REGRESSION MODELS FOR BIVARIATE SURVIVAL DATA

Thesis submitted to the
Cochin University of Science and Technology
for the Award of Degree of

DOCTOR OF PHILOSOPHY

under the Faculty of Science

by

SREEJA V.N.

DEPARTMENT OF STATISTICS
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
COCHIN- 682 022.

MARCH 2008

CERTIFICATE

Certified that the thesis entitled 'Regression models for bivariate survival data' is a bonafide record of work done by Smt. Sreeja V.N. under my guidance in the Department of Statistics, Cochin University of Science and Technology, Cochin-22, Kerala, India and that no part of it has been included anywhere previously for the award of any degree or title.

Cochin-22
14 th March, 2008.

Dr.P.G.Sankaran
(Supervising Guide)

DECLARATION

This thesis contains no material which has been accepted for the award of any Degree or Diploma in any University and to the best of my knowledge and belief, it contains no material previously published by any other person, except where due references are made in the text of the thesis.

Cochin-22
14 th March, 2008.

Sreeja V.N.

Acknowledgements

I deeply express my sincere gratitude and heartfelt indebtedness to my guide Dr. P.G. Sankaran, Reader, Department of Statistics, Cochin University of Science and Technology, for his inspiring guidance, substantial support and encouragement.

I am very much obliged to Dr. N. Balakrishna, Professor and Head, Department of Statistics, Cochin University of Science and Technology, for the valuable suggestions and timely advice.

I am so grateful to each and every teachers of the Department of Statistics who have encouraged and given me moral support.

My most heartfelt thanks go to Prof. J.F. Lawless, University of Waterloo, Canada who helped me with his valuable suggestions in the initial stages of my research work.

It is a pleasure to express my gratitude to Dr. T.M. Jacob, Nirmala College, Muvattupuzha, for his innumerable suggestions and assistance in doing the computational work.

I place on record my profound gratitude to the non-teaching staff of the Department of Statistics for their co-operation and support.

The enthusiastic support and steadfast encouragement got from my family, friends and relatives far exceeds expressions of gratitude. I thank you to everyone who made this possible.

The financial assistance extended by Cochin University of Science and Technology to pursue my research is duly acknowledged.

Above all, for great favours and blessings shown towards me, I now, with a high sense of gratitude, presume to offer my sincere thanks to the Almighty.

Sreeja V.N.

CONTENTS

Chapter One

Preliminaries

1.1. Introduction

Survival data is a term used to refer the data measuring the time to occurrences of certain events. Such data are also referred to as lifetime data or failure time data. Survival data frequently come from medicine, but may come from other applied fields like demography, engineering, economics and social sciences. In the simplest case, the event of interest is death, but the term also covers other events. Survival analysis is the branch of statistics that deals with modeling and analysis of survival data. Some methods of dealing with lifetime data are quite old, but starting about 1970 the field expanded rapidly with respect to methodology, theory, and fields of application.

By survival time, we mean a time from the start of an observation until the occurrence of an event. The event can be death of an individual or failure of some equipment. The following examples illustrate various types of survival data that arise in practical situations.

Example 1.1: In medical studies, the proposed event may be the death of some individual or the occurrence of some disease, which is measured from the date of diagnosis or some other starting point. For example, if we consider the individuals who were diagnosed with AIDS, the event of interest is the time from infection to diagnosis of AIDS (the incubation period).

Example 1.2: In industrial applications, the event is typically time to failure of a unit or a particular component in a unit. For example, Nelson (1972) considered a life test experiment in which specimens of a type of electrical insulating fluid were subject to a constant voltage stress. Then the length of time until each specimen failed or broke down is termed as the event.

Example 1.3: Demographers and social scientists are interested in the duration of certain life 'states' for humans. Consider, for example, the marriages formed during

a period in a particular country. Then the lifetime of a marriage would be its duration; a marriage may end due to annulment, divorce, or death.

1.2. Basic Concepts

Let T be a non-negative random variable representing the time to occurrence of an event. Let $F(t)$ be the distribution function of T, which is absolutely continuous with respect to a Lebesque measure and $f(t)$ be the corresponding probability distribution function (p.d.f.). There are certain basic concepts that characterizes the distribution of T. Survival function, hazard function and mean residual life function are the three concepts, commonly used to explain the physical characteristics of T.

1.2.1. Survival Function

The basic quantity employed to describe time-to-event phenomena is the survival function, which is the probability of an individual surviving beyond time t. Thus the survival function of T, $S(t)$, is given by

$$S(t) = P[T > t] = 1 - F(t). \tag{1.1}$$

When the probability density function of T, $f(t)$ exists, $S(t)$ can be written as

$$S(t) = \int_t^\infty f(u)du. \tag{1.2}$$

The survival function $S(t)$ is a monotone non-increasing, left continuous function with $S(0) = 1$ and $S(\infty) = \lim_{t\to\infty} S(t) = 0$.

It may be noted that

$$f(t) = -\frac{dS(t)}{dt}. \tag{1.3}$$

In the context of analysis of industrial data, $S(t)$ is referred to as the reliability function.

1.2.2. Hazard Function

One of the fundamental concepts in survival analysis is the hazard function. The hazard function of T is defined as

$$\lambda(t) = \lim_{h \to 0} \frac{1}{h} P[t \le T < t + \Delta t \mid T \ge t]. \tag{1.4}$$

The hazard function specifies the instantaneous rate of death or failure of an individual at time t given that the individual survives up to time t. Thus $\lambda(t)\Delta t$ is the approximate probability of death in $[t, t+\Delta t)$, given survival up to t.

When the probability density function (p.d.f.) $f(t)$, exists,

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{-d \ln[S(t)]}{dt}. \tag{1.5}$$

The $\lambda(t)$ indicates the way the risk of failure varies with age or time. Note that $\lambda(t)$ must be non-negative and $\int_0^\infty \lambda(u)du = \infty$.

The cumulative hazard function $\wedge(t)$ is defined as

$$\wedge(t) = \int_0^t \lambda(u)du = -\ln[S(t)]. \tag{1.6}$$

It is well known that $\lambda(t)\ (\wedge(t))$ determines the distribution uniquely by the identity,

$$S(t) = \exp[-\wedge(t)] = \exp\left[-\int_0^t \lambda(u)du\right]. \tag{1.7}$$

The hazard function is also known as conditional failure rate in reliability, the force of mortality in demography, the intensity function in stochastic process, the age-specific failure rate in epidemiology, the inverse of the Mill's ratio in economics, or simply as the hazard function.

1.2.3. Mean Residual Life (MRL) Function

Another basic quantity of interest in survival analysis is the mean residual life function at time t. For a non-negative random variable T, mean residual life function of T is defined as

$$m(t) = E[T - t \mid T > t]. \tag{1.8}$$

The function $m(t)$ represents the average lifetime remaining to a component which is survived up to time t. For a continuous random variable T, $m(t)$ can be written as

$$m(t) = \frac{1}{S(t)} \int_t^\infty S(u)du . \tag{1.9}$$

Note that

$$m(0) = \mu = E(T) = \int_0^\infty S(u)du . \tag{1.10}$$

The function $m(t)$ is related to the hazard function $\lambda(t)$ by

$$\lambda(t) = \frac{1+m'(t)}{m(t)} . \tag{1.11}$$

It is shown that $m(t)$ determine the distribution uniquely by the relationship

$$S(t) = \frac{m(0)}{m(t)} \exp\left[-\int_0^t \frac{du}{m(u)} \right] . \tag{1.12}$$

One set of necessary and sufficient condition for $m(t)$ to be a MRL given by Swartz (1973), is the following

(i) $m(t) \geq 0$

(ii) $m'(t) \geq -1$

and

(iii) $\int_0^\infty \frac{dt}{m(t)}$ should be divergent.

1.3. Censoring

In survival studies, the data may be incomplete due to various reasons. One of the reasons of incompleteness is censoring. Censored data often arises in survival studies because the experimenter is unable to obtain complete information on lifetime of individuals. The study may have terminated before all subjects had experienced an event, or the particular subject may have been lost to the study at some point. The presence of censored observations complicates the analysis of lifetime data. The following are some examples of censored data.

Example 1.4: Censoring is common in clinical trials, since the trial is often terminated before all individuals have failed (died). In addition, individuals may enter a study at various times, and hence may be under observation for different lengths of time. For example, Gehan (1965) has discussed the results of a clinical trial in which the drug 6-mercaptopurine (6-MP) was compared to a placebo with respect to the ability to maintain remission in acute leukemia patients. If the disease was still in a state of remission at the end of the study, those observations are said to be censored.

Example 1.5: In the period 1962-77, 225 patients with malignant melanoma (cancer of the skin) had a radical operation performed at the Department of Plastic Surgery, University Hospital of Odense, Denmark. The tumor was completely removed together with the skin within a distance of about 2.5 cm around it. All patients were followed until the end of 1977, that is, it was noted if and when any of the patients died. Then the lifetime of those individuals are known to be censored who were still alive at the end of 1977.

Example 1.6: Bartholomew (1957) considered a situation in which pieces of equipment were installed in a system at different times. At a later date some of the pieces had failed and the rest were still in use. The first item was installed in June 11 and data were collected up to August 31. At that time, three items had still not failed, and their failure times are therefore censored.

There are three common forms of censoring viz., right censoring, left censoring and interval censoring.

1.3.1. Right Censoring

Right censored survival data is common in clinical trials where some individuals may be lost to follow up for various reasons. We know only the lower bounds on lifetime for those individuals.

Suppose that n individuals have lifetimes represented by $T_1, T_2, ..., T_n$ and the corresponding censoring times are $C_1, C_2, ..., C_n$. Then the observed data contains (t_i, δ_i) with $t_i = \min(T_i, C_i)$ and the indicator function

$$\delta_i = I(T_i = t_i) = \begin{cases} 1 & if \quad T_i = t_i \\ 0 & if \quad T_i > t_i \end{cases}, \; i = 1, 2, ..., n. \tag{1.13}$$

This δ_i is called the censoring or status indicator for t_i, since it tells us if t_i is an observed lifetime ($\delta_i = 1$) or censoring time ($\delta_i = 0$). There are different kinds of right-censored data as described below.

1.3.2. Type I Censoring

A Type I censoring mechanism is said to happen when each individual has a fixed censoring time $C_i > 0$ such that T_i is observed if $T_i \leq C_i$, otherwise, we know only that $T_i > C_i$. Type I censoring often arises when a study is conducted over a specified time period. In clinical trials, there is often staggered entry of individuals to the study combined with a specified end-of-study date.

Assuming that the lifetimes $T_1, T_2, ..., T_n$ are independent and identically distributed (i.i.d.) random variables with common p.d.f. $f(t)$ and survival function $S(t)$, then the likelihood function L under Type I censoring is obtained as

$$L = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}. \tag{1.14}$$

1.3.3. Type II Censoring

Type II censoring refers to the situation where only the r smallest lifetimes $t_{(1)} \leq t_{(2)} \leq ... \leq t_{(r)}$ in a random sample of n are observed; where r is a specified integer between 1 and n. This censoring scheme arises when n individuals start on study at the same time, with the study terminating once r failures have been observed. Although some life tests are formulated with Type II censoring, they have the practical disadvantage that the total time $t_{(r)}$ that the test will run is random and hence unknown at the start of the test.

With Type II censoring, the value of r is chosen before the experiment is performed, and the data consist of the r smallest lifetimes in a random sample $T_1, T_2, ..., T_n$. When the lifetime random variable is continuous, we can ignore the

possibility of ties and denote the r smallest lifetimes as $T_{(1)} < T_{(2)} < ... < T_{(r)}$. If the T_i has p.d.f. $f(t)$ and survival function $S(t)$, then the joint p.d.f. of $T_{(1)}, T_{(2)}..., T_{(r)}$ is given by

$$L = \frac{n!}{(n-r)!}\left[\prod_{i=1}^{r} f(t_i)\right]\left(S(t_{(r)})\right)^{n-r}. \tag{1.15}$$

1.3.4. Progressive Type II Censoring

Progressive Type II censoring is a generalization of Type II censoring. In this context, the first r_1 failures in a life test of n items are observed; then n_1 of the remaining $n - r_1$ unfailed items are removed from the experiment, leaving $n - r_1 - n_1$ items still present when a further r_2 items have failed, n_2 of the still unfailed items are removed, and so on. The experiment terminates after some prearranged series of repetitions of this procedure.

Suppose the censoring has only two stages; at the time of r_1 th failure, n_1 of the remaining $n - r_1$ unfailed items are randomly selected and removed. The experiment then terminates when a further r_2 items have failed. At this point there will be $n - r_1 - n_1 - r_2$ items still unfailed. The observations in this case are the r_1 failure times $T_{(1)} < T_{(2)} < ... < T_{(r_1)}$ in the first stage of the experiment, which we will denote by $T_{(1)}^{*} < T_{(2)}^{*} < ... < T_{(r_2)}^{*}$. When the lifetimes are i.i.d. with common p.d.f. $f(t)$ and survival function $S(t)$, the likelihood function will be

$$L = Cf(t_{(1)})...f(t_{(r_1)})[S(t_{(r_1)})]^{n_1} f(t_{(1)}^{*})...f(t_{(r_2)}^{*})[S(t_{(r_1)})]^{n-r_1-n_1-r_2} \tag{1.16}$$

where

$$C = \frac{n!(n-r_1-n_1)!}{(n-r_1)!(n-r_1-n_1-r_2)!}.$$

1.3.5. Left Censoring

In certain situations, study subjects have experienced the event before study commences. In such situations, we know only the upper bounds on lifetime for these subjects. For instance, we may know that a certain unit failed sometime before 100

hours but not exactly when it happens. In other words, it could have failed any time between 0 and 100 hours.

Example 1.7: Baboons in the Amboseli Reserve, Kenya, sleep in the trees and descend for foraging at some time of the day. Observers often arrive later in the day than this descent and for such days they can only ascertain that descent took place before a particular time, so that the descent times are left censored (see Andersen et al., 1993).

1.3.6. Interval Censoring

Interval censoring occurs when it is not clear when the event occurred, all that is known is that the time to event occurred within some interval $(T_1, T_2]$. Interval censored data reflects uncertainty as to the exact times the subjects failed within an interval. This type of data frequently arises from situations where the objects of interest are not constantly monitored.

Example 1.8: A herd of cows tested bi-weekly for the onset of a disease. Subjects may be seen sporadically due to reasons beyond the investigator's control, such as HIV- positive patients "dropping in" to a health clinic when convenient. In addition, many datasets that have survival recorded in days, weeks, months, and so on, are actually interval censored.

1.4. Truncation

Individuals are sometimes selected and followed prospectively until failure or censoring, but their current lifetime at selection is not at $t = 0$, but some value $u > 0$. The definition of a prospective study is that lifetime information after the time of selection forms the response. Selection of an individual at time u_i thus requires that $T_i \geq u_i$, and the observed data for individual i consist of (u_i, t_i, δ_i), where $t_i \geq u_i$ is a lifetime or censoring time. We say that the lifetime T_i is left truncated at u_i. Left truncation is very common in fields like demography and epidemiology. In many occasions, at least some of the data arises chronologically before the time the individuals are selected for the study. Then the individual is

included in the data when $T_i \leq v_i$ and we say that the individual is right truncated at v_i. Examples of left and right truncation are given below.

Example 1.9: The data of diabetic nephropathy contains all the insulin-dependent patients treated at the Steno Diabetes Center, Denmark, since it opened in 1933 until 1972. The patients were diagnosed between 1933 and 1972, and started treatment at the hospital between 1933 and 1981. They were followed from first visit at the hospital until death, emigration, or January 1, 1984. Thus data are left truncated, both when age and duration of diabetes are used as time scales.

Example 1.10: Kalbfleisch and Lawless (1989) analyzed data on persons infected with HIV via blood transfusion, who were subsequently diagnosed with AIDS. The data were used to estimate the distribution of the time T between HIV infection and AIDS diagnosis. The study group was assembled in 1987 and consisted of individuals who had a diagnosis of AIDS prior to July 1, 1986. For each patient the date of HIV infection could also be ascertained, because the individuals selected were deemed to have contracted the HIV through a blood transfusion on a particular date. The condition for being included in the data set was therefore that $T_i \leq v_i$, where v_i is the time between the individual's HIV infection and July 1, 1986. This is an example of right truncated data.

1.5. Estimation Procedures

One of the objectives in survival analysis is to estimate the survival function. For the estimation of the survival function, there are two common approaches, parametric and non-parametric. In parametric method, it is assumed that the survival time has certain probability distribution $f(t,\theta)$, where the functional form of $f(.)$ is known, but the parameter θ is unknown. There are different procedures for estimating the parameters of the model such as maximum likelihood, method of moments, Bayesian techniques etc. In survival studies, the commonly used parametric models are exponential, Weibull, Pareto, inverse Gaussian, gamma etc. For more details on the estimation procedures of parametric models, one could refer to Martz and Waller (1982), Sinha (1986) and Lawless (2003).

In many situations in survival studies, the given data may not meet the assumptions of the parametric model. In medical studies, sample size may not be adequate for the determination of the parametric model. In addition, censoring and truncation makes problems for the analysis of data using parametric approach. Consequently, non-parametric approach is very common in survival studies.

1.5.1. Non-parametric Estimation

When the data is not censored, the empirical survival function $\hat{S}(t)$, is given by

$$\hat{S}(t) = \frac{\text{Number of observations} \geq t}{\text{n}}, \; t \geq 0 \tag{1.17}$$

is employed to estimate $S(t)$.

When there are censored observations (1.17) can not directly be applied and therefore, Kaplan-Meier (1958) suggested a new non-parametric estimator for the survival function.

1.5.1.1. Kaplan-Meier Estimator

Kaplan and Meier (1958) developed a non-parametric estimator for survival function from censored data. This estimator is also referred as the product limit estimator.

Let (t_i^*, δ_i), $i = 1, 2, ..., n$ represent a censored random sample of lifetimes. Let $d_j = \sum I(t_i^* = t_j, \delta_i = 1)$ represent the number of deaths at t_j, $i, j = 1, 2, ..., n$, $i \neq j$. Then the Kaplan-Meier estimator for $S(t)$ is defined as

$$\hat{S}(t) = \prod_{j:t_j < t} \frac{n_j - d_j}{n_j} \tag{1.18}$$

where, $n_j = \sum I(t_i^* \geq t_j)$ is the number of individuals at risk at t_j, which is the number of individuals alive and uncensored just prior to t_j. The estimate of the variance of $\hat{S}(t)$ is given by

$$\hat{Var}[\hat{S}(t)] = \hat{S}(t)^2 \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)} \tag{1.19}$$

which is referred to as Greenwood's formula.

1.5.1.2. Nelson-Aalen Estimator

From the identity (1.18), it follows that the estimate of the cumulative hazard function can be used for the estimation of the survival function. One could estimate the cumulative hazard function directly using the Nelson-Aalen (NA) estimator. If $t_1, t_2, ..., t_k$ represent the distinct lifetimes at which subjects fails, then the Nelson-Aalen estimate of $\wedge(t)$ is given by

$$\hat{\wedge}(t) = \sum_{j:t_j \le t} \frac{d_j}{n_j}. \tag{1.20}$$

where d_j and n_j are defined as earlier.

This is sometimes called the empirical cumulative hazard function, but is more commonly known as the Nelson-Aalen estimate, having been proposed by Nelson (1969) and by Aalen in a 1972 thesis.

The estimate of the variance of the $\hat{\wedge}(t)$ is given by

$$\hat{Var}[\hat{\wedge}(t)] = \sum_{j:t_j \le t} \frac{d_j(n_j - d_j)}{n_j^3}. \tag{1.21}$$

From (1.7), the estimate of $S(t)$ given by

$$\hat{S}(t) = \exp[-\hat{\wedge}(t)]. \tag{1.22}$$

1.6. Regression Models

In recent years, statistical literature gives considerable interest in specialized methods for the analysis of lifetime data. Much of this interest appears to have been stimulated by problems arising in medical research though rather similar problems arise, for example, in industrial life-testing and demography. A common research question in medical, biological or engineering (lifetime) research is to determine whether or not certain continuous (independent) variables are correlated with the survival times or lifetimes. There are two major reasons why this research issue cannot be addressed via straight forward multiple linear regression techniques. First,

the dependent variable of interest (survival time or lifetime) is most likely not normally distributed. Second, there is the problem of censoring.

The use of explanatory variables, or covariates, in a regression model is an important way to represent heterogeneity in a population. Indeed, the main objective in such studies is to understand and exploit the relationship between lifetime and covariates. For example, in a survival study for lung cancer patients, effect of factors, such as the age and general condition of the patient and the type of tumor in survival time is of interest.

Regression models for lifetimes can be formulated in many ways, and several types are in common use. Regression analysis of lifetimes involves specifications for the distribution of a lifetime T, given a vector of covariates $\underline{z}$. There are two types of covariates: time dependent and time independent. Sometimes a time-varying covariate may be linked physically with the lifetime process. For example, blood pressure may be linked to the time or age at which an individual has a first stroke. Such covariates are termed internal. A covariate which is independent of time is termed as external. Factors such as air pollution or climate conditions, or applied stresses such as voltage or temperature in life test experiments, are examples of such covariates.

The common regression model used in survival studies is the proportional hazards (PH) model introduced by Cox (1972), in which the effect of covariates on the hazard function is studied. There is a vast literature covering the analysis of regression models. For more details, one could refer to Cox and Oaks (1984), Andersen et al. (1993), Klein and Moeschberger (1997), Oakes (2001), Kalbfleisch and Prentice (2002) and Lawless (2003).

1.6.1. Proportional Hazards Model

The proportional hazards model is the most commonly used regression model as it is not based on any assumptions concerning the nature or shape of the underlying survival distribution. Cox (1972) proposed the model as

$$\lambda(t \mid \underline{z}) = \lambda_0(t) r(\underline{z}), \; t > 0 \tag{1.23}$$

where $r(\underline{z})$ and $\lambda_0(t)$ are positive-valued functions and $\lambda(t \mid \underline{z})$ is the hazard function of T given the r-variate covariate vector $\underline{z}$. The function $\lambda_0(t)$ is usually called the baseline hazard function, which is the hazard function for an individual whose covariate vector such that $r(\underline{z}) = 1$. A common specification for $r(\underline{z})$ is $e^{\underline{\beta}'\underline{z}}$. Then (1.23) becomes

$$\lambda(t \mid \underline{z}) = \lambda_0(t) e^{\underline{\beta}'\underline{z}}, \; t > 0. \tag{1.24}$$

where $\underline{\beta}$ is the vector of r - parameters.

The name proportional hazards come from the fact that any two individuals have hazard functions that are constant multiples of one another. The model (1.23) is a semi-parametric since it incorporates the unknown baseline hazard function and parametric vector $\underline{\beta}$. The model specifies a multiplicative relationship between the underlying hazard function and the log-linear function of the covariates. This assumption is also called the 'proportionality assumption'.

Cox's proportional hazards model is a well-recognized statistical model for exploring the relationship between the survival of a patient and several explanatory variables. The model provides an estimate of the treatment effect on survival after adjustment for other explanatory variables. Even if the treatment groups are similar with respect to the variables known to effect survival, using the Cox model with these prognostic variables may produce a more precise estimate of the treatment effect. Interpreting a Cox model involves examining the coefficients for each explanatory variable. A positive regression coefficient for an explanatory variable means that the hazard is higher and thus the prognosis worse, for higher values. A negative regression coefficient implies a better prognosis for patients.

Suppose there are n individuals in the study. $T_1, T_2 ... T_n$ indicates the lifetimes and $C_1, C_2 ..., C_n$ are the corresponding censoring times. Then the observed data contains (t_i, δ_i) with $t_i = \min(T_i, C_i)$ and δ_i the censoring indicator as described in Section 1.3.1 for $i = 1, 2, ..., n$.

The likelihood function for estimating the parameter vector $\underline{\beta}$, which is known as partial likelihood, as suggested by Cox (1972) is given by

$$L(\underline{\beta}) = \prod_{i=1}^{n} \left(\frac{e^{\underline{\beta}' z_i}}{\sum_{l=1}^{n} Y_l(t_i) e^{\underline{\beta}' z_l}} \right)^{\delta_i} \tag{1.25}$$

where $Y_i(t) = I(t_i \geq t)$, $i = 1, 2, ..., n$ is an indicator function.

Then the estimation of $\underline{\beta}$ is obtained by maximizing the partial likelihood (1.25). The generalized Nelson-Aalen estimator is used as the estimator for baseline hazard function.

For the properties of the estimators, one could refer to Kalbfleisch and Prentice (2002) and Lawless (2003).

1.7. Competing Risks Models

In many medical or industrial studies, there are several causes or modes of failure of individuals or components. Such data are commonly referred to as competing risks data, since each causes or modes of failure compete in some sense for the failure of the individual or the component. Following are some examples of competing risks data.

Example 1.11: Consider an experiment in which new models of a small electrical appliance were being tested (Nelson, 1970). The appliances were operated repeatedly by an automatic testing machine. The lifetimes are the number of cycles of use completed until the appliances failed. There were many different ways in which an appliance could fail, which are different possible causes of failure for the appliances.

Example 1.12: Hoel (1972) considered the survival times for two groups of laboratory mice, all of which were exposed to a fixed dose of radiation at an age of 5 to 6 weeks. The first group of mice lived in a conventional lab environment and the second group was kept in a germ-free environment. The causes of death are classified as three for each mouse – thymic lymphoma, reticulum cell sarcoma or other causes.

Two frameworks are used to deal with standard competing risks settings can be observed for an individual:

(i) Cause-specific hazard, $\lambda_j(t)$, formulations, where

$$\lambda_j(t) = \lim_{h \to 0} \frac{1}{h} P[t \le T < t+h, C = j \mid T \ge t], \quad j = 1, 2, ..., k \tag{1.26}$$

and

(ii) Cause-specific sub-distribution function, $F_j(t)$, formulations, where

$$F_j(t) = P(T \le t, C = j), \quad j = 1, 2, ..., k. \tag{1.27}$$

$\lambda_j(t)$ represents the instantaneous rate for failures of type j at time t in the presence of all other failure types, that is, it specifies the rate of type j failures under study conditions. The function $\lambda_j(t)$ is termed 'decremental forces' by the English actuary Makeham (1874) and 'cause-specific hazard function' by Prentice et al. (1978). It is also known as 'force of mortality' and 'force of transition'.

For more details on the analysis of univariate competing risks data, one may refer to Aalen (1976), David and Moeschberger (1978), Anderson et al. (1993), Lin (1997), Cheng et al. (1998), Gooley et al. (1999), Cronin and Feuer (2000), Farley et al. (2001), and Crowder (2001).

When covariate vector $\underline{z}$ is present in the study, the cause-specific hazard and cause-specific sub-distribution functions are defined as

$$\lambda_j(t) = \lim_{h \to 0} \frac{1}{h} P[t \le T < t+h, C = j \mid T \ge t, \underline{z}], \quad j = 1, 2, ..., k \tag{1.28}$$

and

$$F_j(t) = P(T \le t, C = j \mid \underline{z}), \quad j = 1, 2, ..., k. \tag{1.29}$$

The analysis of such data can be done by considering Cox proportional hazards model for different causes. The analysis of competing risks data in the presence of covariates were discussed in literature by different researchers (see Crowder (2001), Fine (2001), Andersen et al. (2002), Kalbfleisch and Prentice (2002) and Lawless(2003)).

1.8. Multivariate Lifetime Data

Multivariate lifetime data arise when each study unit may experience several events or when there exists some natural grouping of subjects, which induces

dependence among lifetimes of the same group. These data are commonly encountered in scientific investigations because each study subject may experience multiple events or because the study involves several members from each group. Examples in biomedical research are the sequence of tumor recurrences or infection episodes, the occurrence of blindness in the left and right eyes, the development of physical symptoms or diseases in several organ systems, the onset of a genetic disease among family members, the initiation of cigarette smoking by classmates, and the appearance of tumors in littermates exposed to a carcinogen. Examples in other area include the repeated breakdowns of a certain type of machinery in industrial reliability, the experiences of different life events by each person in sociological studies, and the purchases of various products by each consumer in market research.

In the bivariate case, let $T=(T_1,T_2)$ be a non-negative random vector having an absolutely continuous distribution function $F(t_1,t_2)$ with respect to a Lebesgue measure. Then the bivariate survival function for T is defined as

$$S(t_1,t_2)=P(T_1>t_1,T_2>t_2). \tag{1.30}$$

The joint probability density function of T is given by

$$f(t_1,t_2)=\frac{\partial^2 S(t_1,t_2)}{\partial t_1 \partial t_2}. \tag{1.31}$$

One can define hazard function of T in more than one way in the bivariate set up. Basu (1971) defined the bivariate hazard function as a scalar quantity and it is given by

$$\lambda(t_1,t_2)=\frac{f(t_1,t_2)}{S(t_1,t_2)}. \tag{1.32}$$

But, the major drawback of the hazard function (1.32) is that it does not determine the joint distribution uniquely. Later, Johnson and Kotz (1975) defined the bivariate hazard function as a vector given by

$$\lambda(t_1,t_2)=\left(\lambda_1(t_1,t_2),\lambda_2(t_1,t_2)\right) \tag{1.33}$$

where

$$\lambda_1(t_1,t_2)=\lim_{h\to 0}\frac{1}{h}P[t_1\le T_1<t_1+h\,|\,T_1\ge t_1,T_2>t_2] \tag{1.34}$$

and

$$\lambda_2(t_1,t_2)=\lim_{h\to 0}\frac{1}{h}P[t_2\le T_2<t_2+h\,|\,T_1>t_1,T_2\ge t_2]. \tag{1.35}$$

$\lambda_1(t_1,t_2)$ is nothing but the instantaneous rate of failure of first individual at time t_1 given that he was alive at the time $T_1 = t_1^-$ and the second individual T_2 survived beyond the time $T_2 = t_2$. The meaning of $\lambda_2(t_1,t_2)$ is similar. It is proved that (1.33) determine the joint distribution of T uniquely.

Dabrowska (1988) provided a representation of bivariate survival function in terms of cumulative hazard function which is a vector of three components that correspond to single and double failures. The cumulative hazard function vector is defined as

$$\wedge(t_1,t_2)=(\wedge_1(t_1,t_2),\wedge_2(t_2,t_1),\wedge_3(t_2,t_1)) \tag{1.36}$$

where

$$\wedge_1(dt_1,t_2)=\frac{P(T_1\in dt_1,T_2>t_2)}{P(T_1\ge t_1,T_2>t_2)}=\frac{-S(dt_1,t_2)}{S(t_1,t_2^-)}$$

$$\wedge_2(t_1,dt_2)=\frac{P(T_1>t_1,T_2\in dt_2)}{P(T_1>t_1,T_2\ge t_2)}=\frac{-S(t_1,dt_2)}{S(t_1^-,t_2)}$$

and

$$\wedge_3(dt_1,dt_2)=\frac{P(T_1\in dt_1,T_2\in dt_2)}{P(T_1\ge t_1,T_2\ge t_2)}=\frac{-S(dt_1,dt_2)}{S(t_1,t_2)}$$

with

$\wedge_1(0,t_2)=\wedge_2(t_1,0)=\wedge_3(0,0)=0$.

When $T=(T_1,T_2)$ has a joint density function, $f(t_1,t_2)$, we have

$$\wedge_1(dt_1,t_2)=\lambda_1(t_1,t_2)dt_1 \tag{1.37}$$

$$\wedge_2(t_1,dt_2)=\lambda_2(t_1,t_2)dt_2 \tag{1.38}$$

and

$$\wedge_3(dt_1,dt_2)=\lambda_3(t_1,t_2)dt_1dt_2 \tag{1.39}$$

with

$$\lambda_3(t_1,t_2)=\lim_{h\to 0}\frac{1}{h}P[t_1\le T_1<t_1+h,t_2\le T_2<t_2+h\,|\,T_1\ge t_1,T_2\ge t_2].$$

The hazard function $\lambda_3(t_1,t_2)$ is the instantaneous rate of failure of both the individuals given that the individuals survived at the time $T_1 = t_1^-$ and $T_2 = t_2^-$.

Then the bivariate survival function is uniquely represented as

$$S(t_1,t_2)=\prod_{u\le t_1}(1-\wedge_1(du,0))\prod_{v\le t_2}(1-\wedge_2(0,dv))\prod_{\substack{u\le t_1\\ v\le t_2}}(1-L(du,dv)) \tag{1.40}$$

where

$$L(du,dv)=\frac{\wedge_1(du,v^-)\wedge_2(u^-,dv)-\wedge_3(du,dv)}{\left(1-\wedge_1(du,v^-)\right)\left(1-\wedge_2(u^-,dv)\right)}.$$

As a natural extension of the univariate definition, Buchanan and Singpurwalla (1977) defined the bivariate mean residual life function as

$$m^*(t_1,t_2)=\frac{1}{S(t_1,t_2)}\int_0^\infty\int_0^\infty S(u_1,u_2)du_1du_2\,.$$

But it does not satisfy the most essential property that it determines the distribution uniquely. A second definition for bivariate mean residual life function is provided by Shanbhag and Kotz (1987) and Arnold and Zahedi (1988), which determine the distribution uniquely.

Estimation of the bivariate survival function when both study units are subject to random censoring in marginal data structures, without covariates, has received a considerable attention in statistical literature. Some of the proposed non-parametric estimators of bivariate survival function are those of Campbell and Foldes (1982), Burke (1988), Dabrowska (1988), Pruitt (1991), Prentice and Cai (1992), van der Laan (1996) and Wang and Wells (1997). Dabrowska (1988), Prentice and Cai (1992) and Pruitt (1991) estimators are not, in general, efficient estimators. Van der Laan's (1996) non-parametric maximum likelihood estimator is globally efficient and typically needs a larger sample size for good performance. Oakes (1989) and Wang and Wells (1999) proposed semi-parametric estimators for the survival function in the presence of covariates. A review on these estimators can

be found in Pruitt (1993) and van der Laan (1997). Quale et al. (2003) proposed a new estimator of the bivariate survival function based on the locally efficient estimation theory. Keles et al. (2004) proposed a bivariate survival function estimator for a general right censored data structure that includes a time dependent covariate process. van der Laan et al. (2002) proposed a locally efficient estimator for multivariate survival function when all the component lifetimes are censored by a common variable, independent of the lifetimes. Akritas and van Keilegom (2003) obtained path-independent bivariate survival function through the estimation of marginal and conditional distributions. Kalbfleisch and Prentice (2002) and Lawless (2003) also discussed different estimation procedures of the bivariate survival function.

The analysis of multivariate lifetime data in the presence of covariates is complicated by the dependence of lifetimes. Lin (1994) provided a detailed description of multivariate lifetime data along with some real biomedical examples. The usual approach is to consider marginal proportional hazards model for each hazard function separately and then apply ideas from generalized estimating function to calculate an appropriate combination of marginal estimates. For the analysis of multivariate lifetime data in the presence of covariates, one could refer to Wei et al. (1989), Prentice and Cai (1992), Spiekerman and Lin (1998), Hougaard (2000), Lin (2000) and Prentice and Kalbfleisch (2003).

1.9. Recurrent Event Data

Many studies in survival analysis involve the recording of times to occurrence of two or more distinct events or failures on each subject. The failures may be repetitions of the same kind of event or may be events of different natures. Such data are referred to as recurrent event data. These multiple events data normally fall into one of two categories, 'parallel' and 'serial'. In the parallel system, several possibly dependent failure processes act concurrently, while in the serial system there is a natural ordering of times of occurrence of events. Medical examples of serial events include the recurrence of a given illness, such as infection episodes and the progression of a disease through successive stages, such as HIV infection $\rightarrow$ AIDS $\rightarrow$ death. Recurrent event data can also be regarded as a specific

type of correlated data. Such data are frequently encountered in health-related studies where longitudinal follow-up designs are commonly employed. In longitudinal follow-up studies, the observation of recurrent events could be terminated at or before the end of the study. Examples of recurrent events in the health and biomedical sciences are repeated hospitalization of patients with chronic diseases, epileptic seizures, multiple opportunistic infections in studies of acquired immunodeficiency syndrome (AIDS) and multiple injuries in ageing studies. In psychiatric studies, the onset of depression and dementia are instances of recurring events; in engineering and reliability setting, the breakdown of mechanical or electronic systems, computer software crashes, stoppages of nuclear power plants, and warranty claims for manufactured products are all examples of recurrent phenomena. Examples in sociology and economics include serious disagreements in a marriage, onset of labor strikes, and auto insurance claims. The development of stochastic models and statistical methods appropriate for the analysis recurrent event data is therefore of considerable importance.

To analyze recurrent event data, the focus can be placed on two types of time scale; the time since entering the study and the time since the last event (gap time). For the situation where the time since study entry is of interest, a variety of statistical methods have been proposed, among them methods proposed by Prentice et al. (1981), Andersen and Gill (1982), Pepe and Cai (1993), Lawless and Nadeau (1995), Lin et al. (2000), Wang et al. (2001) and Pena et al. (2001). These methods consider individuals multiple events as the realization of a counting process and formulate their model based on either the intensity function or the occurrence rate function of the underlying event process (see Cai and Douglas, 2004). The non-parametric estimation of bivariate recurrence time distribution is carried out by Huang and Wang (2005). In the presence of covariates, Chang and Wang (1999) developed inference procedures for the regression models of recurrent event data using conditional approach. Recently, Ebrahimi (2006) introduced marginal proportional hazards models for recurring event times with proper joint density functions.

In many applications, the investigators are more interested in time between consecutive events (gap time) than the total time (Gail et al.1980). For example,

when evaluating the efficacy of a treatment on an episodic illness, it is often important to assess whether or not the treatment delays the time from the initiation of the treatment to the first episode as well as the time from the first episode to the second episode, and so on. The total time from the initiation of the treatment to the second episode is of less interest because a treatment, which delays the first episode, will inevitably lengthen the total time to the second episode even if it becomes ineffective after the first episode. When the study interest is placed on the gap times, the stochastic ordering structure of recurrent events generates challenges for statistical analysis, such as induced dependent censoring and sampling bias and consequently it hampers the development of statistical methods. When the recurrent events are of same type, Wang and Wells (1998) proposed a product limit estimator of the joint survival function of gap times which accounts for the induced dependent censoring. Wang and Chang (1999), then, developed a weighted moment estimator for the inter occurrence time distribution that ignores the last censored observation on all cases experiencing at least one event. Later, Lin et al. (1999) proposed a simple non-parametric estimator for the multivariate distribution function of gap times between the successive events when the follow up time is subject to right censoring.

Recently, in the literature, various statistical methods have also been developed for estimating the joint survival function of series events when events are of different types. These methods can be used for the first pair of recurrent times but such an approach loses efficiency because bivariate recurrent times of higher orders are not used in the estimate. When the events are of different types, various non-parametric methods such as Visser (1996), Hung and Louis (1998), Lin et al. (1999), Huang and Wang (2005) as well as semi parametric methods such as Huang (1999), Chang and Wang (1999) and Chang (2000) have been developed.

In the presence of covariates, Chen et al. (2004) considered the regression problem for gap times using marginal proportional reverse-time hazard function models; when the events are of same type. They used the concept of partial likelihood to develop inference procedures. Recently, Strawderman (2005) introduced an accelerated gap time model for the effect of covariates on the conditional intensity of a recurrent event counting process. For more details on this

topic, one could refer to Pepe and Cai (1993), Lin et al. (2000) and Cai and Douglas (2004). The analysis of multivariate lifetime data in the presence of covariates is usually done by assuming the marginal proportional hazards models for each hazard functions. This procedure would be appropriate in the case of homogeneity among regression coefficients. This brings in the relevance and need of development of new statistical models which are useful for the analysis of different kinds of bivariate (multivariate) lifetime data.

1.10. Present Study

The discussions in previous sections reveal that there has been much research on analyzing various forms of bivariate (multivariate) lifetime data. However, there are various occasions in survival studies where the existing models and methodologies are inadequate for the analysis of bivariate (multivariate) data. The marginal modeling technique existing in literature for the analysis of multivariate survival data is not sufficient to explain the dependence structure of pair of lifetimes on the covariate vector. The objective of the present study is to develop new regression models for the analysis of multivariate lifetime data, arising from different contexts in survival analysis. For simplicity, we consider the bivariate lifetime data, throughout the study.

The thesis is organized as seven chapters of which first chapter is the introductory chapter, where we have pointed out the relevance and scope of the study along with a review of literature. In Chapter 2, we introduce a different approach for modeling bivariate (multivariate) lifetime data, using vector hazard function of Johnson and Kotz (1975). We consider a proportional hazards model in which the covariates under study have different effect on two components of the vector hazard function. The proposed one will be useful in real life situations to study the dependence structure of pair of lifetimes on the covariate vector $\underline{z}$. The univariate model (1.24) can be directly deduced as a special case of the proposed one. Various properties of the model are discussed. We then develop inference procedures for the model. We illustrate the method using two real life data. A simulation work is carried out to study the performance of the estimator.

In Chapter 3, we deal with the regression problem of gap times in which the marginal and conditional hazard functions depend on certain covariates. Since the covariates under study have different effect on marginal and conditional hazard functions, proposed model will be more useful to study the dependence of gap times on covariates. We introduce a bivariate proportional hazards model for gap times. Estimation of parameter vector and baseline hazard function is discussed and asymptotic properties of estimators are also studied. We carried out a simulation study to investigate the finite sample properties of the estimators and their robustness. Finally, we illustrate the procedure with a real life data.

As mentioned earlier, in many fields of application, it is often of interest to analyze the mean residual life function to characterize the stochastic behavior of survival over time. In practical situations, the given recurrent event data may not meet the proportionality assumption among hazard functions. The analysis of gap times for recurrent event data using mean residual life function is an alternative method to analyze the data. In Chapter 4, we propose a bivariate proportional mean residual life model to assess the relationship between mean residual life function and covariates for gap time of recurrent events. Note that the focus will be on the development of the regression model of gap times, when the recurrent events are of same type. Estimators of the parameter vectors as well as baseline mean residual life function are discussed. We apply the model to a kidney dialysis data given in Lawless (2003). A simulation study is carried out to assess the performance of the estimators.

The analysis of multivariate lifetime data is usually done based on the assumption that the lifetime vector and censoring vector are independent. The assumption of independence between the lifetime vector and censoring vector is a strong restriction to apply such models in real life situations. The analysis of duration times under dependent censoring in the presence of covariates is a topic of interest. Motivated by this, in Chapter 5, we consider the regression problem for duration times of successive events under informative censoring. The idea used in Braekers and Veraverbeke (2005) for the analysis of partially informative censored lifetime data in univariate set up, is extended to the analysis of duration times of two successive events. We introduce and study semi-parametric proportional hazards

models for duration times. We estimate the parameters and baseline hazard functions of the model and asymptotic properties of the estimators are studied. A simulation study is carried out to assess the performance of the estimates. We then illustrate the procedure using a real life data.

In many survival studies, we may have multivariate survival data with more than one cause of failure. The analysis of such competing risks data in the presence of covariates is not yet addressed in literature. In Chapter 6, we propose bivariate proportional hazards models for the analysis of competing risks data in the presence of censoring, using vector hazard function of Dabrowska (1988). Estimation of the parameters as well as the cause-specific hazard function is done and various properties of the estimators are discussed. A simulation study is reported to study the finite sample properties of the estimator. The method is illustrated using a real life data.

Finally, Chapter 7 summarizes major conclusions of the study and discusses future works to be carried out in this area.

Chapter Two

Proportional Hazards Model for Bivariate Lifetime Data

2.1. Introduction

Many survival studies record the times to two or more distinct failures on each subject. The failures may be events of different natures or may be repetitions of the same kind. Multivariate lifetime data arise in various forms when individuals are followed to observe the sequence of occurrences of a certain type of event or correlated lifetime when an individual is followed for the occurrence of two or more types of events for which the individual is simultaneously at risk. The analysis of bivariate (multivariate) lifetime data is complicated by the dependence of related lifetimes. One useful approach is to analyze the data using proportional hazards model for marginal distributions (see Wei et al., 1989 and Lee et al., 1992). When each cluster consists of K lifetimes $T_1, T_2, ..., T_K$ with corresponding covariate vectors $\underline{z}_1, \underline{z}_2, ..., \underline{z}_K$, Lin (1994) considered the marginal proportional hazards model and a general methodology adopted for analyzing such data, which is analogous to that of Liang and Zeger (1986) for longitudinal data analysis. Later, Cai and Prentice (1995) modified the approach in Wei et al. (1989) by introducing a weight function into the estimating function for the parameter to improve the efficiency of the estimate. Spiekerman and Lin (1998) extended the marginal modeling technique by allowing separate baseline hazard function among different strata and imposing same baseline hazard function within each stratum. Kalbfleisch and Prentice (2002), Lawless (2003) and Martinussen and Scheike (2006) provide a comprehensive review on this topic.

In many practical situations, as shown in Section 2.2, the marginal modeling

---

The results in this Chapter have been published as entitled "Proportional Hazards Model for Multivariate Failure Time Data", *Communication in Statistics – Theory and Methods*, 36(8), 1627-1642 (see Sankaran and Sreeja (2007)).

technique is not sufficient to explain the dependence structure of pair of lifetimes on the covariate vector. Motivated by this, we introduce a different approach for modeling bivariate (multivariate) lifetime data using vector hazard function of Johnson and Kotz (1975). The univariate model (1.24) can be directly deduced as a special case of the proposed one.

In Section 2.2, we introduce bivariate proportional hazards model and study various properties of the model. In Section 2.3, we develop estimation of the parameters of the model. Estimation of baseline hazard functions is given in Section 2.4. Asymptotic properties of the estimators are discussed in Section 2.5. In Section 2.6, a simulation work is reported to assess the performance of the estimator. We illustrate the method using two real life data in Section 2.7. Finally, Section 2.8 summarizes major conclusion of the study.

2.2. Bivariate Proportional Hazards Model

Let $T = (T_1, T_2)$ be a random vector representing lifetime of pair of study subjects. Let $S(t_1,t_2) = \Pr(T_1 \geq t_1, T_2 \geq t_2)$ denotes the bivariate survival function of T. Then the bivariate hazard vector of Johnson and Kotz (1975) for $T = (T_1, T_2)$ is given by (1.33). The joint survival function of $T = (T_1, T_2)$ can be represented by

$$S(t_1,t_2) = \exp[-\wedge_1(t_1,0) - \wedge_2(t_1,t_2)] \tag{2.1}$$

and

$$S(t_1,t_2) = \exp[-\wedge_1(t_1,t_2) - \wedge_2(0,t_2)] \tag{2.2}$$

where,

$$\wedge_1(t_1,t_2) = \int_0^{t_1} \lambda_1(u,t_2)\,du$$

and

$$\wedge_2(t_1,t_2) = \int_0^{t_2} \lambda_2(t_1,v)\,dv$$

are cumulative hazard functions.

Note that, right side of (2.1) and (2.2) are nothing but the product of marginal survival function of T_i and conditional survival function $T_i \mid T_j \geq t_j$, $i, j = 1,2$, $i \neq j$.

In many practical situations, the conditional hazard functions of T_i given $T_j \geq t_j$ $(i, j = 1,2; i \neq j)$ are better tools than the marginal hazard functions to explain the joint dependence structure of pair of lifetimes on the covariate vector. For an example, we consider the data on Australian Twin Study (Duffy et al., 1990). This study was conducted to compare monozygotic (MZ) and dizygotic (DZ) twins with respect to the strength of dependency of disease risk between pair members, for various diseases. Twin pairs over the age of 17 were asked to provide information on the occurrence, and age at occurrence, of disease- related events, including the occurrence of vermiform appendectomy. Respondents not undergoing appendectomy prior to survey, or suspected of undergoing prophylactic appendectomy, give rise to right- censored times. There are six types of zygosities; 1 = Monozygotic Female-Female pair, 2 = Monozygotic Male-Male pair, 3 = Dizygotic Female-Female pair, 4 = Dizygotic Male-Male pair, 5 = Dizygotic Male-Female pair and 6 = Dizygotic Female-Male pair. The data contains the information of 3808 complete pairs. We now consider two sets, zygosity 1 and 2; each consists of 100 pairs of individuals. T_1 and T_2 represents the two individual's age in the pair at the time of surgery to appendicectomy undergone. Using the method given in Dabrowska (1988), the estimators $\hat{\wedge}_i^{(1)}(t_i, t_j)$ and $\hat{\wedge}_i^{(2)}(t_i, t_j)$ of $\wedge_i(t_i, t_j)$, $i, j = 1,2, i \neq j$ for the two sets are calculated and their ratio are given in Table 2.1.

Table 2.1. Estimates of the cumulative hazard functions, $\hat{\Lambda}_i^{(2)}(t_i,t_j)$

(t_1,t_2)	$\hat{\Lambda}_1^{(1)}$	$\hat{\Lambda}_1^{(2)}$	$\hat{\Lambda}_2^{(1)}$	$\hat{\Lambda}_2^{(2)}$	$\hat{\Lambda}_1^{(1)}/\hat{\Lambda}_1^{(2)}$	$\hat{\Lambda}_2^{(1)}/\hat{\Lambda}_2^{(2)}$
(21,18)	0.1799	0.1183	0.1744	0.1319	1.5207	1.3222
(21,0)	0.2754	0.1508	0	0	1.8263	0
(0,18)	0	0	0.2607	0.1378	0	1.8919
(25,25)	0.2149	0.1454	0.1934	0.1435	1.4779	1.3477
(25,0)	0.3577	0.1830	0	0	1.9546	0
(0,25)	0	0	0.3943	0.1926	0	2.0472
(21,25)	0.1308	0.1248	0.2429	0.1799	1.0481	1.3502
(25,18)	0.2569	0.1364	0.1308	0.1364	1.8834	0.9589

The ratio of the marginal hazards for two sets is not constant. For example, the ratio of the marginal hazards for the sets (21,0) and (25,0) are different. The ratio $\dfrac{\hat{\Lambda}_i^{(1)}(t_i,t_j)}{\hat{\Lambda}_i^{(2)}(t_i,t_j)}$ depends on $t_j, i,j=1,2, i\neq j$. Accordingly, modeling of conditional hazard functions is useful to explain the dependence structure of pair of lifetimes on the covariate vector. With this motivation, in the following, we propose a bivariate proportional hazards model.

The bivariate proportional hazards model is defined as

$$\lambda_i\left(t_i,t_j \mid \underline{z}\right)=\lambda_{i0}\left(t_i,t_j\right)e^{\underline{\beta}_i'(t_j)\underline{z}},\ i,j=1,2,\ i\neq j \tag{2.3}$$

where, $\underline{z}$ is a $r\times 1$ covariate vector, $\underline{\beta}_i(t_j)$ is the $r\times 1$ parameter vector, $\lambda_i\left(t_i,t_j \mid \underline{z}\right)$ is the hazard function of the pair of lifetimes $T=(T_1,T_2)$ given the covariate vector $\underline{z}$ and $\lambda_{i0}\left(t_i,t_j\right), i,j=1,2,\ i\neq j$ is an unspecified baseline hazard function. When $\underline{\beta}_i(t_j)$ is a zero vector, the covariates has no effect on the hazard functions. The model (2.3) implies that for $T_j \geq t_j$, the ratio $\dfrac{\lambda_i(t_i,t_j \mid \underline{z}^{(1)})}{\lambda_i(t_i,t_j \mid \underline{z}^{(2)})}$ of the hazard functions of two pairs with covariate vectors $\underline{z}^{(1)}$ and $\underline{z}^{(2)}$ does not vary with time t_i, $i,j=1,2,\ i\neq j$. Accordingly, $\underline{\beta}_i(t_j)$ depends on t_j, but not on t_i, $i,j=1,2,\ i\neq j$. Thus the covariates under study have different effect on two

components of the vector hazard function. Hereafter, we denote $\underline{\beta}_i(t_j)$ as $\underline{\beta}_i$ for $i, j = 1,2$, $i \neq j$.

From (2.1), (2.2) and (2.3), we can represent survival function of T as

$$S(t_1, t_2 \mid \underline{z}) = \exp[-\wedge_{10}(t_1, 0)e^{\underline{\beta}_1'\underline{z}} - \wedge_{20}(t_1, t_2)e^{\underline{\beta}_2'\underline{z}}] \tag{2.4}$$

and

$$S(t_1, t_2 \mid \underline{z}) = \exp[-\wedge_{10}(t_1, t_2)e^{\underline{\beta}_1'\underline{z}} - \wedge_{20}(0, t_2)e^{\underline{\beta}_2'\underline{z}}] \tag{2.5}$$

where, $\wedge_{i0}(t_i, t_j)$ is the baseline cumulative hazard function corresponding to $\lambda_{i0}(t_i, t_j)$, $i, j = 1,2$, $i \neq j$.

When $t_i \to 0$, $\wedge_{i0}(t_i, t_j) \to 0$ and thus $S(t_1, t_2 \mid \underline{z})$ in (2.4) reduces to the univariate proportional hazards model for $T_j, i, j = 1,2$, $i \neq j$.

2.3. Estimation of Regression Parameters

In this section, we discuss the estimation of parameter vectors $\underline{\beta}_i, i = 1,2$ using Cox's partial likelihood. From (2.4) and (2.5), we can see that the parameter vector $\underline{\beta}_i$ is associated with the conditional variable $T_i \mid T_j \geq t_j$, $i, j = 1,2$, $i \neq j$. As in the univariate set up, we can employ the method of partial likelihood developed by Cox (1975) to estimate parameter vectors $\underline{\beta}_i, i = 1,2$. In the following, we give the procedure for the estimation of $\underline{\beta}_i, i = 1,2$.

We label n pairs $(11,21),(12,22)...(1n,2n)$. For fixed t_j, suppose that there are $n_i (\leq n)$ pairs satisfying the condition $T_j \geq t_j, i, j = 1,2, i \neq j$. For $T_j \geq t_j, j = 1,2$ suppose that k_i items labeled $11,12,...1k_i$ give rise to observed lifetimes $t_{i1} \leq t_{i2} \leq \ldots \leq t_{ik_i}$ of the i^{th} component with corresponding covariate vectors $\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_{1k_i}, i = 1,2$. Then the remaining $n_i - k_i$ individuals of $T_i, i = 1,2$ are right censored. Following the approach given in Kalbfleisch and Prentice (2002), for fixed $T_j \geq t_j$ we can directly obtain the partial likelihood for $\underline{\beta}_i$ as

$$L(\underline{\beta}_i)=\prod_{m=1}^{n_i}\left[\frac{e^{\underline{\beta}_i' z_{im}}}{\sum_{l=1}^{n_i} Y_{il}(t_{im})e^{\underline{\beta}_i' z_{il}}}\right]^{\delta_{im}} \tag{2.6}$$

where, $Y_{im}(t_i)=I(t_{im}\geq t_i), m=1,2,...n_i$ and $\delta_{im}=1$, if t_{im} is a lifetime and $\delta_{im}=0$, if t_{im} is a censoring time for $i=1,2$.

Note that, when $t_j=0$, n_i will be n, $i,j=1,2$, $i\neq j$.

The maximum partial likelihood estimates of $\underline{\beta}_i$ is the value of $\underline{\beta}_i, i=1,2$ that maximizing (2.6); which will be obtained through numerical methods. It is important to note that the estimates of the vector $\underline{\beta}_i$ may be different for different values of $t_i, i=1,2$.

As in the univariate set up, if there are a substantial number of ties, the discrete nature of the lifetimes should be considered. Suppose that, of n pairs under test, for $T_j\geq t_{jm}$, d_{im} units are observed to fail at $t_{im}, m=1,2,...,k_i$ where, $t_{i1}\leq t_{i2}\leq \ldots \leq t_{ik_i}$ and $\sum_{m=1}^{k_i} d_{im}=k_i, i=1,2$.

In this case, the likelihood of $\underline{\beta}_i$ can be approximated by

$$L^*(\underline{\beta}_i)=\prod_{m=1}^{n_i}\left[\frac{e^{\underline{\beta}_i' \underline{s}_{im}}}{\left(\sum_{l=1}^{n_i} Y_{il}(t_{im})e^{\underline{\beta}_i' z_{il}}\right)^{d_{im}}}\right]^{\delta_{im}} \tag{2.7}$$

where, $\underline{s}_{im}=\sum_{r=1}^{d_{im}} \underline{z}_{imr}$ is the sum of the covariate of individual observed to fail at t_{im} with δ_{im}, $Y_l(t_{im})$ and n_i for $i=1,2$ are same as above.

Thus we obtain estimates of $\underline{\beta}_i, i=1,2$ by maximizing the likelihood (2.7).

2.4. Estimation of Baseline Hazard Functions

In this section, we discuss the estimation of baseline hazard function. Since $\wedge_{i0}(t_i,t_j)$ is the cumulative hazard function of random variable $T_i \mid T_j \geq t_j$, $i,j=1,2, i \neq j$, the estimate of baseline cumulative hazard function is given by

$$\hat{\wedge}_{i0}(t_i,t_j)=\sum_{m:t_{im}\leq t_i} \frac{\delta_{im}}{\sum_{l=1}^{n_i} Y_{il}(t_{im}) e^{\hat{\underline{\beta}}_i' \underline{z}_{il}}}. \tag{2.8}$$

When $t_j \to 0$, the cumulative hazard function $\wedge_{i0}(t_i,0)$ for the random variable T_i can be estimate using the generalized Breslow (1974) estimator (see Kalbfleisch and Prentice (2002), p. 104). When $\hat{\underline{\beta}}_i = 0, i=1,2$, (2.8) reduces to Nelson-Aalen estimates of the cumulative hazard function in univariate set up. Then the survival function can be estimated either by

$$\hat{S}_1(t_1,t_2 \mid \underline{z}) = \exp[-\hat{\wedge}_{10}(t_1,0) e^{\hat{\underline{\beta}}_1' \underline{z}} - \hat{\wedge}_{20}(t_1,t_2) e^{\hat{\underline{\beta}}_2' \underline{z}}] \tag{2.9}$$

or by

$$\hat{S}_2(t_1,t_2 \mid \underline{z}) = \exp[-\hat{\wedge}_{10}(t_1,t_2) e^{\hat{\underline{\beta}}_1' \underline{z}} - \hat{\wedge}_{20}(0,t_2) e^{\hat{\underline{\beta}}_2' \underline{z}}]. \tag{2.10}$$

The estimator of $S(t_1,t_2 \mid \underline{z})$ obtained by (2.9) and (2.10) may be different. To get a unique estimator, we consider a convex combination of the two expressions (2.9) and (2.10). Thus the estimator of $S(t_1,t_2 \mid \underline{z})$ is given by

$$\hat{S}(t_1,t_2 \mid \underline{z}) = a(t_1,t_2)\hat{S}_1(t_1,t_2 \mid \underline{z}) + (1-a(t_1,t_2))\hat{S}_2(t_1,t_2 \mid \underline{z}). \tag{2.11}$$

Expressions (2.9) and (2.10) are proper bivariate survival functions in the sense that they assign positive mass to any rectangle in the plane. Hence $\hat{S}(t_1,t_2 \mid \underline{z})$ is also a proper bivariate survival function provided that $a(t_1,t_2)$ is a constant between 0 and 1. $S(t_1,t_2 \mid \underline{z})$ may assign negative mass to certain rectangles when $a(t_1,t_2)$ changes with (t_1,t_2). However, in practice, the weights are chosen in such a way that they do not make abrupt changes (Akritas and van Keilegom (2003)). Choose $a(t_1,t_2)$ in

such a way that the mean square error of $\hat{S}(t_1,t_2 \mid \underline{z})$ is minimum. Mean square error of $\hat{S}(t_1,t_2 \mid \underline{z})$ is given by

$$E[a\hat{S}_1+(1-a)\hat{S}_2-S]^2 = a^2[\sigma_{11}+\sigma_{22}-2\sigma_{12}+\mu_1^2+\mu_2^2-2\mu_1\mu_2]$$
$$-2a[\sigma_{22}+\mu_2^2-\sigma_{12}-\mu_1\mu_2]+\sigma_{22}+\mu_2^2 \qquad (2.12)$$

where, σ_{ij} is the asymptotic covariance between $\hat{S}_i(t_1,t_2)$ and $\hat{S}_j(t_1,t_2)$ and μ_j is the asymptotic bias of $\hat{S}_j(t_1,t_2)$, $i, j = 1,2, i \neq j$. Thus we can obtain $a(t_1,t_2)$ which minimized the mean square error as

$$a(t_1,t_2) = \frac{\sigma_{22}-\sigma_{12}+\mu_2^2-\mu_1\mu_2}{\sigma_{11}+\sigma_{22}-2\sigma_{12}+\mu_1^2+\mu_2^2-2\mu_1\mu_2} \qquad (2.13)$$

To ensure that $\hat{S}(t_1,t_2 \mid \underline{z})$ belongs to the interval [0, 1], we replace $a(t_1,t_2)$ by $\min[1, \max(a(t_1,t_2),0)]$.

In practical situations, we do not know μ_j and σ_{ij}. The simulations reported in Section 2.6 suggest that the bias of $\hat{S}(t_1,t_2 \mid \underline{z})$ is negligible relative to the variance of $\hat{S}(t_1,t_2 \mid \underline{z})$. So we propose the estimate of $a(t_1,t_2)$ by minimizing the average variance of $\hat{S}(t_1,t_2 \mid \underline{z})$ over the data points. This can be achieved by replacing the unknown quantities in the expression of the asymptotic variance of $\hat{S}(t_1,t_2 \mid \underline{z})$ by appropriate estimators. To estimate the variance, one can use the extension of Efron's (1981) bootstrap procedure for one dimensional censored data. Given the data $(T_{1i},T_{2i},\Delta_{1i},\Delta_{2i},\underline{Z}_i)$, $i=1,2,...,n$, where Δ_{ji}, $j=1,2$ is the censoring indicator, generate bootstrap data $(T_{1k}^*,T_{2k}^*,\Delta_{1k}^*,\Delta_{2k}^*,\underline{Z}_k^*)$, $k=1,2,...,n$ from the empirical distribution function

$$\frac{1}{n}\sum_{j=1}^{n} I(T_{1j} \le t_1, T_{2j} \le t_2, \Delta_{1j}=\delta_{1j}, \Delta_{2j}=\delta_{2j}, \underline{Z}_j = \underline{z}_j).$$

For $i=1,2,...,n$ let $Var(S_{1i}(t_1,t_2)^*)$, $Var(S_{2i}(t_1,t_2)^*)$ and $Cov(S_{1i}(t_1,t_2)^*, S_{2i}(t_1,t_2)^*)$ be variance and covariance of the $\hat{S}_1(t_1,t_2)$ and $\hat{S}_2(t_1,t_2)$ in the expression of

$\hat{S}(t_1,t_2)$ given in (2.11), obtained from a large number of resamples. Since the biases are negligible as shown in Section 2.6, we then find the weight $a^*(t_1,t_2)$ as

$$a^*(t_1,t_2)=\frac{\mathrm{var}\left(S_2(t_1,t_2)^*\right)-\mathrm{cov}\left(S_1(t_1,t_2)^*,S_2(t_1,t_2)^*\right)}{\mathrm{var}\left(S_1(t_1,t_2)^*\right)+\mathrm{var}\left(S_2(t_1,t_2)^*\right)-2\mathrm{cov}\left(S_1(t_1,t_2)^*,S_2(t_1,t_2)^*\right)}. \tag{2.14}$$

2.5. Properties of Estimators

In this section, we discuss various properties of estimators. The estimate of the variance of $\hat{\wedge}_{i0}(t_1,t_2)$, $i=1,2$ is obtained in the following way.

From (2.6) the log likelihood function is obtained as

$$l\left(\underline{\beta}_i\right)=\sum_{m=1}^{n}\delta_{im}\left[\underline{\beta}_i{}'\,\underline{z}_{im}-\log\left(\sum_{l=1}^{n_i}Y_{il}\left(t_{im}\right)e^{\underline{\beta}_i{}'\,\underline{z}_{il}}\right)\right]. \tag{2.15}$$

Define, for any $t_i \geq 0$

$$\bar{\mathrm{X}}\left(t_i,\underline{\beta}_i\right)=\frac{\sum_{l=1}^{n_i}Y_{il}\left(t_i\right)e^{\underline{\beta}_i{}'\,\underline{z}_{il}}\,\underline{z}_{il}}{\sum_{l=1}^{n_i}Y_{il}\left(t_i\right)e^{\underline{\beta}_i{}'\,\underline{z}_{il}}} \tag{2.16}$$

and

$$S_i^{(0)}\left(t_{im},\underline{\beta}_i\right)=\sum_{l=1}^{n_i}Y_{il}\left(t_{im}\right)e^{\underline{\beta}_i{}'\,\underline{z}_{il}}\ ,i=1,2. \tag{2.17}$$

Then as in the univariate set up, for fixed t_j, we can prove that $\sqrt{n}\left[\hat{\wedge}_{i0}\left(t_i,t_j\right)-\wedge_{i0}\left(t_i,t_j\right)\right]$ has a limiting normal distribution with mean zero vector and variance $n_i\sigma_i^2$ where σ_i^2 is given by

$$\sigma_i^2=\sum_{m:t_{im}\leq t_i}\frac{\delta_{im}}{\left(S_i^{(0)}\left(t_{im},\underline{\beta}_i\right)\right)^2}+W\left(t_{im}\right)'I\left(\underline{\beta}_i\right)^{-1}W\left(t_{im}\right) \tag{2.18}$$

where,

$$W\left(t_{im}\right)=\sum_{m:t_{im}\leq t_i}\frac{\delta_{im}\,\bar{X}\left(t_{im},\underline{\beta}_i\right)}{S_i^{(0)}\left(t_{im},\underline{\beta}_i\right)},\ i,j=1,2,i\neq j.$$

To study the asymptotic properties of $\hat{\underline{\beta}}_i$, consider the score function,

$$U(\underline{\beta}_i)=\frac{\partial \log L}{\partial \underline{\beta}_i}=\sum_{m=1}^{n}\delta_{im}\left[\underline{z}_{im}-\frac{S_i^{(1)}(\underline{\beta}_i,t_{im})}{S_i^{(0)}(\underline{\beta}_i,t_{im})}\right] \tag{2.19}$$

where,

$$S_i^{(1)}(\underline{\beta}_i,t_{im})=\sum_{l=1}^{n_i}Y_{il}(t_{im})e^{\underline{\beta}'_i\underline{z}_{il}}\underline{z}_{il},\ i=1,2.$$

Then the maximum likelihood estimator $\hat{\underline{\beta}}_i$ is the solution of the score function $U(\underline{\beta}_i)=0$ and hence $\hat{\underline{\beta}}_i$ is a consistent estimator for $\underline{\beta}_i, i=1,2$. For large n_i, the score statistic $U(\underline{\beta}_i)$ is asymptotically r-variate normal with mean zero vector and with covariance matrix $A_i(\underline{\beta}_i)$ where,

$$A_i(\underline{\beta}_i)=\sum_{m=1}^{n}\delta_{im}\left\{\frac{S_i^{(2)}(\underline{\beta}_i,t_{im})}{S_i^{(0)}(\underline{\beta}_i,t_{im})}-\frac{S_i^{(1)}(\underline{\beta}_i,t_{im})S_i^{(1)}(\underline{\beta}_i,t_{im})'}{\left(S_i^{(0)}(\underline{\beta}_i,t_{im})\right)^2}\right\} \tag{2.20}$$

with

$$S_i^{(2)}(\underline{\beta}_i,t_{im})=\sum_{l=1}^{n_i}Y_{il}(t_{im})e^{\underline{\beta}'_i\underline{z}_{il}}\underline{z}_{il}\underline{z}_{il}',\ i=1,2.$$

The covariance matrix can be estimated by substituting the estimate of $\underline{\beta}_i$ in (2.20) for $i=1,2$. Thus $\hat{\underline{\beta}}_i$ is asymptotically r-variate normal with mean vector $\underline{\beta}_i$ and with covariance matrix $A_i^{-1}(\underline{\beta}_i), i=1,2$. The proof of the above asymptotic properties of the estimators follows from Lin (1994).

Asymptotic distribution theory of $\hat{S}(t_1,t_2\mid\underline{z})$ is difficult and the most attractive approach to variance and confidence interval estimation is through resampling method. The bootstrap procedure of resampling the observed data units with replacement, as discussed in Section 2.4, can be employed in such situations.

2.6. Simulation Study

We now carried out a simulation study to evaluate the performance of the aforementioned inference procedures. We consider a bivariate Dirichlet distribution with survival function

$$S(t_1,t_2)=(1-a_1t_1-a_2t_2)^{\alpha+1},\ 0<t_1<\frac{1}{a_1},\ 0<t_2<\frac{1-a_1t_1}{a_2}. \tag{2.21}$$

For (2.21), the bivariate hazard function is given by

$$\lambda_i(t_1,t_2)=\frac{(\alpha+1)a_i}{1-a_1t_1-a_2t_2},\ i=1,2. \tag{2.22}$$

We consider the covariate vector $\underline{z}=(z_1,z_2)$, where covariate z_1 is a binary variable having values 1 and 2 and it is generated with equal probability using Bernoulli distribution. The covariate z_2 is generated from uniform (0, 1). Then paired lifetimes are generated from bivariate Dirichlet distribution for $\alpha=1.5$ and 2 with $a_1=a_2=1$ and for various values of $\underline{\beta}_i$, $i=1,2$. The paired lifetimes are censored independently by a bivariate Dirichlet distribution (2.21) with $\alpha=2$ and $a_1=a_2=1$. We consider $a(t_1,t_2)=0.5$. We used the algorithm given in Gentle (1998, p.111) for generating observations. We estimate the parameter vector and baseline cumulative hazard functions using the procedures given in Sections 2.3 and 2.4. We then compute the estimate of $S(t_1,t_2\mid\underline{z})$ for 1000 simulations and we calculate average bias and variance of the estimate. The bias and variance of the estimator $\hat{S}(t_1,t_2\mid\underline{z})$ for different combinations of n, $\underline{\beta}_1$ and $\underline{\beta}_2$ are given in Table 2.2. From Table 2.2, it follows that the bias of $\hat{S}(t_1,t_2\mid\underline{z})$ is decreasing in n. The variance of the estimate is small irrespective of the sample size.

Next we consider a Gumbel's (1960) bivariate exponential distribution with survival function

$$S(t_1,t_2)=\exp(-t_1-t_2-\lambda t_1t_2),\ t_1,t_2>0,\ 0\le\lambda\le1. \tag{2.23}$$

For (2.23), the bivariate hazard function is

$$\lambda(t_i,t_j)=1+\lambda t_j,\ i,j=1,2,\ i\neq j. \tag{2.24}$$

The covariates are generated as in the case of bivariate Dirichlet distribution. We generated observations from Gumbel's bivariate exponential distribution for λ=0.75 and λ=0.9 and for various values of $\underline{\beta}_i$, $i=1,2$, using algorithm given in Devroye (1986,p.584). The paired lifetimes are censored by a Gumbel's bivariate exponential distribution (2.23) with λ=0.1. We consider $a(t_1,t_2)=0.5$. We estimate

the parameter vector and baseline cumulative hazard functions using the procedures given in Sections 2.3 and 2.4. We then compute the estimate of $S(t_1,t_2 \mid \underline{z})$ for 1000 simulations and we calculate average bias and variance of the estimate. The bias and variance of the estimator $\hat{S}(t_1,t_2 \mid \underline{z})$ for different combinations of n, $\underline{\beta}_1$ and $\underline{\beta}_2$ are given in Table 2.3, which shows that the bias of $\hat{S}(t_1,t_2 \mid \underline{z})$ is decreasing in n. The variance of the estimate is small irrespective of the sample size.

Table 2.2. Bias and variance of the estimator $\hat{S}(t_1,t_2 \mid \underline{z})$ for bivariate Dirichlet distribution at different time points

n	$\underline{\beta}_1$	$\underline{\beta}_2$	(t_1,t_2)	Bias		Variance	
				α=1.5	α=2	α=1.5	α=2
25	(-1, 0.005)	(-0.997, -0.002)	(0.13, 0.12)	0.0622	0.0662	0.0326	0.0850
			(0.05, 0.14)	0.0398	0.0442	0.0818	0.0495
			(0.02, 0.09)	-0.0680	0.0690	0.0253	0.0490
	(0.17, -0.8)	(0.32, -1)	(0.13, 0.12)	0.0704	0.0152	0.0790	0.0247
			(0.05, 0.14)	0.0232	0.0318	0.0650	0.0399
			(0.02, 0.09)	-0.0317	0.0630	0.0734	0.0364
	(-0.03, -0.4)	(-0.8, 2.68)	(0.13, 0.12)	0.0376	0.0702	0.0508	0.0820
			(0.05, 0.14)	0.0816	0.0463	0.0474	0.0530
			(0.02, 0.09)	0.0440	0.0688	0.0560	0.0610
50	(-1, 0.005)	(-0.997, -0.002)	(0.13, 0.12)	0.0283	0.0259	0.0282	0.0287
			(0.05, 0.14)	0.0209	-0.0009	0.0084	0.0398
			(0.02, 0.09)	0.0249	0.0491	0.0202	0.0325
	(0.17, -0.8)	(0.32, -1)	(0.13, 0.12)	0.0666	-0.0097	0.0525	0.0148
			(0.05, 0.14)	0.0184	0.0265	0.0359	0.0362
			(0.02, 0.09)	-0.0179	-0.0420	0.0274	0.0241
	(-0.03, -0.4)	(-0.8, 2.68)	(0.13, 0.12)	0.0285	0.0213	0.0445	0.0365
			(0.05, 0.14)	0.0498	-0.0320	0.0295	0.0367
			(0.02, 0.09)	0.0370	-0.0400	0.0193	0.0233
100	(-1, 0.005)	(-0.997, -0.002)	(0.13, 0.12)	0.0160	-0.0160	0.0054	0.0112
			(0.05, 0.14)	-0.0002	0.0008	0.0058	0.0231
			(0.02, 0.09)	0.0060	-0.0060	0.0073	0.0222
	(0.17, -0.8)	(0.32, -1)	(0.13, 0.12)	0.0232	-0.0073	0.0464	0.0135
			(0.05, 0.14)	0.0067	0.0170	0.0296	0.0321
			(0.02, 0.09)	0.0054	-0.0350	0.0119	0.0155
	(-0.03, -0.4)	(-0.8, 2.68)	(0.13, 0.12)	0.0009	0.0113	0.0202	0.0350
			(0.05, 0.14)	-0.0060	0.0131	0.0189	0.0134
			(0.02, 0.09)	-0.0187	0.0188	0.0009	0.0172

Table 2.3. Bias and variance of the estimator $\hat{S}(t_1, t_2 \mid \underline{z})$ for bivariate Gumbel's exponential distribution at different time points

n	$\underline{\beta}_1$	$\underline{\beta}_2$	(t_1, t_2)	Bias		Variance	
				λ=0.75	λ=0.9	λ=0.75	λ=0.9
25	(-0.2, 0.1)	(-1.3, 2.3)	(2, 1)	0.0230	0.0285	0.0184	0.0280
			(1, 3)	0.0276	-0.0520	0.0274	0.0387
			(2, 2)	0.0380	0.0367	0.0820	0.0490
	(0.25, -0.4)	(-0.5, 0.6)	(2, 1)	0.0429	-0.0145	0.0560	0.0466
			(1, 3)	0.0870	0.0340	0.0704	0.0219
			(2, 2)	-0.0870	0.0460	0.0453	0.0276
	(-0.28, 0.9)	(-0.31, 1)	(2, 1)	0.0594	-0.0610	0.0248	0.0378
			(1, 3)	0.0200	0.0212	0.0258	0.0389
			(2, 2)	0.0207	0.0161	0.0260	0.0151
50	(-0.2, 0.1)	(-1.3, 2.3)	(2, 1)	0.0186	0.0111	0.0176	0.0096
			(1, 3)	0.0192	0.0430	0.0242	0.0353
			(2, 2)	0.0322	-0.0159	0.0113	0.0409
	(0.25, -0.4)	(-0.5, 0.6)	(2, 1)	0.0215	-0.0135	0.0114	0.0273
			(1, 3)	0.0182	-0.0180	0.0338	0.0212
			(2, 2)	0.0177	-0.0310	0.0126	0.0159
	(-0.28, 0.9)	(-0.31, 1)	(2, 1)	0.0526	0.0051	0.0049	0.0201
			(1, 3)	0.0180	0.0199	0.0239	0.0164
			(2, 2)	0.0183	0.0054	0.0119	0.0092
100	(-0.2, 0.1)	(-1.3, 2.3)	(2, 1)	0.0172	-0.0049	0.0170	0.0090
			(1, 3)	-0.0160	-0.0029	0.0121	0.0248
			(2, 2)	0.0278	-0.0106	0.0037	0.0085
	(0.25, -0.4)	(-0.5, 0.6)	(2, 1)	0.0135	-0.0051	0.0009	0.0143
			(1, 3)	-0.0126	-0.0117	0.0264	0.0167
			(2, 2)	-0.0003	0.0041	0.0055	0.0029
	(-0.28, 0.9)	(-0.31, 1)	(2, 1)	0.0209	0.0015	0.0003	0.0179
			(1, 3)	-0.0102	-0.0058	0.0163	0.0080
			(2, 2)	0.0062	-0.0027	0.0088	0.0085

2.7. Data Analysis

For the illustration of the estimation procedure we consider two real life examples.

First, we consider a data on Kalbfleish and Prentice (2002, page 329). The survival times of closely and poorly HLA (human lymphocyte antigen) matched skin grafts on the same burned individual is considered as T_1 and T_2. Amount of burn is considered as a covariate. The data is given in Table 2.4.

Table 2.4. Days of survival of closely and poorly matched grafts on the same person

Case number	4	5	7	8	9	10	11	12	13	15	16
Survival of close match graft	37	19	57^+	93	16	21^+	20	18	63	29	60^+
Survival of poor match graft	29	13	15	26	11	15^+	26	19^+	43	15	38^+
Amount of burn	30	20	25	45	20	18	35	25	50	30	30

+ indicates censoring time.

We compute the estimates $\hat{\beta}_i, i=1,2$ for observed pair of lifetimes, those are given in Table 2.5.

Table 2.5. Estimates of β_1 and β_2

(t_1,t_2)	$(t_1,0)$ $\hat{\beta}_1$	$(0,t_2)$ $\hat{\beta}_2$	(t_1 given t_2) $\hat{\beta}_1$	(t_2 given t_1) $\hat{\beta}_2$
(16, 11)	-0.0777	-0.0706	-0.0777	-0.0706
(19, 13)	-0.0777	-0.0706	-0.0612	-0.0580
(20, 26)	-0.0777	-0.0706	-0.0714	-0.0361
(29, 15)	-0.0777	-0.0706	-0.0411	-0.0562

We then calculate $\hat{\Lambda}_{10}(t_1,t_2)$, $\hat{\Lambda}_{10}(t_1,0)$, $\hat{\Lambda}_{20}(t_1,t_2)$ and $\hat{\Lambda}_{20}(0,t_2)$ as discussed in Section 2.4. The estimates of $\Lambda_{i0}(t_i,t_j)$, $i,j=1,2$, $i\neq j$ are given in the Table 2.6. Based on bootstrap procedure explained in Section 2.4, we obtain $a(t_1,t_2)=0.5$. Finally, Table 2.7 provides the estimates of $S(t_1,t_2 \mid \underline{z})$

Table 2.6. Estimate of the baseline cumulative hazard function

(t_1,t_2)	$\hat{\Lambda}_{10}(t_1,t_2)$	$\hat{\Lambda}_{10}(t_1,0)$	$\hat{\Lambda}_{20}(t_1,t_2)$	$\hat{\Lambda}_{20}(0,t_2)$
(16, 11)	0.7322	0.7322	0.6152	0.6152
(19, 23)	1.2118	2.5881	0.5936	1.3389
(20, 26)	2.5938	3.8393	2.2886	8.1513
(29, 25)	1.4926	5.8962	2.1224	3.0959

Table 2.7. Estimate of the survival function

(t_1,t_2)	$\hat{S}_1(t_1,t_2 \mid \underline{z})$	$\hat{S}_2(t_1,t_2 \mid \underline{z})$	$\hat{S}(t_1,t_2 \mid \underline{z})$
(16, 11)	0.7373	0.7373	0.7373
(19, 23)	0.5054	0.4803	0.4928
(20, 26)	0.4061	0.4067	0.4064
(29, 25)	0.4461	0.3801	0.4131

From Table 2.5, it is clear that $\hat{\beta}_i$'s have negative values. From Table 2.7, it follows that, as the survival of closely and poorly matched grafts increases, the survival time of the corresponding individual's decreases. As amount of burn increases probability of survival decreases, as expected. Figure 2.1 shows the estimates of the bivariate survival function.

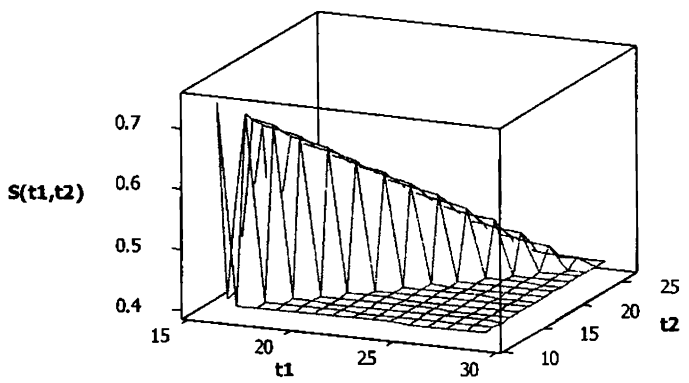

Figure 2.1. Estimates of bivariate survival function

Secondly, we consider the data on Australian Twin Study (Duffy et al., 1990), explained in Section 2.2. The data contains the information of 3808 complete pairs and from that we take only the first 250 pairs for the illustration purpose. T_1 and T_2 represents the pair of individual's age at the time of surgery to appendicectomy undergone. Zygosity is considered as covariate. The estimate of $\underline{\beta}_1$

and $\underline{\beta}_2$ are given in Table 2.8 and the estimates of baseline cumulative hazard function is given in Table 2.9. Since $\hat{S}_1(t_1,t_2 \mid \underline{z})$ and $\hat{S}_2(t_1,t_2 \mid \underline{z})$ are not very different; we take $a(t_1,t_2)=0.5$ and the estimates are given in Table 2.10.

Table 2.8. Estimates of $\underline{\beta}_1$ and $\underline{\beta}_2$

(t_1,t_2)	(z_1,z_2)	$(t_1,0)$ $\hat{\underline{\beta}}_1$	$(0,t_2)$ $\hat{\underline{\beta}}_2$	(t_1 given t_2) $\hat{\underline{\beta}}_1$	(t_2 given t_1) $\hat{\underline{\beta}}_2$
(5, 7)	(1, 1)	-0.0424	0.0229	-0.0424	0.0204
(9, 7)	(1, 1)	-0.0424	0.0229	-0.0424	0.0341
(12, 12)	(6, 6)	-0.0424	0.0229	-0.0644	0.0319
(17, 9)	(5, 5)	-0.0424	0.0229	-0.0555	0.0325
(8, 24)	(1, 1)	-0.0424	0.0229	-0.1172	0.0324
(21, 11)	(1, 1)	-0.0424	0.0229	-0.0694	0.0321
(22, 29)	(1, 1)	-0.0424	0.0229	-0.1495	0.0631
(18, 41)	(5, 5)	-0.0424	0.0229	-0.0731	0.0382
(4, 49)	(3, 3)	-0.0424	0.0229	-0.1393	0.0229
(24, 52)	(4, 4)	-0.0424	0.0229	-0.2115	0.0876

Table 2.9. Estimate of the baseline cumulative hazard function

(t_1,t_2)	(z_1,z_2)	$\hat{\Lambda}_{10}(t_1,t_2)$	$\hat{\Lambda}_{10}(t_1,0)$	$\hat{\Lambda}_{20}(t_1,t_2)$	$\hat{\Lambda}_{20}(0,t_2)$
(5, 7)	(1, 1)	0.0136	0.0135	0.0269	0.0265
(9, 7)	(1, 1)	0.0459	0.455	0.0227	0.0265
(12, 12)	(6, 6)	0.0779	0.0879	0.0522	0.0658
(17, 9)	(5, 5)	0.1755	0.1775	0.0386	0.0499
(8, 24)	(1, 1)	0.0413	0.0362	0.2097	0.2144
(21, 11)	(1, 1)	0.2334	0.2319	0.0487	0.0572
(22, 29)	(1, 1)	0.2154	0.2566	0.2287	0.3092
(18, 41)	(5, 5)	0.1378	0.1984	0.4277	0.4912
(4, 49)	(3, 3)	0.0334	0.0089	0.6567	0.6559
(24, 52)	(4, 4)	0.3699	0.3257	0.4876	0.6832

Table 2.10. Estimate of the survival function

(t_1, t_2)	(z_1, z_2)	$\hat{S}_1(t_1, t_2 \mid \underline{z})$	$\hat{S}_2(t_1, t_2 \mid \underline{z})$	$\hat{S}(t_1, t_2 \mid \underline{z})$
(5, 7)	(1, 1)	0. 9607	0.9605	0.9606
(9, 7)	(1, 1)	0.9314	0.9351	0.9333
(12, 12)	(6, 6)	0.8794	0.8768	0.8781
(17, 9)	(5, 5)	0.8279	0.8278	0.8278
(8, 24)	(1, 1)	0.7739	0.7777	0.7759
(21, 11)	(1, 1)	0.7581	0.7614	0.7598
(22, 29)	(1, 1)	0.6054	0.6129	0.6091
(18, 41)	(5, 5)	0.5239	0.5076	0.5157
(4, 49)	(3, 3)	0.4845	0.4910	0.4878
(24, 52)	(4, 4)	0.4035	0.3802	0.3919

From Table 2.8, it is clear that $\hat{\beta}_1$'s have negative values and $\hat{\beta}_2$'s have positive values. It follows from Table 2.10 that as the time for undergoing surgery to appendectomy increases, the survival probability decreases. Figure 2.2 shows the estimates of the bivariate survival function.

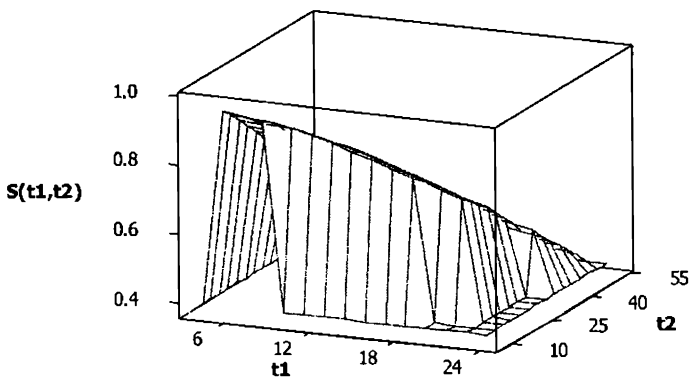

Figure 2.2. Estimates of bivariate survival function

2.8. Conclusion

We developed a bivariate proportional hazards model using vector hazard function of Johnson and Kotz (1975). Since the covariates under study have different effect on two components of the vector hazard function, proposed model will be more useful to study the dependence of lifetime on covariates. The estimators of the parameters as well as the baseline hazard function are developed. Simulation studies showed that the performance of the estimator is good. We illustrated the procedure using two real life data.

The univariate proportional hazards model can be directly deduced as a particular case. Further, the model can be extended to the multivariate set up using multivariate version of vector hazard function of Johnson and Kotz (1975). When $\hat{\underline{\beta}}_1 = \hat{\underline{\beta}}_2 = 0$, the estimator of the bivariate cumulative baseline hazard function is the extension of the well-known Nelson-Aalen estimator of the hazard function in the univariate set up.

Chapter Three

Proportional Hazards Model for Gap Time Distributions of Recurrent Events

3.1. Introduction

As mentioned earlier, in many survival studies, the investigators are more interested in the analysis of gap time than the total time. In these studies, investigators are often interested in the distribution of gap times and how this distribution depends on important predictor variables. The analysis of gap time data is usually done by assuming proportional hazards model for marginal hazard functions. In the present study, we deal with the regression problem for gap time of recurrent events in which the marginal and conditional hazard functions depend on certain covariates. Since the covariates under study have different effect on marginal and conditional hazard functions, proposed model will be more useful to study the dependence of gap times on covariates.

In Section 3.2, we consider bivariate proportional hazards model for gap times. Estimation of parameter vector and baseline hazard functions is discussed in Section 3.3. Asymptotic properties of estimators are also studied. In Section 3.4, we carried out a simulation study to investigate the finite sample properties of the estimators and their robustness. In Section 3.5, we apply the new model to a real life data. Finally, we conclude our study in Section 3.6.

3.2. The Model

Suppose that an individual may experience k consecutive events at times $X_1 < X_2 < ... < X_k$ which are measured from the start of the follow up. We are interested in gap times $T_1 = X_1$, $T_2 = X_2 - X_1$ and $T_k = X_k - X_{k-1}$. We assume that

---

The results in this Chapter have been communicated as entitled "Proportional Hazards Model for Gap Time Distributions of Recurrence Events" (see Sankaran and Sreeja (2008)).

the follow up time is subject to independent right censoring by C which implied that $(X_1, X_2, ...X_k)$ are independent of C. On the other hand, the gap time T_i is subject to right censoring by $C - X_{i-1}, i = 2,3,...,k$, which is naturally correlated with T_i unless T_i is independent of X_{i-1}. We now consider the regression problem in which the marginal and conditional hazard functions of $(T_1, T_2, ...T_k)$ depend on certain covariates. We confine our study for $k = 2$. The extension to higher dimensions is direct.

Suppose that $S_i(t_i) = P[T_i \geq t_i]$ is the marginal recurrence survival function of $T_i, i = 1,2$. Let $S(t_1, t_2) = P[T_1 \geq t_1, T_2 \geq t_2]$ be the joint recurrence survival function of T_1 and T_2. Our objective is to estimate $S(t_1, t_2)$ in the presence of covariates. For this, one possible method is to consider marginal hazard functions of T_1 and T_2 and then apply ideas from generalized estimating function to calculate an appropriate combination of the two marginal estimates. This can be done in the case of homogeneity of the two regression coefficients. Another technique, one could use, is to model T_1 and then consider the conditional distribution of T_2 given $T_1 = t_1$. From Wang and Wells (1998), the survival function $S(t_1, t_2)$ is given by

$$S(t_1, t_2) = -\int_{u > t_1} P[T_2 > t_2 \mid T_1 = t_1]\, S_1(u)du$$

$$= -\int_{u > t_1} \prod_{u} (1 - \lambda^*(u, t_2)) S_1(u)du \qquad (3.1)$$

where $\lambda^*(t_1, t_2)$ is the hazard function of T_2 given $T_1 = t_1$. When X is continuous, estimating $\lambda^*(t_1, t_2)$ requires special smoothing techniques and can be very complicated when the dependent censoring condition is taken into account (see Wang and Wells, 1998).

In the following, we consider a simple method for the analysis using marginal hazard function of T_1 and conditional hazard function of T_2 given $T_1 \geq t_1$.

Let $\lambda_1(t_1)$ be the hazard function of T_1, which is defined as

$$\lambda_1(t_1) = \lim_{\Delta t_1 \to 0} \frac{1}{\Delta t_1} P[t_1 \leq T_1 < t_1 + \Delta t_1 \mid T_1 \geq t_1]. \qquad (3.2)$$

$\lambda_1(t_1)$ is nothing but the instantaneous rate of occurrence of the first event at time t_1 given that he was alive at the time $T_1 \geq t_1$.

Since T_2 depends directly on T_1, we consider the conditional hazard function of T_2 given $T_1 \geq t_1$, which is defined as

$$\lambda_2(t_1,t_2) = \lim_{\Delta t_2 \to 0} \frac{1}{\Delta t_2} P[t_2 \leq T_2 < t_2 + \Delta t_2 \mid T_1 \geq t_1, T_2 \geq t_2]. \tag{3.3}$$

The meaning of $\lambda_2(t_1,t_2)$ is instantaneous rate of occurrence of the second event at time t_2 given that $T_1 \geq t_1$ and $T_2 \geq t_2$.

The cumulative hazard functions respectively are denoted by $\wedge_1(t_1)$ and $\wedge_2(t_1,t_2)$ where,

$$\wedge_1(t_1) = \int_0^{t_1} \lambda_1(u)\, du \tag{3.4}$$

and

$$\wedge_2(t_1,t_2) = \int_0^{t_2} \lambda_2(t_1,u)\, du \tag{3.5}$$

The survival function can be written as

$$S(t_1,t_2) = \exp[-\wedge_1(t_1) - \wedge_2(t_1,t_2)] \tag{3.6}$$

Now we define proportional hazards model for (T_1,T_2) as

$$\lambda_1(t_1 \mid \underline{z}) = \lambda_{10}(t_1) e^{\underline{\beta}_1' \underline{z}} \tag{3.7}$$

and

$$\lambda_2(t_1,t_2 \mid \underline{z}) = \lambda_{20}(t_1,t_2) e^{\underline{\beta}_2' \underline{z}} \tag{3.8}$$

where $\underline{z} = (z_1, z_2, ..., z_r)'$ is a vector of covariates, $\underline{\beta}_1'$ and $\underline{\beta}_2'$ are r-component parameter vectors, independent of both t_1 and t_2 and $\lambda_{10}(t_1)$ and $\lambda_{20}(t_1,t_2)$ are baseline hazard functions. The model (3.7) means that the ratio $\frac{\lambda_1(t_1 \mid \underline{z}^{(1)})}{\lambda_1(t_1 \mid \underline{z}^{(2)})}$ of the hazard functions of two individuals with covariate vectors $\underline{z}^{(1)}$ and $\underline{z}^{(2)}$ does not vary with t_1 and the model (3.8) means that the ratio $\frac{\lambda_2(t_1,t_2 \mid \underline{z}^{(1)})}{\lambda_2(t_1,t_2 \mid \underline{z}^{(2)})}$ of the hazard

functions of two individuals with covariate vectors $\underline{z}^{(1)}$ and $\underline{z}^{(2)}$ does not vary with both t_1 and t_2.

From (3.7), it follows that the marginal hazard functions of T_1 for two-individuals are proportional to one another. The model (3.8) implies that the ratio of conditional hazard functions of T_2 given $T_1 \geq t_1$ for two individuals are independent of both t_1 and t_2.

In many situations, one may be interested in the joint survival function $S(t_1, t_2 \mid \underline{z})$ of gap times. Using (3.6), (3.7) and (3.8), the survival function can be obtained as

$$S(t_1, t_2 \mid \underline{z}) = \exp[-\wedge_{10}(t_1)e^{\underline{\beta}_1'\underline{z}} - \wedge_{20}(t_1, t_2)e^{\underline{\beta}_2'\underline{z}}] \tag{3.9}$$

3.3. Inference Procedures

Suppose now that there are n independent subjects in the study so that $(T_{1i}, T_{2i}, C_i, \underline{z}_i)$, $i = 1, 2, ..., n$ are n independent replicates of $(T_1, T_2, C, \underline{z})$ where $T_1 = X_1$ and $T_2 = X_2 - X_1$. In the presence of censoring, the observable data consists of $(\tilde{X}_{1i}, \tilde{X}_{2i}, \delta_{1i}, \delta_{2i}, \underline{z}_i)$, where $\tilde{X}_{1i} = \min(T_{1i}, C_i)$, $\delta_{1i} = I(T_{1i} < C_i)$, $\tilde{X}_{2i} = \min(T_{2i}, C_i - T_{1i})$, $\delta_{2i} = I(T_{2i} < C_i - T_{1i})$ and $\underline{z}_i$ is the covariate vector for $i = 1, 2, ..., n$ and $j = 1, 2$ with $I(.)$ as indicator function. First, we consider the estimation of regression parameters $\underline{\beta}_1$ and $\underline{\beta}_2$.

The counting process $N_1(t_1) = \{N_{1i}(t_1), t_1 \geq 0\}$ given at time t_1 by $N_1(t_1) = I(\tilde{X}_{1i} \leq t_1, \delta_1 = 1)$ where $\tilde{X}_1 = \min(T_1, C)$ and $\delta_1 = I(T_1 < C)$. For fixed t_1, we also define a counting process $\{N_2(t_1, t_2), t_1 \geq 0, t_2 \geq 0\}$, given at time t_2 by $N_2(t_1, t_2) = I(\tilde{X}_1 \geq t_1, \tilde{X}_2 \leq t_2, \delta_2 = 1)$ where $\tilde{X}_2 = \min(T_2, C - T_1)$ and $\delta_2 = I(T_2 < C - T_1)$. Then we have, for $i = 1, 2, ..., n$,

$$N_{1i}(t_1) = I(\tilde{X}_{1i} \leq t_1)\delta_{1i} \tag{3.10}$$

and

$$N_{2i}(t_1, t_2) = I(\tilde{X}_{1i} \geq t_1, \tilde{X}_{2i} \leq t_2)\delta_{2i}. \tag{3.11}$$

Consider the at-risk processes $Y_1(t_1) = \{Y_{1i}(t_1), t_1 \geq 0\}$ and

$Y_2(t_1,t_2)=\{Y_{2i}(t_1,t_2),t_1\geq 0,t_2\geq 0\}$, where

$$Y_{1i}(t_1)=I(X_{1i}\geq t_1) \tag{3.12}$$

and

$$Y_{2i}(t_1,t_2)=I(X_{1i}\geq t_1,X_{2i}\geq t_2). \tag{3.13}$$

Then we can write

$$E\left[N_{1i}(dt_1)\mid \mathbb{F}_{t_1^-}\right]=Y_{1i}(t_1)\wedge_{1i}(dt_1) \tag{3.14}$$

and for fixed t_1,

$$E\left[N_{2i}(t_1,dt_2)\mid \mathbb{F}_{t_1^-,t_2^-}\right]=Y_{2i}(t_1,t_2)\wedge_{2i}(t_1,dt_2) \tag{3.15}$$

where $\mathbb{F}_{t_1}$ belong to the right-continuous filtration $\{\mathbb{F}_{t_1}:t_1\geq 0\}$ and for fixed t_1, $\mathbb{F}_{t_1,t_2}$ belong to the right-continuous filtration $\{\mathbb{F}_{t_1,t_2}:t_1\geq 0,t_2\geq 0\}$ with $\mathbb{F}_{t_1}$ and $\mathbb{F}_{t_1,t_2}$ are defined by

$$\mathbb{F}_{t_1}=\sigma\{N_{1i}(u),Y_{1i}(u+),\underline{z}_i:0\leq u\leq t_1,i=1,2,...,n\} \tag{3.16}$$

and

$$\mathbb{F}_{t_1,t_2}=\sigma\{N_{2i}(t_1,v),Y_{2i}(t_1,v+),\underline{z}_i:0\leq v\leq t_2,i=1,2,...,n\}. \tag{3.17}$$

Denoting $M_{1i}(t_1)=N_{1i}(t_1)-\int_0^{t_1}Y_{1i}(s)\wedge_{1i}(ds)$ (3.18)

and

$$M_{2i}(t_1,t_2)=N_{2i}(t_1,t_2)-\int_0^{t_2}Y_{2i}(t_1,s)\wedge_{2i}(t_1,ds),\ i=1,2,...,n, \tag{3.19}$$

$M_{1i}(t_1)$ is zero mean $\mathbb{F}_{t_1}$ martingale and for fixed t_1, $M_{2i}(t_1,t_2)$ is zero mean $\mathbb{F}_{t_1,t_2}$ martingale.

Then the score function of $\underline{\beta}_1$ is given by

$$U(\underline{\beta}_1)=\sum_{i=1}^{n}\delta_{1i}\left[\underline{z}_{1i}-\frac{S_1^{(1)}(\underline{\beta}_1,t_{1i})}{S_1^{(0)}(\underline{\beta}_1,t_{1i})}\right] \tag{3.20}$$

where,

$$S_1^{(0)}(t_{1i},\underline{\beta}_1)=\sum_{l=1}^{n_1}Y_{1l}(t_{1i})e^{\underline{\beta}'_1\underline{z}_{1l}}$$

and

$$S_1^{(1)}(\underline{\beta}_1, t_{1i}) = \sum_{l=1}^{n_1} Y_{1l}(t_{1i}) e^{\underline{\beta}_1' \underline{z}_{1l}} \underline{z}_{1l}.$$

The maximum likelihood estimator of $\underline{\beta}_1$ is the solution of the score function $U(\underline{\beta}_1) = 0$.

To obtain the estimate of $\underline{\beta}_2$ consider the score function

$$U(\underline{\beta}_2) = \sum_{i=1}^{n} \delta_{2i} \left[\underline{z}_i - \frac{S_2^{(1)}(t_{1i}, t_{2i}, \underline{\beta}_2)}{S_2^{(0)}(t_{1i}, t_{2i}, \underline{\beta}_2)} \right] \tag{3.21}$$

where,

$$S_2^{(0)}(t_{1i}, t_{2i}, \underline{\beta}_2) = \sum_{l=1}^{n_2} Y_{2l}(t_{1i}, t_{2i}) e^{\underline{\beta}_2' \underline{z}_l}$$

and

$$S_2^{(1)}(t_{1i}, t_{2i}, \underline{\beta}_2) = \sum_{l=1}^{n_2} Y_{2l}(t_{1i}, t_{2i}) e^{\underline{\beta}_2' \underline{z}_l} \underline{z}_l.$$

As in the univariate set up, we can obtain the estimate of $\underline{\beta}_2$ by maximizing the score function (3.21).

Now we discuss the estimation of baseline cumulative hazard functions $\wedge_{10}(t_1)$ and $\wedge_{20}(t_1, t_2)$. From Lawless (2003), the estimator of $\wedge_{10}(t_1)$ is given by

$$\hat{\wedge}_{10}(t_1) = \int_0^{t_1} \frac{I(Y_1(s) > 0)}{S_1^{(0)}(s, \hat{\underline{\beta}}_1)} dN_1(s), \tag{3.22}$$

which can be written as

$$\hat{\wedge}_{10}(t_1) = \sum_{i: t_{1i} \le t_1} \left(\frac{\delta_{1i}}{\sum_{l=1}^{n} Y_{1l}(t_{1i}) e^{\hat{\underline{\beta}}_1' \underline{z}_l}} \right) \tag{3.23}$$

Similarly, by using the counting process approach, the estimator of $\wedge_{20}(t_1, t_2)$ is obtained as

$$\hat{\Lambda}_{20}(t_1,t_2)=\int_0^{t_2}\frac{I(Y_2(t_1,s)>0)}{S_2^{(0)}(t_1,s,\hat{\underline{\beta}}_2)}dN_2(t_1,s) \tag{3.24}$$

which reduces to

$$\hat{\Lambda}_{20}(t_1,t_2)=\sum_{i:t_{2i}\le t_2}\left(\frac{\delta_{2i}}{\sum_{l=1}^{n_2}Y_{2l}(t_{1i},t_{2i})e^{\hat{\underline{\beta}}_2'\underline{z}_l}}\right). \tag{3.25}$$

Now we discuss the asymptotic properties of $\underline{\beta}_1$ and $\underline{\beta}_2$. The asymptotic properties of $\hat{\underline{\beta}}_1$ are well studied in literature (see Lawless, 2003, p.342). Precisely, $\hat{\underline{\beta}}_1$ is asymptotically r-variate normal with the mean vector $\underline{\beta}_1$ and covariance matrix $A_1^{-1}(\underline{\beta}_1)$ where

$$A_1(\underline{\beta}_1)=\frac{1}{n}\sum_{i=1}^{n}\delta_{1i}\left\{\frac{S_1^{(2)}(t_{1i},\underline{\beta}_1)}{S_1^{(0)}(t_{1i},\underline{\beta}_1)}-\frac{S_1^{(1)}(t_{1i},\underline{\beta}_1)S_1^{(1)}(t_{1i},\underline{\beta}_1)'}{\left(S_1^{(0)}(t_{1i},\underline{\beta}_1)\right)^2}\right\} \tag{3.26}$$

with

$$S_1^{(2)}(t_{1i},\underline{\beta}_1)=\sum_{l=1}^{n}Y_{1l}(t_{1i})e^{\underline{\beta}_1'\underline{z}_l}\underline{z}_l\underline{z}_l'.$$

For fixed t_1, the maximum likelihood estimator $\hat{\underline{\beta}}_2$ is the solution of the score function $U(\underline{\beta}_2)=0$ and hence $\hat{\underline{\beta}}_2$ is a consistent estimator for $\underline{\beta}_2$. When t_1 is fixed, the score statistic $U(\underline{\beta}_2)$ is asymptotically r-variate normal with mean zero vector and with covariance matrix $A_2(\underline{\beta}_2)$ where,

$$A_2(\underline{\beta}_2)=\frac{1}{n_2}\sum_{i=1}^{n_2}\delta_{2i}\left\{\frac{S_2^{(2)}(t_{1i},t_{2i},\underline{\beta}_2)}{S_2^{(0)}(t_{1i},t_{2i},\underline{\beta}_2)}-\frac{S_2^{(1)}(t_{1i},t_{2i},\underline{\beta}_2)S_2^{(1)}(t_{1i},t_{2i},\underline{\beta}_2)'}{\left(S_2^{(0)}(t_{1i},t_{2i},\underline{\beta}_2)\right)^2}\right\} \tag{3.27}$$

with n_2 as the number of observed occurrence of the second event and

$$S_2^{(2)}(t_{1i},t_{2i},\underline{\beta}_2)=\sum_{l=1}^{n_2}Y_{2l}(t_{1i},t_{2i})e^{\underline{\beta}_2'\underline{z}_l}\underline{z}_l\underline{z}_l'.$$

Thus $\hat{\underline{\beta}}_2$ is asymptotically r-variate normal with mean vector $\underline{\beta}_2$ and covariance matrix $A_2^{-1}(\underline{\beta}_2)$.

The asymptotic properties of $\hat{\wedge}_{10}(t_1)$ are well discussed in literature (see Lawless, 2003, p. 353). Under the same set of regularity conditions, as required for the asymptotic normality of $\hat{\underline{\beta}}_2$, the $(\hat{\wedge}_{20}(t_1,t_2)-\wedge_{20}(t_1,t_2))$ converges weakly to a mean zero Gaussian process. In particular, for fixed t_1, the variance of $\hat{\wedge}_{20}(t_1,t_2)$ can be consistently estimated by

$$\hat{\text{var}}\left[\hat{\wedge}_{20}(t_1,t_2)\right]=\sum_{i:t_{2i}\le t_2}\frac{\delta_{2i}}{\left(S_2^{(0)}\left(t_{1i},t_{2i},\hat{\underline{\beta}}_2\right)\right)^2}+\hat{W}_2(t_1,t_2)'A_2\left(\hat{\underline{\beta}}_2\right)^{-1}\hat{W}_2(t_1,t_2) \tag{3.28}$$

where,

$$\hat{W}_2(t_1,t_2)=\sum_{i:t_{2i}\le t_2}\frac{\delta_{2i}\bar{X}_2\left(t_{1i},t_{2i},\hat{\underline{\beta}}_2\right)}{S_2^{(0)}\left(t_{1i},t_{2i},\hat{\underline{\beta}}_2\right)}$$

with

$$\bar{\mathrm{X}}_2\left(t_{1i},t_{2i},\underline{\beta}_2\right)=\frac{\sum_{l=1}^{n_2}Y_{2l}(t_{1i},t_{2i})e^{\underline{\beta}_2'\underline{z}_l}\underline{z}_l}{\sum_{l=1}^{n_2}Y_{2l}(t_{1i},t_{2i})e^{\underline{\beta}_2'\underline{z}_l}}.$$

One may often be interested in estimating the survival function $S(t_1,t_2\mid\underline{z}_0)$ of gap times with a fixed covariate $\underline{z}_0$. From (3.9), a natural estimate for $S(t_1,t_2\mid\underline{z}_0)$ is given by

$$\hat{S}(t_1,t_2\mid\underline{z})=\exp[-\hat{\wedge}_{10}(t_1)e^{\hat{\underline{\beta}}_1'\underline{z}_0}-\hat{\wedge}_{20}(t_1,t_2)e^{\hat{\underline{\beta}}_2'\underline{z}_0}]. \tag{3.29}$$

For fixed t_1, t_2 and $\underline{z}_0$, $\{\hat{\wedge}_{10}(t_1),\hat{\wedge}_{20}(t_1,t_2)\}$ asymptotically follows a bivariate normal distribution with covariance matrix

$$W(t_1,t_2)=\begin{bmatrix}W_1(t_1) & W_{12}(t_1,t_2)\\ W_{12}(t_1,t_2) & W_2(t_1,t_2)\end{bmatrix} \tag{3.30}$$

where

$$W_1(t_1) = \sum_{i:t_{1i} \le t_1} \frac{\delta_{1i} \bar{X}_1(t_{1i}, \underline{\beta}_1)}{S_1^{(0)}(t_{1i}, \underline{\beta}_1)}$$

with

$$\bar{X}_1(t_{1i}, \underline{\beta}_1) = \frac{\sum_{l=1}^{n} Y_{1l}(t_{1i}) e^{\underline{\beta}_1' \underline{z}_l} \underline{z}_l}{\sum_{l=1}^{n} Y_{1l}(t_{1i}) e^{\underline{\beta}_1' \underline{z}_l}},$$

$$W_2(t_1, t_2) = \sum_{i:t_{2i} \le t_2} \frac{\delta_{2i} \bar{X}_2(t_{1i}, t_{2i}, \underline{\beta}_2)}{S_2^{(0)}(t_{1i}, t_{2i}, \underline{\beta}_2)}$$

and

$$W_{12}(t_1, t_2) = E\left[\int_0^{t_1}\int_0^{t_2} \frac{I(Y_1(s) > 0)}{Y_1(s)} \frac{I(Y_2(t_1, u) > 0)}{Y_2(t_1, u)} dN_1(s) dN_2(t_1, u)\right].$$

A straight forward application of functional delta method, then, establishes the asymptotic normality of $\hat{S}(t_1, t_2 \mid \underline{z}_0)$ with mean $S(t_1, t_2 \mid \underline{z}_0)$ and variance that can be estimated as follows.

From Andersen et al. (1993, p. 503), the covariance between $\hat{\underline{\beta}}_1$ and $\hat{\Lambda}_{10}(t_1)$ is consistently estimated as

$$C_1^*(t_1, \hat{\underline{\beta}}_1) = -A_1^{-1}(\hat{\underline{\beta}}_1) \int_0^{t_1} S_1^{(1)}(s, \hat{\underline{\beta}}_1) \left(S_1^{(0)}(s, \hat{\underline{\beta}}_1)\right)^{-2} dN_1(s) \tag{3.31}$$

where

$$dN_1(t_1) = \sum_{i=1}^{n} dN_{1i}(t_1)$$

with

$$dN_{1i}(t_1) = I(T_{1i} \in [t_1, t_1 + \Delta t_1); \delta_{1i} = 1).$$

On similar lines, we can also obtain the covariance between $\hat{\underline{\beta}}_2$ and $\hat{\Lambda}_{20}(t_1, t_2)$, which can be consistently estimated by

$$C_2^*(t_1, t_2, \hat{\underline{\beta}}_2) = -A_2^{-1}(\hat{\underline{\beta}}_2) \int_0^{t_2} S_2^{(1)}(t_1, s, \hat{\underline{\beta}}_2) \left(S_2^{(0)}(t_1, s, \hat{\underline{\beta}}_1)\right)^{-2} dN_2(t_1, s). \tag{3.32}$$

The delta method is then used to estimate the covariance matrix of $(\hat{\underline{\beta}}_1, \hat{\underline{\beta}}_2, \hat{\Lambda}_{10}(t_1), \hat{\Lambda}_{20}(t_1, t_2))$. But in practical situations an attractive approach to variance or confidence interval estimation is through resampling methods. The naive bootstrap procedure of resampling the observed data units $(\tilde{X}_{1i}, \tilde{X}_{2i}, \delta_{1i}, \delta_{2i}, \underline{z}_0)$ with replacement will be satisfactory under fairly mild conditions (see Efron and Tibshirani, 1993).

Remark 3.1.

In the absence of covariates ($\beta_1 = 0, \beta_2 = 0$), the expressions (3.23) and (3.25) for cumulative hazard function will reduces to the non-parametric estimates of $\hat{\Lambda}_1(t_1)$ and $\hat{\Lambda}_2(t_1, t_2)$ given in Wang and Wells (1998).

3.4. Simulation Study

In this section, we carried out a simulation study to evaluate the performance of the aforementioned inference procedures. We consider a Gumbel's (1960) bivariate exponential distribution with survival function

$$S(t_1, t_2) = \exp(-t_1 - t_2 - \gamma t_1 t_2), \; t_1, t_2 > 0 \tag{3.33}$$

with hazard functions

$$\lambda_1(t_1) = 1 \text{ and } \lambda_2(t_1, t_2) = (1 + \gamma t_1). \tag{3.34}$$

Two covariates z_1 and z_2 are generated from uniform (0, 1) distribution. We generated observations from Gumbel's bivariate exponential distribution for different values of γ using algorithm given in Devroye (1986). Independent censoring times are generated from the uniform distribution (0, b), where the constant b is taken in such a way that 30% of the observations are censored. We compute estimates for 1000 simulations and we then calculate average bias and variance of the estimates of $\underline{\beta}_1 = (\beta_{11}, \beta_{12})$, $\underline{\beta}_2 = (\beta_{21}, \beta_{22})$ and baseline cumulative hazard functions. The estimates are given in Table 3.1 to Table 3.3. As n increases, both bias and variance of the estimates decreases.

Table 3.1. Bias and variance of $\hat{\beta}_{11}$ and $\hat{\beta}_{12}$

β_{11}	β_{12}	γ	n	Bias $\hat{\beta}_{11}$	Var $\hat{\beta}_{11}$	Bias $\hat{\beta}_{12}$	Var $\hat{\beta}_{12}$
1.1	-0.8	0.7	50	-0.0789	0.0716	-0.0819	0.0884
			250	-0.0486	0.0711	-0.0398	0.0113
		0.8	50	-0.0157	0.0671	-0.0807	0.0934
			250	-0.0134	0.0525	-0.0754	0.0536
0.7	0.9	0.7	50	-0.0801	0.0579	-0.0433	0.0607
			250	-0.0462	0.0444	-0.0102	0.0296
		0.8	50	-0.0826	0.0618	-0.0499	0.0693
			250	-0.0493	0.0329	-0.0104	0.0475

Table 3.2. Bias and variance of $\hat{\beta}_{21}$ and $\hat{\beta}_{22}$

β_{21}	β_{22}	γ	n	Bias $\hat{\beta}_{21}$	Var $\hat{\beta}_{21}$	Bias $\hat{\beta}_{22}$	Var $\hat{\beta}_{22}$
1	-1.3	0.7	50	0.0236	0.0282	0.0289	0.0575
			250	0.0210	0.0185	0.0287	0.0181
		0.8	50	0.0231	0.0297	0.0284	0.0368
			250	0.0188	0.0118	0.0280	0.0157
1.2	0.8	0.7	50	0.0273	0.0313	0.0182	0.0339
			250	0.0239	0.0183	0.0167	0.0275
		0.8	50	0.0236	0.0281	0.0166	0.0162
			250	0.0224	0.0101	0.0153	0.0129

Table 3.3. Bias and variance of estimates of the baseline cumulative hazard functions

β_{11}	β_{12}	β_{21}	β_{22}	(t_1,t_2)	γ	n	Bias $\hat{\Lambda}_{10}(t_1)$	Var $\hat{\Lambda}_{10}(t_1)$	Bias $\hat{\Lambda}_{20}(t_1,t_2)$	Var $\hat{\Lambda}_{20}(t_1,t_2)$
1.1	-0.8	1	-1.3	(1,0.9)	0.7	50	0.0969	0.0691	-0.0685	0.0301
						250	0.0305	0.0101	-0.0275	0.0192
					0.8	50	0.0124	0.0289	-0.0636	0.0364
						250	0.0102	0.0184	-0.0543	0.0156
				(0.8,1.1)	0.7	50	0.0322	0.0554	-0.0515	0.0307
						250	0.0291	0.0161	-0.0448	0.0255
					0.8	50	0.0255	0.0197	-0.0567	0.0500
						250	0.0105	0.0157	-0.0194	0.0454
0.7	0.9	1.2	0.8	(1,0.9)	0.7	50	0.0723	0.0769	-0.0522	0.0589
						250	0.0687	0.0656	-0.0293	0.0208
					0.8	50	0.0694	0.0736	0.0039	0.0764
						250	0.0201	0.0438	0.0003	0.0658
				(0.8,1.1)	0.7	50	0.0586	0.0432	-0.0279	0.0542
						250	0.0524	0.0374	-0.0229	0.0129
					0.8	50	0.0308	0.0619	0.0126	0.0855
						250	0.0127	0.0388	0.0015	0.0782

3.5. Data Analysis

For the illustration of the procedure, we consider a data given in Lawless (2003, p.531). This data show on the recurrent times to infection at the point of insertion of the catheter for 38 persons undergoing kidney dialysis. Data for the first two occurrences of infection are given; either one or both may be censored, because catheters were sometimes removed for causes other than infection. The two covariates considered are sex (1 = male, 2 = female) and kidney disease type (0 = glomerulo nephritis, 1 = acute nephritis, 2 = polycystic kidney disease, 3 = other). T_1 and T_2 represents the first two occurrences of infection.

We compute the estimates of $\underline{\hat{\beta}}_1 = (\hat{\beta}_{11}, \hat{\beta}_{12})$, $\underline{\hat{\beta}}_2 = (\hat{\beta}_{21}, \hat{\beta}_{22})$ by the method given in Section 3.3. The estimates $\underline{\hat{\beta}}_1$ and $\underline{\hat{\beta}}_2$ are $\underline{\hat{\beta}}_1 = (-1.5046, -0.2354)$, $\underline{\hat{\beta}}_2 = (-0.5185, -0.1008)$. It follows that both $\underline{\hat{\beta}}_1$ and $\underline{\hat{\beta}}_2$ have negative values. Thus the covariates in the study have negative effect on the recurrence time of the individuals. We then compute the estimates of baseline cumulative hazard functions

and survival functions. The estimates are given in Table 3.4. From Table 3.4, we can observe that $\hat{\Lambda}_{10}(t_1)$ is increasing in t_1, as expected. However, both $\hat{\Lambda}_{20}(t_1,t_2)$ and $\hat{S}(t_1,t_2 \mid \underline{z})$ are depends on t_1 and t_2. Figure 3.1 shows the estimates of the survival function.

Table 3.4. Estimates of baseline cumulative hazard functions and survival function

(t_1,t_2)	$\underline{z}$	$\hat{\Lambda}_{10}(t_1)$	$\hat{\Lambda}_{20}(t_1,t_2)$	$\hat{S}(t_1,t_2 \mid \underline{z})$
(2,25)	(1,0)	0.3825	0.4547	0.7007
(7,9)	(1,0)	1.2435	0.1630	0.6885
(7,333)	(2,1)	1.2435	7.5370	0.0851
(8,16)	(1,3)	1.7286	0.3486	0.7097
(12,40)	(1,1)	2.2409	1.3795	0.3211
(13,66)	(2,1)	2.8038	1.9359	0.4820
(15,154)	(1,0)	3.9549	2.7114	0.0827
(22,28)	(1,3)	5.3857	0.6930	0.4085

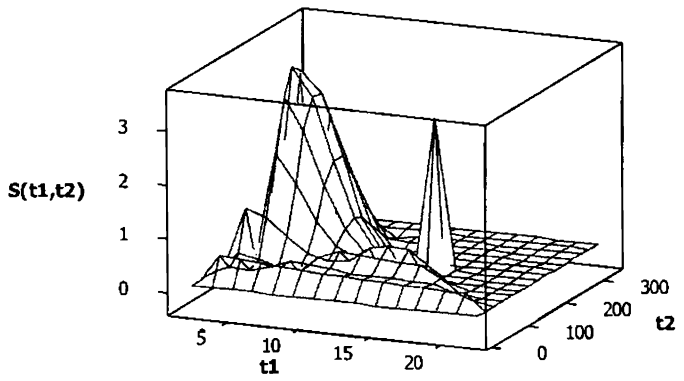

Figure 3.1. Estimates of bivariate survival function

3.6. Conclusion

In this chapter, we developed a bivariate proportional hazards model, using marginal and conditional hazard functions, for gap times of recurrent events. Since the covariates under study have different effect on marginal and conditional hazard functions, proposed model is more useful to study the dependence of gap times on covariates. The estimators of the parameters and the baseline cumulative hazard functions were developed. Asymptotic properties of estimators were studied. Then, we illustrated our procedure with a real life data, given in Lawless (2003).

Chapter Four

Proportional Mean Residual Life Model for Gap Time Distributions of Recurrent Events

4.1. Introduction

In survival studies, it is often of interest to analyze the mean residual life function to characterize the stochastic behavior of survival over time. The mean residual life function $m(t)$ defined in (1.8) is interpreted as the average remaining lifetime of an individual given that the individual has survived up to time t. Although the hazard function, mean residual life function and survival function are in one-to-one correspondence with each other, Muth (1977) considered the mean residual life to be a superior concept than the hazard function on the following grounds:

a) Regarding the ageing phenomena the two concepts are not equivalent. A decreasing mean residual life does not imply an increasing hazard function, though the converse is true. Thus the decreasing mean residual life is more general in character.

b) The hazard function accounts only for the immediate future in assessing failure phenomenon as described by the derivative of $S(t)$, where as the latter is descriptive of the entire future implied through the integral of $S(t)$ over t to ∞. A consequence of this is that a component may experience deterioration though its hazard function may be zero at a certain point.

c) It is advantageous to use the mean residual life function as a decision making criterion for replacement or maintenance policies. The expected remaining life of the component gives an indication of whether to replace or to re-schedule and

---

The results in this Chapter have been published as entitled "Proportional Mean Residual Life Model for Gap Time Distributions of Recurrent Events", *Metron, Vol. LXV, n.3* (see Sreeja and Sankaran (2007)).

this could be more useful than the hazard function to formulate maintenance policies.

Oakes and Dasu (1990) considered a proportional mean residual life model as an alternative to Cox (1972) proportional hazards model to assess the effects of covariates on the survival time. The model is defined by

$$m(t \mid \underline{z}) = m_0(t)e^{\underline{\beta}'\underline{z}} \tag{4.1}$$

where $m(t \mid \underline{z})$ is the mean residual life corresponding to the r-vector of covariates $\underline{z}$, $m_0(t)$ is the unknown baseline mean residual life function when $\underline{z} = 0$ and $\underline{\beta}$ is the vector of regression parameters. Generally, there is no direct relationship between the proportional mean residual life model and the Cox proportional hazards model. However, Oakes and Dasu (1990) proved that, when a model satisfies both the proportional hazards and the proportional mean residual life assumptions, its underlying distribution then belongs to the Hall-Wellner class of distributions with linear mean residual life function (Hall and Wellner, 1981).

Previous work on the mean residual life has focused on single-sample and two-sample cases; see Oakes and Dasu (2003). In regression analysis, Robins and Rotnitzky (1992) and Maguluri and Zang (1994) developed estimation procedures for the model (4.1) under uncensored and censored set up respectively. Recently, Chen and Cheng (2005) developed semi-parametric inference procedures for the regression model (4.1) using martingale theory of counting processes, in the presence of censoring. Later, Chen and Cheng (2006) considered a linear mean residual life model and developed inference procedures under right censoring. The analysis of gap times for recurrent event data using mean residual life function is more appropriate in many practical situations, as seen in Section 4.5. Motivated by this, we propose a bivariate proportional mean residual life model to assess the relationship between mean residual life function and covariates for gap time of recurrent events. Note that the focus will be on the development of the regression model of the duration times, when the recurrent events are of same type.

In Section 4.2, we introduce a bivariate proportional mean residual life model to assess the relationship between mean residual lifetime and covariates for gap time of recurrent events. Estimators of the parameter vectors as well as baseline

mean residual life function are discussed in Section 4.3. Asymptotic properties of the estimators are studied. A simulation study is carried out to assess the performance of the estimators in Section 4.4. In Section 4.5, we illustrate the procedure using kidney dialysis data given in Lawless (2003). Conclusions of the study is given in Section 4.6.

4.2. Bivariate Proportional Mean Residual Life Model

Suppose that an individual may experience k consecutive events at times $X_1 < X_2 < ... < X_k$. Let $T_1, T_2, ..., T_K$ represents the gap times where $T_1 = X_1$, $T_2 = X_2 - X_1$ and $T_k = X_k - X_{k-1}$. As in Chapter 3, we assume that the follow up time is subject to independent right censoring by C which implied that $(X_1, X_2, ...X_k)$ are independent of C. We now consider the regression problem in which the marginal and conditional mean residual lifetime functions of $(T_1, T_2, ...T_k)$ depend on certain covariates. We consider the case where $k = 2$. The extension to higher dimensions is direct.

Let $S(t_1, t_2) = P[T_1 \geq t_1, T_2 \geq t_2]$ be the joint survival function of T_1 and T_2. Let $m_1(t_1)$ be the mean residual life function of T_1, which is defined as

$$m_1(t_1) = E[T_1 - t_1 \mid T_1 \geq t_1]. \tag{4.2}$$

In the context of recurrent events, $m_1(t_1)$ can be interpreted as the expected remaining gap time of T_1 given that T_1 is larger than or equal to t_1. For the recurrent events, the occurrence of the second event depends on the occurrence of the first one. Accordingly, we can consider mean residual life function of T_2 given $T_1 \geq t_1$. The mean residual life function of T_2 given $T_1 \geq t_1$ is defined as

$$m_2(t_1, t_2) = E[T_2 - t_2 \mid T_1 \geq t_1, T_2 \geq t_2]\ . \tag{4.3}$$

The expression (4.3) can be interpreted as the average remaining gap time of T_2 given that T_1 is larger than or equal to t_1 and T_2 is larger than or equal to t_2. We use the term mean residual life time for average remaining gap time through out this chapter. Note that (4.3) is the second component of the vector MRL in the bivariate set up, defined in Arnold and Zahedi (1988).

Then the survival function $S(t_1,t_2)$ can be determined from (4.2) and (4.3) by the identity

$$S(t_1,t_2)=\frac{m_1(0)}{m_1(t_1)}\frac{m_2(t_1,0)}{m_2(t_1,t_2)}\exp\left[-\int_0^{t_1}\frac{du}{m_1(u)}-\int_0^{t_2}\frac{du}{m_2(t_1,u)}\right]. \tag{4.4}$$

Note that the hazard functions given in (1.34) and (1.35), and mean residual life functions are related by the identities

$$\lambda_1(t_1)=\frac{1+m_1'(t_1)}{m_1(t_1)} \tag{4.5}$$

and

$$\lambda_2(t_1,t_2)=\frac{1+m_2'(t_1,t_2)}{m_2(t_1,t_2)}. \tag{4.6}$$

where $m_1'(t_1)$ is the derivative of $m_1(t_1)$ with respect to t_1 and $m_2'(t_1,t_2)$ is the derivative for $m_2(t_1,t_2)$ with respect to t_2.

For the analysis of gap times of recurrent event data using MRL, one possible method is to consider marginal mean residual life functions of T_1 and T_2 and then apply ideas from generalized estimating functions to calculate an appropriate combination of the two marginal estimates. This can be done in the case of homogeneity among two regression coefficients. In many practical situations as shown in Section 4.4, the conditional mean residual life function of T_2 given $T_1 \geq t_1$ is meaningful than the marginal mean residual life function of T_2 to explain the joint dependence structure of pair of lifetimes on the covariate vector. Accordingly, we define a bivariate proportional mean residual life model for T_1 and T_2 as

$$m_1(t_1 \mid \underline{z})=m_{10}(t_1)e^{\underline{\beta}_1'\underline{z}} \tag{4.7}$$

and

$$m_2(t_1,t_2 \mid \underline{z})=m_{20}(t_1,t_2)e^{\underline{\beta}_2'\underline{z}}. \tag{4.8}$$

In model (4.7), $m_1(t_1 \mid \underline{z})$ is the mean residual life function at time t_1 when the $r\times 1$ covariate vector $\underline{z}$ is given and $m_{10}(t_1)$ is the baseline mean residual life function, which is the mean residual life function when $\underline{z}=0$. Here $\underline{\beta}_1$ and $\underline{\beta}_2$ are $r\times 1$

vector of parameters. For the model (4.7), the ratio $\frac{m_1(t_1 \mid \underline{z}^{(1)})}{m_1(t_1 \mid \underline{z}^{(2)})}$ of the mean residual life functions of two individuals with covariate vectors $\underline{z}^{(1)}$ and $\underline{z}^{(2)}$ does not vary with time t_1. The model (4.8) implies that the ratio $\frac{m_2(t_1,t_2 \mid \underline{z}^{(1)})}{m_1(t_1,t_2 \mid \underline{z}^{(2)})}$ of the mean residual life functions of two pairs of individuals with covariate vectors $\underline{z}^{(1)}$ and $\underline{z}^{(2)}$ does not vary with t_1 and t_2.

Using (4.5), (4.6), (4.7) and (4.8), we can have

$$m_{10}(t_1) \wedge_1 (dt_1) = e^{-\underline{\beta}_1' \underline{z}} dt_1 + m_{10}(dt_1) \tag{4.9}$$

and

$$m_{20}(t_1,t_2) \wedge_2 (t_1,dt_2) = e^{-\underline{\beta}_2' \underline{z}} dt_2 + m_{20}(t_1,dt_2) \tag{4.10}$$

where $\wedge_1(t_1)$ and $\wedge_2(t_1,t_2)$ represent the cumulative hazard functions corresponding to $\lambda_1(t_1)$ and $\lambda_2(t_1,t_2)$ respectively.

From (4.4), (4.7) and (4.8), we obtain

$$S(t_1,t_2 \mid \underline{z}) = \frac{m_{10}(0)}{m_{10}(t_1)} \frac{m_{20}(t_1,0)}{m_{20}(t_1,t_2)} \exp\left[-\int_0^{t_1} \frac{du}{m_{10}(u) e^{\underline{\beta}_1' \underline{z}}} - \int_0^{t_2} \frac{du}{m_{20}(t_1,u) e^{\underline{\beta}_2' \underline{z}}} \right]. \tag{4.11}$$

4.3. Inference Procedures

The observed data set consists of n i.i.d. sets of $(\tilde{X}_{ji}, \delta_{ji}, \underline{z}_i)$, $j = 1,2$, where $\tilde{X}_{1i} = \min(T_{1i}, C_i)$, $\tilde{X}_{2i} = \min(T_{2i}, C_i - X_{1i})$, $\delta_{1i} = I(T_{1i} < C_i)$, $\delta_{2i} = I(T_{2i} < C_i - \tilde{X}_{1i})$ and $\underline{z}_i$ is the covariate vector for $i = 1,2,...,n$. Here $I(.)$ is the indicator function. Given $\underline{z}_i$, T_{ij} and C_i's are assumed to be independent. Let

$N_{1i}(t_1) = I(\tilde{X}_{1i} \le t_1)\delta_{1i}$, $N_{2i}(t_1,t_2) = I(\tilde{X}_{1i} \ge t_1, \tilde{X}_{2i} \le t_2)\delta_{2i}$, $Y_{1i}(t_1) = I(\tilde{X}_{1i} \ge t_1)$ and $Y_{2i}(t_1,t_2) = I(\tilde{X}_{1i} \ge t_1, \tilde{X}_{2i} \ge t_2)$, $i = 1,2,...,n$. From Fleming and Harrington (1991), we can write

$$E\left[N_{1i}(dt_1) \mid \mathbb{F}_{t_1^-}; \underline{\beta}_1, m_{10}(.) \right] = Y_{1i}(t_1) \wedge_{1i} (dt_1, \underline{\beta}_1, m_{10}) \tag{4.12}$$

and for fixed t_1,

$$E\left[N_{2i}(t_1,dt_2)\,|\,\mathbb{F}_{t_1^-,t_2^-};\underline{\beta}_2,m_{20}(.)\right]=Y_{2i}(t_1,t_2)\wedge_{2i}(t_1,dt_2,\underline{\beta}_2,m_{20}) \tag{4.13}$$

where $\mathbb{F}_{t_1}$ belongs to the right-continuous filtrations $\{\mathbb{F}_{t_1}:t_1\geq 0\}$ and for fixed t_1, $\mathbb{F}_{t_1,t_2}$ belongs to the right-continuous filtrations $\{\mathbb{F}_{t_1,t_2}:t_1\geq 0,t_2\geq 0\}$ with $\mathbb{F}_{t_1}$ and $\mathbb{F}_{t_1,t_2}$ are defined by

$$\mathbb{F}_{t_1}=\sigma\{N_{1i}(u),Y_{1i}(u+),\underline{z}_i:0\leq u\leq t_1,i=1,2,...,n\}$$

and

$$\mathbb{F}_{t_1,t_2}=\sigma\{N_{2i}(t_1,v),Y_{2i}(t_1,v+),\underline{z}_i:0\leq v\leq t_2,i=1,2,...,n\}.$$

Denoting

$$M_{1i}(t_1,\underline{\beta}_1,m_{10})=N_{1i}(t_1)-\int_0^{t_1}Y_{1i}(s)\wedge_{1i}(ds,\underline{\beta}_1,m_{10})$$

and

$$M_{2i}(t_1,t_2,\underline{\beta}_2,m_{20})=N_{2i}(t_1,t_2)-\int_0^{t_2}Y_{2i}(t_1,s)\wedge_{2i}(t_1,ds,\underline{\beta}_2,m_{20}),\ i=1,2,...,n,$$

$\{M_{1i}(t_1,\underline{\beta}_1,m_{10})\}$ is zero mean $\mathbb{F}_{t_1}$ martingale and for fixed t_1, $\{M_{2i}(t_1,t_2,\underline{\beta}_2,m_{20})\}$ is zero mean $\mathbb{F}_{t_1,t_2}$ martingale. Therefore the estimates of $\underline{\beta}_1$ and $m_{10}(t_1)$ are obtained from the following partial score equations;

$$\sum_{i=1}^{n}\left[m_{10}(t_1)N_{1i}(dt_1)-Y_{1i}(t_1)\left\{e^{-\underline{\beta}_1'\underline{z}_i}dt_1+m_{10}(dt_1)\right\}\right]=0 \tag{4.14}$$

and

$$\sum_{i=1}^{n}\int_0^{\tau_1}\underline{z}_i\left[m_{10}(t_1)N_{1i}(dt_1)-Y_{1i}(t_1)\left\{e^{-\underline{\beta}_1'\underline{z}_i}dt_1+m_{10}(dt_1)\right\}\right]=0,\ 0\leq t_1<\tau_1. \tag{4.15}$$

where $0<\tau_1=\inf\left\{t_1:P\left[\tilde{X}_1>t_1\right]\right\}<\infty$.

It is easy to note that, equation (4.14) is a first order linear ordinary differential equation in $m_{10}(t_1)$, which can be written as

$$\left\{\frac{\sum_{i=1}^{n}N_{1i}(dt_1)}{\sum_{i=1}^{n}Y_{1i}(t_1)}\right\}m_{10}(t_1)-m_{10}(dt_1)=Q_1(t_1,\underline{\beta}_1)dt_1,\ 0\leq t_1\leq\tau_1 \tag{4.16}$$

where $Q_1(t_1, \underline{\beta}_1) = \dfrac{\sum_{i=1}^{n} Y_{1i}(t_1) e^{-\underline{\beta}_1' \underline{z}_i}}{\sum_{i=1}^{n} Y_{1i}(t_1)}$.

Then the solution of (4.16) is obtained as

$$\hat{m}_{10}(t_1, \underline{\beta}_1) = \hat{S}_{NA}^{(1)}(t_1)^{-1} \int_{t_1}^{\tau_1} \hat{S}_{NA}^{(1)}(u) Q_1(u, \underline{\beta}_1) du \tag{4.17}$$

where

$\hat{S}_{NA}^{(1)}(t_1) = \exp\left[-\int_0^{t_1} \dfrac{\sum_{i=1}^{n} N_{1i}(du)}{\sum_{i=1}^{n} Y_{1i}(u)} du \right]$, is the Nelson-Aalen estimator of the survival function for the pooled observations.

To estimate $\underline{\beta}_1$, we replace $m_{10}(t_1)$ with $\hat{m}_{10}(t_1, \underline{\beta}_1)$ in (4.15) and then divide the resulting equation by n. This will leads to score function

$$U_1(\underline{\beta}_1) = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau_1} \{\underline{z}_i - \bar{\underline{z}}_1(t_1)\} \left\{ \hat{m}_{10}(t_1, \underline{\beta}_1) N_{1i}(dt_1) - Y_{1i}(t_1) e^{-\underline{\beta}_1' \underline{z}_i} dt_1 \right\} = 0 \tag{4.18}$$

with $\bar{\underline{z}}_1(t_1) = \dfrac{\sum_{i=1}^{n} Y_{1i}(t_1) \underline{z}_i}{\sum_{i=1}^{n} Y_{1i}(t_1)}$.

Here, $\hat{m}_{10}(t_1)$ serves as a role similar to a weight function on each individual term in the summation. Then $\hat{\underline{\beta}}_1$ is the solution of (4.18). From Chen and Cheng (2005), it follows that under some regularity conditions, the random vector $n^{1/2}(\hat{\underline{\beta}}_1 - \underline{\beta}_1)$ converges weakly to a r-vector normal variable with mean zero and covariance matrix $A_1^{-1} V_1 A_1^{-1}$, where matrices A_1 and V_1 are given as

$$A_1 = \int_0^{\tau_1} E\left[\left\{ \underline{z} - \mu_{\underline{z}}^{(1)}(t_1) \right\}^{\otimes 2} S_*^{(1)}(t_1 \mid \underline{z}) e^{-\underline{\beta}_1^{*\prime} \underline{z}} \right] dt_1 \tag{4.19}$$

and

$$V_1 = \int_0^{\tau_1} E\left[\left\{\underline{z} - \mu_{\underline{z}}^{(1)}(t_1)\right\}^{\otimes 2} S_*^{(1)}(t_1 \mid \underline{z}) m_{10}(t_1)\left\{e^{-\underline{\beta}_1^{*'}\underline{z}} dt_1 + m_{10}(dt_1)\right\}\right]. \tag{4.20}$$

with $S_*^{(1)}(t_1 \mid \underline{z}) = P[X_1 \geq t_1 \mid \underline{z}]$ and $\mu_{\underline{z}}^{(1)}(t_1)$ is the limit of $\bar{z}_1(t_1)$ as $n \to \infty$.

In addition, A_1 and V_1 can be consistently estimated by the empirical counter parts,

$$\hat{A}_1 = \frac{1}{n}\sum_{i=1}^{n} \int_0^{\tau_1} \left\{\underline{z}_i - \bar{z}_1(t_1)\right\}^{\otimes 2} Y_{1i}(t_1) e^{-\underline{\beta}_1'\underline{z}_i} dt_1 \tag{4.21}$$

and

$$\hat{V}_1 = \frac{1}{n}\sum_{i=1}^{n} \int_0^{\tau_1} \left\{\underline{z}_i - \bar{z}_1(t_1)\right\}^{\otimes 2} Y_{1i}(t_1) \hat{m}_{10}(t_1, \underline{\hat{\beta}}_1)\left\{e^{-\underline{\beta}_1'\underline{z}_i} dt_1 + \hat{m}_{10}(dt_1, \underline{\hat{\beta}}_1)\right\} \tag{4.22}$$

respectively, where $v^{\otimes 2} = vv'$ for a vector v. Inferences for $\underline{\beta}_1$ can then be made through this large sample distribution of $\underline{\beta}_1$. As shown by Chen and Cheng (2005), one can increase the efficiency of the estimator of $\underline{\beta}_1$ by using the weighted version of the estimating equation.

To estimate $\underline{\beta}_2$ and $m_{20}(t_1, t_2)$, for fixed t_1, the equations parallel to the partial score equations (4.14) and (4.15) are

$$\sum_{i=1}^{n}\left[m_{20}(t_1, t_2) N_{2i}(t_1, dt_2) - Y_{2i}(t_1, t_2)\left\{e^{-\underline{\beta}_2'\underline{z}} dt_2 + m_{20}(t_1, dt_2)\right\}\right] = 0 \tag{4.23}$$

and

$$\sum_{i=1}^{n} \int_0^{\tau_2} \underline{z}_i \left[m_{20}(t_1, t_2) dN_{2i}(t_1, t_2) - Y_{2i}(t_1, t_2)\left\{e^{-\underline{\beta}_2'\underline{z}_i} dt_2 + m_{20}(t_1, dt_2)\right\}\right] = 0, 0 \leq t_2 \leq \tau_2, \tag{4.24}$$

where for fixed t_1, $0 < \tau_2 = \inf\left\{t_2 : P[X_2 > t_2 \mid X_1 > t_1]\right\} < \infty$.

Similarly, from (4.23), we obtain the first order linear ordinary differential equation in $m_{20}(t_1, t_2)$ as

$$\left\{\frac{\sum_{i=1}^{n} N_{2i}(t_1, dt_2)}{\sum_{i=1}^{n} Y_{2i}(t_1, t_2)}\right\} m_{20}(t_1, t_2) - m_{20}(t_1, dt_2) = Q_2(t_1, t_2, \underline{\beta}_2) dt_2, \; 0 \leq t_2 < \tau_2 \tag{4.25}$$

where $Q_2(t_1,t_2,\underline{\beta}_2)=\dfrac{\sum_{i=1}^{n}Y_{2i}(t_1,t_2)e^{-\underline{\beta}_2'\underline{z}_i}}{\sum_{i=1}^{n}Y_{2i}(t_1,t_2)}$.

Thus, for fixed t_1, the solution of (4.25) is given by

$$\hat{m}_{20}(t_1,t_2,\underline{\beta}_2)=\hat{S}_{NA}^{(2)}(t_1,t_2)^{-1}\int_{t_2}^{\tau_2}\hat{S}_{NA}^{(2)}(t_1,u)Q_2(t_1,u,\underline{\beta}_2)du \tag{4.26}$$

where $\hat{S}_{NA}^{(2)}(t_1,t_2)=\exp\left[-\int_0^{t_2}\dfrac{\sum_{i=1}^{n}N_{2i}(t_1,du)}{\sum_{i=1}^{n}Y_{2i}(t_1,u)}du\right]$.

From (4.24) and (4.26), for fixed t_1, we obtain the score function as

$$U_2(\underline{\beta}_2)=\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau_2}\{\underline{z}_i-\bar{\underline{z}}_2(t_1,t_2)\}\left\{\hat{m}_{20}(t_1,t_2,\underline{\beta}_2)dN_{2i}(t_1,t_2)-Y_{2i}(t_1,t_2)e^{-\underline{\beta}_2'\underline{z}_i}dt_2\right\}=0, \tag{4.27}$$

with $\bar{\underline{z}}_2(t_1,t_2)=\dfrac{\sum_{i=1}^{n}Y_{2i}(t_1,t_2)\underline{z}_i}{\sum_{i=1}^{n}Y_{2i}(t_1,t_2)}$.

The solution of the equation (4.27) provides the estimate of $\underline{\beta}_2$.

For fixed t_1, the asymptotic properties of $\hat{\underline{\beta}}_2$ and $\hat{m}_2(t_1,t_2)$ can be established by extending the proofs for $\hat{\underline{\beta}}_1$ and $\hat{m}_1(t_1)$, given in Chen and Cheng (2005). To see this, let $H(.)$ be the marginal distribution function of $\underline{z}$ and let $S^*(t_2\mid t_1)$ be the conditional survival function of T_2 given $T_1\geq t_1$. Since $m_{20}(t_1,t_2)$ is the mean residual life function of T_2 given $T_1\geq t_1$, it follows from Arnold and Zahedi (1988) and (4.10) that

$$m_{20}(t_1,t_2)S^*(t_2\mid t_1,\underline{z})=e^{-\underline{\beta}_2'\underline{z}}m_{20}(t_1,t_2)S^*(t_2\mid t_1,\underline{z})$$

$$=e^{-\underline{\beta}_2'\underline{z}}\int_{t_2}^{\tau_2}S^*(u\mid t_1,\underline{z})du$$

for any possible $\underline{Z} = \underline{z} \in Supp\{\underline{z} \in \mathbb{R}^p; H(\underline{z})\}$.

For fixed t_1, using Baye's theorem,

$$m_{20}(t_1,t_2) = \frac{1}{S^*(t_2 \mid t_1)} \int_z m_2(t_1,t_2 \mid \underline{z}) S^*(t_2 \mid t_1, \underline{z}) dH(\underline{z})$$

$$= \frac{1}{S^*(t_2 \mid t_1)} \int_z \left\{ e^{-\beta_2' \underline{z}} \int_{t_2}^{\tau_2} S^*(u \mid t_1, \underline{z}) du \right\} dH(\underline{z})$$

$$= \frac{1}{S^*(t_2 \mid t_1)} \int_{t_2}^{\tau_2} \{ S^*(u \mid t_1, \underline{z}) \int_z e^{-\beta_2' \underline{z}} dH(\underline{z} \mid T_2 \geq u) \} du . \qquad (4.28)$$

If we replace $S^*(t_2 \mid t_1)$ and $\int_z e^{-\beta_2' \underline{z}} dH(\underline{z} \mid T_2 \geq u)$ with $S_{NA}^{(2)}(t_1,t_2)$ and $Q_2(t_1,u,\beta_2)$ respectively in (4.28), we obtain the same estimator for $m_{20}(t_1,t_2,\beta_2)$ as (4.26). As in the univariate case, under appropriate regularity conditions, any consistent estimates of these quantities will lead to a consistent estimator for $m_{20}(t_1,t_2)$.

Under certain regularity conditions, the random vector $n^{1/2}(\hat{\underline{\beta}}_2 - \underline{\beta}_2)$ is asymptotically r-vector normal with mean zero vector and covariance matrix $A_2^{-1} V_2 A_2^{-1}$, where the matrices A_2 and V_2 are given by

$$A_2 = \int_0^{\tau_2} E\left[\left\{ \underline{z} - \mu_{\underline{z}}^{(2)}(t_1,t_2) \right\}^{\otimes 2} S^*(t_2 \mid t_1, \underline{z}) e^{-\underline{\beta}_2^{*\prime} \underline{z}} \right] dt_2 \qquad (4.29)$$

and

$$V_2 = \int_0^{\tau_2} E\left[\left\{ \underline{z} - \mu_{\underline{z}}^{(2)}(t_1,t_2) \right\}^{\otimes 2} S_*^{(2)}(t_1,t_2 \mid \underline{z}) m_{20}(t_1,t_2) \left\{ e^{-\underline{\beta}_2^{*\prime} \underline{z}} dt_2 + m_{20}(t_1,dt_2) \right\} \right] \qquad (4.30)$$

with $\mu_{\underline{z}}^{(2)}(t_1,t_2)$ is the limit of $\bar{z}_2(t_1,t_2)$ as $n \to \infty$.

Then A_2 and V_2 can be consistently estimated by their empirical counter parts,

$$\hat{A}_2 = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau_2} \{ \underline{z}_i - \bar{z}_2(t_1,t_2) \}^{\otimes 2} Y_{2i}(t_1,t_2) e^{-\underline{\beta}_2' \underline{z}_i} dt_2 \qquad (4.31)$$

and

$$\hat{V}_2 = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau_2}\{\underline{z}_i - \overline{z}_2(t_1,t_2)\}^{\otimes 2} Y_{2i}(t_1,t_2)\hat{m}_{20}(t_1,t_2,\underline{\hat{\beta}}_2)\left\{e^{-\underline{\hat{\beta}}_2'\underline{z}_i}dt_2 + \hat{m}_{20}(t_1,dt_2,\underline{\hat{\beta}}_2)\right\} \quad (4.32)$$

respectively.

One may often be interested in estimating the survival function $S(t_1,t_2 \mid \underline{z}_0)$ of gap times with a fixed covariate $\underline{z}_0$. From (4.11), the survival function $S(t_1,t_2 \mid \underline{z}_0)$ can be estimated as

$$\hat{S}(t_1,t_2 \mid \underline{z}_0) = \frac{\hat{m}_{10}(0)}{\hat{m}_{10}(t_1)}\frac{\hat{m}_{20}(t_1,0)}{\hat{m}_{20}(t_1,t_2)}\exp\left[-\int_0^{t_1}\frac{du}{\hat{m}_{10}(u)e^{\underline{\hat{\beta}}_1'\underline{z}_0}} - \int_0^{t_2}\frac{du}{\hat{m}_{20}(t_1,u)e^{\underline{\hat{\beta}}_2'\underline{z}_0}}\right]. \quad (4.33)$$

Asymptotic distribution theory is difficult for non-parametric bivariate survival function estimates and the most attractive approach to variance or confidence interval estimation is through resampling methods. The naive bootstrap procedure of resampling the observed data units $(X_{1i}, X_{2i}, \delta_{1i}, \delta_{2i}, \underline{z}_0)$ with replacement will be satisfactory under fairly mild conditions (see Efron and Tibshirani, 1993).

The most challenging part in this procedure is the tail due to potential censoring. If the underlying recurrence times are heavily right censored, it is not possible to estimate the mean residual life functions on the whole positive real line without additional assumptions. One possible approach is to modify the fully unspecified $m_{10}(t_1)$ and $m_{20}(t_1,t_2)$ by including a parametric component in the tail. For example, when τ_0 is a pre-specified truncation time, we assume that

$$m_{10}(t_1) = m_{10}(t_1)I(t_1 < \tau_0) + m_{1a}I(t_1 > \tau_0) \quad (4.34)$$

and for fixed t_1,

$$m_{20}(t_1,t_2) = m_{20}(t_1,t_2)I(t_2 < \tau_0) + m_{2a}I(t_2 > \tau_0)\ , \quad (4.35)$$

where m_{1a} and m_{2a} are some positive constants. Thus the $m_{10}(t_1)$ and $m_{20}(t_1,t_2)$ are unspecified up to the time τ_0, while it becomes exponential after τ_0, so that the techniques discussed earlier can be extended to the whole positive real line.

4.4. Simulation Study

In this section, we carried out a simulation study to evaluate the performance of the aforementioned inference procedures. We consider a Gumbel's (1960) bivariate exponential distribution with survival function

$$S(t_1,t_2)=\exp(-t_1-t_2-\gamma t_1 t_2),\ t_1,t_2>0,\ 0\le\gamma\le 1 \tag{4.36}$$

We considered a single covariate z, which is generated from uniform (0, 1) distribution. We generated observations from Gumbel's bivariate exponential distribution for different values of γ using algorithm given in Devroye (1986). Independent censoring times are generated from the uniform distribution (0, b), where the constant b is taken in such a way that 30% of the observations are censored. We compute estimates of $\underline{\beta}_1$ and $\underline{\beta}_2$ for 1000 simulations and then calculate empirical bias and variance of $\hat{\underline{\beta}}_1$ and $\hat{\underline{\beta}}_2$, which are given in Table 4.1 and Table 4.2. The empirical bias and variance of the estimates of the baseline mean residual functions along with coverage probabilities are given in Table 4.3 and Table 4.4. As n increases, both bias and variance of the estimates decreases.

Table 4.1. Bias and variance of $\hat{\beta}_1$

β_1	β_2	γ	n	Bias $\hat{\beta}_1$	Var $\hat{\beta}_1$	Cov. Prob
0.5	0.6	0.6	50	-0.0448	0.0920	0.939
			250	-0.0408	0.0545	0.949
		0.8	50	-0.0376	0.0964	0.942
			250	-0.0355	0.0626	0.948
-0.6	1.1	0.6	50	0.0577	0.0829	0.939
			250	0.0546	0.0412	0.941
		0.8	50	0.0554	0.0499	0.938
			250	0.0525	0.0289	0.943

Table 4.2. Bias and variance of $\hat{\beta}_2$

β_1	β_2	γ	n	Bias $\hat{\beta}_2$	Var $\hat{\beta}_2$	Cov. Prob
0.5	0.6	0.6	50	-0.0487	0.0406	0.942
			250	-0.0241	0.0194	0.945
		0.8	50	-0.0510	0.0461	0.946
			250	-0.0498	0.0162	0.952
-0.6	1.1	0.6	50	-0.0971	0.0631	0.939
			250	-0.0850	0.0422	0.941
		0.8	50	-0.0958	0.0756	0.948
			250	-0.0897	0.0561	0.949

Table 4.3. Bias and variance of $\hat{m}_{10}(t_1)$

β_1	β_2	t_1	γ	n	Bias $\hat{m}_{10}(t_1)$	Var $\hat{m}_{10}(t_1)$	Cov. Prob
0.5	0.6	0.05	0.6	50	0.0578	0.0206	0.943
				250	0.0102	0.0112	0.945
			0.8	50	0.0989	0.0288	0.947
				250	0.0539	0.0150	0.949
		0.09	0.6	50	0.0526	0.0248	0.951
				250	0.0101	0.0116	0.954
			0.8	50	0.0955	0.0503	0.936
				250	0.0499	0.0218	0.940
-0.6	1.1	0.05	0.6	50	0.0587	0.0736	0.941
				250	0.0105	0.0166	0.950
			0.8	50	0.0561	0.0197	0.942
				250	0.0103	0.0135	0.953
		0.09	0.6	50	0.0581	0.0243	0.949
				250	0.0105	0.0168	0.957
			0.8	50	0.0992	0.0315	0.951
				250	0.0558	0.0199	0.955

Table 4.4. Bias and variance $\hat{m}_{20}(t_1,t_2)$

β_1	β_2	(t_1,t_2)	γ	n	Bias $\hat{m}_{20}(t_1,t_2)$	Var $\hat{m}_{20}(t_1,t_2)$	Cov. Prob
0.5	0.6	(0.05, 0.06)	0.6	50	0.0232	0.0324	0.950
				250	0.0157	0.0129	0.952
			0.8	50	0.0163	0.0138	0.948
				250	0.0123	0.0101	0.953
		(0.09, 0.07)	0.6	50	0.0223	0.0269	0.939
				250	0.0144	0.0177	0.940
			0.8	50	0.0978	0.0216	0.935
				250	0.0156	0.0102	0.939
-0.6	1.1	(0.05, 0.06)	0.6	50	0.0144	0.0226	0.942
				250	0.0105	0.0129	0.948
			0.8	50	0.0189	0.0990	0.947
				250	0.0105	0.0231	0.955
		(0.09, 0.07)	0.6	50	0.0209	0.0239	0.950
				250	0.0134	0.0103	0.954
			0.8	50	0.0999	0.0162	0.955
				250	0.0132	0.0109	0.963

4.5. Data Analysis

For the illustration of the procedure, we consider a data given in Lawless (2003, p.531). The data shows the recurrent times to infection at the point of insertion of the catheter for 38 persons undergoing kidney dialysis. Data for the first two occurrences of infection are given; either one or both may be censored, because catheters were sometimes removed for causes other than infection. The covariate considered in our study is kidney disease type (0 = glomerulo nephritis, 1 = acute nephritis, 2 = polycystic kidney disease, 3 = other). T_1 and T_2 represents the first two occurrences of infection.

We compute the estimates of $\hat{\underline{\beta}}_1$ and $\hat{\underline{\beta}}_2$ by the method given in Section 4.3. The estimates $\hat{\underline{\beta}}_1$ and $\hat{\underline{\beta}}_2$ are given in Table 4.5. We then estimate $m_{10}(t_1)$ and $m_{20}(t_1,t_2)$. Finally, we estimate $S(t_1,t_2 \mid \underline{z})$ by substituting the estimates of $\underline{\beta}_1$, $\underline{\beta}_2$, $m_{10}(t_1)$ and $m_{20}(t_1,t_2)$ in (4.33). The estimates are given in Table 4.5. From Table 4.5, it is easy to see that the values of both $\hat{\underline{\beta}}_1$ and $\hat{\underline{\beta}}_2$ are negative. The disease type

has negative effect on the first and second recurrence times of the individuals respectively. As expected, $\hat{m}_{10}(t_1)$ is decreasing in t_1. However, $\hat{m}_{20}(t_1,t_2)$ depends both on t_1 and t_2.

There are natural restrictions on both $m_{10}(t_1)$ and $m_{20}(t_1,t_2)$ such that $m_{10}(t_1)$ must be monotonically non-decreasing in t_1 and $m_{20}(t_1,t_2)+t_2$ should be monotonically non-decreasing in t_2 for every fixed t_1. These constraints are satisfied in this data example. To check the model adequacy of (4.10) and (4.11), the estimated mean residual life functions of T_1 and that of T_2 given $T_1 \geq t_1$, without adjusting for any of the covariates and the estimated baseline mean residual life functions are plotted. On the log scale, their lowess curves are parallel to each other and their difference is roughly constant (see Figures 4.1 and 4.2). This suggests a reasonable goodness of fit of the proportionality assumption given in (4.7) and (4.8). Figure 4.3 shows the estimates of the survival function.

Table 4.5. Estimates of $\underline{\beta}_1, \underline{\beta}_2$, $m_{10}(t_1)$ and $m_{20}(t_1,t_2)$

(t_1,t_2)	$\hat{\underline{\beta}}_1$	$\hat{\underline{\beta}}_2$	$\underline{z}$	$\hat{m}_{10}(t_1)$	$\hat{m}_{20}(t_1,t_2)$	$\hat{S}(t_1,t_2 \mid \underline{z})$
(7,9)	-0.0774	-0.0440	0	20.3853	11.5136	0.2965
(8,16)	-0.0774	-0.0440	3	19.6691	9.2308	0.2242
(2,25)	-0.0774	-0.0440	0	21.5899	6.1896	0.0108
(30,12)	-0.0774	-0.0440	3	13.7672	11.0371	0.0061
(12,40)	-0.0774	-0.0440	1	18.0828	9.5078	0.0012
(34,30)	-0.0774	-0.0440	1	13.2308	12.0054	0.0010
(22,28)	-0.0774	-0.0440	3	17.0786	3.5833	0.0001

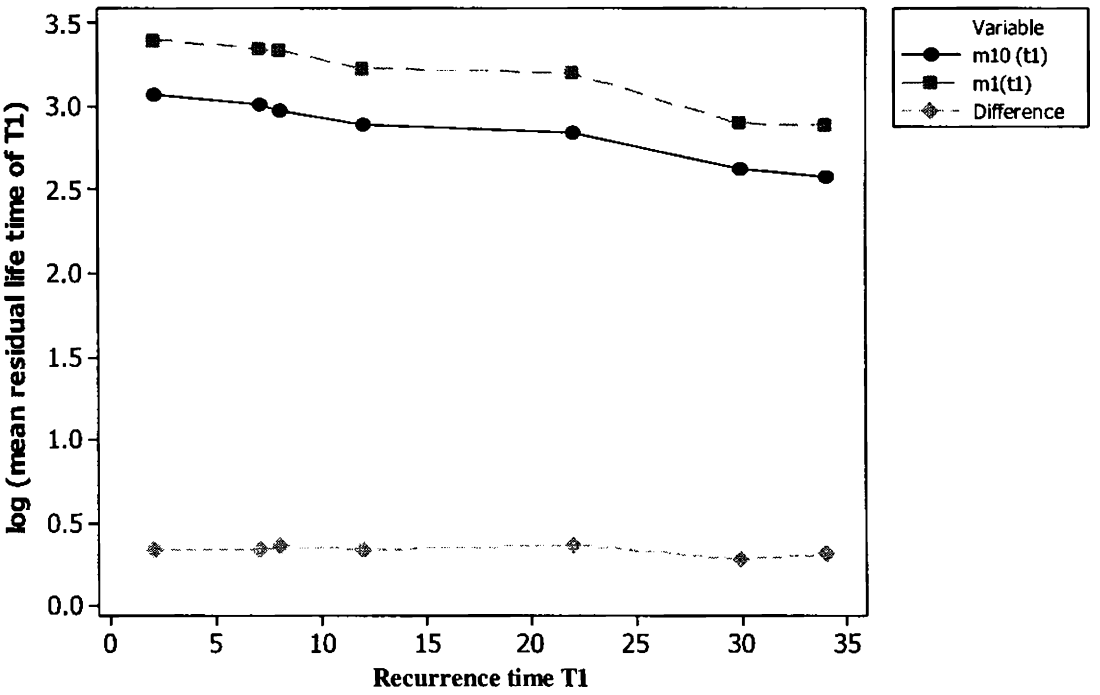

Figure 4. 1. Log of mean residual lifetimes for T_1

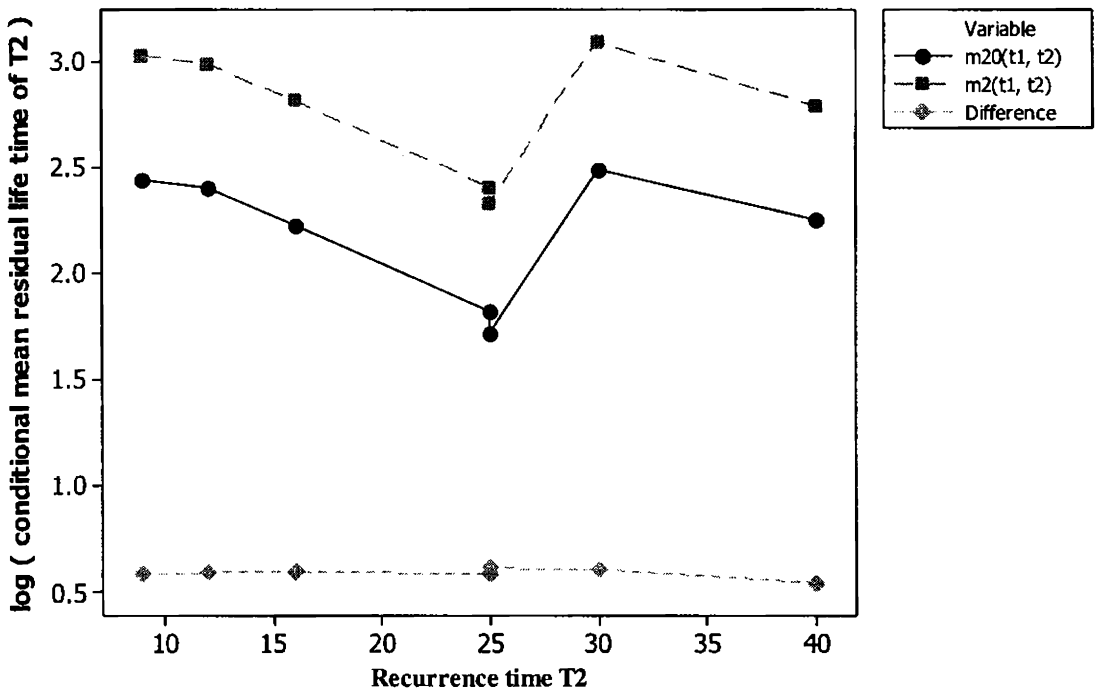

Figure 4. 2. Log of mean residual lifetimes for T_2

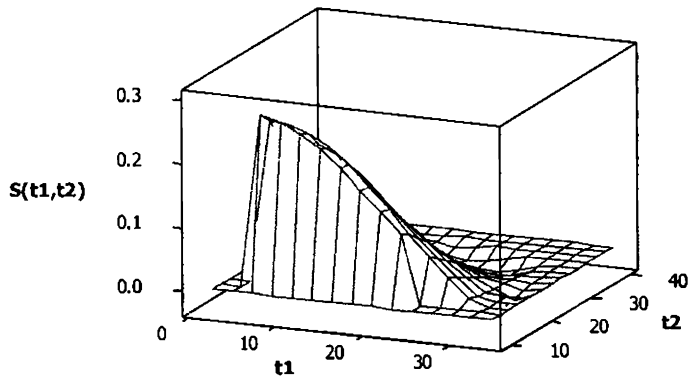

Figure 4.3. Estimates of bivariate survival function

4.6. Conclusion

We introduced a bivariate proportional mean residual life model to asses the effect of covariates on the mean residual life function for the gap time distribution of recurrent events. The model is a transparent extension of the mean residual lifetime model developed by Chen and Cheng (2005) for univariate survival data. The estimation of parameter vectors and baseline mean residual life functions were done using counting process theory. The proposed method can directly be extended to the higher dimensions by considering the multivariate mean residual life function of Arnold and Zahedi (1988).

The efficiency of the proposed method depends on the conditionally i.i.d. assumption on the bivariate recurrence times as well as the independent censoring assumption. In the presence of trend on the bivariate recurrence times, the i.i.d. assumption will be violated and therefore, the proposed method would not be appropriate. The independent assumption could also fail if the observed data is terminated by information drop out or a failure event. Both assumptions should be examined carefully when applying the proposed method.

Chapter Five

Proportional Hazards Model for Successive Duration Times under Informative Censoring

5.1. Introduction

In the analysis of censored lifetime data, it is usually assumed that the lifetime variables are independent of the censoring variables to ensure identifiably of the marginal survival function. This assumption is referred to as 'non-informative censoring'. In many practical situations in survival studies, the 'non-informative censoring' assumption is not realistic. For example, in the analysis of duration times of two successive events, the length of the first duration affects the chance of the second duration being censored. In such cases, the two duration times are correlated and accordingly the second duration time is censored by a dependent variable related to the first duration time. Examples where successive durations arise are numerous. In reliability theory, the first duration might correspond to the time interval separating the moment a new machine starts operating and the moment a fault is detected, and the second duration to the subsequent time interval between detection of the fault and failure of the machine. In economics, the first duration might refer to the time an individual is unemployed, and the second duration refer to the time the individual is employed. In biometrics, the durations might correspond to the successive stages of a disease, or to a sequence of repeated events such as certain cyclic movements in the small bowl. Wang and Wells (1998) provide non-parametric estimation of successive duration times under dependent censoring in absence of covariates. The analysis of such data in the presence of covariate is not yet carried out. Motivated by this, we consider the regression problem for duration

---

The results in this Chapter have been communicated as entitled "Proportional Hazards Model for Successive Duration Times under Informative Censoring" (see Sreeja and Sankaran (2008a)).

times of successive events under informative censoring. The idea used in Braekers and Veraverbeke (2005) for the analysis of partially informative censored lifetime data in univariate set up, is extended to the analysis of duration times of two successive events.

In Section 5.2, we introduce and study semi-parametric proportional hazards models. We estimate the parameters and baseline hazard functions of the models in Section 5.3. In Section 5.4, asymptotic properties of the estimators are studied. A simulation study is carried out in Section 5.5 to assess the performance of the estimators. In Section 5.6, the procedure is illustrated using a real life data. Finally, we conclude our study in Section 5.7.

5.2. The Model

Let (X_1, X_2) be the duration times of two consecutive and adjacent events with joint survival function $S(t_1, t_2)$. Both X_1 and X_2 are subject to right censoring by C_1. Let $\underline{Z}$ be the vector of r-covariates present in the study. Let C_2 be a censoring time corresponding to X_2 in such a way that C_2 relates to C_1 by $C_2 = C_1 - X_1$. One observes the random vector $T = (T_1, T_2)$, where $T_1 = \min(X_1, C_1)$ and $T_2 = \min(X_2, C_2) I(X_1 \le C_1)$, with $I(.)$ as the indicator function. Then the observed data is $(T_1, T_2, \delta, \underline{z})$ where δ , the censoring indicator, is defined by

$$\delta = \begin{cases} 1 & if \quad C_1 < X_1 \\ 2 & if \quad X_1 \le C_1 < X_1 + X_2 \\ 3 & if \quad X_1 + X_2 \le C_1 \end{cases}. \tag{5.1}$$

In the first regime, both durations are censored. In the second regime, X_1 is observed while X_2 is censored. In the third regime, both durations are observed. When $\delta = 1$, there is no information available about X_2.

Let $G_1(t_1)$ denote the survival function of the censoring time C_1. Denote $G_2(t_1, t_2)$ as the survival function of C_2 given $C_1 > t_1$ and $G(t_1, t_2)$ as the joint survival function of C_1 and C_2. Let $S_1(t_1)$ be the marginal survival function of X_1 and $S_2(t_1, t_2)$ be the survival function of X_2 given $X_1 = t_1$. Suppose $g_1(t_1)$, $g_2(t_1, t_2)$,

$f_1(t_1)$ and $f_2(t_1,t_2)$ are the density functions of C_1, C_2 given $C_1 > t_1$, X_1 and X_2 given $X_1 > t_1$ respectively.

Under informative censoring scheme, we assume the following relationships

$$G_1(t_1 \mid \underline{z}) = S_1(t_1 \mid \underline{z})^{\underline{\eta}_z} \tag{5.2}$$

and

$$G_2(t_1,t_2 \mid \underline{z}) = S_2(t_1,t_2 \mid \underline{z})^{\underline{\gamma}_z} \tag{5.3}$$

for some constants $\underline{\eta}_z > 0$, $\underline{\gamma}_z > 0$ depending on the covariate vector $\underline{z}$.

The parameters $\underline{\eta}_z$ and $\underline{\gamma}_z$ satisfies a model

$$\underline{\eta}_z = \phi_1(\underline{z}, \underline{\eta}_0) \tag{5.4}$$

and

$$\underline{\gamma}_z = \phi_2(\underline{z}, \underline{\gamma}_0) \tag{5.5}$$

with ϕ_1 and ϕ_2 are some known functions and $\underline{\eta}_0 = (\eta_{01},...\eta_{0r})$ and $\underline{\gamma}_0 = (\gamma_{01},...\gamma_{0r})$ are vectors of r unknown parameters. We assume that ϕ_1 and ϕ_2 are strictly positive in a neighborhood of $\underline{\eta}_0$ and $\underline{\gamma}_0$ respectively and has partial derivatives of first and second order in its neighborhood, denoted as follows;

$$\phi_{1j}{}' = \frac{\partial \phi_1}{\partial \eta_{0j}},\ \phi_{1ij}{}'' = \frac{\partial^2 \phi_1}{\partial \eta_{0i} \partial \eta_{0j}},\ i,j = 1,...,r. \tag{5.6}$$

$$\phi_{2j}{}' = \frac{\partial \phi_2}{\partial \gamma_{0j}},\ \phi_{2ij}{}'' = \frac{\partial^2 \phi_2}{\partial \gamma_{0i} \partial \gamma_{0j}},\ i,j = 1,...,r. \tag{5.7}$$

From straight forward calculations we obtain,

$$P[X_1 \ \ is \ \ observed \mid \underline{Z} = \underline{z}] = \int_0^\infty G_1(u \mid \underline{z}) f_1(u \mid \underline{z}) du$$

$$= \int_0^\infty S_1(u \mid \underline{z})^{\underline{\eta}_z} f_1(u \mid \underline{z}) du \tag{5.8}$$

$$P[X_1 \ \ is \ \ censored \mid \underline{Z} = \underline{z}] = \int_0^\infty g_1(u \mid \underline{z}) S_1(u \mid \underline{z}) du$$

$$= \underline{\eta}_z \int_0^\infty S_1(u \mid \underline{z})^{\underline{\eta}_z} f_1(u \mid \underline{z}) du \tag{5.9}$$

$$P[\delta = 2 \mid \underline{Z} = \underline{z}] = \int_0^\infty \int_0^\infty f_1(u \mid \underline{z}) S_2(t_1, v \mid \underline{z}) G_1(u \mid \underline{z}) g_2(t_1, v \mid \underline{z}) du dv$$

$$= \underline{\gamma}_z \int_0^\infty \int_0^\infty f_1(u \mid \underline{z}) f_2(t_1, v \mid \underline{z}) S_1(u \mid \underline{z})^{\underline{\eta}_z} S_2(t_1, v \mid \underline{z})^{\underline{\gamma}_z} du dv \tag{5.10}$$

and

$$P[\delta = 3 \mid \underline{Z} = \underline{z}] = \int_0^\infty \int_0^\infty f_1(u \mid \underline{z}) f_2(t_1, v \mid \underline{z}) G_1(u \mid \underline{z}) G_2(t_1, v \mid \underline{z}) du dv$$

$$= \int_0^\infty \int_0^\infty f_1(u \mid \underline{z}) f_2(t_1, v \mid \underline{z}) S_1(u \mid \underline{z})^{\underline{\eta}_z} S_2(t_1, v \mid \underline{z})^{\underline{\gamma}_z} du dv \tag{5.11}$$

Thus the parameter vectors $\underline{\eta}_z$ in (5.4) and $\underline{\gamma}_z$ in (5.5) can be interpreted as

$$\underline{\eta}_z = \frac{P[X_1 \quad is \quad censored \mid \underline{Z} = \underline{z}]}{P[X_1 \quad is \quad observed \mid \underline{Z} = \underline{z}]} \tag{5.12}$$

and

$$\underline{\gamma}_z = \frac{P[\delta = 3 \mid \underline{Z} = \underline{z}]}{P[\delta = 2 \mid \underline{Z} = \underline{z}]}. \tag{5.13}$$

Let $\lambda_1(t_1)$ be the hazard function of X_1 and $\lambda_2(t_1, t_2)$ be the hazard function of X_2 given $X_1 = t_1$.

Now we consider the proportional hazards models,

$$\lambda_1(t_1 \mid \underline{z}) = \lambda_{10}(t_1) e^{\underline{\beta}_1' \underline{z}} \tag{5.14}$$

and

$$\lambda_2(t_1, t_2 \mid \underline{z}) = \lambda_{20}(t_1, t_2) e^{\underline{\beta}_2' \underline{z}} \tag{5.15}$$

where $\lambda_{10}(t_1)$ and $\lambda_{20}(t_1, t_2)$ are baseline hazard functions of X_1 and X_2 given $X_1 = t_1$ and $\underline{\beta}_1$ and $\underline{\beta}_2$ are $r \times 1$ vectors of parameters.

The model (5.14) means that the ratio $\frac{\lambda_1(t_1 \mid \underline{z}^{(1)})}{\lambda_1(t_1 \mid \underline{z}^{(2)})}$ of the hazard functions of two individuals with covariate vectors $\underline{z}^{(1)}$ and $\underline{z}^{(2)}$ does not vary with time t_1. The model (5.15) means that the ratio $\frac{\lambda_2(t_1, t_2 \mid \underline{z}^{(1)})}{\lambda_2(t_1, t_2 \mid \underline{z}^{(2)})}$ of the hazard functions of two individuals with covariate vectors $\underline{z}^{(1)}$ and $\underline{z}^{(2)}$ does not vary with t_1 and t_2.

In many situations, one may be interested in the joint survival function $S(t_1,t_2 \mid \underline{z})$ of duration times.

The survival function can be obtained as

$$S(t_1,t_2 \mid \underline{z}) = \exp[-\wedge_{10}(t_1)e^{\underline{\beta}_1'\underline{z}} - \wedge_{20}(t_1,t_2)e^{\underline{\beta}_2'\underline{z}}] \tag{5.16}$$

where $\wedge_{10}(t_1) = \int_0^{t_1} \lambda_{10}(u)du$ and $\wedge_{20}(t_1,t_2) = \int_0^{t_2} \lambda_{20}(t_1,v)dv$ are baseline cumulative hazard functions.

5.3. Estimation Procedures

Suppose now that $(T_{1i},T_{2i},\delta_i,Z_i)$, $i=1,2,..,n$ are i.i.d. observations having the same distribution as (T_1,T_2,δ,Z). The likelihood contribution of a pair i with $T_{1i}=t_{1i}$, $T_{2i}=t_{2i}$, $\delta_i=d_i$ and the covariate vector $\underline{Z}_i=\underline{z}_i$ is given by

$$L = \begin{cases} g_1(t_{1i} \mid \underline{z}_i)S_1(t_{1i} \mid \underline{z}_i) & if \quad d_i=1 \\ f_1(t_{1i} \mid \underline{z}_i)S_2(t_{1i},t_{2i} \mid \underline{z}_i)G_1(t_{1i} \mid \underline{z}_i)g_2(t_{1i},t_{2i} \mid \underline{z}_i) & if \quad d_i=2 \\ f(t_{1i},t_{2i} \mid \underline{z}_i)G(t_{1i},t_{2i} \mid \underline{z}_i) & if \quad d_i=3 \end{cases}.$$

Then the likelihood function can be written as

$$L = \prod_{\delta_i=1} g_1(t_{1i} \mid \underline{Z}_i)S_1(t_{1i} \mid \underline{Z}_i) \prod_{\delta_i=2} f_1(t_{1i} \mid \underline{Z}_i)S_2(t_{1i},t_{2i} \mid \underline{Z}_i)G_1(t_{1i} \mid \underline{Z}_i)g_2(t_{1i},t_{2i} \mid \underline{Z}_i)$$

$$\prod_{\delta_i=3} f(t_{1i},t_{2i} \mid \underline{Z}_i)G(t_{1i},t_{2i} \mid \underline{Z}_i). \tag{5.17}$$

Substituting (5.4), (5.5), (5.14) and (5.15) in (5.17) and simplifying, we get

$$L = \prod_{\delta_i=1} \phi_1(\underline{Z}_i,\eta_0) \prod_{\delta_i=2} \phi_2(\underline{Z}_i,\gamma_0) \prod_{i=1}^{n} \lambda_{10}(t_{1i})e^{\underline{\beta}_1'\underline{Z}_i} \prod_{i=1}^{n} e^{-(\phi_1(\underline{Z}_i,\eta_0)+1)\wedge_{10}(t_{1i})e^{\underline{\beta}_1'\underline{Z}_i}}$$

$$\prod_{\delta_i\neq 1} \lambda_{20}(t_{1i},t_{2i})e^{\underline{\beta}_2'\underline{Z}_i} \prod_{\delta_i\neq 1} e^{-(\phi_2(\underline{Z}_i,\gamma_0)+1)\wedge_{20}(t_{1i},t_{2i})e^{\underline{\beta}_2'\underline{Z}_i}}. \tag{5.18}$$

Taking logarithm on both sides of the expression (5.18), we obtain

$$\log L = \sum_{\delta_i=1} \log \phi_1(\underline{Z}_i,\eta_0) + \sum_{\delta_i=2} \log \phi_2(\underline{Z}_i,\gamma_0)$$

$$+\sum_{i=1}^{n}\left[\log \lambda_{10}(t_{1i}) + \underline{\beta}_1'\underline{Z}_i\right] - \sum_{i=1}^{n}\left[(\phi_1(\underline{Z}_i,\eta_0)+1)\wedge_{10}(t_{1i})e^{\underline{\beta}_1'\underline{Z}_i}\right]$$

$$+\sum_{\delta_i \neq 1}\left[\log \lambda_{20}(t_{1i}, t_{2i}) + \underline{\beta}_2' \underline{Z}_i\right] - \sum_{\delta_i \neq 1}\left[(\phi_2(\underline{Z}_i, \gamma_0)+1) \wedge_{20}(t_{1i}, t_{2i}) e^{\underline{\beta}_2' \underline{Z}_i}\right]. \quad (5.19)$$

By maximizing the expression (5.18), we get the estimators of $\underline{\beta}_1 = (\beta_{11}, ..., \beta_{1r})$, $\underline{\beta}_2 = (\beta_{21}, ..., \beta_{2r})$, $\underline{\eta}_0 = (\eta_{01}, ..., \eta_{0r})$ and $\underline{\gamma}_0 = (\gamma_{01}, ..., \gamma_{0r})$. In ordinary Cox proportional hazards model, $\underline{\beta}_1$ and $\underline{\beta}_2$ are estimated by maximizing partial likelihood method. But in this situation, the partial likelihood analysis is not possible, due to the presence of unknown parameters $\underline{\eta}_0$ and $\underline{\gamma}_0$. For the estimation of higher dimensional parameters in the semi-parametric models, we can use profile likelihood method as discussed in Murphy and van der Vaart (2000). We first estimate the baseline cumulative hazard functions $\wedge_{10}(t_1)$ and $\wedge_{20}(t_1, t_2)$. Consider the case of estimation of $\wedge_{10}(t_1)$. Let $\tilde{X}_{11} < ... < \tilde{X}_{1N_1}$ be ordered observed first duration times and let $\tilde{X}_{21} < ... < \tilde{X}_{2N_2}$ be ordered observed second duration times given $t_{1i} = \tilde{X}_{1j}$. As in the classical Cox model (see Andersen et al., 1993 and Murphy and van der vaart, 2000), maximization of the full likelihood over arbitrary $\wedge_{10}(t_1)$ leads to maximization over $\hat{\wedge}_{10}(t_1)$ which is a piecewise constant function with jumps at the observed deaths $\tilde{X}_{1j}$ only. Thus the least informative non-parametric estimator is given by

$$\hat{\wedge}_{10}(T_{1i}) = \sum_{j=1}^{N_1} \lambda_{1j} I_1(i \in \Re_{1j}) \quad (5.20)$$

where $\Re_{1j} = \{i : T_{1i} \geq \tilde{X}_{1j}\}$ is the risk set at $\tilde{X}_{1j}^-$. Similarly, we can obtain the estimate of $\wedge_{20}(t_1, t_2)$ as

$$\hat{\wedge}_{20}(T_{1i}, T_{2i}) = \sum_{j=1}^{N_2} \lambda_{2j} I_2(i \in \Re_{2j}) \quad (5.21)$$

where $\Re_{2j} = \{i : T_{2i} \geq \tilde{X}_{2j}, T_{1i} = \tilde{X}_{1j}\}$ is the risk set at $\tilde{X}_{2j}^-$ and $T_{1i} = \tilde{X}_{1j}$. From (5.20) and (5.21), the log likelihood function (5.19) can be written as

$$\log L = \sum_{\delta_i=1}\left[\log \phi_1(\underline{Z}_i, \eta_0)\right] + \sum_{\delta_i=2}\left[\log \phi_2(\underline{Z}_i^*, \gamma_0)\right] + \sum_{j=1}^{N_1}\left[\log \lambda_{1j} + \underline{\beta}_1' \underline{Z}_{(j)}\right]$$

$$-\sum_{i=1}^{n}\left[\left(\phi_1(\underline{Z}_i,\eta_0)+1\right)e^{\underline{\beta}_1'\underline{Z}_i}\sum_{j=1}^{N_1}\lambda_{1j}I_1(i\in\Re_{1j})\right]+\sum_{j=1}^{N_2}\left[\log\lambda_{2j}+\underline{\beta}_2'\underline{Z}^*_{(j)}\right]$$
$$-\sum_{\delta_i\neq 1}\left[\left(\phi_2(\underline{Z}^*_i,\gamma_0)+1\right)e^{\underline{\beta}_2'\underline{Z}_i}\sum_{j=1}^{N_2}\lambda_{2j}I_2(i\in\Re_{2j})\right] \tag{5.22}$$

where $\underline{Z}_{(j)};\ j=1,...,N_1$ are the covariates associated with the ordered $\tilde{X}_{11}<...<\tilde{X}_{1N_1}$ and $\underline{Z}^*_{(j)};\ j=1,...,N_2$ are the covariates associated with the ordered $\tilde{X}_{21}<...<\tilde{X}_{2N_2}$ given $t_{1i}=\tilde{X}_{1j}$.

Maximizing (5.22) with respect to λ_{1j}, we get

$$\hat{\lambda}_{1j}=\frac{1}{\sum_{i\in\Re_{1j}}\left(\phi_1(\underline{Z}_i,\hat{\eta}_0)+1\right)e^{\hat{\underline{\beta}}_1'\underline{Z}_i}}. \tag{5.23}$$

On similar lines, maximization of (5.22) with respect to λ_{2j} provides

$$\hat{\lambda}_{2j}=\frac{1}{\sum_{i\in\Re_{2j}^*}\left(\phi_2(\underline{Z}^*_i,\hat{\gamma}_0)+1\right)e^{\hat{\underline{\beta}}_2'\underline{Z}^*_i}} \tag{5.24}$$

where $\Re_{2j}^*=\{i:i\in\Re_{2j},\delta_i\neq 1\}$.

Substituting (5.23) and (5.24) in (5.22), we get the profile log likelihood as

$$\log L_P=\sum_{\delta_i=1}\left[\log\phi_1(\underline{Z}_i,\eta_0)\right]+\sum_{\delta_i=2}\left[\log\phi_2(\underline{Z}^*_i,\gamma_0)\right]$$
$$+\sum_{j=1}^{N_1}\left[-\log\left(\sum_{i\in\Re_{1j}}\left(\phi_1(\underline{Z}_i,\eta_0)+1\right)e^{\underline{\beta}_1'\underline{Z}_i}\right)+\underline{\beta}_1'\underline{Z}_{(j)}\right]$$
$$+\sum_{j=1}^{N_2}\left[-\log\left(\sum_{i\in\Re_{2j}^*}\left(\phi_2(\underline{Z}^*_i,\gamma_0)+1\right)e^{\underline{\beta}_2'\underline{Z}^*_i}\right)+\underline{\beta}_2'\underline{Z}^*_{(j)}\right]-N_1-N_2. \tag{5.25}$$

which has to be maximized with respect to $\underline{\beta}_1$, $\underline{\beta}_2$, $\underline{\eta}_0$ and $\underline{\gamma}_0$. This is of course equivalent to maximizing

$$\hat{H}(\underline{\beta})=\frac{1}{n}\sum_{\delta_i=1}\left[\log\phi_1(\underline{Z}_i,\eta_0)\right]+\frac{1}{n}\sum_{\delta_i=2}\left[\log\phi_2(\underline{Z}^*_i,\gamma_0)\right]$$
$$-\frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{1}{n}\sum_{k\in\Re_{1i}}\left(\phi_1(\underline{Z}_k,\eta_0)+1\right)e^{\underline{\beta}_1'\underline{Z}_k}\right]$$

$$+\frac{1}{n}\sum_{i=1}^{n}\underline{\beta}_1'\underline{Z}_i-\frac{1}{n}\sum_{\delta_i\neq 1}\log\left[\frac{1}{n}\sum_{k\in\Re_{2i}^{\cdot}}\left(\phi_2(\underline{Z}_k^*,\gamma_0)+1\right)e^{\underline{\beta}_2'\underline{Z}_k^*}\right]$$

$$+\frac{1}{n}\sum_{\delta_i\neq 1}\underline{\beta}_2'\underline{Z}_i^* \tag{5.26}$$

where $\underline{\beta}=(\underline{\beta}_1,\underline{\beta}_2,\underline{\eta}_0,\underline{\gamma}_0)$.

For $j=1,...,r$, the estimators $\hat{\underline{\beta}}_1$, $\hat{\underline{\beta}}_2$, $\hat{\underline{\eta}}_0$ and $\hat{\underline{\gamma}}_0$ are solutions to the following equations

$$\frac{\partial\hat{H}(\underline{\beta})}{\partial\beta_{1j}}=\sum_{i=1}^{n}\underline{Z}_{ij}-\sum_{i=1}^{n}\left(\frac{\sum_{k\in\Re_{1i}}\left(\phi_1(\underline{Z}_k,\eta_0)+1\right)e^{\underline{\beta}_1'\underline{Z}_k}\underline{Z}_{kj}}{\sum_{k\in\Re_{1i}}\left(\phi_1(\underline{Z}_k,\eta_0)+1\right)e^{\underline{\beta}_1'\underline{Z}_k}}\right)=0 \tag{5.27}$$

$$\frac{\partial\hat{H}(\underline{\beta})}{\partial\beta_{2j}}=\sum_{\delta_i\neq 1}\underline{Z}_{ij}-\sum_{\delta_i\neq 1}\left(\frac{\sum_{k\in\Re_{2i}^{\cdot}}\left(\phi_2(\underline{Z}_k^*,\gamma_0)+1\right)e^{\underline{\beta}_2'\underline{Z}_k^*}\underline{Z}_{kj}^*}{\sum_{k\in\Re_{2i}^{\cdot}}\left(\phi_2(\underline{Z}_k^*,\gamma_0)+1\right)e^{\underline{\beta}_2'\underline{Z}_k^*}}\right)=0 \tag{5.28}$$

$$\frac{\partial\hat{H}(\underline{\beta})}{\partial\underline{\eta}_{0j}}=\sum_{\delta_i=1}\frac{\phi_{1j}'(\underline{Z}_i,\eta_0)}{\phi_1(\underline{Z}_i,\eta_0)}-\sum_{i=1}^{n}\left(\frac{\sum_{k\in\Re_{1i}}\phi_{1j}'(\underline{Z}_k,\eta_0)e^{\underline{\beta}_1'\underline{Z}_k}}{\sum_{k\in\Re_{1i}}\left(\phi_1(\underline{Z}_k,\eta_0)+1\right)e^{\underline{\beta}_1'\underline{Z}_k}}\right)=0 \tag{5.29}$$

and

$$\frac{\partial\hat{H}(\underline{\beta})}{\partial\underline{\gamma}_{0j}}=\sum_{\delta_i=2}\frac{\phi_{2j}'(\underline{Z}_i,\gamma_0)}{\phi_2(\underline{Z}_i,\gamma_0)}-\sum_{\delta_i\neq 1}\left(\frac{\sum_{k\in\Re_{2i}^{\cdot}}\phi_{2j}'(\underline{Z}_k^*,\gamma_0)e^{\underline{\beta}_2'\underline{Z}_k^*}}{\sum_{k\in\Re_{2i}^{\cdot}}\left(\phi_2(\underline{Z}_k^*,\gamma_0)+1\right)e^{\underline{\beta}_2'\underline{Z}_k^*}}\right)=0. \tag{5.30}$$

Maximization of the likelihood for a fixed value of $\underline{\beta}=(\underline{\beta}_1,\underline{\beta}_2,\underline{\eta}_0,\underline{\gamma}_0)$ provides the estimator for the cumulative hazard functions as

$$\hat{\Lambda}_1(t_1\mid\underline{Z})=\sum_{j=1}^{n}\frac{I(T_{1j}\le t_1)}{\sum_{i\in\Re_{1j}^{\cdot}}\left(\phi_1(\underline{Z}_i,\hat{\eta}_0)+1\right)e^{\hat{\underline{\beta}}_1'\underline{Z}_i}} \tag{5.31}$$

and

$$\hat{\Lambda}_2(t_1,t_2\mid\underline{Z}^*)=\sum_{j=1}^{n}\frac{I(T_{2j}\le t_2,T_{1j}=t_1,\delta_i\neq 1)}{\sum_{i\in\Re_{2j}^{\cdot}}\left(\phi_2(\underline{Z}_i^*,\hat{\gamma}_0)+1\right)e^{\hat{\underline{\beta}}_2'\underline{Z}_i^*}} \tag{5.32}$$

where $\hat{\underline{\beta}} = (\hat{\underline{\beta}}_1, \hat{\underline{\beta}}_2, \hat{\underline{\eta}}_0, \hat{\underline{\gamma}}_0)$ is the estimate of $\underline{\beta} = (\underline{\beta}_1, \underline{\beta}_2, \underline{\eta}_0, \underline{\gamma}_0)$.

From (5.16), a natural estimate for $S(t_1, t_2 \mid \underline{z}_0)$ is given by

$$\hat{S}(t_1, t_2 \mid \underline{z}) = \exp[-\hat{\Lambda}_{10}(t_1) e^{\hat{\underline{\beta}}_1' \underline{z}} - \hat{\Lambda}_{20}(t_1, t_2) e^{\hat{\underline{\beta}}_2' \underline{z}}]. \tag{5.33}$$

5.4. Properties of the Estimators

To study the properties of the estimators, we define the following functions. For any continuous function $p(.)$, denote

$$E_1(p(\underline{z}), t_1) = \int p(\underline{z}) P(T_1 \geq t_1 \mid \underline{Z} = \underline{z}) q(\underline{z}) d\underline{z}$$

$$E_1^m(p(\underline{z}), t_1) = \int p(\underline{z}) P(T_1 \geq t_1, \delta = m \mid \underline{Z} = \underline{z}) q(\underline{z}) d\underline{z}, \quad m = 1, 2, 3$$

$$E_1^{1,3}(p(\underline{z}), t_1) = \int p(\underline{z}) P(T_1 \geq t_1, \delta \neq 2 \mid \underline{Z} = \underline{z}) q(\underline{z}) d\underline{z}$$

$$E_2(p(\underline{z}), t_2) = \int p(\underline{z}) P(T_2 \geq t_2 \mid T_1 = t_1, \underline{Z} = \underline{z}) q(\underline{z}) d\underline{z}$$

$$E_2^m(p(\underline{z}), t_2) = \int p(\underline{z}) P(T_2 \geq t_2, \delta = m \mid T_1 = t_1, \underline{Z} = \underline{z}) q(\underline{z}) d\underline{z}, \quad m = 1, 2, 3$$

and

$$E_2^{2,3}(p(\underline{z}), t_2) = \int p(\underline{z}) P(T_2 \geq t_2, \delta \neq 1 \mid T_1 = t_1, \underline{Z} = \underline{z}) q(\underline{z}) d\underline{z}$$

where $q(\underline{z})$ is the density function of the covariate $\underline{z}$. The empirical versions of the above functions will be denoted by $\hat{E}_1(p(\underline{z}), t_1)$, $\hat{E}_1^1(p(\underline{z}), t_1)$ etc. For example,

$$\hat{E}_1(p(\underline{z}), t_1) = \frac{1}{n} \sum_{i=1}^{n} p(\underline{z}_i) I(T_{1i} \geq t_1).$$

Now we define

$Q_1(t_1) = P(T_1 \geq t_1)$ and $Q_2(t_2) = P(T_2 \geq t_2, \delta_i \neq 1 \mid T_1 = t_1)$.

Then

$$\hat{Q}_1(t_1) = \frac{1}{n} \sum_{i=1}^{n} I(T_1 \geq t_1) \text{ and } \hat{Q}_2(t_2) = \frac{1}{n} \sum_{i=1}^{n} I(T_2 \geq t_2, \delta_i \neq 1 \mid T_1 = t_1).$$

With the above notations, we can write (5.26) as,

$$\hat{H}(\underline{\beta}) = \hat{E}_1^1(\log \phi_1(\underline{z}, \eta_0), 0) + \hat{E}_1^2(\log \phi_2(\underline{z}^*, \gamma_0), 0)$$

$$+ \int_0^{T_0} \log \hat{E}_1\left((\phi_1(\underline{z}, \eta_0) + 1) e^{\underline{\beta}_1' \underline{z}}, t_1\right) d\hat{Q}_1(t_1)$$

$$+ \hat{E}_1\left(\underline{\beta}_1'\underline{z},0\right) + \int_0^{T_0} \log \hat{E}_2\left(\left(\phi_2(\underline{z}^*,\gamma_0)+1\right)e^{\underline{\beta}_2'\underline{z}^*},t_2\right) d\hat{Q}_2(t_2)$$

$$+ \hat{E}_2^{2,3}\left(\underline{\beta}_2'\underline{z}^*,0\right). \tag{5.34}$$

Then the population version of (5.34) is,

$$H(\underline{\beta}) = E_1^1\left(\log\phi_1(\underline{z},\eta_0),0\right) + E_1^2\left(\log\phi_2(\underline{z}^*,\gamma_0),0\right)$$

$$+ \int_0^{T_0} \log E_1\left(\left(\phi_1(\underline{z},\eta_0)+1\right)e^{\underline{\beta}_1'\underline{z}},t_1\right) dQ_1(t_1)$$

$$+ E_1\left(\underline{\beta}_1'\underline{z},0\right) + \int_0^{T_0} \log E_2\left(\left(\phi_2(\underline{z}^*,\gamma_0)+1\right)e^{\underline{\beta}_2'\underline{z}^*},t_2\right) dQ_2(t_2)$$

$$+ E_2^{2,3}\left(\underline{\beta}_2'\underline{z}^*,0\right). \tag{5.35}$$

Let $I = I(\underline{\beta})$ denote the information matrix of the function $H(\underline{\beta})$.

Furthermore, the first-order partial derivatives of the function H are zero at $\underline{\beta}$; which gives, for $j = 1,2,....r$,

$$\frac{\partial H(\underline{\beta})}{\partial \beta_{1j}} = E_1\left(\underline{z}_j,0\right) + \int_0^{T_0} \frac{E_1\left(\left(\phi_1(\underline{z},\eta_0)+1\right)e^{\underline{\beta}_1'\underline{z}}\underline{z},t_1\right)}{E_1\left(\left(\phi_1(\underline{z},\eta_0)+1\right)e^{\underline{\beta}_1'\underline{z}},t_1\right)} dQ_1(t_1) = 0 \tag{5.36}$$

$$\frac{\partial H(\underline{\beta})}{\partial \beta_{2j}} = E_2^{2,3}\left(\underline{z}_j,0\right) + \int_0^{T_0} \frac{E_2\left(\left(\phi_2(\underline{z}^*,\gamma_0)+1\right)e^{\underline{\beta}_2'\underline{z}^*}\underline{z}^*,t_2\right)}{E_2\left(\left(\phi_2(\underline{z}^*,\gamma_0)+1\right)e^{\underline{\beta}_2'\underline{z}^*},t_2\right)} dQ_2(t_2) = 0 \tag{5.37}$$

$$\frac{\partial H}{\partial \eta_{0j}} = E_1^1\left(\frac{\phi_{1j}'(\underline{z},\eta_0)}{\phi_1(\underline{z},\eta_0)},0\right) + \int_0^{T_0} \frac{E_1\left(\phi_{1j}'(\underline{z},\eta_0)e^{\underline{\beta}_1'\underline{z}},t_1\right)}{E_1\left(\left(\phi_1(\underline{z},\eta_0)+1\right)e^{\underline{\beta}_1'\underline{z}},t_1\right)} dQ_1(t_1) = 0 \tag{5.38}$$

and

$$\frac{\partial H}{\partial \gamma_{0j}} = E_1^2\left(\frac{\phi_{2j}'(\underline{z}^*,\gamma_0)}{\phi_2(\underline{z}^*,\gamma_0)},0\right) + \int_0^{T_0} \frac{E_2\left(\phi_{2j}'(\underline{z}^*,\gamma_0)e^{\underline{\beta}_2'\underline{z}^*},t_2\right)}{E_2\left(\left(\phi_2(\underline{z}^*,\gamma_0)+1\right)e^{\underline{\beta}_2'\underline{z}^*},t_2\right)} dQ_2(t_2) = 0 \tag{5.39}$$

where T_0 is some pre-specified time which in most cases represents the study period.

Now, since

$P[T_1 \geq t_1, \delta = 1 \mid \underline{Z} = \underline{z}] = \underline{\eta}_{\underline{z}} e^{\underline{\beta}_1' \underline{z}} \int_{t_1}^{\infty} P[T_1 \geq u, \delta = 1 \mid \underline{Z} = \underline{z}] \lambda_{10}(u) du$, we can write

$$dE_1^1(p(\underline{z}), t_1) = -\lambda_{10}(t_1) E_1^1\left(p(\underline{z}) \underline{\eta}_{\underline{z}} e^{\underline{\beta}_1' \underline{z}}, t_1\right) dt_1. \tag{5.40}$$

Similarly, since

$$P[T_2 \geq t_2, \delta \neq 1 \mid T_1 = t_1, \underline{Z}^* = \underline{z}^*] = (1 + \underline{\gamma}_{\underline{z}^*}) e^{\underline{\beta}_2' \underline{z}^*} \int_{t_2}^{\infty} P[T_2 \geq u \mid T_1 = t_1, \underline{Z}^* = \underline{z}^*] \lambda_{20}(t_1, u) du,$$

we have

$$dQ_2(t_2) = -\lambda_{20}(t_1, t_2) E_2\left((1 + \underline{\gamma}_{\underline{z}^*}) e^{\underline{\beta}_2' \underline{z}^*}, t_2\right) dt_2. \tag{5.41}$$

These representations can be used to study the asymptotic properties of the estimates.

Theorem 5.1.

Assume that I is positive definite at $\underline{\beta}$. Assume that $E \mid \underline{Z} \mid < \infty$, $E \mid \underline{Z}^* \mid < \infty$ and that $E \mid \log \phi_1(\underline{Z}, \eta_0) \mid$, $E \mid \log \phi_2(\underline{Z}^*, \gamma_0) \mid$, $E\left[\left((\phi_1(\underline{Z}, \eta_0) + 1) e^{\underline{\beta}_1' \underline{Z}}\right)^2\right]$ and $E\left[\left((\phi_2(\underline{Z}^*, \gamma_0) + 1) e^{\underline{\beta}_2' \underline{Z}^*}\right)^2\right]$ are bounded uniformly in a neighborhood of $\underline{\beta}$. Then there exists a sequence of solutions $\hat{\underline{\beta}}$ of equations (5.27) - (5.30) such that $\hat{\underline{\beta}} \to \underline{\beta}$ a.s. as $n \to \infty$.

Proof. The positive definiteness of I at $\underline{\beta}$ implies that the function $H(.)$ has a local maximum at $\underline{\beta}$. For $\underline{\beta}^*$ in a c-neighborhood of $\underline{\beta}$ ($\| \underline{\beta}^* - \underline{\beta} \| \leq c$, with $\| \ \|$ Euclidean distance) we have that

$$H(\underline{\beta}) - H(\underline{\beta}^*) \geq 0 \tag{5.42}$$

with strict inequality if $\| \underline{\beta}^* - \underline{\beta} \| = c$. From the strong law of large numbers together with Lemmas A1 and A2 in Tsiatis (1981), we get

$$\hat{H}(\underline{\beta}) - \hat{H}(\underline{\beta}^*) \to H(\underline{\beta}) - H(\underline{\beta}^*). \tag{5.43}$$

From (5.42) and (5.43), we can prove that, there exists an n_0 such that for all $n \geq n_0$:

$$\hat{H}(\underline{\beta}) - \hat{H}(\underline{\beta}^*) > 0 \text{ for } \| \underline{\beta}^* - \underline{\beta} \| = c. \tag{5.44}$$

Since $\hat{H}$ is continuous and differentiable at $\underline{\beta}$, we get that $\hat{H}$ has a local maximum on $\| \underline{\beta}^* - \underline{\beta} \| \leq c$. Since the maximum cannot be on the boundary ($\| \underline{\beta}^* - \underline{\beta} \| = c$), the first derivatives vanish somewhere on $\| \underline{\beta}^* - \underline{\beta} \| < c$. The value where $\frac{\partial \hat{H}}{\partial \beta_{1i}} = \frac{\partial \hat{H}}{\partial \beta_{2i}} = \frac{\partial \hat{H}}{\partial \eta_{0i}} = \frac{\partial \hat{H}}{\partial \gamma_{0i}} = 0$, $i = 1,2,...r$ is the maximum likelihood estimate $\hat{\underline{\beta}}$.

We can now repeat this argument for c decreasing with n. Thus, we get a sequence $\hat{\underline{\beta}}_n$ with $\hat{\underline{\beta}}_n \to \underline{\beta}$ a.s. as $n \to \infty$, which completes the proof.

Lemma 5.1. Assume that $E\left[\left(\left(\phi_1(\underline{Z}, \eta_0) + 1\right) e^{\underline{\beta}_1' \underline{Z}}\right)^2\right]$ and

$E\left[\left(\left(\phi_2(\underline{Z}^*, \gamma_0) + 1\right) e^{\underline{\beta}_2' \underline{Z}^*}\right)^2\right]$ are bounded uniformly in a neighborhood of $\underline{\beta}$. If

$\hat{\underline{\beta}} = (\hat{\underline{\beta}}_1, \hat{\underline{\beta}}_2, \hat{\underline{\eta}}_0, \hat{\underline{\gamma}}_0)$ is any random sequence with $\hat{\underline{\beta}} \xrightarrow{p} \underline{\beta}$ as $n \to \infty$, then

$$(i) \quad \sup_{0 \leq t_1 \leq T_0} \left| \hat{\wedge}_{10}(t_1) - \wedge_{10}(t_1) \right| \xrightarrow{p} 0$$

and for fixed t_1,

$$(ii) \quad \sup_{0 \leq t_2 \leq T_0} \left| \hat{\wedge}_{20}(t_1, t_2) - \wedge_{20}(t_1, t_2) \right| \xrightarrow{p} 0.$$

Proof. First consider the case (i). From (5.40), it follows that

$$\wedge_{10}(t_1) = \int_0^{t_1} \frac{-dE_1^1\left(p(\underline{z}), u\right)}{E_1^1\left(p(\underline{z}) \underline{\eta}_{\underline{z}} e^{\underline{\beta}_1' \underline{z}}, u\right)}.$$

Then

$$\hat{\wedge}_{10}(t_1) = \int_0^{t_1} \frac{-d\hat{E}_1^1\left(p(\underline{z}), u\right)}{\hat{E}_1^1\left(p(\underline{z}) \hat{\underline{\eta}}_{\underline{z}} e^{\hat{\underline{\beta}}_1' \underline{z}}, u\right)}.$$

We have

$$\sup_{0 \leq t_1 \leq T_0} \left| \hat{E}_1^1\left(p(\underline{z}) \phi_1(\underline{z}, \hat{\eta}_0) e^{\hat{\underline{\beta}}_1' \underline{z}}, t_1\right) - E_1^1\left(p(\underline{z}) \phi_1(\underline{z}, \eta_0) e^{\underline{\beta}_1' \underline{z}}, t_1\right) \right|$$

$$\leq \sup_{0\leq t_1\leq T_0}\left|\hat{E}_1^1\left(p(\underline{z})\phi_1(\underline{z},\hat{\eta}_0)e^{\hat{\underline{\beta}}'\underline{z}},t_1\right)-E_1^1\left(p(\underline{z})\phi_1(\underline{z},\hat{\eta}_0)e^{\hat{\underline{\beta}}'\underline{z}},t_1\right)\right|$$

$$+\sup_{0\leq t_1\leq T_0}\left|E_1^1\left(p(\underline{z})\phi_1(\underline{z},\hat{\eta}_0)e^{\hat{\underline{\beta}}_1'\underline{z}},t_1\right)-E_1^1\left(p(\underline{z})\phi_1(\underline{z},\eta_0)e^{\underline{\beta}_1'\underline{z}},t_1\right)\right|.$$

The first term tends to zero a.s. by Lemma A1 in Tsiatis (1981). The second term tends to zero in probability since $\hat{\underline{\beta}}\xrightarrow{p}\underline{\beta}$ and since the function

$$\sup_{0\leq t_1\leq T_0}\left|E_1^1\left(p(\underline{z})\phi_1(\underline{z},\tilde{\eta}_0)e^{\tilde{\underline{\beta}}_1'\underline{z}},t_1\right)-E_1^1\left(p(\underline{z})\phi_1(\underline{z},\eta_0)e^{\underline{\beta}_1'\underline{z}},t_1\right)\right|$$

is continuous in $\tilde{\underline{\beta}}=(\tilde{\underline{\beta}}_1,\tilde{\underline{\beta}}_2,\tilde{\underline{\eta}}_0,\tilde{\underline{\gamma}}_0)$. This leads to

$$\sup_{0\leq t_1\leq T_0}\left|\hat{\wedge}_{10}(t_1)-\wedge_{10}(t_1)\right|$$

$$\leq \sup_{0\leq t_1\leq T_0}\left|\int_0^{t_1}\frac{-d\hat{E}_1^1\left(p(\underline{z}),u\right)}{\hat{E}_1^1\left(p(\underline{z})\phi_1(\underline{z},\hat{\eta}_0)e^{\hat{\underline{\beta}}_1'\underline{z}},u\right)}-\int_0^{t_1}\frac{-dE_1^1\left(p(\underline{z}),u\right)}{\hat{E}_1^1\left(p(\underline{z})\phi_1(\underline{z},\hat{\eta}_0)e^{\hat{\underline{\beta}}_1'\underline{z}},u\right)}\right|$$

$$+\sup_{0\leq t_1\leq T_0}\left|\int_0^{t_1}\frac{-dE_1^1\left(p(\underline{z}),u\right)}{\hat{E}_1^1\left(p(\underline{z})\phi_1(\underline{z},\hat{\eta}_0)e^{\hat{\underline{\beta}}_1'\underline{z}},u\right)}-\int_0^{t_1}\frac{-dE_1^1\left(p(\underline{z}),u\right)}{E_1^1\left(p(\underline{z})\phi_1(\underline{z},\eta_0)e^{\underline{\beta}_1'\underline{z}},u\right)}\right|$$

$$\leq \frac{\sup_{0\leq t_1\leq T_0}\left|\hat{E}_1^1\left(p(\underline{z}),t_1\right)-E_1^1\left(p(\underline{z}),t_1\right)\right|}{\hat{E}_1^1\left(p(\underline{z})\phi_1(\underline{z},\hat{\eta}_0)e^{\hat{\underline{\beta}}_1'\underline{z}},T_0\right)}$$

$$+\frac{\sup_{0\leq t_1\leq T_0}\left|\hat{E}_1^1\left(p(\underline{z})\phi_1(\underline{z},\hat{\eta}_0)e^{\hat{\underline{\beta}}_1'\underline{z}},t_1\right)-E_1^1\left(p(\underline{z})\phi_1(\underline{z},\eta_0)e^{\underline{\beta}_1'\underline{z}},t_1\right)\right|}{\hat{E}_1^1\left(p(\underline{z})\phi_1(\underline{z},\hat{\eta}_0)e^{\hat{\underline{\beta}}_1'\underline{z}},T_{10}\right)E_1^1\left(p(\underline{z})\phi_1(\underline{z},\eta_0)e^{\underline{\beta}_1'\underline{z}},T_{10}\right)}$$

which proves the case (i).

Similarly from (5.41), we can have

$$\wedge_{20}(t_1,t_2)=\int_0^{t_2}\frac{-dQ_2(v)}{E_2\left((1+\underline{\gamma}_z)e^{\underline{\beta}_2'\underline{z}^*},t_2\right)}.$$

Proceeding as above, we can prove (ii).

Theorem 5.2. Assume that I is positive definite at $\underline{\beta}$. Assume that

$$E|\log\phi_1(\underline{Z},\eta_0)|,\ E|\log\phi_2(\underline{Z}^*,\gamma_0)|,\ E\left(|\underline{Z}|^5\left(\phi_1(\underline{Z},\eta_0)+1\right)^2 e^{2\underline{\beta}_1'\underline{Z}}\right),$$

$$E\left(|\underline{Z}^*|^5\left(\phi_2(\underline{Z}^*,\gamma_0)+1\right)^2 e^{2\underline{\beta}_2'\underline{Z}^*}\right),\ E\left(\underline{Z}^2\left|\phi_{1j}'(\underline{Z},\eta_0)\right|e^{\underline{\beta}_1'\underline{Z}}\right),$$

$$E\left(\underline{Z}^{*2}\left|\phi_{2j}'(\underline{Z}^*,\gamma_0)\right|e^{\underline{\beta}_2'\underline{Z}^*}\right), E\left(\frac{\phi_{1j}'(\underline{Z},\eta_0)\phi_{1j'}'(\underline{Z},\eta_0)e^{2\underline{\beta}_1'\underline{Z}}}{\phi_1(\underline{Z},\eta_0)^2}\right),$$

$$E\left(\frac{\phi_{2j}'(\underline{Z}^*,\gamma_0)\phi_{2j'}'(\underline{Z}^*,\gamma_0)e^{2\underline{\beta}_2'\underline{Z}^*}}{\phi_2(\underline{Z}^*,\gamma_0)^2}\right),\ E\left(\frac{\underline{Z}^2\left|\phi_{1jj'}''(\underline{Z},\eta_0)\right|}{\phi_{1j}(\underline{Z},\eta_0)}\right) \text{ and } E\left(\frac{\underline{Z}^{*2}\left|\phi_{2jj'}''(\underline{Z}^*,\gamma_0)\right|}{\phi_{2j}(\underline{Z}^*,\gamma_0)}\right)$$

are uniformly bounded in a neighborhood of $\underline{\beta}$ for all $j,j'=1,...,r$. Then the solution $\hat{\underline{\beta}}$ given in Theorem 5.1 is asymptotically normal with mean vector zero and variance covariance matrix I^{-1}, where I is the information matrix of the function H.

Proof. We follow the general approach of Murphy and van der Vaart (2000) for verifying the validity of the profile likelihood method. In particular, we check the conditions of the Theorem 5.1, which guarantees that the profile likelihood allows an asymptotic expansion, which then leads to the asymptotic normality of the maximum likelihood estimator $\hat{\underline{\beta}}$. The rest of the proof follows from Theorem 5.1 given in Braekers and Veraverbeke (2005).

In particular, the covariance matrix is $I=(I_{ij})$, $i,j=1,2....r$ is obtained from the second derivatives of H with respect to the parameters and I is a positive definite matrix by assumption. The second derivatives H are as follows;

$$\frac{\partial^2 H}{\partial\beta_{1j}^{\ 2}}=\int_0^{T_0}\left\{\frac{E_1\left(\left(\phi_1(\underline{z},\eta_0)+1\right)e^{\underline{\beta}_1'\underline{z}},t_1\right)E_1\left(\left(\phi_1(\underline{z},\eta_0)+1\right)e^{\underline{\beta}_1'\underline{z}}\underline{z}\underline{z}',t_1\right)}{-\left[E_1\left(\left(\phi_1(\underline{z},\eta_0)+1\right)e^{\underline{\beta}_1'\underline{z}}\underline{z},t_1\right)\right]^2}\right\}dQ_1(t_1)$$

$$\frac{\partial^2 H}{\partial \beta_{2j}{}^2} = \int_0^{T_0} \left\{ \frac{\begin{array}{l} E_2\left(\left(\phi_2(\underline{z}^*,\gamma_0)+1\right)e^{\underline{\beta}_2{}'\underline{z}^*},t_2\right)E_2\left(\left(\phi_2(\underline{z}^*,\gamma_0)+1\right)e^{\underline{\beta}_2{}'\underline{z}^*}\underline{z}^*\underline{z}^*{}',t_2\right) \\ -\left[E_2\left(\left(\phi_2(\underline{z}^*,\gamma_0)+1\right)e^{\underline{\beta}_2{}'\underline{z}^*}\underline{z}^*,t_2\right)\right]^2 \end{array}}{\left[E_2\left(\left(\phi_2(\underline{z}^*,\gamma_0)+1\right)e^{\underline{\beta}_2{}'\underline{z}^*},t_2\right)\right]^2} \right\} dQ_2(t_2)$$

$$\frac{\partial^2 H}{\partial \eta_{0j}{}^2} = E_1{}^1\left(\frac{\phi_1(\underline{z},\eta_0)\phi_{1j}{}''(\underline{z},\eta_0)-\phi_{1j}{}'(\underline{z},\eta_0)^2}{\phi_1(\underline{z},\eta_0)^2},0\right)$$

$$+\int_0^{T_0} \left\{ \frac{\begin{array}{l} E_1\left(\left(\phi_1(\underline{z},\eta_0)+1\right)e^{\underline{\beta}_1{}'\underline{z}},t_1\right)E_1\left(\phi_{1j}{}''(\underline{z},\eta_0)e^{\underline{\beta}_1{}'\underline{z}},t_1\right) \\ -\left[E_1\left(\phi_{1j}{}'(\underline{z},\eta_0)e^{\underline{\beta}_1{}'\underline{z}},t_1\right)\right]^2 \end{array}}{\left[E_1\left(\left(\phi_1(\underline{z},\eta_0)+1\right)e^{\underline{\beta}_1{}'\underline{z}},t_1\right)\right]^2} \right\} dQ_1(t_1)$$

$$\frac{\partial^2 H}{\partial \gamma_{0j}{}^2} = E_1{}^2\left(\frac{\phi_2(\underline{z}^*,\gamma_0)\phi_{2j}{}''(\underline{z}^*,\gamma_0)-\phi_{2j}{}'(\underline{z}^*,\gamma_0)^2}{\phi_2(\underline{z}^*,\gamma_0)^2},0\right)$$

$$+\int_0^{T_0} \left\{ \frac{\begin{array}{l} E_2\left(\left(\phi_2(\underline{z}^*,\gamma_0)+1\right)e^{\underline{\beta}_2{}'\underline{z}^*},t_2\right)E_2\left(\left(\phi_{2j}{}''(\underline{z}^*,\gamma_0)+1\right)e^{\underline{\beta}_2{}'\underline{z}^*},t_2\right) \\ -\left[E_2\left(\left(\phi_{2j}{}'(\underline{z}^*,\gamma_0)+1\right)e^{\underline{\beta}_2{}'\underline{z}^*},t_2\right)\right]^2 \end{array}}{\left[E_2\left(\left(\phi_2(\underline{z}^*,\gamma_0)+1\right)e^{\underline{\beta}_2{}'\underline{z}^*},t_2\right)\right]^2} \right\} dQ_2(t_2)$$

$$\frac{\partial^2 H(\underline{\beta})}{\partial \beta_{1i}\partial \eta_{0j}} = \int_0^{T_0} \left\{ \frac{\begin{array}{l} E_1\left(\left(\phi_1(\underline{z},\eta_0)+1\right)e^{\underline{\beta}_1{}'\underline{z}},t_1\right)E_1\left(\left(\phi_{1j}{}'(\underline{z},\eta_0)+1\right)e^{\underline{\beta}_1{}'\underline{z}}\underline{z},t_1\right) \\ -E_1\left(\left(\phi_1(\underline{z},\eta_0)+1\right)e^{\underline{\beta}_1{}'\underline{z}}\underline{z},t_1\right)E_1\left(\left(\phi_{1j}{}'(\underline{z},\eta_0)+1\right)e^{\underline{\beta}_1{}'\underline{z}},t_1\right) \end{array}}{\left[E_1\left(\left(\phi_1(\underline{z},\eta_0)+1\right)e^{\underline{\beta}_1{}'\underline{z}},t_1\right)\right]^2} \right\} dQ_1(t_1)$$

$$\frac{\partial^2 H(\underline{\beta})}{\partial \beta_{2i} \partial \gamma_{0j}} =$$

$$\int_0^{T_0} \left\{ \frac{\begin{array}{l} E_2\left(\left(\phi_2(\underline{z}^*, \gamma_0)+1\right) e^{\underline{\beta}_2' \underline{z}^*}, t_2\right) E_2\left(\left(\phi_{2j}'(\underline{z}^*, \gamma_0)+1\right) e^{\underline{\beta}_2' \underline{z}^*} \underline{z}^*, t_2\right) \\ -E_2\left(\left(\phi_2(\underline{z}^*, \gamma_0)+1\right) e^{\underline{\beta}_2' \underline{z}^*} \underline{z}^*, t_2\right) E_2\left(\left(\phi_{2j}'(\underline{z}^*, \gamma_0)+1\right) e^{\underline{\beta}_2' \underline{z}^*}, t_2\right) \end{array}}{\left[E_2\left(\left(\phi_2(\underline{z}^*, \gamma_0)+1\right) e^{\underline{\beta}_2' \underline{z}^*}, t_2\right)\right]^2} \right\} dQ_2(t_2)$$

$$\frac{\partial^2 H}{\partial \beta_{1i} \partial \beta_{2j}} = 0 = \frac{\partial^2 H}{\partial \beta_{2i} \partial \beta_{1j}} \text{ and } \frac{\partial^2 H}{\partial \eta_{0i} \partial \gamma_{0j}} = 0 = \frac{\partial^2 H}{\partial \gamma_{0i} \partial \eta_{0j}}.$$

The estimates of the covariance matrix, $\hat{I}$ is obtained by substituting the unknown quantities by their estimators. The well known functional delta method can be used to estimate the variance of the estimates (van der Vaart and Wellner, 1996). Obviously, the expressions for the variance of the estimates are complicated. In practical situations, an attractive approach to variance or confidence interval estimation is through resampling methods. The naive bootstrap procedure of resampling the observed data units $(T_{1i}, T_{2i}, \delta_i, Z_i)$ with replacement (see Efron and Tibshirani, 1993) will be satisfactory under fairly mild conditions.

5.5. Simulation Study

In this section, we carried out a simulation study to evaluate the performance of the aforementioned inference procedures. We consider Arnold and Strauss (1988) bivariate exponential distribution with joint density function

$$f(t_1, t_2) = C \exp(-t_1 - t_2 - \gamma t_1 t_2),\ t_1, t_2 > 0,\ 0 \le \gamma \le 1 \qquad (5.45)$$

where C is the normalizing constant.

We generated a single covariate z from uniform (0, 1) distribution. We generated observations from Arnold and Strauss (1988) bivariate exponential distribution for different values of γ.

We consider $\underline{\eta}_z = \underline{\gamma}_z = e^{a+bz}$. We first estimate a and b and compute the average bias and variance of the estimates, for 1000 simulations, those are given in Table 5.1 and Table 5.5. We then compute estimates of $\underline{\beta}_1$ and $\underline{\beta}_2$ and then calculate average

bias and variance of the estimates of β_1 and β_2, which are given in Table 5.2 and Table 5.6. The average bias and variance of estimates of baseline cumulative hazard functions are given in Table 5.3 and Table 5.7 and those of bivariate survival function are given in Table 5.4 and Table 5.8. As n increases, both bias and variance of the estimates decreases.

Table 5.1. Bias and variance of $\hat{a}$ and $\hat{b}$ when $a=0.5$ and $b=1.5$

β_1	β_2	γ	n	Bias $\hat{a}$	Var $\hat{a}$	Bias $\hat{b}$	Var $\hat{b}$
1.5	0.9	0.7	50	0.0432	0.0370	-0.0405	0.0276
			250	-0.0324	0.0119	0.0110	0.0128
		0.9	50	0.0454	0.0281	0.0129	0.0384
			250	0.0199	0.0189	0.0104	0.0113
0.8	-1	0.7	50	0.0314	0.0370	0.0239	0.0314
			250	0.0145	0.0178	0.0196	0.0212
		0.9	50	-0.0323	0.0287	-0.0375	0.0261
			250	-0.0191	0.0139	-0.0334	0.0129

Table 5.2. Bias and variance of $\hat{\beta}_1$ and $\hat{\beta}_2$ when $a=0.5$ and $b=1.5$

β_1	β_2	γ	n	Bias $\hat{\beta}_1$	Var $\hat{\beta}_1$	Bias $\hat{\beta}_2$	Var $\hat{\beta}_2$
1.5	0.9	0.7	50	-0.0171	0.0383	-0.0344	0.0229
			250	0.0149	0.0179	0.0305	0.0208
		0.9	50	0.0197	0.0321	-0.0370	0.0253
			250	0.0168	0.0286	0.0340	0.0191
0.8	-1	0.7	50	-0.0409	0.0133	0.0181	0.0274
			250	-0.0376	0.0115	0.0149	0.0167
		0.9	50	0.0364	0.0177	-0.0181	0.0289
			250	-0.0248	0.0147	0.0162	0.0117

Table 5.3. Bias and variance of baseline cumulative hazard functions when $a = 0.5$ and $b = 1.5$

β_1	β_2	(t_1, t_2)	γ	n	Bias $\hat{\Lambda}_{10}(t_1)$	Var $\hat{\Lambda}_{10}(t_1)$	Bias $\hat{\Lambda}_{20}(t_1, t_2)$	Var $\hat{\Lambda}_{20}(t_1, t_2)$
1.5	0.9	(1,1.5)	0.7	50	-0.0395	0.0433	-0.0254	0.0239
				250	0.0251	0.0306	0.0244	0.0167
			0.9	50	0.0314	0.0291	-0.0285	0.0134
				250	0.0119	0.0109	0.0260	0.0095
		(2,1.1)	0.7	50	-0.0164	0.0157	-0.0264	0.0263
				250	0.0148	0.0118	0.0213	0.0035
			0.9	50	-0.0158	0.0299	-0.0308	0.0287
				250	0.0147	0.0209	-0.0200	0.0127
0.8	-1	(1,1.5)	0.7	50	0.0304	0.0221	0.0255	0.0112
				250	0.0199	0.0192	0.0206	0.0017
			0.9	50	0.0337	0.0214	0.0283	0.0090
				250	-0.0157	0.0164	-0.0180	0.0062
		(2,1.1)	0.7	50	0.0143	0.0134	-0.0257	0.0070
				250	-0.0119	0.0113	0.0152	0.0042
			0.9	50	0.0145	0.0452	0.0308	0.0021
				250	-0.0137	0.0235	0.0206	0.0015

Table 5.4. Bias and variance of bivariate survival function when $a = 0.5$ and $b = 1.5$

β_1	β_2	(t_1, t_2)	γ	n	Bias $\hat{S}(t_1, t_2)$	Var $\hat{S}(t_1, t_2)$
1.5	0.9	(1,1.5)	0.7	50	0.0301	0.0366
				250	0.0296	0.0243
			0.9	50	0.0367	0.0254
				250	0.0302	0.0121
		(2,1.1)	0.7	50	0.0401	0.0218
				250	0.0356	0.0100
			0.9	50	0.0378	0.0198
				250	0.0322	0.0142
0.8	-1	(1,1.5)	0.7	50	0.0299	0.0254
				250	0.0283	0.0122
			0.9	50	0.0351	0.0187
				250	0.0316	0.0100
		(2,1.1)	0.7	50	0.0300	0.0123
				250	0.0290	0.0098
			0.9	50	0.0288	0.0172
				250	0.0275	0.0128

Table 5.5. Bias and variance of $\hat{a}$ and $\hat{b}$ when $a=1$ and $b=0.8$

β_1	β_2	γ	n	Bias $\hat{a}$	Var $\hat{a}$	Bias $\hat{b}$	Var $\hat{b}$
1.5	0.9	0.7	50	-0.0247	0.0291	0.0163	0.0359
			250	0.0100	0.0268	0.0137	0.0116
		0.9	50	0.0150	0.0202	0.0364	0.0252
			250	-0.0108	0.0193	-0.0163	0.0199
0.8	-1	0.7	50	0.0362	0.0187	-0.0144	0.0284
			250	0.0306	0.0130	-0.0135	0.0198
		0.9	50	0.0392	0.0262	-0.0200	0.0395
			250	0.0323	0.0135	0.0160	0.0281

Table 5.6. Bias and variance of $\hat{\beta}_1$ and $\hat{\beta}_2$ when $a=1$ and $b=0.8$

β_1	β_2	γ	n	Bias $\hat{\beta}_1$	Var $\hat{\beta}_1$	Bias $\hat{\beta}_2$	Var $\hat{\beta}_2$
1.5	0.9	0.7	50	0.0165	0.0241	0.0325	0.0218
			250	0.0151	0.0225	-0.0307	0.0195
		0.9	50	-0.0157	0.0279	0.0347	0.0278
			250	0.0102	0.0253	0.0324	0.0249
0.8	-1	0.7	50	-0.0319	0.0194	0.0350	0.0335
			250	0.0247	0.0147	-0.0335	0.0156
		0.9	50	-0.0299	0.0252	0.0254	0.0143
			250	-0.0178	0.0129	0.0148	0.0106

Table 5.7. Bias and variance of baseline cumulative hazard functions when $a=1$ and $b=0.8$

β_1	β_2	(t_1,t_2)	γ	n	Bias $\hat{\Lambda}_{10}(t_1)$	Var $\hat{\Lambda}_{10}(t_1)$	Bias $\hat{\Lambda}_{20}(t_1,t_2)$	Var $\hat{\Lambda}_{20}(t_1,t_2)$
1.5	0.9	(1,1.5)	0.7	50	0.0297	0.0322	-0.0255	0.0363
				250	0.0169	0.0296	0.0150	0.0236
			0.9	50	-0.0261	0.0187	0.0285	0.0299
				250	-0.0127	0.0156	0.0185	0.0231
		(2,1.1)	0.7	50	0.0158	0.0272	0.0264	0.0258
				250	0.0148	0.0235	-0.0216	0.0113
			0.9	50	0.0138	0.0226	-0.0308	0.0265
				250	-0.0120	0.0164	0.0201	0.0107
0.8	-1	(1,1.5)	0.7	50	0.0378	0.0492	0.0255	0.0368
				250	0.0216	0.0308	0.0250	0.0339
			0.9	50	-0.0299	0.0282	-0.0295	0.0107
				250	0.0174	0.0137	-0.0280	0.0012
		(2,1.1)	0.7	50	0.0168	0.0337	0.0264	0.0356
				250	0.0103	0.0205	-0.0164	0.0197
			0.9	50	-0.0146	0.0231	0.0308	0.0260
				250	0.0119	0.0177	-0.0300	0.0223

Table 5.8. Bias and variance of fivariate survival function when $a=1$ and $b=0.8$

β_1	β_2	(t_1,t_2)	γ	n	Bias $\hat{S}(t_1,t_2)$	Var $\hat{S}(t_1,t_2)$
1.5	0.9	(1,1.5)	0.7	50	0.0389	0.0247
				250	0.0247	0.0229
			0.9	50	0.0301	0.0214
				250	0.0292	0.0210
		(2,1.1)	0.7	50	0.0258	0.0359
				250	0.0221	0.0322
			0.9	50	-0.0298	0.0301
				250	0.0252	0.0254
0.8	-1	(1,1.5)	0.7	50	0.0295	0.0347
				250	0.0243	0.0340
			0.9	50	0.0326	0.0258
				250	-0.0301	0.0197
		(2,1.1)	0.7	50	0.0333	0.0219
				250	0.0298	0.0156
			0.9	50	0.0248	0.0300
				250	0.0215	0.0271

5.6. Data Analysis

For the illustration of the estimation procedure we consider an example of transfusion-related AIDS data given in Kalbfleish and Prentice (2002, page 385). The data gives the infection (transfusion) time in months with 1= January 1978, incubation time (time from infection to diagnosis of AIDS measured in months from time of infection) and the age of individuals at time of infection. We take T_1 as transfusion time, T_2 as incubation time and age as covariate. We consider $\underline{\eta}_z = \underline{\gamma}_z = e^{a+bz}$.

We compute the estimates $\hat{a}$, $\hat{b}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ by the method given in Section 5.3. The estimates of a, b, β_1 and β_2 are $\hat{a} = -101.253$, $\hat{b} = 0.0009$, $\hat{\beta}_1 = 0.0027$ and $\hat{\beta}_2 = -0.0070$. We then estimate $\wedge_{10}(t_1)$, $\wedge_{20}(t_1, t_2)$ and $S(t_1, t_2)$, those are given in Table 5.9. We can observe that $\hat{\beta}_2$ has negative effect on the incubation time and $\hat{\beta}_1$ has positive effect on the transfusion time. From Table 5.9, it follows that as t_1 increases, the estimate $\hat{\wedge}_{10}(t_1)$ increases, but there is no specific pattern for both the estimates $\hat{\wedge}_2(t_1, t_2)$ and $\hat{S}(t_1, t_2)$ and it depends on both t_1 and t_2. Figure 5.1 shows the estimates of the bivariate survival function.

Table 5.9. Estimates of bivariate survival function

(t_1,t_2)	z	$\hat{\Lambda}_{10}(t_1)$	$\hat{\Lambda}_{20}(t_1,t_2)$	$\hat{S}(t_1,t_2)$
(87,9)	57	3.9516	0.0113	0.0098
(12,60)	21	0.0186	3.5073	0.0475
(17,53)	33	0.0328	2.9462	0.0932
(27,59)	60	0.0969	3.0355	0.1217
(72,24)	44	1.6988	0.2519	0.1223
(71,29)	44	1.6351	0.2853	0.1283
(36,60)	66	0.2258	2.6349	0.1454
(71,8)	1	1.6351	0.0455	0.1855
(67,14)	45	1.3526	0.1865	0.1890
(65,23)	66	1.1706	0.3956	0.1918
(58,41)	70	0.8326	0.9174	0.2082
(52,48)	56	0.5696	1.2948	0.2149
(33,41)	67	0.1605	1.9092	0.2502
(40,39)	50	0.2968	1.4763	0.2517
(58,13)	39	0.8326	0.2335	0.3316
(50,24)	52	0.5520	0.6552	0.3358
(56,17)	66	0.7220	0.3244	0.3433
(23,27)	4	0.0717	0.9121	0.3832
(48,15)	63	0.4976	0.3140	0.4525
(36,18)	65	0.2258	0.4751	0.5651

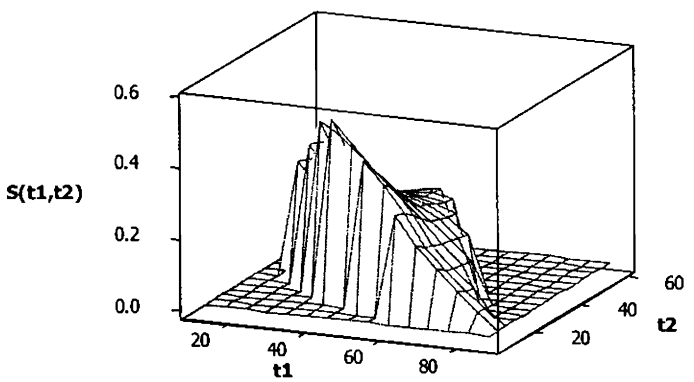

Figure 5.1. Estimates of bivariate survival function

5.7. Conclusion

We introduced proportional hazards model for duration times under informative censoring. Estimators of the parameters and baseline hazard functions are developed and properties of the estimators are discussed. A simulation study is conducted to assess the performance of the estimators. We illustrated the procedure using a real life data. The proposed method is an extension of the work done in Wang and Wells (1998) to the situation where covariate is present, using the idea given in Braekers and Veraverbeke (2005). Obviously, the proposed method depends on the choice of $\underline{\eta}_z$ and $\underline{\gamma}_z$. In a practical situation, one should find optimal choice for $\underline{\eta}_z$ and $\underline{\gamma}_z$, which is subject for future research.

Chapter Six

Proportional Hazards Model for Bivariate Competing Risks Data

6.1. Introduction

In survival studies, when covariates are present, the standard analysis for competing risks data involves modeling the cause-specific hazard functions of the different failure types through proportional hazards assumption (see Lanson 1984; Kalbfleisch and Prentice, 2002 and Lawless, 2003). Gelfand et al. (2000) proposed a modified semi-parametric version of the proportional hazards models which include an arbitrary rich class of continuous baseline hazards, an attractive epidemiological interpretation of the hazard as a latent competing risks model and trivial handling of censoring. Fiocco et al. (2005) introduced a reduced rank proportional hazards model for competing risks and describe an algorithm for estimating the parameters of the model. Recently Gichangi and Vach (2005) provided a guided tour in analyzing competing risks data in medical research.

The failure of systems in the multivariate situations can also classify into different modes. Accordingly, statistical analysis of multivariate competing risks models is a topic of recent interest in survival analysis (see DeMasi, 2000). Non-parametric estimations of bivariate survival function and cause specific distributions under censoring were recently developed by Antony and Sankaran (2005), Sankaran et al. (2006), Sankaran and Antony (2008a) and Sankaran and Antony (2008b). The analysis of multivariate competing risks data in the presence of covariates is not yet discussed in literature. Motivated by this, we introduce multivariate proportional hazards models for the analysis of competing risks data in the presence of censoring. For simplicity, we confine our study to bivariate set up.

The chapter is organized as follows. In Section 6.2, we introduce

---

The results in this Chapter have been communicated as entitled "Proportional Hazards Model for Bivariate Competing Risks Data" (see Sreeja and Sankaran (2008b)).

proportional hazards models for bivariate competing risks data using vector hazard function of Dabrowska (1988). Estimation of the parameters as well as the cause-specific hazard function is done in Section 6.3. In Section 6.4, various properties of the estimators are discussed. A simulation study is reported to assess the performance of the estimator in Section 6.5. In Section 6.6, we apply the models to a real life data that concerning the times to tumor appearance or death for 100 pairs of mice. Finally, the study is concluded in Section 6.7.

6.2. Basic Concepts and Model

Let $T=(T_1,T_2)$ be a random vector representing the lifetime of pair of individuals. Let $S(t_1,t_2)=P[T_1\geq t_1,T_2\geq t_2]$ be the joint survival function of T. Then Dabrowska (1988) defined a cumulative hazard function vector of T as given in (1.35).

Let $C=(C_1,C_2)$ be a set of causes corresponding to the lifetime vector $T=(T_1,T_2)$. Suppose that there are k_1 causes of failure for T_1 and k_2 causes of failure for T_2. Then the cause-specific hazard functions are given by

$$\lambda_{1p}(t_1,t_2)=\lim_{h\to 0}\frac{1}{h}P[t_1\leq T_1<t_1+h,C_1=p\mid T_1\geq t_1,T_2>t_2] \tag{6.1}$$

$$\lambda_{2q}(t_1,t_2)=\lim_{h\to 0}\frac{1}{h}P[t_2\leq T_2<t_2+h,C_2=q\mid T_1>t_1,T_2\geq t_2] \tag{6.2}$$

and

$$\lambda_{3pq}(t_1,t_2)=\lim_{h\to 0}\frac{1}{h}P[t_1\leq T_1<t_1+h,t_2\leq T_2<t_2+h,C_1=p,C_2=q\mid T_1\geq t_1,T_2\geq t_2],$$

$$p=1,2,...,k_1,\ q=1,2,...,k_2. \tag{6.3}$$

Assume that the failure cause C_i must be a unique element of $\{1,2,...,k_i\}, i=1,2$. Then the cumulative hazard functions in (1.36), (1.37) and (1.38) can be expressed in terms of cause-specific cumulative hazard functions as given by

$$\wedge_1(t_1,t_2)=\sum_{p=1}^{k_1}\wedge_{1p}(t_1,t_2)=\sum_{p=1}^{k_1}\int_0^{t_1}\lambda_{1p}(u,t_2)du \tag{6.4}$$

$$\wedge_2(t_1,t_2)=\sum_{q=1}^{k_2}\wedge_{2q}(t_1,t_2)=\sum_{q=1}^{k_2}\int_0^{t_2}\lambda_{2q}(t_1,v)dv \tag{6.5}$$

and

$$\wedge_{3pq}(t_1,t_2)=\sum_{p=1}^{k_1}\sum_{q=1}^{k_2}\wedge_{3pq}(t_1,t_2)=\sum_{p=1}^{k_1}\sum_{q=1}^{k_2}\int_0^{t_1}\int_0^{t_2}\lambda_{3pq}(u,v)dudv\,. \tag{6.6}$$

Then the bivariate survival function can be written as

$$S(t_1,t_2)=\exp[-\wedge_1(t_1,0)-\wedge_2(t_1,t_2)] \tag{6.7}$$

or

$$S(t_1,t_2)=\exp[-\wedge_1(t_1,t_2)-\wedge_2(0,t_2)]\,. \tag{6.8}$$

The cause-specific sub-distribution function $F_{pq}(t_1,t_2)$ in the bivariate case is defined as

$$F_{pq}(t_1,t_2)=P(T_1\le t_1,T_2\le t_2,C_1=p,C_2=q),\ p=1,2,...,k_1,\ q=1,2,...,k_2\,. \tag{6.9}$$

$F_{pq}(t_1,t_2)$ measures the probability that the failure of both the study subjects (T_1,T_2) due to the causes (p,q) prior to (t_1,t_2). In mortality studies, (6.9) is helpful to compare whether death of one is important for the partner's risk of death of other causes.

We can write (6.9) in terms of cumulative cause-specific hazard function as

$$\wedge_{3pq}(dt_1,dt_2)=\frac{F_{pq}(dt_1,dt_2)}{S(t_1,t_2)} \tag{6.10}$$

which provides

$$F_{pq}(t_1,t_2)=\int_0^{t_1}\int_0^{t_2}S(u,v)\wedge_{3pq}(du,dv)\,. \tag{6.11}$$

With covariates, one possible technique for the analysis of bivariate competing risks data is to model marginal cause-specific hazard functions for T_1 and T_2 and then apply ideas from generalized estimating functions to calculate an appropriate combination of the two marginal estimates. This can be done in the case of homogeneity of the two regression coefficients.

We now consider a different approach for modeling bivariate competing risks data using vector hazard function of Dabrowska (1988).

We define proportional hazards models for (T_1, T_2) as

$$\lambda_{1p}(t_1, t_2 \mid \underline{z}) = \lambda_{1p0}(t_1, t_2) e^{\underline{\beta}_{1p}' \underline{z}} \tag{6.12}$$

$$\lambda_{2q}(t_1, t_2 \mid \underline{z}) = \lambda_{2q0}(t_1, t_2) e^{\underline{\beta}_{2q}' \underline{z}} \tag{6.13}$$

and

$$\lambda_{3pq}(t_1, t_2 \mid \underline{z}) = \lambda_{3pq0}(t_1, t_2) e^{\underline{\beta}_{3pq}' \underline{z}},\ p = 1, 2, ..., k_1,\ q = 1, 2, ..., k_2. \tag{6.14}$$

In the models (6.12), (6.13) and (6.14), $\lambda_{1p}(t_1, t_2 \mid \underline{z})$, $\lambda_{2q}(t_1, t_2 \mid \underline{z})$ and $\lambda_{3pq}(t_1, t_2 \mid \underline{z})$ are the cause-specific hazard functions of $T = (T_1, T_2)$ with $r \times 1$ covariate vector $\underline{z}$. $\lambda_{1p0}(t_1, t_2)$, $\lambda_{2q0}(t_1, t_2)$ and $\lambda_{3pq0}(t_1, t_2 \mid \underline{z})$ are the corresponding baseline cause-specific hazard functions and $\underline{\beta}_{1p}$, $\underline{\beta}_{2q}$ and $\underline{\beta}_{3pq}$ are $r \times 1$ vector of parameters.

Due to consistency conditions among bivariate vector hazard functions,

$$\frac{\partial \lambda_i(t_1, t_2)}{\partial t_j} = \lambda_1(t_1, t_2) \lambda_2(t_1, t_2) - \lambda_3(t_1, t_2),\ i, j = 1, 2; i \neq j,$$

we should have $\underline{\beta}_{3pq} = \underline{\beta}_{1p} + \underline{\beta}_{2q}$. When $\underline{\beta}_i$, $i, = 1, 2$ is a zero vector, the covariates has no effect on the hazard functions.

The model (6.12) means that for $T_2 \geq t_2$, the ratio $\dfrac{\lambda_{1p}(t_1, t_2 \mid \underline{z}^{(1)})}{\lambda_{1p}(t_1, t_2 \mid \underline{z}^{(2)})}$ of the cause-specific hazard functions of pair of two individuals with covariate vectors $\underline{z}^{(1)}$ and $\underline{z}^{(2)}$ does not vary with t_1 and t_2. Similar interpretation can be given to the models (6.13) and (6.14). The vector $\underline{\beta}_{1p}$ depends on cause p and $\underline{\beta}_{2q}$ depends on cause q. Thus the covariates under study have different effect on the components of the vector hazard function.

Under the model (6.12) and (6.13), the bivariate survival function can be written as

$$S(t_1, t_2 \mid \underline{z}) = \exp[-\sum_{p=1}^{k_1} \wedge_{1p0}(t_1, 0) e^{\underline{\beta}_{1p}' \underline{z}} - \sum_{q=1}^{k_2} \wedge_{2q0}(t_1, t_2) e^{\underline{\beta}_{2q}' \underline{z}}] \tag{6.15}$$

and

$$S(t_1,t_2 \mid \underline{z}) = \exp[-\sum_{p=1}^{k_1} \wedge_{1p0}(t_1,t_2)e^{\underline{\beta}_{1p}'\underline{z}} - \sum_{q=1}^{k_2} \wedge_{2q0}(0,t_2)e^{\underline{\beta}_{2q}'\underline{z}}] \quad (6.16)$$

where $\wedge_{1p0}(t_1,t_2) = \int_0^{t_1} \lambda_{1p0}(u,t_2)du$ and $\wedge_{2q0}(t_1,t_2) = \int_0^{t_2} \lambda_{2q0}(t_1,v)dv$.

Then the expression for cause-specific sub-distribution function can be written as

$$F_{pq}(t_1,t_2 \mid \underline{z}) = \int_0^{t_1}\int_0^{t_2} S(u,v \mid \underline{z}) \wedge_{3pq}(du,dv \mid \underline{z}). \quad (6.17)$$

6.3. Estimation Procedures

Suppose now that there are n independent pairs of subjects in the study so that $(t_{1i},t_{2i},\delta_{1i},\delta_{2i},C_{1i},C_{2i},\underline{z}_i), i=1,2,...,n$ are n i.i.d. replicates of $(t_1,t_2,\delta_1,\delta_2,C_1,C_2,\underline{z})$ were $t_j = \min(T_j,L_j)$, $j=1,2$ with L_1 and L_2 are censoring times corresponding to T_1 and T_2 and $\delta_1 = I(T_1 \le L_1)$ and $\delta_2 = I(T_2 \le L_2)$ are censoring indicators. To estimate the parameter vectors $\underline{\beta}_{1p}$, $p=1,2,...,k_1$, we have the score function for $\underline{\beta}_{1p}$ as

$$U_{1p}(\underline{\beta}_{1p}) = \sum_{j=1}^{n} \delta^*_{1pj}\left[\underline{z}_j - \overline{z}_{1p}(t_{1j},t_{2j},\underline{\beta}_{1p})\right] \quad (6.18)$$

where

$$\overline{z}_{1p}(t_{1j},t_{2j},\underline{\beta}_{1p}) = \frac{\sum_{l=1}^{n} Y_l(t_{1j},t_{2j})e^{\underline{\beta}_{1p}'\underline{z}_l}\underline{z}_l}{\sum_{l=1}^{n} Y_l(t_{1j},t_{2j})e^{\underline{\beta}_{1p}'\underline{z}_l}}, \quad p=1,2,...,k_1,$$

$Y_l(t_1,t_2) = I(t_{1l} \ge t_1, t_{2i} \ge t_2)$, $l=1,2,...,n$ and δ^*_{1pl} is the censoring indicator, $\delta^*_{1pl} = I(t_{1l} = T_{1l}, C_{1l} = p)$, $l=1,2,...,n$.

We now reformulate $U_{1p}(\underline{\beta}_{1p})$ in terms of counting processes. Let

$$N_{1pi}(t_1,t_2) = I\left(T_{1i} \le t_1, T_{2i} \ge t_2, \delta_{1pi} = 1\right), \quad p=1,2,...,k_1, \; i=1,2,...,n. \quad (6.19)$$

Then (6.18) becomes

$$U_{1p}(\underline{\beta}_{1p}) = \sum_{j=1}^{n} \int_0^{\infty} \left[\underline{z}_j - \overline{z}_{1p}(u,t_2,\underline{\beta}_{1p})\right] dN_{1pi}(u,t_{2j}) \quad (6.20)$$

Similarly, the score functions for $\underline{\beta}_{2q}$ is

$$U_{2q}(\underline{\beta}_{2q}) = \sum_{j=1}^{n} \delta^*_{2qj} \left[\underline{z}_j - \overline{z}_{2q}(t_{1j}, t_{2j}, \underline{\beta}_{2q}) \right] \tag{6.21}$$

where

$$\overline{z}_{2q}(t_{1j}, t_{2j}, \underline{\beta}_{2q}) = \frac{\sum_{l=1}^{n} Y_l(t_{1j}, t_{2j}) e^{\underline{\beta}_{2q}' \underline{z}_l} \underline{z}_l}{\sum_{l=1}^{n_2} Y_l(t_{1j}, t_{2j}) e^{\underline{\beta}_{2q}' \underline{z}_l}}, q = 1, 2, ..., k_2$$

and $\delta^*_{2qj} = I(t_{2j} = T_{2j}, C_{2j} = q)$ is the censoring indicator.

In terms of counting processes $U_{2q}(\underline{\beta}_{2q})$ will be

$$U_{2q}(\underline{\beta}_{2q}) = \sum_{j=1}^{n} \int_0^{\infty} \left[\underline{z}_j - \overline{z}_{2q}(t_{1j}, v, \underline{\beta}_{2q}) \right] dN_{2qi}(t_{1j}, v) \tag{6.22}$$

with

$$N_{2qi}(t_1, t_2) = I\left(T_{1i} \geq t_1, T_{2i} \leq t_2, \delta_{2qi} = 1\right), \; q = 1, 2, ..., k_2, \; i = 1, 2, ..., n. \tag{6.23}$$

Then we can write the score functions for $\underline{\beta}_1$ and $\underline{\beta}_2$ are,

$$U_1(\underline{\beta}_{1p}) = \sum_{j=1}^{n} \delta^*_{1pj} \left[\underline{z}_j - \overline{z}_{1p}(t_{1j}, t_{2j}, \underline{\beta}_{1p}) \right] \tag{6.24}$$

and

$$U_2(\underline{\beta}_{2q}) = \sum_{j=1}^{n} \delta^*_{2qj} \left[\underline{z}_j - \overline{z}_{2q}(t_{1j}, t_{2j}, \underline{\beta}_{2q}) \right], \; p = 1, 2, ..., k_1, \; q = 1, 2, ..., k_2. \tag{6.25}$$

The maximum likelihood estimators $\hat{\underline{\beta}}_{1p}$ and $\hat{\underline{\beta}}_{2q}$ for $\underline{\beta}_{1p}$ and $\underline{\beta}_{2q}$ are solutions of $U_1(\underline{\beta}_{1p}) = 0$ and $U_2(\underline{\beta}_{2q}) = 0$.

We can find the generalized Nelson-Aalen estimate of the baseline cumulative cause-specific hazard functions as

$$\hat{\Lambda}_{1p0}(t_1, t_2) = \sum_{j: t_{1j} \leq t_1, t_{2j} \leq t_2} \left[\frac{\delta^*_{1pj}}{\sum_{l=1}^{n} Y_l(t_{1j}, t_{2j}) e^{\hat{\underline{\beta}}_{1p}' \underline{z}_l}} \right] \tag{6.26}$$

$$\hat{\Lambda}_{2q0}(t_1,t_2)=\sum_{j:t_{1j}\le t_1,t_{2j}\le t_2}\left[\frac{\delta^*_{2qj}}{\sum_{l=1}^{n}Y_l(t_{1j},t_{2j})e^{\hat{\underline{\beta}}_{2q}'\underline{z}_l}}\right] \tag{6.27}$$

and

$$\hat{\Lambda}_{3pq0}(t_1,t_2)=\sum_{j:t_{1j}\le t_1,t_{2j}\le t_2}\left[\frac{\delta^*_{3pqj}}{\sum_{l=1}^{n}Y_l(t_{1j},t_{2j})e^{\hat{\underline{\beta}}_{1p}'\underline{z}_l+\hat{\underline{\beta}}_{2q}'\underline{z}_l}}\right] \tag{6.28}$$

with

$\delta^*_{3pqj}=I(t_{1j}=T_{1j},C_{1j}=p,t_{2j}=T_{2j},C_{2j}=q)$.

One may often be interested in estimating the survival function $S(t_1,t_2\mid\underline{z}_0)$ of lifetimes with a fixed covariate $\underline{z}_0$. A natural estimator for $S(t_1,t_2\mid\underline{z}_0)$ is given by

$$\hat{S}_1(t_1,t_2\mid\underline{z}_0)=\exp[-\hat{\Lambda}_{1p}(t_1,0)e^{\hat{\underline{\beta}}_{1p}'\underline{z}_0}-\hat{\Lambda}_{2q}(t_1,t_2)e^{\hat{\underline{\beta}}_{2q}'\underline{z}_0}] \tag{6.29}$$

and

$$\hat{S}_2(t_1,t_2\mid\underline{z}_0)=\exp[-\hat{\Lambda}_{1p}(t_1,t_2)e^{\hat{\underline{\beta}}_{1p}'\underline{z}_0}-\hat{\Lambda}_{2q}(0,t_2)e^{\hat{\underline{\beta}}_{2q}'\underline{z}_0}] \tag{6.30}$$

The estimator of the survival function $S(t_1,t_2\mid\underline{z}_0)$ obtained by (6.29) and (6.30) may be different. So as in Section 2.4, to get a unique estimator, we consider a convex combination of the two expressions (6.29) and (6.30). Thus the estimator of $S(t_1,t_2\mid\underline{z})$ is given by

$$\hat{S}(t_1,t_2\mid\underline{z}_0)=a(t_1,t_2)\hat{S}_1(t_1,t_2\mid\underline{z}_0)+(1-a(t_1,t_2))\hat{S}_2(t_1,t_2\mid\underline{z}_0). \tag{6.31}$$

Thus $a(t_1,t_2)$ which minimized the mean square error is

$$a(t_1,t_2)=\frac{\sigma_{22}-\sigma_{12}+\mu_2^2-\mu_1\mu_2}{\sigma_{11}+\sigma_{22}-2\sigma_{12}+\mu_1^2+\mu_2^2-2\mu_1\mu_2} \tag{6.32}$$

where, σ_{ij} is the asymptotic covariance between $\hat{S}_i(t_1,t_2)$ and $\hat{S}_j(t_1,t_2)$ and μ_j is the asymptotic bias of $\hat{S}_j(t_1,t_2)$, $i,\ j=1,2, i\neq j$.

To ensure that $\hat{S}(t_1,t_2\mid\underline{z})$ belongs to the interval [0, 1], we replace $a(t_1,t_2)$ by $\min[1,\max(a(t_1,t_2),0)]$.

In practice, we estimate $a(t_1,t_2)$ using the variance and covariance of the estimators $\hat{S}_1(t_1,t_2 \mid \underline{z})$ and $\hat{S}_2(t_1,t_2 \mid \underline{z})$. Accordingly, the estimate of $a(t_1,t_2)$ is given by minimizing the average variance of $\hat{S}(t_1,t_2 \mid \underline{z})$ over the data points. Simulation study reported in Section 6.5 shows that biases are negligible, To estimate the variance, we can use the extension of Efron's (1981) bootstrap procedure for one dimensional censored data. Given the data $(t_{1i},t_{2i},\Delta_{1i},\Delta_{2i},C_{1i},C_{2i},\underline{Z}_i)$, $i=1,2,...,n$, where Δ_{ji}, $j=1,2$ is the censoring indicator, generate the bootstrap data $(t_{1k}^*,t_{2k}^*,\Delta_{1k}^*,\Delta_{2k}^*,\underline{Z}_k^*)$, $k=1,2,...,n$ from the empirical distribution function

$$\frac{1}{n}\sum_{j=1}^{n} I\left(T_{1j} \le t_1, T_{2j} \le t_2, \Delta_{1j}=\delta_{1j}, \Delta_{2j}=\delta_{2j} C_{1i}=p, C_{2j}=q, \underline{Z}_j=\underline{z}_j\right).$$

For $i=1,2,...,n$ let $Var(S_{1i}(t_1,t_2)^*)$, $Var(S_{2i}(t_1,t_2)^*)$ and $Cov(S_{1i}(t_1,t_2)^*, S_{2i}(t_1,t_2)^*)$ be variance and covariance of the $\hat{S}_1(t_1,t_2)$ and $\hat{S}_2(t_1,t_2)$ in the expression of $\hat{S}(t_1,t_2)$ given in (6.31), obtained from a large number of resamples. Since the biases are negligible as shown in Section 2.7, we then find the weight $a^*(t_1,t_2)$ as

$$a^*(t_1,t_2)=\frac{\mathrm{var}\left(S_2(t_1,t_2)^*\right)-\mathrm{cov}\left(S_1(t_1,t_2)^*, S_2(t_1,t_2)^*\right)}{\mathrm{var}\left(S_1(t_1,t_2)^*\right)+\mathrm{var}\left(S_2(t_1,t_2)^*\right)-2\mathrm{cov}\left(S_1(t_1,t_2)^*, S_2(t_1,t_2)^*\right)}. \qquad (6.33)$$

Substituting (6.28) and (6.31) in (6.17), the estimate of $F_{pq}(t_1,t_2)$ will be obtained as

$$\hat{F}_{pq}(t_1,t_2 \mid \underline{z}) = \int_0^{t_1}\int_0^{t_2} \hat{S}(u,v \mid \underline{z})\, \hat{\Lambda}_{3pq}(du,dv \mid \underline{z}). \qquad (6.34)$$

6.4. Properties of the Estimators

Define for any $t_1,t_2 \ge 0$

$$S_{1p}^{(0)}(t_1,t_2,\underline{\beta}_{1p}) = \sum_{l=1}^{n} Y_l(t_1,t_2) e^{\underline{\beta}_{1p}\cdot \underline{z}_l}. \qquad (6.35)$$

Then for fixed t_1 and for fixed p, we can easily prove that $\sqrt{n}\left[\hat{\wedge}_{1p0}(t_1,t_2)-\wedge_{1p0}(t_1,t_2)\right]$ has a limiting normal distribution with mean zero vector and variance $n\sigma_1^2$ where σ_1^2 is

$$\sigma_1^2=\sum_{j:t_{1j}\le t_1}\frac{\delta^*_{1pj}}{\left(S_{1p}^{(0)}\left(t_{1j},t_{2j},\underline{\beta}_{1p}\right)\right)^2}+W_{1p}(t_1,t_2)'I_{1p}\left(\underline{\beta}_{1p}\right)^{-1}W_{1p}(t_1,t_2) \tag{6.36}$$

where

$$W_{1p}(t_1,t_2)=\sum_{j:t_{1j}\le t_1,t_{2j}\le t_2}\frac{\delta^*_{1pj}\overline{z}_{1p}(t_{1j},t_{2j},\underline{\beta}_{1p})}{S_{1p}^{(0)}(t_{1j},t_{2j},\underline{\beta}_{1p})}.$$

Since $\hat{\wedge}_{10}(t_1,t_2)=\sum_{p=1}^{k_1}\hat{\wedge}_{1p0}(t_1,t_2)$, the asymptotic normality of $\sqrt{n}[\hat{\wedge}_{10}(t_1,t_2)-\wedge_{10}(t_1,t_2)]$ can easily be obtained.

Similarly for fixed t_1, t_2 and fixed q, $\sqrt{n}\left[\hat{\wedge}_{2q0}(t_1,t_2)-\wedge_{2q0}(t_1,t_2)\right]$ has a limiting normal distribution with mean zero vector and variance $n\sigma_2^2$, where σ_2^2 is given by

$$\sigma_2^2=\sum_{j:t_{1j}\le t_1,t_{2j}\le t_2}\frac{\delta^*_{2qj}}{\left(S_{2q}^{(0)}\left(t_{1j},t_{2j},\underline{\beta}_{2q}\right)\right)^2}+W_{2q}(t_1,t_2)'I_{2q}\left(\underline{\beta}_{2q}\right)^{-1}W_{2q}(t_1,t_2) \tag{6.37}$$

where

$$S_{2q}^{(0)}(t_1,t_2,\underline{\beta}_{2q})=\sum_{l=1}^{n}Y_{2l}(t_1,t_2)e^{\underline{\beta}_{2q}'z_l}\text{, for any } t_2\ge 0$$

and

$$W_{2q}(t_1,t_2)=\sum_{j:t_{1j}\le t_1,t_{2j}\le t_2}\frac{\delta^*_{2qj}\overline{z}_{2q}(t_{1j},t_{2j},\underline{\beta}_{2q})}{S_{2q}^{(0)}(t_{1j},t_{2j},\underline{\beta}_{2q})}.$$

This provides the asymptotic normality of $\sqrt{n}[\hat{\wedge}_{20}(t_1,t_2)-\wedge_{20}(t_1,t_2)]$, since $\hat{\wedge}_{20}(t_1,t_2)=\sum_{q=1}^{k_2}\hat{\wedge}_{2q0}(t_1,t_2)$.

We can also prove that $\sqrt{n}\left[\hat{\wedge}_{3pq0}(t_1,t_2)-\wedge_{3pq0}(t_1,t_2)\right]$ has a limiting normal distribution with mean zero vector and variance $n\sigma_3^2$, where σ_3^2 is given by

$$\sigma_3^2 = \sum_{j:t_{1j}\le t_1, t_{2j}\le t_2} \frac{\delta^*_{3pqj}}{\left(S_{3pq}^{(0)}\left(t_{1j}, t_{2j}, \underline{\beta}_{1p}, \underline{\beta}_{2q}\right)\right)^2} + W_{3pq}(t_1,t_2)' I_{3pq}\left(\underline{\beta}_{1p}, \underline{\beta}_{2q}\right)^{-1} W_{3pq}(t_1,t_2) \tag{6.38}$$

where

$$W_{3pq}(t_1,t_2) = \sum_{j:t_{1j}\le t_1, t_{2j}\le t_2} \frac{\delta^*_{3pqj}\bar{z}_{3pq}(t_{1j}, t_{2j}, \underline{\beta}_{1p}, \underline{\beta}_{2q})}{S_{3pq}^{(0)}(t_{1j}, t_{2j}, \underline{\beta}_{1p}, \underline{\beta}_{2q})}$$

with

$$\bar{z}_{3pq}(t_1,t_2,\underline{\beta}_{1p},\underline{\beta}_{2q}) = \frac{\sum_{l=1}^{n} Y_l(t_1,t_2) e^{\underline{\beta}_{3pq}'\underline{z}_l}\underline{z}_l}{\sum_{l=1}^{n} Y_l(t_1,t_2) e^{\underline{\beta}_{3pq}'\underline{z}_l}}$$

and

$$S_{3pq}^{(0)}(t_1,t_2,\underline{\beta}_{1p},\underline{\beta}_{2q}) = \sum_{l=1}^{n} Y_l(t_1,t_2) e^{\underline{\beta}_{1p}'\underline{z}_l + \underline{\beta}_{2q}'\underline{z}_l},\ p=1,2,...,k_1,\ q=1,2,...,k_2.$$

Thus the asymptotic normality of $\sqrt{n}\left[\hat{\wedge}_{30}(t_1,t_2) - \wedge_{30}(t_1,t_2)\right]$ can easily be established.

We then study the asymptotic properties of $\underline{\hat{\beta}}_{1p}$ and $\underline{\hat{\beta}}_{2q}$. Using (6.25), the score function of $\underline{\beta}_{1p}$ can be written as,

$$U_{1p}\left(\underline{\beta}_{1p}\right) = \sum_{j=1}^{n} \delta^*_{1pj}\left[\underline{z}_j - \frac{S_{1p}^{(1)}\left(t_{1j}, t_{2j}, \underline{\beta}_{1p}\right)}{S_{1p}^{(0)}(t_{1j}, t_{2j}, \underline{\beta}_{1p})}\right] \tag{6.39}$$

where

$$S_{1p}^{(1)}\left(t_1,t_2,\underline{\beta}_{1p}\right) = \sum_{l=1}^{n} Y_l(t_1,t_2) e^{\underline{\beta}_{1p}'\underline{z}_l}\underline{z}_l.$$

Then the maximum likelihood estimator $\underline{\hat{\beta}}_{1p}$ is the solution of the score function $U_{1p}\left(\underline{\hat{\beta}}_{1p}\right) = 0$ and hence $\underline{\hat{\beta}}_{1p}$ is a consistent estimator for $\underline{\beta}_{1p}$. For large n, the score statistic $U_{1p}\left(\underline{\beta}_{1p}\right)$ is asymptotically r-variate normal with mean zero vector and covariance matrix $A_{1p}\left(\underline{\beta}_{1p}\right)$ where

$$A_{1p}\left(\underline{\hat{\beta}}_{1p}\right)=\sum_{j=1}^{n}\delta^{*}{}_{1pj}\left\{\frac{S_{1p}{}^{(2)}\left(t_{1j},t_{2j},\underline{\hat{\beta}}_{1p}\right)}{S_{1p}{}^{(0)}\left(t_{1j},t_{2j},\underline{\hat{\beta}}_{1p}\right)}-\frac{S_{1p}{}^{(1)}\left(t_{1j},t_{2j},\underline{\hat{\beta}}_{1p}\right)S_{1p}{}^{(1)}\left(t_{1j},t_{2j},\underline{\hat{\beta}}_{1p}\right)'}{\left(S_{1p}{}^{(0)}\left(t_{1j},t_{2j},\underline{\hat{\beta}}_{1p}\right)\right)^{2}}\right\} \tag{6.40}$$

with

$$S_{1p}{}^{(2)}\left(t_1,t_2,\underline{\beta}_{1p}\right)=\sum_{l=1}^{n}Y_l(t_1,t_2)e^{\underline{\beta}_{1p}'\underline{z}_l}\underline{z}_l\underline{z}_l'.$$

Thus $\underline{\hat{\beta}}_{1p}$ is asymptotically r-variate normal with mean vector $\underline{\beta}_{1p}$ and covariance matrix $A_{1p}{}^{-1}\left(\underline{\beta}_{1p}\right)$.

Similarly we have,

$$U_{2q}\left(\underline{\beta}_{2q}\right)=\sum_{j=1}^{n}\delta_{2qj}\left[\underline{z}_j-\frac{S_{2q}{}^{(1)}\left(t_1,t_2,\underline{\beta}_{2q}\right)}{S_{2q}{}^{(0)}\left(t_1,t_2,\underline{\beta}_{2q}\right)}\right] \tag{6.41}$$

where

$$S_{2q}{}^{(1)}\left(t_1,t_2,\underline{\beta}_{2q}\right)=\sum_{l=1}^{n}Y_l(t_1,t_2)e^{\underline{\beta}_{2q}'\underline{z}_l}\underline{z}_l.$$

Then $\underline{\hat{\beta}}_{2q}$ is the solution of $U_{2q}\left(\underline{\hat{\beta}}_{2q}\right)=0$. Since $\underline{\hat{\beta}}_{2q}$ is the maximum likelihood estimator, $\underline{\hat{\beta}}_{2q}$ is a consistent estimator for $\underline{\beta}_{2q}$. For large n, $U_{2q}\left(\underline{\hat{\beta}}_{2q}\right)$ is asymptotically r-variate normal with mean zero vector and covariance matrix $A_{2q}\left(\underline{\beta}_{2q}\right)$, where

$$A_{2q}\left(\underline{\hat{\beta}}_{2q}\right)=\sum_{j=1}^{n}\delta^{*}{}_{2qj}\left\{\frac{S_{2q}{}^{(2)}\left(t_{1j},t_{2j},\underline{\hat{\beta}}_{2q}\right)}{S_{2q}{}^{(0)}\left(t_{1j},t_{2j},\underline{\hat{\beta}}_{2q}\right)}-\frac{S_{2q}{}^{(1)}\left(t_{1j},t_{2j},\underline{\hat{\beta}}_{2q}\right)S_{2q}{}^{(1)}\left(t_{1j},t_{2j},\underline{\hat{\beta}}_{2q}\right)'}{\left(S_{2q}{}^{(0)}\left(t_{1j},t_{2j},\underline{\hat{\beta}}_{2q}\right)\right)^{2}}\right\} \tag{6.42}$$

with

$$S_{2q}{}^{(2)}\left(t_1,t_2,\underline{\beta}_{2q}\right)=\sum_{l=1}^{n}Y_l(t_1,t_2)e^{\underline{\beta}_{2q}'\underline{z}_l}\underline{z}_l\underline{z}_l'.$$

Thus $\underline{\hat{\beta}}_{2q}$ is asymptotically r-variate normal with mean vector $\underline{\beta}_{2q}$ and covariance matrix $A_{2q}{}^{-1}\left(\underline{\beta}_{2q}\right)$.

The well known functional delta method can be used to prove the asymptotic properties of $\hat{S}(t_1,t_2 \mid \underline{z})$. Obviously, the expression of the variance of the estimate is complex. In practical situations, an attractive approach for variance or confidence interval estimation is through resampling methods. The naïve bootstrap procedure of resampling the observed data units $(t_{1i},t_{2i},\delta_{1i},\delta_{2i},C_{1i},C_{2i},\underline{z}_i)$, $i=1,2,...,n$ with replacement will be satisfactory under fairly mild condition (see Efron and Tibshirani, 1993).

6.5. Simulation Study

In this section, we carried out a simulation study to evaluate the performance of the aforementioned inference procedures. We consider a Gumbel's (1960) bivariate exponential distribution with survival function

$$S(t_1,t_2)=\exp(-t_1-t_2-\gamma t_1 t_2),\ t_1,t_2>0,\ 0\le\gamma\le 1. \tag{6.43}$$

The model has hazard functions

$$\lambda_i(t_1,t_2)=(1+\gamma t_j)t_i,\ i,j=1,2;i\neq j. \tag{6.44}$$

For the simplicity of the analysis, we take only one covariate z. Covariate z is generated from uniform (0, 1) distribution. Corresponding to each T_1 and T_2, we consider two causes say 1 and 2. The observations are generated from the model (6.43) with cause-specific hazard functions

$$\wedge_{1p}(t_1,t_2 \mid \underline{z})=(1+\gamma t_2)t_1 e^{\beta_{1p}\underline{z}} \tag{6.45}$$

and

$$\wedge_{2q}(t_1,t_2 \mid \underline{z})=(1+\gamma t_1)t_2 e^{\beta_{2q}\underline{z}} \tag{6.46}$$

for various values of γ, β_{1p} and β_{2q}, $p,q=1,2$.

We used the algorithm given in Devroye (1986) for generating the observations. The paired lifetimes are censored by a Gumbel's bivariate exponential distribution (6.43) with λ=0.6. We compute estimates for 1000 simulations for sample size $n=50$ and $n=100$. Average bias and variance of the estimates of $\underline{\beta}_1,\underline{\beta}_2$ and baseline cumulative cause-specific hazard functions are calculated and are given in Table 6.1 to Table 6.10. Bias and variance of the estimates of bivariate survival function $S(t_1,t_2 \mid \underline{z})$ and cause-specific sub-distribution functions $F_{pq}(t_1,t_2 \mid \underline{z})$, $p,q=1,2$ are

given in Table 6.11 to Table 6.15. From the tables, it follows that as n increases, both bias and variance of the estimates decreases.

Table 6.1. Bias and variance of $\hat{\beta}_{11}$ and $\hat{\beta}_{12}$

β_{11}	β_{12}	γ	n	Bias $\hat{\beta}_{11}$	Var $\hat{\beta}_{11}$	Bias $\hat{\beta}_{12}$	Var $\hat{\beta}_{12}$
1	-0.5	0.8	50	-0.0541	0.0600	0.0549	0.0442
			250	0.0429	0.0553	0.0317	0.0416
		0.6	50	0.0483	0.0389	0.0473	0.0591
			250	-0.0396	0.0301	-0.0400	0.0511
0.9	0.6	0.8	50	-0.0461	0.0403	-0.0391	0.0369
			250	0.0412	0.0392	0.0300	0.0297
		0.6	50	0.0359	0.0421	-0.0522	0.0463
			250	0.0321	0.0392	0.0425	0.0432

Table 6.2. Bias and variance of $\hat{\beta}_{21}$ and $\hat{\beta}_{22}$

β_{21}	β_{22}	γ	n	Bias $\hat{\beta}_{21}$	Var $\hat{\beta}_{21}$	Bias $\hat{\beta}_{22}$	Var $\hat{\beta}_{22}$
0.8	1.2	0.8	50	0.0491	0.0394	-0.0446	0.0511
			250	0.0372	0.0301	0.0411	0.0498
		0.6	50	-0.0473	0.0395	-0.0368	0.0211
			250	-0.0302	0.0286	0.0300	0.0203
-0.6	1	0.8	50	0.0456	0.0400	0.5112	0.0483
			250	0.0325	0.0369	0.0458	0.0415
		0.6	50	0.0385	0.0421	-0.0369	0.0349
			250	0.0289	0.0112	0.0257	0.0300

Table 6.3. Bias and variance of estimates of the baseline cumulative cause-specific hazard function $\wedge_{110}(t_1,t_2)$

β_{11}	β_{12}	β_{21}	β_{22}	(t_1,t_2)	γ	n	Bias $\hat{\wedge}_{110}(t_1,t_2)$	Var $\hat{\wedge}_{110}(t_1,t_2)$
1	-0.5	0.8	1.2	(1,1.2)	0.8	50	0.0423	0.0221
						250	0.0325	0.0200
					0.6	50	-0.0470	0.0329
						250	0.0369	0.0228
				(1.5,0.8)	0.8	50	0.0226	0.0190
						250	0.0189	0.0153
					0.6	50	0.0358	0.0371
						250	-0.0300	0.0254
0.9	0.6	-0.6	1	(1,1.2)	0.8	50	0.0425	0.0221
						250	0.0301	0.0156
					0.6	50	0.0328	0.0221
						250	0.0283	0.0200
				(1.5,0.8)	0.8	50	0.0344	0.0298
						250	0.0226	0.0255
					0.6	50	0.0364	0.0211
						250	0.0301	0.0177

Table 6.4. Bias and variance of estimates of the baseline cumulative cause-specific hazard function $\wedge_{120}(t_1,t_2)$

β_{11}	β_{12}	β_{21}	β_{22}	(t_1,t_2)	γ	n	Bias $\hat{\wedge}_{120}(t_1,t_2)$	Var $\hat{\wedge}_{120}(t_1,t_2)$
1	-0.5	0.8	1.2	(1,1.2)	0.8	50	-0.0559	0.0591
						250	0.0405	0.0301
					0.6	50	0.0324	0.0389
						250	-0.0202	0.0144
				(1.5,0.8)	0.8	50	0.0322	0.0454
						250	0.0271	0.0361
					0.6	50	0.0355	0.0297
						250	0.0205	0.0147
0.9	0.6	-0.6	1	(1,1.2)	0.8	50	-0.0523	0.0569
						250	0.0487	0.0356
					0.6	50	0.0694	0.0436
						250	0.0401	0.0238
				(1.5,0.8)	0.8	50	-0.0516	0.0472
						250	-0.0324	0.0174
					0.6	50	0.0408	0.0419
						250	0.0327	0.0308

Table 6.5. Bias and variance of estimates of the baseline cumulative cause-specific hazard function $\wedge_{210}(t_1,t_2)$

β_{11}	β_{12}	β_{21}	β_{22}	(t_1,t_2)	γ	n	Bias $\hat{\wedge}_{210}(t_1,t_2)$	Var $\hat{\wedge}_{210}(t_1,t_2)$
1	-0.5	0.8	1.2	(1,1.2)	0.8	50	0.0382	0.0339
						250	0.0167	0.0205
					0.6	50	-0.0350	0.0398
						250	0.0230	0.0257
				(1.5,0.8)	0.8	50	0.0441	0.0316
						250	-0.0240	0.0238
					0.6	50	0.0536	0.0302
						250	0.0310	0.0285
0.9	0.6	-0.6	1	(1,1.2)	0.8	50	0.0384	0.0428
						250	0.0280	0.0317
					0.6	50	-0.0466	0.0305
						250	0.0330	0.0266
				(1.5,0.8)	0.8	50	-0.0204	0.0568
						250	0.0110	0.0357
					0.6	50	0.0493	0.0409
						250	0.0355	0.0305

Table 6.6. Bias and variance of estimates of the baseline cumulative cause-specific hazard functions $\wedge_{220}(t_1,t_2)$

β_{11}	β_{12}	β_{21}	β_{22}	(t_1,t_2)	γ	n	Bias $\hat{\wedge}_{220}(t_1,t_2)$	Var $\hat{\wedge}_{220}(t_1,t_2)$
1	-0.5	0.8	1.2	(1,1.2)	0.8	50	0.0516	0.0471
						250	0.0311	0.0428
					0.6	50	0.0411	0.0518
						250	0.0302	0.0439
				(1.5,0.8)	0.8	50	-0.0505	0.0524
						250	-0.0413	0.0441
					0.6	50	-0.0489	0.0494
						250	-0.0404	0.0346
0.9	0.6	-0.6	1	(1,1.2)	0.8	50	0.0502	0.0594
						250	0.0396	0.0453
					0.6	50	0.0554	0.0521
						250	0.0416	0.0505
				(1.5,0.8)	0.8	50	-0.0593	0.0478
						250	0.0515	0.0329
					0.6	50	0.0462	0.0511
						250	0.0374	0.0408

Table 6.7. Bias and variance of estimates of the baseline cumulative cause-specific hazard function $\wedge_{3110}(t_1,t_2)$

β_{11}	β_{12}	β_{21}	β_{22}	(t_1,t_2)	γ	n	Bias $\hat{\wedge}_{3110}(t_1,t_2)$	Var $\hat{\wedge}_{3110}(t_1,t_2)$
1	-0.5	0.8	1.2	(1,1.2)	0.8	50	0.0554	0.0281
						250	0.0429	0.0189
					0.6	50	-0.0519	0.0314
						250	0.0426	0.0242
				(1.5,0.8)	0.8	50	0.0336	0.0394
						250	0.0210	0.0202
					0.6	50	0.0473	0.0416
						250	0.0334	0.0380
0.9	0.6	-0.6	1	(1,1.2)	0.8	50	-0.0410	0.0328
						250	0.0329	0.0304
					0.6	50	0.0470	0.0239
						250	0.0408	0.0196
				(1.5,0.8)	0.8	50	0.0339	0.0404
						250	0.0310	0.0359
					0.6	50	-0.0300	0.0209
						250	0.0278	0.0196

Table 6.8. Bias and variance of estimates of the baseline cumulative cause-specific hazard function $\wedge_{3120}(t_1,t_2)$

β_{11}	β_{12}	β_{21}	β_{22}	(t_1,t_2)	γ	n	Bias $\hat{\wedge}_{3120}(t_1,t_2)$	Var $\hat{\wedge}_{3120}(t_1,t_2)$
1	-0.5	0.8	1.2	(1,1.2)	0.8	50	0.0549	0.0479
						250	-0.0397	0.0321
					0.6	50	0.0486	0.0440
						250	0.0398	0.0325
				(1.5,0.8)	0.8	50	0.0569	0.0299
						250	-0.0489	0.0117
					0.6	50	0.0679	0.0374
						250	0.0486	0.0234
0.9	0.6	-0.6	1	(1,1.2)	0.8	50	-0.0557	0.0507
						250	0.0433	0.0481
					0.6	50	0.0362	0.0317
						250	0.0267	0.0263
				(1.5,0.8)	0.8	50	0.0408	0.0305
						250	0.0375	0.0262
					0.6	50	0.0479	0.0205
						250	0.0321	0.0170

Table 6.9. Bias and variance of estimates of the baseline cumulative cause-specific hazard function $\wedge_{3210}(t_1,t_2)$

β_{11}	β_{12}	β_{21}	β_{22}	(t_1,t_2)	γ	n	Bias $\hat{\wedge}_{3210}(t_1,t_2)$	Var $\hat{\wedge}_{3210}(t_1,t_2)$
1	-0.5	0.8	1.2	(1,1.2)	0.8	50	0.0395	0.0333
						250	0.0201	0.0206
					0.6	50	0.0408	0.0521
						250	-0.0306	0.0415
				(1.5,0.8)	0.8	50	0.0405	0.0312
						250	0.0306	0.0217
					0.6	50	0.0304	0.0221
						250	0.0299	0.0192
0.9	0.6	-0.6	1	(1,1.2)	0.8	50	0.0357	0.0304
						250	-0.0218	0.0213
					0.6	50	0.0457	0.0314
						250	0.0218	0.0310
				(1.5,0.8)	0.8	50	0.0357	0.0264
						250	-0.0243	0.0134
					0.6	50	0.0306	0.0344
						250	0.0191	0.0285

Table 6.10. Bias and variance of estimates of the baseline cumulative cause-specific hazard function $\wedge_{3220}(t_1,t_2)$

β_{11}	β_{12}	β_{21}	β_{22}	(t_1,t_2)	γ	n	Bias $\hat{\wedge}_{3220}(t_1,t_2)$	Var $\hat{\wedge}_{3220}(t_1,t_2)$
1	-0.5	0.8	1.2	(1,1.2)	0.8	50	-0.0347	0.0291
						250	0.0200	0.0208
					0.6	50	-0.0463	0.0459
						250	0.0337	0.0316
				(1.5,0.8)	0.8	50	-0.0307	0.0300
						250	0.0235	0.0219
					0.6	50	0.0462	0.0387
						250	0.0396	0.0230
0.9	0.6	-0.6	1	(1,1.2)	0.8	50	0.0250	0.0352
						250	0.0108	0.0243
					0.6	50	-0.0333	0.0200
						250	0.0202	0.0185
				(1.5,0.8)	0.8	50	-0.0433	0.0400
						250	0.0382	0.0355
					0.6	50	0.0363	0.0299
						250	0.0207	0.0116

Table 6.13. Estimates of the cause-specific sub-distribution function $F_{12}(t_1,t_2 \mid \underline{z})$

β_{11}	β_{12}	β_{21}	β_{22}	(t_1,t_2)	γ	n	Bias $\hat{F}_{12}(t_1,t_2) \mid \underline{z}$	Var $\hat{F}_{12}(t_1,t_2 \mid \underline{z})$
1	-0.5	0.8	1.2	(1,1.2)	0.8	50	-0.0577	0.0529
						250	-0.0546	0.0412
					0.6	50	0.0359	0.0455
						250	0.0243	0.0408
				(1.5,0.8)	0.8	50	-0.0601	0.0304
						250	0.0545	0.0210
					0.6	50	0.0496	0.0422
						250	0.0394	0.0357
0.9	0.6	-0.6	1	(1,1.2)	0.8	50	0.0457	0.0582
						250	-0.0365	0.0482
					0.6	50	0.0577	0.0329
						250	0.0506	0.0312
				(1.5,0.8)	0.8	50	-0.0360	0.0227
						250	0.0353	0.0127
					0.6	50	0.0260	0.0327
						250	0.0153	0.0293

Table 6.14. Estimates of the cause-specific sub-distribution function $F_{21}(t_1,t_2 \mid \underline{z})$

β_{11}	β_{12}	β_{21}	β_{22}	(t_1,t_2)	γ	n	Bias $\hat{F}_{21}(t_1,t_2 \mid \underline{z})$	Var $\hat{F}_{21}(t_1,t_2 \mid \underline{z})$
1	-0.5	0.8	1.2	(1,1.2)	0.8	50	0.0578	0.0406
						250	-0.0402	0.0312
					0.6	50	0.0592	0.0315
						250	0.0558	0.0299
				(1.5,0.8)	0.8	50	0.0439	0.0350
						250	0.0326	0.0248
					0.6	50	-0.0655	0.0503
						250	0.0599	0.0418
0.9	0.6	-0.6	1	(1,1.2)	0.8	50	0.0492	0.0415
						250	0.0358	0.0399
					0.6	50	0.0505	0.0229
						250	-0.0489	0.0190
				(1.5,0.8)	0.8	50	0.0405	0.0529
						250	0.0389	0.0490
					0.6	50	-0.0232	0.0324
						250	0.0157	0.0209

Table 6.15. Estimates of the cause-specific sub-distribution function $F_{22}(t_1,t_2 \mid \underline{z})$

β_{11}	β_{12}	β_{21}	β_{22}	(t_1,t_2)	γ	n	Bias $\hat{F}_{22}(t_1,t_2 \mid \underline{z})$	Var $\hat{F}_{22}(t_1,t_2 \mid \underline{z})$
1	-0.5	0.8	1.2	(1,1.2)	0.8	50	0.0509	0.0495
						250	0.0473	0.0418
					0.6	50	0.0395	0.0429
						250	-0.0275	0.0373
				(1.5,0.8)	0.8	50	0.0605	0.0379
						250	-0.0509	0.0220
					0.6	50	0.0497	0.0412
						250	0.0349	0.0336
0.9	0.6	-0.6	1	(1,1.2)	0.8	50	0.0437	0.0502
						250	0.0255	0.0462
					0.6	50	0.0509	0.0406
						250	0.0415	0.0319
				(1.5,0.8)	0.8	50	0.0374	0.0301
						250	0.0242	0.0215
					0.6	50	-0.0384	0.0208
						250	-0.0342	0.0195

6.6. Data Analysis

We now present a potential application of the proposed method to a litter-matched time-to-response data which is given in Mantel et al. (1977). Different versions of this data have been studied in Mantel and Ciminera (1979), Ying and Wei (1994), Kalbfliesh and Prentice (2002), Lawless (2003) and Sankaran et al. (2006). The purpose here is to illustrate a possible application of the proposed techniques rather than provide a definite analysis of the data. The study consists of 300 rats divided into 50 male litters and 50 female litters, all litters of which were of size 3. In the data, T_1 and T_2 represent lifetimes (in weeks) for a pair of mice, and C_j $(j=1,2)$ indicates whether the failure was the appearance of a tumor $(C_j=1)$ or the occurrence of death prior to tumor appearance $(C_j=2)$. The censored observations are denoted by $C_j=0$. Since the study terminates after 104 weeks, 104 is the common censoring time. Gender is considered as covariate $\underline{z}$. The data is given in Table 6.16.

We find the estimates using the procedure given in Section 6.3. The estimates of β_{1p} and β_{2q}, $p,q=1,2$ are given in Table 6.17. The estimates of baseline cumulative cause-specific hazard functions are given in Table 6.18 to Table 6.20. Estimates of the bivariate survival function $S(t_1,t_2 \mid \underline{z})$ are given in Table 6.21 and the estimates of cause-specific sub-distribution functions are given in Table 6.22. From Table 6.17, we can see that β_{1p} has positive effect on $\wedge_{1p0}(t_1,t_2)$, $p=1,2$. However, β_{2q} has negative effect on $\wedge_{2q0}(t_1,t_2)$, $q=1,2$. From Table 6.18 and Table 6.19, it follows that values of baseline cumulative cause-specific hazard functions corresponding to the causes (1,1) and (2,1) are negligible, but those corresponding to the causes (2,1) and (2,2) have high values for the most of the time pair. The estimator of the baseline cumulative cause-specific hazard function $\wedge_{3pq0}(t_1,t_2)$ has negligible values for the set of causes (1,2) and (2,1). From Table 6.22, we can see that $F_{12}(t_1,t_2 \mid \underline{z})$ and $F_{22}(t_1,t_2 \mid \underline{z})$ have much larger values compared to $F_{11}(t_1,t_2 \mid \underline{z})$ and $F_{21}(t_1,t_2 \mid \underline{z})$; as expected from the data. Figure 6.1 shows the estimates of the survival function. Figures 6.2 to 6.5 shows the estimates of cause-specific sub-distribution function for different causes.

We then estimate $\pi_{pq}=P(C_1=p,C_2=q)$ as $\hat{\pi}_{pq}=\hat{F}_{pq}(\infty,\infty)$. Since the several pairs are censored, we normalize the estimate as $\hat{\pi}_{pq}^{*}=\dfrac{\hat{\pi}_{pq}}{\sum_k\sum_l \hat{\pi}_{kl}}$. We also estimate the marginal probabilities $\pi_p^{(q)}=P(C_q=p)$, $p,q=1,2$ using the estimate of the marginal sub-distribution function. The estimates $\hat{\pi}_p^{(q)*}$ and $\hat{\pi}_{pq}^{*}$ are given in Table 6.23, which shows that the two causes are not independent.

Table 6.16. Data concerning the times to tumor appearance or death for 100 pairs of mice

Litter no.	z	T_1	C_1	T_2	C_2	Litter no.	z	T_1	C_1	T_2	C_2
01	2	49	1	104	0	02	1	104	0	102	2
03	2	102	2	104	0	04	1	104	0	102	2
05	2	104	0	104	0	06	1	62	2	77	2
07	2	97	2	79	2	08	1	98	2	76	2
09	2	104	0	104	0	10	1	104	0	98	2
11	2	96	1	104	0	12	1	71	2	91	2
13	2	94	2	77	1	14	1	104	0	99	2
15	2	104	0	104	0	16	1	88	2	85	2
17	2	77	2	104	0	18	1	104	0	102	2
19	2	104	0	77	2	20	1	104	0	102	2
21	2	91	2	90	2	22	1	80	2	92	2
23	2	70	2	92	2	24	1	104	0	101	2
25	2	45	2	50	1	26	1	53	2	102	2
27	2	69	2	91	2	28	1	104	0	91	2
29	2	104	0	103	2	30	1	104	0	75	2
31	2	72	2	104	0	32	1	100	2	102	2
33	2	63	2	104	0	34	1	104	0	95	2
35	2	104	0	74	2	36	1	104	0	102	2
37	2	104	0	69	2	38	1	93	2	80	2
39	2	104	0	68	1	40	1	98	2	83	2
41	2	104	0	104	0	42	1	89	2	89	2
43	2	104	0	104	0	44	1	32	2	51	2
45	2	83	2	40	1	46	1	98	2	78	2
47	2	104	0	104	0	48	1	104	0	102	2
49	2	104	0	104	0	50	1	104	0	94	2
51	2	104	0	104	0	52	1	104	0	102	2
53	2	104	0	104	0	54	1	91	2	102	2
55	2	81	1	64	1	56	1	104	0	55	2
57	2	55	1	94	2	58	1	104	0	102	2
59	2	104	0	54	1	60	1	104	0	102	2
61	2	87	2	74	2	62	1	104	0	102	2
63	2	73	1	84	1	64	1	71	1	90	2
65	2	104	0	80	2	66	1	51	2	102	2
67	2	104	0	73	2	68	1	83	2	102	2
69	2	79	2	104	0	70	1	104	0	96	2
71	2	104	0	104	0	72	1	84	2	94	2
73	2	104	0	104	0	74	1	104	0	99	2
75	2	101	1	94	2	76	1	94	2	102	2
77	2	84	1	78	1	78	1	103	2	102	2
79	2	81	1	76	2	80	1	104	0	91	2
81	2	95	1	104	0	82	1	98	2	102	2
83	2	104	0	66	1	84	1	54	2	39	2
85	2	104	0	102	1	86	1	84	2	54	2
87	2	98	2	73	2	88	1	104	0	87	2
89	2	104	0	104	0	90	1	82	2	102	2
91	2	83	2	77	2	92	1	104	0	102	2
93	2	104	0	104	0	94	1	89	2	77	2
95	2	79	2	99	2	96	1	69	2	102	2
97	2	91	2	104	0	98	1	75	1	64	2
99	2	104	0	79	1	100	1	104	0	102	2

Table 6.17. Estimates of β_1 and β_2 for different causes

(t_1,t_2)	$\underline{z}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
(98,83)	1	1.393	40.02	-0.260	-1.102
(84,94)	1	1.393	40.02	-0.260	-1.102
(98,78)	1	1.393	40.02	-0.260	-1.102
(89,89)	1	1.393	40.02	-0.260	-1.102
(97,79)	2	1.393	40.02	-0.260	-1.102
(93,80)	1	1.393	40.02	-0.260	-1.102
(80,92)	1	1.393	40.02	-0.260	-1.102
(98,76)	1	1.393	40.02	-0.260	-1.102
(91,90)	2	1.393	40.02	-0.260	-1.102
(88,85)	1	1.393	40.02	-0.260	-1.102
(98,73)	2	1.393	40.02	-0.260	-1.102
(94,77)	2	1.393	40.02	-0.260	-1.102
(79,99)	2	1.393	40.02	-0.260	-1.102
(89,77)	1	1.393	40.02	-0.260	-1.102
(71,91)	1	1.393	40.02	-0.260	-1.102
(84,78)	2	1.393	40.02	-0.260	-1.102
(55,94)	2	1.393	40.02	-0.260	-1.102
(70,92)	2	1.393	40.02	-0.260	-1.102
(87,74)	2	1.393	40.02	-0.260	-1.102
(71,90)	1	1.393	40.02	-0.260	-1.102
(69,91)	2	1.393	40.02	-0.260	-1.102
(73,84)	2	1.393	40.02	-0.260	-1.102
(83,77)	2	1.393	40.02	-0.260	-1.102
(81,76)	2	1.393	40.02	-0.260	-1.102
(81,64)	2	1.393	40.02	-0.260	-1.102
(84,54)	1	1.393	40.02	-0.260	-1.102
(83,40)	2	1.393	40.02	-0.260	-1.102
(62,77)	1	1.393	40.02	-0.260	-1.102
(75,64)	1	1.393	40.02	-0.260	-1.102
(54,39)	1	1.393	40.02	-0.260	-1.102
(45,50)	2	1.393	40.02	-0.260	-1.102
(32,51)	1	1.393	40.02	-0.260	-1.102

Table 6.18. Estimates of the baseline cumulative cause-specific hazard functions

(t_1,t_2)	$\underline{z}$	$\hat{\Lambda}_{110}(t_1,t_2)$	$\hat{\Lambda}_{110}(t_1,0)$	$\hat{\Lambda}_{120}(t_1,t_2)$	$\hat{\Lambda}_{120}(t_1,0)$
(98,83)	1	0.0031	0.0111	0.3355	0.7768
(84,94)	1	0	0.0094	0.2066	0.3776
(98,78)	1	0.0056	0.0111	0.2616	0.7768
(89,89)	1	0	0.0094	0.2040	0.4625
(97,79)	2	0.0056	0.0111	0.2102	0.6492
(93,80)	1	0.0039	0.0094	0.1846	0.5524
(80,92)	1	0	0.0055	0.1282	0.2611
(98,76)	1	0.0056	0.0111	0.1560	0.7768
(91,90)	2	0	0.0094	0.2262	0.5291
(88,85)	1	0	0.0094	0.1610	0.4196
(98,73)	2	0.0079	0.0111	0.1096	0.7768
(94,77)	2	0.0039	0.0094	0.1850	0.5996
(79,99)	2	0	0.0055	0.1278	0.2429
(89,77)	1	0.0039	0.0094	0.1614	0.4625
(71,91)	1	0	0.0032	0.0934	0.1724
(84,78)	2	0.0039	0.0094	0.1191	0.3776
(55,94)	2	0	0.0020	0.4491	0.0753
(70,92)	2	0	0.0020	0.0933	0.1557
(87,74)	2	0.0051	0.0094	0.1049	0.3985
(71,90)	1	0	0.0032	0.0607	0.1724
(69,91)	2	0	0.0020	0.0768	0.1393
(73,84)	2	0	0.0043	0.0607	0.1895
(83,77)	2	0.0025	0.0080	0.0991	0.3377
(81,76)	2	0.0025	0.0080	0.0449	0.2611
(81,64)	2	0.0060	0.0080	0.0449	0.2611
(84,54)	1	0.0084	0.0094	0.0841	0.3776
(83,40)	2	0.0080	0.0080	0.0347	0.3377
(62,77)	1	0	0.0020	0.0607	0.0910
(75,64)	1	0.0035	0.0055	0.0449	0.1895
(54,39)	1	0.0010	0.0010	0.0154	0.0753
(45,50)	2	0	0	0.0148	0.0295
(32,51)	1	0	0	0.0147	0.0147

Table 6.19. Estimates of the baseline cumulative cause-specific hazard functions

(t_1,t_2)	z	$\hat{\Lambda}_{210}(t_1,t_2)$	$\hat{\Lambda}_{210}(0,t_2)$	$\hat{\Lambda}_{220}(t_1,t_2)$	$\hat{\Lambda}_{220}(0,t_2)$
(98,83)	1	0	3.69×10^{-36}	0.7944	1.1608
(84,94)	1	5.65×10^{-37}	4.26×10^{-36}	0.8577	2.2217
(98,78)	1	0	3.18×10^{-36}	0.6151	0.9220
(89,89)	1	0	4.26×10^{-36}	0.5825	1.3532
(97,79)	2	0	3.69×10^{-36}	0.5129	0.9807
(93,80)	1	0	3.69×10^{-36}	0.5137	1.0997
(80,92)	1	5.65×10^{-37}	4.26×10^{-36}	0.5491	1.9126
(98,76)	1	0	2.22×10^{-36}	0.3955	0.6482
(91,90)	2	0	4.26×10^{-36}	0.7165	1.4871
(88,85)	1	0	4.26×10^{-36}	0.4629	1.2235
(98,73)	2	0	2.22×10^{-36}	0.2384	0.3878
(94,77)	2	0	2.68×10^{-36}	0.4542	0.8648
(79,99)	2	1.08×10^{-36}	4.26×10^{-36}	0.6420	2.7490
(89,77)	1	0	2.68×10^{-36}	0.4542	0.8648
(71,91)	1	2.04×10^{-36}	4.26×10^{-36}	0.3507	1.7632
(84,78)	2	0	3.18×10^{-36}	0.3487	0.9220
(55,94)	2	3.18×10^{-36}	4.26×10^{-36}	0.1689	2.2217
(70,92)	2	2.04×10^{-36}	4.26×10^{-36}	0.2895	1.9126
(87,74)	2	0	2.22×10^{-36}	0.2389	0.4905
(71,90)	1	2.04×10^{-36}	4.26×10^{-36}	0.2127	1.4871
(69,91)	2	2.04×10^{-36}	4.26×10^{-36}	0.2147	1.7632
(73,84)	2	2.04×10^{-36}	4.26×10^{-36}	0.1457	1.1608
(83,77)	2	0	2.68×10^{-36}	0.3016	0.8648
(81,76)	2	0	2.22×10^{-36}	0.1933	0.6482
(81,64)	2	0	1.45×10^{-36}	0.1404	0.2357
(84,54)	1	0	1.07×10^{-36}	01387	0.1387
(83,40)	2	0	3.50×10^{-37}	0.0452	0.0452
(62,77)	1	1.60×10^{-36}	2.68×10^{-36}	0.1457	0.8648
(75,64)	1	0	1.45×10^{-36}	0.1405	0.2357
(54,39)	1	0	0	0.0452	0.0452
(45,50)	2	3.57×10^{-37}	7.08×10^{-37}	0	0.0452
(32,51)	1	7.08×10^{-37}	7.08×10^{-37}	0.0464	0.0916

Table 6.20. Estimators of the baseline cumulative cause-specific hazard functions

(t_1,t_2)	$\underline{z}$	$\hat{\Lambda}_{3110}(t_1,t_2)$	$\hat{\Lambda}_{3120}(t_1,t_2)$	$\hat{\Lambda}_{3210}(t_1,t_2)$	$\hat{\Lambda}_{3220}(t_1,t_2)$
(98,83)	1	$7.24\text{x}10^{-38}$	0.0179	$2.63\ \text{x}10^{-36}$	1.3016
(84,94)	1	$1.16\ \text{x}10^{-37}$	0.0412	$1.40\ \text{x}10^{-36}$	0.9507
(98,78)	1	$7.24\ \text{x}10^{-38}$	0.0179	$2.63\ \text{x}10^{-36}$	0.9295
(89,89)	1	$1.16\ \text{x}10^{-37}$	0.0179	$1.40\ \text{x}10^{-36}$	0.8112
(97,79)	2	$7.24\ \text{x}10^{-38}$	0.0179	$2.63\ \text{x}10^{-36}$	0.7025
(93,80)	1	$7.24\ \text{x}10^{-38}$	0.0179	$1.40\ \text{x}10^{-36}$	0.6949
(80,92)	1	$4.32\ \text{x}10^{-38}$	0.0194	$6.02\ \text{x}10^{-37}$	0.6412
(98,76)	1	$2.92\ \text{x}10^{-38}$	0.0179	$1.40\ \text{x}10^{-36}$	0.5354
(91,90)	2	$1.16\ \text{x}10^{-37}$	0.0293	$1.40\ \text{x}10^{-36}$	0.9357
(88,85)	1	$1.16\ \text{x}10^{-37}$	0.0179	$1.40\ \text{x}10^{-36}$	0.5895
(98,73)	2	$2.92\ \text{x}10^{-38}$	0.0080	$1.40\ \text{x}10^{-36}$	0.3248
(94,77)	2	$2.92\ \text{x}10^{-38}$	0.0179	$2.63\ \text{x}10^{-36}$	0.5769
(79,99)	2	$4.32\ \text{x}10^{-38}$	0.0313	$6.02\ \text{x}10^{-37}$	0.6699
(89,77)	1	$2.92\ \text{x}10^{-38}$	0.0179	$1.40\text{x}10^{-36}$	0.5769
(71,91)	1	0	0.0114	$6.02\ \text{x}10^{-37}$	0.4136
(84,78)	2	$7.24\ \text{x}10^{-38}$	0.0179	$1.40\ \text{x}10^{-36}$	0.3832
(55,94)	2	0	0.0119	$6.02\ \text{x}10^{-37}$	0.1296
(70,92)	2	0	0	$6.02\ \text{x}10^{-37}$	0.4220
(87,74)	2	$2.92\ \text{x}10^{-38}$	0.0080	$1.40\ \text{x}10^{-36}$	0.3055
(71,90)	1	0	0.0114	$6.02\ \text{x}10^{-37}$	0.2074
(69,91)	2	0	0	$6.02\ \text{x}10^{-37}$	0.3085
(73,84)	2	$4.32\ \text{x}10^{-38}$	0	$6.02\ \text{x}10^{-37}$	0.2074
(83,77)	2	$2.92\ \text{x}10^{-38}$	0.0179	$1.40\ \text{x}10^{-36}$	0.3001
(81,76)	2	$2.92\ \text{x}10^{-38}$	0.0179	$6.02\ \text{x}10^{-37}$	0.1296
(81,64)	2	$2.92\ \text{x}10^{-38}$	0.0080	$6.02\ \text{x}10^{-37}$	0.1296
(84,54)	1	0	0	$1.40\ \text{x}10^{-36}$	0.2128
(83,40)	2	0	0	$7.97\ \text{x}10^{-37}$	0.0659
(62,77)	1	0	0	$6.02\ \text{x}10^{-37}$	0.2074
(75,64)	1	0	0.0080	$6.02\ \text{x}10^{-37}$	0.1296
(54,39)	1	0	0	0	0.0659
(45,50)	2	0	0	$6.02\ \text{x}10^{-37}$	0
(32,51)	1	0	0	0	0.0637

Table 6.21. Estimates of the bivariate survival function with $a = 0.5$

(t_1, t_2)	$\underline{z}$	$\hat{S}(t_1, t_2 \mid \underline{z})$
(98,83)	1	0.4611
(84,94)	1	0.4745
(98,78)	1	0.5084
(89,89)	1	0.5503
(97,79)	2	0.5613
(93,80)	1	0.5615
(80,92)	1	0.5732
(98,76)	1	0.5799
(91,90)	2	0.5806
(88,85)	1	0.5929
(98,73)	2	0.6042
(94,77)	2	0.6139
(79,99)	2	0.6152
(89,77)	1	0.6160
(71,91)	1	0.6438
(84,78)	2	0.6595
(55,94)	2	0.6774
(70,92)	2	0.6805
(87,74)	2	0.6911
(71,90)	1	0.6939
(69,91)	2	0.6954
(73,84)	2	0.6976
(83,77)	2	0.7005
(81,76)	2	0.7514
(81,64)	2	0.7666
(84,54)	1	0.7762
(83,40)	2	0.7769
(62,77)	1	0.7985
(75,64)	1	0.8438
(54,39)	1	0.9477
(45,50)	2	0.9550
(32,51)	1	0.9664

Table 6.22. Estimates of the cause-specific sub-distribution functions $F_{pq}(t_1, t_2 \mid \underline{z})$

(t_1, t_2)	$\underline{z}$	$\hat{F}_{11}(t_1, t_2 \mid \underline{z})$	$\hat{F}_{21}(t_1, t_2 \mid \underline{z})$	$\hat{F}_{12}(t_1, t_2 \mid \underline{z})$	$\hat{F}_{22}(t_1, t_2 \mid \underline{z})$
(98,83)	1	3.34×10^{-38}	1.21×10^{-36}	0.0082	0.6002
(84,94)	1	5.49×10^{-38}	6.63×10^{-37}	0.0195	0.0451
(98,78)	1	3.68×10^{-38}	1.34×10^{-36}	0.0091	0.4725
(89,89)	1	6.36×10^{-38}	7.69×10^{-37}	0.0098	0.4465
(97,79)	2	4.07×10^{-38}	1.47×10^{-36}	0.0100	0.3943
(93,80)	1	4.07×10^{-38}	7.85×10^{-37}	0.0100	0.3902
(80,92)	1	2.48×10^{-38}	3.45×10^{-37}	0.0111	0.3698
(98,76)	1	1.69×10^{-38}	8.11×10^{-37}	0.0104	0.3105
(91,90)	2	6.71×10^{-38}	8.12×10^{-37}	0.0170	0.5433
(88,85)	1	6.86×10^{-38}	8.29×10^{-37}	0.0106	0.3495
(98,73)	2	1.76×10^{-38}	8.45×10^{-37}	0.0048	0.1962
(94,77)	2	1.79×10^{-38}	1.61×10^{-36}	0.0109	0.3542
(79,99)	2	2.66×10^{-38}	3.70×10^{-37}	0.0193	0.4121
(89,77)	1	1.80×10^{-38}	8.61×10^{-37}	0.0110	0.3553
(71,91)	1	0	3.87×10^{-37}	0.0074	0.2663
(84,78)	2	4.78×10^{-38}	9.22×10^{-37}	0.0118	0.2527
(55,94)	2	0	4.07×10^{-37}	0.0081	0.0878
(70,92)	2	0	4.09×10^{-37}	0	0.2872
(87,74)	2	2.02×10^{-38}	9.66×10^{-37}	0.0055	0.2111
(71,90)	1	0	4.17×10^{-37}	0.0079	0.1439
(69,91)	2	0	4.18×10^{-37}	0	0.2145
(73,84)	2	3.02×10^{-38}	4.20×10^{-37}	0	0.1447
(83,77)	2	2.05×10^{-38}	9.79×10^{-37}	0.0125	0.2102
(81,76)	2	2.19×10^{-38}	4.52×10^{-37}	0.0134	0.0974
(81,64)	2	2.24×10^{-38}	4.61×10^{-37}	0.0061	0.0994
(84,54)	1	0	1.09×10^{-36}	0	0.1652
(83,40)	2	0	6.19×10^{-37}	0	0.0512
(62,77)	1	0	4.80×10^{-37}	0	0.1656
(75,64)	1	0	5.08×10^{-37}	0.0068	0.1094
(54,39)	1	0	0	0	0.0624
(45,50)	2	0	5.74×10^{-37}	0	0
(32,51)	1	0	0	0	0.0616

Table 6.23. Estimates of $\pi_p^{(q)*}$ and π_{pq}^{*}

$\hat{\pi}_1^{(1)*}$	$\hat{\pi}_2^{(1)*}$	$\hat{\pi}_1^{(2)*}$	$\hat{\pi}_2^{(2)*}$	$\hat{\pi}_{11}^{*}$	$\hat{\pi}_{12}^{*}$	$\hat{\pi}_{21}^{*}$	$\hat{\pi}_{22}^{*}$
$\frac{11}{52}$	$\frac{41}{52}$	$\frac{11}{78}$	$\frac{67}{78}$	$\frac{3}{43}$	$\frac{5}{43}$	$\frac{3}{43}$	$\frac{32}{43}$

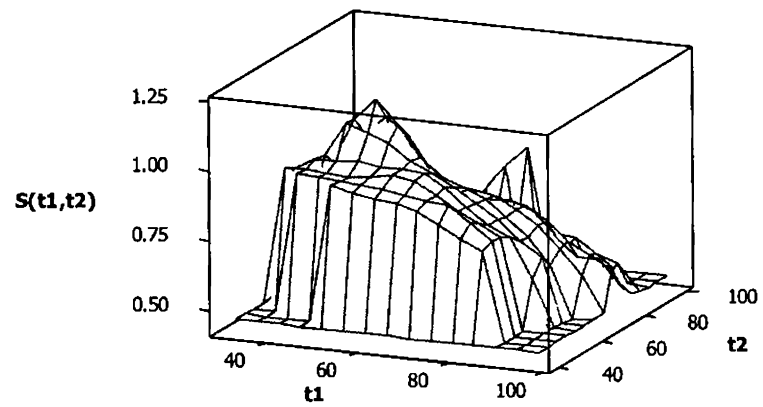

Figure 6.1. Estimates of bivariate survival function with $a = 0.5$

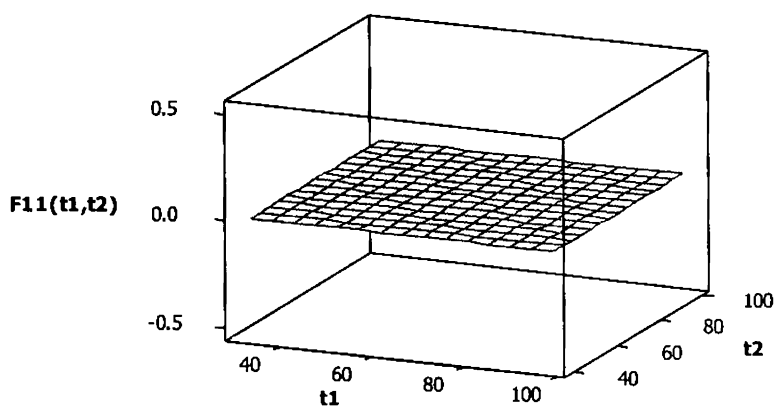

Figure 6.2. Estimates of cause-specific sub-distribution function for the cause (1, 1)

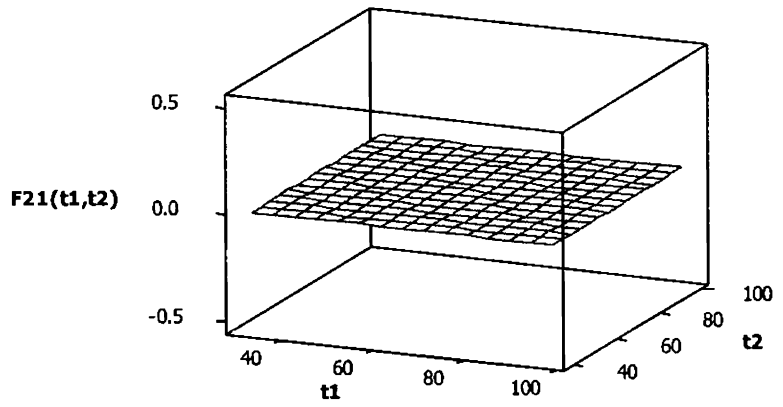

Figure 6.3. Estimates of cause-specific sub-distribution function for the cause (2, 1)

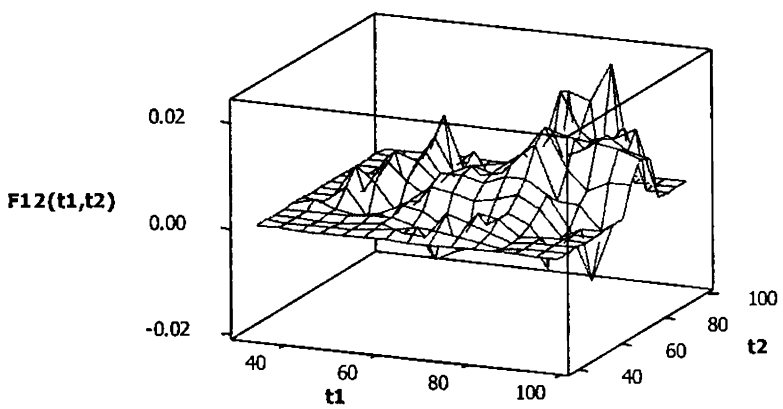

Figure 6.4. Estimates of cause-specific sub-distribution function for the cause (1, 2)

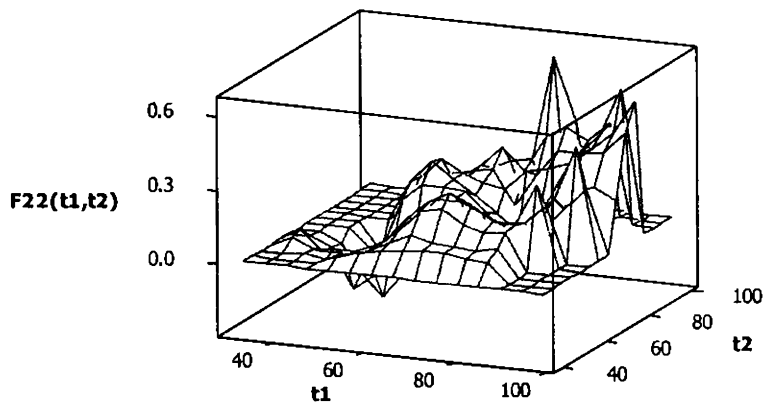

Figure 6.5. Estimates of cause-specific sub-distribution function for the cause (2, 2)

6.7. Conclusion

This chapter has discussed statistical analysis of bivariate competing risks models in the presence of covariates. We introduced proportional hazards models for cause-specific hazard function. Estimation of regression parameters and baseline cause-specific hazard functions were developed using counting process approach. Strong consistency and asymptotic normality of the estimator were established. The proposed method can be extended to higher dimensions by considering the vector hazard functions of Johnson and Kotz (1975) in multivariate set up. The proposed method can also be extended to time dependent covariates.

Cause-specific sub-distribution function is more helpful in many survival studies as it directly provides the failure probabilities due to a particular set of causes (Fine, 2001). The development of semi-parametric proportional hazards models for the cause-specific sub-distribution function, extending the model of Fine and Gray (1999) to the multivariate set up is a topic of further research.

Chapter Seven

Conclusion

7.1. Introduction

Multivariate lifetime data arise in various forms including recurrent event data when individuals are followed to observe the sequence of occurrences of a certain type of event; correlated lifetime when an individual is followed for the occurrence of two or more types of events, or when distinct individuals have dependent event times. In most studies there are covariates such as treatments, group indicators, individual characteristics, or environmental conditions, whose relationship to lifetime is of interest. This leads to a consideration of regression models.

The well known Cox proportional hazards model and its variations, using the marginal hazard functions employed for the analysis of multivariate survival data in literature are not sufficient to explain the complete dependence structure of pair of lifetimes on the covariate vector. Motivated by this, in Chapter 2, we introduced a bivariate proportional hazards model using vector hazard function of Johnson and Kotz (1975), in which the covariates under study have different effect on two components of the vector hazard function. The proposed model is useful in real life situations to study the dependence structure of pair of lifetimes on the covariate vector $\underline{z}$. The well known partial likelihood approach is used for the estimation of parameter vectors. We then introduced a bivariate proportional hazards model for gap times of recurrent events in Chapter 3. The model incorporates both marginal and joint dependence of the distribution of gap times on the covariate vector $\underline{z}$. In many fields of application, mean residual life function is considered superior concept than the hazard function. Motivated by this, in Chapter 4, we considered a new semi-parametric model, bivariate proportional mean residual life time model, to assess the relationship between mean residual life and covariates for gap time of recurrent events. The counting process approach is used for the inference procedures

of the gap time of recurrent events. In many survival studies, the distribution of lifetime may depend on the distribution of censoring time. In Chapter 5, we introduced a proportional hazards model for duration times and developed inference procedures under dependent (informative) censoring. In Chapter 6, we introduced a bivariate proportional hazards model for competing risks data under right censoring. The asymptotic properties of the estimators of the parameters of different models developed in previous chapters, were studied. The proposed models were applied to various real life situations.

7.2. Future works

Multivariate lifetime data are frequently encountered in longitudinal studies when subjects may experience several events or when there is a grouping of individuals into a cluster. The heterogeneity among variables may be due to certain unobserved common risk factors present in the data. To model such unobserved factors, frailty models are usually employed in survival analysis. Frailty models are basically random effects models for survival data, where one of the random effects is specified by means of the hazard function. The extension of our models to frailty set up is an area for further research. As the literature on the analysis of multivariate data under informative censoring is limited, one can do further research on the development of new stochastic models based on the relationship between lifetime vector and censoring vector. In survival studies, the covariates under study may change their values over time. The extension of the models developed in previous chapters to the time dependent covariate set up is an area of research to be explored. Testing equality of survival functions in multivariate set up in the presence of covariates is not yet carried out. The non-parametric Bayesian estimation technique for the analysis of multivariate survival data is a topic of research interest. In survival studies, there are situations where the exact lifetime of an event is not known, but it known to lie in some interval. A variety of univariate models have been developed for the analysis of such interval censored data. The analysis of interval censored data in the bivariate (multivariate) set up is complicated and the research work in this direction will be worth exploring.

In all the works mentioned in previous chapters, we have considered either proportional hazards model or proportional mean residual life model to study effect of covariates on lifetime. There are many life test situations in which these models are not adequate. The analysis of multivariate survival data using proportional odds and accelerated failure time models is an area of research to be explored. Apart from censoring, truncation is very common in life test experiments. As the literature on the analysis of truncated data in multivariate set up is limited, the extension of our models to the truncated case is a topic of future work. There are many situations in survival studies when the covariates in the study are missing and our models can be extended to this situation. In biostatistical applications, there are situations were one can only observe the lifetime T belongs to certain interval $(0, C]$ or (C, ∞) where C is known as status time. Then the data structure is called current status data. Very few works are done in multivariate current status data when covariates are present. The models developed here can be extended to this set up, which is not straight forward. There are situations in the analysis of competing risks data, where the exact failure cause cannot be identified. Then we say that the cause of failure is masked. The analysis of the model given in Chapter 6 can be extended to the masking situation, which is a topic of future study.

References

1. Aalen, O. (1976) Nonparametric inference in connection with multiple decrement models, Scanadian Journal of Statistics, **3**, 15-27.
2. Akritas, M.G. and van Keilegom, I.V. (2003) Estimation of bivariate and marginal distributions with censored data, Journal of the Royal Statistical Society B, **65**, 457-471.
3. Andersen, P.K., Abildstrom, S.Z. and Rosthoj, S. (2002) Competing risks as a multi-state model, Statistical Methods in Medical Research, **11**, 203-215.
4. Andersen, P.K., Borgan, O., Gill and R.D., Keiding, N. (1993) *Statistical Models Based on Counting Processes*, Springer-Verlag, New York.
5. Andersen, P.K. and Gill, R.D. (1982) Cox's regression model for counting processes: a large sample study, The Annals of Statistics, **10**, 1100-1120.
6. Antony, A.A. and Sankaran, P.G. (2005) Estimation of bivariate survivor function under masked causes of failure, Journal of Statistical Theory and Applications, 4, 401-423.
7. Arnold, B.C. and Strauss, D. (1988) Bivariate distributions with exponential conditionals, Journal of the American Statistical Association, **83**, 522-527.
8. Arnold, B.C. and Zahedi, H. (1988) On multivariate mean remaining life functions, Journal of Multivariate Analysis, **25**, 1-9.
9. Bartholomew, D.J. (1957) A problem in life testing, Journal of the American Statistical Association, **52**, 350-355.
10. Basu, A.P. (1971) Bivariate failure rate, Journal of the American Statistical Association, **66**, 103-104.
11. Braekers, R. and Veraverbeke, N. (2005) Cox's regression model under partially informative censoring, Communications in Statistics-Theory and Methods, **34**, 1793-1811.
12. Buchanan, W.B. and Singpurwalla, N.D. (1977) Some stochastic characterization of multivariate survival, Theory and Applications of Reliability, **1**, Eds. C.P. Tsokos and I.N.Shimi, Academic Press, New York, 329-348.

13. Burke, M.D. (1988) Estimation of a distribution function under random censorship, Biometrika, **75**, 379-382.
14. Cai, J. and Douglas, E.S. (2004) Analysis of recurrent event data, Handbook of Statistics, **23**, 603-623.
15. Cai, J. and Prentice, R.L. (1995) Estimating equations for hazard ratio parameters based on correlated failure time data, Biometrika, **82**, 151-164.
16. Campbell, G. and Foldes, A. (1982) Large sample properties of nonparametric bivariate estimators with censored data, Colloquia Mathemetica-Societatis, Janos Bolyai, **32**, 103-121.
17. Chang, S.H. (2000) A two-sample comparison for multiple ordered event data, Biometrics, **56**, 183-189.
18. Chang, S. H. and Wang, M. C. (1999) Conditional regression analysis for recurrence time data, Journal of the American Statistical Association, **94**, 1221-1230.
19. Chen, Y.Q. and Cheng, S. (2005) Semiparametric regression analysis of mean residual life with censored survival data, Biometrika, **92**, 19-29.
20. Chen, Y.Q. and Cheng, S. (2006) Linear life expectancy regression with censored data, Biometrika, **93**, 303-313.
21. Chen, Y. Q., Wang, M.C. and Huang, Y. (2004) Semi parametric regression analysis on longitudinal pattern of recurrent gap times, Biostatistics, **5**, 277-290.
22. Cheng, S.C., Fine, J.P. and Wei, L.J. (1998) Prediction of cumulative incidence function under the proportional hazards model, Biometrics, **54**, 219-228.
23. Cox, D.R. (1972) Regression models and life tables (with discussion), Journal of the Royal Statistical Society B, **34**,187-220.
24. Cox, D.R. (1975) Partial likelihood, Biometrika, 62, 269-276.
25. Cox, D.R. and Oakes, D. (1984) *Analysis of Survival Data*, Chapman and Hall, London.
26. Cronin, K.A. and Feuer, E.J. (2000) Cumulative Cause-Specific mortality for cancer patients in the presence of other causes: A crude analogue of relative survival, Statistics in Medicine, **19**, 1729-1740.
27. Crowder, M. (2001) *Classical competing risks*, Chapman and Hall, London.
28. Dabrowska, D.M. (1988) Kaplan- Meier estimate on the plane, The Annals of Statistics, **16**, 1475-1489.

29. David, H.A. and Moeschberger, M.L. (1978) *Theory of Competing Risks.* Griffin, London.

30. DeMasi, R.A. (2000) Statistical methods for multivariate failure time data and competing risks, Handbook of Statistics, **18**, 749-781.

31. Devroye, L. (1986) *Non-Uniform Random Variate Generation,* Springer-Verlag, New York

32. Duffy, D.L., Martin N.G. and Matthews, J.D. (1990) Appendectomy in Australian twins, The American Journal of Human Genetics, **47**, 590-592.

33. Ebrahimi, N. (2006) Models for recurring events with marginal proportional hazards, Biometrika, **93**, 481-485.

34. Efron, B. (1981) Censored data and the bootstrap, Journal of the American Statistical Association, **76**, 312-319.

35. Efron, B. and Tibshirani, R.J. (1993) *An introduction to the bootstrap*, Chapman and Hall.

36. Farley, T.M., Ali, M.M. and Slaymaker, E. (2001) Competing approaches to analysis of failure times with competing risks, Statistics in Medicine, **120**, 3601-3610.

37. Fine, J.P. (2001) Regression modeling of competing crude failure probabilities, Biostatistics, **2**, 85-97.

38. Fine, J.P. and Gray, R.J. (1999) A proportional hazards model for the sub distribution of a competing risk, Journal of the American Statistical Association, **94**, 496-509.

39. Fiocco, M., Putter, H. and Van Houwelingen (2005) Reduced rank proportional hazards model for competing risks, Biostatistics, **6**, 465-478.

40. Fleming, T. R. and Harrington, D. P. (1991) *Counting Processes and Survival Analysis*, Wiley, New York.

41. Gail, M.H., Santner, T.J. and Brown, C.C. (1980) An analysis of comparative carcinogenesis experiments based on multiple times to tumor, Biometrics, **36**, 255-266.

42. Gehan, E.A. (1965) A generalized Wilcoxon test for comparing arbitrarily singly-censored samples, Biometrika, **52**, 203-233.

43. Gelfand, A.E., Ghosh, S.K., Christensen, C., Soumerai, S.B. and McLaughlin, T.J. (2000) Proportional hazards models: a latent competing risk approach, Applied Statistics, **49**, 385-397.

44. Gentle, J.E. (1998) *Random number generation and Monte Carlo methods* (Statistics and computing), Springer-Verlag, New York.

45. Gichangi, A. and Vach, W. (2005) The analysis of competing risks data: A guided tour. Technical report, Department of Statistics, University of Southern denmark, Campusvej 55, 5230 Odense M, 5000 Odense C. Denmark.

46. Gooley, T.A., Leisenring, W., Crowley, J. and Storer, B.E. (1999) Estimation of failure probabilities in the presence of competing risks: New representations of old estimators, Statistics in Medicine, **18**, 695-706.

47. Gumbel, E.J.(1960) Bivariate exponential distributions, Journal of the American Statistical Association, **55**, 698-707.

48. Hall, W.J. and Wellner, J.A. (1981) Mean residual life. In Proceedings of International Symposiam on Statistics and Related Topics, Ed. M. Csorgo, D.A. Dawson, J. N. K. Rao and A. K. Md. E. Saleh, pp. 169-181, Amsterdam: North-Holland.

49. Hoel, D.G. (1972) A representation of mortality by competing risks, Biometrics, **28**, 475-488.

50. Hougaard, P. (2000) *Analysis of Multivariate Survival Data*, Springer-Verlag, New York.

51. Huang, Y. (1999) The two-sample problem with induced dependent censorship, Biometrics, **55**, 1108-1113.

52. Huang, Y. and Louis, T.A. (1998) Nonparametric estimation of the joint distribution of survival time and mark variable, Biometrika, **85**, 785-798.

53. Huang, C.Y. and Wang, M.C. (2005) Nonparametric estimation of the bivariate recurrence time distribution, Biometrics, **61**, 392-402.

54. Johnson, N.L. and Kotz, S. (1975) A vector valued multivariate hazard rate, Journal of Multivariate Analysis, **5**, 53-66.

55. Kalbfleisch, J.D. and Lawless, J.F. (1989) Inference based on retrospective ascertainment: An analysis of the data on transfusion-related AIDS, Journal of the American Statistical Association. **84**, 360-372.

56. Kalbfleisch, J.D. and Prentice, R.L. (2002) *The Statistical Analysis of Failure Time Data*, 2nd ed., John Wiley and Sons, New York.

57. Kaplan.E.L. and Meier, P. (1958) Nonparametric estimation from incomplete observations, Journal of the American Statistical Association, 53, 457-481.

58. Keles, S., van der Laan, M.J. and Robins, J.M. (2004) Estimation of the bivariate survival function with generalized bivariate right censored data structures, Handbook of Statistics, **23,** 143-173.

59. Klein, J.P. and Moeschberger, M.L. (1997) *Survival Analysis.* Springer-Verlag, New York.

60. Lanson, M.G. (1984) Covariate analysis of competing risk models with long linear models, Biometrics, **40**, 459-469.

61. Lawless, J.F. (2003) *Statistical Models and Methods for Lifetime Data,* 2nd ed., John Wiley and Sons, New York.

62. Lawless, J.F. and Nadeau, C. (1995) Some simple robust methods for the analysis of recurrent events, Technometrics, **37**, 158-168.

63. Lee, E.W., Wei, L.J. and Amato, D.A. (1992) Cox-type regression analysis for large numbers of small groups of correlated failure time observations, in Survival Analysis: State of the Art, eds. J.P.Klein and P.K. Goel, Dordrecht: Kluwer Academic:237-247.

64. Liang, K.Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models, Biometrika, **73**:13-22.

65. Lin, D.Y. (1994) Cox regression analysis of multivariate failure time data: The marginal approach, Statistics in Medicine, **13**: 2233-2247.

66. Lin, D.Y. (1997) Non-parametric inference for cumulative incidence functions in competing risks studies, Statistics in Medicine, **16**, 901-910.

67. Lin, D. Y. (2000) Linear regression analysis of censored medical costs, Biostatistics, **1**, 35-47.

68. Lin, D. Y., Sun, W. and Ying, Z. (1999) Nonparametric estimation of gap time distributions for serial events with censored data, Biometrika, **86**, 59-70.

69. Lin, D.Y., Wei, L.J., Yang, I. and Ying, Z. (2000) Semi-parametric regression for the mean and rate functions of recurrent events, Journal of the Royal Statistical Society B, **62**, 711-730.

70. Maguluri, G. and Zhang, C.H. (1994) Estimation in the mean residual life regression model, Journal of the Royal Statistical Society B, **56**, 477-489.

71. Makeham, W.M. (1874) On an application of the theory of the composition of decremental forces, Journal of the Institute of Actuaries (London), **18**, 317-322.

72. Mantel, N., Bohidar, N.R. and Ciminera, J.L. (1977) Mantel-haenszel analyses of litter-matched time-to-response data, with modifications for recovery of interlitter information, Cancer Research, **37**, 3863-3868.

73. Mantel, N. and Ciminera, J.L. (1979) Use of log rank series in the analysis of litter-matched data on time to tumor appearance, Cancer Research, **39**, 4308-4315.

74. Martinussen, T. and Scheike, T.H. (2006) *Dynamic Regression Models for Survival Data*, Springer verlag, New York.

75. Martz, H.F. and Waller, R.A. (1982) *Bayesian Reliability Analysis*, John Wiley and Sons, New York.

76. Murphy, S.A. and van der Vaart, A.W. (2000) On profile likelihood, Journal of the American Statistical Society, **95**, 449-485.

77. Muth, E.J. (1977) Reliability models with positive memory derived from the mean residual life function, In theory and applications in reliability, eds. C.P. Tsokos and I.N. Shimi, Academic Press, New York.

78. Nelson, W.B. (1969) Hazard plotting for incomplete failure data, Journal of Quality and Technology, **1**, 27-52.

79. Nelson, W.B. (1970) Hazard plotting methods for analysis of life data with different failure modes, Journal of Quality and Technology, **2**, 126-149.

80. Nelson, W.B. (1972) Graphical analysis of accelerated life test data with the inverse power law model, IEEE Transactions on Reliability, **21**, 2-11.

81. Oakes, D. (1989) Bivariate survival models induced by frailties, Journal of the American Statistical Association, **84**, 487-493.

82. Oakes, D. (2001) Biometrika centenary: Survival analysis, Biomatrika, **88**, 99-142.

83. Oakes, D. and Dasu, T. (1990) A note on residual life, Biometrika, **77**, 409-410.

84. Oakes, D. and Dasu, T. (2003) Inference for the proportional mean residual life model, In crossing boundaries: Statistical essays in honor of Jack Hall, Institute

of Mathematical Statistics lecture notes monograph series, **43**, Ed. J. E. Kolassa and D. Oakes,. Hayward, CA: Institute of Mathematical Statistics, 105-116.

85. Pena, E.A., Strawderman, R.L. and Hollander, M. (2001) Nonparametric estimation with recurrent event data, Journal of the American Statistical Association, **96**, 1299-1315.
86. Pepe, M.S. and Cai, J. (1993) Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates, Journal of the American Statistical Association, **88**, 811-820.
87. Prentice, R.L. and Cai, J. (1992) Covariance and survival function estimation using censored multivariate failure time data, Biometrika, **79**, 495-512.
88. Prentice, R.L. and Kalbfleisch, J.D. (2003) Aspects of the analysis of multivariate failure time data, Statistics and Operations Research Transactions, **27**, 65-78.
89. Prentice, R.L., Kalbfleisch, J.D. , Peterson, A.V., Flournoy, N., Farewell, V.T. and Breslow, N.E. (1978) The analysis of failure times in the presence of competing risks, Biometrics, **34**, 541-554.
90. Prentice, R.L.,Williams, B.J. and Peterson, A.V. (1981) On the regression analysis of multivariate failure time data, Biometrika, **68**, 373-379.
91. Pruitt, R.C. (1991) Strong consistency of self-consistent estimators: general theory and an application to bivariate survival analysis, Technical Report **543**, Univarsity of Minneapolis, U.S.A.
92. Pruitt, R.C. (1993) Small sample comparisons of five bivariate survival estimators. Journal of Statistical Computing and Simulation, **45**,147-167.
93. Quale C.M., van der Laan M.J. and Robins J.M. (2003) Locally efficient estimation with bivariate right censored data. University of California, Berkeley, Department of Statistics, Technical Report.
94. Robins, J.M., Rotnitzky, A. (1992) Recovery of information and adjustment of dependent censoring using surrogate markers. In AIDS Epidemiology. Methodological issues, Ed. N.P.Jewell, K. Dietz and V. Farewell, 297-331. Boston: Birkhauser.
95. Sankaran, P.G. and Ansa, A.A. (2008a) Non parametric estimation of lifetime distribution of competing risk models when censoring times are missing, Statistical Papers (to appear).

96. Sankaran P.G., and Ansa, A.A. (2008b) Non-parametric estimation of bivariate survivor function under masked causes of failure, Journal of Nonparametric Statistics, **20**, 77-89.

97. Sankaran, P.G., Lawless, J.F., Abraham, B. and Ansa, A. A. (2006) Non-parametric estimation of distribution functions in bivariate competing risk models, Biometrical Journal, **48(3)**, 399-410.

98. Sankaran, P.G. and Sreeja, V.N. (2007) Proportional hazards model for multivariate failure time data, Communications in Statistics-Theory and Methods **36(8)**, 1627- 1642.

99. Sankaran, P.G. and Sreeja, V.N. (2008) Proportional Hazards Model for Gap Time Distributions of Recurrence Events, Communicated.

100. Shanbag, D.N. and Kotz, S. (1987) Some new approaches to multivariate probability distributions, Journal of Multivariate Analysis, **22**, 189-211.

101. Sinha, S.K. (1986) *Reliability and Life Testing*, Wiley Eastern Limited, New Delhi.

102. Spiekerman, C.F. and Lin, D. Y. (1998) Marginal regression models for multivariate failure time data, Journal of the American Statistical Association, **93**, 1164-1175.

103. Sreeja, V.N. and Sankaran, P.G. (2007) Proportional Mean Residual Life Model for Gap Time Distributions of Recurrent Events, Metron, **LXV**, n.3 (to appear).

104. Sreeja, V.N. and Sankaran, P.G. (2008a) Proportional Hazards Model for Successive Duration Times under Informative Censoring, Communicated.

105. Sreeja, V.N. and Sankaran, P.G. (2008b) Proportional Hazards Model for Bivariate Competing Risks Data, Communicated.

106. Strawderman, R.L. (2005) The accelerated gap times model, Biometrika, **92**, 647- 666.

107. Swartz, G.B. (1973) The mean residual life time function, IEEE Transactions on Reliability, **22**, 108.

108. Tsiatis, A.A. (1981) A large sample study of Cox's regression model, The Annals of Statistics, 9, 93- 108.

109. van der Laan, M. J. (1996) Efficient estimation in the bivariate censoring model and repairing NPMLE, The Annals of Statistics, **24**, 596-627.

110. van der Laan, M. J. (1997) Non parametric estimator of the bivariate survival function under random censoring, Statistica Neerlandica, **51**, 178-200.

111. van der Laan, M. J., Hubbard, A.E. and James, M.R. (2002) Locally efficient estimation of a multivariate survivor function in longitudinal studies, Journal of the American Statistical Association, **97**, 494-507.

112. Visser, M.(1996) Nonparametric estimation of the bivariate survival function with an application to vertically transmitted AIDS, Biometrika, **83**, 507-518.

113. Wang, M. C. and Chang, S. H. (1999) Nonparametric estimation of a recurrent survival function, Journal of the American Statistical Association, **94**, 146-153.

114. Wang, M. C., Qin, J. and Chiang, C. T. (2001) Analyzing recurrent event data with informative censoring, Journal of the American Statistical Association, **96**, 1057-1065.

115. Wang, W. and Wells. M.T. (1997) Non-parametric estimators of the bivariate survival function under simplified censoring conditions, Biometrika, **84**, 863-880.

116. Wang, W. and Wells, M.T. (1998) Nonparametric estimation of successive duration times under dependent censoring, Biometrika, **86**, 59-70.

117. Wang, W. and Wells. M.T. (1999) Semi-parametric estimation of the bivariate survival function, Technical Report, Institute of Statistics, Academia Sinica, Taipei.

118. Wei, L.J., Lin, D.Y. and Weissfeld, L. (1989) Regression analysis of multivariate incomplete failure time data by modeling marginal distributions, Journal of the American Statistical Association, **84**, 1065-1073.

119. Ying, Z. and Wei, L.J. (1994) The Kaplan-Meier estimate for dependent failure time observations, Journal of Multivariate Analysis, **50**, 17-29.