

Profile Likelihood Estimation Applied To The Semi-parametric Models With Survival Data

by

Khandoker Akib Mohammad

A thesis

submitted to the Victoria University of Wellington

in fulfilment of the

requirements for the degree of

Doctor of Philosophy

in Statistics

Victoria University of Wellington

2021

Acknowledgements

At first, I would like to acknowledge my honor and gratitude to almighty Allah for enabling me to accomplish this thesis paper successfully.

Next, my reverent gratefulness and indebtedness go to my respectable supervisor, Dr. Yuichi Hirose, for his beneficial guidance and imperative supervision. His continuous encouragement, important suggestions and necessary corrections helped me selecting the content of this study. With his experienced and constructive thoughts and ideas, directions and sincere guidance, I have been able to conduct this thesis paper in time successfully.

I express my warm gratitude to my co-supervisors Dr. Budhi Surya and Dr. Yuan Yao for helping me from the very early stage to the conclusion of this thesis. They have always been available to answer my questions. Their vast expertise in survival analysis and statistical modelling have been a crucial contribution to this thesis.

I would like to express my very great appreciation to the three examiners, Prof. Thomas Lumley, Dr. Eleni-Rosalina Andrinopoulou and Dr. Nokuthaba Sibanda, for their encouragement and their critical comments about this thesis.

I would like to acknowledge Victoria University of Wellington (VUW) for the financial support through Victoria Doctoral Scholarship. I am con-

IV

fidant that a PhD in Statistics from VUW will help me fortify my research career in the field of statistics.

I am grateful to all the academics and staff from the School of Mathematics and Statistics at VUW for supporting me when I needed it.

A special thanks to my dear officemates Lindsay Morris, Kien Tran, Erandi Nanayakkara Halloluwaage, Lingyu Li, Kyuhan Kim and all my friends around the world who have always believed that I could do it.

I owe my deepest gratitude to my parents (Khandoker Golam Mohammad and Anju Ara Begum) who gave me the opportunity of an education and have supported me throughout my life. I am specially grateful to my beloved parents and my brother (Khandoker Asif Mohammad) for their unconditional love and support encouragement throughout my entire study period in New Zealand.

Last but not the least, I would like to thank my wife Tanjila Tabassum for her love and constant support during the process of this thesis. I feel lucky to have her in my life and cannot imagine it without her.

Khandoker Akib Mohammad

Abstract

In this thesis, we have investigated the efficiency of profile likelihood in the estimation of parameters from the Cox Proportional Hazards (PH) cure model and joint model of longitudinal and survival data. For the profile likelihood approach in the joint model of longitudinal and survival data, Hsieh et al. (2006) stated “No distributional or asymptotic theory is available to date, and even the standard errors (SEs), defined as the standard deviations of the parametric estimators, are difficult to obtain”. The reason behind this difficulty is the estimator of baseline hazard which involves implicit function in the profile likelihood estimation (Hirose and Liu, 2020). Hence finding the estimated SE of the parametric estimators from the Cox PH cure model and joint model using profile likelihood approach is a great challenge. Therefore, bootstrap method has been suggested to get the estimated standard errors while using the profile likelihood approach (Hsieh et al., 2006).

To solve the difficulty, we have expanded the profile likelihood function directly without assuming the derivative of the profile likelihood score function and obtain the explicit form of the SE estimator using the profile likelihood score function. Our proposed alternative approach gives us not only analytical understanding of the profile likelihood estimation, but also provides closed form formula to compute the standard error of the profile likelihood maximum likelihood estimator in terms of profile likelihood score function. To show the advantage of our proposed approach in medi-

cal and clinical studies, we have analysed the simulated and real-life data, and compared our results with the output obtained from the `smcure`, `JM` (method: 'Cox-PH-GH') and `joineRML` R-packages. The outputs suggest that the bootstrap method and our proposed approach have provided similar and comparable results. In addition, the average computation times of our approach are much less compared to the above mentioned R-packages.

Acronyms

AIDS: Acquired Immune Deficiency Syndrome

CLT: Central Limit Theorem

ECOG: Eastern Cooperative Oncology Group

EM: Expectation Maximization

FS: Fisher Scoring

GEE: Generalized Estimating Equations

GH: Gauss Hermite

GLMM: Generalized Linear Mixed Model

HIV: Human Immunodeficiency Virus

KM: Kaplan-Meier

LLN: Law of Large Number

LME: Linear Mixed Effects

MCEM: Monte Carlo Expectation Maximization

MLE: Maximum Likelihood Estimator

MSE: Mean Square Error

NCTTG: North Central Cancer Treatment Group

NPMLE: Non-parametric Maximum Likelihood Estimator

NR: Newton-Raphson

PH: Proportional Hazards

PSA: Prostate Specific Antigen

RFS: Relapse Free Survival

SE: Standard Error

Contents

1	Introduction	1
1.1	Survival Data	1
1.2	Cox Proportional Hazards (PH) model	3
1.3	Profile Likelihood Approach	3
1.4	Cure Model	4
1.4.1	Cox PH Cure Model	6
1.5	Longitudinal Data	8
1.6	Joint Model of Longitudinal and Survival Data	9
1.7	Motivation of the Research	11
1.8	Research Objectives	17
1.9	Organization of the research	18
2	Asymptotic Normality of MLE	21
2.1	Finite Dimensional Parameter	21
2.2	Finite Dimensional Nuisance Parameter	24
2.2.1	Score Function	25
2.2.2	The Efficient Score Function	25
2.2.3	The Efficient Influence Function	26
2.3	Infinite Dimensional Nuisance Parameter	27
2.3.1	Score Function	28
2.3.2	The Efficient Score Function	29

3	Profile Likelihood	31
3.1	Preface	31
3.2	Profile Likelihood Approach	32
3.3	Efficiency of Profile Likelihood	33
4	Cox PH Model	35
4.1	Estimation of Regression Parameter	36
4.1.1	Estimate Regression Parameters using Partial Likelihood	37
4.1.2	Estimate Regression Parameters using Profile Likelihood	38
4.2	Efficiency of Profile Likelihood	39
4.3	Appendix	41
5	Efficient Estimation For Cox PH Cure Model	47
5.1	Introduction	47
5.2	Cox PH Cure Model	50
5.3	Estimation: Profile Likelihood with EM Algorithm	52
5.3.1	Baseline Hazard Estimation and Profile Likelihood Function	54
5.3.2	The EM Algorithm	56
5.3.3	Asymptotic Normality of the MLE	56
5.3.4	Score Functions	58
5.4	Simulation Study	59
5.5	Application to ECOG Data	63
5.6	Discussion	64
6	Theorems and Lemmas for Cox PH Cure Model	75
6.1	Theorem- 6.1 with proof	76
6.2	Lemma- 6.1 with Proof	77
6.3	Lemma- 6.2 with Proof	81
6.4	Theorem- 6.2 with proof	82

<i>CONTENTS</i>	XI
6.5 Lemma- 6.3 with proof	83
6.6 Theorem- 6.3 with proof	87
6.7 Efficient Score Function using Projection Theory	90
7 Efficient Estimation For The Joint Model	95
7.1 Introduction	95
7.2 Joint Model of Longitudinal and Survival Data	99
7.3 Estimation: Profile Likelihood with EM Algorithm	100
7.3.1 Baseline Hazard Estimation	102
7.3.2 Profile Likelihood Function	103
7.3.3 The EM Algorithm	104
7.3.4 Asymptotic Normality of the MLE	105
7.4 Score Functions	107
7.5 Application to Real Life Data	109
7.5.1 Application to AIDS Data	109
7.5.2 Application to PBC Data	113
7.6 Simulation Study	115
7.7 Discussion	118
8 Theorems and Lemmas for Joint Model	129
8.1 Theorem-8.1 with proof	130
8.2 Lemma-8.1 with Proof	132
8.3 Theorem- 8.2 with proof	132
8.4 Lemma- 8.3 with proof	133
8.5 Theorem- 8.3 with proof	133
8.6 Efficient Score Function using Projection Theory	133
9 Conclusion	139
9.1 Limitation and future research	143

Chapter 1

Introduction

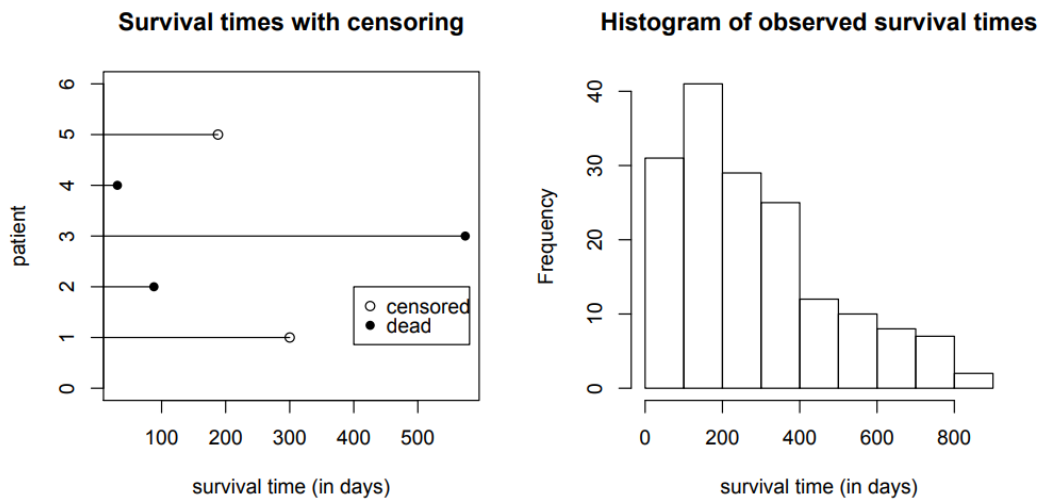
1.1 Survival Data

Survival data or time-to-event data refer times to occur an event of interest such as time to failure of a machine or time to infection of a disease (Yu, 2018). In most of the medical or clinical studies, it is seen that the event of interest is only observed for a proportion of individuals and for others, the event time is unobserved i.e., event time is greater than the last available follow-up time. This phenomenon is considered as censoring which is very common in survival data (Kleinbaum and Klein, 2011; Sullivan, 2012).

Survival analysis (deals with time-to-event data which can be days, weeks, months or years from the beginning of follow-up of the individuals) is a statistical method widely used in medical literature that explores the time period from a certain point until the occurrence of the event of interest (Allison, 2010; Bewick et al., 2004; Cox and Oakes, 1984; Putter et al., 2007; Southern et al., 2006). In medical setup, an event may be death, occurrence of disease, relapse from remission or any destined experience of interest that may occur to a subject (Andersen and Keiding, 2006; Keiding, 2007; Lawless, 2003).

For example, we can consider lung cancer data from the North Central Cancer Treatment Group (NCTTG) where the event of interest is the death of a patient. This data contain survival times (times to deaths) of patients with advanced lung cancer (Loprinzi et al., 1994). Figure 1.1 shows the survival times of five randomly selected patients where event occurred for three patients and the rest are censored. Moreover, the histogram of all observed survival times are provided to show the distribution of observed survival times (Yu, 2018).

Figure 1.1: Survival times of randomly selected patients from NCTTG data (left side) and histogram of observed survival times for all patients (right side)



In regression models for survival analysis, we are interested in modelling and determining the relationship between survival time and covariates. We know that the parameters in parametric and non-parametric models are defined in the finite-dimensional and infinite-dimensional space respectively. However, the semi-parametric models are widely used in survival analysis as they contain both finite dimensional and infinite dimensional parameters (Andersen and Keiding, 2006).

1.2 Cox Proportional Hazards (PH) model

In classical survival analysis, Cox PH model is very popular semi parametric model which has attracted attention for decades (Fox and Weisberg, 2018). In Cox PH model, the baseline hazard is an unknown function of survival time (infinite dimensional parameter) and the regression coefficient is an unknown vector (finite dimensional parameter). The Cox PH model can be expressed as

$$\lambda(t|Z) = \lambda(t)e^{\beta'Z}, \quad (1.1)$$

where $Z = (Z_1, \dots, Z_p)'$ represents a set of covariates and $\beta = (\beta_1, \dots, \beta_p)'$ is a vector of unknown regression coefficients. This model gives an expression for the hazard at time t for an individual with a given set of explanatory variables denoted by the Z . Moreover, the baseline hazard function, $\lambda(t)$ is unspecified which implies that the distribution of lifetime random variable is unknown. To estimate the regression parameters, Cox (1972) proposed a partial likelihood approach to avoid the estimation of $\lambda(t)$ (Ahmed et al., 2007; Lewis, 2016). Cox PH model assumes that the hazards ratio is constant over time (Bewick et al., 2004), therefore, it is important to check the proportional hazard assumption (Persson, 2002). More details on Cox PH model are given in Chapter-4.

1.3 Profile Likelihood Approach

Profile likelihood estimation technique is a familiar methodology in the presence of an infinite-dimensional nuisance parameter. Moreover, this approach is very popular because of reducing the infinite-dimensional estimation problem to a finite-dimensional one (Murphy and Van der Vaart, 2000). For example, semi-parametric maximum likelihood estimator (MLE) from different case control studies have been studied using the profile likelihood approach (Scott and Wild, 1997, 2001).

From equation (1.1), apart from the Cox PH partial likelihood, the regression parameters can also be estimated through profile likelihood where we profile out the baseline hazard $\lambda(t)$ using non-parametric maximum likelihood estimator (NPMLE) approach. Let us assume $\lambda(t)$ has (possible) non-zero values on the observed time t_i :

$$\lambda(t_i) = \lambda_i \geq 0 ; i = 1, 2, \dots, n$$

and $\lambda(t) = 0$ for the rest of time (Murphy and Van der Vaart, 2000). Under this assumption, the cumulative baseline hazard function, $\Lambda(t) = \int_0^t \lambda(s)ds$ is replaced with the following sum which considers the step function with jumps at all event times

$$\Lambda(t) = \sum_{i=1}^n \lambda_i 1\{t_i \leq t\}.$$

Suppose β is fixed but Λ can vary, so we can estimate Λ by maximizing the log-likelihood function $l(\beta, \Lambda)$ with respect to Λ as

$$\hat{\Lambda}_\beta = \arg \max_{\Lambda} l(\beta, \Lambda),$$

where $\hat{\Lambda}_\beta$ depends on β . Then we plugged $\hat{\Lambda}_\beta$ in $l(\beta, \Lambda)$ and construct the profile log-likelihood function $l(\beta, \hat{\Lambda}_\beta)$ and maximize it with respect to β as

$$\hat{\beta} = \arg \max_{\beta} l(\beta, \hat{\Lambda}_\beta),$$

where we profiled out Λ . More details on profile likelihood are provided in Chapter-3.

1.4 Cure Model

Although the Cox PH model is well developed for exploring time-to-event data, there are complex situations where the model is not appropriate for

instance a situation where a fraction of subjects do not experience the event of interest. These subjects are considered as cured due to their infinite survival/ event times (Amico and Van Keilegom, 2018). The cure fraction is incorporated in statistical models to analyse the effect of a treatment on the event of interest (Robinson, 2014). Survival models taking this feature are referred as cure models. Therefore, if only a subset of population is expected to experience the event of interest, then the cure model is more realistic and appropriate to use compared to the Cox PH model due to separate modelling of the probability of cure and the time to the event.

Cure models are widely used in medical studies where the main interest is in the time until the recurrence of a specific disease and it is possible that some individuals will never suffer a relapse of a given disease and therefore they are considered to be cured. From this point of view, we can say that the population is a mixture of two subpopulations: (1) cured individuals and (2) susceptible or uncured individuals (Patilea and Van Keilegom, 2017).

The works of Peng et al. (2007); Sy and Taylor (2000); Withers et al. (1995) have provided an example of head and neck cancer study where the mixture cure model has been illustrated beautifully. In this study, radiation therapy was used as a treatment for the patients with localized disease of the tonsil. The aim of the radiation therapy was to kill the cancerous cells within the tumour. Moreover, the event or the endpoint of this study was recurrence of the cancer within the tonsil region. It is well known that the majority of local recurrences (for tonsil cancer) occur within three years and rarely after five years. Therefore, after five years of follow-up, patients can be referred as locally cured. This is an ideal situation of mixture model where patients can be thought as cured or uncured at the beginning of the study by the treatment. However, the cure or uncure status can not be observed at the beginning of the study and possible to reveal itself later.

The mixture cure model was first proposed to analyze survival data with a cured fraction because of its easy-to-use mixture model structure (Boag, 1949). Moreover, it is popular for its appealing interpretation and the ease of generalization to more complex situations using parametric approaches (Berkson and Gage, 1952; Boag, 1949; Haybittle, 1965). The works of Berkson and Gage (1952); Boag (1949); Haybittle (1965) have used the log-normal and exponential models for the survival function of uncured subjects whereas incidence was modelled as a constant. Farewell (1977) showed that in the presence of covariates, the probability of cured subjects can be modelled by assuming a logistic model. Later Farewell (1982) used weibull distribution for the survival function of the uncured subjects.

1.4.1 Cox PH Cure Model

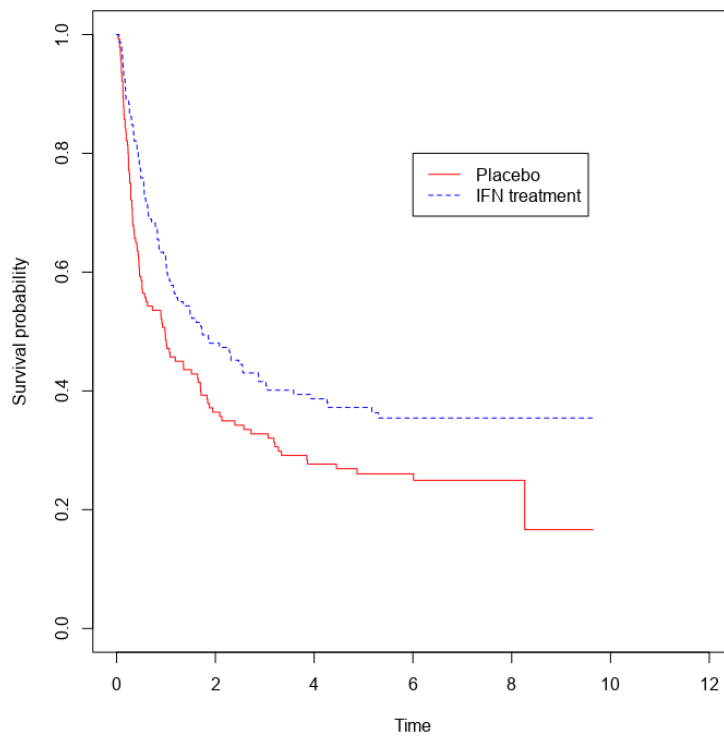
Kuk and Chen (1992) first proposed the Cox PH cure model as a semiparametric generalization of Farewell's model (Farewell, 1982) where a combination of Cox PH model and logistic regression has been used to study the survival times of uncured subjects and cure rate respectively. We know that the population is a mixture of cured and uncured subjects in the Cox PH cure model where the information on cure status are unobserved. Therefore, Expectation Maximization (EM) algorithm has been used to estimate the parameters due to treating cure status as missing data.

In this mixture model, the Cox PH partial likelihood approach developed by Cox (1972) is no longer valid to estimate the regression parameters because it is not difficult to see that a mixture of PH functions is no longer proportional (Amico and Van Keilegom, 2018; Cai, 2013; Sy and Taylor, 2000). Therefore, one of the possible solutions is to incorporate profile likelihood approach based on Breslow's estimator (Breslow, 1972) in the Cox PH cure model to estimate the parameters (Klein, 1992; Sy and Taylor, 2000,

2001). We can see the application of Cox PH cure model in various disciplines such as biology, clinical trials, genetic engineering etc. More details on Cox PH cure model have provided in Chapter-5.

For example, we can consider the Eastern Cooperative Oncology Group (ECOG) phase III clinical trial data which has been used in the works of Chen et al. (1999, 2002); Corbiere and Joly (2007); Ibrahim et al. (2001); Kirkwood et al. (1996). The purpose of the study was to investigate the effects of high dose interferon alpha-2b (IFN) regimen against the placebo as the postoperative adjuvant therapy. In this study, observed relapse-free-survival (RFS) time has defined as failure time (in years) and the recurrence of melanoma has defined as the event of interest.

Figure 1.2: Survival curves of RFS for IFN treatment and placebo group



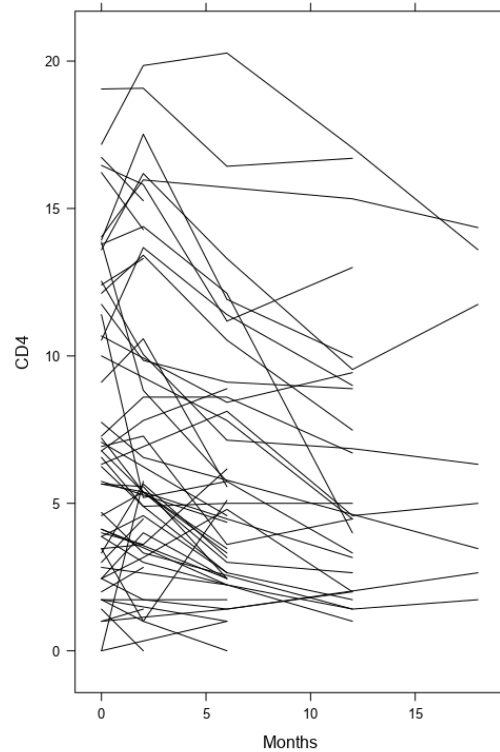
From Figure-1.2, we can see that the estimated RFS curves for IFN treatment and placebo groups where the survival probability of patients from the IFN group is higher compared to the placebo group. From the figure, it is observed that the curves for both IFN and placebo groups level off at a value substantially greater than zero after about 8 years. Therefore, it can be concluded that some patients didn't experience the recurrence of melanoma due to the presence of cured patients in both IFN and placebo group. So for this example, Cox PH cure model is more appropriate (compared to Cox PH model) to estimate the probability of recurrence of melanoma as well as the time to recurrence of melanoma.

1.5 Longitudinal Data

Longitudinal data refer to measurements on variables that are collected repeatedly over time from a group of subjects (Rizopoulos, 2012; Yu, 2018). In medical or clinical studies, longitudinal data play a prominent role to understand the development and persistence of disease. By measuring the subjects repeatedly throughout the duration of the study, longitudinal data permit the direct assessment of changes in the response variable over time which is a distinguishing feature of longitudinal studies (Rizopoulos, 2012). When patients are randomly assigned to take different drugs/ treatments at the beginning of a longitudinal study and followed over time, we can investigate both (1) cross-sectional effect (how treatment means differ at the end of the study) and (2) longitudinal effect (how treatment means change over time) .

As an example of longitudinal study, we can consider Acquired Immune Deficiency Syndrome (AIDS) data where CD4 (a type of immune cell that protects a person's immune system) is measured repeatedly to monitor disease progression of Human Immunodeficiency Virus (HIV) infected patient

Figure 1.3: CD4 trajectories of 70 randomly selected subjects



(Abrams et al., 1994). The reason for measuring CD4 repeatedly is that when a person is infected by the HIV, it attacks the immune system by reducing the CD4 cells which is one of the possible reasons for developing AIDS.

Figure 1.3 shows the longitudinal trajectories of CD4 counts of 70 randomly selected HIV-infected subjects from AIDS data. From Figure 1.3, we can see a large variation between measurements across individuals and different number of repeated measurements across individuals.

1.6 Joint Model of Longitudinal and Survival Data

In biomedical research and clinical studies, joint modelling technique is frequently used to model longitudinal and time-to-event outcomes simulta-

neously (Wulfsohn and Tsiatis, 1997). The main goal of the joint model is to study the effect of longitudinal covariates on the survival outcome. The widely used example of joint model is HIV study where infected patients are monitored until developing AIDS or death (Brown et al., 2005; Faucett et al., 2002; Wulfsohn and Tsiatis, 1997). Moreover, the condition of the infected patients immune system are regularly monitored using biomedical markers such as CD4 lymphocyte count or the estimated viral load (Guo and Carlin, 2004; Wu et al., 2012; Wulfsohn and Tsiatis, 1997). Similar example of joint model can be found in cancer studies where death or metastasis is the survival outcome (Pauker and Finkelstein, 2002; Proust-Lima and Taylor, 2009).

In cancer studies, patients need to provide repeated measurements of antibody levels or of other biomedical markers of carcinogenesis such as Prostate Specific Antigen (PSA) measurements for prostate cancer (Pauker and Finkelstein, 2002). In clinical trials, one important example of longitudinal study is Alzheimer's disease where height, weight, blood pressure measurements are repeatedly collected to determine disease aetiology (Yang and Gao, 2013). Even cardiovascular disease patients need to regularly monitor their blood pressure at regular intervals to ensure health safety (Rothwell et al., 2010).

In the above situations, separate modelling of longitudinal and time-to-event outcomes are inappropriate to discover potential disease mechanisms (Yang, 2014). We know that linear mixed effects (LME) models are well-known methods to analyse longitudinal data based on maximum and restricted maximum likelihood approach (Laird and Ware, 1982), where longitudinal measurements can be both discrete and continuous (Salkind, 2008). In addition, generalized estimating equations (GEE) used in marginal and transitional models are very popular as well for analysing longitudinal data

(Liang and Zeger, 1986). On the other hand, Cox PH model is commonly used semi-parametric model for analysing survival data (Cox, 1972).

The association between longitudinal and survival outcomes is ignored while analysing separately (Terrera et al., 2011). Therefore, to identify the potential relationships between the longitudinal measures and event outcome, an alternative framework has been introduced known as joint model of longitudinal and survival data where primary interest is to couple the survival model with a suitable model for the longitudinal measurements (Faucett and Thomas, 1996; Henderson et al., 2000; Rizopoulos, 2012; Tsiatis and Davidian, 2001; Wulfsohn and Tsiatis, 1997). In summary, the joint model is used to incorporate the association or correlation between longitudinal and survival outcomes (De Gruttola and Tu, 1994; Faucett and Thomas, 1996; Guo and Carlin, 2004; lavalley and Deguttola, 1996; Pawitan and Self, 1993; Taylor et al., 1994; Tsiatis et al., 1995; Wulfsohn and Tsiatis, 1997). More details on joint model of longitudinal and survival data are provided in Chapter-7.

1.7 Motivation of the Research

In this thesis, we have shown asymptotic normality of the maximum profile likelihood estimator (profile likelihood MLE) in the Cox PH cure model and joint model of longitudinal and survival data with EM-algorithm. Moreover, we have derived the expressions for the standard error of the estimators based on the profile likelihood score function.

While using the profile likelihood approach in the joint model of longitudinal and survival data, we profile out the baseline hazard function by plugging in the estimator of the hazard function in the likelihood function. However, the key problem is the estimator of the baseline hazard which is

an implicit function [Rizopoulos (2012); page-67]. For the joint model of longitudinal and survival data, Hsieh et al. (2006) stated “No distributional or asymptotic theory is available to date, and even the standard errors (SEs), defined as the standard deviations of the parametric estimators, are difficult to obtain”. Similar kind of problem exists for the Cox PH cure model as well. Therefore, in this research, we have considered the Cox PH cure model and the joint model of longitudinal and survival data where the common difficulty is both of them are dealing with an implicit function.

Murphy and Van der Vaart (2000) proposed a solution to the implicit function problem in the profile likelihood estimation. They have used an ‘approximate least favorable submodel’ as a mathematical device to show the asymptotic normality of profile likelihood estimator, however they didn’t express any analytical formula to calculate the estimated standard error of estimators. Therefore, they have used the numerical approximation to show asymptotic variance of the estimator as the inverse of the information matrix. According to the Corollary 3 of Murphy and Van der Vaart (2000), if $pl_n(\theta)$ is the profile log-likelihood function for θ , then the (p, q) th element of the information matrix of $\hat{\theta}$ can be estimated as

$$-\varepsilon^{-2}\{pl_n(\hat{\theta} + \varepsilon_n e_p + \varepsilon_n e_q) - pl_n(\hat{\theta} + \varepsilon_n e_p - \varepsilon_n e_q) - pl_n(\hat{\theta} - \varepsilon_n e_p + \varepsilon_n e_q) + pl_n(\hat{\theta})\}, \quad (1.2)$$

where, ε_n is a constant of order $n^{-1/2}$. Moreover, e_p and e_q are the p th and q th canonical vectors respectively.

In the Cox PH cure model, a non-parametric maximum likelihood approach along with EM algorithm has been used for the estimation procedure (Lu, 2008; Peng and Dear, 2000; Sy and Taylor, 2000, 2001). Sy and Taylor (2000, 2001) have used the profile approach along with EM algorithm to estimate the MLE of the parameters, however, for the standard error estimation, baseline hazard function has been treated as a finite dimensional parameter.

As a result, they have considered the Cox PH cure model as finite dimensional parametric model and calculate the observed information using the Hessian matrix. On the other hand, Lu (2008) has used the equation (1.2) to get the information matrix. Though the works of Peng and Dear (2000); Sy and Taylor (2000, 2001) were largely relied on the EM algorithm for MLE computation, however, the large sample properties of the estimators for the Cox PH cure model remain to be established. Fang et al. (2005) studied the theoretical properties of the estimators under Cox PH cure model, however, they didn't use the profile likelihood approach.

The profile likelihood approach has also been used in the estimation of parameters from the joint model of longitudinal and survival data. Due to the presence of random effects in the joint model of longitudinal and survival data, Cox PH partial likelihood can no longer be used to estimate the parameters. Therefore, as a possible solution, Wulfsohn and Tsiatis (1997) have proposed the profile approach in the joint model of longitudinal and survival data. Later based on the work of Murphy and Van der Vaart (2000), a non-parametric maximum likelihood estimator of the baseline hazard function has been considered i.e., baseline hazard function replaced by a step function with jumps at the event time points in the joint model of longitudinal and survival data (Hickey et al., 2018; Hsieh et al., 2006; Rizopoulos, 2012; Zeng et al., 2005). Under this situation, the estimator of baseline hazard is an implicit function.

Zeng et al. (2005); Zeng and Lin (2007a, 2010) have adopted profile likelihood approach proposed by Wulfsohn and Tsiatis (1997) in the joint model of longitudinal and survival data. However, dealing with the implicit function in the profile likelihood function has been avoided in the above mentioned joint model studies (Hirose and Liu, 2020). Zeng et al. (2005) showed the asymptotic normality of the profile likelihood estimator using the equiv-

alence between the maximum likelihood estimator and the profile likelihood estimator. Then, for the estimation of the variance of the profile likelihood estimator, they used the method proposed by Murphy and Van der Vaart (2000) given in the equation (1.2). These two papers represent mathematical proofs of the asymptotic normality of the profile likelihood estimator. There are many other papers following this line of approaches. However, they did not give an analytic method to calculate the information matrix for the profile likelihood function. As a result there is a gap between theoretical work and applied work in this area.

Based on the above discussions, we can say that it is difficult to find the estimated standard error of the profile likelihood estimator from the Cox PH cure model and joint model of longitudinal and survival data. Hence, finding the asymptotic distributions of the estimators using profile likelihood approach is a great challenge due to the presence of an implicit function (that appear in the profile likelihood estimation). Therefore, to get the estimated standard errors from the models dealing with implicit function, Hsieh et al. (2006) suggested to use bootstrap method (Efron and Tibshirani, 1994) while using the profile likelihood approach.

Our work aims to close the gap by proposing a new approach to show the asymptotic normality of the profile likelihood estimator with implicit function. This enables us to have a closed form expression for the standard error of the profile likelihood estimator. The main idea of our approach is to use the weak derivative. The score function for the profile likelihood function may not be differentiable due to the existence of implicit function in the profile likelihood score function. However, it is weakly differentiable in a sense that

$$E \left[\sqrt{n} \left\{ \phi(\hat{\theta}_n) - \phi(\theta_0) \right\} \right] = -E \left[\phi(\theta_0) \phi'(\theta_0) \right] \left\{ \sqrt{n}(\hat{\theta}_n - \theta_0) \right\} + o_p(1), \quad (1.3)$$

where θ_0 and $\hat{\theta}_n$ are the true value and MLE of θ , and ϕ is the profile likelihood score function. One of the advantages of our proposed methodology is to show equation (1.3) when the score function is not differentiable. Using this weak differentiability of the score function, we are able to show the asymptotic normality of the profile likelihood estimator and get a closed form expression for the information matrix. Therefore we can solve the theoretical challenge of the Cox PH cure model and the joint model of longitudinal and survival data by using equation (1.3).

Our contribution is that we have expanded the profile log-likelihood function and used the asymptotic expansion of profile likelihood function to get the estimated SE of the profile likelihood MLE and obtain the explicit form of the SE estimator using the profile likelihood score function. Our proposed approach gives us not only analytical understanding of the profile likelihood estimation process, but also provides closed form formula to compute the standard errors of the profile likelihood MLE using the profile likelihood score function.

In addition to the theoretical challenges, some R-packages have also faced difficulties while calculating the standard errors from the profile likelihood approach. For example, Cai et al. (2012) developed a R package (*smcure*) to fit the Cox PH cure model which has received much attention in recent years (Peng and Taylor, 2014; Amico and Van Keilegom, 2018). This package has adopted the Breslow's estimator (Breslow, 1972) with EM algorithm to estimate the regression coefficients. However, the standard errors of the estimated parameters are not directly available due to the presence of implicit function in the estimating equation. Hence, bootstrap samples (Efron and Tibshirani, 1994) have been used in *smcure* R-package to calculate the standard errors.

In the case of the joint model of longitudinal and survival data, Rizopoulos (2010) developed a R package (JM) where one of the method is 'Cox-PH-GH' based on NPMLE of the baseline hazard function. In JM (method: 'Cox-PH-GH') package, the score function has been calculated without differentiating the implicit function within the log-likelihood function. Then forward or central difference approximation (numerical derivative) has been used to differentiate the score function (for the standard error estimation). However, the standard errors of the estimated parameters based on 'Cox-PH-GH' method can be underestimated while calculated from the score vector (Rizopoulos, 2012). Later, Hickey et al. (2018) have developed another R-package (joineRML) to fit the joint model with multiple longitudinal responses. Similar to the smcure and JM (method: 'Cox-PH-GH') R-packages, joineRML has also adopted the NPMLE of the baseline hazard function. In addition, Monte Carlo EM (MCEM) algorithm and bootstrap process have been considered for estimation procedure, where they have estimated the fixed effect coefficients by ignoring the survival part from the estimation process.

In summary, it can be said that all three above mentioned R-packages have incorporated Breslow's estimator for the baseline hazard and therefore, due to the presence of the implicit function in the profile likelihood function, these R-packages have faced difficulty while calculating the standard errors directly. Hence smcure and joineRML R-packages use the bootstrap method to compute the standard error in the profile likelihood estimation which have theoretical advantages, but computationally expensive. Moreover, bootstrap method may overestimate the standard errors (Hickey et al., 2018; Hsieh et al., 2006; Xu and Zeger, 2001; Xu et al., 2014) whereas the JM (method: 'Cox-PH-GH') R-package uses a improper profile score function which gives underestimated standard error. None of these work used the equation (1.2) for standard error calculation proposed by Murphy

and Van der Vaart (2000).

To show the application of our proposed approach in biomedical and clinical research, we have compared our results with the output obtained from *smcure* R-package (for the Cox PH cure model) in Chapter 5 and with the output obtained from JM (method: 'Cox-PH-GH') and *joinerML* R-packages (for the joint model of longitudinal and survival data) in Chapter 7. The simulation studies suggest that the estimated standard errors from bootstrap samples and our approach are providing similar and comparable results. In addition, in order to assess and compare the computational time between our proposed approach and *smcure*, JM (method: 'Cox-PH-GH') and *joinerML* R-packages, the average time elapsed to get the estimates of the parameters and estimated standard errors have recorded. We observed that the average computation times for our proposed approach are much less compared to the *smcure*, JM (method: 'Cox-PH-GH') and *joinerML* R-packages.

1.8 Research Objectives

The primary objectives of this research are as follows:

1. Expand the profile likelihood function directly to show the asymptotic normality of the profile likelihood estimators for the Cox PH cure model and express the standard errors in terms of the profile likelihood score function.
2. Perform simulation study to compare and contrast the *smcure* R-package with our proposed approach by assessing parameter and standard error estimation.
3. Apply our proposed approach to the Eastern Cooperative Oncology

Group data and compare with the output obtained from *smcure* R-package.

4. Use the asymptotic expansion of the profile likelihood function to get the asymptotic normality of the profile likelihood estimators for the joint model of longitudinal and survival data and obtain the explicit form of SE estimators using the profile likelihood score function.
5. Perform simulation study to compare and contrast the results of JM (method: 'Cox-PH-GH') and *joineRML* R-packages with our proposed approach.
6. Apply our proposed approach to the AIDS and Primary biliary cirrhosis (PBC) datasets (extracted from JM R-package) and compare the results with the output obtained from JM (method: 'Cox-PH-GH') and *joineRML* R-packages.

1.9 Organization of the research

This thesis paper is organized in nine chapters:

Chapter-1 introduces a short definition of survival data, Cox PH model, Profile likelihood approach, Cure model, Cox PH cure model, longitudinal data, joint model of longitudinal and survival data, narration of the main purposes of the study and highlight the primary objectives of the research.

Chapter-2 discusses briefly on estimation of a finite-dimensional parameter, estimation with a finite-dimensional nuisance parameter and estimation with a infinite-dimensional nuisance parameter.

Chapter-3 explains the concept of profile likelihood with mathematical notation and the literature on profile likelihood.

Chapter-4 introduces the Cox PH model along with the likelihood function formulation procedure for parameter estimation, efficient score function for Cox PH model and the efficiency of profile likelihood in Cox PH model.

Chapter-5 explains the Cox PH cure model, asymptotic normality of the profile likelihood MLE with asymptotic variance, derivation of efficient score function in the Cox PH cure model based on our proposed approach. Moreover, for simulation study and real-life example, we have compared our results with the output obtained from *smcure* R-package.

Chapter-6 explains the proofs of Theorems and Lemmas for the Cox PH cure model.

Chapter-7 discusses the basic joint model of longitudinal and survival data, asymptotic normality of the profile likelihood MLE with the standard errors expressed in terms of profile likelihood score function, derivation of efficient score function in the joint model based on our proposed approach. Moreover, for simulation study and real-life example, our results have been compared with the output obtained from JM (method: 'Cox-PH-GH') and *joineRML* R-packages.

Chapter-8 explains the proofs of Theorems and Lemmas for the joint Model of Longitudinal and Survival Data.

Chapter-9 concludes with a brief discussion.

Chapter 2

Asymptotic Normality of MLE

In this chapter, we are going to review the theory of MLE with finite dimensional parameters. In addition, we are going to cover the estimation process with finite and infinite dimensional nuisance parameters. To show the asymptotic normality of MLE, we review the concept of the influence function and its relationship to the asymptotic variance of the estimator. The references which have used in this chapter are Begun et al. (1983); Bickel et al. (1993); Newey (1990); Tsiatis (2007).

2.1 Estimation of a Finite Dimensional Parameter

Let us consider a parametric model, $P = \{p(x; \theta) : \theta \in \Theta \subset \mathbb{R}^p\}$ with density function $p(x; \theta)$. So the log-likelihood function $l(x; \theta)$ can be defined as

$$l(x; \theta) = \sum_{i=1}^n \log p(x_i; \theta).$$

Now the score function is defined as

$$\frac{\partial l(x; \theta)}{\partial \theta} = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log p(x_i; \theta) = \sum_{i=1}^n \phi(x_i; \theta).$$

We will get $\hat{\theta}_n$ (which is the MLE of θ) by solving the equation

$$\sum_{i=1}^n \phi(x_i; \hat{\theta}_n) = 0.$$

Lemma 2.1: The score function, $\phi(x; \theta)$ has mean zero and variance, $I(\theta)$; i.e.,

$$E[\phi(x; \theta)] = 0 \text{ and } Var[\phi(x; \theta)] = I(\theta),$$

where $I(\theta)$ can be expressed as

$$I(\theta) = E[\phi(x; \theta)\phi'(x; \theta)] = -E\left[\frac{\partial}{\partial\theta'}\phi(x; \theta)\right].$$

In order to prove the consistency and asymptotic normality of $\hat{\theta}_n$, we need to consider $E[\frac{\partial}{\partial\theta'}\phi(x; \theta_0)]$ be non-singular and the equation

$$\frac{1}{n} \sum_{i=1}^n \phi(x_i; \theta) \xrightarrow{P} E\left[\phi(x; \theta_0)\right],$$

holds uniformly in θ in a neighbourhood of θ_0 and from the regularity conditions we can show that $\hat{\theta}_n$ is consistent, that is

$$\hat{\theta}_n \xrightarrow{P} \theta_0.$$

So assuming the regularity conditions hold, the influence function for $\hat{\theta}_n$ can be obtained using Taylor's theorem as follows

$$0 = \sum_{i=1}^n \phi(x_i; \hat{\theta}_n) = \sum_{i=1}^n \phi(x_i; \theta_0) + \sum_{i=1}^n \frac{\partial}{\partial\theta'}\phi(x_i; \theta_n^*)(\hat{\theta}_n - \theta_0),$$

where θ_n^* is an intermediate value between $\hat{\theta}_n$ and θ_0 . Because of assuming sufficient regularity conditions for the consistency of $\hat{\theta}_n$, from Uniform Law of Large Numbers (LLN) we can write that

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial\theta'}\phi(x_i; \theta_n^*) \xrightarrow{P} E\left[\frac{\partial}{\partial\theta'}\phi(x; \theta_0)\right].$$

Therefore,

$$\begin{aligned}
0 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(x_i; \theta_0) + \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta'} \phi(x_i; \theta_n^*) \sqrt{n}(\hat{\theta}_n - \theta_0) \\
\sqrt{n}(\hat{\theta}_n - \theta_0) &= - \left[\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta'} \phi(x_i; \theta_n^*) \right]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(x_i; \theta_0) \\
&= - \left\{ \left[E \left(\frac{\partial}{\partial \theta'} \phi(x; \theta_0) \right) \right]^{-1} + o_p(1) \right\} \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(x_i; \theta_0) \\
&= - \left[E \left(\frac{\partial}{\partial \theta'} \phi(x; \theta_0) \right) \right]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(x_i; \theta_0) + o_p(1) \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(x_i; \theta_0) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\phi}(x_i; \theta_0) + o_p(1) \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(x_i; \theta_0),
\end{aligned} \tag{2.1}$$

where we deduce the influence function of $\hat{\theta}_n$ by

$$\tilde{\phi}(x; \theta_0) = - \left[E \left(\frac{\partial}{\partial \theta'} \phi(x; \theta_0) \right) \right]^{-1} \phi(x, \theta_0).$$

We know that $\frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(x_i; \theta_0)$ is bounded in probability, so from equation (2.1), we can write

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(x_i; \theta_0) = O_p(1).$$

Lemma 2.2: If $X_n = o_p(1)$ and $Y_n = O_p(1)$, then $X_n Y_n = o_p(1)$.

So, from equation (2.1) we can conclude that

$$o_p(1) \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(x_i; \theta_0) = o_p(1) O_p(1) = o_p(1).$$

Finally we can express equation (2.1) as

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\phi}(x_i; \theta_0) + o_p(1), \tag{2.2}$$

Now we know that

$$E[\tilde{\phi}(x; \theta_0)] = 0,$$

and the variance of the influence function of $\hat{\theta}_n$ can be expressed as

$$\begin{aligned} Var[\tilde{\phi}(x; \theta_0)] &= \left[E \left(\frac{\partial}{\partial \theta'} \phi(x; \theta_0) \right) \right]^{-1} Var[\phi(x; \theta_0)] \left[E \left(\frac{\partial}{\partial \theta'} \phi(x; \theta_0) \right) \right]^{-1} \\ &= I^{-1}(\theta_0) I(\theta_0) I^{-1}(\theta_0) \\ &= I^{-1}(\theta_0), \end{aligned}$$

where $Var[\tilde{\phi}(x; \theta_0)] < \infty$. From Central Limit Theorem (CLT) we know that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\tilde{\phi}(x_i; \theta_0) - E \left\{ \tilde{\phi}(x; \theta_0) \right\} \right] \xrightarrow{D} N(0, Var[\tilde{\phi}(x; \theta_0)]),$$

or equivalently

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\phi}(x_i; \theta_0) \xrightarrow{D} N(0, I^{-1}(\theta_0)).$$

So finally from equation (2.2), by Slutsky's theorem we can write

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, I^{-1}(\theta_0)).$$

2.2 Estimation with a Finite Dimensional Nuisance Parameter

Let us define a parametric model as a set of probability densities

$$P = \{p(x; \theta) : \theta \in \Theta \subset \mathbb{R}^p\},$$

where, the dimension p is some finite positive integer and the set Θ is defined to be open so we are able to define derivatives with respect to each $\theta \in \Theta$. Let us partition the parameter θ as $(\beta', \eta')'$ where β is a m -dimensional vector and η is a $p - m$ dimensional vector. Then,

$$\Theta = \Theta_\beta \times \Theta_\eta = \{(\beta, \eta) : \beta \in \Theta_\beta \subset \mathbb{R}^m, \eta \in \Theta_\eta \subset \mathbb{R}^{p-m}\},$$

where $\Theta_\beta = \{\beta : (\beta, \eta) \in \Theta\}$ and $\Theta_\eta = \{\eta : (\beta, \eta) \in \Theta\}$. We want to estimate β (parameter of interest) and consider η as the nuisance parameter. The true value of θ is denoted as $\theta_0 = (\beta'_0, \eta'_0)'$. So the parametric model can be expressed as

$$P = \{p(x; \beta, \eta) : \beta \in \Theta_\beta, \eta \in \Theta_\eta\}.$$

2.2.1 Score Function

The score function for θ is partitioned into

$$\phi_\theta(x; \beta, \eta) = \begin{pmatrix} \phi_\beta(x; \beta, \eta) \\ \phi_\eta(x; \beta, \eta) \end{pmatrix},$$

where

$$\phi_\beta(x; \beta, \eta) = \frac{\partial}{\partial \beta} \log p(x; \beta, \eta),$$

is the score function for β and

$$\phi_\eta(x; \beta, \eta) = \frac{\partial}{\partial \eta} \log p(x; \beta, \eta),$$

is the score function for η .

As a result, the Fisher Information matrix, $I(\theta)$ can be partitioned as

$$I(\theta) = \begin{pmatrix} E_{\beta_0, \eta_0}(\phi_\beta \phi'_\beta) & E_{\beta_0, \eta_0}(\phi_\beta \phi'_\eta) \\ E_{\beta_0, \eta_0}(\phi_\eta \phi'_\beta) & E_{\beta_0, \eta_0}(\phi_\eta \phi'_\eta) \end{pmatrix}.$$

We know that if $\hat{\theta}_n$ is the MLE of θ , then we can write

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\phi}(x_i; \theta_0) + o_p(1), \quad (2.3)$$

where the influence function of $\hat{\theta}_n$ can be written as

$$\tilde{\phi}(x_i; \theta_0) = I^{-1}(\theta_0) \phi(x_i; \theta_0).$$

2.2.2 The Efficient Score Function

Now if we partition θ as $(\beta', \eta')'$, then we can partition the equation (2.3) as

$$\begin{aligned} \sqrt{n} \left(\begin{pmatrix} \hat{\beta}_n \\ \hat{\eta}_n \end{pmatrix} - \begin{pmatrix} \beta_0 \\ \eta_0 \end{pmatrix} \right) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} I_{\beta\beta} & I_{\beta\eta} \\ I_{\eta\beta} & I_{\eta\eta} \end{pmatrix}^{-1} \begin{pmatrix} \phi_\beta(x_i; \beta, \eta) \\ \phi_\eta(x_i; \beta, \eta) \end{pmatrix} + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} I^{\beta\beta} & I^{\beta\eta} \\ I^{\eta\beta} & I^{\eta\eta} \end{pmatrix} \begin{pmatrix} \phi_\beta(x_i; \beta, \eta) \\ \phi_\eta(x_i; \beta, \eta) \end{pmatrix} + o_p(1). \end{aligned} \quad (2.4)$$

From the block matrix form of inverse, it can be shown that

$$\begin{aligned} I^{\beta\beta} &= \left[E(\phi_\beta \phi'_\beta) - E(\phi_\beta \phi'_\eta) E(\phi_\eta \phi'_\eta)^{-1} E(\phi_\eta \phi'_\beta) \right]^{-1} \\ &= [I_{\beta\beta} - I_{\beta\eta} I_{\eta\eta}^{-1} I_{\eta\beta}]^{-1}, \end{aligned}$$

and

$$\begin{aligned} I^{\beta\eta} &= - \left[E(\phi_\beta \phi'_\beta) - E(\phi_\beta \phi'_\eta) E(\phi_\eta \phi'_\eta)^{-1} E(\phi_\eta \phi'_\beta) \right]^{-1} E(\phi_\beta \phi'_\eta) E(\phi_\eta \phi'_\eta)^{-1} \\ &= - [I_{\beta\beta} - I_{\beta\eta} I_{\eta\eta}^{-1} I_{\eta\beta}]^{-1} I_{\beta\eta} I_{\eta\eta}^{-1} \\ &= - I^{\beta\beta} I_{\beta\eta} I_{\eta\eta}^{-1}. \end{aligned}$$

So, from the equation (2.4), we can write

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[I^{\beta\beta} \phi_\beta(x_i; \beta, \eta) + I^{\beta\eta} \phi_\eta(x_i; \beta, \eta) \right] + o_p(1). \quad (2.5)$$

So from equation (2.5), we get

$$\begin{aligned} \sqrt{n}(\hat{\beta}_n - \beta_0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[I^{\beta\beta} \phi_\beta(x_i; \beta, \eta) - I^{\beta\beta} I_{\beta\eta} I_{\eta\eta}^{-1} \phi_\eta(x_i; \beta, \eta) \right] + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[I^{\beta\beta} \{ \phi_\beta(x_i; \beta, \eta) - I_{\beta\eta} I_{\eta\eta}^{-1} \phi_\eta(x_i; \beta, \eta) \} \right] + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n I^{\beta\beta} \phi_{\text{eff}}(x_i; \beta, \eta) + o_p(1), \end{aligned} \quad (2.6)$$

where

$$\phi_{\text{eff}}(x; \beta, \eta) = \phi_\beta(x; \beta, \eta) - I_{\beta\eta} I_{\eta\eta}^{-1} \phi_\eta(x; \beta, \eta), \quad (2.7)$$

is the efficient score function for β where $I_{\beta\eta} I_{\eta\eta}^{-1} \phi_\eta = E[\phi_\beta \phi_\eta] E[\phi_\eta \phi_\eta]^{-1} \phi_\eta$ is the projection of ϕ_β on the space spanned by ϕ_η (Begun et al., 1983; Bickel et al., 1993).

2.2.3 The Efficient Influence Function

From equation (2.6), we can write

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\phi}_{\text{eff}}(x_i; \beta, \eta) + o_p(1), \quad (2.8)$$

where we deduce the efficient influence function of $\hat{\beta}_n$ by

$$\tilde{\phi}_{\text{eff}}(x; \beta, \eta) = I^{\beta\beta} \phi_{\text{eff}}(x; \beta, \eta),$$

where

$$\begin{aligned} I^{\beta\beta} &= \left[E \left\{ \phi_\beta(x; \beta, \eta) - I_{\beta\eta} I_{\eta\eta}^{-1} \phi_\eta(x; \beta, \eta) \right\} \left\{ \phi_\beta(x; \beta, \eta) - I_{\beta\eta} I_{\eta\eta}^{-1} \phi_\eta(x; \beta, \eta) \right\}' \right]^{-1} \\ &= \left[E \left\{ \phi_{\text{eff}}(x; \beta, \eta) \phi_{\text{eff}}'(x; \beta, \eta) \right\} \right]^{-1}. \end{aligned}$$

From CLT, we can write

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\tilde{\phi}_{\text{eff}}(x_i; \beta, \eta) - E \left\{ \tilde{\phi}_{\text{eff}}(x; \beta, \eta) \right\} \right] \xrightarrow{D} N(0, I^{\beta\beta}),$$

or equivalently

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n [\tilde{\phi}_{\text{eff}}(x_i; \beta, \eta)] \xrightarrow{D} N(0, I^{\beta\beta}).$$

So finally from equation (2.8), by Slutsky's theorem we can write

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{D} N(0, I^{\beta\beta}).$$

2.3 Estimation with a Infinite Dimensional Nuisance Parameter

Let us define a semi-parametric model as a set of probability densities having both finite-dimensional and infinite-dimensional parameters.

$$\mathcal{P} = \{P_{\beta, \eta} : \beta \in \Theta_\beta \subset \mathbb{R}^m, \eta \in \Theta_\eta\},$$

where, the parameter of interest β is an m -dimensional parameter and the nuisance parameter η is an infinite-dimensional parameter. Here we assume that Θ_β is an open neighborhood of β_0 in \mathbb{R}^m and Θ_η is an open neighborhood of η_0 of some Banach space, B_η . Let $p(x; \beta, \eta)$ be the density for the probability measure $P_{\beta, \eta}$.

A parametric submodel of a semi-parametric model P around P_{β_0, η_0} is a subset of P which has a finite-dimensional smooth parametrization and contains the true distribution P_{β_0, η_0} .

For each r -dimensional surface $t \rightarrow \eta_t, t \in \Theta_t$, passing through η_0 ,

$$\mathcal{P}_{\beta, \eta_t} = \{P_{\beta, \eta_t} : \beta \in \Theta_\beta \subset \mathbb{R}^m, t \in \Theta_t \subset \mathbb{R}^r\},$$

is a parametric submodel with and r -dimensional nuisance parameter where $p(x; \beta, \eta_t)$ is a density for P_{β, η_t} .

2.3.1 Score Function

First we have to define the score function for β and score operator for η in the following ways

Score Function for β

The score function for β can be obtained as

$$\phi_\beta = \frac{\partial}{\partial \beta} \log p(x; \beta, \eta).$$

Score Operator for η

Take a measurable function which is bounded such as $f : [0, \tau] \rightarrow R$, where f is defined in the interval $[0, \tau]$ where τ is a finite number and η is also restricted within this interval. The path defined by $d\eta_t = (1 + tf)d\eta$ is a sub-model passing through η at $t = 0$. By the boundedness of f , we are assured that $(1 + tf)d\eta$ is positive for sufficiently small t .

The derivative of log-likelihood function with respect to t can be expressed as $\frac{\partial}{\partial t} \log p(x; \beta, \eta_t)$ and when $t = 0$, the derivative yields

$$B_\eta f = \left. \frac{\partial}{\partial t} \right|_{t=0} \log p(x; \beta, \eta_t),$$

where B_η is the score operator for η .

2.3.2 The Efficient Score Function

The efficient score function from equation (2.7) motivate us to define for the semi-parametric model also. Moreover, we have defined the score function for β and score operator for η in section (2.3.1).

We define the efficient score function in the semi-parametric model by

$$\phi_{eff} = \left[I - B_\eta (B_\eta^* B_\eta)^{-1} B_\eta^* \right] \phi_\beta, \quad (2.9)$$

where ϕ_β and B_η are the score function and score operator with respect to β and η respectively. B_η^* is the adjoint of the operator B_η . Moreover, $B_\eta (B_\eta^* B_\eta)^{-1} B_\eta^* \phi_\beta$ is the projection of ϕ_β on the space spanned by B_η and I denotes the identity operator (Begun et al., 1983; Bickel et al., 1993; Newey, 1990; Tsiatis, 2007). This definition is a natural extension of the efficient score function in the parametric model (equation 2.7).

Chapter 3

Profile Likelihood

3.1 Preface

Profile likelihood approach is usually used for the models with higher dimensional nuisance parameters (semi-parametric models). Ordinary likelihood function of parametric models and profile likelihood function in semi-parametric models can be used in the same way. The profile likelihood can be useful to estimate parameters and produce confidence intervals for non-linear models with better coverage (Royston et al., 2007). For example, to estimate regression coefficients and frailty parameters in the correlated gamma-frailty model, an inference has been made by using profile likelihood approach (Chang et al., 2007). Profile likelihood approach has been also used for estimating the parameters from the complex models such as generalized linear mixed models (GLMM) with factor structures using standard software and minimal programming (Jeon and Rabe-Hesketh, 2012).

The efficiency of profile likelihood estimator in semi-parametric models has been done by introducing the approximate least favourable sub-model (Murphy and Van der Vaart, 2000). Later, while investigating the efficiency of profile likelihood estimator, an improvement has been established directly through the quadratic expansion of the profile likelihood (Hi-

rose, 2011b). To illustrate this approach, we apply to the Cox PH model in Chapter-4.

3.2 Profile Likelihood Approach

Let us define the unknown parameter as θ which can be partitioned as $\theta' = (\beta', \eta')$. Here, β is a p -dimensional vector (parameter of interest) and η is infinite dimensional nuisance parameter. Though our interest lies only in β , but we have to estimate both β and η . To achieve our goal, first we profile out the nuisance parameter η . Now we will describe the process of estimating β and η in two stages to motivate the profile likelihood estimation procedure.

Let us consider the density function, $p(x; \beta, \eta)$ from which we estimate β and η . So the log-likelihood function can be expressed as

$$l(\beta, \eta) = \sum_{i=1}^n \log p(x_i; \beta, \eta).$$

To estimate β and η , we can use

$$(\hat{\beta}, \hat{\eta}) = \arg \max_{\beta, \eta} l(\beta, \eta).$$

However, for some cases it is very difficult to directly maximize the above equation. For this reason we consider an approach which can be easier compared to above the equation. Suppose β is fixed but η can vary, so to estimate η , we can maximize $l(\beta, \eta)$ with respect to η as

$$\hat{\eta}_\beta = \arg \max_{\eta} l(\beta, \eta),$$

where $\hat{\eta}_\beta$ depends on β . Now for estimating β , we use the profile log-likelihood function:

$$\hat{\beta} = \arg \max_{\beta} l(\beta, \hat{\eta}_\beta),$$

where we profiled out the nuisance parameter η .

3.3 Efficiency of Profile Likelihood Approach in Semi-Parametric Models

Let us define a semi-parametric model as a set of probability densities

$$P = \{p(x; \beta, \eta) : \beta \in \Theta \subset \mathbb{R}^p, \eta \in H\},$$

where, β is p dimensional parameter of interest and η is a nuisance parameter which is infinite dimensional. Let (β_0, η_0) be the true value of (β, η) . Moreover, we assume Θ is a compact set containing an open neighbourhood β_0 in \mathbb{R}^p and H is a convex set containing η_0 in Banach space, B .

To show the alternative approach of Murphy and Van der Vaart (2000), the definition of profile likelihood has extended by Hirose (2011b). According to Hirose (2011b), if there exist a function $\hat{\eta}(\beta, F)$, where F is cdf such that $\hat{\eta}(\beta_0, F_0) = \eta_0$ and the derivative

$$\left. \frac{\partial}{\partial \beta} \right|_{\beta=\beta_0} \log P(x; \beta, \hat{\eta}(\beta, F_0)),$$

is the efficient score function where F_0 is the cdf of the density function $P(x; \beta_0, \eta_0)$. Then the profile log-likelihood function for β (for the empirical cdf F_n) can be written as

$$l(\beta, \hat{\eta}(\beta, F_n)) = \sum_{i=1}^n \log P(x_i; \beta, \hat{\eta}(\beta, F_n)).$$

The main purpose of Hirose (2011b) is to have an additional parameter F in the function $\hat{\eta}(\beta, F)$, so that estimating equation can be written as

$$\phi(x; \beta, F_n) = \sum_{i=1}^n \frac{\partial}{\partial \beta} \log P(x_i; \beta, \hat{\eta}(\beta, F_n)) = 0, \quad (3.1)$$

where $\phi(x; \beta, F_n)$ is the profile likelihood score function. Therefore, equation (3.1) can be expressed as an explicit function of sample size n , through F_n . Moreover, $\hat{\beta}_n$ corresponding to equation (3.1) is efficient (Hirose, 2011b).

Assumptions

To show the consistency and efficiency of $\hat{\beta}_n$, Hirose (2011b) has considered the following assumptions:

A1: $\hat{\eta}(\beta_0, F_0) = \eta_0$ and

$$\phi_{\text{eff}}(x; \beta_0, F_0) = \left. \frac{\partial}{\partial \beta} \log P(x; \beta, \hat{\eta}(\beta, F)) \right|_{\beta=\beta_0, F=F_0},$$

is the efficient score function.

A2: The empirical cdf F_n is $n^{1/2}$ consistent such that $n^{1/2}|F_n - F_0| = O_p(1)$. Moreover for all x , $\log P(x; \beta, \hat{\eta}(\beta, F))$ is twice continuous differentiable with respect to β and Hadamard differentiable with respect to F for each $(\beta, F) \in \Theta \times F$.

A3: The efficient information matrix $I^* = E(\phi_{\text{eff}} \phi'_{\text{eff}})$ is invertible.

A4: There exists a $\rho > 0$ such that the class of functions $\{\phi(x; \beta, F) : (\beta, F) \in \Theta \times \zeta_\rho\}$ is Donsker type with square-integrable envelope function, and the class of functions $\{\frac{\partial}{\partial \beta} \phi(x; \beta, F) : (\beta, F) \in \Theta \times \zeta_\rho\}$ is Glivenko-Cantelli with integrable envelope function.

Theorem 3.1: Based on the assumptions $\{A1, A2, A3, A4\}$, it can be showed that a consistent estimator $\hat{\beta}_n$ corresponding to the following estimating equation

$$\sum_{i=1}^n \phi(x_i; \hat{\beta}_n, F_n) = 0,$$

is an asymptotically linear estimator for β_0 with the efficient influence function $(I^*)^{-1} \phi_{\text{eff}}(x; \beta_0, F_0)$. Therefore, we can write

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (I^*)^{-1} \phi_{\text{eff}}(x_i; \beta_0, F_0) + o_p(1) \xrightarrow{D} N\{0, (I^*)^{-1}\}, \quad (3.2)$$

where $N\{0, (I^*)^{-1}\}$ is a normal distribution with mean zero and variance $(I^*)^{-1}$. So from equation (3.2), we can say that the profile likelihood MLE $\hat{\beta}_n$ is efficient (Hirose, 2011b).

Chapter 4

Cox PH Model

Cox PH model is the most widely used semi-parametric regression model where the baseline hazard is an unknown function of survival time and the regression coefficient is an unknown vector. Different kinds of semi-parametric models such as proportional odds models and linear transformation models have been developed to analyze time to event data in survival analysis after the establishment of Cox PH model (Hanson and Yang, 2007; Wang and Dunson, 2010; Zeng and Lin, 2007a). In addition, numerous mathematical techniques such as counting process theory and empirical process theory have been used to develop asymptotic theory of the estimators (Fleming and Harrington, 1991; Van der Vaart and Wellner, 1996).

Let T denote the failure time and $Z = (Z_1, \dots, Z_k)'$ represents a set of available covariates. The covariates Z may include quantitative, qualitative and time dependent variables. We are interested in modeling and determining the relationship between Z and lifetime random variable (T). These models assume that covariates have a multiplicative effect on the hazard for an event, and they are formulated as,

$$\lambda(t|Z) = \lambda(t)e^{\beta'Z},$$

where the form of baseline hazard function $\lambda(t)$ is unspecified which im-

plies that the distribution of lifetime random variable is unknown and β is a vector of regression coefficients (Cox, 1972).

The Cox PH model is preferred over the logistic regression model when the survival time information is available and there is censoring in the data set. The Cox PH model is very popular because it is a robust model so the results obtained from the Cox PH model will closely approximate the results obtained from a correct parametric model. Even though the baseline hazard function $\lambda(t)$ is unspecified or unknown, it is possible to estimate the regression coefficient vector β and the hazard ratio is easily estimable from the estimate of β .

4.1 Estimation of Regression Parameter

Suppose we observe n observations $(T_i, Z_i, \delta_i); i = 1, 2, \dots, n$ where T_i is the length of time a subject was observed, Z_i a covariates whose value is measured at the beginning of the study. Moreover, δ_i indicates whether the observed time is censored or not

$$\delta_i = \begin{cases} 1 & \text{for } T_i = \text{event time} \\ 0 & \text{for } T_i = \text{censored time} \end{cases}$$

The likelihood of observed $(T_i = t_i, Z_i, \delta_i = 1)$ is $f(t_i|Z_i)$ and the likelihood of censored $(T_i = t_i, Z_i, \delta_i = 0)$ is $S(t_i|Z_i)$, where $f(\cdot)$ and $S(\cdot)$ are probability distribution function and survival function respectively. Therefore the likelihood of an observation (t_i, Z_i, δ_i) is generally given by

$$f(t_i|Z_i)^{\delta_i} S(t_i|Z_i)^{1-\delta_i}.$$

The likelihood of all n observations is

$$L(\beta) = \prod_{i=1}^n f(t_i|Z_i)^{\delta_i} S(t_i|Z_i)^{1-\delta_i}.$$

We know that $f(t|Z) = \lambda(t|Z)S(t|Z)$ and $S(t|Z) = \exp(-\int_0^t \lambda(s|Z)ds)$, so we can write the likelihood as

$$L(\beta) = \prod_{i=1}^n \lambda(t_i|Z_i)^{\delta_i} S(t_i|Z_i).$$

For the Cox PH model, we know that

$$\lambda(t|Z) = \lambda(t)e^{\beta'Z}, \quad S(t|Z) = [S_0(t)]^{e^{\beta'Z}}.$$

where $S_0(t)$ is the baseline survival function. So the likelihood function can be written as

$$L(\beta) = \prod_{i=1}^n \left[\{\lambda(t_i)e^{\beta'Z_i}\}^{\delta_i} \exp\left(-\int_0^{t_i} \lambda(s)ds\right)^{e^{\beta'Z_i}} \right].$$

4.1.1 Estimate Regression Parameters using Partial Likelihood

The likelihood, $L(\beta)$ require $\lambda(t)$ to be known and therefore an estimator of $\lambda(t)$ is needed. To avoid the estimation of $\lambda(t)$, Cox proposed the partial likelihood function (Cox, 1975).

Suppose we observe (T, δ, Z) in time interval $[0, \tau]$, $Z \subset R^k$ is a regression covariate and T is a right-censored failure time related to cumulative hazard, $\Lambda(t|Z) = e^{\beta'Z}\Lambda(t)$. Let us assume there is no tie in the observation times; so one and only one individual failed at time t_i . Let $Y_i(t) = 1\{T_i \geq t\}$. Then the partial likelihood function is given by

$$L(\beta) = \prod_{i=1}^n \left(\frac{e^{\beta'Z_i}}{\sum_{l=1}^n Y_l(t_i)e^{\beta'Z_l}} \right)^{\delta_i}, \quad (4.1)$$

and the log partial likelihood function can be written as

$$\log L(\beta) = \sum_{i=1}^n \delta_i \left[Z_i\beta - \log \sum_{l=1}^n Y_l(t_i)e^{Z_l\beta} \right].$$

The corresponding score function is

$$\frac{\partial}{\partial \beta} \log L(\beta) = \sum_{i=1}^n \delta_i \left[Z_i - \frac{\sum_{l=1}^n Y_l(t_i) Z_l e^{Z_l \beta}}{\sum_{l=1}^n Y_l(t_i) e^{Z_l \beta}} \right].$$

The variance of $\hat{\beta}$ is estimated by $\widehat{Var}(\hat{\beta}) = \hat{I}^{-1}(\hat{\beta})$, where

$$\hat{I}(\hat{\beta}) = - \frac{\partial^2}{\partial \beta \partial \beta'} \bigg|_{\beta=\hat{\beta}} \log L(\beta).$$

4.1.2 Estimate Regression Parameters using Profile Likelihood

Suppose the Cox model, $\lambda(t|Z) = \lambda(t)e^{\beta'Z}$ and also assume that the baseline hazard $\lambda(t)$ has (possible) non-zero values on t_i

$$\lambda(t_i) = \lambda_i \geq 0; i = 1, 2, \dots, n$$

and $\lambda(t) = 0$ for the rest of time (Murphy and Van der Vaart, 2000). Under the assumption, the baseline cumulative hazard function, $\Lambda(t) = \int_0^t \lambda(s)ds$ can be expressed as

$$\Lambda(t) = \sum_{i=1}^n \lambda_i 1\{t_i \leq t\}.$$

Then the likelihood function is

$$\begin{aligned} L(\beta, \Lambda) &= \prod_{i=1}^n \left[\left\{ \lambda(t_i) e^{\beta' Z_i} \right\}^{\delta_i} \exp \left(- \int_0^{t_i} \lambda(s) ds \right)^{e^{\beta' Z_i}} \right] \\ &= \prod_{i=1}^n \left[\left\{ \lambda_i e^{\beta' Z_i} \right\}^{\delta_i} \exp \left(- \sum_{j=1}^n \lambda_j 1\{t_j \leq t_i\} \right)^{e^{\beta' Z_i}} \right], \end{aligned}$$

and the log-likelihood function can be written as

$$\log L(\beta, \Lambda) = \sum_{i=1}^n \left[\delta_i \{ \log \lambda_i + \beta' Z_i \} - e^{\beta' Z_i} \sum_{j=1}^n \lambda_j 1\{t_j \leq t_i\} \right].$$

Now the score equation with respect to λ_k is

$$\frac{\partial}{\partial \lambda_k} \log L(\beta, \Lambda) = \frac{\delta_k}{\lambda_k} - \sum_{i=1}^n 1\{t_k \leq t_i\} e^{\beta' Z_i} = 0.$$

So the estimate of λ_k is given by

$$\hat{\lambda}_k(\beta) = \frac{\delta_k}{\sum_{l=1}^n 1\{t_k \leq t_l\} e^{\beta' Z_l}} = \frac{\delta_k}{\sum_{l=1}^n Y_l(t_k) e^{\beta' Z_l}},$$

and the estimate of cumulative baseline hazard can be written as

$$\hat{\Lambda}_\beta(t) = \sum_{i=1}^n \frac{\delta_i 1\{t_i \leq t\}}{\sum_{l=1}^n Y_l(t) e^{\beta' Z_l}}. \quad (4.2)$$

Then we can estimate β as the maximizer of a profile likelihood function for β . The profile log likelihood function can be written as

$$\begin{aligned} \log L(\beta, \hat{\Lambda}_\beta) &= \sum_{i=1}^n \left[\delta_i \{ \log \hat{\lambda}_i(\beta) + \beta' Z_i \} - e^{\beta' Z_i} \sum_{j=1}^n \hat{\lambda}_j(\beta) 1\{t_j \leq t_i\} \right] \\ &= \sum_{i=1}^n \left[\delta_i \left\{ \log \frac{\delta_i}{\sum_{l=1}^n Y_l(t_i) e^{\beta' Z_l}} + \beta' Z_i \right\} - e^{\beta' Z_i} \sum_{j=1}^n \frac{\delta_j 1\{t_j \leq t_i\}}{\sum_{l=1}^n Y_l(t_i) e^{\beta' Z_l}} \right]. \end{aligned}$$

Now the score function based on profile likelihood function is

$$\frac{\partial}{\partial \beta} \log L(\beta, \hat{\Lambda}_\beta) = \sum_{i=1}^n \delta_i \left\{ Z_i - \frac{\sum_{l=1}^n Y_l(t_i) Z_l e^{\beta' Z_l}}{\sum_{l=1}^n Y_l(t_i) e^{\beta' Z_l}} \right\}.$$

The variance of $\hat{\beta}$ is estimated by $\widehat{Var}(\hat{\beta}) = \hat{I}^{-1}(\hat{\beta})$, where

$$\hat{I}(\hat{\beta}) = - \frac{\partial^2}{\partial \beta \partial \beta'} \bigg|_{\beta=\hat{\beta}} \log L(\beta, \hat{\Lambda}_\beta).$$

So it can be observed that the partial likelihood score function is identical to the profile likelihood score function. Thus, the profile likelihood estimator of the regression coefficient is the same as the partial likelihood estimator (Hirose, 2011a; Yan and Yi, 2015).

4.2 Efficiency of Profile Likelihood

An estimator of the baseline cumulative hazard function in the counting process notation (Fleming and Harrington, 1991) can be written from equation (4.2) as

$$\hat{\Lambda}(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u) e^{\beta' Z_i}}, \quad (4.3)$$

where $N(t) = 1\{T \leq t, \delta = 1\}$ and $Y(t) = 1\{T \geq t\}$.

We denote the empirical cdf of the observations $\{T_i, \delta_i, Z_i : i = 1, \dots, n\}$ by F_n . Let us denote $E_{F_n} f = \int f dF_n$. Then $\hat{\Lambda}(t)$ can be expressed as

$$\hat{\Lambda}_{\beta, F_n}(t) = \int_0^t \frac{E_{F_n} dN(u)}{E_{F_n} Y(u) e^{\beta' Z}}. \quad (4.4)$$

The profile log-likelihood function for a single observation can be written as

$$\log P(T, \delta | \beta, \hat{\Lambda}_{\beta, F_n}) = \delta \left\{ \beta' Z + \log \frac{E_{F_n} dN(T)}{E_{F_n} Y(T) e^{\beta' Z}} \right\} - e^{\beta' Z} \int_0^T \frac{E_{F_n} dN(u)}{E_{F_n} Y(u) e^{\beta' Z}}. \quad (4.5)$$

The score function can be obtained as

$$\begin{aligned} \phi_{pl}(T, \delta | \beta, F_n) &= \frac{\partial}{\partial \beta} \log P(T, \delta | \beta, \hat{\Lambda}_{\beta, F_n}) \\ &= \delta \left[Z - \frac{E_{F_n} Y(t) Z e^{\beta' Z}}{E_{F_n} Y(t) e^{\beta' Z}} \right] - e^{\beta' Z} \int_0^T \left[Z - \frac{E_{F_n} Y(u) Z e^{\beta' Z}}{E_{F_n} Y(u) e^{\beta' Z}} \right] d\hat{\Lambda}_{\beta, F_n}(u). \end{aligned} \quad (4.6)$$

Replace F_n by F_0 , we get from equation (4.4)

$$\hat{\Lambda}_{\beta_0, F_0}(t) = \int_0^t \frac{E[dN(u)]}{E[Y(u) e^{\beta_0' Z}]}, \quad (4.7)$$

where E is the expectation with respect to the true distribution F_0 . At the true value of the parameters (β, F) we can write

$$E[dN(u)] = E[Y(u) e^{\beta_0' Z}] d\Lambda_0(u). \quad (4.8)$$

Therefore from equation (4.7), we can write

$$\hat{\Lambda}_{\beta_0, F_0}(t) = \Lambda_0(t),$$

where $\Lambda_0(t)$ is the true cumulative baseline hazard function. Now the score function for Cox PH model at the true value of the parameters (β, F) can be expressed as

$$\begin{aligned} \phi_{pl}(T, \delta | \beta_0, F_0) &= \frac{\partial}{\partial \beta} \log P(T, \delta | \beta, \hat{\Lambda}_{\beta, F}) \Big|_{\beta=\beta_0, F=F_0} \\ &= \left\{ \delta \left[Z - \frac{E[Y(T) Z e^{\beta_0' Z}]}{E[Y(T) e^{\beta_0' Z}]} \right] - e^{\beta_0' Z} \int_0^T \left[Z - \frac{E[Y(u) Z e^{\beta_0' Z}]}{E[Y(u) e^{\beta_0' Z}]} \right] d\Lambda_0(u) \right\}, \end{aligned} \quad (4.9)$$

which is the profile likelihood score function for the Cox PH model. In section 4.3 (Appendix), we show that the profile likelihood score function based on equation (4.9) is the efficient score function.

Asymptotic normality of MLE

Using the Theorem 3.1 from Chapter 3, it can be shown that a consistent estimator $\hat{\beta}_n$ is an asymptotically linear estimator for β_0 and we can write

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (I^*)^{-1} \phi_{pl}(T, \delta | \beta_0, F_0) + o_p(1) \xrightarrow{D} N\{0, (I^*)^{-1}\}.$$

Here $I^* = E[\phi_{pl} \phi_{pl}']$ is the efficient information matrix and $N\{0, (I^*)^{-1}\}$ is a normal distribution with mean zero and variance $(I^*)^{-1}$. So from the above equation, we can say that the profile likelihood MLE $\hat{\beta}_n$ is efficient (Hirose, 2011a,b).

4.3 Appendix

Efficient Score Function using Projection Theory

In this section, we are going to use the equation (2.9) for the calculation of efficient score function based on the projection theory (Begun et al., 1983; Bickel et al., 1993). Moreover, we are going to show that the score function for Cox PH model based on profile likelihood and the efficient score function based on the projection theory are identical.

To get the efficient score function using the projection theory, we assume the parameters (β, Λ) are evaluated at the true values β_0, Λ_0 and omit subscript “0” for brevity.

The log-likelihood function of Cox PH model for one observation can be written as

$$\log P(T, \delta | \beta, \Lambda) = \left\{ \delta (\log \lambda(t) + \beta' Z) - e^{\beta' Z} \Lambda(t) \right\}.$$

Score Function for β

$$\phi_\beta = \frac{\partial}{\partial \beta} \log P(T, \delta | \beta, \Lambda) = \left\{ Z (\delta - e^{\beta' Z} \Lambda(t)) \right\}.$$

Score Operator for Λ

Let us take a bounded continuous function $g : [0, \tau] \rightarrow R$, where g is defined in the interval $[0, \tau]$ because Λ is also restricted within this interval. We define a path by

$$d\Lambda_s = (1 + sg)d\Lambda$$

which defines a submodel passing through Λ at $s = 0$. By the boundedness of g , we are assured that $(1 + sg)d\Lambda$ is positive for sufficiently small s .

The corresponding path for the baseline hazard function is

$$\lambda_s(t) = \frac{d\Lambda_s}{dt} = (1 + sg) \frac{d\Lambda}{dt} = (1 + sg)\lambda(t),$$

and the score operator is given by the derivative of the log-likelihood function with respect to s (at $s = 0$) can be expressed as

$$B_\Lambda(T, \delta | \beta, \Lambda)g = \left. \frac{\partial}{\partial s} \right|_{s=0} \log P(T, \delta | \beta, \Lambda_s) = \left\{ \delta g(t) - e^{\beta' Z} \int_0^t g(u) d\Lambda(u) \right\}.$$

Information Operator $B_\Lambda^* B_\Lambda$ and its Inverse $(B_\Lambda^* B_\Lambda)^{-1}$

Let us start with the information operator $B_\Lambda^* B_\Lambda$ and take two arbitrary functions f and g . By definition of the adjoint, we can write

$$\langle B_\Lambda^* B_\Lambda f, g \rangle_{L_2(\Lambda)} = \langle B_\Lambda f, B_\Lambda g \rangle_{L_2(P)}. \quad (4.10)$$

The path defined by $d\Lambda_{r,s} = (1 + rf + sg + rsfg)d\Lambda$ is positive for small r and s . It can be written as $d\Lambda_{r,s} = (1 + rf)(1 + sg)d\Lambda$. The corresponding path for the baseline hazard function is

$$\lambda_{r,s}(t) = \frac{d\Lambda_{r,s}}{dt} = (1 + rf + sg + rsfg) \frac{d\Lambda}{dt} = (1 + rf + sg + rsfg)\lambda(t).$$

Now we can write

$$\left. \frac{\partial}{\partial r} \right|_{(r,s)=(0,0)} \log P(T, \delta | \beta, \Lambda_{r,s}) = B_{\Lambda} f, \quad (4.11)$$

and

$$\left. \frac{\partial}{\partial s} \right|_{(r,s)=(0,0)} \log P(T, \delta | \beta, \Lambda_{r,s}) = B_{\Lambda} g. \quad (4.12)$$

Using (4.11) and (4.12) we have that

$$\begin{aligned} \langle B_{\Lambda} f, B_{\Lambda} g \rangle_{L_2(P)} &= E \left\{ (B_{\Lambda} f)(B_{\Lambda} g) \right\} \\ &= - E \left\{ \left. \frac{\partial^2}{\partial r \partial s} \right|_{(r,s)=(0,0)} \log P(T, \delta | \beta, \Lambda_{r,s}) \right\} \\ &= E \left\{ e^{\beta' Z} \int_0^T f(\xi) g(\xi) d\Lambda(\xi) \right\}. \end{aligned} \quad (4.13)$$

Now we manipulate the integral involving the function ξ , we deduce

$$\int_0^T f(\xi) g(\xi) d\Lambda(\xi) = \int_0^T I(\xi \leq T) f(\xi) g(\xi) d\Lambda(\xi).$$

Indeed, if $\xi > T$, then the contribution will be 0 to the integral. So the last term in equation (4.13) can be expressed as

$$E \left\{ e^{\beta' Z} \int_0^T f(\xi) g(\xi) d\Lambda(\xi) \right\} = E \left\{ e^{\beta' Z} \int_0^T I(\xi \leq T) f(\xi) g(\xi) d\Lambda(\xi) \right\}. \quad (4.14)$$

Using Fubini's theorem, equation (4.14) can be written as

$$E \left\{ e^{\beta' Z} \int_0^T I(\xi \leq T) f(\xi) g(\xi) d\Lambda(\xi) \right\} = \left\langle E \{ e^{\beta' Z} I(\xi \leq T) f(\xi) \}, g(\xi) \right\rangle_{L_2(\Lambda)}. \quad (4.15)$$

From equation (4.10) we can write

$$\left\langle B_{\Lambda}^* B_{\Lambda} f, g \right\rangle_{L_2(\Lambda)} = \left\langle E \{ e^{\beta' Z} I(t \leq T) f \}, g \right\rangle_{L_2(\Lambda)}.$$

So, the information operator is

$$B_{\Lambda}^* B_{\Lambda} f = E \left\{ e^{\beta' Z} I(t \leq T) \right\} f(t).$$

It follows that the inverse of information operator is

$$(B_{\Lambda}^* B_{\Lambda})^{-1} f(t) = \left[E \{ e^{\beta' Z} I(t \leq T) \} \right]^{-1} f(t).$$

The Action of the Adjoint Score Operator B_{Λ}^* on the Score Function ϕ_{β}

Assume the differentiable paths $(r, s) \mapsto P(T, \delta | \beta + ru, \Lambda_s)$ can be exploited with the path $d\Lambda_s = (1 + sg)d\Lambda$. Now we can write

$$\left. \frac{\partial}{\partial r} \right|_{(r,s)=(0,0)} \log P(T, \delta | \beta + ru, \Lambda_s) = u' \phi_{\beta}. \quad (4.16)$$

and

$$\left. \frac{\partial}{\partial s} \right|_{(r,s)=(0,0)} \log P(T, \delta | \beta + ru, \Lambda_s) = B_{\Lambda} g. \quad (4.17)$$

Using equation (4.16) and (4.17) we can write

$$\begin{aligned} \left\langle u' \phi_{\beta}, B_{\Lambda} g \right\rangle &= E \left\{ (u' \phi_{\beta})(B_{\Lambda} g) \right\} \\ &= - E \left\{ \left. \frac{\partial^2}{\partial r \partial s} \right|_{(r,s)=(0,0)} \log P(T, \delta | \beta + ru, \Lambda_s) \right\} \\ &= u' E \left\{ Z e^{\beta' Z} \int_0^t g(\xi) d\Lambda(\xi) \right\}. \end{aligned} \quad (4.18)$$

Now by manipulating the integral involving the function ξ , the equation (4.18) can be expressed as

$$\left\langle u' \phi_{\beta}, B_{\Lambda} g \right\rangle = u' E \left\{ Z e^{\beta' Z} \int_0^T I(\xi \leq T) g(\xi) d\Lambda(\xi) \right\}. \quad (4.19)$$

Using the Fubini's theorem, we can conclude that

$$u' E \left\{ Z e^{\beta' Z} \int_0^T I(\xi \leq T) g(\xi) d\Lambda(\xi) \right\} = \left\langle u' E \left\{ Z e^{\beta' Z} I(\xi \leq T) \right\}, g(\xi) \right\rangle_{L_2(\Lambda)}. \quad (4.20)$$

We know that

$$\left\langle u' B_{\Lambda}^* \phi_{\beta}, g \right\rangle_{L_2(P)} = \left\langle u' \phi_{\beta}, B_{\Lambda} g \right\rangle_{L_2(\Lambda)}.$$

So we can write

$$B_{\Lambda}^* \phi_{\beta} = E \left\{ Z e^{\beta' Z} I(t \leq T) \right\}.$$

Efficient Score Function ϕ_{eff} :

Finally the efficient score function can be expressed as

$$\begin{aligned} \phi_{eff} &= \phi_{\beta} - B_{\Lambda} (B_{\Lambda}^* B_{\Lambda})^{-1} B_{\Lambda}^* \phi_{\beta} \\ &= \left\{ \delta Z - Z e^{\beta' Z} \Lambda(T) - \left[\delta - e^{\beta' Z} \int_0^T d\Lambda(u) \right] \frac{E[Z e^{\beta' Z} I(t \leq T)]}{E[e^{\beta' Z} I(t \leq T)]} \right\} \\ &= \left\{ \delta \left[Z - \frac{E[Y(T) Z e^{\beta' Z}]}{E[Y(T) e^{\beta' Z}]} \right] - e^{\beta' Z} \int_0^T \left[Z - \frac{E[Y(u) Z e^{\beta' Z}]}{E[Y(u) e^{\beta' Z}]} \right] d\Lambda(u) \right\}. \end{aligned} \tag{4.21}$$

The efficient score function based on equation (4.21) and the profile likelihood score function from equation (4.9) are identical.

Chapter 5

Efficient Estimation For The Cox PH Cure Model

5.1 Introduction

Mixture cure models are commonly used in medical and clinical studies where the population has cured and uncured subjects/ individuals. For analyzing mixture cure model, Berkson and Gage (1952) have used a mixture of exponential regression with a constant cure fraction and Farewell (1982) adopted the weibull distribution for survival function and the logistic model for the cure fraction. Later logistic regression for the mixture proportion along with a generalized F-statistic has been applied for the survival function of the uncured subjects (Peng et al., 1998). Ghitany et al. (1994) proposed a combination of logistic and exponential distribution for the mixture cure model. As an extension of Farewell's model (Farewell, 1977), a more flexible mixture cure model has been introduced by retaining the conditional survival distribution for uncured individuals as a completely unspecified function (Taylor, 1995). A class of transformation models for the event time has been proposed to use generalized estimating equations for parameter estimation. Lu and Ying (2004) have used the counting process and its associated martingale theory to show the asymptotic properties of mixture

cure models. To avoid the restrictions imposed by parametric conditional survival functions, semi-parametric mixture cure models have been used as well. From the literature, it is seen that covariate effect has expressed both in parametric and semi-parametric mixture cure models (Chappell et al., 1995; Goldman, 1984; Kuk and Chen, 1992; Maller and Zhou, 1996; Peng, 2003; Peng and Dear, 2000; Sposto et al., 1992; Sy and Taylor, 2000; Yu and Peng, 2008).

Kuk and Chen (1992) first proposed the Cox PH cure model which has attracted attention for decades. In clinical settings, Cox PH cure model has been widely used for modelling the failure time data for various types of cancer studies such as breast cancer, head and neck cancer, leukemia, prostate cancer, melanoma etc. (Amico and Van Keilegom, 2018; Othus et al., 2012; Peng and Dear, 2000; Peng and Taylor, 2014; Sy and Taylor, 2000, 2001; Zhao and Zhou, 2006).

While using the profile likelihood approach, we profile out the baseline hazard function from the Cox PH cure model and plugged the estimator in the likelihood function. However the problem is that the estimator of the baseline hazard function is an implicit function (see equations (5.12) and (5.16) and Remark-5.1 in section 5.3.3). Therefore, it is very challenging to find the estimated SE of the profile likelihood estimator. For this reason, Hsieh et al. (2006) proposed to use bootstrap method to get the standard errors while using the profile likelihood approach. However, in some studies, it is mentioned that the standard errors estimated by bootstrap method can be overestimated (Hickey et al., 2018; Hsieh et al., 2006; Xu et al., 2014).

Sy and Taylor (2000, 2001) have used NPMLE approach along with EM algorithm to estimate both regression parameters and cumulative baseline hazard function from the Cox PH cure model. However, Sy and Taylor

(2000, 2001) have treated the baseline hazard function as a finite dimensional parameter for the standard error calculation, therefore, they have considered the Cox PH cure model as finite dimensional parametric model and calculated the observed information using the Hessian matrix. On the other hand, Lu (2008) has used Breslow-type estimator and extended the approach developed by Murphy et al. (1994) and Murphy (1995) for MLE calculation and used the result of Murphy and Van der Vaart (2000) for the information matrix of profile likelihood using the equation (1.2) from Chapter-1.

In summary, it can be said that all the existing works of Cox PH cure model based on profile likelihood approach have avoided dealing with the implicit function (in profile likelihood function) to show the asymptotic normality of profile likelihood MLE.

Cai et al. (2012) developed the *smcure* R-package to fit the semiparametric mixture cure model which has received much attention in recent years (Amico and Van Keilegom, 2018; Peng and Taylor, 2014; Robinson, 2014). In this package, Cai et al. (2012) have used the melanoma data from the ECOG phase III clinical trial e1684 where Breslow-type estimator along with EM algorithm has been used in the Cox PH cure model to estimate the regression coefficients and cumulative baseline hazard function. However, the estimating equation was not simple due to the presence of implicit function, therefore they didn't calculate the standard errors directly. As a result, bootstrap samples have been used in order to obtain the variance estimator (Cai et al., 2012).

Our contribution is that we have shown the asymptotic normality of the profile likelihood estimator with implicit function for the Cox PH cure model. This enables us to have a closed form expression for the standard

error of the profile likelihood estimator (Lemma-6.3 and Theorem-6.3 in Chapter-6). This is an alternative methodology to Murphy and Van der Vaart (2000). Moreover, we have calculated the estimated information matrix as

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n \phi(x_i; \hat{\theta}) \phi'(x_i; \hat{\theta}),$$

where, ϕ is the profile score function.

Our proposed alternative approach gives us not only the direct asymptotic expansion of the profile likelihood, but also closed form formula to compute the efficient information matrix for the Cox PH cure model using the profile likelihood score function. Based on our proposed approach, we have calculated the estimated standard error of the profile likelihood estimator which is illustrated in the simulation study (Section-5.4) and real-life example (Section-5.5). For the real-data example, we have used the same data (ECOG phase III clinical trial e1684) from `smcure` package and compared our results with the output obtained from the `smcure` package.

This chapter is organized as follows. A brief discussion on Cox PH cure model has been given in Section-5.2. In Section-5.3, we describe the estimation procedure to show that the profile likelihood estimators are consistent and asymptotically normal. Results obtained from the simulation study and real-life data application are shown in Section-5.4 and Section-5.5 respectively. This chapter concludes in Section-5.6 with a short discussion.

5.2 Cox PH Cure Model

Let us define a binary variable V , where $V = 0$ indicates an individual that will be a long-term survivor (never experience the event of interest) and $V = 1$ indicates an individual that will experience the event. For an individual with covariate vector $W = (1, W_1, \dots, W_n)$, the distribution of $V = 1$ can be

expressed as a logistic model

$$p = Pr(V = 1, W; b) = \frac{e^{b'W}}{1 + e^{b'W}}, \quad (5.1)$$

where p is the probability of being susceptible (often called incidence of the model), b is a vector parameter and W include the intercept. The time to experience the event among individuals for which $V = 1$ can be modelled by Cox PH model

$$\lambda(t|V = 1, Z; \beta) = \lambda(t|V = 1)e^{\beta'Z}, \quad (5.2)$$

where we observe another set of covariate Z without intercept and $\lambda(t|V = 1)$ is the baseline hazard function. The two sets of covariates may be identical, or partially or completely different from each other (Kuk and Chen, 1992).

An individual who experience the event at time t contributes a likelihood factor

$$pf(t|V = 1, Z; \lambda, \beta),$$

which is the probability of death at time t (Kuk and Chen, 1992). On the other hand, an individual who has been followed to time t without experiencing the event contributes a likelihood factor

$$(1 - p) + pS(t|V = 1, Z; \lambda, \beta),$$

which is the probability of long-term survivor (cure) plus the probability of experiencing the event after time t (Kuk and Chen, 1992). In addition $S(t|V = 1, Z; \lambda, \beta) = S_0(t|V = 1)e^{\beta'Z}$ is the conditional survival function of the susceptibles (often called the latency) where $S_0(t|V = 1) = \exp(-\Lambda(t|V = 1)) = \exp(-\int_0^t \lambda(s|V = 1)ds)$ is the baseline survival function and $\Lambda(t|V = 1)$ is the cumulative baseline hazard function.

5.3 Estimation: Profile Likelihood with EM Algorithm

In this section, first we present the likelihood function for the Cox PH cure model. In Section 5.3.1, we profile out the baseline hazard function using NPMLE approach and construct the profile likelihood function. In Section 5.3.2, we estimate the profile likelihood MLE through EM algorithm and in Section 5.3.3, we prove the asymptotic normality of the profile likelihood MLE.

Suppose the observed data for individual i can be denoted by $(T_i, \delta_i, Z_i); i = 1, 2, \dots, n$ where T_i is the length of time a subject was observed, Z_i is a vector of covariates. Moreover, δ_i indicates whether the observed time is censored or not

$$\delta_i = \begin{cases} 1 & \text{for } T_i = \text{event time} \\ 0 & \text{for } T_i = \text{censored time} \end{cases}$$

For convenience, let $W_i = (1, Z_i)'$, although the covariates in W_i and Z_i do not have to be equal.

The likelihood for n observations will be

$$L(b, \beta, \lambda) = \prod_{i=1}^n \left\{ p_i f(t_i | V = 1, Z_i; \lambda, \beta) \right\}^{\delta_i} \left\{ (1-p_i) + p_i S(t_i | V = 1, Z_i; \lambda, \beta) \right\}^{1-\delta_i}, \quad (5.3)$$

where p_i is the probability of i th individual being susceptible. Using

$$f(t | V = 1, Z; \lambda, \beta) = \lambda(t | V = 1, Z; \beta) S(t | V = 1, Z; \lambda, \beta),$$

the likelihood function can be written as

$$L(b, \beta, \Lambda) = \prod_{i=1}^n \left[p_i \lambda(t_i | V = 1, Z_i; \beta) S(t_i | V = 1, Z_i; \lambda, \beta) \right]^{\delta_i} \left[(1 - p_i) + p_i S(t_i | V = 1, Z_i; \lambda, \beta) \right]^{1-\delta_i}. \quad (5.4)$$

Here we want to obtain the estimates of b and β that maximize $L(b, \beta, \Lambda)$. For maximizing $L(b, \beta, \Lambda)$, we are going to apply profile likelihood technique in which $\Lambda(t)$ is profiled out from the likelihood.

Let us define the complete data by $(t_i, \delta_i, Z_i, v_i)$, $i = 1, \dots, n$ which includes the observed data and unobserved v_i , where v_i is the value taken by the variable V_i . It follows that if $\delta_i = 1$ then $v_i = 1$ and if $\delta_i = 0$ then v_i is unobserved. The choice for using EM algorithm (Dempster et al., 1977) is justified by the fact that the model depends on a latent variable, v_i (cure status). So the complete data likelihood can be written as

$$L_c(b, \beta, \Lambda; v) = \prod_{i=1}^n \left[p_i \lambda(t_i | V = 1, Z_i; \beta) S(t_i | V = 1, Z_i; \lambda, \beta) \right]^{\delta_i v_i} \times \prod_{i=1}^n \left[p_i S(t_i | V = 1, Z_i; \lambda, \beta) \right]^{(1-\delta_i)v_i} \times \prod_{i=1}^n \left[1 - p_i \right]^{(1-\delta_i)(1-v_i)}. \quad (5.5)$$

The above equation can be rewritten as the product of a logistic and a PH component.

$$L_c(b, \beta, \Lambda; v) = \prod_{i=1}^n p_i^{v_i} (1 - p_i)^{1-v_i} \times \prod_{i=1}^n \lambda(t_i | V = 1, Z_i; \beta)^{\delta_i v_i} S(t_i | V = 1, Z_i; \lambda, \beta)^{v_i}. \quad (5.6)$$

So it is possible to estimate the incidence and the latency separately (Amico and Van Keilegom, 2018). Now the expected complete data log-likelihood under $p(V|T, \delta, Z)$ is

$$\sum_{i=1}^n \left\{ \gamma(V_i) \log p_i + (1 - \gamma(V_i)) \log(1 - p_i) \right\} + \sum_{i=1}^n \gamma(V_i) \left\{ \delta_i \log \lambda(t_i | V = 1, Z_i; \beta) + \log S(t_i | V = 1, Z_i; \lambda, \beta) \right\}, \quad (5.7)$$

where $\gamma(V_i)$ can be defined as

$$\gamma(V_i) = E(V_i | T_i, \delta_i, Z_i) = \left(\frac{p_i S(t_i | V = 1, Z_i; \lambda, \beta)}{1 - p_i + p_i S(t_i | V = 1, Z_i; \lambda, \beta)} \right)^{1-\delta_i}. \quad (5.8)$$

Here, for censored cases $\gamma(V_i) = E(V_i|T_i, \delta_i, Z_i)$ and for uncensored cases $\gamma(V_i) = 1$. To estimate all parameters and the baseline hazards simultaneously, we combine the EM algorithm and profile likelihood approach. From equation (5.6), it can be observed that the likelihood function for the logistic component is same as for a classical logistic regression model.

5.3.1 Baseline Hazard Estimation and Profile Likelihood Function

Before starting the EM algorithm, we profile out the baseline hazard function $\lambda(t)$ using NPMLE. We assume that the baseline hazard, $\lambda(t)$ has (possible) non-zero values on the observed time t_i :

$$\lambda(t_i) = \lambda_i \geq 0 ; i = 1, 2, \dots, n$$

and $\lambda(t) = 0$ for the rest of time (Murphy and Van der Vaart, 2000). Under the assumption, the integral, $\Lambda(t) = \int_0^t \lambda(s)ds$ is replaced with the sum

$$\Lambda(t) = \sum_{i=1}^n \lambda_i 1\{t_i \leq t\}.$$

Now the survival part of equation (5.7) can be separately maximized with respect to λ using the log-likelihood:

$$\sum_{i=1}^n \gamma(V_i) \left[\delta_i \{ \log \lambda_i + \beta' Z_i \} - e^{\beta' Z_i} \sum_{j=1}^n \lambda_j 1\{t_j \leq t_i\} \right]. \quad (5.9)$$

The score equation with respect to λ_k is

$$\frac{\partial}{\partial \lambda_k} \sum_{i=1}^n \gamma(V_i) \left[\delta_i \{ \log \lambda_i + \beta' Z_i \} - e^{\beta' Z_i} \sum_{j=1}^n \lambda_j 1\{t_j \leq t_i\} \right] = 0.$$

By solving the above equation, the estimate of λ_k is

$$\hat{\lambda}_k = \frac{\delta_k}{\sum_{l=1}^n \gamma(V_l) 1\{t_k \leq t_l\} e^{\beta' Z_l}},$$

and the estimate of the cumulative baseline hazard, $\Lambda(t)$ is

$$\hat{\Lambda}(t|V = 1; \beta) = \sum_{i=1}^n \frac{\delta_i 1\{t_i \leq t\}}{\sum_{l=1}^n \gamma(V_l) 1\{t \leq t_l\} e^{\beta' Z_l}}. \quad (5.10)$$

Now the estimate of the baseline survival function can be expressed as

$$\hat{S}_0(t|V = 1; \beta) = \exp(-\hat{\Lambda}(t|V = 1; \beta)) = \exp\left(-\sum_{i=1}^n \frac{\delta_i 1\{t_i \leq t\}}{\sum_{l=1}^n \gamma(V_l) 1\{t \leq t_l\} e^{\beta' Z_l}}\right).$$

Note: $\hat{S}_0(t|V = 1; \beta)$ may not converge to 0 as $t \rightarrow \infty$, therefore in smcure R-package, Cai et al. (2012) have considered $\hat{S}_0(t|V = 1; \beta) = 0$ for all the time points after the last event time. To compare our results with the output obtained from smcure package, we have considered the same assumption.

Profile Likelihood Function

An estimator of the baseline cumulative hazard function in the counting process notation can be written from equation (5.10) as

$$\hat{\Lambda}(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n \gamma(V_i) Y_i(u) e^{\beta' Z_i}}, \quad (5.11)$$

where $N(t) = 1\{T \leq t, \delta = 1\}$ and $Y(t) = 1\{T \geq t\}$.

Let us denote $E_{F_n} f = \int f dF_n$. Then $\hat{\Lambda}(t)$ can be expressed as

$$\hat{\Lambda}_{\beta, F_n}(t) = \int_0^t \frac{E_{F_n} dN(u)}{E_{F_n} \gamma(V) Y(u) e^{\beta' Z}}. \quad (5.12)$$

Now we plug the baseline hazard estimator ($\hat{\Lambda}_{\beta, F_n}$ in the place of Λ) in equation (5.7) and obtain the expected complete data profile log-likelihood function:

$$\sum_{i=1}^n \left\{ \log P(V_i|b) + \log P(T_i, \delta_i | \beta, \hat{\Lambda}_{\beta, F_n}) \right\}, \quad (5.13)$$

where $\log P(V_i|b)$ and $\log P(T_i, \delta_i | \beta, \hat{\Lambda}_{\beta, F_n})$ are the expected complete data profile log-likelihood functions (for one observation) for logistic and Cox PH component respectively. Now we can express the components as

$$\begin{aligned} \log P(V_i|b) &= \{ \gamma(V_i) \log p_i + (1 - \gamma(V_i)) \log(1 - p_i) \} \\ &= \gamma(V_i) b' W_i - \log(1 + e^{b' W_i}), \end{aligned} \quad (5.14)$$

and

$$\begin{aligned}
& \log P(T_i, \delta_i | \beta, \hat{\Lambda}_{\beta, F_n}) \\
&= \gamma(V_i) \left\{ \delta_i \log \hat{\lambda}(t_i | V = 1, Z_i; \beta) + \log S(t_i | V = 1, Z_i; \hat{\Lambda}_{\beta, F_n}, \beta) \right\} \\
&= \gamma(V_i) \left[\delta_i \left\{ \log \frac{E_{F_n} dN(T_i)}{E_{F_n} \gamma(V) Y(T_i) e^{\beta' Z}} + \beta' Z_i \right\} - e^{\beta' Z_i} \int_0^{T_i} \frac{E_{F_n} dN(u)}{E_{F_n} \gamma(V) Y(u) e^{\beta' Z}} \right].
\end{aligned} \tag{5.15}$$

For simplicity, we can express $\log P(T_i, \delta_i | \beta, \hat{\Lambda}_{\beta, F_n})$ as $\log P(T_i, \delta_i | \beta, F_n)$.

5.3.2 The EM Algorithm

In this section, we will apply the EM algorithm to maximize the expected complete data profile log-likelihood function to find the profile likelihood MLE. First, we replace Λ with $\hat{\Lambda}_{\beta, F_n}$ in equation (5.8).

The E-step

In the E-step, we use the current parameter estimates b and β to find the expected values of V_i :

$$\gamma(V_i) = E(V_i | T_i, \delta_i, Z_i) = \left(\frac{p_i S(t_i | V = 1, Z_i; \beta, \hat{\Lambda}_{\beta, F_n})}{1 - p_i + p_i S(t_i | V = 1, Z_i; \beta, \hat{\Lambda}_{\beta, F_n})} \right)^{1 - \delta_i}. \tag{5.16}$$

The M-step

Now we maximize the equation (5.13) with respect to b and β to obtain \hat{b}_n and $\hat{\beta}_n$ respectively. The updated parameters from the M-step are returned into E-step until the values of \hat{b}_n and $\hat{\beta}_n$ converge.

In M step of EM algorithm, we have used non-linear optimization routine ("Nelder-Mead" method) to maximize the complete data log likelihood function.

5.3.3 Asymptotic Normality of the MLE

Here we outline the results of asymptotic normality of the profile likelihood estimator in the Cox PH cure model.

The difficulty of the problem is the profile likelihood function involve the implicit function. From equations (5.12) and (5.16), we see that $\gamma(V)$ and $\hat{\Lambda}$ have an iterative relationship. As a result, the estimator of the baseline hazard function is an implicit function. Similar type of problem has been identified in the joint model of longitudinal and survival data [Rizopoulos (2012); page-67].

We have successfully proved the asymptotic normality of the profile likelihood estimator based on two results:

1. The score functions defined in equations (5.19), (5.20) and (5.21) in Section 5.3.4 don't involve differentiation of the implicit function and
2. Without assuming the differentiability of the profile likelihood score function (ϕ_s) with respect to β and F , we have showed that

$$\begin{aligned} E \left[\sqrt{n} \left\{ \phi_s(T, \delta | \hat{\beta}_n, F_0) - \phi_s(T, \delta | \beta_0, F_0) \right\} \right] \\ = -E \left[\phi_s(T, \delta | \beta_0, F_0) \phi'_s(T, \delta | \beta_0, F_0) \right] \left\{ \sqrt{n}(\hat{\beta}_n - \beta_0) \right\} + o_p(1). \end{aligned}$$

The proof for the above equation is provided in Lemma-6.3 of Chapter-6. By combining the above two results, we are able to show not only the asymptotic normality of the estimator but also derive the closed form expression for information matrices (equations 5.17 and 5.18) without differentiating the implicit function in the model.

The asymptotic properties of $\hat{\beta}_n$ and \hat{b}_n are proved in Theorem-6.1, Theorem-6.2 and Theorem-6.3 (in Chapter-6). The EM estimators \hat{b}_n and $\hat{\beta}_n$ have asymptotic normal distribution which can be expressed as

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{D} N\{0, (I_s^*)^{-1}\}$$

and

$$\sqrt{n}(\hat{b}_n - b_0) \xrightarrow{D} N\{0, (I_l^*)^{-1}\},$$

where (β_0, b_0) are the true values of (β, b) and, $I_s^* = E[\phi_s \phi_s']$ and $I_l^* = E[\phi_l \phi_l']$ are the efficient information matrices for survival and logistic part respectively. ϕ_l and ϕ_s given in (5.20) and (5.21) are the score functions for logistic and survival part respectively. In Theorem-6.1 it is shown that at the true value of parameters, ϕ_l and ϕ_s are efficient score functions. The consistency of \hat{b}_n and $\hat{\beta}_n$ are provided in Theorem-6.2 and the asymptotic normality of \hat{b}_n and $\hat{\beta}_n$ are given in Theorem-6.3. Moreover, the estimated information matrices can be computed as

$$\hat{I}_l = \frac{1}{n} \sum_{i=1}^n \phi_l(V_i | \hat{b}_n) \phi_l'(V_i | \hat{b}_n), \quad (5.17)$$

and

$$\hat{I}_s = \frac{1}{n} \sum_{i=1}^n \phi_s(T_i, \delta_i | \hat{\beta}_n, F_n) \phi_s'(T_i, \delta_i | \hat{\beta}_n, F_n). \quad (5.18)$$

Note: Assumptions, Lemmas and Theorems are provided in Chapter-6.

Remark-5.1: The existing approaches don't involve the differentiation of the profile likelihood function. These methods have used an 'approximate least favorable submodel' proposed by Murphy and Van der Vaart (2000) to avoid the differentiation of profile likelihood function and therefore they have used the equation (1.2) from Chapter-1 to calculate the efficient information matrix.

5.3.4 Score Functions

The score functions for the profile likelihood are

$$\phi(V_i, T_i, \delta_i | b, \beta, F_n) = \phi_l(V_i | b) + \phi_s(T_i, \delta_i | \beta, F_n), \quad (5.19)$$

where $\phi_l(V_i | b)$ is the score function for logistic component which can be expressed as

$$\phi_l(V_i | b) = \frac{\partial}{\partial b} \log P(V_i | b) = \gamma(V_i) W_i - \frac{W_i e^{b' W_i}}{1 + e^{b' W_i}}, \quad (5.20)$$

and $\phi_s(T_i, \delta_i|\beta, F_n)$ is the score function for survival component which can be written as

$$\begin{aligned}\phi_s(T_i, \delta_i|\beta, F_n) &= \frac{\partial}{\partial \beta} \log P(T_i, \delta_i|\beta, F_n) \\ &= \gamma(V_i) \left\{ \delta_i \left[Z_i - \frac{E_{F_n} \gamma(V) Y(T_i) Z e^{\beta' Z}}{E_{F_n} \gamma(V) Y(T_i) e^{\beta' Z}} \right] \right. \\ &\quad \left. - e^{\beta' Z_i} \int_0^{T_i} \left[Z_i - \frac{E_{F_n} \gamma(V) Y(u) Z e^{\beta' Z}}{E_{F_n} \gamma(V) Y(u) e^{\beta' Z}} \right] d\hat{\Lambda}_{\beta, F_n}(u) \right\}. \quad (5.21)\end{aligned}$$

Now we will calculate the score operator $B(T_i, \delta_i|\beta, F)$, which is Hadamard derivative with respect to F . For an integrable function h with the same domain as F , we can express

$$\begin{aligned}B(T_i, \delta_i|\beta, F)h &= d_F \log P(T_i, \delta_i|\beta, \hat{\Lambda}_{\beta, F})h \\ &= \gamma(V_i) \left\{ \delta_i \left[\frac{E_h dN(T_i)}{E_F dN(T_i)} - \frac{E_h \gamma(V) Y(T_i) e^{\beta' Z}}{E_F \gamma(V) Y(T_i) e^{\beta' Z}} \right] - e^{\beta' Z_i} \int_0^{T_i} \frac{E_h dN(u)}{E_F \gamma(V) Y(u) e^{\beta' Z}} \right. \\ &\quad \left. + e^{\beta' Z_i} \int_0^{T_i} \frac{E_F dN(u) E_h Y(u) \gamma(V) e^{\beta' Z}}{(E_F Y(u) \gamma(V) e^{\beta' Z})^2} \right\}.\end{aligned}$$

where, $d_F \log P(T_i, \delta_i|\beta, \hat{\Lambda}_{\beta, F})$ represents the Hadamard derivative of $\log P(T_i, \delta_i|\beta, \hat{\Lambda}_{\beta, F})$ with respect to F (Hirose, 2011b).

5.4 Simulation Study

We are going to perform a simulation study where our goal is to compare and contrast the smcure package with our approach by assessing parameter and standard error estimation. Survival times were generated from Weibull proportional hazards model using 'simsurv' R-package (Brilleman et al., 2021). Moreover, censoring times were generated from uniform distribution, $U[a, b]$ with a and b defined in order to provide chosen censoring rates. Simulation results for Cox PH cure model were evaluated with two covariates (fixed by design), one binary covariate from binomial distribu-

tion with probability 0.5 and one continuous covariate generated from normal distribution $N(1, 1)$. Therefore, the covariate vectors for logistic and survival components were $W = (1, W_1, W_2)$ and $Z = (Z_1, Z_2)$ respectively.

The cure rates were varied through the coefficients (b) corresponding to W . The slight cure rates for the treatment group ($W_1 = 1$) and control group ($W_1 = 0$) were 20% and 8% respectively, resulting from $b = (2.1, -1, 0.3)$. The moderate cure rates for the treatment and control groups were 43% and 21% respectively, resulting from $b = (1, -1, 0.3)$. Moreover, The substantial cure rates for the treatment and control groups were 69% and 45% respectively, resulting from $b = (-0.1, -1, 0.3)$. For each configuration, mean was chosen as the value of the continuous covariate (W_2). Moreover, the coefficient vector for survival part was $\beta = (-1, 0.5)$. For each setting, the censoring rates relative to the defined cure rates and their effects were also evaluated. These results include a sample of 150 and 300 individuals with 500 replications from both *smcure* package and our approach. The computation time required for different sample size and configurations has also compared between *smcure* R-package and our proposed approach.

The results from simulation studies such as mean bias, empirical SE (= the standard deviation of the parameter estimates), mean SE (=the mean of standard error estimates calculated for each fitted model), mean square error (MSE) and 95% confidence interval coverage probabilities (95% CI cap) for each configuration are given in Tables 5.1-5.6. Details of mean SE calculation is given later in this section (see Note).

The explicit explanation of the above mentioned statistics are given here. Let T be the estimator for the parameter of interest θ . From our simulation study ($K = 500$ replicates), we obtain K estimates T_1, \dots, T_K for each simulated data. Based on these estimators, we can compute K biases such as

$\text{bias}_1 = T_1 - \theta, \text{bias}_2 = T_2 - \theta, \dots, \text{bias}_K = T_K - \theta$. Moreover, we obtain K estimates of \hat{SE} : $\hat{SE}_1, \hat{SE}_2, \dots, \hat{SE}_K$. Now we can calculate the following sample statistics for the estimates:

- Sample mean, $\bar{T} = \sum_{i=1}^K T_i$
- Mean bias = $\frac{1}{K} \sum_{i=1}^K (T_i - \bar{T})$
- Empirical standard error = $\sqrt{\frac{1}{K-1} \sum_{i=1}^K (T_i - \bar{T})^2}$
- $MSE = \frac{1}{K} \sum_{i=1}^K (T_i - \theta)^2$
- mean $\hat{SE} = \frac{1}{K} \sum_{i=1}^K \hat{SE}_i$
- 95% CI cap = proportion of confidence intervals contains the true value θ

Note: Method of mean SE calculation

In *smcure* package, due to the presence of implicit function in the estimating equation, the SE of estimated parameters are not directly available. As a result, the package has used the bootstrap samples to compute the standard errors of estimated parameters. Here, 100 bootstrap samples have been considered for all simulation settings. On the other hand, in our proposed approach, we have used the profile likelihood score function to compute the estimated SE analytically through the inverse of equations (5.17) and (5.18). For each simulated data, we can express the observed information matrix as

$$\hat{I} = \frac{1}{n} \sum_{m=1}^n \phi_m \phi_m' ,$$

where ϕ is the profile score function and n is the number of individuals in each simulated data. Moreover, the estimated SE for each fitted model can be expressed as

$$\hat{SE} = \sqrt{\hat{I}^{-1}}.$$

Results

From Tables 5.1-5.6, we can observe that for both *smcure* package and our approach, the parameter estimates are close to the true values and mean biases are very small (consistently less than 0.1 with most less than 0.05). For all settings in both approaches, the bias and MSE decrease with increasing sample size as expected and with only a few exceptions, the mean SE of the parameters are very close for both *smcure* package and our approach. The capture rates based on the 95% confidence interval are also relatively similar for both *smcure* package and our approach.

Computation Time Analysis

In order to assess and compare the computational time between our proposed approach and *smcure* package, the average time elapsed to get the estimates of the parameters and estimated standard errors has recorded. For all configurations, 500 datasets have generated as described before. The average computation times for parameter estimates and standard errors for both approaches are given in Tables 5.1-5.6.

We can see from the Tables 5.1-5.6 that the the average computation times for our proposed approach are much less compared to *smcure* package. The reason for taking more time in *smcure* package is incorporating bootstrap process for standard error estimation. On the other hand, we have used profile likelihood score function to compute the estimated SE which provides us a closed form expression to compute the efficient information matrix, therefore the average computation time for our approach is much less compared to the *smcure* package.

5.5 Application to Eastern Cooperative Oncology Group (ECOG) Data

We have used the melanoma data (ECOG phase III clinical trial e1684) from *smcure* package as a numerical example to compare our results with the output obtained from *smcure* package.

In the dataset, the subjects had melanoma cancer and were treated with interferon alpha-2b (IFN) regimen. The main aim of this research was to investigate the effects of high dose interferon alpha-2b (IFN) regimen against the placebo as the postoperative adjuvant therapy. In this example, recurrence of melanoma cancer has been defined as the event and observed relapse free-survival has defined as failure time. A total number of 284 observations has been used for the statistical analysis. Three covariates are considered: gender (0=male,1=female), treatment (0=control,1=treatment) and age (continuous variable which is centered to the mean) for both the incidence and latency parts.

Out of 284 individuals, 196 had melanoma cancer recurring (approximately 31% censoring rate). The observed follow-up time of the individuals ranged from 0.032 to 9.643 years. The parameter estimates, standard errors and 95% CI using *smcure* package and our approach (for logistic and Cox PH components) are given in Table 5.7 and Table 5.8 respectively.

From Table 5.7, we observed that in both *smcure* package and our approach, only intercept is significant in determining the long term incidence. The result for treatment (insignificant negative estimates) suggests that the probability of recurring melanoma for control group is higher compared to the treatment group. However, age and sex both are insignificant on *smcure* package and our approach. The positive estimates for age indicates a higher

(though insignificant) melanoma recurrence rate for the older patients. On the other hand, from Table 5.8 it is observed that in both `smcure` package and our approach, all the covariates have insignificant effect on latency. The insignificant negative estimates for treatment and age indicate later recurrence times of melanoma for IFN group and older patients respectively. In the case of sex, insignificant positive estimate indicates earlier recurrence times of melanoma for females.

Though both methods (`smcure` package and our approach) have provided similar parameter estimates, however, the estimated SE are little bit different from each other. The reason is the calculation procedure of SE which we mentioned in the previous section (simulation study). In `smcure` package, the SE of estimated parameters have calculated using bootstrap samples, whereas we have used the profile likelihood score function and got the closed form for the information matrix.

5.6 Discussion

Over the years, many techniques have been proposed where the direct expansion of the profile likelihood function has been avoided by Cai et al. (2012); Lu (2008); Peng and Dear (2000); Sy and Taylor (2000, 2001). Therefore, these approaches didn't derive the closed form solution for standard error calculation.

In this chapter, we have expanded the profile log-likelihood function directly to show the asymptotic normality of the maximum profile likelihood estimator and derived the closed form expression for the empirical observed information matrix for the Cox PH cure model to estimate the standard errors. This is an additional method to compute standard errors for the maximum profile likelihood estimator which shows how natural and easy our

approach is to compute the estimate of the SE estimator.

For the application, we have performed the simulation study and analysed the real-life data, and compared our results with the output obtained from *smcure* package. Our proposed approach and the *smcure* package have provided similar and comparable parameter estimates and the estimated SE of the estimators. In this chapter, we have used the profile likelihood score function to calculate the estimated SE and got the explicit form of the SE estimator, whereas the *smcure* package has used bootstrap samples which are computationally expensive. Therefore, the computation time for our proposed approach is much less compared to the *smcure* R-package.

Table 5.1: Simulation results for Cox PH cure model with cure rate 20% for treatment group and 08% for control group (number of individuals=150)

Parameter	True value	Censoring rate	SMCURE package						Our approach					
			Mean Bias	Mean SE	Empirical SE	MSE	95% CI cap	Average Computation Time (seconds)	Mean Bias	Mean SE	Empirical SE	MSE	95% CI cap	Average Computation Time (seconds)
b0	2.1	U[0,25] 16.01%	0.0949	1.2354	0.4981	0.2567	0.954	3.9270	0.0948	0.4969	0.4979	0.2564	0.972	0.8121
b1	-1		-0.0835	1.3345	0.5575	0.3172	0.974		-0.0833	0.5538	0.5574	0.3170	0.972	
b2	0.3		0.0136	0.2520	0.2406	0.0580	0.968		0.0135	0.2539	0.2406	0.0580	0.968	
β_1	-1		-0.0317	0.2058	0.1943	0.0387	0.966		-0.0318	0.2023	0.1944	0.0387	0.966	
β_2	0.5		0.0058	0.1021	0.1024	0.0105	0.946		0.0057	0.1006	0.1024	0.0105	0.954	

Table 5.2: Simulation results for Cox PH cure model with cure rate 20% for treatment group and 08% for control group (number of individuals=300)

Parameter	True value	Censoring rate	SMCURE package						Our approach					
			Mean Bias	Mean SE	Empirical SE	MSE	95% CI cap	Average Computation Time (seconds)	Mean Bias	Mean SE	Empirical SE	MSE	95% CI cap	Average Computation Time (seconds)
b0	2.1	U[0,25] 17.31%	0.0334	0.3184	0.3198	0.1032	0.954	7.6673	0.0332	0.3278	0.3198	0.1031	0.966	2.8907
b1	-1		-0.0193	0.4099	0.3684	0.1358	0.978		-0.0191	0.3780	0.3684	0.1358	0.964	
b2	0.3		0.0184	0.1756	0.1750	0.0309	0.952		0.0185	0.1739	0.1750	0.0308	0.946	
β_1	-1		-0.0106	0.1444	0.1475	0.0218	0.936		-0.0105	0.1428	0.1475	0.0218	0.934	
β_2	0.5		0.0065	0.0735	0.0740	0.0055	0.942		0.0066	0.0725	0.0740	0.0055	0.944	

Table 5.3: Simulation results for Cox PH cure model with cure rate 43% for treatment group and 21% for control group (number of individuals=150)

Parameter	True value	Censoring rate	SMCURE package						Our approach					
			Mean Bias	Mean SE	Empirical SE	MSE	95% CI cap	Average Computation Time (seconds)	Mean Bias	Mean SE	Empirical SE	MSE	95% CI cap	Average Computation Time (seconds)
b0	1	U[0,25] 33.55%	0.0308	0.3441	0.3526	0.1250	0.952	3.8135	0.0309	0.3673	0.3526	0.1250	0.972	0.9389
b1	-1		-0.0247	0.4038	0.3866	0.1498	0.968		-0.0246	0.3832	0.3866	0.1498	0.958	
b2	0.3		-0.0025	0.2400	0.2264	0.0511	0.972		-0.0026	0.2332	0.2263	0.0510	0.966	
β_1	-1		-0.0175	0.2420	0.2452	0.0603	0.950		-0.0176	0.2342	0.2452	0.0604	0.938	
β_2	0.5		0.0144	0.1411	0.1387	0.0194	0.954		0.0144	0.1357	0.1387	0.0193	0.948	

Table 5.4: Simulation results for Cox PH cure model with cure rate 43% for treatment group and 21% for control group (number of individuals=300)

Parameter	True value	Censoring rate	SMCURE package						Our approach					
			Mean Bias	Mean SE	Empirical SE	MSE	95% CI cap	Average Computation Time (seconds)	Mean Bias	Mean SE	Empirical SE	MSE	95% CI cap	Average Computation Time (seconds)
b0	1	U[0.25] 33.63%	0.0165	0.2157	0.2570	0.0662	0.920	6.2447	0.0166	0.2447	0.2570	0.0661	0.950	2.6732
b1	-1		-0.0106	0.2717	0.2770	0.0767	0.944		-0.0107	0.2671	0.2770	0.0768	0.944	
b2	0.3		-0.0008	0.1433	0.1371	0.0188	0.958		-0.0007	0.1415	0.1371	0.0188	0.954	
β_1	-1		-0.0129	0.1627	0.1556	0.0243	0.944		-0.0130	0.1610	0.1556	0.0242	0.952	
β_2	0.5		0.0019	0.0820	0.0824	0.0068	0.940		0.0018	0.0815	0.0824	0.0067	0.954	

Table 5.5: Simulation results for Cox PH cure model with cure rate 69% for treatment group and 45% for control group (number of individuals=150)

Parameter	True value	Censoring rate	SMCURE package						Our approach					
			Mean Bias	Mean SE	Empirical SE	MSE	95% CI cap	Average Computation Time (seconds)	Mean Bias	Mean SE	Empirical SE	MSE	95% CI cap	Average Computation Time (seconds)
b0	-0.1	U[0,30] 57.95%	-0.0224	0.2327	0.2708	0.0737	0.910	3.5046	-0.0222	0.2805	0.2708	0.0738	0.964	1.0191
b1	-1		-0.0147	0.3678	0.3599	0.1295	0.950		-0.0146	0.3584	0.3599	0.1294	0.946	
b2	0.3		0.0174	0.1811	0.1707	0.0294	0.968		0.0175	0.1765	0.1707	0.0294	0.966	
β_1	-1		-0.0280	0.3402	0.3338	0.1120	0.954		-0.0281	0.3180	0.3338	0.1121	0.944	
β_2	0.5		0.0093	0.1618	0.1418	0.0202	0.970		0.0092	0.1524	0.1419	0.0201	0.966	

Table 5.6: Simulation results for Cox PH cure model with cure rate 69% for treatment group and 45% for control group (number of individuals=300)

Parameter	True value	Censoring rate	SMCURE package						Our approach					
			Mean Bias	Mean SE	Empirical SE	MSE	95% CI cap	Average Computation Time (seconds)	Mean Bias	Mean SE	Empirical SE	MSE	95% CI cap	Average Computation Time (seconds)
b0	-0.1	U[0,30] 57.86%	-0.0141	0.1729	0.2080	0.0434	0.896	5.2545	-0.0142	0.2075	0.2080	0.0433	0.950	2.9784
b1	-1		-0.0163	0.2522	0.2370	0.0563	0.964		-0.0161	0.2485	0.2370	0.0564	0.968	
b2	0.3		0.0140	0.1335	0.1329	0.0178	0.948		0.0141	0.1315	0.1329	0.0178	0.946	
β_1	-1		-0.0174	0.2187	0.2235	0.0501	0.940		-0.0173	0.2129	0.2235	0.0501	0.944	
β_2	0.5		0.0077	0.1084	0.1010	0.0102	0.958		0.0078	0.1054	0.1010	0.0101	0.960	

Table 5.7: Results for logistic component from ECOG data

Covariates	SMCURE package			Our approach		
	Estimates	SE	95% CI	Estimates	SE	95% CI
Intercept	1.3649	0.2877	(0.8012, 1.9286)	1.3665	0.2708	(0.8356, 1.8973)
Treatment	-0.5884	0.3065	(-1.1889, 0.0121)	-0.5891	0.3029	(-1.1829, 0.0048)
Age	0.0203	0.0145	(-0.0081, 0.0487)	0.0204	0.0118	(-0.0027, 0.0435)
Sex	-0.0869	0.3291	(-0.7319, 0.5581)	-0.0869	0.3027	(-0.6801, 0.5063)

Table 5.8: Results for survival component from ECOG data

Covariates	SMCURE package			Our approach		
	Estimates	SE	95% CI	Estimates	SE	95% CI
Treatment	-0.1535	0.1721	(-0.4908, 0.1838)	-0.1538	0.1469	(-0.4417, 0.1341)
Age	-0.0077	0.0067	(-0.0208, 0.0005)	-0.0076	0.0057	(-0.0188, 0.0035)
Sex	0.0995	0.1908	(-0.2744, 0.4734)	0.0991	0.1488	(-0.1926, 0.3907)

Chapter 6

Proofs of Theorems and Lemmas for Cox PH Cure Model

The proof of Theorems and Lemmas to show the asymptotic normality of profile likelihood estimator in the Cox PH cure model are given in this chapter. In Section 5.3.3 of Chapter-5, we have summarized the results of this chapter. To show the asymptotic normality of the MLE and its asymptotic variance, we have considered the following assumptions:

On the set of cdf functions F , we use the sup-norm, i.e., for $F, F_0 \in F$,

$$\|F - F_0\|_\infty = \sup_x |F(x) - F_0(x)|.$$

For $\rho > 0$, let

$$\zeta_\rho = \{F \in F : \|F - F_0\|_\infty < \rho\}.$$

The assumptions are given below

A1: We assume the event-time $T < \tau$ where τ a finite number ($\tau > 0$) such that $S(\tau) = P(T > \tau) = E[Y(\tau)] > 0$.

A2: The range of Z is bounded and β is in the compact set Θ which follows $\|Z\| \leq M$ and $\|\beta\| \leq M$ for some $0 < M < \infty$.

A3: The empirical cdf F_n is \sqrt{n} consistent i.e. $\sqrt{n}|F_n - F_0| = O_p(1)$.

A4: The efficient information matrix $I^* = E[\phi\phi']$ is invertible where ϕ is the efficient score function.

A5: True hazard function, $\lambda(t)$ is bounded and positive in $(0, \tau)$.

6.1 Theorem- 6.1 with proof

Theorem 6.1: At the true value of (b, β, F) , we are going to prove the followings

1. $\hat{\Lambda}_{\beta_0, F_0}(t) = \Lambda_0(t)$, the true cumulative hazard and
2. The score functions $\phi_t(V|b_0)$ and $\phi_s(T, \delta|\beta_0, F_0)$ defined in the equations (5.20) and (5.21) of Chapter-5 are the efficient score functions where we drop the subscript i .

Proof:

1. Replace F_n by F_0 , we get from equation (5.12)

$$\hat{\Lambda}_{\beta_0, F_0}(t) = \int_0^t \frac{E[dN(u)]}{E[\gamma(V)Y(u)e^{\beta'_0 Z}]},$$

where E is the expectation with respect to the true distribution F_0 . At the true value of the parameters (β, F) we can write

$$E[dN(u)] = E[\gamma(V)Y(u)e^{\beta'_0 Z}d\Lambda_0(u)]. \quad (6.1)$$

So from this point of view, we have $\hat{\Lambda}_{\beta_0, F_0}(t) = \Lambda_0(t)$. Hence (1) is proven.

2. We know that the logistic part is a parametric model that does not involve Λ , so we will work on the score function of survival part only. Now the score function for the survival part at the true value of the parameters (β, F) can be expressed as

$$\begin{aligned} \phi_s(T, \delta|\beta_0, F_0) &= \frac{\partial}{\partial \beta} \log P(T, \delta|\hat{\Lambda}_{\beta, F}, \beta) \Big|_{\beta=\beta_0, F=F_0} \\ &= \gamma(V) \left\{ \delta \left[Z - \frac{E[\gamma(V)Y(T)Ze^{\beta'_0 Z}]}{E[\gamma(V)Y(T)e^{\beta'_0 Z}]} \right] \right. \\ &\quad \left. - e^{\beta'_0 Z} \int_0^T \frac{E[dN(u)]}{E[\gamma(V)Y(u)e^{\beta'_0 Z}]} \left[Z - \frac{E[\gamma(V)Y(u)Ze^{\beta'_0 Z}]}{E[\gamma(V)Y(u)e^{\beta'_0 Z}]} \right] \right\}. \end{aligned}$$

Let $M_1(u) = E[\gamma(V)Y(u)Ze^{\beta'_0 Z}]$ and $M_0(u) = E[\gamma(V)Y(u)e^{\beta'_0 Z}]$. So by using equation (6.1), the above equation can be expressed as

$$\phi_s(T, \delta|\beta_0, F_0) = \gamma(V) \left\{ \delta \left[Z - \frac{M_1(T)}{M_0(T)} \right] - e^{\beta'_0 Z} \int_0^T \left[Z - \frac{M_1(u)}{M_0(u)} \right] d\Lambda_0(u) \right\}, \quad (6.2)$$

which is the profile likelihood score function for the survival part of the Cox PH cure model. In section 6.7, we showed that the profile likelihood score function is the efficient score function.

6.2 Lemma- 6.1 with Proof

Lemma-6.1: If the assumptions (A1-A5) hold, then

- (i) $P(T, \delta|\beta, F)$ is bounded away from zero.
- (ii) The class of functions $\left\{ \log P(T, \delta|\beta, F) : \beta \in \Theta, F \in \zeta_\rho \right\}$ is uniformly bounded Donsker.
- (iii) The class of functions $\left\{ \phi_s(T, \delta|\beta, F) : \beta \in \Theta, F \in \zeta_\rho \right\}$ is uniformly bounded Donsker.

Proof: For (i), we know

$$P(T, \delta|\beta, F) = \left[\frac{E_F dN(T)}{E_F \gamma(V) Y(T) e^{\beta'_0 Z}} \right]^{\delta \gamma(V)} \left\{ \left[\exp - \int_0^T \frac{E_F dN(u)}{E_F \gamma(V) Y(u) e^{\beta'_0 Z}} \right]^{e^{\beta'_0 Z}} \right\}^{\gamma(V)}. \quad (6.3)$$

Since the map $(f, F) \rightarrow E_F(f) = \int f dF$ is continuous, there is a constant $c > 0$, such that for all $F \in \zeta_\rho$ (based on A1), we can write

$$E_F[Y(\tau)] \geq c > 0.$$

We know $\gamma(V) = \left(\frac{pS(\tau)}{1-p+pS(\tau)} \right)^{1-\delta}$, so we have

$$0 < \frac{cp}{1-p+cp} \leq \gamma(V) \leq 1. \quad (6.4)$$

On the basis of **A2**, we can write $e^{-M^2} \leq e^{\beta'Z} \leq e^{M^2}$. So the upper bound of $E_F\gamma(V)Y(u)e^{\beta'Z}$ can be expressed as

$$|E_F\gamma(V)Y(u)e^{\beta'Z}| \leq |E_Fe^{\beta'Z}| \leq e^{M^2}. \quad (6.5)$$

Now by using equation (6.4), we can write

$$E_F\gamma(V)Y(u)e^{\beta'Z} \geq \frac{c^2p}{1-p+cp}e^{-M^2} > 0. \quad (6.6)$$

For some constant $c_1 > 0$, we can write $0 < c_1 \leq E_FdN(u) \leq 1$. Since $E_F\gamma(V)Y(u)e^{\beta'Z}$ is bounded away from zero (equation 6.6), we get

$$0 < \frac{c_1}{e^{M^2}} \leq \frac{E_FdN(u)}{E_F\gamma(V)Y(u)e^{\beta'Z}} \leq \frac{e^{M^2}(1-p+cp)}{c^2p}.$$

When $\delta = 1$, from (6.3) we get

$$P(T, \delta|\beta, F) = \left[\frac{E_FdN(T)}{E_F\gamma(V)Y(T)e^{\beta'Z}} e^{\beta'Z} \right] \left[\exp - \int_0^T \frac{E_FdN(u)}{E_F\gamma(V)Y(u)e^{\beta'Z}} \right] e^{\beta'Z}$$

$$P(T, \delta|\beta, F) \geq \frac{c_1}{e^{M^4}} \left\{ e^{-\frac{e^{M^4}(1-p+cp)}{c^2p}} \right\},$$

and when $\delta = 0$,

$$P(T, \delta|\beta, F) = \left\{ \left[\exp - \int_0^T \frac{E_FdN(u)}{E_F\gamma(V)Y(u)e^{\beta'Z}} \right] e^{\beta'Z} \right\}^{\gamma(V)}$$

$$P(T, \delta|\beta, F) \geq \left\{ e^{-\frac{e^{M^4}(1-p+cp)}{c^2p}} \right\}.$$

From the above equations, we can write

$$P(T, \delta|\beta, F) \geq \min \left\{ e^{-\frac{e^{M^4}(1-p+cp)}{c^2p}}, \left(e^{-\frac{e^{M^4}(1-p+cp)}{c^2p}} \right) \frac{c_1}{e^{M^4}} \right\} > 0. \quad (6.7)$$

So finally we can say that $P(T, \delta|\beta, F)$ is bounded away from zero and hence (i) is proved.

For (ii), the profile log-likelihood function of the survival part for Cox PH cure model is

$$\log P(T, \delta|\beta, F) = \gamma(V) \left[\delta \left\{ \log \frac{E_FdN(T)}{E_F\gamma(V)Y(T)e^{\beta'Z}} + \beta'Z \right\} - e^{\beta'Z} \int_0^T \frac{E_FdN(u)}{E_F\gamma(V)Y(u)e^{\beta'Z}} \right]$$

We know the set of cdf functions F is uniformly bounded Donsker. Hence the subset $\zeta_\rho \subset F$ is uniformly bounded Donsker. The class of functions $\{N(t) : t \in [0, \tau]\}$ and $\{Y(t) : t \in [0, \tau]\}$ are uniformly bounded Donsker (Theorem 2.10.6 in Van der Vaart and Wellner (1996)).

The class of functions $\{\beta'Z : \beta \in \Theta\}$ is Lipschitz in β . So, by Theorem 2.10.6 in Van der Vaart and Wellner (1996), the class of functions $\{\beta'Z : \beta \in \Theta\}$ is uniformly bounded Donsker.

Since $e^{\beta'Z}$ is a Lipschitz continuous function, so by Theorem 2.10.6 in Van der Vaart and Wellner (1996), the class of functions $\{e^{\beta'Z} : \beta \in \Theta\}$ is uniformly bounded Donsker.

Since $\{Y(t) : t \in [0, \tau]\}$ and $\{e^{\beta'Z} : \beta \in \Theta\}$ are uniformly bounded Donsker, so by Example 2.10.8 of Van der Vaart and Wellner (1996), the class of functions $\{\gamma(V)Y(t)e^{\beta'Z} : t \in [0, \tau], \beta \in \Theta\}$ is uniformly bounded Donsker.

Since $E_F(f) = \int f dF$ is Lipschitz, so for the class of functions $\{E_F \gamma(V)Y(t)e^{\beta'Z} : t \in [0, \tau], \beta \in \Theta, F \in \zeta_\rho\}$, we can write

$$\begin{aligned}
& \left| E_{F_1} \gamma(V)Y(t)e^{\beta_1'Z} - E_{F_2} \gamma(V)Y(t)e^{\beta_2'Z} \right| \\
&= \int \gamma(V)Y(t) \left| e^{\beta_1'Z} - e^{\beta_2'Z} \right| dF_1 + \int \gamma(V)Y(t)e^{\beta_2'Z} d\|F_1 - F_2\| \\
&\leq \int \left| e^{\beta_1'Z} - e^{\beta_2'Z} \right| dF_1 + \int e^{\beta_2'Z} d\|F_1 - F_2\| \\
&\leq M e^{M^2} \|\beta_1 - \beta_2\| + e^{M^2} \|F_1 - F_2\|.
\end{aligned}$$

Let $\max\left(M e^{M^2}, e^{M^2}\right) = M e^{M^2}$, then the above equation can be expressed

as

$$\left| E_{F_1} \gamma(V) Y(t) e^{\beta'_1 Z} - E_{F_2} \gamma(V) Y(t) e^{\beta'_2 Z} \right| \leq M e^{M^2} \left(\|\beta_1 - \beta_2\| + \|F_1 - F_2\| \right), \quad (6.8)$$

which is Lipschitz in parameters (β, F) . So by Theorem 2.10.6 in Van der Vaart and Wellner (1996), the class of functions $\{E_F \gamma(V) Y(t) e^{\beta' Z} : t \in [0, \tau], \beta \in \Theta, F \in \zeta_\rho\}$ is uniformly bounded Donsker. Similarly the class of functions $\{E_F N(t) : t \in [0, \tau]\}$ is uniformly bounded Donsker.

Since $\{E_F \gamma(V) Y(t) e^{\beta' Z} : t \in [0, \tau], \beta \in \Theta, F \in \zeta_\rho\}$ is uniformly bounded Donsker and $E_F \gamma(V) Y(u) e^{\beta' Z}$ is bounded away from zero (equation 6.6), by Example 2.10.9 in Van der Vaart and Wellner (1996), the class of functions

$$\left\{ \frac{1}{E_F \gamma(V) Y(t) e^{\beta' Z}} : t \in [0, \tau], \beta \in \Theta, F \in \zeta_\rho \right\}$$

is uniformly bounded Donsker.

Since the map $(f, F) \rightarrow E_F(f) = \int f dF$ is Lipschitz, by Theorem 2.10.6 of Van der Vaart and Wellner (1996), the class of functions

$$\left\{ \int_0^t \frac{E_F dN(u)}{E_F \gamma(V) Y(u) e^{\beta' Z}} : t \in [0, \tau], \beta \in \Theta, F \in \zeta_\rho \right\}$$

is uniformly bounded Donsker.

Since $\{e^{\beta' Z} : \beta \in \Theta\}$ is uniformly bounded Donsker, so by Example 2.10.8 of Van der Vaart and Wellner (1996), the class of functions

$$\left\{ e^{\beta' Z} \int_0^t \frac{E_F dN(u)}{E_F \gamma(V) Y(u) e^{\beta' Z}} : t \in [0, \tau], \beta \in \Theta, F \in \zeta_\rho \right\}$$

is uniformly bounded Donsker.

Since the class $\{\beta' Z : \beta \in \Theta\}$ is uniformly bounded Donsker, by Example 2.10.7 (Van der Vaart and Wellner, 1996), the class of functions

$$\left\{ \log \frac{E_F dN(T)}{E_F \gamma(V) Y(T) e^{\beta' Z}} + \beta' Z : \beta \in \Theta, F \in \zeta_\rho \right\}$$

is uniformly bounded Donsker.

Since the map $(f, F) \rightarrow E_F(f) = \int f dF$ is Lipschitz, so by Theorem 2.10.6 in Van der Vaart and Wellner (1996), the class of functions $\left\{ \log P(T, \delta|\beta, F) : \beta \in \Theta, F \in \zeta_\rho \right\}$ is uniformly bounded Donsker. So (ii) is proven.

For (iii), we know the score function of the survival part for Cox PH cure model is

$$\phi_s(T, \delta|\beta, F) = \gamma(V) \left\{ \delta \left[Z - \frac{E_F \gamma(V) Y(T) Z e^{\beta' Z}}{E_F \gamma(V) Y(T) e^{\beta' Z}} \right] - e^{\beta' Z} \int_0^T \frac{E_F dN(u)}{E_F \gamma(V) Y(u) e^{\beta' Z}} \left[Z - \frac{E_F \gamma(V) Y(u) Z e^{\beta' Z}}{E_F \gamma(V) Y(u) e^{\beta' Z}} \right] \right\}.$$

Similar proof to (ii), we can show that the class of functions $\left\{ \phi_s(T, \delta|\beta, F) : \beta \in \Theta, F \in \zeta_\rho \right\}$ is uniformly bounded Donsker.

6.3 Lemma- 6.2 with Proof

Lemma-6.2: If the assumptions (A1-A5) hold, then

$$\left\| \phi_s(T, \delta|\hat{\beta}_n, F_n) - \phi_s(T, \delta|\beta_0, F_0) \right\| \leq M'' \left(\|\hat{\beta}_n - \beta_0\| + \|F_n - F_0\| + \|\hat{\Lambda}_{\hat{\beta}_n, F_n} - \hat{\Lambda}_{\beta_0, F_0}\| \right),$$

for some positive constant M'' .

Proof: From equation (5.21) of Chapter-5, the score function for the survival part is

$$\phi_s(T, \delta|\beta, F_n) = \gamma(V) \left\{ \delta \left[Z - \frac{E_{F_n} \gamma(V) Y(T) Z e^{\beta' Z}}{E_{F_n} \gamma(V) Y(T) e^{\beta' Z}} \right] - e^{\beta' Z} \int_0^T \left[Z - \frac{E_{F_n} \gamma(V) Y(u) Z e^{\beta' Z}}{E_{F_n} \gamma(V) Y(u) e^{\beta' Z}} \right] d\hat{\Lambda}_{\beta, F_n}(u) \right\}.$$

Define

$$\phi_s(T, \delta|\beta, F, \Lambda) = \gamma(V) \left\{ \delta \left[Z - \frac{E_F \gamma(V) Y(T) Z e^{\beta' Z}}{E_F \gamma(V) Y(T) e^{\beta' Z}} \right] - e^{\beta' Z} \int_0^T \left[Z - \frac{E_F \gamma(V) Y(u) Z e^{\beta' Z}}{E_F \gamma(V) Y(u) e^{\beta' Z}} \right] d\Lambda(u) \right\}.$$

Then the function is differentiable with respect to β , F and Λ . Now we have

$$\phi_s(T, \delta|\beta, F_n) = \phi_s(T, \delta|\beta, F_n, \hat{\Lambda}_{\beta, F_n}).$$

Similar to the proof of Lemma-6.1, we can show that the derivative of the score function will also be uniformly bounded with the bound M'' .

From these we can say that the class of functions $\{\phi_s(T, \delta|\beta, F, \Lambda) : \beta \in \Theta, F \in \zeta_\rho, \Lambda \in H\}$ is Lipschitz in parameters (β, F, Λ) and the result follows:

$$\begin{aligned} & \left\| \phi_s(T, \delta|\hat{\beta}_n, F_n) - \phi_s(T, \delta|\beta_0, F_0) \right\| \\ &= \left\| \phi_s(T, \delta|\hat{\beta}_n, F_n, \hat{\Lambda}_{\hat{\beta}_n, F_n}) - \phi_s(T, \delta|\beta_0, F_0, \hat{\Lambda}_{\beta_0, F_0}) \right\| \\ &\leq M'' \left(\|\hat{\beta}_n - \beta_0\| + \|F_n - F_0\| + \|\hat{\Lambda}_{\hat{\beta}_n, F_n} - \hat{\Lambda}_{\beta_0, F_0}\| \right). \end{aligned}$$

6.4 Theorem- 6.2 with proof

Theorem 6.2: If the assumptions (A1-A5) hold, then

1. $\hat{\Lambda}_{\beta, F}$ is continuously differentiable function of (β, F) and $\hat{\Lambda}_{\beta_0, F_0} = \Lambda_0$
2. $\hat{\beta}_n \xrightarrow{P} \beta_0$ and $\hat{b}_n \xrightarrow{P} b_0$ as $n \rightarrow \infty$

Proof: Proof of (1): In Theorem-6.1 it is shown that $\hat{\Lambda}_{\beta_0, F_0} = \Lambda_0$.

Since $(f, F) \rightarrow E_F(f) = \int f dF$ is continuously differentiable, so clearly the function

$$(\beta, F) \rightarrow \hat{\Lambda}_{\beta, F} = \int_0^\tau \frac{E_F dN(u)}{E_F \gamma(V) Y(u) e^{\beta' Z}}. \quad (6.9)$$

is continuously differentiable.

Proof of (2): The profile log-likelihood function for one observation can be expressed as $\log P(V, T, \delta|b, \beta, \hat{\Lambda}_{\beta, F})$ which is continuously differentiable. Now, we set

$$M_n(b, \beta) = \frac{1}{n} \sum_{i=1}^n \log P(V_i, T_i, \delta_i|b, \beta, \hat{\Lambda}_{\beta, F_n})$$

and

$$M(b, \beta) = E \left[\log P(V, T, \delta | b, \beta, \hat{\Lambda}_{\beta, F_0}) \right]$$

and apply the proof of Theorem-5.7 (Van der Vaart, 2000) to the Cox PH cure model. Based on the results of Theorem-5.7 from Van der Vaart (2000), it can be shown that

$$\hat{\beta}_n \xrightarrow{P} \beta_0 \text{ and } \hat{b}_n \xrightarrow{P} b_0 \text{ as } n \rightarrow \infty.$$

6.5 Lemma- 6.3 with proof

Lemma 6.3: Suppose for assumptions **(A1-A5)**, $\hat{\beta}_n \xrightarrow{P} \beta_0$ and $F_n \xrightarrow{P} F_0$ as $n \rightarrow \infty$, then we have

$$\begin{aligned} & E \left[\sqrt{n} \left\{ \phi_s(T, \delta | \hat{\beta}_n, F_0) - \phi_s(T, \delta | \beta_0, F_0) \right\} \right] \\ &= -E \left[\phi_s(T, \delta | \beta_0, F_0) \phi'_s(T, \delta | \beta_0, F_0) \right] \left\{ \sqrt{n}(\hat{\beta}_n - \beta_0) \right\} + o_p(1) \quad (6.10) \end{aligned}$$

and

$$\begin{aligned} & E \left[\sqrt{n} \left\{ \phi_s(T, \delta | \hat{\beta}_n, F_n) - \phi_s(T, \delta | \hat{\beta}_n, F_0) \right\} \right] \\ &= -E \left[\phi_s(T, \delta | \beta_0, F_0) B(T, \delta | \beta_0, F_0) \right] \left\{ \sqrt{n}(F_n - F_0) \right\} \\ &+ o_p(1 + \sqrt{n}(\hat{\beta}_n - \beta_0)). \quad (6.11) \end{aligned}$$

(Proof is similar to Theorem 1 of Hirose and Liu (2020)).

Remark: The results are obtained without assuming the existence of derivative of the score functions $\frac{\partial}{\partial \beta} \phi_s(T, \delta | \beta, F)$ and $d_F B(T, \delta | \beta, F)$. This result give us asymptotic expansion of profile likelihood without differentiating the score function that involve implicit function.

Proof: Based on Lemma-6.1, we know $P(T, \delta | \beta_0, F_0) > \delta > 0$ for some positive constant $\delta > 0$. So by the differentiability of $P(T, \delta | \beta, F)$ with re-

spect to β and F , we have

$$\frac{\sqrt{n}\{P(T, \delta|\hat{\beta}_n, F_0) - P(T, \delta|\beta_0, F_0)\}}{P(T, \delta|\beta_0, F_0)} = \phi_s(T, \delta|\beta_0, F_0)\{\sqrt{n}(\hat{\beta}_n - \beta_0)\} + o_p(1), \quad (6.12)$$

$$\frac{\sqrt{n}\{P(T, \delta|\hat{\beta}_n, F_n) - P(T, \delta|\hat{\beta}_n, F_0)\}}{P(T, \delta|\beta_0, F_0)} = B(T, \delta|\beta_0, F_0)\{\sqrt{n}(F_n - F_0)\} + o_p(1). \quad (6.13)$$

In Lemma-1, we showed the class of functions $\{\phi_s(T, \delta|\beta, F) : \beta \in \Theta, F \in \zeta_\rho\}$ is uniformly bounded. Similarly, we can show the class of functions $\{B(T, \delta|\beta, F) : \beta \in \Theta, F \in \zeta_\rho\}$ is uniformly bounded. From these results, it follows that there is a P_0 -square integrable function, such that

$$\frac{P(T, \delta|\beta', F') - P(T, \delta|\beta, F)}{P(T, \delta|\beta, F)} \leq M'(\|\beta' - \beta\| + \|F' - F\|), \quad (6.14)$$

where M' is a P_0 -square integrable function $\forall \beta, \forall \beta' \in \Theta$ and $\forall F, \forall F' \in \zeta_\rho$.

First we start with (6.10), for each n , the equality

$$\begin{aligned} 0 &= \sqrt{n} \left\{ \int \phi_s(T, \delta|\hat{\beta}_n, F_0) P(T, \delta|\hat{\beta}_n, F_0) dF - \int \phi_s(T, \delta|\beta_0, F_0) P(T, \delta|\beta_0, F_0) dF \right\} \\ &= \sqrt{n} \left\{ \int \phi_s(T, \delta|\hat{\beta}_n, F_0) P(T, \delta|\beta_0, F_0) dF - \int \phi_s(T, \delta|\beta_0, F_0) P(T, \delta|\beta_0, F_0) dF \right. \\ &\quad \left. + \int \phi_s(T, \delta|\hat{\beta}_n, F_0) P(T, \delta|\hat{\beta}_n, F_0) dF - \int \phi_s(T, \delta|\hat{\beta}_n, F_0) P(T, \delta|\beta_0, F_0) dF \right\}, \end{aligned}$$

holds and we can express the above equation as

$$\begin{aligned} &\int \sqrt{n} \left\{ \phi_s(T, \delta|\hat{\beta}_n, F_0) - \phi_s(T, \delta|\beta_0, F_0) \right\} P(T, \delta|\beta_0, F_0) dF \\ &= - \int \phi_s(T, \delta|\hat{\beta}_n, F_0) \sqrt{n} \left\{ P(T, \delta|\hat{\beta}_n, F_0) - P(T, \delta|\beta_0, F_0) \right\} dF. \end{aligned} \quad (6.15)$$

By the dominated convergence theorem with (6.12), the right hand side of (6.15) can be expressed as, when $n \rightarrow \infty$

$$\begin{aligned} &- \int \phi_s(T, \delta|\hat{\beta}_n, F_0) \sqrt{n} \left\{ P(T, \delta|\hat{\beta}_n, F_0) - P(T, \delta|\beta_0, F_0) \right\} dF \\ &= - \int \phi_s(T, \delta|\hat{\beta}_n, F_0) \frac{\sqrt{n}\{P(T, \delta|\hat{\beta}_n, F_0) - P(T, \delta|\beta_0, F_0)\}}{P(T, \delta|\beta_0, F_0)} P(T, \delta|\beta_0, F_0) dF \\ &= - \int \phi_s(T, \delta|\beta_0, F_0) \phi'_s(T, \delta|\beta_0, F_0) P(T, \delta|\beta_0, F_0) \{\sqrt{n}(\hat{\beta}_n - \beta_0)\} + o_p(1) \\ &= -E \left[\phi_s(T, \delta|\beta_0, F_0) \phi'_s(T, \delta|\beta_0, F_0) \right] \left\{ \sqrt{n}(\hat{\beta}_n - \beta_0) \right\} + o_p(1). \end{aligned} \quad (6.16)$$

So from (6.15) and (6.16), we can write

$$\begin{aligned} & \int \sqrt{n} \left\{ \phi_s(T, \delta | \hat{\beta}_n, F_0) - \phi_s(T, \delta | \beta_0, F_0) \right\} P(T, \delta | \beta_0, F_0) dF \\ &= -E \left[\phi_s(T, \delta | \beta_0, F_0) \phi'_s(T, \delta | \beta_0, F_0) \right] \left\{ \sqrt{n}(\hat{\beta}_n - \beta_0) \right\} + o_p(1). \end{aligned}$$

So (6.10) is proven. Now we prove (6.11) by following the similar idea of proving (6.10). For each n , the following equation holds

$$\begin{aligned} 0 &= \sqrt{n} \left\{ \int \phi_s(T, \delta | \hat{\beta}_n, F_n) P(T, \delta | \hat{\beta}_n, F_n) dF - \int \phi_s(T, \delta | \hat{\beta}_n, F_0) P(T, \delta | \hat{\beta}_n, F_0) dF \right\} \\ &= \sqrt{n} \left\{ \int \phi_s(T, \delta | \hat{\beta}_n, F_n) P(T, \delta | \hat{\beta}_n, F_n) dF - \int \phi_s(T, \delta | \hat{\beta}_n, F_0) P(T, \delta | \hat{\beta}_n, F_n) dF \right. \\ &\quad \left. + \int \phi_s(T, \delta | \hat{\beta}_n, F_0) P(T, \delta | \hat{\beta}_n, F_n) dF - \int \phi_s(T, \delta | \hat{\beta}_n, F_0) P(T, \delta | \hat{\beta}_n, F_0) dF \right\}. \end{aligned}$$

We can express the above equation as

$$\begin{aligned} & \int \sqrt{n} \left\{ \phi_s(T, \delta | \hat{\beta}_n, F_n) - \phi_s(T, \delta | \hat{\beta}_n, F_0) \right\} P(T, \delta | \hat{\beta}_n, F_n) dF \\ &= - \int \phi_s(T, \delta | \hat{\beta}_n, F_0) \sqrt{n} \left\{ P(T, \delta | \hat{\beta}_n, F_n) - P(T, \delta | \hat{\beta}_n, F_0) \right\} dF. \end{aligned} \tag{6.17}$$

By using the dominated convergence theorem with (6.14) and Lemma-2,

when $n \rightarrow \infty$, the left hand side of (6.17) can be derived as

$$\begin{aligned}
& \left\| \int \sqrt{n} \left\{ \phi_s(T, \delta | \hat{\beta}_n, F_n) - \phi_s(T, \delta | \hat{\beta}_n, F_0) \right\} P(T, \delta | \hat{\beta}_n, F_n) dF \right. \\
& \quad \left. - \int \sqrt{n} \left\{ \phi_s(T, \delta | \hat{\beta}_n, F_n) - \phi_s(T, \delta | \hat{\beta}_n, F_0) \right\} P(T, \delta | \beta_0, F_0) dF \right\| \\
&= \left\| \int \left\{ \phi_s(T, \delta | \hat{\beta}_n, F_n) - \phi_s(T, \delta | \hat{\beta}_n, F_0) \right\} \times \right. \\
& \quad \left. \sqrt{n} \left\{ P(T, \delta | \hat{\beta}_n, F_n) - P(T, \delta | \beta_0, F_0) \right\} dF \right\| \\
&= \left\| \int \left\{ \phi_s(T, \delta | \hat{\beta}_n, F_n) - \phi_s(T, \delta | \hat{\beta}_n, F_0) \right\} \times \right. \\
& \quad \left. \frac{\sqrt{n} \{ P(T, \delta | \hat{\beta}_n, F_n) - P(T, \delta | \beta_0, F_0) \}}{P(T, \delta | \beta_0, F_0)} P(T, \delta | \beta_0, F_0) dF \right\| \\
&\leq \left| \int \left\{ M'' \left(\|F_n - F_0\| + \|\hat{\Lambda}_{\hat{\beta}_n, F_n} - \hat{\Lambda}_{\hat{\beta}_n, F_0}\| \right) \right\} \times \right. \\
& \quad \left. \left\{ \sqrt{n} M' \left(\|\hat{\beta}_n - \beta_0\| + \|F_n - F_0\| \right) \right\} P(T, \delta | \beta_0, F_0) dF \right| \\
&= \left\{ \left| \int M'' M' P(T, \delta | \beta_0, F_0) dF \right| \right\} \times \\
& \quad \left\{ \sqrt{n} \left(\|\hat{\beta}_n - \beta_0\| + \|F_n - F_0\| \right) \left(\|F_n - F_0\| + \|\hat{\Lambda}_{\hat{\beta}_n, F_n} - \hat{\Lambda}_{\hat{\beta}_n, F_0}\| \right) \right\} \\
&= O \left\{ \sqrt{n} \left(\|\hat{\beta}_n - \beta_0\| + \|F_n - F_0\| \right) \left(\|F_n - F_0\| + \|\hat{\Lambda}_{\hat{\beta}_n, F_n} - \hat{\Lambda}_{\hat{\beta}_n, F_0}\| \right) \right\} \\
&= \sqrt{n} (\hat{\beta}_n - \beta_0) O(o_p(1)) + O(O_p(1) \cdot o_p(1)) \\
&= \sqrt{n} (\hat{\beta}_n - \beta_0) \cdot o_p(1) + o_p(1) \\
&= o_p(1 + \sqrt{n} (\hat{\beta}_n - \beta_0)), \tag{6.18}
\end{aligned}$$

where we used $F_n - F_0 = o_p(1)$ from assumption **(A3)**, $\hat{\beta}_n - \beta_0 = o_p(1)$ and $\hat{\Lambda}_{\hat{\beta}_n, F_n} - \hat{\Lambda}_{\beta_0, F_0} = o_p(1)$ from Theorem-2.

By the dominated convergence theorem with (6.13), the right hand side

of (6.17) can be written as, when $n \rightarrow \infty$

$$\begin{aligned}
& - \int \phi_s(T, \delta|\hat{\beta}_n, F_0) \sqrt{n} \left\{ P(T, \delta|\hat{\beta}_n, F_n) - P(T, \delta|\hat{\beta}_n, F_0) \right\} dF \\
& = - \int \phi_s(T, \delta|\hat{\beta}_n, F_0) \frac{\sqrt{n} \{ P(T, \delta|\hat{\beta}_n, F_n) - P(T, \delta|\hat{\beta}_n, F_0) \}}{P(T, \delta|\beta_0, F_0)} P(T, \delta|\beta_0, F_0) dF \\
& = - \int \phi_s(T, \delta|\beta_0, F_0) B(T, \delta|\beta_0, F_0) P(T, \delta|\beta_0, F_0) \sqrt{n} (F_n - F_0) dF + o_p(1) \\
& = - E \left[\phi_s(T, \delta|\beta_0, F_0) B(T, \delta|\beta_0, F_0) \right] \sqrt{n} (F_n - F_0) + o_p(1). \tag{6.19}
\end{aligned}$$

So by combining (6.18) and (6.19), the equality (6.17) is equivalent to

$$\begin{aligned}
& \int \sqrt{n} \left\{ \phi_s(T, \delta|\hat{\beta}_n, F_n) - \phi_s(T, \delta|\hat{\beta}_n, F_0) \right\} P(T, \delta|\beta_0, F_0) dF \\
& = - E \left[\phi_s(T, \delta|\beta_0, F_0) B(T, \delta|\beta_0, F_0) \right] \sqrt{n} (F_n - F_0) + o_p(1 + \sqrt{n}(\hat{\beta}_n - \beta_0)).
\end{aligned}$$

So equation (6.11) is also proven. Hence, we proved Lemma-6.3.

6.6 Theorem- 6.3 with proof

Theorem-6.3: The estimator $\hat{\beta}_n$ from EM algorithm is an asymptotically linear estimator for β_0 with the efficient influence function $(I_s^*)^{-1} \phi_s(T, \delta|\beta_0, F_0)$, so that

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (I_s^*)^{-1} \phi_s(T_i, \delta_i|\beta_0, F_0) + o_p(1) \xrightarrow{D} N\{0, (I_s^*)^{-1}\},$$

where $I_s^* = E[\phi_s \phi_s']$ and $N\{0, (I_s^*)^{-1}\}$ is a normal distribution with mean zero and variance $(I_s^*)^{-1}$. So the estimator $\hat{\beta}_n$ is efficient.

In addition \hat{b}_n from EM algorithm is an asymptotically linear estimator for b_0 with the influence function $(I_l^*)^{-1} \phi_l(V|b_0)$, so that

$$\sqrt{n}(\hat{b}_n - b_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (I_l^*)^{-1} \phi_l(V_i|b_0) + o_p(1) \xrightarrow{D} N\{0, (I_l^*)^{-1}\},$$

where $I_l^* = E[\phi_l \phi_l']$, so the estimator \hat{b}_n is efficient.

(Proof is similar to Theorem 2 of Hirose and Liu (2020)).

Proof: The profile likelihood maximum likelihood estimators $\hat{\beta}_n$ and \hat{b}_n from EM algorithm described in Chapter-5 (Section 5.3) are solutions to the following system of estimating equations

$$\sum_{i=1}^n \phi_s(T_i, \delta_i | \hat{\beta}_n, F_n) = 0,$$

and

$$\sum_{i=1}^n \phi_l(V_i | \hat{b}_n) = 0.$$

Here, $\phi_s(T_i, \delta_i | \beta, F_n)$ and $\phi_l(V_i | b)$ are the profile likelihood score functions of β and b defined in equations (5.21) and (5.20). Since $\phi_l(V_i | b)$ is the score function corresponding to logistic part (which is a parametric model that doesn't involve Λ), so we are going to focus on the survival part only.

Since $\phi_s(T, \delta | \beta, F)$ is uniformly bounded Donsker (Lemma-6.1), so by Lemma 19.24 (Van der Vaart, 2000), we can write

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \phi_s(T_i, \delta_i | \hat{\beta}_n, F_n) - \phi_s(T_i, \delta_i | \beta_0, F_0) \right\} \\ &= \sqrt{n} E \left[\phi_s(T, \delta | \hat{\beta}_n, F_n) - \phi_s(T, \delta | \beta_0, F_0) \right] + o_p(1), \end{aligned} \quad (6.20)$$

From (6.10), it follows that

$$\begin{aligned} & \sqrt{n} E \left[\phi_s(T, \delta | \hat{\beta}_n, F_0) - \phi_s(T, \delta | \beta_0, F_0) \right] \\ &= -E \left[\phi_s(T, \delta | \beta_0, F_0) \phi'_s(T, \delta | \beta_0, F_0) \right] \sqrt{n} (\hat{\beta}_n - \beta_0) + o_p(1) \\ &= -I_s^* \sqrt{n} (\hat{\beta}_n - \beta_0) + o_p(1). \end{aligned} \quad (6.21)$$

From (6.11), it follows that

$$\begin{aligned} & \sqrt{n} E \left[\phi_s(T, \delta | \hat{\beta}_n, F_n) - \phi_s(T, \delta | \hat{\beta}_n, F_0) \right] \\ &= -E \left[\phi_s(T, \delta | \beta_0, F_0) B(T, \delta | \beta_0, F_0) \right] \sqrt{n} (F_n - F_0) \\ &+ o_p(1 + \sqrt{n} (\hat{\beta}_n - \beta_0)). \end{aligned} \quad (6.22)$$

Since $B(T, \delta|\beta_0, F_0)$ is in the nuisance tangent space and $\phi_s(T, \delta|\beta_0, F_0)$ is the efficient score function, so we have

$$E[\phi_s(T, \delta|\beta_0, F_0)B(T, \delta|\beta_0, F_0)] = 0. \quad (6.23)$$

Now using (6.21), (6.22) and (6.23), the right hand side of (6.20) can be expressed as

$$\begin{aligned} & \sqrt{n}E\left[\phi_s(T, \delta|\hat{\beta}_n, F_n) - \phi_s(T, \delta|\beta_0, F_0)\right] \\ &= \sqrt{n}E\left[\phi_s(T, \delta|\hat{\beta}_n, F_0) - \phi_s(T, \delta|\beta_0, F_0)\right] + \\ & \quad \sqrt{n}E\left[\phi_s(T, \delta|\hat{\beta}_n, F_n) - \phi_s(T, \delta|\hat{\beta}_n, F_0)\right] \\ &= -I_s^* \sqrt{n}(\hat{\beta}_n - \beta_0) + o_p(1 + \sqrt{n}(\hat{\beta}_n - \beta_0)). \end{aligned} \quad (6.24)$$

We know that $\frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_s(T_i, \delta_i|\hat{\beta}_n, F_n) = 0$, so using (6.24), the equation (6.20) can be written as

$$(I_s^* + o_p(1))\sqrt{n}(\hat{\beta}_n - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_s(T_i, \delta_i|\beta_0, F_0) + o_p(1). \quad (6.25)$$

By Central Limit Theorem, we can write $\frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_s(T_i, \delta_i|\beta_0, F_0) + o_p(1) = O_p(1)$. Since I_s^* is invertible, we have $(I_s^* + o_p(1))^{-1} = O_p(1)$.

So from (6.25) we can write $\sqrt{n}(\hat{\beta}_n - \beta_0) = (I_s^* + o_p(1))^{-1} O_p(1) = O_p(1)$.

Finally we can express (6.20) as

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n I_s^{*-1} \phi_s(T_i, \delta_i|\beta_0, F_0) + o_p(1).$$

It follows that the large sample distribution of the estimator $\hat{\beta}_n$ can be expressed as

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{D} N\{0, (I_s^*)^{-1}\},$$

where $I_s^* = E[\phi_s \phi_s']$ is the efficient information (ϕ_s is defined as efficient score function in Theorem- 6.1).

6.7 Efficient Score Function for Cox PH Cure Model using Projection Theory

To get the efficient score function using the projection theory, we assume the parameters (β, Λ) are evaluated at the true values β_0, Λ_0 and omit subscript “0” for brevity.

The log-likelihood function of the survival part for one observation can be written as

$$\log P(T, \delta | \beta, \Lambda) = \gamma(V) \left\{ \delta (\log \lambda(t) + \beta' Z) - e^{\beta' Z} \Lambda(t) \right\}.$$

Score Function for β

$$\phi_\beta(T, \delta | \beta, \Lambda) = \frac{\partial}{\partial \beta} \log P(T, \delta | \beta, \Lambda) = \gamma(V) \left\{ Z (\delta - e^{\beta' Z} \Lambda(t)) \right\}.$$

Score Operator for Λ

Let us take a measurable function which is bounded such as $g : [0, \tau] \rightarrow R$, where g is defined in the interval $[0, \tau]$ because Λ is also restricted within this interval. The path can be defined as

$$d\Lambda_s = (1 + sg)d\Lambda.$$

The corresponding path for the baseline hazard function is

$$\lambda_s(t) = \frac{d\Lambda_s}{dt} = (1 + sg) \frac{d\Lambda}{dt} = (1 + sg) \lambda(t).$$

The derivative of the log-likelihood function with respect to s can be expressed as

$$B_\Lambda(T, \delta | \beta, \Lambda)g = \frac{\partial}{\partial s} \Big|_{s=0} \log P(T, \delta | \beta, \Lambda_s) = \gamma(V) \left\{ \delta g(t) - e^{\beta' Z} \int_0^T g(u) d\Lambda(u) \right\}.$$

Information Operator $B_{\Lambda}^* B_{\Lambda}$ and its Inverse $(B_{\Lambda}^* B_{\Lambda})^{-1}$

Let us start with the information operator $B_{\Lambda}^* B_{\Lambda}$ and take two arbitrary functions f and g . By definition of the adjoint, we can write

$$\langle B_{\Lambda}^* B_{\Lambda} f, g \rangle_{L_2(\Lambda)} = \langle B_{\Lambda} f, B_{\Lambda} g \rangle_{L_2(P)}. \quad (6.26)$$

The path defined by $d\Lambda_{r,s} = (1 + rf + sg + rsfg)d\Lambda$ is positive for small r and s . The corresponding path for the baseline hazard function is

$$\lambda_{r,s}(t) = \frac{d\Lambda_{r,s}}{dt} = (1 + rf + sg + rsfg) \frac{d\Lambda}{dt} = (1 + rf + sg + rsfg)\lambda(t).$$

Now we can write

$$\left. \frac{\partial}{\partial r} \right|_{(r,s)=(0,0)} \log P(T, \delta | \beta, \Lambda_{r,s}) = B_{\Lambda} f, \quad (6.27)$$

and

$$\left. \frac{\partial}{\partial s} \right|_{(r,s)=(0,0)} \log P(T, \delta | \beta, \Lambda_{r,s}) = B_{\Lambda} g. \quad (6.28)$$

Using (6.27) and (6.28) we can write

$$\begin{aligned} \langle B_{\Lambda} f, B_{\Lambda} g \rangle_{L_2(P)} &= E \left\{ (B_{\Lambda} f)(B_{\Lambda} g) \right\} \\ &= - E \left\{ \left. \frac{\partial^2}{\partial r \partial s} \right|_{(r,s)=(0,0)} \log P(T, \delta | \beta, \Lambda_{r,s}) \right\} \\ &= E \left\{ \gamma(V) e^{\beta' Z} \int_0^T f(\xi) g(\xi) d\Lambda(\xi) \right\}. \end{aligned} \quad (6.29)$$

Now we manipulate the integral involving the function ξ , we deduce

$$\int_0^T f(\xi) g(\xi) d\Lambda(\xi) = \int_0^T I(\xi \leq T) f(\xi) g(\xi) d\Lambda(\xi).$$

Indeed, if $\xi > T$, then the contribution will be 0 to the integral. So the last term in equation (6.29) can be expressed as

$$E \left\{ \gamma(V) e^{\beta' Z} \int_0^T f(\xi) g(\xi) d\Lambda(\xi) \right\} = E \left\{ \gamma(V) e^{\beta' Z} \int_0^T I(\xi \leq T) f(\xi) g(\xi) d\Lambda(\xi) \right\}. \quad (6.30)$$

Using Fubini's theorem, equation (6.30) can be written as

$$E \left\{ \gamma(V) e^{\beta' Z} \int_0^T I(\xi \leq T) f(\xi) g(\xi) d\Lambda(\xi) \right\} = \left\langle E \{ \gamma(V) e^{\beta' Z} I(\xi \leq T) f(\xi) \}, g(\xi) \right\rangle_{L_2(\Lambda)}. \quad (6.31)$$

From equation (6.26) we can write

$$\left\langle B_{\Lambda}^* B_{\Lambda} f, g \right\rangle_{L_2(\Lambda)} = \left\langle E \{ \gamma(V) e^{\beta' Z} I(t \leq T) f \}, g \right\rangle_{L_2(\Lambda)}.$$

So, the information operator is

$$B_{\Lambda}^* B_{\Lambda} f = E \left\{ \gamma(V) e^{\beta' Z} I(t \leq T) \right\} f(t).$$

It follows that the inverse of information operator is

$$(B_{\Lambda}^* B_{\Lambda})^{-1} f(t) = \left[E \{ \gamma(V) e^{\beta' Z} I(t \leq T) \} \right]^{-1} f(t).$$

The Action of the Adjoint Score Operator B_{Λ}^* on the Score Function ϕ_{β}

Assume the differentiable paths $(r, s) \mapsto P(T, \delta | \beta + ru, \Lambda_s)$ can be exploited with the path $d\Lambda_s = (1 + sg)d\Lambda$. Now we can write

$$\left. \frac{\partial}{\partial r} \right|_{(r,s)=(0,0)} \log P(T, \delta | \beta + ru, \Lambda_s) = u' \phi_{\beta}. \quad (6.32)$$

and

$$\left. \frac{\partial}{\partial s} \right|_{(r,s)=(0,0)} \log P(T, \delta | \beta + ru, \Lambda_s) = B_{\Lambda} g. \quad (6.33)$$

Using equation (6.32) and (6.33) we can write

$$\begin{aligned} \left\langle u' \phi_{\beta}, B_{\Lambda} g \right\rangle &= E \left\{ (u' \phi_{\beta})(B_{\Lambda} g) \right\} \\ &= - E \left\{ \left. \frac{\partial^2}{\partial r \partial s} \right|_{(r,s)=(0,0)} \log P(T, \delta | \beta + ru, \Lambda_s) \right\} \\ &= u' E \left\{ \gamma(V) Z e^{\beta' Z} \int_0^T g(\xi) d\Lambda(\xi) \right\}. \end{aligned} \quad (6.34)$$

Now by manipulating the integral involving the function ξ , the equation (6.34) can be expressed as

$$\left\langle u' \phi_\beta, B_\Lambda g \right\rangle = u' E \left\{ \gamma(V) Z e^{\beta' Z} \int_0^\tau I(\xi \leq T) g(\xi) d\Lambda(\xi) \right\}. \quad (6.35)$$

Using the Fubini's theorem, we can conclude that

$$u' E \left\{ \gamma(V) Z e^{\beta' Z} \int_0^\tau I(\xi \leq T) g(\xi) d\Lambda(\xi) \right\} = \left\langle u' E \left\{ \gamma(V) Z e^{\beta' Z} I(\xi \leq T) \right\}, g(\xi) \right\rangle_{L_2(\Lambda)}. \quad (6.36)$$

We know that

$$\left\langle u' B_\Lambda^* \phi_\beta, g \right\rangle_{L_2(P)} = \left\langle u' \phi_\beta, B_\Lambda g \right\rangle_{L_2(\Lambda)}.$$

So we can write

$$B_\Lambda^* \phi_\beta = E \left\{ \gamma(V) Z e^{\beta' Z} I(t \leq T) \right\}.$$

Efficient Score Function ϕ_{eff} :

Finally the efficient score function can be expressed as

$$\begin{aligned} \phi_{eff} &= \phi_\beta - B_\Lambda (B_\Lambda^* B_\Lambda)^{-1} B_\Lambda^* \phi_\beta \\ &= \gamma(V) \left\{ \delta Z - Z e^{\beta' Z} \Lambda(T) - \left[\delta - e^{\beta' Z} \int_0^T d\Lambda(u) \right] \frac{E[\gamma(V) Z e^{\beta' Z} I(t \leq T)]}{E[\gamma(V) e^{\beta' Z} I(t \leq T)]} \right\} \\ &= \gamma(V) \left\{ \delta \left[Z - \frac{M_1(T)}{M_0(T)} \right] - e^{\beta' Z} \int_0^T \left[Z - \frac{M_1(u)}{M_0(u)} \right] d\Lambda(u) \right\}, \end{aligned} \quad (6.37)$$

where $M_0(T)$ and $M_1(T)$ were defined in the proof of Theorem 6.1.

Here we have shown that the profile likelihood score function from equation (6.2) and the efficient score function based on the projection theory from equation (6.37) are identical.

Chapter 7

Efficient Estimation For The Joint Model of Longitudinal and Survival Data

7.1 Introduction

Many medical or clinical studies often deal with two types of outcome which are longitudinal response measurements and the survival outcome such as death, development of disease etc. (Rizopoulos, 2010, 2012). Therefore, the joint model of longitudinal and survival data is considered one of the powerful and popular techniques due to incorporating the association between longitudinal and time-to-event outcomes (Diaz and Sampredo, 2014; Guler et al., 2014; Ibrahim et al., 2010; Sattar et al., 2012).

The longitudinal outcome and the event outcome often modelled separately by using the mixed effect model and the Cox PH model respectively. Several approaches have been introduced to measure the association between longitudinal and survival data. One of the widely used approach is the Extended Cox model where time-dependent longitudinal covariate has

been used to investigate the association with the event outcome (Andersen and Gill, 1982; Andersen et al., 2012; Fleming and Harrington, 1991). However, this model assumes that the time dependent longitudinal outcomes do not have any measurement error (Balakrishnan and Rao, 2004; Dupuy and Mesbah, 2002; Li and Ma, 2013; Rizopoulos, 2012; Sattar et al., 2012). Moreover, the model also assumes that the covariates change at follow-up time and are unchanged between the follow-up times (Kalbfleisch and Prentice, 2011; Rizopoulos, 2012; Wang, 2004). However, in reality, we know that its usual to consider measurement errors while measuring the covariates, therefore, ignoring measurement errors in the longitudinal measurements is one of the main limitations of this model (Bang et al., 2013). To overcome these difficulties, the joint model of longitudinal and time-to-event outcomes was proposed in biomedical research (Sweeting and Thompson, 2011).

To see the effect of the longitudinal outcome as a time-dependent covariate on event outcome, Faucett and Thomas (1996) and Wulfsohn and Tsiatis (1997) have proposed the joint models of longitudinal and survival outcomes where the longitudinal outcome and the event outcome were modelled jointly rather than separately. This joint modelling approach reduced biases while doing parameter estimation (Ibrahim et al., 2010; McCrink et al., 2011). Tsiatis and Davidian (2001) and Yu et al. (2004) reviewed the early works of joint models done by Henderson et al. (2000); Wang and Taylor (2001); Wulfsohn and Tsiatis (1997). In clinical settings, the joint model of longitudinal and survival data has been widely used for various types of medical studies such as HIV and cancer (Brown and Ibrahim, 2003; Hanson and Yang, 2007; Hsieh et al., 2006; Ratcliffe et al., 2004; Rizopoulos, 2010; Viviani et al., 2014).

The profile likelihood approach in the joint model of longitudinal and

survival data with random effects has been proposed by Wulfsohn and Tsiatis (1997) where the parameters were estimated using the EM algorithm. Later Zeng and Lin (2007a,b, 2010) have followed the work of Wulfsohn and Tsiatis (1997) for parameter estimation. In the profile likelihood approach, we profile out the baseline hazard function and plugged the estimate of baseline hazard in the likelihood function of the joint model of longitudinal and survival data. Here the estimate of the baseline hazard function is an implicit function [see equations (7.10) and (7.12) and Remark-7.1 in Section-7.3.4] which is similar problem to the Cox PH cure model. However, the works of Wulfsohn and Tsiatis (1997); Zeng and Lin (2007a,b, 2010) did not address the computation with implicit function in the profile likelihood estimation. Hsieh et al. (2006) first pointed out the problem of implicit function in the joint model of longitudinal and survival data and suggested bootstrap method as a possible solution to compute the estimated standard errors while using the profile likelihood approach.

The efficiency and asymptotic distribution of semiparametric maximum likelihood estimators using a NPMLE approach have been studied by Zeng et al. (2005), however, they didn't use the asymptotic expansion of profile likelihood function to get the asymptotic normality of the profile likelihood MLE. Apart from the NPMLE approach, bayesian estimation techniques have also been used to estimate the parameters from the joint model of longitudinal and survival data (Brown and Ibrahim, 2003; Chi and Ibrahim, 2006; Guo and Carlin, 2004; Hanson and Yang, 2007; Henderson et al., 2000; Huang et al., 2011; Ibrahim et al., 2004; Tsiatis and Davidian, 2001; Wang and Taylor, 2001; Xu and Zeger, 2001).

The works of Zeng et al. (2005); Zeng and Lin (2007a,b, 2010) based on profile approach have avoided dealing with the implicit function and used the equation (1.2) from Chapter-1 for the calculation of information matrix

which didn't require the direct expansion of the profile likelihood function.

Rizopoulos (2010) developed a R package (JM) to fit the joint model of longitudinal and survival data and mentioned that in 'Cox-PH-GH' method, the baseline hazard estimator is an implicit function and the standard errors of the estimated parameters can be underestimated while calculating from the score function (Rizopoulos, 2012). The possible reason for underestimation can be the score function which has been calculated without differentiating the implicit function within the log-likelihood function. Later, Hickey et al. (2018) developed *joinerML* R-package where multivariate longitudinal outcomes have been used for joint modelling and considered bootstrap process to estimate the standard errors which are computationally expensive.

In this chapter, we have shown an alternative approach to the methodologies of Zeng et al. (2005); Zeng and Lin (2007a, 2010) where the asymptotic normality of the profile likelihood estimator has been studied for the joint model of longitudinal and survival data. Similar to Chapter-5, here our contribution is that we have used the direct asymptotic expansion of profile likelihood for the joint model of longitudinal and survival data and obtain the explicit form of the variance estimator using the profile likelihood score function where the estimated information matrix can be expressed as

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n \phi(x_i; \hat{\eta}) \phi'(x_i; \hat{\eta}).$$

Here, ϕ is the profile score function and $\hat{\eta}$ is the MLE of η . Therefore, we have solved the problem of estimating standard errors using the observed information matrix based on the profile likelihood score function.

Our proposed alternative method provides us the the way to expand the profile likelihood function directly. Moreover, we have calculated the

estimated standard errors of the profile likelihood estimators which is illustrated in the simulation study (Section-7.4) and in real-data example (Section-7.5). For the real-data examples, we have used the datasets ('AIDS' and 'PBC') from JM package and computed the standard errors of the estimated parameters from the profile likelihood score function. Moreover, we have compared our results with the output obtained from JM (method: 'Cox-PH-GH') and joineRML packages.

This chapter is organized as follows. A brief discussion on joint model of longitudinal and survival data has been given in Section-7.2. In Section-7.3, we describe the estimation procedure to show that the profile likelihood estimators are consistent and asymptotically normal. Results obtained from the simulation study and real-life data are shown in Section-7.4 and Section-7.5 respectively. This chapter concludes in Section-7.6 with a short discussion.

7.2 Joint Model of Longitudinal and Survival Data

Let y_{ij} be the response of subject $i, i = 1, 2, \dots, n$ observed at time $t_{ij}, j = 1, 2, \dots, n_i$, with fixed effects covariate vector $x_i(t_{ij})$ and random effects covariate vector $z_i(t_{ij})$. Now we assume a linear mixed model for the process as

$$y_{ij} = m_i(t_{ij}) + \varepsilon_i(t_{ij})$$

where $m_i(t_{ij}) = x_i'(t_{ij})\beta + z_i'(t_{ij})b_i$ is the longitudinal process with the fixed effects β and the random effects b_i . In addition, we assume

$$\begin{cases} b_i \sim N(0, D) \\ \varepsilon_i(t) \sim N(0, \sigma^2) \end{cases}$$

where random effects follows bivariate normal distribution with mean 0 and covariance matrix D . Moreover, the error $\varepsilon_i(t)$ follows normal distribu-

tion with mean 0 and variance σ^2 . The random effects are assumed independent of $\varepsilon_i(t)$.

For the survival data, let us assume T_i^* and C_i be the survival and censoring times, respectively. Moreover, T_i be the observed event time which can be defined as $T_i = \min(T_i^*, C_i)$ and $\delta_i = I(T_i^* \leq C_i)$ be the censoring indicator. We observe (T_i, δ_i, w_i) , where w_i be a vector of baseline covariates.

To measure the strength of association between $m_i(t)$ and the risk of the event, we can express the relative risk models as

$$\lambda(t|b_i) = \lambda(t) \exp \{ \gamma' w_i + \alpha' m_i(t) \}, \quad (7.1)$$

where $\lambda(t)$ is the baseline hazard function and γ is the vector of regression coefficients corresponding to w_i . Similarly α represents the effect of the underlying longitudinal outcome to the risk of the event. Let $\Lambda(t) = \int_0^t \lambda(s) ds$ be the cumulative baseline hazard function. Then the survival function can be written as

$$S_i(t|b_i) = \exp \left(- \int_0^t \lambda(s) \exp \{ \gamma' w_i + \alpha' m_i(s) \} ds \right).$$

7.3 Estimation: Profile Likelihood with EM Algorithm

In this section, first we present the likelihood function for the joint model of longitudinal and survival data. In Section 7.3.1, we profile out the baseline hazard using NPMLE approach and in Section 7.3.2 construct the profile likelihood function. Later in Section 7.3.3, we estimate the profile likelihood MLE through EM algorithm and finally in Section 7.3.4, we prove the asymptotic normality of the profile likelihood MLE.

We assume that the time-independent random effects b_i underlies in both longitudinal and event outcomes. Therefore, b_i considers not only the association between survival and longitudinal outcomes, but also the correlation between the repeated measurements of the longitudinal outcome (Rizopoulos, 2012). The joint likelihood for the observed data can be written as

$$\begin{aligned} L(\theta, \sigma^2, D, \Lambda) &= \prod_{i=1}^n \int_{b_i} p(T_i, \delta_i, y_i, b_i; \theta, \sigma^2, D, \Lambda) db_i \\ &= \prod_{i=1}^n \int_{b_i} p(T_i, \delta_i | b_i; \theta, \Lambda) p(y_i | b_i; \beta, \sigma^2) p(b_i; D) db_i \end{aligned}$$

where

$$\begin{aligned} p(T_i, \delta_i | b_i; \theta, \Lambda) &= [\lambda(T_i) \exp \{ \gamma' w_i + \alpha' m_i(T_i) \}]^{\delta_i} \\ &\quad \times \exp \left(- \int_0^{T_i} \exp \{ \gamma' w_i + \alpha' m_i(s) \} d\Lambda(s) \right) \end{aligned} \quad (7.2)$$

$$p(y_i | b_i; \beta, \sigma^2) = (2\pi\sigma^2)^{-n_i/2} \exp \left\{ - \frac{\sum_{j=1}^{n_i} [y_i(t_{ij}) - m_i(t_{ij})]^2}{2\sigma^2} \right\}, \quad (7.3)$$

and

$$p(b_i; D) = (2\pi)^{-r/2} \det(D)^{-1/2} \exp \left\{ - \frac{b_i' D^{-1} b_i}{2} \right\}. \quad (7.4)$$

Here r is the dimension of b_i and $(\theta', \sigma^2, D)'$ be the all parameters in the joint model of longitudinal and event data where $\theta' = (\beta', \gamma', \alpha')$. In our joint model, $(\theta', \sigma^2, D)'$ is the main parameter of interest whereas the cumulative baseline hazard function Λ is the nuisance parameter.

Let us define the complete data by $(t_i, \delta_i, y_i, b_i)$, $i = 1, \dots, n$ which includes the observed data and unobserved b_i (treating b_i as missing data). The choice for using EM algorithm is justified by the fact that the model depends on a latent variable, b_i . Moreover, the aim of EM algorithm is to maximize observed data likelihood from a complete data likelihood (Dempster

et al., 1977). So the complete data log-likelihood function can be written as

$$\log L_c(\theta, \sigma^2, D, \Lambda; b) = \sum_{i=1}^n \log \left\{ p(T_i, \delta_i | b_i; \theta, \Lambda) p(y_i | b_i; \beta, \sigma^2) p(b_i; D) \right\}.$$

Now the expected complete data log-likelihood under $p(b|T, \delta, y)$ will be

$$\begin{aligned} E_b[\log L_c(\theta, \sigma^2, D, \Lambda; b)] &= \sum_{i=1}^n \int_{b_i} \log \left\{ p(T_i, \delta_i | b_i; \theta, \Lambda) p(y_i | b_i; \beta, \sigma^2) p(b_i; D) \right\} p(b_i | T_i, \delta_i, y_i; \theta, \sigma^2, D, \Lambda) db_i \\ &= \sum_{i=1}^n \left\{ E_b[\log p(T_i, \delta_i | b_i; \theta, \Lambda)] + E_b[\log p(y_i | b_i; \beta, \sigma^2)] + E_b[\log p(b_i; D)] \right\}, \end{aligned} \quad (7.5)$$

where E_b is the expectation with respect to $b \sim p(b|T, \delta, y; \theta, \sigma^2, D, \Lambda)$ which can be expressed as

$$\begin{aligned} p(b|T, \delta, y; \theta, \sigma^2, D, \Lambda) &= \frac{p(T, \delta | b; \theta, \Lambda) p(y | b; \beta, \sigma^2) p(b; D)}{\int_b p(T, \delta | b; \theta, \Lambda) p(y | b; \beta, \sigma^2) p(b; D) db} \\ &\propto p(T, \delta | b; \theta, \Lambda) p(y | b; \beta, \sigma^2) p(b; D). \end{aligned} \quad (7.6)$$

7.3.1 Baseline Hazard Estimation

Before starting the EM algorithm, we profile out the baseline hazard function $\lambda(t)$ using NPMLE. We assume that the baseline hazard, $\lambda(t)$ has (possible) non-zero values on the observed time t_i :

$$\lambda(t_i) = \lambda_i \geq 0; \quad i = 1, 2, \dots, n$$

and $\lambda(t) = 0$ for the rest of time (Murphy and Van der Vaart, 2000). Under the assumption, the integral, $\Lambda(t) = \int_0^t \lambda(s) ds$ is replaced with the sum

$$\Lambda(t) = \sum_{i=1}^n \lambda_i 1\{t_i \leq t\}.$$

Now the survival part of equation (7.5) can be separately maximized with respect to λ using the log-likelihood:

$$\sum_{i=1}^n E_b \left[\delta_i \left\{ \log \lambda_i + \gamma' w_i + \alpha' m_i(t) \right\} - \exp \{ \gamma' w_i + \alpha' m_i(t) \} \sum_{j=1}^n \lambda_j 1\{t_j \leq t_i\} \right], \quad (7.7)$$

and the score equation with respect to λ_k will be

$$\frac{\partial}{\partial \lambda_k} \sum_{i=1}^n E_b \left[\delta_i \{ \log \lambda_i + \gamma' w_i + \alpha' m_i(t) \} - \exp \{ \gamma' w_i + \alpha' m_i(t) \} \sum_{j=1}^n \lambda_j 1 \{ t_j \leq t_i \} \right] = 0.$$

By solving the above equation, the estimate of λ_k is

$$\hat{\lambda}_k(t) = \frac{\delta_k}{\sum_{l=1}^n 1 \{ t_k \leq t_l \} E_b [\exp \{ \gamma' w_l + \alpha' m_l(t_k) \}]},$$

and the estimate of the cumulative baseline hazard can be written as

$$\hat{\Lambda}(t; \theta) = \sum_{i=1}^n \frac{\delta_i 1 \{ t_i \leq t \}}{\sum_{l=1}^n 1 \{ t \leq t_l \} E_b [\exp \{ \gamma' w_l + \alpha' m_l(t) \}]}. \quad (7.8)$$

7.3.2 Profile Likelihood Function

The NPMLE of the baseline cumulative hazard function in the counting process notation can be written from equation (7.8) as

$$\hat{\Lambda}(t; \theta) = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u) E_b [\exp \{ \gamma' w_i + \alpha' m_i(u) \}]}, \quad (7.9)$$

where $N_i(u) = 1 \{ T_i \leq u, \delta = 1 \}$ and $Y_i(u) = 1 \{ T_i \geq u \}$.

Let us denote $E_{F_n} f = \int f dF_n$. Then $\hat{\Lambda}(t)$ can be expressed as

$$\hat{\Lambda}_{\theta, F_n}(t) = \int_0^t \frac{E_{F_n} dN(u)}{E_{F_n} Y(u) E_b [\exp \{ \gamma' w + \alpha' m(u) \}]}. \quad (7.10)$$

Now we plug the baseline hazard estimator ($\hat{\Lambda}_{\theta, F_n}$ in the place of Λ) in the likelihood function of the joint model to obtain the profile likelihood function. The log-profile likelihood functions (for one observation) for Cox PH, longitudinal and random effect components can be denoted as

$\log p(T_i, \delta_i | b_i; \theta, \hat{\Lambda}_{\theta, F_n})$, $\log p(y_i | b_i; \beta, \sigma^2)$ and $\log p(b_i; D)$ respectively which are

$$\begin{aligned} \log p(T_i, \delta_i | b_i; \theta, \hat{\Lambda}_{\theta, F_n}) &= \delta_i \left\{ \log \frac{E_{F_n} dN(T_i)}{E_{F_n} Y(T_i) E_b [\exp \{ \gamma' w + \alpha' m(T_i) \}]} + \gamma' w_i + \alpha' m_i(T_i) \right\} \\ &\quad - \exp \left(\gamma' w_i + \alpha' m_i(T_i) \right) \int_0^{T_i} \frac{E_{F_n} dN(u)}{E_{F_n} Y(u) E_b [\exp \{ \gamma' w + \alpha' m(u) \}]}, \\ \log p(y_i | b_i; \beta, \sigma^2) &= -\frac{1}{2} \left\{ n_i \log(2\pi) + \log(\sigma^2) + (\sigma^2)^{-1} (y_i - x_i \beta - z_i b_i)' (y_i - x_i \beta - z_i b_i) \right\}, \end{aligned}$$

and

$$\log p(b_i; D) = -\frac{1}{2} \left\{ r \log(2\pi) + \log |D| + b_i' D^{-1} b_i \right\}.$$

For simplicity, we can express $\log p(T_i, \delta_i | b_i; \theta, \hat{\Lambda}_{\theta, F_n})$ as $\log p(T_i, \delta_i | b_i; \theta, F_n)$. Now we can write the expected complete data profile log-likelihood function as

$$\begin{aligned} E_b[\log L_c(\theta, \sigma^2, D, \hat{\Lambda}_{\theta, F_n}; b)] \\ = \sum_{i=1}^n \left\{ E_b[\log p(T_i, \delta_i | b_i; \theta, F_n)] + E_b[\log p(y_i | b_i; \beta, \sigma^2)] + E_b[\log p(b_i; D)] \right\}. \end{aligned} \quad (7.11)$$

7.3.3 The EM Algorithm

In this section, we will apply the EM algorithm to find the profile likelihood MLE. First, we replace Λ with $\hat{\Lambda}_{\theta, F_n}$ in equation (7.6) and express the posterior distribution as

$$p(b_i | T_i, \delta_i, y_i; \theta, \sigma^2, D, \hat{\Lambda}_{\theta, F_n}) \propto p(T, \delta | b; \theta, F_n) p(y | b; \beta, \sigma^2) p(b; D).$$

The E-step

In the E-step, we use the current parameter estimates to find the expected values of b as

$$\tilde{b}_i = \int_{b_i} b_i p(b | T, \delta, y; \hat{\theta}, \hat{\sigma}^2, \hat{D}, \hat{\Lambda}_{\hat{\theta}, F_n}) db_i, \quad (7.12)$$

and the posterior variance of the random effects can be expressed as

$$\text{var}(b_i) = \int_{b_i} (b_i - \tilde{b}_i)^2 p(b | T, \delta, y; \hat{\theta}, \hat{\sigma}^2, \hat{D}, \hat{\Lambda}_{\hat{\theta}, F_n}) db_i.$$

Here, the integral with respect to random effects is computationally challenging as its dimension increases. Therefore we have used the Gauss-Hermite quadrature (Rizopoulos, 2012) to evaluate the integrals in the estimation procedure.

The M-step

Now we maximize the equation (7.11) with respect to (θ, σ^2, D) to obtain $(\hat{\theta}, \hat{\sigma}^2, \hat{D})$. The estimated parameters from the M-step are returned into E-step then repeat the M-step and E-step until the values of the parameters converge.

In the M-step, we can use the closed form solutions for the estimators of measurement error variance (σ^2) from longitudinal part and covariance matrix (D) from random effect part. The closed form expressions for the estimates of σ^2 and D are

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^n (y_i - x_i \beta)' (y_i - x_i \beta - 2z_i \tilde{b}_i) + \text{tr}(z_i' z_i \text{var}(b_i)) + \tilde{b}_i' z_i z_i' \tilde{b}_i$$

and

$$\hat{D} = \frac{1}{n} \sum_{i=1}^n \text{var}(b_i) + \tilde{b}_i' \tilde{b}_i,$$

where $N = \sum_i n_i$. However, we couldn't obtain the closed form solutions of the score function for $\theta = (\beta', \gamma', \alpha')$. Thus for these parameters, in the M-step, we have implemented a one-step Fisher-scoring update i.e.,

$$\hat{\theta}_n^{(k+1)} = \hat{\theta}_n^{(k)} + [\hat{I}^{(k)}]^{-1} \frac{1}{n} \sum_{i=1}^n \phi(T_i, \delta_i, y_i | \hat{\theta}_n, F_n)^{(k)},$$

where $\phi(T_i, \delta_i, y_i | \hat{\theta}_n, F_n)^{(k)}$ represents the score function at current iteration calculated by using equation (7.19) and \hat{I}^k is the estimated information matrix at current iteration which can be expressed as

$$\hat{I}^{(k)} = \frac{1}{n} \sum_{i=1}^n \phi(T_i, \delta_i, y_i | \hat{\theta}_n, F_n)^{(k)} \phi'(T_i, \delta_i, y_i | \hat{\theta}_n, F_n)^{(k)}.$$

7.3.4 Asymptotic Normality of the MLE

Here we outline the results of asymptotic normality of the profile likelihood estimator in the joint model of longitudinal and survival data.

The difficulty of the problem is the profile likelihood function involve the implicit function which is similar to Cox PH cure model. We can see a circular relationship between the equations (7.10) and (7.12). This relation shows the estimator of the baseline hazard function is an implicit function [Rizopoulos (2012); page-67].

We have successfully proved the asymptotic normality of the profile likelihood estimator in the joint model of longitudinal and survival data based on the following two results:

1. The score functions defined in equations (7.15), (7.16) and (7.17) in Section-7.4 don't involve differentiation of the implicit function and
2. Without assuming the differentiability of the profile likelihood score function (ϕ) with respect to θ and F , we have shown that

$$\begin{aligned} E \left[\sqrt{n} \left\{ \phi(T, \delta | \hat{\theta}_n, F_0) - \phi(T, \delta | \theta_0, F_0) \right\} \right] \\ = -E \left[\phi(T, \delta | \theta_0, F_0) \phi'(T, \delta | \theta_0, F_0) \right] \left\{ \sqrt{n}(\hat{\theta}_n - \theta_0) \right\} + o_p(1). \end{aligned}$$

The proof of the above equation is similar to Lemma-6.3 from Chapter 6. By combining the above two results, we can show not only the asymptotic normality of the estimators but also derived the closed form expression for information matrix in equation (7.13) .

The asymptotic properties of the profile likelihood estimator $\hat{\theta}_n$ with proofs are provided in Theorem-8.1, Theorem-8.2 and Theorem-8.3 (please see Chapter-8, where we have presented the results). The estimator $\hat{\theta}_n$ has asymptotic normal distribution which can be expressed as

$$\hat{\theta}_n \sim N \left\{ \theta_0, \frac{(I^*)^{-1}}{n} \right\},$$

where θ_0 is the true values of θ and $I^* = E[\phi\phi']$ is the efficient information matrix of the joint model, where ϕ given in equation (7.19) is the profile likelihood score function. In Theorem-8.1, it is shown that at the true value of

parameters, ϕ is efficient score function. The consistency of the estimator $\hat{\theta}_n$ are provided in Theorem-8.2 and the asymptotic normality of the estimator $\hat{\theta}_n$ are given in Theorem-8.3. The proofs of Theorem-8.2 and Theorem-8.3 are similar to the Theorem-6.2 and Theorem-6.3 from Chapter-6. Moreover, using equation (7.19), the estimated information matrix can be computed as

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n \phi(T_i, \delta_i, y_i | \hat{\theta}_n, F_n) \phi'(T_i, \delta_i, y_i | \hat{\theta}_n, F_n). \quad (7.13)$$

Note: Assumptions, Lemmas and Theorems are provided in Chapter-8.

Remark-7.1: The existing approaches of the joint model of longitudinal and survival data don't involve direct asymptotic expansion of the profile likelihood function. All the works of Zeng et al. (2005); Zeng and Lin (2007a,b, 2010) have used an 'approximate least favorable submodel' proposed by Murphy and Van der Vaart (2000) to get the variance estimator. Therefore they have used the equation (1.2) from Chapter-1 to calculate the efficient information matrix.

7.4 Score Functions

The score function of β for longitudinal part can be expressed as

$$\phi_{l\beta}(y_i | \beta, \sigma^2) = \frac{\partial}{\partial \beta} E_b[\log p(y_i | b_i; \beta, \sigma^2)] = \left[x_i(y_i - x_i\beta - z_i\bar{b}_i) \right] / \sigma^2. \quad (7.14)$$

The score function of β for survival part can be expressed as

$$\begin{aligned} \phi_{s\beta}(T_i, \delta_i | \theta, F_n) &= \frac{\partial}{\partial \beta} E_b[\log p(T_i, \delta_i | b_i; \theta, F_n)] \\ &= E_b \left[\delta_i \left\{ \alpha x_i(T_i) - \frac{L_{1\beta}(T_i)}{L_0(T_i)} \right\} \right] \\ &\quad - E_b \left[\int_0^{T_i} \left\{ \alpha x_i(s) - \frac{L_{1\beta}(s)}{L_0(s)} \right\} \exp \{ \gamma' w_i + \alpha' m_i(s) \} d\hat{\Lambda}_{\theta, F_n}(s) \right], \end{aligned} \quad (7.15)$$

where

$$L_0(u) = E_{F_n} [Y(u) E_b \exp \{ \gamma' w + \alpha' m(u) \}],$$

$$L_{1\beta}(u) = E_{F_n} [Y(u) E_b \alpha x(u) \exp \{\gamma' w + \alpha' m(u)\}].$$

The score function of α for survival part can be expressed as

$$\begin{aligned} \phi_\alpha(T_i, \delta_i | \theta, F_n) &= \frac{\partial}{\partial \alpha} E_b [\log p(T_i, \delta_i | b_i; \theta, F_n)] \\ &= E_b \left[\delta_i \left\{ m_i(T_i) - \frac{L_{1\alpha}(T_i)}{L_0(T_i)} \right\} \right] \\ &\quad - E_b \left[\int_0^{T_i} \left\{ m_i(s) - \frac{L_{1\alpha}(s)}{L_0(s)} \right\} \exp \{\gamma' w_i + \alpha' m_i(s)\} d\hat{\Lambda}_{\theta, F_n}(s) \right], \end{aligned} \quad (7.16)$$

where

$$L_{1\alpha}(u) = E_{F_n} [Y(u) E_b m(u) \exp \{\gamma' w + \alpha' m(u)\}].$$

The score function of γ for survival part can be expressed as

$$\begin{aligned} \phi_\gamma(T_i, \delta_i | \theta, F_n) &= \frac{\partial}{\partial \gamma} E_b [\log p(T_i, \delta_i | b_i; \theta, F_n)] \\ &= E_b \left[\delta_i \left\{ w_i - \frac{L_{1\gamma}(T_i)}{L_0(T_i)} \right\} \right] \\ &\quad - E_b \left[\int_0^{T_i} \left\{ w_i - \frac{L_{1\gamma}(s)}{L_0(s)} \right\} \exp \{\gamma' w_i + \alpha' m_i(s)\} d\hat{\Lambda}_{\theta, F_n}(s) \right], \end{aligned} \quad (7.17)$$

where

$$L_{1\gamma}(u) = E_{F_n} [Y(u) E_b w \exp \{\gamma' w + \alpha' m(u)\}].$$

From equations (7.14) and (7.15), the score function of β can be written as

$$\phi_\beta(T_i, \delta_i, y_i | \theta, F_n) = \phi_{l\beta}(y_i | \beta, \sigma^2) + \phi_{s\beta}(T_i, \delta_i | \theta, F_n). \quad (7.18)$$

Finally, by combining equations (7.18), (7.16) and (7.17), the score function for the joint model can be expressed as

$$\phi(T_i, \delta_i, y_i | \theta, F_n) = \begin{pmatrix} \phi_\beta(T_i, \delta_i, y_i | \theta, F_n) \\ \phi_\alpha(T_i, \delta_i | \theta, F_n) \\ \phi_\gamma(T_i, \delta_i | \theta, F_n) \end{pmatrix}. \quad (7.19)$$

Now we will calculate the score operator $B(T_i, \delta_i | \theta, F)$, which is Hadamard derivative with respect to F . For an integrable function h with the same domain as F , we can express

$$\begin{aligned} B(T_i, \delta_i | \theta, F)h &= E_b[d_F \log p(T_i, \delta_i | b_i; \theta, F)]h \\ &= E_b \left\{ \delta_i \left[\frac{E_h dN(T_i)}{E_F dN(T_i)} - \frac{E_h Y(T_i) E_b[\exp\{\gamma'w + \alpha'm(T_i)\}]}{E_F Y(T_i) E_b[\exp\{\gamma'w + \alpha'm(T_i)\}]} \right] \right. \\ &\quad - \exp\{\gamma'w_i + \alpha'm_i(T_i)\} \int_0^{T_i} \frac{E_h dN(u)}{E_F Y(u) E_b[\exp\{\gamma'w + \alpha'm(u)\}]} \\ &\quad \left. + \exp\{\gamma'w_i + \alpha'm_i(T_i)\} \int_0^{T_i} \frac{E_F dN(u) E_h Y(u) E_b[\exp\{\gamma'w + \alpha'm(u)\}]}{(E_F Y(u) E_b[\exp\{\gamma'w + \alpha'm(u)\}])^2} \right\}. \end{aligned}$$

where, $E_b[d_F \log p(T_i, \delta_i | b_i; \theta, F)]$ represents the Hadamard derivative of $E_b[\log p(T_i, \delta_i | b_i; \theta, F)]$ with respect to F .

7.5 Application to Real Life Data

We have used the AIDS and PBC datasets from JM package as real data examples to compare our results with the output obtained from JM (“Cox-PH-GH” method) and joineRML packages.

7.5.1 Application to AIDS Data

We have 467 patients in the AIDS dataset having advanced human immunodeficiency virus infection during antiretroviral treatment. The study was conducted to compare the efficiency of two antiretroviral drugs: didanosine (ddI) and zalcitabine (ddC) in the time-to-death. Drug ddI or ddC were assigned randomly to the patients and CD4 cell counts were recorded at study entry randomly at 2, 6, 12, and 18 months thereafter (Abrams et al., 1994). Out of 467 individuals, 188 died (approximately 60% censoring rate).

Here our main aim is to analyze the joint model to study the association between the CD4 count and the risk for death for advanced HIV-

infected patients.. In particular, the CD4 cells are a type of white blood cells which are part of the infection-fighting system. Therefore, a decrease in the CD4 cell count over time is an indication to the worse condition of the immune system of the patient, and therefore higher susceptibility to infection. CD4 represents the CD4 cell count measurements and constitutes an important marker of the strength of the immune system, obstime- the time points at which the corresponding longitudinal response was recorded and treatment- a factor with levels ddC and ddI.

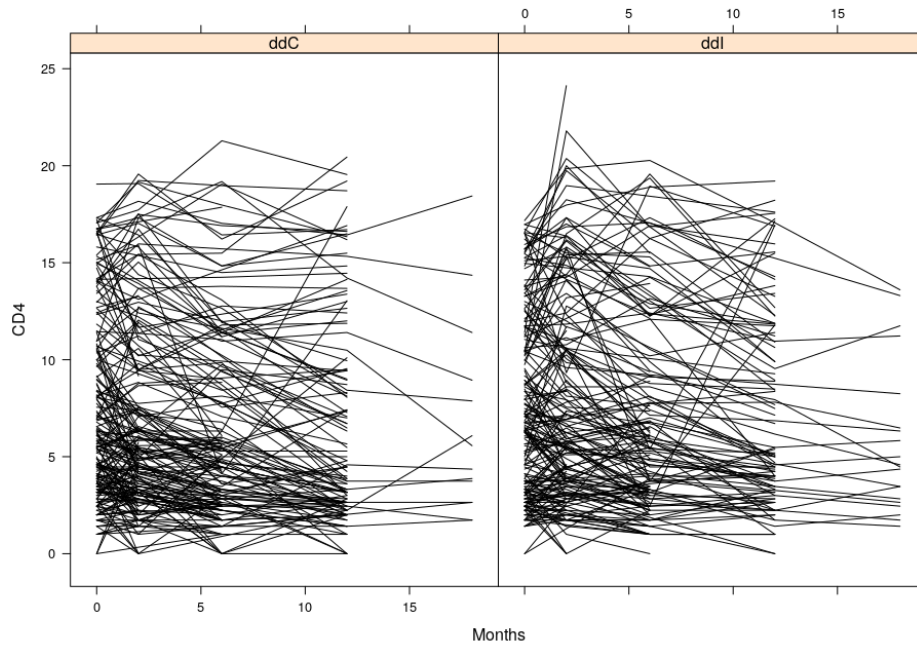
First we perform a mixed model analysis where the model postulates that the CD4 cell counts of all patients have different evolution in time from baseline, i.e., each patient has own intercept and slope with some patients starting with a higher CD4 count and some with lower. As a primary analysis we presented the subject-specific longitudinal profiles (Figure 7.1) and the Kaplan-Meier estimate of survival probability (Figure 7.2).

For mixed model, in the fixed-effects part we include the main effect of obstime and the interaction between treatment and obstime, and in the random-effects design matrix we include an intercept and obstime term. For the survival submodel, we include treatment as a time-independent covariate and the true underlying profile of the CD4 cell count estimated from the longitudinal model as time-dependent covariate.

The Cox model fitted with aids.id dataset (from JM package)- containing only the survival information (i.e., single row per patient). In the results for the survival process, γ is the regression coefficient corresponding to baseline covariate (treatment) and α (association parameter) measures the association between the risk for death and longitudinal covariate.

The parameter estimates, estimated standard errors and 95% CI based on

Figure 7.1: Subject-specific evolutions in time of the CD4 cell count measurements separately for ddC and ddi



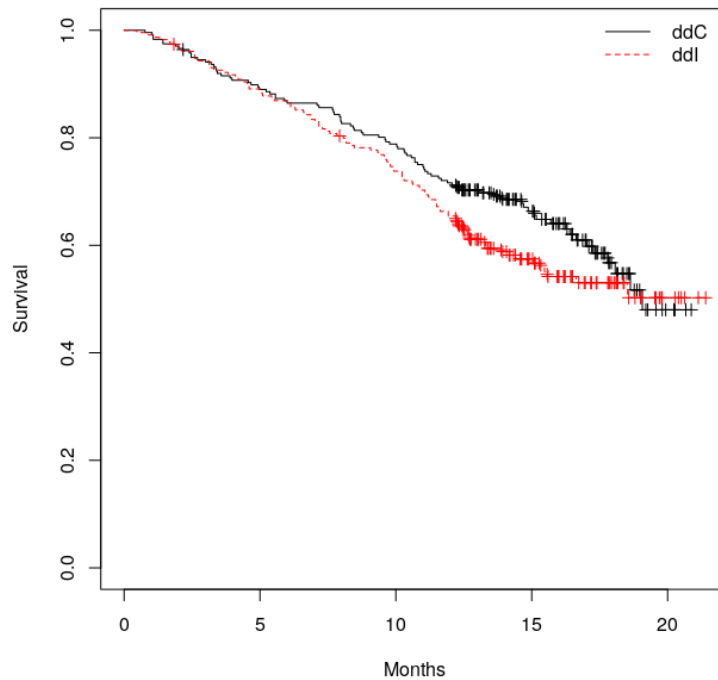
JM package (method:“Cox-PH-GH”), joinerML package and our approach are given in Table 7.1.

From Table 7.1, it is observed that for all three approaches, intercept and obstime have significant effects on CD4 in the longitudinal process. On the other hand, from survival process it is seen that for JM (method:“Cox-PH-GH”) package and our approach, the treatment and CD4 cell counts have significant effects on death of patients. Since α is significant in all three approaches, therefore it can be said that the joint model finds a significant strong association between the CD4 cell count and the risk for death.

Calculation procedure of estimated SE

Though all three approaches have incorporated NPMLE approach for parameter estimation, however, the estimated SE are little bit different from

Figure 7.2: Kaplan-Meier estimate of survival probability for the ddC and ddI treatment groups



each other. The main reason is incorporating different methods to calculate score function which is briefly explained as follows:

While calculating the score functions from JM (method:“Cox-PH-GH”) package, Rizopoulos (2012) has ignored differentiating the estimator of baseline hazard function (within the log-likelihood function) and treated it as an independent parameter. Then forward or central difference approximations have been used to calculate the derivative of the score vector (which was used to estimate the SE of estimated parameters). In the case of *joinerML* package, Hickey et al. (2018) considered the Monte Carlo EM (MCEM) algorithm and bootstrap process for estimation procedure, where they have estimated the fixed effect coefficients by ignoring the survival part from the estimation process.

The advantage of our approach is that we can express the efficient information matrix as the variance of the profile likelihood score function (using equation 7.19). We have calculated the estimated SE of parameter estimates through the square root of the inverse of equation (7.13).

7.5.2 Application to PBC Data

PBC is a fatal and chronic but rare liver disease which eventually leads to cirrhosis of the liver. This study originally conducted by the Mayo Clinic between 1974 and 1984 (Murtaugh et al., 1994) and we are considering 312 patients, where D-penicillamine and placebo were assigned randomly to 158 and 154 patients respectively. In this study, 140 died out of 312 patients (censoring rate over 55%). Patients survival was the primary interest of this study where they also wanted to test the efficiency of D-penicillamine. In addition, follow-up measurements for several biomarkers such as serum bilirubin, prothrombin time, the presence of spiders (blood vessel malformations in the skin), hepatomegaly (enlarged liver) etc. were also recorded.

In this section, we perform a joint model to test whether serum bilirubin is a strong indicator of disease progression or not. From the Q-Q plots of residuals while fitting linear mixed model separately for serum bilirubin, it has been confirmed that log transformation of serum bilirubin is standard to use (Hickey et al., 2018).

First we perform a mixed model analysis where model postulates that the log transformation of serum bilirubin of all patients have different evolution in time from baseline. As a primary analysis we presented the subject-specific longitudinal profiles (log transformation of serum bilirubin) in Figure-7.4 and the Kaplan-Meier estimate of survival probability in Figure-7.5. Our main aim is to study the association between the log transformation of serum

bilirubin and the risk for death.

From Figure-7.3, we can see that some of the patients have non-linear profiles. Therefore, for mixed model, in both fixed-effect and random-effect parts, we consider the natural cubic splines for year (number of years between enrolment and this visit date, remaining values on the line of data refer to this visit) with two internal knots corresponding to 33.3% and 66.7% of the observed follow-up times (Andrinopoulou and Rizopoulos, 2016). For the survival submodel, we include treatment as a baseline covariate and time-dependent one as the true underlying profile of the log transformation of serum bilirubin as estimated from the longitudinal model.

In the results for the survival process, γ is the regression coefficient corresponding to baseline covariate (treatment) and α (association parameter) measures the association between longitudinal covariate (log transformation of serum bilirubin) and the risk for death.

The parameter estimates, standard errors and 95% CI using JM package (method: "Cox-PH-GH"), joineRML package and our approach are given in Table 7.2.

From Table 7.2, we have observed that the intercept and year have significant effects in longitudinal process (in all three approaches). On the other hand, in survival process, only log-transformation of serum bilirubin has significant effect on death of patients. Based on the results from all approaches, it can be said that the joint model finds a significant strong association between the log-transformation of serum bilirubin and the risk for death. From Table 7.2, it is seen that the estimated parameters and SE values are different which similar case to Table 7.1. The possible reason for these differences is explained in the section 7.5.1.

7.6 Simulation Study

We are going to perform the simulation study where our goal is to compare and contrast the JM (“Cox-PH-GH” method) and joineRML packages with our proposed approach by assessing parameter and standard error estimation. For simulating data, we have considered AIDS data from JM package. In the AIDS dataset, there is a variable called ‘prevOI’ which has two levels: ‘aids’ and ‘no-aids’ where the first one denotes previous opportunistic infection (AIDS diagnosis) at study entry and the later one is denoting no previous infection. In this section, we are going to use the subset of patients with ‘aids’ factor from ‘prevOI’ variable for simulating datasets using a function called ‘simulate’ from JM package. This function provides a data frame that contains the simulated responses for the longitudinal process, the simulated event times, the event indicator, the subject identification number and related covariates (Rizopoulos, 2012).

A brief description about the dataset and other variables have provided in section 7.5.1 (Application to AIDS Data). Here, first we perform a mixed model where the model postulates that the CD4 cell counts of patients have different evolution in time from baseline. For mixed model, in the fixed-effects part we include the main effect of ‘obstime’ and ‘treatment’, and in the random-effects design matrix we include an intercept and ‘obstime’ term. For the survival submodel, we have considered ‘treatment’ as time-independent covariate and the true underlying profile of the CD4 cell count (time dependent covariate) estimated from the longitudinal model. Then we fit a joint model using Weibull proportional hazards model with Gauss-Hermite quadrature (method: “weibull-PH-GH”) to describe the evolution in time of the CD4 cell count of the patients with previous opportunistic infection from the AIDS dataset. Using the results from this fitted joint model, we have simulated longitudinal and survival data for our simulation purpose.

The results include samples of 200 and 300 individuals with 500 replications where the results obtained from our approach have compared with the output obtained from JM (“Cox-PH-GH” method) and joineRML packages. The true parameter values and the results from the simulation studies such as mean bias, mean SE (the mean of SE estimates calculated for each fitted model) and MSE for each configuration are given in Table-7.3 and Table-7.4. The explicit explanation of the above mentioned statistics are given below

Let U be the estimator for the parameter of interest θ . From our simulation study ($J = 500$ replicates), we obtain J estimates U_1, U_2, \dots, U_J for each data generating condition. Based on these estimators, we can compute J biases: $\text{bias}_1 = U_1 - \theta, \text{bias}_2 = U_2 - \theta, \dots, \text{bias}_J = U_J - \theta$. Moreover, we obtain J estimates of \hat{SE} ; $\hat{SE}_1, \hat{SE}_2, \dots, \hat{SE}_J$. Now we can calculate the following sample statistics for the estimates:

- Sample mean, $\bar{U} = \sum_{i=1}^J U_i$
- Mean bias = $\frac{1}{J} \sum_{i=1}^J (U_i - \bar{U})$
- $MSE = \frac{1}{J} \sum_{i=1}^J (U_i - \theta)^2$
- mean $\hat{SE} = \frac{1}{J} \sum_{i=1}^J \hat{SE}_i$

Note: Method of mean SE calculation based on our approach

We have calculated the estimated variance analytically through the inverse of equation (7.13). For each simulated data, we can express the observed information matrix as

$$\hat{I} = \frac{1}{p} \sum_{q=1}^p \phi_q \phi_q',$$

where p is the number of individuals in each simulated data and ϕ is the profile score function. Then the estimated SE for each fitted model can be expressed as

$$\hat{SE} = \sqrt{\hat{I}^{-1}}.$$

Results

In Table-7.3 and Table-7.4, the fixed effect coefficients for the longitudinal process are defined as $(\beta_0, \beta_1, \beta_2)$. Moreover, the time independent baseline covariate and the association parameter between longitudinal process and survival outcome are defined as γ and α respectively.

For all configurations in Table-7.3 and Table-7.4, with a few exceptions, the mean bias estimates are very small (the parameter estimates are close to the true values), consistently less than 0.15 with most less than 0.05. For all three approaches, with a few exceptions, bias and MSE for each parameter estimate decrease with increasing sample size.

Computation Time Analysis

In order to assess and compare the computational time between our approach, JM (“Cox-PH-GH” method) and joineRML packages, the average time elapsed to get the estimates of the parameters and estimated standard errors has recorded. For all configurations, 500 datasets have generated as described before. The average computation times for parameter estimates and standard errors for all three approaches are given in Table-7.5.

From Table 7.5, we can see that the average computation time is longest for the JM (“Cox-PH-GH” method) package. The possible reason is the computation process of the derivative of the score function. On the other hand, joineRML package has incorporated bootstrap process for standard error estimation which is comparatively faster than JM (“Cox-PH-GH” method) package.

From Table-7.5, we can see that the average computation times are much less for our approach compared to JM (“Cox-PH-GH” method) and join-

eRML packages. The reason for taking shortest time is our computation approach where the explicit form of the efficient score function has been found via profile likelihood score function and hence we computed the SE through the closed form of the efficient information matrix.

7.7 Discussion

It is difficult to find the estimated standard errors of profile likelihood estimators from the joint model of longitudinal and survival data. The works of Wulfsohn and Tsiatis (1997); Zeng et al. (2005); Zeng and Lin (2007a,b, 2010) have adopted the profile likelihood approach, however, they have avoided expanding the profile likelihood function directly. Therefore, they didn't involve the profile likelihood score function to calculate the standard error.

In this chapter, we have expanded the profile likelihood function directly and found the estimated SE of the profile likelihood MLE in the joint model of longitudinal and survival data. Here, we have shown the estimated SE in terms of the profile likelihood score function. Our approach gave us not only the analytical understanding of the profile likelihood estimation in the joint model of longitudinal and survival data but also an alternative method to compute the estimated SE for profile estimators based on profile likelihood score function.

We have analysed the simulated and real-life data, and compared our results with the output obtained from JM ("Cox-PH-GH" method) and join-eRML packages. Though our approach and the JM ("Cox-PH-GH" method) package provided similar estimates of the parameter, however, the standard errors of the estimated parameters are quite different. In "Cox-PH-GH" method of JM package, the estimated SE are calculated based on the derivative of the score function where the implicit function within the score

functions haven't differentiated. On the other hand, we have used the profile likelihood score function and got the explicit form of the information matrix. In the case of `joineRML` package, bootstrap samples have been used to calculate the SE of estimated parameters. Another advantage of our proposed approach is the computation time which is faster compared to the JM ("Cox-PH-GH" method) and `joineRML` packages.

Table 7.1: Results for joint model using AIDS data

Results for longitudinal process

Parameters	Our approach			JM (method: "Cox-PH-GH")			joinerRML		
	Estimate	SE	95% CI	Estimate	SE	95% CI	Estimate	SE	95% CI
Intercept (β_0)	7.3638	0.1468	(7.0760, 7.6516)	7.3632	0.1400	(7.0888, 7.6376)	7.1914	0.2355	(6.7298, 7.6529)
Obstime (β_1)	-0.1705	0.0276	(-0.2248, -0.1162)	-0.1706	0.0218	(-0.2133, -0.1278)	-0.1897	0.0257	(-0.2400, -0.1393)
Obstime*Treatment: ddl (β_2)	-0.0086	0.0282	(-0.0639, 0.0465)	-0.0086	0.0295	(-0.0664, 0.0492)	0.0137	0.0312	(-0.0474, 0.0748)

Results for survival process

Parameters	Our approach			JM (method: "Cox-PH-GH")			joinerRML		
	Estimate	SE	95% CI	Estimate	SE	95% CI	Estimate	SE	95% CI
Treatment: ddl (γ)	0.3202	0.1667	(0.0067, 0.6471)	0.3206	0.1317	(0.0624, 0.5787)	0.3096	0.1582	(-0.0004, 0.6196)
Association (α)	-0.2660	0.0301	(-0.3250, -0.2070)	-0.2659	0.0252	(-0.3153, -0.2165)	-0.2896	0.0246	(-0.3378, -0.2413)

Figure 7.3: Smooth longitudinal measurements of log transformation of serum bilirubin from 9 randomly selected patients

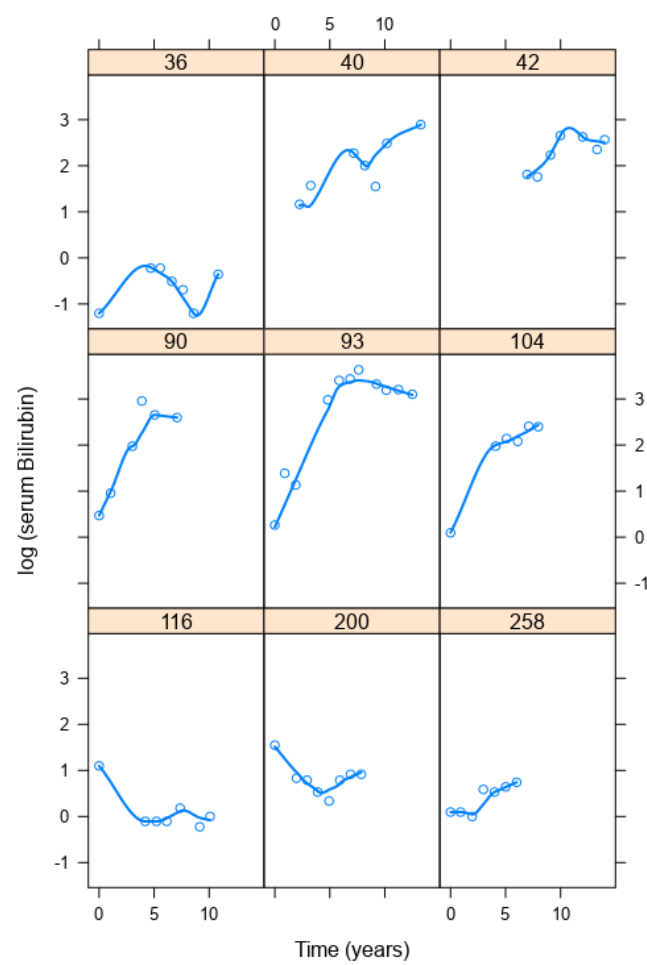


Figure 7.4: Subject-specific evolutions in time of the log transformation of serum bilirubin separately for D-penicillamine and placebo

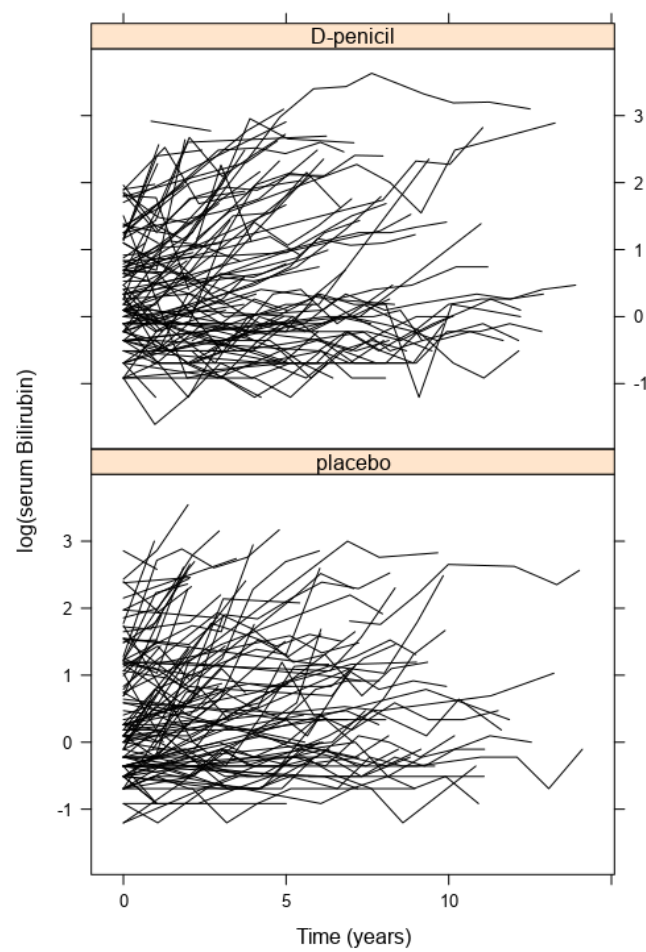


Figure 7.5: Kaplan-Meier estimate of survival probability for the D-penicillamine and placebo treatment groups

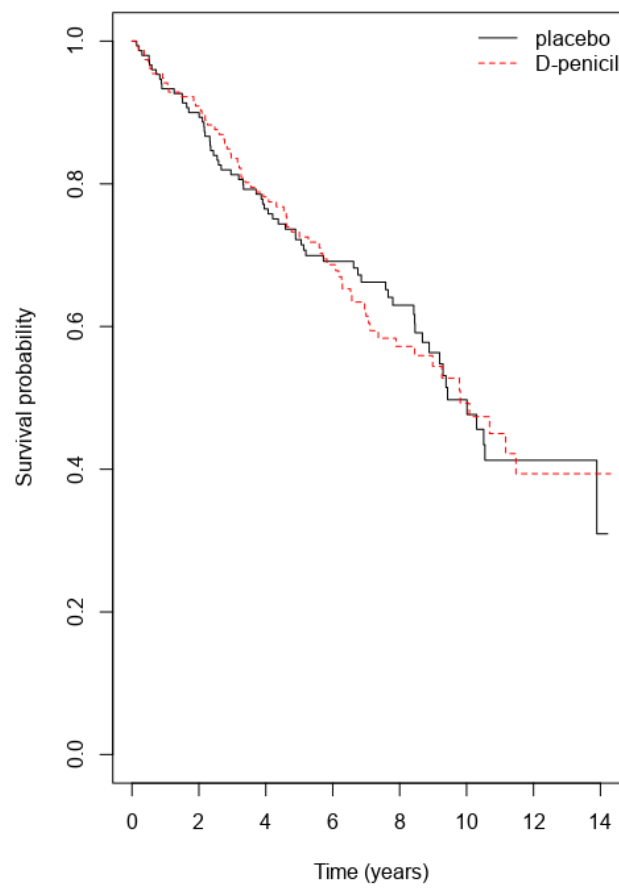


Table 7.2: Results for joint model using PBC data

Results for longitudinal process

Parameters	Our approach			JM package (method: "Cox-PH-GH")			joinerRML package		
	<i>Estimate</i>	<i>SE</i>	<i>95% CI</i>	<i>Estimate</i>	<i>SE</i>	<i>95% CI</i>	<i>Estimate</i>	<i>SE</i>	<i>95% CI</i>
Intercept (β_0)	0.3861	0.0455	(0.2970,0.4753)	0.3865	0.0454	(0.2975,0.4754)	0.5156	0.0608	(0.3964,0.6347)
ns (year, 2)1 (β_1)	1.3954	0.0945	(1.2101,1.5808)	1.3950	0.0914	(1.2158,1.5741)	1.7919	0.1450	(1.5077,2.0761)
ns (year, 2)2 (β_2)	1.6737	0.0586	(1.5586,1.7886)	1.6735	0.0948	(1.4876,1.8593)	1.3988	0.2174	(0.9726,1.8249)

Results for survival process

Parameters	Our approach			JM package (method: "Cox-PH-GH")			joinerRML package		
	<i>Estimate</i>	<i>SE</i>	<i>95% CI</i>	<i>Estimate</i>	<i>SE</i>	<i>95% CI</i>	<i>Estimate</i>	<i>SE</i>	<i>95% CI</i>
Treatment: D-penicil (γ)	-0.0132	0.1834	(-0.3726,0.3463)	-0.0136	0.1691	(-0.3450,0.3178)	0.0047	0.1781	(-0.3443,0.3537)
Association (α)	1.1152	0.0889	(0.9408,1.2895)	1.1147	0.0570	(1.0029,1.2264)	1.1165	0.0785	(0.9626,1.2703)

Table 7.3: Simulation results for Joint model (number of individuals=200,
number of simulations=500)

Longitudinal process

Parameters	True values	Our approach			JM package (method: "Cox-PH-GH")			joinerRML package		
		<i>Bias</i>	<i>SE</i>	<i>MSE</i>	<i>Bias</i>	<i>SE</i>	<i>MSE</i>	<i>Bias</i>	<i>SE</i>	<i>MSE</i>
β_0	5.6847	-0.3857	0.3720	0.3605	-0.3997	0.3226	0.3608	-0.3986	0.3499	0.2892
β_1	-0.1821	-0.0444	0.0267	0.0027	-0.0449	0.0224	0.0028	-0.0442	0.0214	0.0025
β_2	-0.1283	0.1354	0.4495	0.4484	0.1417	0.4132	0.4347	0.1288	0.4867	0.2478

Survival process

Parameters	True values	Our approach			JM package (method: "Cox-PH-GH")			joinerRML package		
		<i>Bias</i>	<i>SE</i>	<i>MSE</i>	<i>Bias</i>	<i>SE</i>	<i>MSE</i>	<i>Bias</i>	<i>SE</i>	<i>MSE</i>
γ	0.1974	-0.0010	0.2445	0.0425	-0.0010	0.1630	0.0425	-0.0021	0.2393	0.0576
α	-0.3232	0.0416	0.0455	0.0037	0.0412	0.0377	0.0036	0.0408	0.0396	0.0035

Table 7.4: Simulation results for Joint model (number of individuals=300,
number of simulations=500)

Longitudinal process

Parameters	True values	Our approach			JM package (method: "Cox-PH-GH")			joinerRML package		
		<i>Bias</i>	<i>SE</i>	<i>MSE</i>	<i>Bias</i>	<i>SE</i>	<i>MSE</i>	<i>Bias</i>	<i>SE</i>	<i>MSE</i>
β_0	5.5808	-0.4137	0.3107	0.3118	-0.4209	0.2739	0.3151	-0.4210	0.2843	0.2595
β_1	-0.1843	-0.0327	0.0203	0.0014	-0.0330	0.0175	0.0014	-0.0317	0.0166	0.0013
β_2	-0.0165	0.0485	0.3733	0.2915	0.0543	0.3555	0.2891	0.0564	0.4009	0.1668

Survival process

Parameters	True values	Our approach			JM package (method: "Cox-PH-GH")			joinerRML package		
		<i>Bias</i>	<i>SE</i>	<i>MSE</i>	<i>Bias</i>	<i>SE</i>	<i>MSE</i>	<i>Bias</i>	<i>SE</i>	<i>MSE</i>
γ	0.1848	0.0066	0.2021	0.0255	0.0065	0.1332	0.0256	-0.0063	0.1978	0.0359
α	-0.3406	0.0360	0.0369	0.0025	0.0358	0.0311	0.0025	0.0357	0.0310	0.0025

Table 7.5: Average computation time for the model parameter estimation
 Average computation time (number of individuals=200)

	Average computation time (seconds)		
Censoring rate 43.01%	Our approach	JM package (method: "Cox-PH-GH")	joinerML package
	16.10	43.96	41.27

Average computation time (number of individuals=300)

	Average computation time (seconds)		
Censoring rate 42.45%	Our approach	JM package (method: "Cox-PH-GH")	joinerML package
	36.54	95.02	71.64

Chapter 8

Proofs of Theorems and Lemmas for Joint Model of Longitudinal and Survival Data

The proof of Theorems and Lemmas to show the asymptotic normality of profile likelihood estimator in the joint model of longitudinal and survival data are given in this chapter. We have summarized the results of this chapter in Section 7.3.4 of Chapter-7. To show the asymptotic normality of the MLE and its asymptotic variance, we consider the following assumptions:

On the set of cdf functions F , we use the sup-norm, i.e., for $F, F_0 \in F$,

$$\|F - F_0\|_\infty = \sup_u |F(u) - F_0(u)|.$$

For $\rho > 0$, let

$$\zeta_\rho = \{F \in F : \|F - F_0\|_\infty < \rho\}.$$

The assumptions are given below

A1: $P(T, \delta | b; \theta_0, F_0) > \delta > 0$ for some positive constant $\delta > 0$ and $S(\tau) = P(T > \tau) = E[Y(\tau)] > 0$.

A2: The range of x , z and w are bounded and $\theta = (\alpha, \beta, \gamma)$ is in the compact set Θ which follows $\|x\| \leq M, \|z\| \leq M, \|w\| \leq M$ and $\|\beta\| \leq M$,

$\|\alpha\| \leq M, \|\gamma\| \leq M$ for some $0 < M < \infty$.

A3: The empirical cdf F_n is \sqrt{n} consistent i.e. $\sqrt{n}|F_n - F_0| = O_p(1)$.

A4: The efficient information matrix, $I^* = E[\phi\phi']$ is invertible where ϕ' is the efficient score function.

A5: True hazard function, $\lambda(t)$ is bounded and positive in $(0, \tau)$.

8.1 Theorem-8.1 with proof

Theorem 8.1: At the true value of (θ, F) , we are going to proof the followings

1. $\hat{\Lambda}_{\theta_0, F_0}(t) = \Lambda_0(t)$, the true cumulative hazard and
2. The score functions for survival part $\phi_{s\beta}$, ϕ_α and ϕ_γ defined in equations (7.15), (7.16) and (7.17) are efficient score functions (we drop the subscript i).

Proof:

1. Replace F_n by F_0 , we get from (7.10)

$$\hat{\Lambda}_{\theta_0, F_0}(t) = \int_0^t \frac{E[dN(u)]}{E[Y(u)E_b \exp\{\gamma'_0 w + \alpha'_0 m(u)\}]}, \quad (8.1)$$

where E is the expectation with respect to the true distribution F_0 . At the true value of the parameters (θ, F) we can write

$$E[dN(u)] = E[Y(u)E_b \exp\{\gamma'_0 w + \alpha'_0 m(u)\}]d\Lambda_0(u). \quad (8.2)$$

So from this point of view, we have $\hat{\Lambda}_{\theta_0, F_0}(t) = \Lambda_0(t)$. Hence (1) is proven.

2. We know that the longitudinal and random effect parts are parametric models those do not involve Λ , so we will work on the score function of the survival part only. Now the score functions of the survival part at the true value of the parameters (θ, F) can be expressed as

$$\begin{aligned}\phi_\alpha(T_i, \delta_i | \theta_0, F_0) &= E_b \left[\delta \left\{ m(T) - \frac{L_{1\alpha_0}(T)}{L_0(T)} \right\} \right] \\ &\quad - E_b \left[\int_0^T \left\{ m(s) - \frac{L_{1\alpha_0}(s)}{L_0(s)} \right\} \exp \{ \gamma'_0 w + \alpha'_0 m(s) \} d\Lambda_0(s) \right]\end{aligned}\tag{8.3}$$

$$\begin{aligned}\phi_{s\beta}(T_i, \delta_i | \theta_0, F_0) &= E_b \left[\delta \left\{ \alpha_0 x(T) - \frac{L_{1\beta_0}(T)}{L_0(T)} \right\} \right] \\ &\quad - E_b \left[\int_0^T \left\{ \alpha_0 x(s) - \frac{L_{1\beta_0}(s)}{L_0(s)} \right\} \exp \{ \gamma'_0 w + \alpha'_0 m(s) \} d\Lambda_0(s) \right]\end{aligned}\tag{8.4}$$

$$\begin{aligned}\phi_\gamma(T_i, \delta_i | \theta_0, F_0) &= E_b \left[\delta \left\{ w - \frac{L_{1\gamma_0}(T)}{L_0(T)} \right\} \right] \\ &\quad - E_b \left[\int_0^T \left\{ w - \frac{L_{1\gamma_0}(s)}{L_0(s)} \right\} \exp \{ \gamma'_0 w + \alpha'_0 m(s) \} d\Lambda_0(s) \right]\end{aligned}\tag{8.5}$$

Now by combining (8.3) , (8.4) and (8.5), we can write the profile likelihood score function for the survival part as

$$\begin{aligned}\phi(T, \delta | \theta_0, F_0) &= \begin{pmatrix} \phi_\alpha(T_i, \delta_i | \theta_0, F_0) \\ \phi_{s\beta}(T_i, \delta_i | \theta_0, F_0) \\ \phi_\gamma(T_i, \delta_i | \theta_0, F_0) \end{pmatrix} \\ &= E_b \left[\delta \left\{ \begin{pmatrix} m(T) \\ \alpha_0 x(T) \\ w \end{pmatrix} - \frac{L_1(T)}{L_0(T)} \right\} \right] \\ &\quad - E_b \left[\int_0^T \left\{ \begin{pmatrix} m(s) \\ \alpha_0 x(s) \\ w \end{pmatrix} - \frac{L_1(s)}{L_0(s)} \right\} \exp \{ \gamma'_0 w + \alpha'_0 m(s) \} d\Lambda(s) \right],\end{aligned}\tag{8.6}$$

where

$$\begin{aligned}
 L_0(s) &= E[Y(s)E_b \exp\{\gamma'_0 w + \alpha'_0 m(s)\}], \\
 L_1(s) &= \begin{pmatrix} L_{1\alpha_0} \\ L_{1\beta_0} \\ L_{1\gamma_0} \end{pmatrix} \\
 &= E \left[Y(s) E_b \begin{pmatrix} m(s) \\ \alpha_0 x(s) \\ w \end{pmatrix} \exp\{\gamma'_0 w + \alpha'_0 m(s)\} \right]. \quad (8.7)
 \end{aligned}$$

Here ϕ_s in equation (8.6) is the profile likelihood score function for the survival part of the joint model of longitudinal and survival data. In section 8.6, we show that the equation (8.6) is same as the efficient score function based on the projection theory given in equation (8.21).

8.2 Lemma-8.1 with Proof

Lemma-8.1: If the assumptions (A1-A5) hold, then

- (i) The class of functions $\left\{ E_b[\log p(T, \delta|b; \theta, F)] : \theta \in \Theta, F \in \zeta_\rho \right\}$ is uniformly bounded Donsker.
- (ii) The class of functions $\left\{ \phi(T, \delta|\theta, F) : \theta \in \Theta, F \in \zeta_\rho \right\}$ is uniformly bounded Donsker.

Proof: Proof is similar to Lemma-6.1 of Chapter-6.

8.3 Theorem- 8.2 with proof

Theorem 8.2: If the assumptions (A1-A5) hold, then

1. $\hat{\Lambda}_{\theta, F}$ is continuously differentiable function at (θ, F) and $\hat{\Lambda}_{\hat{\theta}_n, F_n} = \Lambda_0$
2. $\hat{\theta}_n \xrightarrow{P} \theta_0$ as $n \rightarrow \infty$

Proof: Proof is similar to Theorem- 6.2 of Chapter-6.

8.4 Lemma- 8.3 with proof

Lemma 8.3: Suppose for assumptions **(A1-A5)**, $\hat{\theta}_n \xrightarrow{P} \theta_0$ and $F_n \xrightarrow{P} F_0$ as $n \rightarrow \infty$, then we have

$$\begin{aligned} & E \left[\sqrt{n} \left\{ \phi(T, \delta | \hat{\theta}_n, F_0) - \phi(T, \delta | \theta_0, F_0) \right\} \right] \\ &= -E \left[\phi(T, \delta | \theta_0, F_0) \phi'(T, \delta | \theta_0, F_0) \right] \left\{ \sqrt{n}(\hat{\theta}_n - \theta_0) \right\} + o_p(1) \end{aligned} \quad (8.8)$$

and

$$\begin{aligned} & E \left[\sqrt{n} \left\{ \phi(T, \delta | \hat{\theta}_n, F_n) - \phi(T, \delta | \hat{\theta}_n, F_0) \right\} \right] \\ &= -E \left[\phi(T, \delta | \theta_0, F_0) B(T, \delta | \theta_0, F_0) \right] \left\{ \sqrt{n}(F_n - F_0) \right\} \\ &+ o_p(1 + \sqrt{n}(\hat{\theta}_n - \theta_0)). \end{aligned} \quad (8.9)$$

Proof: Proof is similar to Lemma- 6.3 of Chapter-6.

8.5 Theorem- 8.3 with proof

Theorem 8.3: The estimator $\hat{\theta}_n$ from EM algorithm is an asymptotically linear estimator for θ_0 with the efficient influence function $(I^*)^{-1} \phi(T, \delta | \theta_0, F_0)$, so that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (I^*)^{-1} \phi(T_i, \delta_i | \theta_0, F_0) + o_p(1) \xrightarrow{D} N\{0, (I_s^*)^{-1}\},$$

where $I^* = E[\phi \phi']$ and $N\{0, (I^*)^{-1}\}$ is a normal distribution with mean zero and variance $(I^*)^{-1}$. So the estimator $\hat{\theta}_n$ is efficient.

Proof: Proof is similar to Theorem- 6.3 of Chapter-6.

8.6 Efficient Score Function for Joint Model using Projection Theory

To get the efficient score function using the projection theory, we assume the parameters (θ, Λ) are evaluated at the true values θ_0, Λ_0 and omit subscript

“0” for brevity.

The log-likelihood function of the survival part for one observation can be written as

$$\log P(T, \delta|b; \theta, \Lambda) = E_b \left\{ \delta \left(\log \lambda(T) + \gamma'w + \alpha'm(T) \right) - e^{\gamma'w + \alpha'm(T)} \Lambda(T) \right\}.$$

Score Function for θ

$$\phi_\theta(T, \delta|\theta, \Lambda) = \frac{\partial}{\partial \theta} \log P(T, \delta|b; \theta, \Lambda) = E_b \left\{ \begin{pmatrix} m(T) \\ \alpha x(T) \\ w \end{pmatrix} \left(\delta - e^{\gamma'w + \alpha'm(T)} \Lambda(T) \right) \right\}.$$

Score Operator for Λ

Let us take a measurable function which is bounded such as $g : [0, \tau] \rightarrow R$, where g is defined in the interval $[0, \tau]$ because Λ is also restricted within this interval. The path can be defined as

$$d\Lambda_s = (1 + sg)d\Lambda.$$

The corresponding path for the baseline hazard function is

$$\lambda_s(t) = \frac{d\Lambda_s}{dt} = (1 + sg) \frac{d\Lambda}{dt} = (1 + sg)\lambda(t).$$

The derivative of the log-likelihood function with respect to s can be expressed as

$$B_\Lambda(T, \delta|\theta, \Lambda)g = \frac{\partial}{\partial s} \Big|_{s=0} \log P(T, \delta|b; \theta, \Lambda_s) = E_b \left\{ \delta g(T) - e^{\gamma'w + \alpha'm(T)} \int_0^T g(u) d\Lambda(u) \right\}.$$

Information Operator $B_\Lambda^* B_\Lambda$ and its Inverse $(B_\Lambda^* B_\Lambda)^{-1}$

Let us start with the information operator $B_\Lambda^* B_\Lambda$ and take two arbitrary functions f and g . By definition of the adjoint, we can write

$$\langle B_\Lambda^* B_\Lambda f, g \rangle_{L_2(\Lambda)} = \langle B_\Lambda f, B_\Lambda g \rangle_{L_2(P)}. \quad (8.10)$$

The path defined by $d\Lambda_{r,s} = (1 + rf + sg + rsfg)d\Lambda$ is positive for small r and s . It can be written as $d\Lambda_{r,s} = (1 + rf)(1 + sg)d\Lambda$. The corresponding path for the baseline hazard function is

$$\lambda_{r,s}(T) = \frac{d\Lambda_{r,s}}{dt} = (1 + rf + sg + rsfg) \frac{d\Lambda}{dt} = (1 + rf + sg + rsfg)\lambda(T).$$

Now we can write

$$\left. \frac{\partial}{\partial r} \right|_{(r,s)=(0,0)} \log P(T, \delta | b; \theta, \Lambda_{r,s}) = B_{\Lambda} f, \quad (8.11)$$

and

$$\left. \frac{\partial}{\partial s} \right|_{(r,s)=(0,0)} \log P(T, \delta | b; \theta, \Lambda_{r,s}) = B_{\Lambda} g. \quad (8.12)$$

Using (8.11) and (8.12) we can write

$$\begin{aligned} \langle B_{\Lambda} f, B_{\Lambda} g \rangle_{L_2(P)} &= E \left\{ (B_{\Lambda} f)(B_{\Lambda} g) \right\} \\ &= - E \left\{ \left. \frac{\partial^2}{\partial r \partial s} \right|_{(r,s)=(0,0)} \log P(T, \delta | b; \theta, \Lambda_{r,s}) \right\} \\ &= E \left\{ E_b e^{\gamma'w + \alpha'm(T)} \int_0^T f(\xi)g(\xi)d\Lambda(\xi) \right\}. \end{aligned} \quad (8.13)$$

Now we manipulate the integral involving the function ξ , we deduce

$$\int_0^T f(\xi)g(\xi)d\Lambda(\xi) = \int_0^T I(\xi \leq T)f(\xi)g(\xi)d\Lambda(\xi).$$

Indeed, if $\xi > T$, then the contribution will be 0 to the integral. So the last term in equation (8.13) can be expressed as

$$\begin{aligned} &E \left\{ E_b e^{\gamma'w + \alpha'm(T)} \int_0^T f(\xi)g(\xi)d\Lambda(\xi) \right\} \\ &= E \left\{ E_b e^{\gamma'w + \alpha'm(T)} \int_0^T I(\xi \leq T)f(\xi)g(\xi)d\Lambda(\xi) \right\}. \end{aligned} \quad (8.14)$$

Using Fubini's theorem, equation (8.14) can be written as

$$\begin{aligned} &E \left\{ E_b e^{\gamma'w + \alpha'm(T)} \int_0^T I(\xi \leq T)f(\xi)g(\xi)d\Lambda(\xi) \right\} \\ &= \left\langle E \left\{ E_b e^{\gamma'w + \alpha'm(T)} I(\xi \leq T)f(\xi) \right\}, g(\xi) \right\rangle_{L_2(\Lambda)}. \end{aligned} \quad (8.15)$$

From equation (8.10) we can write

$$\left\langle B_{\Lambda}^* B_{\Lambda} f, g \right\rangle_{L_2(\Lambda)} = \left\langle E \left\{ E_b e^{\gamma' w + \alpha' m(T)} I(t \leq T) f \right\}, g \right\rangle_{L_2(\Lambda)}.$$

So, the information operator is

$$B_{\Lambda}^* B_{\Lambda} f = E \left\{ E_b e^{\gamma' w + \alpha' m(T)} I(t \leq T) \right\} f(t).$$

It follows that the inverse of information operator is

$$(B_{\Lambda}^* B_{\Lambda})^{-1} f(t) = \left[E \left\{ E_b e^{\gamma' w + \alpha' m(T)} I(t \leq T) \right\} \right]^{-1} f(t).$$

The Action of the Adjoint Score Operator B_{Λ}^* on the Score Function ϕ_{θ}

Assume the differentiable paths $(r, s) \mapsto P(T, \delta|b; \theta + ru, \Lambda_s)$ can be exploited with the path $d\Lambda_s = (1 + sg)d\Lambda$. Now we can write

$$\left. \frac{\partial}{\partial r} \right|_{(r,s)=(0,0)} \log P(T, \delta|b; \theta + ru, \Lambda_s) = u' \phi_{\theta}. \quad (8.16)$$

and

$$\left. \frac{\partial}{\partial s} \right|_{(r,s)=(0,0)} \log P(T, \delta|b; \theta + ru, \Lambda_s) = B_{\Lambda} g. \quad (8.17)$$

Using equation (8.16) and (8.17) we can write

$$\begin{aligned} \left\langle u' \phi_{\theta}, B_{\Lambda} g \right\rangle &= E \left\{ (u' \phi_{\theta})(B_{\Lambda} g) \right\} \\ &= -E \left\{ \left. \frac{\partial^2}{\partial r \partial s} \right|_{(r,s)=(0,0)} \log P(T, \delta|b; \theta + ru, \Lambda_s) \right\} \\ &= u' E \left\{ E_b \begin{pmatrix} m(T) \\ \alpha x(T) \\ w \end{pmatrix} e^{\gamma' w + \alpha' m(T)} \int_0^T g(\xi) d\Lambda(\xi) \right\}. \end{aligned} \quad (8.18)$$

Now by manipulating the integral involving the function ξ , the equation (8.18) can be expressed as

$$\left\langle u' \phi_\theta, B_\Lambda g \right\rangle = u' E \left\{ E_b \begin{pmatrix} m(T) \\ \alpha x(T) \\ w \end{pmatrix} e^{\gamma' w + \alpha' m(T)} \int_0^\tau I(\xi \leq T) g(\xi) d\Lambda(\xi) \right\}. \quad (8.19)$$

Using the Fubini's theorem, we can conclude that

$$\begin{aligned} & u' E \left\{ E_b \begin{pmatrix} m(T) \\ \alpha x(T) \\ w \end{pmatrix} e^{\gamma' w + \alpha' m(T)} \int_0^\tau I(\xi \leq T) g(\xi) d\Lambda(\xi) \right\} \\ &= \left\langle u' E \left\{ E_b \begin{pmatrix} m(T) \\ \alpha x(T) \\ w \end{pmatrix} e^{\gamma' w + \alpha' m(T)} I(\xi \leq T) \right\}, g(\xi) \right\rangle_{L_2(\Lambda)}. \end{aligned} \quad (8.20)$$

We know that

$$\left\langle u' B_\Lambda^* \phi_\theta, g \right\rangle_{L_2(P)} = \left\langle u' \phi_\theta, B_\Lambda g \right\rangle_{L_2(\Lambda)}.$$

So we can write

$$B_\Lambda^* \phi_\theta = E \left\{ E_b \begin{pmatrix} m(T) \\ \alpha x(T) \\ w \end{pmatrix} e^{\gamma' w + \alpha' m(T)} I(t \leq T) \right\}.$$

Efficient Score Function ϕ_{eff} :

Finally the efficient score function can be expressed as

$$\begin{aligned} \phi_{eff} &= \phi_\theta - B_\Lambda (B_\Lambda^* B_\Lambda)^{-1} B_\Lambda^* \phi_\theta \\ &= E_b \left[\delta \left\{ \begin{pmatrix} m(T) \\ \alpha x(T) \\ w \end{pmatrix} - \frac{L_1(T)}{L_0(T)} \right\} \right] \\ &\quad - E_b \left[\int_0^T \left\{ \begin{pmatrix} m(s) \\ \alpha x(s) \\ w \end{pmatrix} - \frac{L_1(s)}{L_0(s)} \right\} \exp \{ \gamma' w + \alpha' m(s) \} d\Lambda(s) \right]. \end{aligned} \quad (8.21)$$

where $L_0(T)$ and $L_1(T)$ were defined in the proof of Theorem 8.1 (equation 8.7).

Here we have shown that the profile likelihood score function from equation (8.6) and the efficient score function based on the projection theory from equation (8.21) are identical.

Chapter 9

Conclusion

The Cox PH cure model and joint model of longitudinal and survival data have received much attention in recent years in the field of public health, medical studies and biostatistics. The classical survival models assume that all patients will eventually experience the event of interest. However, mixture cure model such as Cox PH cure models take into account for a proportion of subjects who will never experience the event. Therefore, Cox PH cure model has been used to model the cured and uncured individuals simultaneously. In the case of joint model of longitudinal and survival data, we model the longitudinal process and survival outcome simultaneously. The objective of the joint model is to investigate the effect of time dependent longitudinal process on survival outcome. Therefore, joint model is an useful statistical tool to study the association between repeated measurements and survival outcome.

It is difficult to find the estimated SE for the parametric estimators using profile likelihood approach from the Cox PH cure model and joint model due to the presence of an implicit function in the profile likelihood function which is difficult to differentiate (Hirose and Liu, 2020). In this thesis, we have solved the theoretical challenge for these models by showing the equation (1.3) without assuming the differentiability of the profile like-

likelihood score function. We have expanded the profile likelihood function directly and found the estimated SE of the profile likelihood MLE in the Cox PH cure model and joint model of longitudinal and survival data. Moreover, we have expressed the standard errors in terms of the profile likelihood score function.

The goal of this thesis is to propose an alternative approach of Murphy and Van der Vaart (2000) to analyse the Cox PH cure model and the joint model of longitudinal and survival data based on profile likelihood approach. Due to the theoretical and computational complexity in the method of Murphy and Van der Vaart (2000), the R-packages: `smcure`, `JM` (“Cox-PH-GH” method) and `joineRML` have avoided the method of Murphy and Van der Vaart (2000) to calculate the standard errors. Among these packages, `smcure` and `joineRML` have used bootstrap process to calculate the estimated standard errors (Cai et al., 2012; Hickey et al., 2018). On the other hand, `JM` (“Cox-PH-GH” method) package has used the derivative of the score function (without differentiating the implicit function) for SE calculation. From computational perspective, to show the advantage of our proposed approach, we have performed the simulation studies and analysed the real-life data to compare our results with the output obtained from `smcure`, `JM` (“Cox-PH-GH” method) and `joineRML` R-packages.

In **Chapter-2**, we have reviewed the theory of estimation with finite dimensional parameters and the estimation process with finite and infinite dimensional nuisance parameters. Moreover, we have discussed the concept of the influence function and its relationship to the asymptotic variance of the estimator and showed the asymptotic normality of MLE for all configurations.

In **Chapter-3**, a brief explanation on the profile likelihood concept has

been highlighted. Moreover, we have reviewed the work of Hirose (2011b) where the efficiency of profile likelihood approach in semi-parametric models has been highlighted.

In **Chapter-4**, we briefly described the Cox PH model with estimation procedure based on partial likelihood and profile likelihood approaches. Later, we computed the score function of Cox PH model based on profile likelihood approach, and calculated the efficient score function based on the projection theory. Finally, we have shown that the score function based on profile likelihood and projection theory are identical. Therefore, the score function calculated from profile likelihood approach is efficient score function.

In **Chapter-5**, we have shown the asymptotic normality of the maximum profile likelihood estimator via asymptotic expansion of the profile likelihood and derived the estimated information matrix for the Cox PH cure model (without assuming the differentiability of the profile likelihood score function) to estimate the standard errors. Here, we have expressed the efficient information matrix of the Cox PH cure model as the variance of the profile likelihood score function. Later we have performed simulation study and compared the bias, SE, MSE, coverage rate and computation time with the output obtained from `smcure` R-package. We found that for all configurations, with only a few exceptions, both `smcure` package and our proposed approach have provided similar results. In the case of computation time, our proposed approach takes less time compared to the `smcure` package. Moreover, for real-life data comparison, we have used Eastern Cooperative Oncology Group data. Due to the presence of implicit function in the EM algorithm, `smcure` package has used bootstrap process to calculate the standard errors of the estimated parameters, whereas we have found the explicit form of the efficient score function via profile likelihood score function and

computed the SE analytically through the efficient information matrix.

In **Chapter-6**, we have stated the necessary assumptions and proved the theorems and lemmas which were used to show the asymptotic normality of the profile likelihood estimator in the Cox PH cure model.

In **Chapter-7**, the asymptotic normality of the maximum profile likelihood estimator in the joint model of longitudinal and survival data has been shown via asymptotic expansion of the profile likelihood function. Similar to the Cox PH cure model, we have expanded the profile likelihood function and expressed the efficient information matrix without assuming the differentiability of the profile likelihood score function. For the purpose of analysis, we have used the AIDS and PBC datasets from JM package as real-life data examples to compare our results with the output obtained from JM ("Cox-PH-GH" method) and joinerML packages. In JM (method: 'Cox-PH-GH') package, score function has been calculated from the derivative of the log-likelihood function (by ignoring the derivative of the implicit function within the log-likelihood function) for standard error estimation whereas joinerML package has considered the bootstrap samples. In the case of our approach, we have expressed the efficient information matrix as the variance of the profile likelihood score function. We have also performed the simulation study and compared the bias, SE, MSE and computation time with the output obtained from JM ("Cox-PH-GH" method) and joinerML packages. We found that the computation time for our proposed approach are much less compared to the JM ("Cox-PH-GH" method) and joinerML packages.

In **Chapter-8**, we have proven the theorems and lemmas which were used to show the asymptotic normality of the profile likelihood estimator in the joint model of longitudinal and survival data with random effects.

In summary, we have expanded the profile likelihood function directly for the Cox PH cure model and joint model of longitudinal and survival data. Moreover, we have found the estimated SE of the profile likelihood MLE and expressed the efficient information matrix in terms of the profile likelihood score function. Our proposed technique gave us not only the analytical understanding of the profile likelihood estimation but also provided closed form expression to compute the SE of the profile likelihood MLE without assuming the differentiability of the profile likelihood score function.

9.1 Limitation and future research

In this thesis, we have used e1684 data in Chapter 5, and AIDS and PBC data in Chapter 7 for real-life data analysis. These historic datasets are publicly available and easy to extract from the `smcure`, `JM` ('Cox-PH-GH') and `join-eRML` R-packages. Hence we have used the above mentioned datasets to show the better performance and computational efficiency of our proposed approach over the existing R-packages. However, it is a fact that the examples with publicly available data can be out of date, therefore, for future research we will use modern and up-to-date clinical and medical datasets for example newly released Demographic Health Survey (DHS) datasets. Since the thesis emphasizes the practical importance of the cure model and joint model in the field of medical science, therefore, it will be great to see the plausibility of our proposed approach while using up-to-date clinical datasets.

Another limitation is that we have considered the application of our approach to the interval censored data with the proportional hazard model. However, we are unable to generalize our proposed profile likelihood ap-

proach to the interval censored data. This is because the profile likelihood approach requires a closed form expression of the estimator of the baseline hazard function which is not available in the setting of interval censoring.

For the Cox PH cure model, we have considered only time-independent covariates. The Cox PH cure model is identifiable because the cured subpopulation leads to a specific pattern of decrease in hazard ratios over time. However, the scenarios with time dependent covariates and time-dependent coefficients remain to be established. Therefore, for our future research, we can consider the above mentioned scenarios to check whether it is possible to distinguish the model from one where the effects of covariates just decrease over time.

On the other hand, we can consider several extensions of the joint model of longitudinal and survival data for future research which includes several families of parametrizations for the association structure between the longitudinal and survival outcomes, joint models with multiple longitudinal responses and incorporating stratification factors (either observed or latent). One possible limitation of fitting multivariate joint models is the numerical integration with respect to the random effects which will create the computational complexity. More specifically, when the number of longitudinal outcomes increased from moderate to large, it is quite evident that the dimensionality of the random effects vector is going to be increase as well. Moreover, the situation will become more complex when the subject-specific longitudinal profiles are non-linear (which require even higher-dimensional random effects vectors). Therefore, it will be difficult to handle the complex situation such as joint model (with high-dimensional data) while using the profile likelihood approach and hence we may need to come up with a new approach to handle more complex settings.

Bibliography

- Abrams, D. I., Goldman, A. I., Launer, C., Korvick, J. A., Neaton, J. D., Crane, L. R., Grodesky, M., Wakefield, S., Muth, K., Kornegay, S., et al. (1994). A comparative trial of didanosine or zalcitabine after treatment with zidovudine in patients with human immunodeficiency virus infection. *New England Journal of Medicine*, 330(10):657–662.
- Ahmed, F. E., Vos, P. W., and Holbert, D. (2007). Modeling survival in colon cancer: a methodological review. *Molecular Cancer*, 6(1):15.
- Allison, P. D. (2010). *Survival analysis using SAS: a practical guide*. Sas Institute.
- Amico, M. and Van Keilegom, I. (2018). Cure models in survival analysis. *Annual Review of Statistics and Its Application*, 5:311–342.
- Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (2012). *Statistical models based on counting processes*. Springer Science & Business Media.
- Andersen, P. K. and Gill, R. D. (1982). Cox’s regression model for counting processes: a large sample study. *The annals of statistics*, pages 1100–1120.
- Andersen, P. K. and Keiding, N. (2006). *Survival and event history analysis*. Wiley Chichester.
- Andrinopoulou, E.-R. and Rizopoulos, D. (2016). Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assum-

- ing different association structures. *Statistics in medicine*, 35(26):4813–4823.
- Balakrishnan, N. and Rao, C. R. (2004). *Advances in survival analysis*. Elsevier.
- Bang, H., Chiu, Y.-L., Kaufman, J. S., Patel, M. D., Heiss, G., and Rose, K. M. (2013). Bias correction methods for misclassified covariates in the cox model: comparison of five correction methods by simulation and data analysis. *Journal of statistical theory and practice*, 7(2):381–400.
- Begun, J. M., Hall, W., Huang, W.-M., Wellner, J. A., et al. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *The Annals of Statistics*, 11(2):432–452.
- Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 47(259):501–515.
- Bewick, V., Cheek, L., and Ball, J. (2004). Statistics review 12: survival analysis. *Critical care*, 8(5):389.
- Bickel, P. J., Klaassen, C. A., Bickel, P. J., Ritov, Y., Klaassen, J., Wellner, J. A., and Ritov, Y. (1993). *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins University Press Baltimore.
- Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(1):15–53.
- Breslow, N. (1972). Discussion of regression models and life-tables by cox, dr. *J. Roy. Statist. Assoc., B*, 34:216–217.
- Brilleman, S. L., Wolfe, R., Moreno-Betancur, M., and Crowther, M. J. (2021). Simulating survival data using the *simsurv* r package. *Journal of Statistical Software*, 97(1):1–27.

- Brown, E. R. and Ibrahim, J. G. (2003). Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine trials. *Biometrics*, 59(3):686–693.
- Brown, E. R., Ibrahim, J. G., and DeGruttola, V. (2005). A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61(1):64–73.
- Cai, C. (2013). Advanced methodology developments in mixture cure models.
- Cai, C., Zou, Y., Peng, Y., and Zhang, J. (2012). smcure: An r-package for estimating semiparametric mixture cure models. *Computer methods and programs in biomedicine*, 108(3):1255–1260.
- Chang, I.-S., Wen, C.-C., and Wu, Y.-J. (2007). A profile likelihood theory for the correlated gamma-frailty model with current status family data. *Statistica Sinica*, pages 1023–1046.
- Chappell, R., Nondahl, D. M., and Fowler, J. F. (1995). Modeling dose and local control in radiotherapy. *Journal of the American Statistical Association*, 90(431):829–838.
- Chen, M.-H., Harrington, D. P., and Ibrahim, J. G. (2002). Bayesian cure rate models for malignant melanoma: a case-study of eastern cooperative oncology group trial e1690. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51(2):135–150.
- Chen, M.-H., Ibrahim, J. G., and Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94(447):909–919.
- Chi, Y.-Y. and Ibrahim, J. G. (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62(2):432–445.

- Corbiere, F. and Joly, P. (2007). A sas macro for parametric and semiparametric mixture cure models. *Computer methods and programs in biomedicine*, 85(2):173–180.
- Cox, D. R. (1972). mregression models and life tables (with discussion). *n Journal of the Royal Statistical Society, Series B*, 34(2).
- Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.
- Cox, D. R. and Oakes, D. (1984). Analysis of survival data. 1984. *Chapman&Hall, London*.
- De Gruttola, V. and Tu, X. M. (1994). Modelling progression of cd4-lymphocyte count and its relationship to survival time. *Biometrics*, pages 1003–1014.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.
- Díaz, L. C. and Sampedro, F. G. (2014). Joint modelling for longitudinal and time-to-event data. application to liver transplantation data.
- Dupuy, J.-f. and Mesbah, M. (2002). Joint modeling of event time and non-ignorable missing longitudinal data. *Lifetime data analysis*, 8(2):99–115.
- Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.
- Fang, H.-b., Li, G., and Sun, J. (2005). Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model. *Scandinavian Journal of Statistics*, 32(1):59–75.
- Farewell, V. T. (1977). A model for a binary variable with time-censored observations. *Biometrika*, 64(1):43–46.

- Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, pages 1041–1046.
- Faucett, C. L., Schenker, N., and Taylor, J. M. (2002). Survival analysis using auxiliary variables via multiple imputation, with application to aids clinical trial data. *Biometrics*, 58(1):37–47.
- Faucett, C. L. and Thomas, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a gibbs sampling approach. *Statistics in medicine*, 15(15):1663–1685.
- Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis*, volume 169. John Wiley & Sons.
- Fox, J. and Weisberg, S. (2018). *An R companion to applied regression*. Sage publications.
- Ghitany, M., Maller, R. A., and Zhou, S. (1994). Exponential mixture models with long-term survivors and covariates. *Journal of multivariate Analysis*, 49(2):218–241.
- Goldman, A. I. (1984). Survivorship analysis when cure is a possibility: a monte carlo study. *Statistics in Medicine*, 3(2):153–163.
- Guler, I., Calaza-Díaz, L., Faes, C., Cadarso-Suárez, C., Giraldez, E., and Gude, F. (2014). Joint modelling for longitudinal and time-to-event data: Application to liver transplantation data. In *International Conference on Computational Science and Its Applications*, pages 580–593. Springer.
- Guo, X. and Carlin, B. P. (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, 58(1):16–24.
- Hanson, T. and Yang, M. (2007). Bayesian semiparametric proportional odds models. *Biometrics*, 63(1):88–95.

- Haybittle, J. (1965). A two-parameter model for the survival curve of treated cancer patients. *Journal of the American Statistical Association*, 60(309):16–26.
- Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480.
- Hickey, G. L., Philipson, P., Jorgensen, A., and Kolamunnage-Dona, R. (2018). joinerml: a joint model and software package for time-to-event and multivariate longitudinal outcomes. *BMC medical research methodology*, 18(1):50.
- Hirose, Y. (2011a). Asymptotic linear expansion of profile likelihood in the cox mode. *Mathematical Methods of Statistics*, 20(3):224.
- Hirose, Y. (2011b). Efficiency of profile likelihood in semi-parametric models. *Annals of the Institute of Statistical Mathematics*, 63(6):1247–1275.
- Hirose, Y. and Liu, I. (2020). Statistical generalized derivative applied to the profile likelihood estimation in a mixture of semiparametric models. *Entropy*, 22(3):278.
- Hsieh, F., Tseng, Y.-K., and Wang, J.-L. (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*, 62(4):1037–1043.
- Huang, Y., Dagne, G., and Wu, L. (2011). Bayesian inference on joint models of hiv dynamics for time-to-event and longitudinal data with skewness and covariate measurement errors. *Statistics in Medicine*, 30(24):2930–2946.
- Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). Bayesian semiparametric models for survival data with a cure fraction. *Biometrics*, 57(2):383–388.

- Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2004). Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica*, pages 863–883.
- Ibrahim, J. G., Chu, H., and Chen, L. M. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28(16):2796.
- Jeon, M. and Rabe-Hesketh, S. (2012). Profile-likelihood approach for estimating generalized linear mixed models with factor structures. *Journal of Educational and Behavioral Statistics*, 37(4):518–542.
- Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*, volume 360. John Wiley & Sons.
- Keiding, N. (2007). Survival and event history analysis.
- Kirkwood, J. M., Strawderman, M. H., Ernstoff, M. S., Smith, T. J., Borden, E. C., and Blum, R. H. (1996). Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the eastern cooperative oncology group trial est 1684. *Journal of clinical oncology*, 14(1):7–17.
- Klein, J. P. (1992). Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics*, pages 795–806.
- Kleinbaum, D. and Klein, M. (2011). Survival analysis: A self-learning text, 2005. New York, NY: Spring Science+ Business Media.
- Kuk, A. Y. and Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79(3):531–541.
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, pages 963–974.

- lavalley, M. P. and Deguttola, V. (1996). Models for empirical bayes estimators of longitudinal cd4 counts. *Statistics in Medicine*, 15(21):2289–2305.
- Lawless, J. (2003). Event history analysis and longitudinal surveys. *Analysis of Survey data*, pages 221–243.
- Lewis, T. H. (2016). *Complex survey data analysis with SAS*. CRC Press.
- Li, J. and Ma, S. (2013). *Survival analysis in medicine and genetics*. CRC Press.
- Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.
- Loprinzi, C. L., Laurie, J. A., Wieand, H. S., Krook, J. E., Novotny, P. J., Kugler, J. W., Bartel, J., Law, M., Bateman, M., and Klatt, N. E. (1994). Prospective evaluation of prognostic variables from patient-completed questionnaires. north central cancer treatment group. *Journal of Clinical Oncology*, 12(3):601–607.
- Lu, W. (2008). Maximum likelihood estimation in the proportional hazards cure model. *Annals of the Institute of Statistical Mathematics*, 60(3):545–574.
- Lu, W. and Ying, Z. (2004). On semiparametric transformation cure models. *Biometrika*, 91(2):331–343.
- Maller, R. A. and Zhou, X. (1996). *Survival analysis with long-term survivors*. John Wiley & Sons.
- McCrink, L., Marshall, A. H., Cairns, K., Fogarty, D., and Casula, A. (2011). Joint modelling of longitudinal and survival data: A comparison of joint and independent models. In *Proc 58th World Statistical Congress*.
- Murphy, S. A. (1995). Asymptotic theory for the frailty model. *The annals of statistics*, pages 182–198.

- Murphy, S. A. et al. (1994). Consistency in a proportional hazards model incorporating a random effect. *The Annals of Statistics*, 22(2):712–731.
- Murphy, S. A. and Van der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95(450):449–465.
- Murtaugh, P. A., Dickson, E. R., Van Dam, G. M., Malinchoc, M., Grambsch, P. M., Langworthy, A. L., and Gips, C. H. (1994). Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. *Hepatology*, 20(1):126–134.
- Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of applied econometrics*, 5(2):99–135.
- Othus, M., Barlogie, B., LeBlanc, M. L., and Crowley, J. J. (2012). Cure models as a useful statistical tool for analyzing survival. *Clinical Cancer Research*, 18(14):3731–3736.
- Patilea, V. and Van Keilegom, I. (2017). A general approach for cure models in survival analysis. *arXiv preprint arXiv:1701.03769*.
- Pauler, D. K. and Finkelstein, D. M. (2002). Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. *Statistics in medicine*, 21(24):3897–3911.
- Pawitan, Y. and Self, S. (1993). Modeling disease marker processes in aids. *Journal of the American Statistical Association*, 88(423):719–726.
- Peng, Y. (2003). Fitting semiparametric cure models. *Computational statistics & data analysis*, 41(3-4):481–490.
- Peng, Y. and Dear, K. B. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, 56(1):237–243.
- Peng, Y., Dear, K. B., and Denham, J. (1998). A generalized f mixture model for cure rate estimation. *Statistics in medicine*, 17(8):813–830.

- Peng, Y. and Taylor, J. M. (2014). Cure models. *Handbook of survival analysis*, pages 113–134.
- Peng, Y., Taylor, J. M., and Yu, B. (2007). A marginal regression model for multivariate failure time data with a surviving fraction. *Lifetime data analysis*, 13(3):351–369.
- Persson, I. (2002). *Essays on the assumption of proportional hazards in Cox regression*. PhD thesis, Acta Universitatis Upsaliensis.
- Proust-Lima, C. and Taylor, J. M. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modeling approach. *Biostatistics*, 10(3):535–549.
- Putter, H., Fiocco, M., and Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine*, 26(11):2389–2430.
- Ratcliffe, S. J., Guo, W., and Ten Have, T. R. (2004). Joint modeling of longitudinal and survival data via a common frailty. *Biometrics*, 60(4):892–899.
- Rizopoulos, D. (2010). Jm: An r package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software (Online)*, 35(9):1–33.
- Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. CRC press.
- Robinson, M. (2014). Mixture cure models: simulation comparisons of methods in r and sas.
- Rothwell, P. M., Howard, S. C., Dolan, E., O’Brien, E., Dobson, J. E., Dahlöf, B., Sever, P. S., and Poulter, N. R. (2010). Prognostic significance of visit-to-visit variability, maximum systolic blood pressure, and episodic hypertension. *The Lancet*, 375(9718):895–905.

- Royston, P. et al. (2007). Profile likelihood for estimation and confidence intervals. *Stata Journal*, 7(3):376–387.
- Salkind, N. J. (2008). *Encyclopedia of educational psychology*. SAGE publications.
- Sattar, A., Sinha, S. K., Argyropoulos, C., and Unruh, M. (2012). Joint modeling of all-cause mortality and longitudinally measured serum albumin. *Progress in Applied Mathematics*, 4(2):182–195.
- Scott, A. and Wild, C. (2001). Maximum likelihood for generalised case-control studies. *Journal of Statistical Planning and Inference*, 96(1):3–27.
- Scott, A. J. and Wild, C. J. (1997). Fitting regression models to case-control data by maximum likelihood. *Biometrika*, 84(1):57–71.
- Southern, D. A., Faris, P. D., Brant, R., Galbraith, P. D., Norris, C. M., Knudtson, M. L., Ghali, W. A., Investigators, A., et al. (2006). Kaplan–meier methods yielded misleading results in competing risk scenarios. *Journal of clinical epidemiology*, 59(10):1110–1114.
- Spoto, R., Sather, H. N., and Baker, S. A. (1992). A comparison of tests of the difference in the proportion of patients who are cured. *Biometrics*, pages 87–99.
- Sullivan, L. (2012). *Essentials of biostatistics in public health*. burlington. MA: Jones & Bartlett Learning.
- Sweeting, M. J. and Thompson, S. G. (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, 53(5):750–763.
- Sy, J. and Taylor, J. (2001). Standard errors for the cox proportional hazards cure model. *Mathematical and computer modelling*, 33(12-13):1237–1251.

- Sy, J. P. and Taylor, J. M. (2000). Estimation in a cox proportional hazards cure model. *Biometrics*, 56(1):227–236.
- Taylor, J. M. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics*, pages 899–907.
- Taylor, J. M., Cumberland, W., and Sy, J. (1994). A stochastic model for analysis of longitudinal aids data. *Journal of the American Statistical Association*, 89(427):727–736.
- Terrera, G. M., Piccinin, A. M., Johansson, B., Matthews, F., and Hofer, S. M. (2011). Joint modeling of longitudinal change and survival. *GeroPsych*.
- Tsiatis, A. (2007). *Semiparametric theory and missing data*. Springer Science & Business Media.
- Tsiatis, A. A. and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88(2):447–458.
- Tsiatis, A. A., Degruetola, V., and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*, 90(429):27–37.
- Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.
- Van der Vaart, A. W. and Wellner, J. A. (1996). Weak convergence. In *Weak convergence and empirical processes*, pages 16–28. Springer.
- Viviani, S., Alfó, M., and Rizopoulos, D. (2014). Generalized linear mixed joint model for longitudinal and survival outcomes. *Statistics and Computing*, 24(3):417–427.

- Wang, L. and Dunson, D. B. (2010). Semiparametric bayes multiple testing: Applications to tumor data. *Biometrics*, 66(2):493–501.
- Wang, W. (2004). Proportional hazards regression models with unknown link function and time-dependent covariates. *Statistica Sinica*, pages 885–905.
- Wang, Y. and Taylor, J. M. G. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, 96(455):895–905.
- Withers, H. R., Peters, L. J., Taylor, J. M., Owen, J. B., Morrison, W. H., Schultheiss, T. E., Keane, T., O’Sullivan, B., van Dyk, J., Gupta, N., et al. (1995). Local control of carcinoma of the tonsil by radiation therapy: an analysis of patterns of fractionation in nine institutions. *International Journal of Radiation Oncology* Biology* Physics*, 33(3):549–562.
- Wu, L., Liu, W., Yi, G. Y., and Huang, Y. (2012). Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *Journal of Probability and Statistics*, 2012.
- Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, pages 330–339.
- Xu, C., Baines, P. D., and Wang, J.-L. (2014). Standard error estimation using the em algorithm for the joint modeling of survival and longitudinal data. *Biostatistics*, 15(4):731–744.
- Xu, J. and Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3):375–387.
- Yan, Y. and Yi, G. Y. (2015). A corrected profile likelihood method for survival data with covariate measurement error under the cox model. *Canadian Journal of Statistics*, 43(3):454–480.

- Yang, L. (2014). *Joint models for longitudinal and survival data*. PhD thesis.
- Yang, L. and Gao, S. (2013). Bivariate random change point models for longitudinal outcomes. *Statistics in medicine*, 32(6):1038–1053.
- Yu, B. and Peng, Y. (2008). Mixture cure models for multivariate survival data. *Computational Statistics & Data Analysis*, 52(3):1524–1532.
- Yu, M., Law, N. J., Taylor, J. M., and Sandler, H. M. (2004). Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, pages 835–862.
- Yu, T. (2018). *Joint modelling of complex longitudinal and survival data, with applications to HIV studies*. PhD thesis, University of British Columbia.
- Zeng, D., Cai, J., et al. (2005). Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *The Annals of Statistics*, 33(5):2132–2163.
- Zeng, D. and Lin, D. (2007a). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):507–564.
- Zeng, D. and Lin, D. (2007b). Semiparametric transformation models with random effects for recurrent events. *Journal of the American Statistical Association*, 102(477):167–180.
- Zeng, D. and Lin, D. (2010). A general asymptotic theory for maximum likelihood estimation in semiparametric regression models with censored data. *Statistica Sinica*, 20(2):871.
- Zhao, X. and Zhou, X. (2006). Proportional hazards models for survival data with long-term survivors. *Statistics & probability letters*, 76(15):1685–1693.