0.001)*	0.003(0.005)*	0.012(0.002)*
`rr`	0.013(0.002)	-0.015(0.003)*	0.016(0.014)*	0.010(0.006)
`cc`	0.133(0.010)*	-0.127(0.011)*	0.071(0.056)*	0.117(0.021)*

Table 7.19: Maximum likelihood estimates for the Discrete Time Vertical Model fitted to Training data(with standard errors) (* denotes $P< 0.05$).

	Vertical Model				
Coefficient	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$\hat{\beta}$
T1	-1.330(0.121)*	1.896(0.114)*	-1.366(0.225)*	0.190(0.138)*	-1.939(0.032)*
T2	-0.754(0.279)*	1.643(0.196)*	-1.143(0.341)*	0.723(0.213)*	-3.308(0.059)*
T3	-3.408(1.025)*	0.869(0.261)*	-1.708(0.486)*	0.053(0.286)*	-3.815(0.096)*
T4	-0.707(0.656)*	-1.185(0.483)*	-0.687(0.669)*	1.135(0.470)*	-4.089(0.137)*
female	0.258(0.224)*	0.188(0.214)	-0.033(0.387)*	0.032(0.244)*	-0.156(0.057)*
age	-0.004(0.005)*	-0.004(0.005)*	0.016(0.008)*	0.004(0.005)*	0.018(0.001)*
penetrating	0.512(0.188)	-1.267(0.201)	0.128(0.345)	0.174(0.209)	-0.535(0.054)*
penetrating & blunt	0.843(0.281)	0.600(0.272)*	0.959(0.419)*	0.906(0.295)*	0.009(0.061)*
injurytime	-0.085(0.039)*	0.013(0.036)	0.045(0.056)	0.042(0.039)*	-0.043(0.010)*
hr	0.002(0.003)*	-0.003(0.003)*	-0.006(0.006)*	0.002(0.004)*	0.011(0.001)*
rr	0.009(0.009)	-0.016(0.009)*	0.001(0.016)*	0.001(0.010)*	0.016(0.003)*
cc	0.067(0.045)*	0.011(0.045)*	0.098(0.086)*	-0.045(0.048)*	0.158(0.011)*

For a subject described by vector $\boldsymbol{x}_i$ that failed at time t_i, for example, $h_j(t|\boldsymbol{x}_i)\,(j = 1, 2\ldots, J)$, will be computed for times $1, 2, \ldots, t_i$. The *root mean square error*(RMSE) is calculated for each failure cause to determine how close are predicted risks $\hat{h}_{ijs}$ to d_{ijs} for each failure cause.

We have listed the results of fitting the multinomial model, the proposed *discrete time mixture model* and the *discrete time vertical model* to training data in Table 7.17, 7.18 and 7.19. We found the c-indexes to be 0.934, 0.916 and 0.928 for the multinomial model, the proposed *mixture model* and the *vertical model*, respectively. Clearly, these figures are very comparable and they indicate a high degree of *discrimination* for these models. If there are concerns that the multinomial model and the vertical model seem to much closer to each other in terms of indexes, i.e., the index for the mixture model seems to be a bit out, or if there is a wish to improve all these indexes, simply, we would have to re-fit the models with the view to improve their fit by, for example, considering methods such as variable selection methods (Friedman et al., 2010).

We have listed the *calibration* results in Table 7.20. Upon examination of these results it is clear that all three models produced essentially comparable results for all five failure causes. In particular, the errors were identical for death by *vascular occlusion*, while the mixture model edged the other two models with respect to *bleeding*, *multi organ failure* and *other* causes of death. The multinomial was ahead with regard to *head injury*. All in all, though, these figures suggest that all three models compare favourably with one another with regards to *calibration*. In general, these figures are relatively small which suggest that the differences between the predicted CSHs, $\hat{h}_{ijs}(t)$ and observed d_{ijs} are not significant.

We then moved to *prediction errors*. Here, we concentrated on the time dependent prediction errors $\mathrm{PE}_j(t)$ for each failure cause because the (IPCW) weights will be same for each model as they are all applied to the same data.

Table 7.20: The *calibration* figures for the Multinomial Model, the Discrete Time Mixture Model and the Discrete Time Vertical Model .

	RMSE		
Death Type	Multinomial Model	Mixture Model	Vertical Model
`bleeding`	0.0150	0.0147	0.0172
`head injury`	0.0130	0.0118	0.0108
`vascular occlusion`	0.0049	0.0049	0.0049
`multi organ failure`	0.0070	0.0054	0.0065
`other`	0.0043	0.0033	0.0043

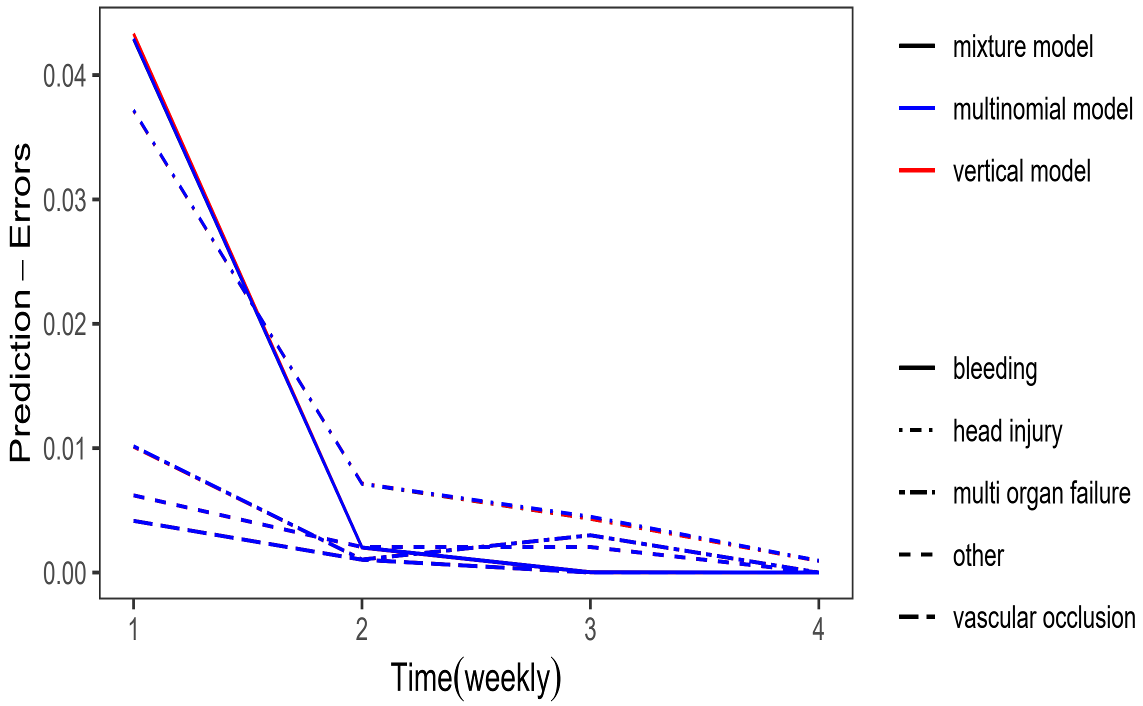

Figure 7.9: Prediction Errors.

If there are any differences in cause-specific *prediction errors* PE_j it will due to these quantities;

$$\mathrm{PE}_j(t) = \frac{1}{n}\sum_{i=1}^{n}\left[I(T_i \le t, \Delta_i = j) - F_j(t))\right]^2 \tag{7.3.7}$$

We see from examining Figure 7.9 that all three models produce almost identical time dependent *prediction errors*. Evidently, this ties up with what we saw when we examined the RMSE's for *calibration*. Here, the largest errors are associated with *bleeding* and *head injury*, precisely what we saw when we examined Table 7.20, the larger errors there are also associated with *bleeding* and *head injury* with the former exhibiting larger errors by all models. Clearly, if we wished to improve our models, one area to focus on is *bleeding* and *head injury* failure causes. Perhaps this is where the mixture model has an advantage over the multinomial model and the vertical model. All failure causes are modeled on the same set of covariates via the CSHs for the multinomial model and through the relative and total hazards for the vertical, while, each failure cause is modeled individually through the component hazards when the mixture model is assumed. This means we can propose different models only for *bleeding* and *head injury* failure causes and leave the models for other failure causes intact if we wished to improve the overall fit of the mixture model. Another model that has the same advantage as the mixture model in this respect is the binomial model (Lee et al., 2018) because each failure cause is also modeled separately for this model.

When we demonstrated the application of the proposed ordinary competing risks models in Chapters 3, 4, and 5 we missed out on the opportunity to test these models in terms of their predictive accuracy with regard to unemployment data. We are of the opinion that it might also be instructive to also conduct that exercise. Recall that the unemployment data came with 574 subjects that had missing failure causes. In testing the proposed models using this data we will include these subjects and create an extra failure cause for them. This is one option to deal with data that has missing failure causes and it is referred to as the *extra failure cause* approach. Of course, the estimates that were obtained excluding these subjects will now be biased downwards, but that is not a real concern to us.

Table 7.21: Maximum likelihood estimates for the Discrete Time Mixture Model and the Multinomial Model fitted to Training Data(with standard errors) (* denotes $P< 0.05$).

	Discrete Mixture Model (Latency Model)		Multinomial Model			
Coefficient	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$
T1	-1.115(0.072)*	-1.587(0.142)*	-1.591(0.116)*	-1.737(0.078)*	-3.029(0.143)*	-2.741(0.119)*
T2	-1.368(0.094)*	-1.587(0.155)*	-1.350(0.118)*	-2.027(0.099)*	-2.980(0.158)*	-2.506(0.122)*
T3	-1.427(0.110)*	-2.062(0.211)*	-1.243(0.130)*	-2.088(0.115)*	-3.405(0.242)*	-2.457(0.122)*
T4	-1.987(0.158)*	-2.270(0.253)*	-2.158(0.223)*	-2.689(0.162)*	-3.606(0.256)*	-3.453(0.224)*
T5	-1.290(0.127)*	-1.6801(0.210)*	-1.314(0.168)*	-1.963(0.132)*	-2.959(0.215)*	-2.581(0.171)*
T6	-2.030(0.198)*	-2.949(0.416)*	-2.101(0.267)*	-2.779(0.201)*	-4.264(0.418)*	-3.437(0.269)*
T7	-0.903(0.131)*	-2.122(0.299)*	-1.207(0.197)*	-1.652(0.139)*	-3.331(0.303)*	-2.492(0.199)*
T8	-2.192(0.272)*	-2.488(0.387)*	-2.183(0.340)*	-3.061(0.275)*	-3.727(0.389)*	-3.545(0.343)*
T9	-1.554(0.216)*	-2.944(0.507)*	-1.593(0.278)*	-2.419(0.219)*	-4.145(0.509)*	-2.959(0.280)*
T10	-3.535(0.579)*	-3.614(0.711)*	-2.557(0.453)*	-4.361(0.581)*	-4.745(0.714)*	-3.887(0.455)*
T11	-1.365(0.215)*	-2.615(0.455)*	-2.267(0.415)*	-2.166(0.221)*	-3.682(0.458)*	-3.559(0.418)*
T12	-3.201(0.580)*	-3.387(0.713)*	-2.115(0.415)*	-4.070(0.582)*	-4.467(0.714)*	-3.422(0.418)*
T13	-1.069(0.221)*	-2.191(0.416)*	-1.849(0.385)*	-1.907(0.227)*	-3.202(0.422)*	-3.099(0.389)*
T14	-0.706(0.217)*	-3.155(0.714)*	-1.648(0.386)*	-1.907(0.225)*	-3.202(0.715)*	-2.878(0.390)*
T15	-1.059(0.298)*	-3.033(0.711)*	-2.747(0.714)*	-2.095(0.391)*	-3.881(0.715)*	-3.928(0.715)*
T16	-1.367(0.386)*	-2.478(0.584)*	-1.035(0.343)*	-2.431(0.385)*	-3.273(0.588)*	-2.202(0.340)*
T17	-1.261(0.411)*	-3.345(1.006)*	-1.839(0.587)*	-2.358(0.423)*	-4.150(1.008)*	-3.066(0.590)*
T18	-0.992(0.410)*	-2.516(0.716)*	-2.824(1.006)*	-2.185(0.425)*	-3.296(0.720)*	-4.000(1.008)*
T19	0.601(0.272)*	-0.942(0.368)*	-0.489(0.363)*	-0.631(0.278)*	-1.404(0.387)*	-1.404(0.383)*

Table 7.22: Maximum likelihood estimates for the Discrete Time Mixture Model fitted to Training Data (with standard errors) (* denotes $P < 0.05$).

(Incidence Model)

Coefficient	$\hat{\gamma}_1$	$\hat{\gamma}_2$
Constant	0.070(0.064)*	-1.405(0.080)*
`ui`	-0.211(0.088)*	0.274(0.105)*
`dr`	-1.894(0.635)*	-0.198(0.735)*
`rr`	1.289(0.581)*	-0.202(0.658)*
`age`	-0.006(0.004)*	0.015(0.005)*
`wage`	0.614(0.125)*	-0.422(0.142)*
`tenure`	0.010(0.008)*	0.009(0.009)*

Of paramount importance to us is that approaching our analysis in this fashion will allow us an oportunity to compare the proposed models in terms of their predictive accuracy where there are three causes of failure instead of two.

Arguably, the advantage of unemployment data over *crash2* data vis-a-vis the assessment of predictive accuracy is that the failure times are not as sparsely populated. This means that there is more latitude in terms of how we split data into training and test data without having to concern ourselves as to whether we will be able to fit the mixture model to training data as the case was with *crash2*, in fact we will proceed to split the data following the 75%:25% (training:test) split rule. In fitting the multinomial model and the two proposed models, i.e., the *vertical model* and the *discrete time mixture model* we continue to include all the variables that came with the unemployment data where, now, the benefit non-recipients are regarded as the reference category and all the continuous covariates are centered at their respective averages. We also continue to regard full-time and part-time employment as cause 1 and cause 2, respectively, while the subjects with missing failure causes are regarded as, say, casual employment (cause 3). The results of fitting these models to training data are displayed Table 7.21 through to Table 7.23.

Table 7.23: Maximum likelihood estimates for the Discrete Vertical Model fitted to Training Data (with standard errors) (* denotes $P< 0.05$).

	Discrete Time Vertical Model		
Coefficient	$\hat{\boldsymbol{\phi}}_1$	$\hat{\boldsymbol{\phi}}_2$	$\hat{\boldsymbol{\beta}}$
T1	0.934(0.132)*	-0.370(0.174)*	-1.220(0.063)*
T2	0.447(0.151)*	-0.511(0.192)*	-1.301(0.073)*
T3	0.333(0.175)*	-0.972(0.251)*	-1.380(0.084)*
T4	0.627(0.282)*	-0.198(0.345)*	-2.041(0.120)*
T5	0.639(0.223)*	-0.398(0.286)*	-1.289(0.098)*
T6	0.700(0.345)*	-0.807(0.505)*	-2.293(0.153)*
T7	0.813(0.246)*	-0.861(0.368)*	-1.142(0.111)*
T8	0.354(0.447)	-3.846(0.528)*	-2.278(0.191)*
T9	0.283(0.369)*	-0.241(0.595)*	-1.822(0.167)*
T10	-0.363(0.747)	-0.721(0.854)*	-3.136(0.322)*
T11	1.430(0.480)*	-0.049(0.627)*	-1.759(0.184)*
T12	-0.793(0.733)	-1.182(0.846)*	-2.754(0.309)*
T13	1.157(0.457)*	-0.102(0.578)*	-1.428(0.184)*
T14	1.249(0.453)*	-1.210(0.816)*	-1.297(0.194)*
T15	1.751(0.782)*	0.053(1.015)*	-1.792(0.265)*
T16	-0.198(0.524)	-1.005(0.685)*	-1.409(0.249)*
T17	0.797(0.724)	-1.041(1.169)*	-1.824(0.334)*
T18	1.639(1.096)	0.621(1.235)*	-1.760(0.352)*
T19	0.721(0.450)	-0.071(0.523)	0.051(0.219)*
ui	-0.007(0.145)	-0.053(0.194)*	-1.138(0.060)*
dr	-3.204(0.942)	-2.089(1.219)*	-0.633(0.430)*
rr	1.116(0.857)*	0.478(1.073)	0.251(0.385)*
age	0.002(0.007)*	0.013(0.009)	-0.014(0.003)*
wage	0.598(0.183)*	-0.270(0.230)	0.250(0.081)*
tenure	0.041(0.016)*	0.038(0.019)	-0.009(0.006)*

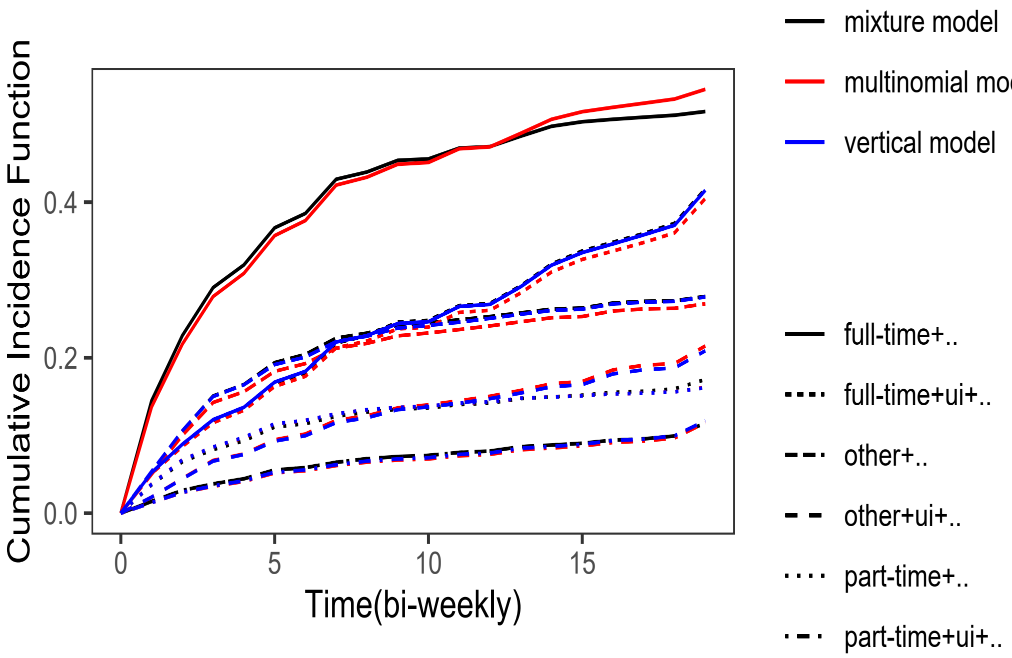

Figure 7.10: The CIFs of exit to full-time, part-time and other employment with effect of `ui` for the Multinomial Model, Vertical Model and the Discrete Time Mixture Model.

First of all, to compare the three models in relation to being fitted to data with three failure causes we will assess the effect of unemployment benefits on re-employment prospects. We have plotted the CIFs from each model with the effect of `ui` in Figure 7.10. Clearly, all three models when fitted to the training data suggest that the provision of unemployment benefits has the effect of reducing re-employment prospects. According to the *discrete time mixture model*, about 46.5% of unemployed individuals that do not receive benefits are re-employed full-time, about 24.4% part-time and about 33.6% into other types of employment, while, for those that do receive the benefits, about 41.7% are re-employed full-time, 11.5% are re-employed part-time and about 22.1% are re-employed to other types of employment. For those that do not receive benefits, the multinomial model and the *discrete time vertical model* suggest that about 54. 52% and 53.9%, are re-employed full-time, respectively, 16.2% and 16.1% are re-employed part-time, while 26.9% and 27.8% are respectively re-employed in other types of work. For recipients of these benefits, both the multinomial and the vertical suggest a drop in full-time re-employment to about

40.5% and 41.5%, respectively, and another drop to about 11.7% and 11.9%, respectively as chances of re-employment part-time, while the probability of re-employment to other types of employment also drops to about 21.5% and 20.9%, respectively. We do see a repeat of what we witnessed in Chapter 4 and 5 where the vertical model and the multinomial model were found to be more comparable in terms of the estimates for CIFs in contrast to the mixture models. It is unfortunate that we could not include the other mixture model in the analysis, it would have been instructive to observe how the two mixture models compare to one another and also to the other two models in terms of CIFs as well.

We then proceeded to assess these models in terms of their *c-Indexes*, i.e., their discriminative ability and *calibration*. We found that Indexes for the multinomial and the vertical model were 0.5734 and 0.5705, respectively, while the same metric was found to be 0.577 for the mixture model. Clearly, there is nothing that separates these models in terms of their discriminative ability, they all performed poorly in respect of unemployment data given how their respective parameters are modeled on data because these figures are barely larger than 0.5. The issue these lowly figures is a separate one because we can work to improve the fit of each model, but of critical importance to us is the fact that both the proposed models, that is, the *discrete time vertical model* and the *discrete time vertical* displayed the same discriminative ability when compared to the multinomial model with regards to this data set just as it was the case with the other data set.

Table 7.24: The *calibration* figures for the Multinomial Model, the Discrete Time Mixture Model and the Discrete Time Vertical Model .

RMSE

Death Type	Multinomial Model	Vertical Model	Mixture Model
`full-time`	0.0139	0.0139	0.0149
`part-time`	0.0059	0.0055	0.0108
`other`	0.0063	0.0063	0.0237

We then conducted the *calibration* exercise and the results are listed in Table 7.24. We also notice here that the multinomial model tends to have performed more similarly to the vertical model than to the mixture model in terms of *calibration*. The RMSE for the mixture model with regard to *other* and *part-time* failure types seem to be outliers. This is one area that may require working on in terms of reducing these errors by improving the fit of the mixture model with respect to just these two failure types.

This concludes the exercise of assessing the vertical and the mixture models against the multinomial model in term of predictive accuracy. The question, however, that remains outstanding is that given the differences between the mixture model on one hand and the pair of the multinomial model and the vertical model on the other hand in terms of the estimates for CIFs then, which model(s) should we go with if all three models compare favourably with one another in terms of *discriminative* ability, (though, the mixture model has the slightest of edges over the other two models), but the mixture model fares poorly in so far as *calibration* is concerned. Posing this question differently, can we, for instance, choose the multinomial model or the vertical model over the mixture model, as things stand, particularly on basis of the *calibration* results because all these models are almost on par in relation to *discrimination*. These are some of the questions that require further inquiry together with the issue of assessing the predictive accuracy for the *continuous time mixture model*.

7.4 Discussion

When we advanced the proposed models in Chapters 2, 3 and 4 we demonstrated their application by using data that had only two failure types. In this chapter, we applied all these models to data that has five failure causes to demonstrate that the model could be applied to model data with more than two failure causes. We commenced this exercise by fitting all three proposed models to this data, that is, the two mixture models and the vertical model, together with the multinomial model. For illustrative purposes, we assessed the effect of injury type on the risks of dying by these five failure causes. We found that all three proposed models lead to the same conclusions as the multinomial model. Because our data did not come with its own set of subjects with missing failure causes we were compelled to generate these subjects from data. Also, due to the structure of the data we were forced to collapse two failure causes into one failure cause. We then fitted the proposed missing failure causes models together with the missing failure causes multinomial model and we found that all proposed models together with the multinomial model led to the same conclusions again. We then moved to validate two of the proposed models and the multinomial in terms of *discrimination* and *calibration* with respect to this new data set and we found that the proposed models compared favourably to the multinomial model. We also conducted a similar exercise with respect to the old unemployment data, but with three failure causes. Here we found that the two proposed models compared favourably to the multinomial model in terms of *discriminative* ability. We also found the proposed mixture model to fall somewhat short when compared to the vertical model and the multinomial model in terms of *calibration*, a minor issue that can easily be addressed by improving the fit of the mixture model. Despite these minor challenges we have shown that two of the proposed models compare favourably to the multinomial model in terms of validation with regard to the new data set and the old unemployment data set.

The following chapter is the final chapter of this thesis where we summarise our finding and discuss future directions.

CHAPTER 8

Conclusions

The main thrust of this thesis was to advance discrete time models for the analysis of ordinary competing risks data as alternatives to the multinomial model and the binomial model, and to upscale the proposed models into models that can handle missing failure causes and cured subjects. Towards that end, we focussed on two continuous time competing risks models, that is, the mixture model and the vertical model with a view to re-formulate these models as discrete time competing risks models. Both these models come as mixtures of two components, a component that relates to failure time and another one that is concerned with failure type. These mixtures follow from the factorization of the bivariate distribution of (T, D) that are suggested by these models. The mixture model proposes a decomposition of this bivariate distribution as; $P(T, D) = P(T|D)P(D)$ which, amongst other things, leads to characterization of data in terms of component hazards and failure type probabilities. The vertical model proposes total hazards and relative hazards for modelling data. This follows from the following decomposition; $P(T, D) = P(T|D)P(D)$ which is proposed by the vertical model.

Two mixture models were advanced for application in discrete time. For the first model, we demonstrated how the mixture model, with the specifications as suggested by its authors, can be applied in discrete time by adjusting the failure/censoring times. For the second mixture model, we advanced a truly discrete time version of the mixture model by specifying discrete time component hazards. For regression purposes, these quantities were modeled within the GLM framework while the failure type distribution was modeled with a multinomial distribution. The vertical model was re-formulated as a discrete time model by specifying discrete time total hazards. Likewise, the total hazards were modeled within the GLM framework as well while the failure type distribution conditional on failure type was also modeled via a multinomial distribution for regression purposes.

While the proposed models produced comparable results upon application, they also differed in certain aspects. One area where there were marked differences was in respect of estimation methods. For application to ordinary competing risks data, the proposed mixture models rely on an EM Algorithm for estimation purposes. This is true at a nonparametric level as well as for regression purposes. On the other hand, when ordinary competing risks data is modeled with the proposed vertical model then standard estimation methods are implemented which are typically applied when data is modeled with the existing discrete time models. It was shown that the vertical model is invariant to the presence or absence of missing failure causes. This means that the same estimation methods that were applied for ordinary data are also applicable when data comes with missing failure causes while the existing discrete time model relies on the MI for handling missing failure causes. An EM Algorithm at a higher level of complexity is implemented for handling missing failure causes when data is modeled with the mixture model. It may be worth noting that there is no guarantee that the MLE that are obtained via an EM Algorithm are global in general. That is the reason that it is often recommended to initialize the EM Algorithm at various starting points, especially, for regression analysis. This issue did not arise when we applied the mixture models as the results that were produced by these models were found to be comparable to the results that were obtained by applying the existing discrete time models. This is also not an issue with regard to the proposed vertical model as standard packages are implemented for for regression analysis.

Another area where the proposed models differed was in the interpretation of the regression coefficients. The proposed vertical model allowed for a limited assessment of the covariate effects on the CSHs from the regression coefficients in the models for total hazards and relative hazards while the regression model for the CIF that arises under the proposed vertical model complicates the assessment of covariate effects from the regression coefficients. The proposed mixture models also allowed for limited evaluation of covariate effects on the CIFs from the regression coeffients in the models for component hazards and failure type probabilities while the regression model for CSHs is complicated.

As a final model in this thesis, we advanced a nonparametric mixture vertical cure model. While the primary objective of advancing this model was to suggest a discrete time nonparametric model for handling cured subjects, we also demonstrated the risks that are associated with a naive application of standard analysis models that do not account for these subjects. One of the notable results of the nonparametric vertical for handling cured subjects is the observation that the conditional relative hazards are identical to the population relative hazards. This is a significant result because it means that the relative hazards are also invariant to the presence or absence of cured subjects. Thus, if relative hazards are the preferred quantities for comparing the risks of failure according to the failure causes that are under consideration, say, conditional on covariates, these quantities can easily be determined by fitting a multinomial distribution to the original data excluding censored subjects whether data comes with missing failure or cured subjects. In simple terms, the advantage of the relative hazards approach is that the EM Algorithm is conveniently bypassed even when data comes with cured subject.

Whilst we are still discussing the nonparametric vertical model for handling cured subjects recall that in the estimation of the proportion of uncured subjects and conditional total hazards, the E-Step in the $(r+1)^{\text{th}}$ iteration of the Algorithm was expressed as a sum of;

$$Q(p|p^{(r)}) = \sum_{i=1}^{n} g_i^{(r)} \log p + (1 - g_i^{(r)}) \log(1 - p)$$

and,

$$Q(\boldsymbol{h}|\boldsymbol{h}^{(r)}) = \sum_{i=1}^{n} \sum_{s=1}^{t_i} d_{is} \log h_u(s) + (g_{is}^{(r)} - d_{is}) \log(1 - h_u(s)).$$

Clearly, both $Q(p|p^{(r)})$ and $Q(\boldsymbol{h}|\boldsymbol{h}^{(r)})$ are recognisable as kernels of a binomial log-likelihood function. This means that upgrading the proposed nonparametric into a regression model should not be a difficult exercise as both the conditional total hazards and the proportion of uncured subjects can be modeled within the GLM framework while the relative hazards continue to be linked to data via the multinomial model. This is one area that requires further exploration.

It should also be noted that all subjects contribute towards $Q(\boldsymbol{h}|\boldsymbol{h}^{(r)})$, that is, all failures and censored subjects contribute towards this log-likelihood function. It should not be a difficult exercise also to show that in the presence of subjects with missing failure causes, these subjects should contribute to $Q(\boldsymbol{h}|\boldsymbol{h}^{(r)})$ as well as part of general failures in much the same way it was shown that missing failures contribute to the failure time likelihood function in the absence of cured subjects. There is no room for confusing cured subjects with subjects that come with missing failure causes. The subjects with missing failure causes are failures while cured subjects are mixed with censored subjects and as such these subjects are identifiable from one another. Thus, yet another area that requires exploration in the future is to upscale the nonparametric vertical model for handling cured subjects into a model that can also deal with missing failure causes. Recall that the data that was used for illustration of the proposed vertical model, that is, the unemployment data, came with missing failure causes as well. These subjects were excluded when we applied the proposed model. If it can be shown that the same vertical that we have advanced in this thesis can also handle missing failure causes it would mean that there would be no need to discard the valuable information that comes with these subjects.

Finally, the mixture model that was originally advanced for application to ordinary competing risks data has been extended into a model that can handle cured subjects in the continuous time realm. In part one of this thesis we have demonstrated how the mixture model, as advanced by its authors, can be applied in discrete time. We have also advanced a discrete time version of this model. Another possibility that may be considered in the future is to upscale these models into models that can be applied to data that has a sizable proportion of cured subjects in discrete time as we have managed to demonstrate regarding the vertical model, albeit, at a nonparametric level.

In the final chapter of this thesis, we applied all three models to another data set. The difference between this new data set and the previous data set was that the new data set came with five failure causes. The intention here was to demonstrate that the proposed models which were shown to be easily applicable to data with only two failure causes can equally apply when there are more than two failure causes. When we applied the three proposed ordinary competing risks to this data set we found that all three models compared favourably to the multinomial model by way of examining the CIFs. We took this validation exercise a step further by additionally evaluating two of the proposed models in terms of *discrimination* and *calibration*. At this stage, we could only assess the *discrete time mixture model* and the *vertical* model. We found these two proposed models to compare favorably with the multinomial model.

Having re-formulated both the mixture model and the vertical model as discrete time competing risks models in the first part of this thesis and having extended the discrete time competing risks models that were advanced in the first part of the thesis into missing failure causes models and, finally, having advanced a discrete time vertical model for handling cured subjects nonparametrically, we conclude this thesis with some measure of confidence that all its objectives as set out in the introduction were met satisfactorily.

Table 8.1: Summary of contributions

Contributions	Publications
1. A continuous time mixture model for application in discrete time	Ndlovu, B.D., Melesse, S.F, and Zewotir, T. (2019). A mixture model with application to discrete competing risks data. *South African Statistical Journal*, 2:73-86
2. A continuous time mixture model for application in discrete time in the presence of missing failure causes	Ndlovu, B.D., Melesse, S.F, and Zewotir, T. (2022). An EM model for analysis of discrete time competing risks data with missing failure causes. *Model Assisted Statistics and Applications*, 17:167-174
3. A nonparametric model for handling missing failure causes in discrete time	Ndlovu, B.D., Melesse, S.F, and Zewotir, T. (2020). A nonparametric vertical model: an application to discrete time competing risks data with missing failure causes. *South African Statistical Journal*, 2:231-241
4. A vertical regression model for ordinary discrete time competing risks data	Ndlovu, B.D., Melesse, S.F, and Zewotir, T. (2022). A regression analysis of discrete time competing risks data using a vertical model approach. *South African Statistical Journal*, 1:21-36
5. A verticalnonparametric model for ordinary discrete time competing risks data	Ndlovu, B.D., Melesse, S.F, and Zewotir, T. (2023). A nonparametric analysis of discrete time competing risks data:A comparison of the cause-specific-hazards approach and the vertical approach. *Statistics in Transition*, 24:61-76

Appendix A

The continuous time mixture model:- standard errors for the CIF

A Complete Case

We rely on the multivariate delta method to determine the variance for the CIF estimate. Suppose that $\boldsymbol{\theta}^0$ is the MLE of $\boldsymbol{\theta}$ in the M-Step at convergence. Let $\boldsymbol{\eta}_j^\beta(t) = (\eta_j^\beta(1), \ldots, \eta_j^\beta(t))^T$, where $\eta_j^\beta(s) = \beta_{0js} + \boldsymbol{x}^T\boldsymbol{\beta}_{1j}$ and $\boldsymbol{\eta}^\gamma = (\eta_1^\gamma, \ldots \eta_{J\text{-}1}^\gamma)^T$. where $\eta_j^\gamma = \gamma_{0j} + \boldsymbol{x}^T\boldsymbol{\gamma}_{1j}$. Recall that;

$$\hat{F}_j(t|\boldsymbol{x}, \hat{\boldsymbol{\theta}}) = \hat{\pi}_j(\boldsymbol{x}, \hat{\boldsymbol{\gamma}})(1 - \hat{S}_j(t|\boldsymbol{x}, \hat{\boldsymbol{\beta}}_j))$$

The variance $\hat{F}_j(t|\boldsymbol{x}, \hat{\boldsymbol{\theta}})$ for can be written as;

$$V(\hat{F}_j(t|\boldsymbol{x}; \hat{\boldsymbol{\theta}})) = \Gamma^T\Sigma\Gamma\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

where Γ is vector of partial derivatives;

$$\Gamma = \begin{bmatrix} \frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\eta}^\gamma} \\ \frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\eta}_j^\beta(t)} \end{bmatrix}$$

and,

$$\Sigma = \begin{bmatrix} \mathrm{V}(\boldsymbol{\eta}^\gamma) & \mathbf{0} \\ \mathbf{0} & \mathrm{V}(\boldsymbol{\eta}_j^\beta(t)) \end{bmatrix}$$

The off diagonal elements of Σ are zero because;

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\eta}^\gamma \partial \boldsymbol{\eta}_j^\beta(q)} = \mathbf{0}$$

The components of $\dfrac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\eta}^\gamma}$ are;

$$\frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial \eta_j^{\gamma}} = \pi_j(\boldsymbol{x},\boldsymbol{\gamma})(1-\pi_j(\boldsymbol{x},\boldsymbol{\gamma}))(1-S_j(t|\boldsymbol{x},\boldsymbol{\beta}_j)) \quad j=1,2,\ldots,J\text{-}1$$

and the components of $\dfrac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\eta}_j^{\beta}(t)}$ are;

$$\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \eta_j^{\beta}(s)} = -\pi_j(\boldsymbol{x},\boldsymbol{\gamma})\lambda_j(s|\boldsymbol{x},\boldsymbol{\beta}_j)S_j(t|\boldsymbol{x},\boldsymbol{\beta}_j) \quad s=1,2\ldots,t$$

Then;

$$\begin{aligned}\text{Var}(\hat{F}_j(t|\boldsymbol{x},\boldsymbol{\theta})) &= \sum_{j=1}^{J\text{-}1}\sum_{k=1}^{J\text{-}1} \frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \eta_j^{\gamma}} \frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \eta_k^{\gamma}} \text{Cov}(\eta_j^{\gamma};\eta_k^{\gamma}) \\ &+ \sum_{s=1}^{t}\sum_{l=1}^{t} \frac{\partial F_j(t|\boldsymbol{x}),\boldsymbol{\theta}}{\partial \eta_j^{\beta}(s)} \frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \eta_j^{\beta}(l)} \text{Cov}(\eta_j^{\beta}(s);\eta_j^{\beta}(l))\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\end{aligned}$$

where;

$$\text{Cov}(\eta_j^{\beta}(s);\eta_j^{\beta}(l)) = \text{Cov}(\beta_{0js};\beta_{0jl}) + \sum_{w}^{p} x_w\big(\text{Cov}(\beta_{0js};\beta_{1jw}) + \text{Cov}(\beta_{0jl};\beta_{1jw})\big) + \boldsymbol{x}^T\text{Var}(\boldsymbol{\beta}_{1j})\boldsymbol{x}$$

and,

$$\text{Cov}(\eta_j^{\gamma};\eta_k^{\gamma}) = \text{Cov}(\gamma_{0j};\gamma_{0k}) + \sum_{w}^{p} x_w\big(\text{Cov}(\gamma_{0j};\gamma_{1kw}) + \text{Cov}(\gamma_{0k};\gamma_{1jw})\big) + \boldsymbol{x}^T\text{Cov}(\boldsymbol{\gamma}_{1j};\boldsymbol{\gamma}_{1k})\boldsymbol{x}$$

The covariance matrices produced by statistical packages are based on the complete data i.e V($\boldsymbol{\gamma}$) and V($\boldsymbol{\beta}_j$) are complete data covariance matrices. We apply the Oakes (1999) approach and it is given by

$$\mathcal{I}_y(\boldsymbol{\theta}^0) = -\frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} + \frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{0T}}$$

where $\boldsymbol{\theta}^0$ is MLE of $\boldsymbol{\theta}$ at the convergence of EM algorithm. The first term of the above equation is the complete data information and second term is the missing information that we have to compute. We begin with $\boldsymbol{\beta}_j$, the vector of cause j failure time parameters, and assume that the subjects have been re-indexed so that the first k subjects are uncensored and the remaining $(n-k)$ are censored. For convenience, let $S_{ij} = S_j(s|\boldsymbol{x};\boldsymbol{\beta}_j)$, $\pi_{ij} = \pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma})$, and $\lambda_{ijs} = \lambda_j(s|\boldsymbol{x};\boldsymbol{\beta}_j)$. The complete data Q score functions can be written as;

$$\frac{\partial \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{0js}^0} = \sum_{i=1}^{k} d_{ijs} - \epsilon_{is}\lambda_{ijs} + \sum_{i=k+1}^{n} d_{ijs} - v_{ij}\epsilon_{is}\lambda_{ijs}$$

$$\frac{\partial \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{1ja}^0} = \sum_{i=1}^{k}\sum_{s=1}^{t_i}(d_{ijs} - \epsilon_{is}\lambda_{ijs})x_{ia} + \sum_{i=k+1}^{n}\sum_{s=1}^{t_i}(d_{ijs} - v_{ij}\epsilon_{is}\lambda_{ijs})x_{ia}$$

Suppressing the superscripts until the final result and using the chain rule, the partial derivatives of the pseudo-variable v_{ij} are given by

$$\begin{aligned}
\frac{\partial v_{ij}^0}{\partial \beta_{0js}^0} &= \frac{\partial v_{ij}^0}{\partial S_{ij}^0}\frac{\partial S_{ij}}{\beta_{ijs}^0} = \frac{\big(\sum_{l=1}^{J}\pi_{ij}S_{ij}\big)\pi_{ij} - \pi_{ij}S_{ij}(\pi_{ij})}{\big(\sum_{j=1}^{J}\pi_{ij}S_{ij}\big)^2} \times -S_{ij}\epsilon_{is}\lambda_{ijs}(1-\lambda_{ijs}) \\
&= -v_{ij}^0(1-v_{ij}^0)\epsilon_{is}\lambda_{ijs}^0(1-\lambda_{ijs}^0) \\
\frac{\partial v_{ij}}{\partial \beta_{1ja}} &= \frac{\partial v_{ij}}{\partial S_{ij}}\frac{\partial S_{ij}}{\partial \beta_{ija}} = -\frac{\big(\sum_{l=1}^{J}\pi_{il}S_{il}\big)\pi_{ij} - \pi_{ij}S_{ij}\pi_{ij}}{(\sum_{l=1}^{J}\pi_{il}S_{il})^2}(S_{ij}x_{ia})\big(\sum_{s=1}^{t_i}\epsilon_{is}\lambda_{ijs}\big) \\
&= -v_{ij}(1-v_{ij})\sum_{s=1}^{t_i}\epsilon_{is}\lambda_{ijs}
\end{aligned}$$

It then follows that;

$$\begin{aligned}
\frac{\partial^2 \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{0js}\partial \beta_{0js}^0} &= \sum_{i=k+1:s\le t_i}^{n} -\epsilon_s\lambda_{ijs}^0\frac{\partial v_{ij}}{\partial \beta_{0js}} \\
&= \sum_{i=k+1:s\le t_i}^{n} \epsilon_{is}\lambda_{ijs}^0 v_{ij}(1-v_{ij})\epsilon_{is}\lambda_{ijs}
\end{aligned}$$

since $\frac{\partial v_{ij}}{\partial \beta_{0js}} = 0$ when $s > t_i$. Therefore the missing information corresponding to β_{0js} is given by

$$\frac{\partial^2 \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{0js}\partial \beta_{0js}^0}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \sum_{i=k+1:s\le t_i}^{n} (\epsilon_{is}\lambda_{ijs}^0)^2 v_{ij}^0(1-v_{ij}^0) \quad s=1,\ldots,q$$

Regarding the regression coefficients, we have;

$$\begin{aligned}
\frac{\partial^2 \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{1jb}\partial \beta_{1ja}^0} &= \sum_{i=k+1}^{n}\sum_{s=1}^{t_i}(-\epsilon_{is}\lambda_{ijs}^0)x_{ia}\frac{\partial v_{ij}}{\partial \beta_{1jb}} \\
&= \sum_{i=k+1}^{n}\sum_{s=1}^{t_i}(\epsilon_{is}\lambda_{ijs}^0)x_{ia}x_{ib}v_{ij}(1-v_{ij})\sum_{s=1}^{t_i}\epsilon_{is}\lambda_{ijs} \\
&= \sum_{i=k+1}^{n} v_{ij}(1-v_{ij})x_{ia}x_{ib}\sum_{s=1}^{t_i}\epsilon_{is}\lambda_{ijs}^0\sum_{s=1}^{t_i}\epsilon_{is}\lambda_{ijs}
\end{aligned}$$

so that the missing information component regarding $\text{Cov}(\beta_{1ja};\beta_{1jb})$ is;

$$\frac{\partial^2 \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{1jb}\partial \beta_{1ja}^0}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \sum_{i=k+1}^{n} v_{ij}^0(1-v_{ij}^0)x_{ia}x_{ib}\bigg(\sum_{s=1}^{t_i}\epsilon_{is}\lambda_{ijs}^0\bigg)^2$$

Regarding Cov(β_{0js}, β_{1ja});

$$\begin{aligned}
\frac{\partial^2 \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial\beta_{0js}\partial\beta_{1ja}^0} &= \sum_{i=k+1}^{n}(-\epsilon_{is}\lambda_{ijs}^0)x_{ia}\frac{\partial v_{ij}}{\partial\beta_{0js}} \\
&= \sum_{i=k+1}^{n}(\epsilon_{is}\lambda_{ijs}^0)x_{ia}v_{ij}(1-v_{ij})\sum_{s=1}^{t_i}\epsilon_{is}\lambda_{ijs} \\
&= \sum_{i=k+1}^{n} v_{ij}(1-v_{ij})x_{ia}\epsilon_{is}\lambda_{ijs}^0\sum_{s=1}^{t_i}\epsilon_{is}\lambda_{ijs}
\end{aligned}$$

The corresponding missing information is, therefore;

$$\left.\frac{\partial^2 \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial\beta_{0js}\partial\beta_{1ja}^0}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \sum_{i=k+1}^{n} v_{ij}^0(1-v_{ij}^0)x_{ia}\epsilon_{is}\lambda_{ijs}^0\sum_{s=1}^{t_i}\epsilon_{is}\lambda_{ijs}^0$$

We now focus on the variance for $\boldsymbol{\gamma}$. We begin by establishing these results;

$$\begin{aligned}
\frac{\partial v_{ij}}{\partial\gamma_{1ja}} &= \frac{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il}+\pi_J S_{iJ}\big)\pi_{ij}(1-\pi_{ij})S_{ij}x_{ia} - \pi_{ij}S_{ij}\big(\pi_{ij}(1-\pi_{ij})S_{ij} - \pi_{ij}\sum_{l\neq j}^{J}\pi_{il}S_{il}\big)x_{ia}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il}+\pi_J S_{iJ}\big)^2} \\
&= \frac{\pi_{ij}S_{ij}\big\{\sum_{l\neq j}^{J}\pi_{il}S_{il}(1-\pi_{ij}) + \pi_{ij}\sum_{l\neq j}^{J}\pi_{il}S_{il}\big\}x_{ia}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il}+\pi_J S_{iJ}\big)^2} \\
&= \frac{\pi_{ij}S_{ij}\times\sum_{l\neq j}^{J}\pi_{il}S_{il}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il}+\pi_J S_{iJ}\big)^2}x_{ia} \\
&= v_{ij}(1-v_{ij})x_{ia} \\
\frac{\partial v_{ij}}{\partial\gamma_{1kb}} &= \frac{\big(-\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il}+\pi_{iJ}S_{iJ}\big)\pi_{ik}\pi_{ij}S_{ij}x_{ib} - \big(\pi_{ik}(1-\pi_{ik})S_{ik} - \pi_{ik}\sum_{l\neq k}^{J}\pi_{il}S_{il}\big)\pi_{ij}S_{ij}x_{ib}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il}+\pi_{iJ}S_{iJ}\big)^2} \\
&= \frac{\pi_{ij}\pi_{ik}S_{ij}x_{ib}\big\{-\sum_{l=1}^{j}\pi_{il}S_{il} - S_{ik} + \sum_{l=1}^{J}\pi_{il}S_{il}\big\}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il}+\pi_{iJ}S_{iJ}\big)^2} \\
&= -\frac{\pi_{ij}S_{ij}}{\sum_{l=1}^{J}\pi_{il}S_{il}}\frac{\pi_{ik}S_{ik}}{\sum_{l=1}^{J}\pi_{il}S_{il}}x_{ib} \\
&= -v_{ij}v_{ik}x_{ib}
\end{aligned}$$

We again re-index the subjects as before, and write the "complete" data score function as follows;

$$\frac{\partial Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial\gamma_{1ja}} = \sum_{i=1}^{k}(d_{ij}-\pi_{ij}^0)x_{ia} + \sum_{i=k+1}^{n}(v_{ij}-\pi_{ij}^0)x_{ia}$$

Then,

$$\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial\gamma_{1ja}\partial\gamma^0_{1ja}} = \sum_{i=k+1}^{n} \frac{\partial v_{ij}}{\partial\gamma_{1ja}} x_{ia} = \sum_{i=k+1}^{n} v_{ij}(1-v_{ij})x^2_{ia}$$
$$\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial\gamma_{1kb}\partial\gamma^0_{1ja}} = \sum_{i=k+1}^{n} \frac{\partial v_{ij}}{\partial\gamma_{1jb}} x_{ia} = -\sum_{i=k+1}^{n} v_{ij}v_{ik}x_{ia}x_{ib}$$

Thus, the missing information components in respect of Var(γ_{1ja}) and Cov$(\gamma_{1kb};\gamma_{1ja})$ are given by

$$\left.\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial^2\gamma_{1ja}\partial\gamma^0_{1ja}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \sum_{i=k+1}^{n} v^0_{ij}(1-v^0_{ij})x^2_{ia}$$
$$\left.\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial\gamma_{1kb}\partial\gamma^0_{1ja}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = -\sum_{i=k+1}^{n} v^0_{ij}v^0_{ik}x_{ia}x_{ib}$$

B Missing Failure Causes

Suppose that $\boldsymbol{\theta}^0$ is the MLE of $\boldsymbol{\theta}$ in the M-Step at convergence. We rely on the multivariate delta method to determine the variance for the CIF estimates, where;

$$\hat{F}_j(t|\boldsymbol{x},\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\pi}}_j(\boldsymbol{x},\hat{\boldsymbol{\gamma}}_j)(1-\hat{S}_j(t|\boldsymbol{x},\hat{\boldsymbol{\beta}}_j))$$

Let $\boldsymbol{\eta}_{\beta_j} = (\eta_{\beta_{j1}},\eta_{\beta_{j1}},\ldots\eta_{\beta_{jt}})^T$ and $\boldsymbol{\eta}_\gamma = (\eta_{\gamma_1},\eta_{\gamma_2},\ldots\eta_{\gamma_{(J-1)}})^T$, where $\eta_{\beta_{js}} = \beta_{0js} + \boldsymbol{x}^T\boldsymbol{\beta}_{1j}$ and $\eta_{\gamma_j} = \gamma_{0j} + \boldsymbol{x}^T\boldsymbol{\gamma}_{1j}$, then $V(\hat{F}_j(t|\boldsymbol{x};\hat{\boldsymbol{\theta}}))$ can be written as;

$$V(\hat{F}_j(t|\boldsymbol{x};\boldsymbol{\theta})) = \left(\frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\eta}_\gamma},\frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\eta}_{\beta_j}}\right)^T \begin{bmatrix} \mathrm{V}(\boldsymbol{\eta}_\gamma) & \mathbf{0} \\ \mathbf{0} & \mathrm{V}(\boldsymbol{\eta}_{\beta_j}) \end{bmatrix} \left.\left(\frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\eta}_\gamma},\frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\eta}_{\beta_j}}\right)\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$$

since $\dfrac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\eta}_\beta\partial\boldsymbol{\eta}_\gamma} = 0$, where $\boldsymbol{\eta}_\beta = (\boldsymbol{\eta}^T_{\beta_1},\boldsymbol{\eta}^T_{\beta_2},\ldots\boldsymbol{\eta}^T_{\beta_J})^T$. The partial derivatives are given as

$$\frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial\eta_{\gamma_j}} = (1-\pi_j(\boldsymbol{x};\boldsymbol{\gamma}))\pi_k(\boldsymbol{x};\boldsymbol{\gamma})Q_j(t|\boldsymbol{x};\boldsymbol{\beta}_j)$$
$$\frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial\eta_{\beta_{js}}} = -\pi_j(\boldsymbol{x};\boldsymbol{\gamma})\lambda_j(s|\boldsymbol{x};\boldsymbol{\beta}_j)Q_j(t|\boldsymbol{x};\boldsymbol{\beta}_j)$$

where, $Q_j(t|\boldsymbol{x};\boldsymbol{\beta}_j) = 1 - S_j(t|\boldsymbol{x};\boldsymbol{\beta}_j)$. Thus,

$$\text{Var}(\hat{F}_j(t|\boldsymbol{x})) = \sum_{m=1}^{J\text{-}1}\sum_{n=1}^{J\text{-}1} \frac{\partial F_j(t|\boldsymbol{x})}{\partial \eta_{\gamma_m}} \frac{\partial F_j(t|\boldsymbol{x})}{\partial \eta_{\gamma_n}} \text{Cov}(\eta_{\gamma_m};\eta_{\gamma_n}) + \sum_{s=1}^{t}\sum_{k=1}^{t} \frac{\partial F_j(t|\boldsymbol{x})}{\partial \eta_{\beta_{js}}} \frac{\partial F_j(t|\boldsymbol{x})}{\partial \eta_{\beta_{jk}}} \text{Cov}(\eta_{\beta_{js}};\eta_{\beta_{jk}})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$$

where,

$$\text{Cov}(\eta_{\beta_{js}};\eta_{\beta_{jk}}) = \text{Cov}(\beta_{0js};\beta_{0jk}) + \sum_{w}^{p} x_w\big(\text{Cov}(\beta_{0js};\beta_{1jw}) + \text{Cov}(\beta_{0jk};\beta_{1jw})\big) + \boldsymbol{x}^T\text{Var}(\boldsymbol{\beta}_{1j})\boldsymbol{x}$$

and,

$$\text{Cov}(\eta_{\gamma_m};\eta_{\gamma_n}) = \text{Cov}(\gamma_{0m};\gamma_{0n}) + \sum_{w}^{p} x_w\big(\text{Cov}(\gamma_{0m};\gamma_{1nw}) + \text{Cov}(\gamma_{0n};\gamma_{1mw})\big) + \boldsymbol{x}^T\text{Cov}(\boldsymbol{\gamma}_{1m};\boldsymbol{\gamma}_{1n})\boldsymbol{x}$$

Since we have applied an EM algorithm, the components of $\text{V}(\boldsymbol{\eta}_\gamma)$ and $\text{V}(\boldsymbol{\eta}_{\beta_j})$ require adjustments. We apply the Oakes (1999) approach which is given by

$$\mathcal{I}_y(\boldsymbol{\theta}^0) = -\frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} + \frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}\partial{\boldsymbol{\theta}^0}^T} \tag{B.1}$$

The first term of the above equation is the complete data information and second term is the missing information. In computing the parameter standard errors, the packages assume complete data information which must then be adjusted according to (B.1). We begin with $\boldsymbol{\beta}_j$, and compute the missing information. We assume that the subjects have been re-indexed so that the first k subjects are uncensored, the next l subjects have missing failure causes and the remaining $(n-k-l)$ are censored. For convenience, let $S_{ij} = S_j(s|\boldsymbol{x};\boldsymbol{\beta}_j^0)$, $\pi_{ij} = \pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma}^0)$, and $\lambda_{ijs} = \lambda_j(s|\boldsymbol{x};\boldsymbol{\beta}_j^0)$. The complete data Q score functions can be written as;

$$\frac{\partial \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j)}{\partial\beta_{0js}} = \sum_{i=1}^{k} d_{ijs} - \epsilon_{is}\lambda_{ijs} + \sum_{i=k+1}^{l+k} d_{i*s}u_{ij}^0 - u_{ij}^0\epsilon_{is}\lambda_{ijs} + \sum_{i=k+l+1}^{n} d_{ijs} - v_{ij}^0 v_{is}\lambda_{ijs}$$

$$\frac{\partial \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial\beta_{1ja}} = \sum_{i=1}^{k}\sum_{s=1}^{t_i}(d_{ijs} - \epsilon_{is}\lambda_{ijs})x_{ia} + \sum_{i=k+1}^{l+k}\sum_{s=1}^{t_i}(d_{i*s}u_{ij}^0 - u_{ij}^0\epsilon_{is}\lambda_{ijs})x_{ia} + \sum_{i=k+l+1}^{n}\sum_{s=1}^{t_i}(d_{ijs} - v_{ij}^0\epsilon_{is}\lambda_{ijs})x_{ia}$$

Since $d_{i*s} = 1$ for $s = t_i$ and 0 otherwise, then, we can, therefore, re-write; $\dfrac{\partial \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j)}{\partial \beta_{0js}}$ and $\dfrac{\partial \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}^0_j)}{\partial \beta^0_{1ja}}$ in respect of u^0_{ij}as;

$$\begin{aligned}
\frac{\partial \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j)}{\partial \beta_{0js}} &= \sum_{i=1}^{k} d_{ijs} - \epsilon_{is}\lambda_{ijs} + \sum_{i=k+1}^{k+l}(1-\epsilon_{is}\lambda_{ijs})u^0_{ij} + \sum_{i=k+l+1}^{n} d_{ijs} - v^0_{ij}\epsilon_{is}\lambda_{ijs} \quad \text{when } s = t_i \\
&= \sum_{i=1}^{k} d_{ijs} - \epsilon_{is}\lambda_{ijs} - \sum_{i=k+1}^{k+l}\epsilon_{is}\lambda_{ijs}u^0_{ij} + \sum_{i=k+l+1}^{n} d_{ijs} - v^0_{ij}\epsilon_{is}\lambda_{ijs} \quad \text{when } s \neq t_i \\
\frac{\partial \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}^0_j)}{\partial \beta_{1ja}} &= \sum_{i=1}^{k}\sum_{s=1}^{t_i}(d_{ijs} - \epsilon_{is}\lambda_{ijs})x_{ia} + \sum_{i=k+1}^{l+k}(1-\sum_{s=1}^{t_i}\epsilon_{is}\lambda_{ijs})x_{ia}u^0_{ij} \\
&+ \sum_{i=k+l+1}^{n}\sum_{s=1}^{t_i}(d_{ijs} - \underline{u}^0_{ij}\epsilon_{is}\lambda_{ijs})x_{ia}
\end{aligned}$$

Using the chain rule, the partial derivatives of the pseudo-variables are given by

$$\begin{aligned}
\frac{\partial v^0_{ij}}{\partial \beta^0_{0js}} &= \frac{\partial v^0_{ij}}{\partial S_{ij}}\frac{\partial S_{ij}}{\partial \beta^0_{0js}} = \frac{\big(\sum_{l=1}^{J}\pi_{ij}S_{ij}\big)\pi_{ij} - \pi_{ij}S_{ij}(\pi_{ij})}{\big(\sum_{j=1}^{J}\pi_{ij}S_{ij}\big)^2} \times -S_{ij}\epsilon_{is}\lambda_{ijs} = -v^0_{ij}(1-v^0_{ij})\epsilon_{is}\lambda^0_{ijs} \\
\frac{v^0_{ij}}{\partial \beta^0_{1ja}} &= \frac{\partial v^0_{ij}}{\partial S_{ij}}\frac{\partial S_{ij}}{\partial \beta^0_{ija}} = -\frac{\big(\sum_{l=1}^{J}\pi_{il}S_{il}\big)\pi_{ij} - \pi_{ij}S_{ij}\pi_{ij}}{(\sum_{l=1}^{J}\pi_{il}S_{il})^2}(S_{ij}x_{ia})\big(\sum_{s=1}^{t_i}\epsilon_{is}\lambda_{ijs}\big) \\
&= -v^0_{ij}(1-v^0_{ij})x_{ia}\sum_{s=1}^{t_i}\epsilon_{is}\lambda^0_{ijs} \\
\frac{\partial u^0_{ij}}{\partial \beta^0_{0js}} &= \frac{\partial u^0_{ij}}{\partial \lambda_{ij}}\frac{\partial \lambda_{ij}}{\partial \beta^0_{0js}} = \frac{\big(\sum_{l=1}^{J}\pi_{ij}\lambda_{ij}\big)\pi_{ij} - \pi_{ij}\lambda_{ij}(\pi_{ij})}{\big(\sum_{j=1}^{J}\pi_{ij}\lambda_{ij}\big)^2} \times \lambda_{ij}(1-\lambda_{ij}) = (1-\lambda_{ij})u^0_{ij}(1-u^0_{ij}) \\
\frac{\partial u^0_{ij}}{\partial \beta^0_{1ja}} &= \frac{\partial u^0_{ij}}{\partial \lambda_{ij}}\frac{\partial \lambda_{ij}}{\partial \beta^0_{ija}} = -\frac{\big(\sum_{l=1}^{J}\pi_{il}\lambda_{il}\big)\pi_{ij} - \pi_{ij}\lambda_{ij}\pi_{ij}}{(\sum_{l=1}^{J}\pi_{il}\lambda_{il})^2}\lambda_{ij}(1-\lambda_{ij})x_{ia} = (1-\lambda_{ij})u^0_{ij}(1-u^0_{ij})x_{ia}
\end{aligned}$$

It then follows that;

$$\begin{aligned}
\frac{\partial^2 \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}^0_j)}{\partial \beta^0_{0js}\partial \beta_{0js}} &= \sum_{i=k+1}^{k+l}(1-\epsilon_{is}\lambda_{ijs})u^0_{ij}(1-u^0_{ij})(1-\lambda^0_{ijs}) \\
&+ \sum_{i=k+l+1}^{n}\epsilon_{is}\lambda_{ijs}v^0_{ij}(1-v^0_{ij})\epsilon_s\lambda^0_{ijs} \quad \text{when } s = t_i \\
&= \sum_{i=k+1}^{k+l}\epsilon_{is}\lambda_{ijs}u^0_{ij}(1-u^0_{ij})(1-\lambda^0_{ijs}) + \sum_{i=k+l+1}^{n}\epsilon_{is}\lambda_{ijs}v^0_{ij}(1-v^0_{ij})\epsilon_s\lambda^0_{ijs} \quad \text{when } s \neq t_i
\end{aligned}$$

since $\frac{\partial v_{ij}}{\partial \beta_{0js}} = 0$ when $s > t_i$. Therefore the missing information corresponding to β_{0js} is;

$$\begin{aligned}
\left.\frac{\partial^2 \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{0js}^0 \partial \beta_{0js}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} &= \sum_{i=k+1}^{k+l} (1-\epsilon_{is}\lambda_{ijs}^0) u_{ij}^0 (1-u_{ij}^0)(1-\lambda_{ijs}^0) \\
&+ \sum_{i=k+l+1}^{n} v_{ij}^0 (1-v_{ij}^0) \big(\epsilon_s \lambda_{ijs}^0\big)^2 \quad \text{when } s = t_i \\
&= \sum_{i=k+1}^{k+l} \epsilon_{is}\lambda_{ijs}^0 u_{ij}^0 (1-u_{ij}^0)(1-\lambda_{ijs}^0) + \sum_{i=k+l+1}^{n} v_{ij}^0 (1-v_{ij}^0) \big(\epsilon_s \lambda_{ijs}^0\big)^2 \quad \text{when } s < t_i
\end{aligned}$$

Regarding the regression coefficients, we have;

$$\begin{aligned}
\frac{\partial^2 \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{1jb}^0 \partial \beta_{1ja}} &= \sum_{i=k+1}^{l+k} (1-\sum_{s=1}^{t_i} \epsilon_{is}\lambda_{ijs}) x_{ia} \frac{\partial u_{ij}^0}{\partial \beta_{1jb}} + \sum_{i=k+1}^{n} \sum_{s=1}^{t_i} (-\epsilon_{is}\lambda_{ijs}^0) x_{ia} \frac{\partial v_{ij}^0}{\partial \beta_{1jb}} \\
&= \sum_{i=k+1}^{l+k} (1-\sum_{s=1}^{t_i} \epsilon_{is}\lambda_{ijs}) x_{ia} u_{ij}^0 (1-u_{ij}^0)(1-\lambda_{ijs}^0) x_{ib} \\
&+ \sum_{i=k+1}^{n} \sum_{s=1}^{t_i} (\epsilon_{is}\lambda_{ijs}^0) x_{ia} x_{ib} u_{ij}^0 (1-u_{ij}^0)(1-\lambda_{ijs}^0) \sum_{s=1}^{t_i} \epsilon_{is}\lambda_{ijs} \\
&= \sum_{i=k+1}^{l+k} (1-\sum_{s=1}^{t_i} \epsilon_{is}\lambda_{ijs}) x_{ia} u_{ij}^0 (1-u_{ij}^0) x_{ib} (1-\lambda_{ijs}^0) \\
&+ \sum_{i=k+1}^{n} v_{ij}(1-v_{ij}) x_{ia} x_{ib} \sum_{s=1}^{t_i} \epsilon_{is}\lambda_{ijs}^0 \sum_{s=1}^{t_i} \epsilon_{is}\lambda_{ijs}
\end{aligned}$$

The missing information component regarding Cov($\beta_{1ja}; \beta_{1jb}$) is;

$$\begin{aligned}
\left.\frac{\partial^2 \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{1jb} \partial \beta_{1ja}^0}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} &= \sum_{i=k+1}^{l+k} (1-\sum_{s=1}^{t_i} \epsilon_{is}\lambda_{ijs}^0) x_{ia} u_{ij}^0 (1-u_{ij}^0) x_{ib} (1-\lambda_{ijs}^0) \\
&+ \sum_{i=k+1}^{n} v_{ij}^0 (1-v_{ij}^0) x_{ia} x_{ib} \left(\sum_{s=1}^{t_i} \epsilon_{is}\lambda_{ijs}^0 \right)^2
\end{aligned}$$

We now focus on $\boldsymbol{\gamma}$. We begin by establishing these results;

$$\begin{aligned}
\frac{\partial v_{ij}^0}{\partial \gamma_{1ja}} &= \frac{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il}+\pi_J S_{iJ}\big)\pi_{ij}(1-\pi_{ij})S_{ij}x_{ia}-\pi_{ij}S_{ij}\big(\pi_{ij}(1-\pi_{ij})S_{ij}-\pi_{ij}\sum_{l\neq j}^{J}\pi_{il}S_{il}\big)x_{ia}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il}+\pi_J S_{iJ})^2} \\
&= \frac{\pi_{ij}S_{ij}\big\{\sum_{l\neq j}^{J}\pi_{il}S_{il}(1-\pi_{ij})+\pi_{ij}\sum_{l\neq j}^{J}\pi_{il}S_{il}\big\}x_{ia}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il}+\pi_J S_{iJ}\big)^2} = \frac{\pi_{ij}S_{ij}\times\sum_{l\neq j}^{J}\pi_{il}S_{il}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il}+\pi_J S_{iJ}\big)^2}x_{ia} \\
&= v_{ij}^0(1-v_{ij}^0)x_{ia} \\
\frac{\partial v_{ij}^0}{\partial \gamma_{1kb}} &= \frac{\big(-\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il}+\pi_{iJ}S_{iJ})\pi_{ik}\pi_{ij}S_{ij}x_{ib}-\big(\pi_{ik}(1-\pi_{ik})S_{ik}-\pi_{ik}\sum_{l\neq k}^{J}\pi_{il}S_{il}\big)\pi_{ij}S_{ij}x_{ib}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il}+\pi_{iJ}S_{iJ}\big)^2} \\
&= \frac{\pi_{ij}\pi_{ik}S_{ij}x_{ib}\big\{-\sum_{l=1}^{j}\pi_{il}S_{il}-S_{ik}+\sum_{l=1}^{J}\pi_{il}S_{il}\big\}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il}+\pi_{iJ}S_{iJ}\big)^2} = -\frac{\pi_{ij}S_{ij}}{\sum_{l=1}^{J}\pi_{il}S_{il}}\frac{\pi_{ik}S_{ik}}{\sum_{l=1}^{J}\pi_{il}S_{il}}x_{ib} \\
&= -v_{ij}^0 v_{ik}^0 x_{ib} \\
\frac{\partial u_{ij}^0}{\partial \gamma_{1ja}} &= \frac{\big(\sum_{l=1}^{J}\pi_{il}\lambda_{il}\big)\pi_{ij}(1-\pi_{ij})\lambda_{ij}x_{ia}-\pi_{ij}\lambda_{ij}\pi_{ij}(1-\pi_{ij})\lambda_{ij}x_{ia}}{\big(\sum_{l=1}^{J}\pi_{il}\lambda_{il}\big)^2} \\
&= \frac{\pi_{ij}(1-\pi_{ij})\lambda_{ij}\big\{\sum_{l\neq j}^{J}\pi_{il}\lambda_{il}\big\}x_{ia}}{\big(\sum_{l=1}^{J}\pi_{il}\lambda_{il}\big)^2} = (1-\pi_{ij})\frac{\pi_{ij}\lambda_{ij}\times\sum_{l\neq j}^{J}\pi_{il}\lambda_{il}}{\big(\sum_{l=1}^{J}\pi_{il}\lambda_{il}\big)^2}x_{ia} \\
&= (1-\pi_{ij}^0)u_{ij}^0(1-u_{ij}^0)x_{ia} \\
\frac{\partial u_{ij}^0}{\partial \gamma_{1kb}} &= \frac{-\pi_{ij}\lambda_{ij}(1-\pi_{ik})\pi_{ik}\lambda_{ik}x_{ib}}{\big(\sum_{l=1}^{J}\pi_{il}\lambda_{il}\big)^2} = -u_{ij}^0 u_{ik}^0(1-\pi_{ik}^0)x_{ib}
\end{aligned}$$

We again re-index the subjects as before, and write the "complete" data score function as follows;

$$\frac{\partial Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial \gamma_{1ja}} = \sum_{i=1}^{k}(d_{ij}-\pi_{ij})x_{ia} + \sum_{i=k+1}^{k+l}(u_{ij}-\pi_{ij})x_{ia} + \sum_{i=k+l+1}^{n}(v_{ij}-\pi_{ij})x_{ia}$$

Then,

$$\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial\gamma_{1ja}\partial\gamma_{1ja}^0} = \sum_{i=k+1}^{k+1} \frac{\partial u_{ij}}{\partial\gamma_{1ja}} x_{ia} + \sum_{i=k+l+1}^{n} \frac{\partial v_{ij}}{\partial\gamma_{1ja}} x_{ia} = \sum_{i=k+1}^{n} (1-\pi_{ij}^0) u_{ij}^0 (1-u_{ij}^0) x_{ia}^2$$
$$+ \sum_{i=k+l+1}^{n} v_{ij}^0 (1-v_{ij}^0) x_{ia}^2$$
$$\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial\gamma_{1kb}\partial\gamma_{1ja}^0} = \sum_{i=k+1}^{k+l} \frac{\partial u_{ij}}{\partial\gamma_{1jb}} x_{ia} + \sum_{i=k+l+1}^{n} \frac{\partial v_{ij}}{\partial\gamma_{1jb}} x_{ia} = -\sum_{i=k+1}^{k+l} u_{ij}^0 (1-\pi_{ij}^0) u_{ik}^0 x_{ia} x_{ib}$$
$$- \sum_{i=k+l+1}^{n} v_{ij}^0 v_{ik}^0 x_{ia} x_{ib}$$

Thus, the missing information components in respect of $\mathrm{Var}(\gamma_{1ja})$ and $\mathrm{Cov}(\gamma_{1kb};\gamma_{1ja})$ are given by

$$\left.\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial^2\gamma_{1ja}\partial\gamma_{1ja}^0}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \sum_{i=k+1}^{n} (1-\pi_{ij}^0) u_{ij}^0 (1-u_{ij}^0) x_{ia}^2 + \sum_{i=k+l+1}^{n} v_{ij}^0 (1-v_{ij}^0) x_{ia}^2$$
$$\left.\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial\gamma_{1kb}\partial\gamma_{1ja}^0}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = -\sum_{i=k+1}^{k+l} u_{ij}^0 (1-\pi_{ij}^0) u_{ik}^0 x_{ia} x_{ib} - \sum_{i=k+l+1}^{n} v_{ij}^0 v_{ik}^0 x_{ia} x_{ib}$$

Appendix B

The discrete time mixture regression model:- standard errors for the CIF

A Complete Case

In this section, we develop the CIF estimate standard errors also via the delta method where;

$$F_j(t|\boldsymbol{x}, \boldsymbol{\theta}) = \pi_j(\boldsymbol{x}, \boldsymbol{\gamma})(1 - S_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j))$$

Suppose that $\boldsymbol{\theta}^0$ is the MLE of $\boldsymbol{\theta}$ in the M-Step at convergence. Let $\boldsymbol{\eta}_j^{\beta}(t) = (\eta_{j1}^{\beta}, \ldots, \eta_{jt}^{\beta})^T$, where $\eta_{js}^{\beta} = \beta_{0js} + \boldsymbol{x}^T \boldsymbol{\beta}_{1j}$ and $\boldsymbol{\eta}^{\gamma} = (\eta_1^{\gamma}, \ldots, \eta_{J\text{-}1}^{\gamma})^T$, where $\eta_j^{\gamma} = \gamma_{0j} + \boldsymbol{x}^T \boldsymbol{\gamma}_{1j}$ Furthermore, let

$$\begin{aligned}
\frac{\partial F_j(t|\boldsymbol{x}, \boldsymbol{\theta})}{\partial \eta_j^{\gamma}} &= \pi_j(\boldsymbol{x}, \boldsymbol{\gamma})(1 - \pi_j(\boldsymbol{x}, \boldsymbol{\gamma}))(1 - S_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j)) \\
\frac{\partial F_j(t|\boldsymbol{x}, \boldsymbol{\theta})}{\partial \eta_{js}^{\gamma}} &= \pi_j(\boldsymbol{x}, \boldsymbol{\gamma})\lambda_j(s|\boldsymbol{x}, \boldsymbol{\beta}_j) S_j(t|\boldsymbol{x}, \boldsymbol{\beta}_j)
\end{aligned}$$

Then,

$$\begin{aligned}
\text{Var}(\hat{F}_j(t|\boldsymbol{x}, \boldsymbol{\theta})) = &\sum_{m=1}^{J\text{-}1} \sum_{n=1}^{J\text{-}1} \frac{\partial F_j(t|\boldsymbol{x}, \boldsymbol{\theta})}{\partial \eta_m^{\gamma}} \frac{\partial F_j(t|\boldsymbol{x}, \boldsymbol{\theta})}{\partial \eta_n^{\gamma}} \text{Cov}(\eta_m^{\gamma}; \eta_n^{\gamma}) \\
&+ \sum_{s=1}^{t} \sum_{k=1}^{t} \frac{\partial F_j(t|\boldsymbol{x}, \boldsymbol{\theta})}{\partial \eta_{js}^{\beta}} \frac{\partial F_j(t|\boldsymbol{x}, \boldsymbol{\theta})}{\partial \eta_{jk}^{\beta}} \text{Cov}(\eta_{js}^{\beta}; \eta_{jk}^{\beta}) \bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}^0}
\end{aligned}$$

where,

$$\text{Cov}(\eta_{js}^{\beta}; \eta_{jk}^{\beta}) = \text{Cov}(\beta_{0js}; \beta_{0jk}) + \sum_{w}^{p} x_w \big(\text{Cov}(\beta_{0js}; \beta_{1jw}) + \text{Cov}(\beta_{0jk}; \beta_{1jw})\big) + \boldsymbol{x}^T \text{Var}(\boldsymbol{\beta}_{1j}) \boldsymbol{x}$$

and,

$$\text{Cov}(\eta_m^{\gamma}; \eta_n^{\gamma}) = \text{Cov}(\gamma_{0m}; \gamma_{0n}) + \sum_{w}^{p} x_w \big(\text{Cov}(\gamma_{0m}; \gamma_{1nw}) + \text{Cov}(\gamma_{0n}; \gamma_{1mw})\big) + \boldsymbol{x}^T \text{Cov}(\boldsymbol{\gamma}_{1m}; \boldsymbol{\gamma}_{1n}) \boldsymbol{x}$$

Here, again, the statistical packages produce complete data standard errors for $\boldsymbol{\theta}^0$. We adjust these standard errors via the Oakes (1999) method which is given by

$$\mathcal{I}_0(\boldsymbol{\theta}^0) = -\left(\frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} + \frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}^0\partial\boldsymbol{\theta}^T}\right)\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$$

$$\text{Observed data information } = -\Big(\text{Complete data information } + \text{Missing information }\Big)$$

Re-order the data such the first r subjects are uncensored and the next $(n-r)$ are censored. Let $S_{ij} = S_j(s|\boldsymbol{x}_i;\boldsymbol{\beta}_j^0)$, $\pi_{ij} = \pi_j(\boldsymbol{x}_i;\boldsymbol{\gamma}^0)$, and $\lambda_{ijs} = \lambda_j(s|\boldsymbol{x}_i;\boldsymbol{\beta}_j^0)$. We begin by establishing the following results;

$$\begin{aligned}
\frac{\partial v_{ij}^0}{\partial \gamma_{1ja}^0} &= \frac{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il} + \pi_J S_{iJ}\big)\pi_{ij}(1-\pi_{ij})S_{ij}x_{ia}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il} + \pi_J S_{iJ}\big)^2} \\
&- \frac{\big(\pi_{ij}S_{ij}\big(\pi_{ij}(1-\pi_{ij})S_{ij} - \pi_{ij}\sum_{l\neq j}^{J\text{-}1}\pi_{il}S_{il} - \pi_{ij}(1-\pi_{ij})S_{iJ} + \pi_{ij}S_{iJ}\sum_{l\neq j}^{J\text{-}1}\pi_{il}\big)\big)x_{ia}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il} + \pi_J S_{iJ}\big)^2} \\
&= \frac{\pi_{ij}S_{ij}\times\sum_{l\neq j}^{J}\pi_{il}S_{il}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il} + \pi_J S_{iJ}\big)^2}x_{ia} \\
&= v_{ij}^0(1-v_{ij}^0)x_{ia} \\
\frac{\partial v_{ij}^0}{\partial \gamma_{1kb}^0} &= -\frac{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il} + \pi_J S_{iJ}\big)\pi_{ij}\pi_{ik}S_{ij}x_{ib}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il} + \pi_J S_{iJ}\big)^2} \\
&- \frac{\big(\pi_{ij}S_{ij}\big(\pi_{ik}(1-\pi_{ik})S_{ij}\big) - \pi_{ik}\sum_{l\neq k}^{J\text{-}1}\pi_{il}S_{il} - \pi_{ik}(1-\pi_{ik})S_{iJ} + \pi_{ik}S_{iJ}\sum_{l\neq k}^{J\text{-}1}\pi_{il}\big)\big)x_{ib}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il} + \pi_J S_{iJ}\big)^2} \\
&= -\frac{\pi_{ij}S_{ij}}{\sum_{l=1}^{J}\pi_{il}S_{il}}\frac{\pi_{ik}S_{ik}}{\sum_{l=1}^{J}\pi_{il}S_{il}}x_{ib} \\
&= -v_{ij}^0 v_{ik}^0 x_{ib} \\
\frac{\partial v_{ij}^0}{\partial \beta_{0js}^0} &= \frac{\big(\sum_{l=1}^{J}\pi_{il}S_{il}\big)\pi_{ij}(-\lambda_{ijs}S_{ij}) - \pi_{ij}S_{ij}(\pi_{ij}(-\lambda_{ijs}S_{ij}))}{\big(\sum_{l=1}^{J}\pi_{il}S_{il}\big)^2} \\
&= \frac{\big\{\sum_{l=1}^{J}\pi_{il}S_{il} - \pi_{ij}S_{ij}\big\}\big(\pi_{ij}(-\lambda_{ijs}S_{ij}\big)}{\big(\sum_{l=1}^{J}\pi_{il}S_{il}\big)^2} = -\frac{\big(\sum_{l\neq j}^{J}\pi_{il}S_{il}\big)(\lambda_{ijs}\pi_{ij}S_{ij})}{\big(\sum_{l=1}^{J}\pi_{il}S_{il}\big)^2} \\
&= -v_{ij}^0(1-v_{ij}^0)\lambda_{ijs}
\end{aligned}$$

$$\frac{\partial \underline{v}_{ij}^0}{\partial \beta_{1jb}^0} = \frac{\big(\sum_{l=1}^{J}\pi_{il}S_{il}\big)\pi_{ij}(-\sum_{s=1}^{t_i}\lambda_{ijs})S_{ij}x_{ib} - \pi_{ij}S_{ij}\big(\pi_{ij}(-\sum_{s=1}^{t_i}\lambda_{ijs})S_{ij}\big)x_{ib}}{\big(\sum_{l=1}^{J}\pi_{il}S_{il}\big)^2}$$
$$= -\frac{\big(\sum_{l\neq j}^{J}\pi_{il}S_{il}\big)\big(\sum_{s=1}^{t_i}\lambda_{ijs}\big)\pi_{ij}S_{ij}}{\big(\sum_{l=1}^{J}\pi_{il}S_{il}\big)^2}x_{ib} = -v_{ij}^0(1-v_{ij}^0)x_{ib}\sum_{s=1}^{t_i}\lambda_{ijs}$$

Let $\acute{\pi}_{ij} = \pi_j(\boldsymbol{x}_i;\boldsymbol{\gamma})$, and $\acute{\lambda}_{ijs} = \lambda_j(s|\boldsymbol{x}_i;\boldsymbol{\beta}_j)$. The Q score functions are given by

$$\frac{\partial Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial \gamma_{1ja}} = \sum_{i=1}^{r}(d_{ij}-\acute{\pi}_{ij})x_{ia} + \sum_{i=r+1}^{n}(v_{ij}^0-\acute{\pi}_{ij})x_{ia}$$
$$\frac{\partial Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{0js}} = \sum_{i=1}^{r} d_{ijs}-\acute{\lambda}_{ijs} + \sum_{i=r+1}^{n} d_{ijs}-v_{ij}^0\acute{\lambda}_{ijs}$$
$$\frac{\partial Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{1ja}} = \sum_{i=1}^{r}\sum_{s=1}^{t_i}(d_{ijs}-\acute{\lambda}_{ijs})x_{ia} + \sum_{i=r+1}^{n}\sum_{s=1}^{t_i}(d_{ijs}-v_{ij}^0\acute{\lambda}_{ijs})x_{ia}$$

Since $\frac{\partial v_{ij}^0}{\partial \beta_{0js}} = 0$ for $s < t_i$, then,

$$\frac{\partial Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{0jk}^0\partial \beta_{0js}} = \text{-}\sum_{i=r+1:s\geq t_i}^{n}\frac{\partial v_{ijs}^0}{\partial \beta_{0jk}^0}\acute{\lambda}_{ijs} = \text{-}\sum_{i=r+1:s\geq t_i}^{n} v_{ij}^0(1-v_{ij}^0))\lambda_{ijk}\acute{\lambda}_{ijs}$$

Thus,

$$\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial \gamma_{1jb}^0\partial \gamma_{1ja}} = \sum_{i=r+1}^{n}\frac{\partial v_{ij}^0}{\partial \gamma_{1jb}^0}x_{ia} = \sum_{i=r+1}^{n} v_{ij}^0(1-v_{ij}^0)x_{ia}x_{ib}$$
$$\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial \gamma_{1kb}^0\partial \gamma_{1ja}} = \sum_{i=r+1}^{n}\frac{\partial v_{ij}^0}{\partial \gamma_{1jb}^0}x_{ia} = -\sum_{i=r+1}^{n} v_{ij}^0 v_{ik}^0 x_{ia}x_{ib}$$
$$\frac{\partial Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{0jk}^0\partial \beta_{0js}} = \text{-}\sum_{i=r+1:s\geq t_i}^{n}\frac{\partial v_{ijs}^0}{\partial \beta_{0jk}^0}\acute{\lambda}_{ijs} = \text{-}\sum_{i=r+1:s\geq t_i}^{n} v_{ij}^0(1-v_{ij}^0))\lambda_{ijk}\acute{\lambda}_{ijs}$$
$$\frac{\partial Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{1jb}^0\partial \beta_{1ja}} = -\sum_{i=r+1}^{n}\frac{\partial v_{ij}^0}{\partial \beta_{1jb}^0}x_{ia}\sum_{s=1}^{t_i}\acute{\lambda}_{ijs} = \sum_{i=r+1}^{n} v_{ij}^0(1-v_{ij}^0)x_{ib}x_{ia}\sum_{s=1}^{t_i}\lambda_{ijs}\sum_{s=1}^{t_i}\acute{\lambda}_{ijs}$$
$$\frac{\partial Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{1jb}^0\partial \beta_{0js}} = -\sum_{i=r+1}^{n}\frac{\partial v_{ij}}{\partial \beta_{1jb}^0}\acute{\lambda}_{ijs} = \sum_{i=r+1}^{n} v_{ij}^0(1-v_{ij}^0)\left(\sum_{s=1}^{t_i}\lambda_{ijs}\right)x_{ib}\acute{\lambda}_{ijs}$$

The missing information components are, therefore;

$$\begin{aligned}
\left.\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial\gamma^0_{1jb}\partial\gamma_{1ja}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} &= \sum_{i=r+1}^{n} v^0_{ij}(1-v^0_{ij})x_{ia}x_{ib} \\
\left.\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial\gamma^0_{1kb}\partial\gamma_{1ja}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} &= -\sum_{i=r+1}^{n} v^0_{ij}v^0_{ik}x_{ia}x_{ib} \\
\left.\frac{\partial Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}^0_j)}{\partial\beta^0_{0jk}\partial\beta_{0js}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} &= \text{-} \sum_{i=r+1:s\geq t_i}^{n} v^0_{ij}(1-v^0_{ij}))\lambda_{ijk}\lambda_{ijs} \\
\left.\frac{\partial Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}^0_j)}{\partial\beta^0_{1jb}\partial\beta_{1ja}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} &= \sum_{i=r+1}^{n} v^0_{ij}(1-v^0_{ij})x_{ib}x_{ia}\left(\sum_{s=1}^{t_i}\lambda_{ijs}\right)^2 \\
\left.\frac{\partial Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}^0_j)}{\partial\beta^0_{1jb}\partial\beta_{0js}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} &= \sum_{i=r+1}^{n} v^0_{ij}(1-v^0_{ij})\left(\sum_{s=1}^{t_i}\lambda_{ijs}\right)x_{ib}\lambda_{ijs}
\end{aligned}$$

B Missing Failure Causes

Suppose that $\boldsymbol{\theta}^0$ is the MLE of $\boldsymbol{\theta}$ in the M-Step at convergence. We rely on the multivariate delta method to determine the variance for the CIF estimates, where;

$$\hat{F}_j(t|\boldsymbol{x},\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\pi}}_j(\boldsymbol{x},\hat{\boldsymbol{\gamma}}_j)(1-\hat{S}_j(t|\boldsymbol{x},\hat{\boldsymbol{\beta}}_j))$$

Let $\boldsymbol{\eta}_{\beta_j} = (\eta_{\beta_{j1}}, \eta_{\beta_{j1}}, \ldots, \eta_{\beta_{jt}})^T$ and $\boldsymbol{\eta}_{\gamma} = (\eta_{\gamma_1}, \eta_{\gamma_2}, \ldots, \eta_{\gamma_{(J-1)}})^T$, where $\eta_{\beta_{js}} = \beta_{0js} + \boldsymbol{x}^T\boldsymbol{\beta}_{1j}$ and $\eta_{\gamma_j} = \gamma_{0j} + \boldsymbol{x}^T\boldsymbol{\gamma}_{1j}$, then $V(\hat{F}_j(t|\boldsymbol{x};\hat{\boldsymbol{\theta}}))$ can be written as;

$$V(\hat{F}_j(t|\boldsymbol{x};\boldsymbol{\theta})) = \left(\frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\eta}_\gamma}, \frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\eta}_{\beta_j}}\right)^T \begin{bmatrix} \mathrm{V}(\boldsymbol{\eta}_\gamma) & \mathbf{0} \\ \mathbf{0} & \mathrm{V}(\boldsymbol{\eta}_{\beta_j}) \end{bmatrix} \left.\left(\frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\eta}_\gamma}, \frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\eta}_{\beta_j}}\right)\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$$

since $\dfrac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\eta}_\beta\partial\boldsymbol{\eta}_\gamma} = 0$, where $\boldsymbol{\eta}_\beta = (\boldsymbol{\eta}^T_{\beta_1}, \boldsymbol{\eta}^T_{\beta_2}, \ldots, \boldsymbol{\eta}^T_{\beta_J})^T$. The partial derivatives are given as

$$\begin{aligned}
\frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial\eta_{\gamma_j}} &= (1-\pi_j(\boldsymbol{x};\boldsymbol{\gamma}))\pi_k(\boldsymbol{x};\boldsymbol{\gamma})Q_j(t|\boldsymbol{x};\boldsymbol{\beta}_j) \\
\frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial\eta_{\beta_{js}}} &= -\pi_j(\boldsymbol{x};\boldsymbol{\gamma})\lambda_j(s|\boldsymbol{x};\boldsymbol{\beta}_j)Q_j(t|\boldsymbol{x};\boldsymbol{\beta}_j)
\end{aligned}$$

where, $Q_j(t|\boldsymbol{x};\boldsymbol{\beta}_j) = 1 - S_j(t|\boldsymbol{x};\boldsymbol{\beta}_j)$. Thus,

$$\mathrm{Var}(\hat{F}_j(t|\boldsymbol{x})) = \sum_{m=1}^{J\text{-}1}\sum_{n=1}^{J\text{-}1} \frac{\partial F_j(t|\boldsymbol{x})}{\partial \eta_{\gamma_m}} \frac{\partial F_j(t|\boldsymbol{x})}{\partial \eta_{\gamma_n}} \mathrm{Cov}(\eta_{\gamma_m};\eta_{\gamma_n}) + \sum_{s=1}^{t}\sum_{k=1}^{t} \frac{\partial F_j(t|\boldsymbol{x})}{\partial \eta_{\beta_{js}}} \frac{\partial F_j(t|\boldsymbol{x})}{\partial \eta_{\beta_{jk}}} \mathrm{Cov}(\eta_{\beta_{js}};\eta_{\beta_{jk}})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$$

where,

$$\mathrm{Cov}(\eta_{\beta_{js}};\eta_{\beta_{jk}}) = \mathrm{Cov}(\beta_{0js};\beta_{0jk}) + \sum_{w}^{p} x_w\big(\mathrm{Cov}(\beta_{0js};\beta_{1jw}) + \mathrm{Cov}(\beta_{0jk};\beta_{1jw})\big) + \boldsymbol{x}^T\mathrm{Var}(\boldsymbol{\beta}_{1j})\boldsymbol{x}$$

and,

$$\mathrm{Cov}(\eta_{\gamma_m};\eta_{\gamma_n}) = \mathrm{Cov}(\gamma_{0m};\gamma_{0n}) + \sum_{w}^{p} x_w\big(\mathrm{Cov}(\gamma_{0m};\gamma_{1nw}) + \mathrm{Cov}(\gamma_{0n};\gamma_{1mw})\big) + \boldsymbol{x}^T\mathrm{Cov}(\boldsymbol{\gamma}_{1m};\boldsymbol{\gamma}_{1n})\boldsymbol{x}$$

Since we have applied an EM algorithm, the components of $\mathrm{V}(\boldsymbol{\eta}_\gamma)$ and $\mathrm{V}(\boldsymbol{\eta}_{\beta_j})$ require adjustments. Again, we apply the Oakes (1999) approach which is given by

$$\mathcal{I}_y(\boldsymbol{\theta}^0) = -\frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} + \frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}\partial{\boldsymbol{\theta}^0}^T} \tag{B.1}$$

We assume that the subjects have been re-indexed so that the first k subjects are uncensored, the next l subjects have missing failure causes and the remaining $(n-k-l)$ are censored. For convenience, let $S_{ij} = S_j(s|\boldsymbol{x};\boldsymbol{\beta}_j^0)$, $\pi_{ij} = \pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma}^0)$, and $\lambda_{ijs} = \lambda_j(s|\boldsymbol{x};\boldsymbol{\beta}_j^0)$. The complete data Q score functions can be written as;

$$\frac{\partial Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j)}{\partial\beta_{0js}} = \sum_{i=1}^{k} d_{ijs} - \lambda_{ijs} + \sum_{i=k+1}^{l+k} d_{i*s}u_{ij}^0 - u_{ij}^0\lambda_{ijs} + \sum_{i=k+l+1}^{n} d_{ijs} - v_{ij}^0 v_{is}\lambda_{ijs}$$

$$\frac{\partial Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial\beta_{1ja}} = \sum_{i=1}^{k}\sum_{s=1}^{t_i}(d_{ijs} - \epsilon_{is}\lambda_{ijs})x_{ia} + \sum_{i=k+1}^{l+k}\sum_{s=1}^{t_i}(d_{i*s}u_{ij}^0 - u_{ij}^0\lambda_{ijs})x_{ia} + \sum_{i=k+l+1}^{n}\sum_{s=1}^{t_i}(d_{ijs} - v_{ij}^0\lambda_{ijs})x_{ia}$$

Since $d_{i*s} = 1$ for $s = t_i$ and 0 otherwise, then, we can, therefore, re-write; $\frac{\partial Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j)}{\partial\beta_{0js}}$ and $\frac{\partial Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial\beta_{1ja}^0}$ in respect of u_{ij}^0as;

$$\frac{\partial Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j)}{\partial \beta_{0js}} = \sum_{i=1}^{k} d_{ijs} - \lambda_{ijs} + \sum_{i=k+1}^{k+l} (1-\lambda_{ijs})u_{ij}^0 + \sum_{i=k+l+1}^{n} d_{ijs} - v_{ij}^0\lambda_{ijs} \quad \text{when } s = t_i$$

$$= \sum_{i=1}^{k} d_{ijs} - \epsilon_{is}\lambda_{ijs} - \sum_{i=k+1}^{k+l} \lambda_{ijs}u_{ij}^0 + \sum_{i=k+l+1}^{n} d_{ijs} - v_{ij}^0\lambda_{ijs} \quad \text{when } s \neq t_i$$

$$\frac{\partial Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{1ja}} = \sum_{i=1}^{k}\sum_{s=1}^{t_i}(d_{ijs} - \lambda_{ijs})x_{ia} + \sum_{i=k+1}^{l+k} (1 - \sum_{s=1}^{t_i} \lambda_{ijs})x_{ia}u_{ij}^0$$

$$+ \sum_{i=k+l+1}^{n}\sum_{s=1}^{t_i}(d_{ijs} - \underline{u}_{ij}^0\lambda_{ijs})x_{ia}$$

Using the chain rule, the partial derivatives of the pseudo-variables are given by

$$\frac{\partial v_{ij}^0}{\partial \beta_{0js}^0} = \frac{\partial v_{ij}^0}{\partial S_{ij}}\frac{\partial S_{ij}}{\partial \beta_{0js}^0} = \frac{\big(\sum_{l=1}^{J} \pi_{ij}S_{ij}\big)\pi_{ij} - \pi_{ij}S_{ij}(\pi_{ij})}{\big(\sum_{j=1}^{J} \pi_{ij}S_{ij}\big)^2} \times -S_{ij}\lambda_{ijs} = -v_{ij}^0(1-v_{ij}^0)\lambda_{ijs}^0$$

$$\frac{v_{ij}^0}{\partial \beta_{1ja}^0} = \frac{\partial v_{ij}^0}{\partial S_{ij}}\frac{\partial S_{ij}}{\partial \beta_{ija}^0} = -\frac{\big(\sum_{l=1}^{J} \pi_{il}S_{il}\big)\pi_{ij} - \pi_{ij}S_{ij}\pi_{ij}}{(\sum_{l=1}^{J} \pi_{il}S_{il})^2}(S_{ij}x_{ia})\big(\sum_{s=1}^{t_i} \lambda_{ijs}\big)$$

$$= -v_{ij}^0(1-v_{ij}^0)x_{ia}\sum_{s=1}^{t_i} \lambda_{ijs}^0$$

$$\frac{\partial u_{ij}^0}{\partial \beta_{0js}^0} = \frac{\partial u_{ij}^0}{\partial \lambda_{ij}}\frac{\partial \lambda_{ij}}{\partial \beta_{0js}^0} = \frac{\big(\sum_{l=1}^{J} \pi_{ij}\lambda_{ij}\big)\pi_{ij} - \pi_{ij}\lambda_{ij}(\pi_{ij})}{\big(\sum_{j=1}^{J} \pi_{ij}\lambda_{ij}\big)^2} \times \lambda_{ij}(1-\lambda_{ij}) = (1-\lambda_{ij})u_{ij}^0(1-u_{ij}^0)$$

$$\frac{\partial u_{ij}^0}{\partial \beta_{1ja}^0} = \frac{\partial u_{ij}^0}{\partial \lambda_{ij}}\frac{\partial \lambda_{ij}}{\partial \beta_{ija}^0} = -\frac{\big(\sum_{l=1}^{J} \pi_{il}\lambda_{il}\big)\pi_{ij} - \pi_{ij}\lambda_{ij}\pi_{ij}}{(\sum_{l=1}^{J} \pi_{il}\lambda_{il})^2}\lambda_{ij}(1-\lambda_{ij})x_{ia} = (1-\lambda_{ij})u_{ij}^0(1-u_{ij}^0)x_{ia}$$

It then follows that;

$$\frac{\partial^2 Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{0js}^0 \partial \beta_{0js}} = \sum_{i=k+1}^{k+l} (1-\lambda_{ijs})u_{ij}^0(1-u_{ij}^0)(1-\lambda_{ijs}^0)$$

$$+ \sum_{i=k+l+1}^{n} \lambda_{ijs}v_{ij}^0(1-v_{ij}^0)\lambda_{ijs}^0 \quad \text{when } s = t_i$$

$$= \sum_{i=k+1}^{k+l} \lambda_{ijs}u_{ij}^0(1-u_{ij}^0)(1-\lambda_{ijs}^0) + \sum_{i=k+l+1}^{n} \lambda_{ijs}v_{ij}^0(1-v_{ij}^0)\lambda_{ijs}^0 \quad \text{when } s \neq t_i$$

since $\frac{\partial v_{ij}}{\partial \beta_{0js}} = 0$ when $s > t_i$. Therefore the missing information corresponding to β_{0js} is;

$$\begin{aligned}
\left.\frac{\partial^2 Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{0js}^0 \partial \beta_{0js}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} &= \sum_{i=k+1}^{k+l} (1-\lambda_{ijs}^0) u_{ij}^0 (1-u_{ij}^0)(1-\lambda_{ijs}^0) \\
&+ \sum_{i=k+l+1}^{n} v_{ij}^0 (1-v_{ij}^0) \big(\lambda_{ijs}^0\big)^2 \quad \text{when } s = t_i \\
&= \sum_{i=k+1}^{k+l} \lambda_{ijs}^0 u_{ij}^0 (1-u_{ij}^0)(1-\lambda_{ijs}^0) + \sum_{i=k+l+1}^{n} v_{ij}^0 (1-v_{ij}^0) \big(\lambda_{ijs}^0\big)^2 \quad \text{when } s < t_i
\end{aligned}$$

Regarding the regression coefficients, we have;

$$\begin{aligned}
\frac{\partial^2 Q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{1jb}^0 \partial \beta_{1ja}} &= \sum_{i=k+1}^{l+k} (1-\sum_{s=1}^{t_i} \lambda_{ijs}) x_{ia} \frac{\partial u_{ij}^0}{\partial \beta_{1jb}} + \sum_{i=k+1}^{n} \sum_{s=1}^{t_i} (-\epsilon_{is} \lambda_{ijs}^0) x_{ia} \frac{\partial v_{ij}^0}{\partial \beta_{1jb}} \\
&= \sum_{i=k+1}^{l+k} (1-\sum_{s=1}^{t_i} \lambda_{ijs}) x_{ia} u_{ij}^0 (1-u_{ij}^0)(1-\lambda_{ijs}^0) x_{ib} \\
&+ \sum_{i=k+1}^{n} \sum_{s=1}^{t_i} (\lambda_{ijs}^0) x_{ia} x_{ib} u_{ij}^0 (1-u_{ij}^0)(1-\lambda_{ijs}^0) \sum_{s=1}^{t_i} \epsilon_{is} \lambda_{ijs} \\
&= \sum_{i=k+1}^{l+k} (1-\sum_{s=1}^{t_i} \lambda_{ijs}) x_{ia} u_{ij}^0 (1-u_{ij}^0) x_{ib} (1-\lambda_{ijs}^0) \\
&+ \sum_{i=k+1}^{n} v_{ij} (1-v_{ij}) x_{ia} x_{ib} \sum_{s=1}^{t_i} \lambda_{ijs}^0 \sum_{s=1}^{t_i} \epsilon_{is} \lambda_{ijs}
\end{aligned}$$

The missing information component regarding $\text{Cov}(\beta_{1ja}; \beta_{1jb})$ is;

$$\begin{aligned}
\left.\frac{\partial^2 \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial \beta_{1jb} \partial \beta_{1ja}^0}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} &= \sum_{i=k+1}^{l+k} (1-\sum_{s=1}^{t_i} \lambda_{ijs}^0) x_{ia} u_{ij}^0 (1-u_{ij}^0) x_{ib} (1-\lambda_{ijs}^0) \\
&+ \sum_{i=k+1}^{n} v_{ij}^0 (1-v_{ij}^0) x_{ia} x_{ib} \bigg(\sum_{s=1} \lambda_{ijs}^0\bigg)^2
\end{aligned}$$

We now focus on $\boldsymbol{\gamma}$. We begin by establishing these results;

$$\frac{\partial v_{ij}^0}{\partial \gamma_{1ja}} = \frac{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il} + \pi_J S_{iJ}\big)\pi_{ij}(1-\pi_{ij})S_{ij}x_{ia} - \pi_{ij}S_{ij}\big(\pi_{ij}(1-\pi_{ij})S_{ij} - \pi_{ij}\sum_{l\neq j}^{J}\pi_{il}S_{il}\big)x_{ia}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il} + \pi_J S_{iJ})^2}$$

$$= \frac{\pi_{ij}S_{ij}\big\{\sum_{l\neq j}^{J}\pi_{il}S_{il}(1-\pi_{ij}) + \pi_{ij}\sum_{l\neq j}^{J}\pi_{il}S_{il}\big\}x_{ia}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il} + \pi_J S_{iJ}\big)^2} = \frac{\pi_{ij}S_{ij}\times\sum_{l\neq j}^{J}\pi_{il}S_{il}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il} + \pi_J S_{iJ}\big)^2}x_{ia}$$

$$= v_{ij}^0(1-v_{ij}^0)x_{ia}$$

$$\frac{\partial v_{ij}^0}{\partial \gamma_{1kb}} = \frac{\big(-\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il} + \pi_{iJ}S_{iJ})\pi_{ik}\pi_{ij}S_{ij}x_{ib} - \big(\pi_{ik}(1-\pi_{ik})S_{ik} - \pi_{ik}\sum_{l\neq k}^{J}\pi_{il}S_{il}\big)\pi_{ij}S_{ij}x_{ib}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il} + \pi_{iJ}S_{iJ}\big)^2}$$

$$= \frac{\pi_{ij}\pi_{ik}S_{ij}x_{ib}\big\{-\sum_{l=1}^{j}\pi_{il}S_{il} - S_{ik} + \sum_{l=1}^{J}\pi_{il}S_{il}\big\}}{\big(\sum_{l=1}^{J\text{-}1}\pi_{il}S_{il} + \pi_{iJ}S_{iJ}\big)^2} = -\frac{\pi_{ij}S_{ij}}{\sum_{l=1}^{J}\pi_{il}S_{il}}\frac{\pi_{ik}S_{ik}}{\sum_{l=1}^{J}\pi_{il}S_{il}}x_{ib}$$

$$= -v_{ij}^0 v_{ik}^0 x_{ib}$$

$$\frac{\partial u_{ij}^0}{\partial \gamma_{1ja}} = \frac{\big(\sum_{l=1}^{J}\pi_{il}\lambda_{il}\big)\pi_{ij}(1-\pi_{ij})\lambda_{ij}x_{ia} - \pi_{ij}\lambda_{ij}\pi_{ij}(1-\pi_{ij})\lambda_{ij}x_{ia}}{\big(\sum_{l=1}^{J}\pi_{il}\lambda_{il}\big)^2}$$

$$= \frac{\pi_{ij}(1-\pi_{ij})\lambda_{ij}\big\{\sum_{l\neq j}^{J}\pi_{il}\lambda_{il}\big\}x_{ia}}{\big(\sum_{l=1}^{J}\pi_{il}\lambda_{il}\big)^2} = (1-\pi_{ij})\frac{\pi_{ij}\lambda_{ij}\times\sum_{l\neq j}^{J}\pi_{il}\lambda_{il}}{\big(\sum_{l=1}^{J}\pi_{il}\lambda_{il}\big)^2}x_{ia}$$

$$= (1-\pi_{ij}^0)u_{ij}^0(1-u_{ij}^0)x_{ia}$$

$$\frac{\partial u_{ij}^0}{\partial \gamma_{1kb}} = \frac{-\pi_{ij}\lambda_{ij}(1-\pi_{ik})\pi_{ik}\lambda_{ik}x_{ib}}{\big(\sum_{l=1}^{J}\pi_{il}\lambda_{il}\big)^2} = -u_{ij}^0 u_{ik}^0(1-\pi_{ik}^0)x_{ib}$$

We again re-index the subjects as before, and write the "complete" data score function as follows;

$$\frac{\partial Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial \gamma_{1ja}} = \sum_{i=1}^{k}(d_{ij}-\pi_{ij})x_{ia} + \sum_{i=k+1}^{k+l}(u_{ij}-\pi_{ij})x_{ia} + \sum_{i=k+l+1}^{n}(v_{ij}-\pi_{ij})x_{ia}$$

Then,

$$\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial\gamma_{1ja}\partial\gamma^0_{1ja}} = \sum_{i=k+1}^{k+1}\frac{\partial u_{ij}}{\partial\gamma_{1ja}}x_{ia} + \sum_{i=k+l+1}^{n}\frac{\partial v_{ij}}{\partial\gamma_{1ja}}x_{ia} = \sum_{i=k+1}^{n}(1-\pi^0_{ij})u^0_{ij}(1-u^0_{ij})x^2_{ia}$$
$$+ \sum_{i=k+l+1}^{n} v^0_{ij}(1-v^0_{ij})x^2_{ia}$$
$$\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial\gamma_{1kb}\partial\gamma^0_{1ja}} = \sum_{i=k+1}^{k+l}\frac{\partial u_{ij}}{\partial\gamma_{1jb}}x_{ia} + \sum_{i=k+l+1}^{n}\frac{\partial v_{ij}}{\partial\gamma_{1jb}}x_{ia} = -\sum_{i=k+1}^{k+l} u^0_{ij}(1-\pi^0_{ij})u^0_{ik}x_{ia}x_{ib}$$
$$- \sum_{i=k+l+1}^{n} v^0_{ij}v^0_{ik}x_{ia}x_{ib}$$

Thus, the missing information components in respect of $\mathrm{Var}(\gamma_{1ja})$ and $\mathrm{Cov}(\gamma_{1kb};\gamma_{1ja})$ are given by

$$\left.\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial^2\gamma_{1ja}\partial\gamma^0_{1ja}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \sum_{i=k+1}^{n}(1-\pi^0_{ij})u^0_{ij}(1-u^0_{ij})x^2_{ia} + \sum_{i=k+l+1}^{n} v^0_{ij}(1-v^0_{ij})x^2_{ia}$$
$$\left.\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial\gamma_{1kb}\partial\gamma^0_{1ja}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = -\sum_{i=k+1}^{k+l} u^0_{ij}(1-\pi^0_{ij})u^0_{ik}x_{ia}x_{ib} - \sum_{i=k+l+1}^{n} v^0_{ij}v^0_{ik}x_{ia}x_{ib}$$

Appendix C

The discrete time vertical regression model (complete case):- standard errors for the CIF

In this section, we derive the expression for standard error of the CIF estimate via the delta method, where;

$$\hat{F}_j(t|\boldsymbol{x},\hat{\boldsymbol{\theta}}) = \sum_{s=1}^{t} \hat{S}(s\text{-}1|\boldsymbol{x},\hat{\boldsymbol{\beta}})\hat{h}(s|\boldsymbol{x},\hat{\boldsymbol{\beta}})\hat{\Pi}_j(s|\boldsymbol{x},\hat{\boldsymbol{\phi}}))$$

Let $\boldsymbol{\eta}_t = (\boldsymbol{\eta}_1^T,\ldots,\boldsymbol{\eta}_t^T)^T$, where $\boldsymbol{\eta}_s = (\eta_{1s},\ldots,\eta_{J\text{-}1s})^T$ and $\eta_{ks} = \gamma_{0ks}+\boldsymbol{x}^T\boldsymbol{\gamma}_{1k}$. Furthermore, let $\boldsymbol{\zeta}_t = (\zeta_1,\ldots,\zeta_t)^T$ where $\zeta_s = \beta_{0s} + \boldsymbol{x}^T\boldsymbol{\beta}_1$. The expression for the standard error is then given by

$$\begin{aligned}
\text{Var}(F_j(t|\boldsymbol{x},\boldsymbol{\theta}) &= \begin{bmatrix} \frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\eta}_t} \\ \frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\zeta}_t} \end{bmatrix}^T \begin{bmatrix} \text{Var}(\boldsymbol{\eta}_t) & \mathbf{0} \\ \mathbf{0} & \text{Var}(\boldsymbol{\zeta}_t) \end{bmatrix} \begin{bmatrix} \frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\eta}_t} \\ \frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\zeta}_t} \end{bmatrix} \\
&= \left(\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\eta}_t}\right)^T \text{Var}(\boldsymbol{\eta}_t)\left(\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\eta}_t}\right) \\
&+ \left(\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\zeta}_t}\right)^T \text{Var}(\boldsymbol{\zeta}_t)\left(\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\zeta}_t}\right)
\end{aligned}$$

Since $\text{Cov}\big(\boldsymbol{\eta}_q;\boldsymbol{\zeta}_q\big) = \mathbf{0}$ because $\dfrac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\eta}_q\partial\boldsymbol{\zeta}_q} = 0$, see (Yu et al., 2011)

The partial derivatives of $F_j(t|\boldsymbol{x},\boldsymbol{\theta})$ w.r.t. ζ_s and η_{js} are given by

$$\begin{aligned}
\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \zeta_s} &= \lambda(s|\boldsymbol{x},\boldsymbol{\beta})\big\{S(s|\boldsymbol{x},\boldsymbol{\beta})\pi_j(s|\boldsymbol{x},\boldsymbol{\gamma}) - \big(F_j(t|\boldsymbol{x},\boldsymbol{\theta}) - F_j(s|\boldsymbol{x},\boldsymbol{\theta})\big)\big\} \\
\frac{\partial F_j(t)}{\partial \eta_{js}} &= \pi_j(s|\boldsymbol{x},\boldsymbol{\gamma})(1-\pi_j(s|\boldsymbol{x},\boldsymbol{\gamma})) \\
\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \eta_{ks}} &= -\pi_j(s|\boldsymbol{x},\boldsymbol{\gamma})\pi_k(s|\boldsymbol{x},\boldsymbol{\gamma}) \quad j \neq k
\end{aligned}$$

The expression given in (Ambrogi et al., 2009) for the cumulative incidence function is modified to give;

$$\begin{aligned}\mathrm{V}(\hat{F}_j(t|\boldsymbol{x},\hat{\boldsymbol{\theta}})) &= \sum_{s=1}^{t}\sum_{l=1}^{t}\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \zeta_s}\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \zeta_l}\mathrm{Cov}(\zeta_s;\zeta_l)\\ &+ \sum_{j=1}^{J\text{-}1}\sum_{k=1}^{J\text{-}1}\sum_{s=1}^{t}\sum_{l=1}^{t}\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \eta_{js}}\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \eta_{kl}}\mathrm{Cov}(\eta_{js};\eta_{kl})\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\end{aligned}$$

Appendix D

The nonparametric discrete time mixture vertical model:- standard errors for the CIF

In this section we determine the standard errors for $\hat{F}_{uj}(t)$, where;

$$\hat{F}_{uj}(t) = \sum_{s=1}^{t} \hat{S}_u(s\text{-}1)\hat{h}_u(s)\hat{\Pi}_{uj}(s)$$

Suppose that $\boldsymbol{\theta}_2^0$ is the MLE for $\boldsymbol{\theta}_2$ at convergence of the EM algorithm. Borrowing the expression for the variance of $\hat{F}_j(t)$ when data is modelled with the ordinary nonparametric vertical model of Chapter 4, we write the expression for the variance of $\hat{F}_{uj}(t)$ as;

$$\begin{aligned}\mathrm{V}(\hat{F}_{uj}(t)) = &\sum_{s=1}^{t} \mathrm{V}(S_u(s\text{-}1)h_u(s)\Pi_{uj}(s)) \\ &+ 2\sum_{s=1}^{q}\sum_{k=s+1}^{q} \mathrm{Cov}(S_u(s\text{-}1)h_u(s)\Pi_{uj}(s), S_u(k\text{-}1)h_u(k)\Pi_{uj}(k))\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\end{aligned} \qquad (.1)$$

where $\boldsymbol{\theta}^0 = (\boldsymbol{\Pi}_u^{0^T}, \boldsymbol{h}_u^{0^T}; p^0)^T$ is the estimate for $\boldsymbol{\theta}$ at convergence of the EM Algorithm bearing in mind that $\boldsymbol{\Pi}_u$ was not estimated via the algorithm and $\boldsymbol{\Pi}_u = \boldsymbol{\Pi}$. We use the multivariate delta method to determine both $\mathrm{V}(S_u(s\text{-}1)h_u(s)\Pi_{uj}(s))$as well as $\mathrm{Cov}(S_u(s\text{-}1)h_u(s)\Pi_{uj}(s), S_u(k\text{-}1)h_u(k)\Pi_{uj}(k))$. Note that $\mathrm{Cov}(h_u(m), h_u(n)) = 0$ $m \neq n$.

Furthermore, $\dfrac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\Pi}_u\partial\boldsymbol{h}_u} = 0 \implies \text{Cov}(\boldsymbol{h}_u, \boldsymbol{\Pi}_u)$. Therefore;

$$\begin{aligned}
\text{V}(S_u(s\text{-}1)h_u(s)\Pi_{uj}(s)) &= (S_u(s\text{-}1)h_u(s)\Pi_{uj}(s))^2\sum_{l=1}^{s\text{-}1}\frac{\text{V}(h_u(l))}{(1-h_u(l))^2} \\
&+ (S_u(s\text{-}1)\Pi_{uj}(s))^2\text{V}(h_u(s)) + (S_u(s\text{-}1)h_u(s))^2\text{V}(\Pi_{uj}(s)) \\
&= (S_u(s\text{-}1)h_u(s)\Pi_j(s))^2\Bigg(\sum_{l=1}^{s\text{-}1}\frac{\text{V}(h_u(l))}{(1-h_u(l))^2} + \frac{\text{V}(h_u(s))}{(h_u(s))^2} \\
&+ \frac{\text{V}(\Pi_{uj}(s))}{(\Pi_{uj}(s))^2}\Bigg) \qquad (.2)
\end{aligned}$$

$$\begin{aligned}
\text{Cov}(S_u(s\text{-}1)h_u(s)\Pi_{uj}(s), S_u(k\text{-}1)h_u(k)\Pi_{uj}(k)) &= (S_u(s\text{-}1)h_u(s)\Pi_{uj}(s)S_u(s\text{-}1)h_u(s)\Pi_{uj}(s)) \\
&\times \sum_{l=1}^{s\text{-}1}\frac{\text{V}(h_u(l))}{(1-h_u(l))^2} \\
&- (S(s\text{-}1)\Pi_{uj}(s), S(k\text{-}1)h_u(k)\Pi_{uj}(k))\frac{\text{V}(h_u(s))}{1-h_u(s)} \\
&= (S(s\text{-}1)h_u(s)\Pi_{uj}(s)S(k\text{-}1)h_u(k)\Pi_{uj}(k)) \\
&\times \left(\sum_{l=1}^{s\text{-}1}\frac{\text{V}(h_u(l))}{(1-h_u(l))^2} - \frac{\text{V}(h_u(s))}{(h_u(s)(1-h_u(s)))}\right)
\end{aligned} \qquad (.3)$$

Since $\text{V}(\boldsymbol{\theta}_2^0)$ is computed from "complete data", we need to scale it down to observed data variance. We use the method suggested by Oakes (1999);

Observed data Information = Complete data information − Missing data Information

$$\begin{aligned}
\mathcal{I}_o &= \mathcal{I}_c - \mathcal{I}_m \\
&= -\frac{\partial^2 Q(\boldsymbol{\theta}_2|\boldsymbol{\theta}_2^0)}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^T} - \frac{\partial Q(\boldsymbol{\theta}_2|\boldsymbol{\theta}_2^0)}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^{0T}}\Bigg|_{\boldsymbol{\theta}_2=\boldsymbol{\theta}_2^0}
\end{aligned}$$

We begin by re-writing $Q(\boldsymbol{h}|\boldsymbol{h}^{(r)})$ as;

$$Q(\boldsymbol{h}|\boldsymbol{h}^0) = \sum_{s=1}^{q} d_{(s)}\log h_u(s) + (R_{(s)}^0 - d_{(s)})\log(1-h_u(s)) \qquad (.4)$$

where $R_{(s)}^0 = \sum_{l=s}^{q} d_{(l)} + c_{(l)}y^0(l)$ and; $y^0(s) = \dfrac{p^0S_u^0(s)}{p^0S_u^0(s) + (1-p^0)}$ The complete data information is given by

$$\frac{\partial^2 Q(\boldsymbol{h}|\boldsymbol{h}^0)}{\partial^2 h_u(s)} = -\frac{d_{(s)}}{(h_u(s))^2} - \frac{R_{(s)}^0 - d_{(s)}}{(1-h_u(s))^2}$$

To determine the "missing information" we require the following result in respect of $y^0(s)$;

$$\frac{\partial y^0(s)}{\partial h_u^0(s)} = \frac{\partial y^0(s)}{\partial S_u^0(s)}\frac{\partial S_u^0(s)}{\partial h_u^0(s)} = \frac{(p^0 S_u^0(s) + (1-p^0)p^0 - p^0 S_u^0(s))p^0}{(p^0 S_u^0(s) + (1-p^0))^2} \times (-)\frac{S_u^0(t)}{1-h_u^0(s)}$$
$$= -\frac{y^0(s)(1-y^0(s))}{1-h_u^0}$$

The missing information is given by

$$\left.\frac{\partial^2 Q(\boldsymbol{h}|\boldsymbol{h}^0)}{\partial h_u(s)\partial h_u^0(s)}\right|_{\boldsymbol{\theta}_2=\boldsymbol{\theta}_2^0} = -\frac{1}{1-h_u(s)}\left.\frac{\partial R_{(s)}^0}{\partial h_u^0(s)}\right|_{\boldsymbol{\theta}_2=\boldsymbol{\theta}_2^0} = \frac{1}{1-h_u^0(s)}\sum_{l=s}^{q} c_{(l)}\frac{y^0(l)(1-y^0(l))}{1-h_u^0(l)}$$

The observed data information for $\hat{h}_u(s)$ is, therefore, given by

$$\mathcal{I}_o(\hat{h}_u(s)) = \frac{d_{(s)}}{(h_u^0(s))^2} + \frac{R_{(s)}^0 - d_{(s)}}{(1-h_u^0(s))^2} - \frac{1}{1-h_u^0(s)}\sum_{l=s}^{q} c_{(l)}\frac{y^0(l)(1-y^0(l))}{1-h_u^0(l)}$$

Using the delta method again, the variance of the population CIF estimate is given by

$$\mathrm{V}(\hat{F}_j(t)) = \mathrm{V}(\hat{p})(\hat{F}_{uj}(t))^2 + (\hat{p})^2\mathrm{V}(\hat{F}_{uj}(t))$$

Again, we need to adjust $\mathrm{V}(\hat{p})$. The complete data information for $\hat{p}$ is given by

$$\frac{\partial^2 Q(p|p^0)}{\partial^2 p} = \sum_{i=1}^{n} -\frac{g_i^0}{p^2} - \frac{1-g_i^0}{(1-p)^2}$$

Suppose that the first r subjects are failures and the remaining $(n-r)$ are censored, then;

$$\frac{\partial Q(p|p^0)}{\partial p} = \sum_{i=1}^{r}\frac{d_i}{p} - \frac{1-d_i}{(1-p)} + \sum_{i=r+1}^{n}\frac{y_i^0(t_i)}{p} - \frac{1-y_i^0(t_i)}{(1-p)}$$

The derivative of $y_i^0(t_i)$ with respect to p^0 is given by

$$\frac{\partial y_i^0(t_i)}{\partial p} = \frac{(p^0 S_u^0(t_i) + (1-p^0))S_u^0(t_i) - p^0 S_u^0(t_i)S_u^0(t_i)}{(pS_u(t_i) + (1-p))^2} = \frac{y_i^0(t_i)(1-y_i^0(t_i))(t_i)}{p^0}$$

Thus;

$$\frac{\partial^2 Q(p|p^0)}{\partial p \partial p^0} = \sum_{i=r+1}^{n}\frac{1}{p}\frac{\partial y_i^0(t_i)}{\partial p^0} - \frac{1}{1-p}\frac{\partial y_i^0(t_i)}{\partial p^0} = \sum_{i=r+1}^{n}\frac{y_i^0(t_i)(1-y_i^0(t_i))}{p(1-p)p^0}$$

The observed information in respect of $\hat{p}$ is given by

$$\mathcal{I}_0(\hat{p}) = \sum_{i=1}^{n}\frac{g_i^0}{(p^0)^2} + \frac{1-g_i^0}{(1-p^0)^2} - \sum_{i=r+1}^{n}\frac{y_i^0(t_i)(1-y_i^0(t_i))}{(1-p^0)(p^0)^2}$$

Bibliography

Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories. *Sociological Methodology*, 13:61–98.

Ambrogi, F., Biganzoli, E., and Boracchi, P. (2009). Estimating Crude Cumulative incidences through Multinomial Logit Regression on Discrete Cause Specific Hazard . *Computational Statistics and Data Analysis*, 53:2767–2779.

Anderson, J., Goetghebeur, E., and Ryan, L. (1996). Missing cause of death information in the analysis of survial data. *Statistics in Medicine*, 16:2191–2201.

Bakoyannis, G., Siannis, F., and Touloumi, G. (2010). Modelling competing risks data with missing cause of failure. *Statistics in Medicine*. DOI: 10.1002/sim.4133.

Beger, M. and Schmid, M. (2018). Semiparametric regression for discrete time-to-event data. *Statistical Modelling*, 18:1–24.

Begg, C. B. and Gray, R. (1984). Calculation of polytomous logistic regression parameters using individualized regressions. *Biometrika*, 71:11–18.

Berger, M., Schmid, M., Schmitz-Valckenberg, Welchowski, T., and Bayermann, S. (2020). Subdistribution hazard models for competing risks in discrete time. *Biostatistics*, 21:449–466.

Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Joumal of the American Statistical Association*, 11:15–44.

Blanche, L., Proust-Lima, C., Loubere, L., Berr, L., Dartigues, J., and Jacqmin-Gadda, H. (2015). Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics*, 71:102–113.

Boag, J. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical, Society B*, 11:15–53.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78:1–3.

Canty, A. and Ripley, B. (2017). *boot: Bootstrap R (S-Plus) Functions.* R package version 1.3-19.

Choi, C. (2002). *On the Mixture Models in Survival Analysis with Competing Risks and Covariates.* PhD thesis, Hong Kong polytechnic University.

Cleric, R., Giraldo, A., and Meggiolaro, S. (2014). The determinants of academic outcomes in acompeting risks approach: evidence from italy. *Studies in Higher Education*, 22:84–99.

Cox, D. (1972). Regression Models and Life Tables. *Journal of Royal Statistical Society B*, 34:187–220.

Croissant, Y. and Graves, S. (2020). *Ecdat: Data Sets for Econometrics.* R package version 0.3-7.

Davis, T. P. and Lawrance, A. J. (1989). The likelihood for competing risk survival analysis. *Board of the Foundation of the Scandinavian Journal of Statistics*, 16:23–28.

Dempters, A., Liard, N., and Rubin, D. (1977). Journal of the royal statistical society. series b (methodological). *Journal of the Royal Statistical Society. Series B*, 39:1–38.

Dewanji, A. (1992). A note on a test for competing risks with missing failure. *Biometrica*, 79:855–857.

Dinse, G. E. (1982). Nonparametric estimation for partially-complete time and of failure data. *Biometrics*, 38:417–431.

Dirick, L., Claesken, G., and Baesens, B. (2015). An akaike information criterion for multiple event mixture cure models. *European Journal of Operational Research*, 241:449–457.

Escarala, G. and Bowater, R. J. (2008). Fitting a Semi-Parametric Mixture Model for Competing Risks in Survival Data. *Communications in Statistics - Theory and Methods*, 37:277–293.

Escarela, Francis, B., and Soothill, K. (2000). Competing Risks, Persistence, and Desistance in analyzing Recidivism. *Journal of Quantitative Criminology*, 16:385–414.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 38:1041–1046.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94:496–509.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent,. *Journal of Statistical Software*, 33:1–22.

Gaynor, J. J., Feuer, E. J., Tan, C. C., Wu, D. H., Little, C. R., Straus, D. J., Clarkson, B. D., and Brennan, M. F. (1993). On the use of cause-specific failure and conditional failure probabilities. examples from clinical oncology data. *American Statistical Association*, 88:400–409.

Goetghebeur, E. and Ryan, L. (1990). A modified logrank test for competing risks with missing failure type. *Biometrika*, 77:207–211.

Goetghebeur, E. and Ryan, L. (1995). Analysis of competing risks survival data when some failure types are missing. *Biometrika*, 82:821–834.

Graf, R. S. E. and Schumacher, M. (2008). Quantifying the predictive performance of prognostic models for censored survival data with timedependent covariates. *Biometrics*, 64:603–610.

Haller, B. (2014). *The Analysis of Competing Risks Data with a Focus on Estimation of Cause-Specific and Subdistribution Hazard Ratios from a Mixture Model.* PhD thesis, Ludwig -Maximiliäns-Universitat München.

Han, A. and Hausman, J. (1990). Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics*, 5:1–28.

Heyard, R., Timsit, J., and Held, L. (2019). Validation of discrete time-to-event prediction models in the presence of competing risks. *Biometrical Journal*. DOI: 10.1002/bimj.201800293.

Holford, T. (1980). The Analysis of rates and of survivorship using log-linear models. *Biometrics*, 36:299–305.

Klein, J. P. and Anderson, P. K. (2005). Regression modelling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, 61:223–229.

Kuk, A. Y. C. and Chen, C. (1992). A Mixture Model Combining Logistic Regression with Proportional Hazards Regression. *Biometrika Trust*, 79:531–541.

Laird, N. and Oliver, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *Biometrica*, 76:231–240.

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica*, 4:939–959.

Larson, M. G. and Dinse, G. E. (1985). A Mixture Model for the Regression Analysis of Competing Risks Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 34:201–211.

Lau, B., Cole, S. R., and Gange, S. J. (2011). Parametric mixture models to evaluate and summarize hazard ratios in the presence of competing risks with time-dependent hazards and delayed entry. *Stat Med*, 30:654–665.

Lau, B., Cole, S. R., Moore, R. D., and Gange, S. J. (2008). Evaluating competing adverse and beneficial outcomes using a mixture model. *Stat Med*, 27:4313–.

Lee, M., Feuer, E., and Fine, J. (2018). On the analysis of discrete time competing risks data. *Biometrics*. DOI:10.1111/biom.12881.

Li, L., Greene, T., and Hu, B. (2018). A simple method to estimate the time-dependent receiver operating characteristic curve and the area under the curve with right censored data. *Statistical Methods in Medical Research*, 27:2264–2278.

Little, R. J. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: John and Wiley Sons, 1st edition.

Lu, K. and Tsiatis, A. (2001). Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics*, 57:1191–1197.

Maller, R. and Zhou, X. (2002). Analysis of parametric models for competing risks. *Statistica Sinica*, 12:725–750.

Maller, R. A. and Zhou, S. (1995). Testing for the presence of immune or cured individuals in censored survival data. *Biometrics*, 51:1197–1205.

McCall, B. P. (1996). Unemployment insurance rules, joblessness, and part-time work. *Econometric Society*, 64:647–682.

Möst, M., Pößecker, W., and Tutz, G. (2016). Variable selection for discrete competing risks models. *Quality & Quantity*, 50:1589–1610.

Narendranathan, W., Nickell, S., and Stern, J. (1985). Unemployment benefits revisited. *Economic Journal*, 95:307–329.

Ng, S. K. and McLachan, G. J. (2003). An EM-based semi-parametric mixture model approach to the regression analysis of competing-risks data. *Stat Med*, 22:1097–1111.

Nickell, S. (1979). Estimating the probability of leaving unemployment. *Econometrica*, 47:1249–66.

Nicolaie, M., Taylor, J., and Legrand, C. (2018). Vertical modeling: analysis of competing risks data with a cure fraction. *Lifetime Data Analysis*.

Nicolaie, M., van Houwelingen, H., and Putter, H. (2010). Vertical modeling: A pattern mixture approach for competing risks modeling. *Statistics in Medicine*, 29:1190–1205.

Nicolaie, M., van Houwelingen, H., and Putter, H. (2015). Vertical modelling: Analysis of competing risks data with missing causes of failure. *Statistical Methods in Medicine*, 29:1190–1205.

Oakes, D. (1999). Direct calculation of the information matrix via the em algorithm. *Journal of Royal Statistical Society, Series B*, 61:479–482.

Peng, Y. and Taylor, J. M. G. (2014). Cure models. In Klein, J., van Houwelingen, H., Ibrahim, J. G., and Scheke, T. H., editors, *Handbook of Survival Analysis*, chapter 6, pages 113–114. Chapman & Hall, BocaRaton, FL, USA.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978). The Analysis of Failure Times in the Presence of Competing Risks. *Biometrics,*, 34:541–554.

Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys.* Wiley: New York, 1987.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63:581–592.

Saha, P. and Heagerty, P. (2010). Time-dependent predictive accuracy in the presence of competing risks. *Biometrics*, 66:999–1011.

Scheike, T. H. and Gerds, M. (2008). Predicting cumulative incidence probabilty by direct binomial regression. *Biometrika*, 1(95):205–220.

Schmid, M., Tu, G., and Welchowski, T. (2018). Discrimination measures for discrete time-to-event predictions. *Econometric Statitics*, 7:153–164.

Schoop, R., Bayersmann, J., Schumacher, M., and Binder, H. (2011a). Quantifying the predictive accuracy of time-to-event models in the presence of competing risks. *Biometrical Journal,*, 53:88–112.

Schoop, R., Schumacher, M., and Graf, E. (2011b). Measures of prediction error for survival data with longitudinal covariates. *biometrical Journal*, 53:275–293.

Scott, M. A. and Kennedy, B. (2005). Pitfalls in Pathways: Some Perspectives on Competing Risks Event History Analysis in Education Research. *Journal of Educational and Behavioral Statistics*, 30:413–442.

Singer, J. and Willet, J. (2003). *Applied Longitudinal Data analysis.* Oxford University Press, New York.

Steele, F., Goldstein, H., and Browne, W. (2004). A general multilevel multistate competing risks model for event history data, with an application to a study of contraceptive use dynamics. *Statistical Modelling*, 4:1548–1556.

Steyerberg, E. (2009). *Clinical prediction models.* New Yor: Springer.

Tutz, G. and Schmid, M. (2016). *Modeling Discrete Time-to-Event Data.* Springer.

Vallejos, C. and Steel, M. (2017). Bayesian survival modelling of university outcomes. *Journal of the Royal Statistical Society. Series A*, 180:613–631.

Welchowski, T. and Schmid, M. (2019). *discSurv: Discrete Time Survival Analysis.* R package version 1.4.1.

Yee, T. (2010). The vgam package for categorical data analysis. *Journal of Statistical Software*, 32. http://www.jstatsoft.org/.

Yu, B., Tiwari, R. C., and Feuer, E. J. (2011). Estimating the personal cure of cancer patients using population-based grouped cancer survival data. *Stat Methods Res*, 20:261–274.

Zheng, P. H. Y. (2005). Survival model predictive accuracy and roc curves. *Biometrics*, 61:92–105.

Zhiping, T. (2011). *On Competing Risks data with covariates and Long-term Survivors.* PhD thesis, Hong Kong Polytechnic University.