2006

Generalized estimating equations for clustered survival data

Xiaohong Zhang
Iowa State University

Generalized estimating equations for clustered survival data

by

Xiaohong Zhang

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Kenneth Koehler, Co-major Professor
Terry Therneau, Co-major Professor
Mervyn Marasinghe
Max Morris
Richard Evans

Iowa State University

Ames, Iowa

2006

UMI Number: 3243545

UMI Microform 3243545

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

To my parents

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Kenneth Koehler for his guidance, helpful suggestions, long hours of discussion, and thorough review of this work within his busy schedule. Thanks to Dr. Terry Therneau for providing this interesting topic and his helpful comments. Special thanks to Dr. Max Morris for numerous encouraging conversations that meant a lot to me, and Dr. Mervy Marasinghe for his encouragement. Thanks to Dr. Richard Evans for serving my committee, and Ted Peterson for his help on the computational aspects on the cluster. Thanks are extended to all the faculty and staff of the Department of Statistics and to all my friends.

My sincere thank goes to Dr. Luis Escobar at Louisiana State University for his generous help, continuous encouragement and inspiration. I also would like to thank Dr. Lynn LaMotte and Dr. Julia Volaufova for their guidance at the early stages of my statistical studies.

Thanks to Dr. Xiaoming Sun for his valuable help at the beginning of the project, which had a great impact on the research direction.

My deepest gratitude is reserved for my family. Thanks to my parents for everything they have given me. Thanks to my sister for always believing in me without hesitation and always being there to support me in difficult times through all these years. Thanks to my husband for his tremendous help, support, understanding, patience, and inspiration. Thanks to my husband's family for their support. Without them, fulfilling this journey would be impossible.

CHAPTER 1. General introduction

1.1 Overview

This thesis investigates several approaches to develop new sets of estimating equations for improving the efficiency of estimators of regression parameters of Cox proportional hazards models applied to clustered survival data. It is organized into five chapters. Chapter 1 presents a general introduction. First the motivation and objectives of the project are reviewed. Then the basic concepts and notation used in this manuscript are introduced. Finally, a literature review is provided.

In Chapter 2, the weighted estimating equations proposed by Cai and Prentice (1995) are reviewed. Their method of estimating weights is very computationally intensive, and it is not clear how to generalize their method to cases involving continuous covariates. A modified algorithm to calculate the weights is introduced by using the inverse of the estimated martingale residual correlation matrix. In addition, the score equation ignoring the within cluster dependence can be written in another form, which allows a new set of weighted estimating equations to be obtained by inserting weight matrices in the score equation in a different manner. The estimators from both sets of weighted estimating equations are examined. It is found that the two types of weighted estimating equations produce an equivalent estimator when there is a single binary covariate and the exchangeable correlation structure is used, although they are not always equivalent in other cases. The efficiency gain is evaluated by the ratio of the variances of the independent working model estimator and the weighted estimating equation estimator.

Simulation results showed that compared to the independent working model estimator there is a substantial efficiency gain when within cluster correlation is strong and the censoring rate is low. The calculation of the weight matrix is much easier than for the Cai and Prentice approach. Simulation results show that a bootstrap method yields reasonable variance estimates for the estimated parameters.

In Chapter 3, another set of estimating equations is derived from a generalized estimating equation (GEE) approach. The GEE methodology proposed by Liang and Zeger (1986) is modified to apply to a counting process methodology. The weight matrix is estimated as in Chapter 2. In this chapter, simulation studies are used to investigate the effects of strength of within cluster correlation and censoring rates on properties of the parameter estimators. All simulation studies use 100 clusters of size 2. Simulation results with one dichotomous covariate demonstrate a gain in efficiency when within correlation is strong and the censoring rate is moderate. Compared to the results using the estimating equations developed in Chapter 2, there is a slightly larger gain in efficiency.

Results from additional simulation studies are presented in Chapter 4 for more general situations including larger cluster sizes and continuous covariates. The effect of the number of clusters and cluster size are discussed. The number of clusters and cluster size have effect on gains in efficiency, but there is no obvious pattern. The weight matrix proposed in Chapter 2 is general enough to estimate weights when the covariates are continuous. Simulation results for cases involving a single continuous covariate and cases involving one continuous and one dichotomous covariate show that the efficiency gains follow a similar pattern as in the case involving a single dichotomous covariate with respect to strength of correlation and censoring rate. There are potentially greater efficiency gains for these two cases compared to the case involving only a single dichotomous covariate. For cluster sizes larger than two, both exchangeable and first-order autoregressive correlation structures are considered. It is found that gains in efficiency can be achieved with an over simplified correlation structure, but greater gain in efficiency are achieved when the weight matrices are based on the correct correlation structure.

All the computations in Chapter 2, 3 and 4 were implemented using the R package "survgee" developed by the author which provides a familiar interface for the end user that is consistent with other model fitting functions in the S language. The main fitting routine in the package written in C. A detailed help file from the package is provided in the appendix.

Chapter 5 gives a general summary of main findings of this research. It summarizes the conclusions in Chapters 2, 3 and 4. Future work is also discussed.

1.2 Motivation and objectives

Clustered survival data are widely encountered in clinical trial studies. An important assumption of clustered data is that the observations from different clusters are independent, while observations within a cluster may be correlated. For Cox proportional hazards models applied to clustered survival data, a set of estimating equations is provided by the standard partial likelihood equations that incorrectly ignore within cluster correlations. Huster et al. (1989) prove that the estimators from estimating equations derived from an incorrect independence assumption are consistent under regularity conditions similar to the usual regularity conditions for maximum likelihood estimation that also assume correct specification of marginal survival models. Royall (1986) shows how to obtain a robust "sandwich" estimator of the limiting covariance matrix. These are applications of a more general theory developed by Huber (1967). In the context of the Cox proportional hazards model, this estimation procedure is available in the *coxph* function in the *survival* package in R and a library of the same name in S-Plus. This function provides a practical tool for human health researchers to handle correlated survival data. Although it generally provides consistent estimators of regression parameters in the Cox model and reliable standard errors for those estimators, more efficient estimators can be obtained by incorporating better weighting into the estimating equations.

Cai and Prentice (1995) developed weighted estimating equations by introducing a weight matrix into partial likelihood score equations for the Cox (1972) proportional hazards model derived from an incorrect working assumption of within cluster independence. The weight matrix is the inverse of a martingale correlation matrix. Cai and Prentice show that under some regularity conditions the regression estimators are consistent with an asymptotic normal distribution and the efficiency of the estimator is improved by introducing the weight matrix. Simulation studies show that efficiency can be improved substantially if the dependency is strong and the censoring rate is moderate. However, Therneau and Grambsch (2000) point out the method is very computer intensive. It requires the estimation of a large number of parameters relative to the number of clusters. This limits the practical use of this method. Consequently, applications of the method have been rare.

The objective of this project is to develop weighted estimating equations that give comparable improvement in efficiency while reducing complexity and computational burden. The first approach we consider is based on an alternative placement of the weight matrix in the estimating equations. An empirical estimator of martingale correlation structure is developed to provide efficient weighting without overwhelming computational burden. A second approach derives a different set of weighted estimating equations from the partial likelihood score equations. The relative performance of the estimators provided by the two sets of weighted estimating equations, using the same weight matrix, are investigated through simulation studies.

Liang and Zeger (1986) proposed generalized estimating equations (GEEs) to handle within cluster correlations in longitudinal data. Another idea considered in this manuscript is using approximate GEEs to establish the estimating equations. The responses are changes in counting processes within small time intervals. They are treated as Poisson random variables that may be correlated within clusters. The performance of the estimators using the approximate GEE approach is also examined through simulation.

1.3 Notation and basic concepts

Survival data is a term used to describe data that measure the time to some event. Let T^* denote a random variable representing failure time. Denote the distribution function by $F(t) = Pr(T^* \leq t)$. Then the corresponding survival function is $S(t) = 1 - F(t) = Pr(T^* > t)$. The hazard function or failure rate, denoted by $\lambda(t)$, is defined as

$$\lambda(t) = \lim_{\Delta t \to 0+} \frac{Pr(t \leq T^* < t + \Delta t | t \leq T^*)}{\Delta t} = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \log S(t),$$

and $\lambda(t)\Delta t$ is the conditional probability of failure in a small time interval $[t, t + \Delta t)$ given survival until time t. The cumulative hazard function is defined as

$$\Lambda(t) = \int_0^t \lambda(w) dw = -logS(t),$$

which leads to the expression

$$S(t) = \exp\{-\int_0^t \lambda(w) dw\}$$

It is not always the case that a failure time can be observed. In a medical study, for example, some patients may be lost to follow up before the event of interest occurs, and other patients may live beyond the end of a study. This is called censoring, and it produces observations for which it is only known that the time to the event of interest exceeded the observed censoring time.

There are three main types of censoring, Type I, Type II, and random censoring. Let T_1^*, ..., T_n^* be the actual failure times. Then, the three types of censoring are defined as follows:

1. Type I Censoring

   Let t_c be some fixed censoring time. Instead of observing T_1^*, ..., T_n^*, we observe $T_1, \cdots, T_n$, where

$$T_i = \begin{cases} T_i^* & \text{if } T_i^* \le t_c \\ t_c & \text{if } T_i^* > t_c \end{cases}$$

2. Type II Censoring

   Let r be fixed, where $r < n$, and let $T_{(1)}^*$,...,$T_{(n)}^*$ be the ordered actual failure times. The study stops after r failures are observed, and the full ordered observed sample is

$$\begin{aligned} &Y_{(1)} = T_{(1)}^*, \\ &\vdots \\ &Y_{(r)} = T_{(r)}^*, \\ &Y_{(r+1)} = T_{(r)}^*, \\ &\vdots \\ &Y_{(n)} = T_{(r)}^*. \end{aligned}$$

3. Random Censoring

   Let C_i^* be the censoring time associated with the *ith* case and true failure time T_i^*. The pair (T_i, δ_i) is what we actually observe, where

$$T_i = min(T_i^*, C_i^*),$$

   and

$$\delta_i = I(T_i^* \le C_i^*) = \begin{cases} 1 & \text{if } T_i^* \le C_i^*, \text{ and } T_i^* \text{ is not censored,} \\ 0 & \text{if } T_i^* > C_i^*, \text{ and } T_i^* \text{ is censored.} \end{cases}$$

There are other commonly used ways of classifying censoring. One is right censoring and left censoring. The three types of censoring reviewed above are all examples of right censoring. In random left censoring, we can only observe the pair (T_i, ϵ_i), where

$$T_i = max(T_i^*, C_i^*),$$
$$\epsilon_i = I(C_i^* \leq T_i^*) = 1 - \delta_i.$$

Both right and left censoring are special cases of interval censoring, in which we only know that a failure occurred in a time interval instead of observing an exact failure time. In this thesis, we only consider Type I and random censoring.

Kaplan-Meier and Nelson-Aalen estimators are two common nonparametric methods used to estimate survival functions. The Kaplan-Meier estimator is of the form

$$\hat{S}_{KM}(t) = \prod_{i=1}^{K} \left(1 - \frac{d_i}{r_i}\right),$$

where d_i is the number of deaths at time t_i, and r_i is the number of subjects at risk at time t_i. The large sample variance of the Kaplan-Meier estimator of $S(t)$ can be estimated with the Greenwoods formula

$$Var(\hat{S}_{KM}(t)) = (\hat{S}_{KM}(t))^2 \sum_{i=1}^{K} \frac{d_i}{r_i(r_i - d_i)}.$$

The Nelson-Aalen estimator of the cumulative hazard function is

$$\hat{\Lambda}_{NA}(t) = \sum_{i=1}^{K} \frac{d_i}{r_i},$$

and the corresponding estimator of the survival function is $\hat{S}_{NA}(t) = exp(-\hat{\Lambda}(t))$. The large sample variance of $\hat{S}_{NA}(t)$ can be estimated by

$$Var(\hat{S}_{NA}(t)) = (\hat{S}_{NA}(t))^2 \sum_{i=1}^{K} \frac{d_i}{r_i^2}.$$

When risk sets are large relative to the number of events, the two estimators are essentially the same.

1.4 Partial likelihood estimation for the Cox model

1.4.1 Univariate survival data

1.4.1.1 Cox model

Cox (1972) proposed a proportional hazards model for incorporating effects of covariates on hazard functions, i.e.,

$$\lambda(t; Z) = \lambda_0(t)e^{\boldsymbol{\beta}'\mathbf{Z}}, \tag{1.1}$$

where $\boldsymbol{\beta}$ is a vector of unknown parameters, $\mathbf{Z}$ is a vector of covariate values, and $\lambda_0(t)$ is a common baseline hazard function. This model yields the following results:

$$\begin{aligned} S(t) &= \exp\{-\Lambda_0(t)e^{\boldsymbol{\beta}'\mathbf{Z}}\}, \\ F(t) &= 1 - \exp\{-\Lambda_0(t)e^{\boldsymbol{\beta}'\mathbf{Z}}\}, \\ f(t) &= \lambda_0(t)\exp\{\boldsymbol{\beta}'\mathbf{Z} - \Lambda_0(t)e^{\boldsymbol{\beta}'\mathbf{Z}}\}, \end{aligned} \tag{1.2}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ is the cumulative baseline hazard function.

Cox (1972) proposed a conditional likelihood function that provides an estimate of β without specification of λ_0. Let $t_{(1)} < t_{(2)} < ... < t_{(k)}$ be distinct ordered failure times. Define the risk set at the ith failure time,

$$\mathcal{R}_i = \{j : t_j \geq t_{(i)}\},$$

as the set of all individuals at risk immediately before time $t_{(i)}$. Conditionally on the risk set $\mathcal{R}(t_{(i)})$, the likelihood that an individual with covariate values $\mathbf{Z}_i$ fails at time $t_{(i)}$ is

$$\frac{\exp\{\boldsymbol{\beta}'\mathbf{Z}_i\}}{\sum_{l\in\mathcal{R}(t_{(i)})}\exp\{\boldsymbol{\beta}'\mathbf{Z}_l\}}. \tag{1.3}$$

Thus, the conditional likelihood for k independent failure times is

$$L(\beta) = \prod_{i=1}^{k} \frac{\exp\{\boldsymbol{\beta}'\mathbf{Z}_i\}}{\sum_{l\in\mathcal{R}(t_{(i)})}\exp\{\boldsymbol{\beta}'\mathbf{Z}_l\}}. \tag{1.4}$$

Cox (1975) showed that the conditional likelihood in (1.4) coincides with a partial likelihood. Let Y be a random variable with density $f_Y(y, \beta)$, and suppose that Y can be transformed into the sequence $(X_1, S_1, ..., X_m, S_m)$. The full likelihood of the sequence is

$$\prod_{j=1}^{m} f_{X_j|X^{(j-1)},S^{(j-1)}}(x_j|x^{(j-1)}, s^{(j-1)}; \beta) \prod_{j=1}^{m} f_{S_j|X^{(j)},S^{(j-1)}}(s_j|x^{(j)}, s^{(j-1)}; \beta), \tag{1.5}$$

where $x^{(j)} = (x_1, ..., x_j)$ and $s^{(j)} = (s_1, ..., s_j)$. The second product is called the partial likelihood based on S in the sequence $\{X_j, S_j\}$. Cox suggested that the partial likelihood contains most of the information about β and the first product can be ignored without much loss of efficiency. Cox also pointed out that in regression with censored data the partial likelihood coincides with conditional likelihood, and (1.4) is in fact a partial likelihood. Thus it is reasonable to estimate β by finding the value of β that maximizes the natural logarithm of (1.4).

1.4.1.2 Tied data

The partial likelihood defined by (1.4) is for continuous data. Observed survival times are generally discrete due to limits on how accurately time can be measured or because subjects are not continuously monitored. When tied failure times are observed, Cox (1972) suggests using

$$\begin{aligned} L_c &= \prod_{i=1}^{r} P\{\mathcal{D}_{(i)}|\mathcal{R}_{(i)}, d_i\} \\ &= \prod_{i=1}^{r} \frac{\exp(\sum_{j\in\mathcal{D}_{(i)}} \boldsymbol{\beta}'\mathbf{Z}_j)}{\sum_{\mathcal{D}^*_{(i)}} \exp(\sum_{j\in\mathcal{D}^*_{(i)}} \boldsymbol{\beta}'\mathbf{Z}_j)}, \end{aligned} \tag{1.6}$$

where $\mathcal{R}_{(i)}$ and $\mathcal{D}_{(i)}$ are the risk set and death set at ith distinct ordered failure time, respectively, and d_i is the number of failures observed at time $t_{(i)}$. The denominator of equation (1.6) is the risk score summing over all possible subsets $\mathcal{D}^*_{(i)}$ of $\mathcal{R}_{(i)}$, such that $\mathcal{D}^*_{(i)}$ contains exactly d_i subjects. For failure time i, there are $\binom{n_i}{d_i}$ subsets to consider. This approach is computationally intensive even for small numbers of ties.

Other methods of dealing with ties have been proposed by Breslow (1974), Efron (1977), Kalbfleisch and Prentice (1973), and Peto (1972). The Breslow approximation to (1.6) is

$$L_{c,B} = \prod_{i=1}^{r} \frac{\exp\left(\sum_{j\in\mathcal{D}_{(i)}} \boldsymbol{\beta}'\mathbf{Z}_j\right)}{\left(\sum_{j\in\mathcal{R}_{(i)}} e^{\boldsymbol{\beta}'\mathbf{Z}_j}\right)^{d_i}}. \tag{1.7}$$

It is easy to compute and works well when the number of ties is not too large and the β's are sufficiently small. A better approximation to (1.6) is the Efron approximation given by

$$L_{c,E} = \prod_{i=1}^{r} \left(\frac{\exp\left(\sum_{j\in D_{(i)}} \boldsymbol{\beta}'\mathbf{Z}_j\right)}{\prod_{k=1}^{d_i} \left[\sum_{j\in R_{(i)}} \boldsymbol{\beta}'\mathbf{Z}_j - \dfrac{k-1}{d_i} \sum_{l\in D_{(i)}} e^{\boldsymbol{\beta}'\mathbf{Z}_l}\right]} \right) \tag{1.8}$$

The estimator of $\boldsymbol{\beta}$ that maximizes (1.8) generally has less bias than the estimator of $\boldsymbol{\beta}$ obtained from (1.7).

Prentice and Gloeckler (1978) propose a full likelihood of the form

$$\begin{aligned} L &= \prod_{i=1}^{n} P(Y_i = j, \delta_i) \\ &= \prod_{i=1}^{n} \left(\left(\prod_{k=1}^{j-1} \alpha_k^{\exp(\boldsymbol{\beta}'\mathbf{Z}_i)} \right) \left(1 - \alpha_j^{\exp(\boldsymbol{\beta}'\mathbf{Z}_i)}\right)^{\delta_i} \right), \end{aligned}$$

where

$$\alpha_i = \exp\left(- \int_{\alpha_{j-i}}^{\alpha_j} \lambda_0(t)dt \right),$$

is the conditional probability of surviving during the interval of $[\alpha_{j-1}, \alpha_j]$, given surviving during the period $[\alpha_{j-2}, \alpha_{j-1}]$, under the condition of $x = 0$. This likelihood strictly adheres to the proportional hazards model. By using the transformation $\gamma_j = log(-log\alpha_j)$ for the evaluation of maximum likelihood estimates of parameters, computation becomes simpler and convergence is faster.

1.4.1.3 Large-sample theory

Cox (1975) asserts that large-sample properties of maximum likelihood estimators can be applied to estimators derived from a partial likelihood. For simplicity, assume β to be one dimensional. Let

$$U = \frac{\partial log \mathcal{L}(\beta)}{\partial \beta} = \sum_{i=1}^{n} U_i(\beta)$$

Under the usual regularity conditions that the second derivative of the likelihood is a smooth function, and the third derivative exists, and bounded in absolute value by an integrable function,

$$\begin{aligned} E(U) &= 0 \\ var(U) &= I^{-1}(\beta) = -\left[E\left(\frac{\partial log^2 \mathcal{L}(\beta)}{\partial \beta^2}\right)\right]^{-1} \end{aligned}$$

Tsiatis (1981) developed a limited asymptotic theory of the partial likelihood estimator of β in the Cox model. He assumes that covariates are random with a given distribution that satisfies given smoothness conditions, and random censorship. The regularity conditions require $E[Zexp(\beta Z)]^2$ to be bounded uniformly in a neighborhood of β, and a positive probability that any subject survives beyond the time of termination of the study. Under these conditions, Tsiatis shows that the solution to the partial likelihood is almost surely inside a neighborhood of β, and this neighborhood becomes smaller as the sample size increases. A Taylor series expansion of the score function is used to show that $\sqrt{n}(\hat{\beta} - \beta)$ converges to a normal distribution, and to derive the formula for the limiting covariance matrix.

Bailey (1983) proves the same results under a more general set of conditions. He assumes that covariates and censoring times are fixed sequences. The regularity conditions in his proof are that the covariates are bounded by some constant, and the expected information matrix increases proportionally with n. Bailey considered two events E_n^1 and E_n^2. Define a distance $r_n = n^{-1/2+\epsilon}$, where $0 < \epsilon < 1/2$, $E_n^1 = \{|\hat{\beta} - \beta| > r_n\}$, and

$E_n^2 = \{log[L(\beta + r_n u_n)] > log[L(\beta)]\}$, where $u_n = \dfrac{\hat{\beta} - \beta}{|\hat{\beta} - \beta|}$, $L(\beta)$ is the partial likelihood function, and $\hat{\beta}$ is the solution to the partial likelihood score equation. Since $logL$ is strictly concave, $E_n^1 \subset E_n^2$, and therefore $P(E_n^1) < P(E_n^2)$. By proving the probability of E_n^2 goes to zero as n increases, the consistency of $\hat{\beta}$ is established. The derivation of the variance of the limiting normal distribution is obtained from a Taylor series expansion of the score function about $\hat{\beta}$.

1.4.2 Counting processes

Under suitable regularity conditions, the asymptotic Gaussian distribution of the partial likelihood estimator for the regression coefficients can be established using counting processes and martingale theory. Counting process analysis is based on a history of the process, often called the filtration, denoted $\{\mathcal{F}_t; t > 0\}$. A natural choice is the history of the process up to t.

1.4.2.1 Basic concepts

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\{X_n\}_{n\geq 1}$ be a sequence of random variables with $E|X_n| < \infty$, for all $n \geq 1$. Let $\mathcal{F}_1 \subset \mathcal{F}_2 \subset ...$ be sub-σ-fields of $\mathcal{F}$.

1. $\{\mathcal{F}_n\}_{n\geq 1}$ is called a filtration.

2. $\{X_n\}_{n\geq 1}$ is adapted to $\{\mathcal{F}_n\}_{n\geq 1}$ if X_n is $< \mathcal{F}_n, \mathcal{R} >$ -measurable, for all $n \geq 1$.

3. $\{X_n, \mathcal{F}_n\}_{n\geq 1}$ is called a martingale/submartingale/supermartingale, if $\{X_n\}_{n\geq 1}$ is adapted to $\{\mathcal{F}_n\}_{n\geq 1}$, and for each $n \in \mathcal{N}$,

$$\begin{aligned} E(X_{n+1}|\mathcal{F}_n) &= X_n \text{ for a martingale} \\ &\geq X_n \text{ for a submartingale} \\ &\leq X_n \text{ for a supermartingale.} \end{aligned}$$

A (real-value) stochastic process is a family of random variables $X = \{X(t) : t \in \Gamma\}$ indexed by Γ, such that all are defined on the same probability space $(\Omega, \mathcal{F}, \mathcal{P})$. A counting process is a stochastic process $\{N(t) : t \geq 0\}$ adapted to a filtration $\{\mathcal{F}_t : t \geq 0\}$ with $N(0) = 0$ and $N(t) < \infty$ a.s., and for which the possible paths are with probability one right-continuous, piecewise constant, and have only jump discontinuities, with positive jumps of size 1.

Doob-Meyer Decomposition: Let X be a right-continuous nonnegative sub-martingale with respect to a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P)$. Then there exists a right-continuous martingale M and an increasing right-continuous predictable process A such that $E(A(t)) < \infty$ and

$$X(t) = M(t) + A(t) \text{ a.s.}$$

for any $t \geq 0$. If $A(0) = 0$ a.s., and if $X = M' + A'$ is another such decomposition with $A'(0) = 0$, then for any $t \geq 0$,

$$P\{M'(t) \neq M(t)\} = P\{A'(t) \neq A(t)\} = 0.$$

A process X is predictable with respect to $\{\mathcal{F}_t : t \geq 0\}$ if can only if it is measurable with respect to the smallest σ-algebra on $R^+ \times \Omega$ generated by the adapted left-continuous processes.

In the counting process approach to survival analysis, the time and censoring indicator pair (T_i, δ_i) is replaced by the pair $(N_i(t), Y_i(t))$, where

$$N_i(t) = \text{ the number of observed events in [0, t] for unit i}$$

and

$$Y_i(t) = \begin{cases} 1 & \text{if unit i is under observation and at risk at time } t \\ 0 & \text{otherwise} \end{cases}$$

According to the Doob-Meyer decomposition theorem, any counting process may be uniquely decomposed as the sum of a martingale and a predictable, right continuous process that is 0 at time 0, called the compensator. The counting process $N_i(t)$ has compensator $A_i(t) = \int_0^t Y_i(s)\lambda_i(s)ds$, where $\lambda_i(t)$ is the hazard function. Then

$$N_i(t) = A_i(t) + M_i(t).$$

where

$$M_i(t) = N_i(t) - A_i(t) = N_i(t) - \int_0^t Y_i(s)\lambda_i(s)ds$$

is a counting process martingale with respect to the history given above.

Let $dN_i(t)$, a counting process differential, denote the increment in N_i over the infinitesimal time interval $[t, t+dt]$. Then

$$dN_i(t) = \begin{cases} 1 & \text{if a failure occurs in } [t, t+dt] \\ 0 & \text{otherwise} \end{cases}$$

$\mathcal{F}_{t-}$ contains all the information on [0,t), and

$$E(dN_i(t)|\mathcal{F}_{t-}) = dA_i(t) = Y_i(t)\lambda_i(t)dt.$$

Martingale increments have mean 0, i.e.,

$$E(dM_i(t)|\mathcal{F}_{t-}) = 0 \text{ for any } t > 0.$$

1.4.2.2 Asymptotic theory

Andersen and Gill (1982) use counting processes and martingale theory to prove the consistency and asymptotic normality of partial likelihood regression estimators. One basic assumption is that the counting process $N(t)$ has intensity of the form $\lambda(t) = \lambda_0(t)\exp\beta Z$. Taylor series expansion is used to expand the first derivative of the partial likelihood $U(\beta, t)$ around the true parameter, β_0. The asymptotic normality of $n^{-1/2}(\hat{\beta} -$

β_0) is obtained by proving weak convergence to a Gaussian process of the local martingale $n^{-1/2}U(\beta_0)$. The central limit theorems for local martingales given by Rebolledo (1980) are used.

1.4.3 Extensions to correlated survival data

A key assumption in the asymptotic theory derived by Andersen and Gill (1982), Tsiatis (1981) and Bailey (1983) is that all survival times are mutually independent. Multivariate survival data analysis is used when the assumption of independent survival times is violated. Some situations in which this can happen are:

1. Times to different events are monitored on the same subject;
2. Repeated measures are taken on the same subject over time;
3. Subjects within a group give correlated responses arising from genetic relationships or common environmental or social effects.

Various methods have been proposed for analyzing multivariate survival data. A key assumption for situations 1 and 2 is that subjects respond independently, and a corresponding assumption for situation 3 is that groups are independent. One approach is to initially ignore the correlations among the survival times and fit models assuming the responses are independent. Then robust covariance estimation is used to obtain appropriate standard errors for estimated parameters. This is often called the independence working model approach. More efficient estimates of model parameters can be obtained by incorporating information on correlations into the estimating equations. This is called the generalized estimating equations approach. Neither of these approaches requires a complete specification of the joint distribution of the correlated survival times. Fully parametric methods have also been proposed. These include frailty models which incorporate random group effects to induce correlations among response times as proposed

by Vaupel et al. (1979), Clayton and Cuzick (1985), Hougaard (1995), and estimation of joint survival functions by Koehler and Symanowski (1995) and Prentice and Cai (1992)
.

1.4.3.1 Independence working model (IWM) approach

When the main interest of a study is to estimate the marginal effects of explanatory variables or survival rates, within group correlations among subjects become nuisance parameters. Huster et al. (1989) propose an independence working model (IWM) for paired survival data. They specify parametric models for the marginal distributions and use the incorrect assumption that the members in each pair respond independently to obtain a set of estimating equations. The resulting IWM likelihood is the product of the marginal likelihoods over all pairs. The parameters of the marginal models can be estimated by maximizing this likelihood. If the marginal distributions of the IWM likelihood are the same as those of the true bivariate model, then the IWM estimator is shown to be consistent by showing that the expectation of the IWM score function is zero.

However, the inverse of the information matrix for the IWM likelihood does not provide a consistent estimator for the covariance matrix for the limiting normal distribution of the IWM parameter estimates. Huber (1967) studied consistency and asymptotic normality of estimators obtained from maximizing incorrect likelihoods under more general conditions. No second or higher order derivatives of the likelihood function are required for the regularity conditions. Royall (1986) shows how to obtain a consistent estimator of the covariance matrix, known as a robust or “sandwich estimator”, using the results developed by Huber.

Let $\boldsymbol{\beta}$ denote parameters of the marginal model, and let n denote the number of observations in the study. Then the IWM likelihood for estimating $\boldsymbol{\beta}$ is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} f_{\boldsymbol{\beta}}(\mathbf{Z}_i).$$

The IWM score vector is

$$U = \frac{\partial log L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} U_i(\boldsymbol{\beta}),$$

where $U_i(\boldsymbol{\beta}) = \frac{\partial log f_{\boldsymbol{\beta}}(\mathbf{z}_i)}{\partial \boldsymbol{\beta}}$ and the information matrix is

$$I(\boldsymbol{\beta}) = E_{IWM}\left[-\frac{\partial^2 log L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta'}}\right],$$

where the expectation is with respect to the incorrect IWM distribution.

Huber and Royall show that under some regularity conditions, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges to a Gaussian distribution with mean 0 and variance $V(\boldsymbol{\beta})$, as $n \to \infty$, where

$$V(\boldsymbol{\beta}) = lim_{n \to \infty} I^{-1}(\boldsymbol{\beta}) E[U(\boldsymbol{\beta})U'(\boldsymbol{\beta})] I^{-1}(\boldsymbol{\beta}).$$

and $E[U(\hat{\boldsymbol{\beta}})U'(\hat{\boldsymbol{\beta}})]$ denotes expectation with respect to the true joint survival distribution. It can be estimated by

$$\hat{V}(\hat{\boldsymbol{\beta}}_{IWM}) = I^{-1}(\hat{\boldsymbol{\beta}}_{IWM}) \sum_{i=1}^{n} [U_i(\hat{\boldsymbol{\beta}}_{IWM}) U_i'(\hat{\boldsymbol{\beta}}_{IWM})] I^{-1}(\hat{\boldsymbol{\beta}}_{IWM}),$$

where $\hat{\boldsymbol{\beta}}_{IWM}$ is the IWM estimator. If the members of the pairs are truly independent, then $E\left(\sum [U_i(\hat{\boldsymbol{\beta}}) U_i'(\hat{\boldsymbol{\beta}})]\right) = I^{-1}(\hat{\boldsymbol{\beta}})$, and $\hat{V}(\hat{\boldsymbol{\beta}})$ approaches $I^{-1}(\hat{\boldsymbol{\beta}})$, which agrees with asymptotic theory for the case with completely independent observations.

Wei et al. (1989) apply the same approach to multivariate survival data. They model the marginal distribution of failure time variables using a Cox proportional hazards model without specification of the dependence structure of failure times measured on the same individual. Lee et al. (1992) show that IWM estimators of regression parameters are consistent under the regularity conditions that the second derivative of likelihood function is a smooth function, and the third derivative exists.

The IWM estimation procedure is conceptually straightforward. It avoids explicitly modeling the structure of dependency, but some efficiency may be lost. Huster et al. (1989) use the asymptotic relative efficiency (ARE) of the IWM estimators to measure the amount of precision lost by the use of the IWM instead of the true bivariate parametric model. They study paired survival data with the model in Clayton (1978) and Oakes (1982). The ARE for a given parameter is defined as the ratio of the appropriate diagonal element of the inverse of the expected information matrix for the true bivariate model to the corresponding term in $V(\beta)$ give by (1.9) for the IWM approach. Both discordant pairs and concordant pairs are examined. Members of a discordant pair have different covariate vectors, while members of concordant pair have identical covariate vectors. Results show that when correlation between the survival times of the members of each pair is 0.64, and discordance is 100%, the ARE of the IWM estimator is 90%, while for 100% concordance, the IWM loses 30% efficiency. The IWM estimator can lose 50% efficiency for intermediate discordance. Under certain circumstances, the IWM estimator may loose a substantial amount of efficiency.

The robust estimate of variance for the IWM estimator of coefficients in a Cox model can also be obtained by a jackknife procedure. Lipsitz et al. (1990) found that an approximate jackknife estimate of variance, $D'D$, is preferred, where $D = UI^{-1}$, U is the logarithm of the partial likelihood, and I^{-1} is the inverse of the observed information matrix. The ith row of D is the approximate change in coefficient estimates if observation i is removed. Therneau and Grambsch (2000) examined a jackknifed estimate of the covariance matrix for coefficient estimates for correlated data. For data with within group correlations and independent groups, they use a grouped jackknife estimate that leaves out an entire group of correlated observations rather than leaving out a single observation at a time. The robust variance can be estimated using an approximate grouped jackknife estimator, $\tilde{D}\tilde{D}$. Each row of $\tilde{D}$ is the summation of rows of D contributed by correlated observations. This approximate grouped jackknife approach provides a computationally efficient procedures for evaluating a robust sandwich estimator of the large sample covariance matrix for coefficient estimates because of the availability of D in standard softwares.

Loughin and Koehler (1997) use an IWM approach to obtain consistent estimates of the parameters when the marginal models have the form of the Cox proportional hazards model. They apply bootstrap resampling procedures to estimate covariance matrices. Loughin and Koehler compare the bootstrap and robust estimation of the true variance matrix $V(\hat{\boldsymbol{\beta}})$. The simulation results show that when there is only one parameter, both methods provide good approximations. However, when several parameters are estimated, the bootstrap may provide more reliable estimates of standard errors than robust covariance estimation in smaller samples. All of these approaches use potentially inefficient IWM estimators.

1.4.3.2 Linear regression weighted estimating equation

Consider a linear model for some function of the failure time,

$$g(T^*_{kj}) = \mathbf{Z}'_{kj}\boldsymbol{\beta} + \epsilon_{kj},$$

where $g(.)$ is a specified transformation, often the natural logarithm. Tsiatis (1990) proposed the following estimating equations based on the results of Louis (1981)

$$\sum_{k=1}^{K}\sum_{j=1}^{n_k} \delta_{kj} W(\varepsilon^*_{kj})\{\mathbf{Z}_{kj} - \bar{Z}(e^*_{kj}, \boldsymbol{\beta})\}, \tag{1.9}$$

where $e^*_{kj} = g(T_{kj}) - \mathbf{Z}'{}_{kj}\boldsymbol{\beta}$, $\bar{Z}(u, \beta) = \sum_{lq} \mathbf{Z}_{lq} I(e^*_{lq} \geq u) / \sum_{lq} I(e^*_{lq} \geq u)$ and $W(\varepsilon^*_{kj})$ is a matrix of weights. Tsiatis (1990) gives asymptotically optimal weights for equation (1.9), but the results cannot be adapted to censored data due to the fact that the score function is not continuous, and it is generally not monotone when censoring exists. Gray (2000) proposed a one-step Newton-like update approach to get a more efficient estimator. Gray (2003) extends the method to clustered data with censoring.

In the context of counting processes, define

$$
\begin{aligned}
N_{kj}(u,\boldsymbol{\beta}) &= I(e^*_{kj} \le u, \delta_{kj}=1),\\
Y_{kj}(u,\boldsymbol{\beta}) &= I(e^*_{kj} \ge u),\\
\hat{M}_{kj}(u,\boldsymbol{\beta}) &= N_{kj}(u,\boldsymbol{\beta}) - \int_{-\infty}^{u} Y_{kj}(w)d\hat{\lambda}_0(w),\\
a_{kj}(u,\boldsymbol{\beta}) &= \sum_i Z_{kj} w_{kji}(u,\boldsymbol{\beta}),\\
\bar{a}(u,\boldsymbol{\beta}) &= \sum_{lq} Y_{lq}(u,\beta) a_{lq}(u,\boldsymbol{\beta}) / \sum_{lq} Y_{lq}(u,\boldsymbol{\beta}),
\end{aligned}
$$

and

$$
A(u,\boldsymbol{\beta}) = \sum_{k=1}^{K}\sum_{j=1}^{n_k}\{a_{kj}(u,\boldsymbol{\beta}) - \bar{a}(u,\boldsymbol{\beta})\} Z'_{kj} Y_{kj}(u,\boldsymbol{\beta}).
$$

Then an one-step estimator is

$$
\boldsymbol{\beta}^1 = \tilde{\boldsymbol{\beta}}^g - \hat{G}^{-1} S_N(\tilde{\boldsymbol{\beta}}^g),
$$

where

$$
S_N(\boldsymbol{\beta}) = \sum_k \sum_{j=1}^{n_i}\sum_{i=1}^{n_k} \int w_{kji} Z_{kj} d\hat{M}_{kj}(u,\boldsymbol{\beta}),
$$

and

$$
\hat{G} = \sum_i \hat{\lambda}(e^*_{(i)})\{A(e^*_{(i)}, \tilde{\boldsymbol{\beta}}^g) A(e^*_{(i+1)}, \tilde{\boldsymbol{\beta}}^g)\}.
$$

Gray (2003) shows that under certain regularity conditions, $\hat{\boldsymbol{\beta}}^1$ is consistent and asymptotically normal distributed with variance $KG(\boldsymbol{\beta}_0)^{-1}Var(S_N(\boldsymbol{\beta}_0))\{G(\boldsymbol{\beta}_0)^{-1}\}$ as K goes to infinity, where $G(\beta) = \int A(u,\beta)\lambda_0(u)du$. The variance of $\hat{\boldsymbol{\beta}}^1$ can be estimated by $\hat{G}^{-1}\left(\sum_{k=1}^{K} \hat{S}_{n_k}\hat{S}'_{n_k}\right)(\hat{G}^{-1})'$, where $\hat{S}_{n_k} = \sum_{j=1}^{n_k}\int\{a_{kj}(u,\hat{\beta}^1) - \bar{a}(u,\hat{\beta}^1) d\hat{M}_{kj}(u,\hat{\beta}^1)\}$. It is shown that by incorporating weight matrices, the efficiency of the estimators can be improved. The simulation results showed that the variance estimator performed poorly in some settings.

1.4.3.3 Frailty model

Frailty models have been extensively developed in recent years. Frailty models contain continuous random variables to explain variation in risk, like group variation, that is not explained by the observed covariates. The model is the usual proportional hazards model conditional on frailty. The conditional hazard of the jth individual in the kth group is

$$\lambda_{0j}(t)\exp(\boldsymbol{\beta}'\mathbf{Z}_{kj} + \boldsymbol{\omega}'_{kj}\mathbf{X}_{kj}), \tag{1.10}$$

where $\mathbf{Z}_{kj}$ is the vector of observed covariates, $\boldsymbol{\omega}_{kj}$ is a vector of random effects or frailties.

If the main interest is to estimate regression coefficients, we do not need to specify the frailty distribution. But if the dependence is also of interest, a frailty distribution needs to be given. The most commonly used distribution is the Gamma distribution. In proportional hazards regression models with Gamma frailty, the marginal hazards do not satisfy the proportionality criterion. In this case, Hougaard (2000) shows that the estimates might have larger bias than IWM estimates because of non-proportional hazard, even though frailty takes consideration of dependence. Positive stable distributions maintain the proportional hazard assumption in the marginal models, which is an appealing theoretical property, but computations are more difficult due to the complicated form of the derivatives of Laplace transformation. The positive stable distribution cannot be extended to yield negative dependence, since the density would have to be negative in that case. Hougaard (2000) gives examples of fits that are not satisfactory for some applications. The power variance function (PVF) family is a natural exponential family, for which the variance is a power function of the mean. The PVF model is a more general family of models that includes the Gamma and positive stable distributions. It can provide better fit, but because of two parameters, it is more difficult to implement.

Computationally, a frailty model can use a penalized likelihood approach as an estimation tool. The likelihood is the product of a partial likelihood with frailty parameters and a penalty function. The estimation procedure assumes the parameters of frailty known, and maximizes the penalized likelihood to obtain estimates of coefficients. Then the parameters of frailty are estimated. The approach is computationally faster.

Some frailty models are available to handle more complicated dependence structure. The multiplicative stable model seems to have interesting theoretically properties, but the fit to real data is often not very satisfactory. The additive frailty model can handle more general dependence structure by creating a multivariate frailty variable, but needs more parameters. Multivariate lognormal frailty has more flexibility to model complicated dependence structures. Further development and applications are needed. Hougaard (2000) points out that research on multivariate frailty models is still at early stage.

1.4.3.4 Multivariate survival function

One direct way to analyze multivariate survival data is to estimate a joint survival function. Clayton (1978) considers a bivariate joint distribution of failure times for studies of familial tendency in chronic disease incidence. Estimating equations are simply obtained from partial derivatives of the log likelihood. Koehler and Symanowski (1995) present a multivariate generalization of a copula method for constructing multivariate distributions with a specific set of univariate marginal distributions. The joint cumulative density function and joint density function can be expressed explicitly as functions of univariate cumulative density function and marginal density function. An application of the method to obtain multivariate survivor functions is provided.

Prentice and Cai (1992) use counting processes to characterize the dependence of bivariate failure data. The covariance function for failure time T_1 and T_2 is defined as

$$C(t_1, t_2) = cov\{M_1(t_1), M_2(t_2)\} = E\{M_1(t_1)M_2(t_2)\},$$

for all $t_1, t_2 \geq 0$, where $M_i(t) = N_i(t) - \Lambda_i(t \wedge T_i)$. And the joint survival function is expressed as

$$F(t_1, t_2) = F_1(t_1)F_2(t_2)\left[1 + \int_0^{t_1}\int_0^{t_2}\{F_1(s_1)F_2(s_2)\}^{-1}C(ds_1, ds_2)\right].$$

The correlation function is given as

$$\rho(t_1, t_2) = \frac{C(t_1, t_2)}{\{varM_1(t_1)varM_2(t_2)\}^{1/2}}.$$

They also derive consistent estimates of $C(t_1, t_2)$ and $F(t_1, t_2)$. This model can be extended to the regression case by defining $M_i(t) = N_i(t) - \Lambda_i(t \wedge T_j; Z)$. Extensions to higher dimensions are also given in the paper.

It is generally computationally intensive to optimize a log-likelihood obtained from a copula construction of a joint distribution. And there is no software available to fit the wide variety of models. Users need to develop their own software depending on how the joint distributions are defined. Therefore the application of high dimensional joint distributions has been limited.

CHAPTER 2. Weighted estimating equations

2.1 Introduction

In this chapter, we modify the approach to estimating weight matrices for estimating equations proposed by Cai and Prentice (1997) to reduce the complexity and computational burden. Also, a new set of estimating equations is investigated.

Cai and Prentice introduced weights into partial likelihood score equations to improve the efficiency of the estimators of coefficients in Cox proportional hazards models applied to clustered data with within cluster correlations among survival times. The inverses of correlation matrices of martingales are used as weights. Correlation matrices are estimated by a nonparametric method that requires estimation of a large number of parameters relative to the number of clusters. Emura and Tsukuma (2003) simplify the method by estimating correlations with simple correlation functions of martingales conditional on all possible combinations of covariate values. For example, for a binary covariate taking a value of 0 or 1 with clusters of size 2, three correlations of covariate pairs (0,0), (1,1) and (1,0) or (0,1) need to be calculated. The method requires replication of all the possible covariate value combinations. If there are several binary covariates, the number of observations available for estimating some correlations might be small. Also, if there is a continuous covariate, the covariate values need to be grouped. It might not be straightforward to find a proper way of grouping. Therefore we consider using common weights for all the clusters. We investigate the performance of the estimators using weights conditional on covariate pairs, and the estimators using common weights. Simulation results show that estimators do not lose much efficiency by using common weights even when the data are correlated with different dependence levels. The simulation results presented in Section 2.6.4 use the common weight approach to form weighted estimating equations.

If the within cluster dependence is ignored, the partial likelihood score equations can be written as a different form from what Cai and Prentice use. Weight matrices can be inserted in a different way to form a new set of estimating equations. The behavior of the estimators of this new set of estimating equations is also investigated. Bootstrap resampling methods are used to estimate the variances of coefficients in the Cox model, and construct confidence intervals. Simulation studies are used to assess the performance of the estimators of the partial likelihood score equations ignoring the within cluster dependence, the estimators of Cai and Prentice estimating equations with modified weight matrices and the estimators of the new set of estimating equations with the same weight matrices.

2.2 Introduction to counting process notation

Counting process analysis is based on a history of the process, often called the filtration, denoted $\{\mathcal{F}_t; t > 0\}$. A natural choice is the history of the process up to time t.

A counting process is a stochastic process $\{N(t) : t \geq 0\}$ adapted to a filtration $\{\mathcal{F}_t : t \geq 0\}$ with $N(0) = 0$ and $N(t) < \infty$ a.s., and for which the possible paths are with probability one right-continuous, piecewise constant, and have only jump discontinuities, with positive jumps of size 1.

In the counting process approach to survival analysis, the time and censoring indicator pair (T_i, δ_i) is replaced by the pair $(N_i(t), Y_i(t))$, where

$$N_i(t) = \text{ the number of observed events in [0, t] for unit i}$$

and

$$Y_i(t) \;=\; \begin{cases} 1 & \text{if unit i is under observation and at risk at time } t \\ 0 & \text{otherwise} \end{cases}$$

According to the Doob-Meyer decomposition theorem, any counting process may be uniquely decomposed as the sum of a martingale and a predictable, right continuous process that is 0 at time 0, called the compensator. The counting process $N_i(t)$ has compensator $A_i(t) = \int_0^t Y_i(s)\lambda_i(s)ds$, where $\lambda_i(t)$ is the hazard function. Then

$$N_i(t) = A_i(t) + M_i(t).$$

where

$$M_i(t) = N_i(t) - A_i(t) = N_i(t) - \int_0^t Y_i(s)\lambda_i(s)ds$$

is a counting process martingale with respect to the history given above.

Let $dN_i(t)$, a counting process differential, denote the increment in N_i over the infinitesimal time interval $[t, t + dt]$. Then

$$dN_i(t) \;=\; \begin{cases} 1 & \text{if a failure occurs in } [t, t + dt] \\ 0 & \text{otherwise} \end{cases}$$

$\mathcal{F}_{t-}$ contains all the information on [0,t), and

$$E(dN_i(t)|\mathcal{F}_{t-}) = dA_i(t) = Y_i(t)\lambda_i(t)dt.$$

Martingale increments have mean 0, i.e.,

$$E(dM_i(t)|\mathcal{F}_{t-}) = 0 \text{ for any } t > 0.$$

2.3 Review of Cai and Prentice estimating equations

2.3.1 Cai and Prentice estimating equations

Consider a Cox proportional hazards model for clustered data with a common baseline hazard function $\lambda_0(t)$. The hazard function of the *jth* subject in the *kth* cluster is

$$\lambda_{kj}(t) = \lambda_0(t)\exp(\boldsymbol{\beta}'\mathbf{Z_{kj}}),$$

where $\mathbf{Z}_{kj}$ is a covariate vector, $k = 1, ..., K$, and $j = 1, ..., n_k$. For notational simplicity, we will assume $n_k = n$, equal cluster sizes. The estimation procedure can easily be modified to apply to clusters of unequal size, by defining $n = max(n_1, ..., n_K)$, and introducing an indicator variable $\xi_{kj} = 1$ if $j \leq n_k$ and 0 otherwise.

Assuming both within and between cluster independence, the partial likelihood using counting process notation is

$$L(\boldsymbol{\beta}, t) \;=\; \prod_{k=1}^{K}\prod_{j=1}^{n}\prod_{0\leq u\leq t}\left[\frac{\exp(\boldsymbol{\beta}'\mathbf{Z_{kj}})}{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u)\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})}\right]^{dN_{kj}(u)}.$$

and the logarithm of the partial likelihood is

$$\begin{aligned} l(\boldsymbol{\beta}, t) = \log L(\boldsymbol{\beta}, t) \;&=\; \sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t \boldsymbol{\beta}'\mathbf{Z}_{kj}dN_{kj}(u) \\ &\quad -\sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t \log\left(\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u)\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})\right) dN_{kj}(u), \end{aligned}$$

where $dN_{kj}(u)$ is a counting process differential, and

$$E(dN_{kj}(u)) = dA_{kj}(u) = Y_{kj}(u)\exp(\boldsymbol{\beta}'\mathbf{Z}_{kj})d\Lambda_0(u).$$

Note that $d\Lambda_0(u)$ can be estimated by the Nelson-Aalen estimator as

$$d\hat{\Lambda}_0(u) = \frac{\sum_{l=1}^{K}\sum_{q=1}^{n} dN_{lq}(u)}{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u)\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})}.$$

First order partial derivatives of the log partial likelihood are

$$
\begin{aligned}
\frac{\partial \log L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t \left\{ \mathbf{Z}_{kj} - \frac{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u)\mathbf{Z}_{lq}\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})}{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u)\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})} \right\} dN_{kj}(u) \\
&= \sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t \mathbf{Z}_{kj} dN_{kj}(u) \\
&\quad - \sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t \frac{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u)\mathbf{Z}_{lq}\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})}{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u)\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})} dN_{kj}(u) \\
&= \sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t \mathbf{Z}_{kj} dN_{kj}(u) \\
&\quad - \int_0^t \frac{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u)\mathbf{Z}_{lq}\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})}{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u)\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})} \sum_{k=1}^{K}\sum_{j=1}^{n} dN_{kj}(u) \\
&= \sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t \mathbf{Z}_{kj} \{ dN_{kj}(u) \\
&\quad \left. - Y_{kj}(u)\exp(\boldsymbol{\beta}'\mathbf{Z}_{kj}) \frac{\sum_{l=1}^{K}\sum_{q=1}^{n} dN_{lq}(u)}{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u)\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})} \right\} \\
&= \sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t \mathbf{Z}_{kj} \left\{ dN_{kj}(u) - Y_{kj}(u)\exp(\boldsymbol{\beta}'\mathbf{Z}_{kj}) d\hat{\Lambda}_0(u) \right\} \\
&= \sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t \mathbf{Z}_{kj} \left\{ dN_{kj}(u) - d\hat{A}_{kj}(u) \right\} \\
&= \sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t \mathbf{Z}_{kj} d\hat{M}_{kj}(u).
\end{aligned}
\tag{2.1}
$$

The partial likelihood score equations can be expressed in vector notation as

$$
\sum_{k=1}^{K}\int_0^t \mathbf{Z}_k' d\hat{\mathbf{M}}_k(u) = \sum_{k=1}^{K}\int_0^t \mathbf{Z}_k' d(\mathbf{N}_k(u) - \hat{\mathbf{A}}_k(u)) = 0 \tag{2.2}
$$

where $\mathbf{Z}_k' = (\mathbf{Z}_{k1}, \mathbf{Z}_{k2}, ..., \mathbf{Z}_{kn})$ is a $p \times n$ matrix, of covariate values for subjects in the

kth cluster, and

$$\hat{\mathbf{M}}_k(u) = \begin{pmatrix} N_{k1}(u) - \int_0^u Y_{k1}(s)\exp(\boldsymbol{\beta}'\mathbf{Z}_{k1}) \dfrac{\sum_{l=1}^K \sum_{q=1}^n dN_{lq}(s)}{\sum_{l=1}^K \sum_{q=1}^n Y_{lq}(s)\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})} \\ N_{k2}(u) - \int_0^u Y_{k2}(s)\exp(\boldsymbol{\beta}'\mathbf{Z}_{k2}) \dfrac{\sum_{l=1}^K \sum_{q=1}^n dN_{lq}(s)}{\sum_{l=1}^K \sum_{q=1}^n Y_{lq}(s)\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})} \\ \vdots \\ N_{kn}(u) - \int_0^u Y_{kn}(s)\exp(\boldsymbol{\beta}'\mathbf{Z}_{kn}) \dfrac{\sum_{l=1}^K \sum_{q=1}^n dN_{lq}(s)}{\sum_{l=1}^K \sum_{q=1}^n Y_{lq}(s)\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})} \end{pmatrix}$$

is a $n \times 1$ vector.

Cox (1972) suggested that a consistent estimator of the asymptotic covariance matrix is provided by the inverse of the Fisher information matrix

$$I(\hat{\beta}) = ((I_{xy}(\hat{\beta}))),$$

where

$$I_{xy}(\hat{\beta}) = -E(\frac{\partial^2 logL(\beta)}{\partial\beta_x \partial\beta_y}).$$

It can be estimated by

$$\hat{I}_{xy}(\hat{\beta}) = -\frac{\partial^2 logL(\beta)}{\partial\beta_x \partial\beta_y}\Big|_{\beta=\hat{\beta}}.$$

This will generally not provide a consistent estimator for the sample covariance matrix for $\hat{\beta}$ when the data contain clusters of correlated survival times, although $\hat{\beta}$ may still be a consistent estimator for β with a limiting normal distribution. Huster et al. (1989) propose an independent working model approach for paired survival data. They specify parametric models for the marginal distributions and obtain a "working" likelihood from the incorrect assumption that the members of each pair respond independently. Then the working likelihood is the product of the marginal likelihoods over all pairs. This is called an independent working model (IWM) likelihood. The parameters of the marginal models can be estimated by maximizing the IWM likelihood. The resulting IWM estimator is equivalent to the solution to equation (2.2). It is denoted by $\hat{\beta}_u$. When the within cluster independence assumption is violated, but between cluster independence holds, $\hat{\beta}_u$ can be a consistent estimator for β with a limiting normal distribution, but the covariance matrix for the IWM model does not provide a consistent estimator for the covariance matrix of the limiting normal distribution. It must be modified to reflect the within cluster correlation. Royall (1986) shows how to obtain a consistent estimator of the covariance matrix, known as a robust or "sandwich" estimator. Wei et al. (1989) apply this approach to multivariate survival data. They model the marginal distribution of failure time variables using a Cox proportional hazards model without specifying the dependence structure of repeated failure times monitored on the same individual. Lee et al. (1992) also show that IWM estimators of regression parameters are consistent. The IWM approach is conceptually straightforward. It avoids explicitly modeling the dependence structure of multivariate survival times, but some efficiency may be lost.

Cai and Prentice (1997) introduce weight matrices into the IWM score equations (2.2) to create a potentially more efficient set of weighted estimating equations for survival analysis. The modified equations are

$$\sum_{k=1}^{K} \int_0^t \mathbf{Z}_k' \mathbf{W}_k d\hat{\mathbf{M}}_k(u) = 0, \tag{2.3}$$

where $\mathbf{W}_k$ is a weight matrix of the form

$$\mathbf{W}_k = \begin{pmatrix} 1 & \rho_{k12} & \cdots & \rho_{k1n} \\ \rho_{k12} & 1 & \cdots & \rho_{k2n} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{kn1} & \rho_{kn2} & \cdots & 1 \end{pmatrix}^{-1},$$

with

$$\rho_{kij} \quad = \quad corr\{M_{ki}(C_{ki}), M_{kj}(C_{kj}) | C_{ki}, C_{kj}, Z_{ki}, Z_{kj}\} \tag{2.4}$$

and $\mathbf{C}_{ki}$ is a censoring time for the ith subject in the kth cluster, $i = 1, ..., n$. The solution to equation (2.3) denoted by $\hat{\beta}_{cw}$ is shown to be consistent with an asymptotic normal distribution under regularity conditions given by Cai and Prentice (1997).

Elements of the weighted matrix given by equation (2.4) are calculated conditional on known censoring times. In some cases when failures occur, censoring times may not be observed and the weights cannot be obtained. To make the weighted equations usable in those cases, estimated weight matrices must be used in equation (2.3) to form estimating equations,

$$\sum_{\mathbf{k=1}}^{K} \int_0^t \mathbf{Z}_k' \hat{\mathbf{W}}_k d\hat{\mathbf{M}}_k(u) = 0, \tag{2.5}$$

where $\hat{\mathbf{W}}_k$ is the estimate of the true weight.

Cai and Prentice (1997) provide a nonparametric method to estimate weights. Elements of the weight matrix are estimated conditional on all the possible pairs of the covariates. The solution to equation (2.5) with estimated weights by the Cai and Prentice approach is denoted by $\hat{\beta}_{ce}$. This method of estimating weights is quite complex and results in an overwhelming computational burden. A large number of parameters must be estimated from relatively small samples in order to estimate weights by this approach. This is noted by Therneau and Grambsch (2000) page 170.

2.3.2 Asymptotic results

Cai and Prentice (1997) prove that under sufficient regularity conditions, $\hat{\beta}_{cw}$ is a consistent estimator of the true parameter vector, β_0, and $K^{\frac{1}{2}}(\hat{\beta}_{cw} - \beta_0)$ is asymptotically normally distributed with mean zero and covariance matrix $\Sigma = A_w^{-1}(\beta_0)\Sigma_w A_w^{-1}(\beta_0)$. To consider the regularity conditions, define

$$S_j^{(d)}(\beta, t) = K^{-1} \sum_{k=1}^{K} Y_{kj}(t) Z_{kj}^d(t) \exp\{\beta' Z_{kj}(t)\}, \ d = 0, 1$$

$$S_j^{(d)}(\beta,t) = K^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n} Z_{ki}(t)w_{kij}(\beta,t)Y_{kj}(t)\{Z'_{kj}(t)\}^{d-2}\exp\{\beta' Z_{kj}(t)\},\ d=2,3$$

$$Q(\beta,t) = \frac{\sum_{j=1}^{n} S_j^{(2)}(\beta,t)}{\sum_{j=1}^{n} S_j^{(0)}(\beta,t)}$$

and

$$V(\beta,t) = \frac{\sum_{j=1}^{n} S_j^{(3)}(\beta,t)}{\sum_{j=1}^{n} S_j^{(0)}(\beta,t)} - \frac{\left(\sum_{j=1}^{n} S_j^{(2)}(\beta,t)\right)\left(\sum_{j=1}^{n} S_j^{(1)}(\beta,t)\right)'}{\left(\sum_{j=1}^{n} S_j^{(0)}(\beta,t)\right)^2}$$

In a neighborhood of the true value β_0, $S_j^{(d)}$, d=0,..,3, satisfies the conditions that

$$sup_{\beta,t}\|S_j^{(d)}(\beta,t) - s_j^{(d)}(\beta,t)\| \to 0,$$

and there exists a positive definite matrix $\Sigma_w = \Sigma_w(\beta)$ such that

$$K^{-1}\sum_{k=1}^{K} var D_k \to \Sigma_w,$$

where

$$D_k = \sum_{j=1}^{n}\int_0^1 \{\sum_{i=1}^{n} Z_{ki}(u)w_{kij}(\beta_0,u) - q(\beta_0,u)\}M_{kj}(du),$$

and

$$q(\beta,u) = (\sum_{j=1}^{n} s_j^{(2)}(\beta,u))/(\sum_{j=1}^{n} s_j^{(0)}(\beta,u)).$$

Define

$$A_w(\beta) = \int_0^1 v(\beta,u)\sum_{j=1}^{n} s_j^{(0)}(\beta,u)\lambda_0(u)du,$$

where

$$v = \frac{\sum_{j=1}^{n} s_j^{(3)}}{\sum_{j=1}^{n} s_j^{(0)}} - \frac{\left(\sum_{j=1}^{n} s_j^{(2)}\right)\left(\sum_{j=1}^{n} s_j^{(1)}\right)'}{\left(\sum_{j=1}^{n} s_j^{(0)}\right)^2}$$

With additional regularity conditions on estimated weight matrices, $\hat{\beta}_{ce}$ is a consistent estimator of the true value β_0, and $K^{1/2}(\hat{\beta}_{ce} - \hat{\beta}_0)$ has the same asymptotic distribution as $K^{1/2}(\hat{\beta}_{cw} - \hat{\beta}_0)$.

The asymptotic variance matrix of $\hat{\beta}_{cw}$ or $\hat{\beta}_{ce}$ can be estimated by

$$\hat{\Sigma} = \hat{A}_w^{-1}(\hat{\beta})\hat{\Sigma}_w\hat{A}_w^{-1}(\hat{\beta}), \tag{2.6}$$

where $\hat{\beta} = \hat{\beta}_{cw}$ or $\hat{\beta}_{ce}$,

$$\hat{A}_w(\beta) = -K^{-1}\sum_{k=1}^{K}\sum_{j=1}^{n}\delta_{kj}\left[\frac{\sum_{l=1}^{n}\hat{S}_l^{(3)}(\beta;T_{kj})}{\sum_{l=1}^{n}\hat{S}_l^{(0)}(\beta;T_{kj})} - \frac{\left(\sum_{l=1}^{n}\hat{S}_l^{(2)}(\beta;T_{kj})\right)\left(\sum_{l=1}^{n}\hat{S}_l^{(1)}(\beta;T_{kj})\right)'}{\left(\sum_{l=1}^{n}\hat{S}_l^{(0)}(\beta;T_{kj})\right)^2}\right]$$

and

$$\hat{\Sigma}_w(\beta) = K^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{D}_{ki}(\beta)\hat{D}'_{kj}(\beta),$$

where

$$\begin{aligned}\hat{D}_{kj}(\beta) \quad = \quad & \delta_{kj}\left[\sum_{i=1}^{n}Z_{ki}(T_{kj})\hat{w}_{kij}(\beta;T_{kj}) - \hat{Q}(\beta;T_{kj})\right] \\ & -K^{-1}\sum_{l=1}^{K}\sum_{m=1}^{n}\delta_{lm}Y_{kj}(T_{lm})e^{\beta' Z_{kj}(T_{lm})}\left\{\sum_{l=1}^{n}\hat{S}_t^{(0)}(\beta;T_{lm})\right\}^{-1} \\ & \left[\sum_{i=1}^{n}Z_{ki}(T_{lm})\hat{w}_{kij}(\beta;T_{lm}) - \hat{Q}(\beta;T_{lm})\right]\end{aligned}$$

and δ_{kj} is a censoring indicator of time T_{kj}, being one if a failure happens and zero otherwise.

2.3.3 Finite sample properties

Cai and Prentice (1997) studied the case for which there is only one binary covariate corresponding to a treatment effect. Treatments are randomly assigned to subjects within each cluster, so that both subjects within a pair have probability 0.5 of receiving different treatments. The joint survival distribution is the Copula model examined by Clayton and Cuzick (1985) with an association parameter θ corresponding to the level of dependence of failure times where $\theta \to 0$ corresponds to independence, and $\theta \to \infty$ corresponds maximal positive dependence. More details are provided in Section 2.5.2. Their simulation studies show that compared to $\hat{\beta}_u$, the gain of efficiency of $\hat{\beta}_{ce}$

is substantial when within cluster correlation is strong, and the censoring rate is low. Relative to the IWM estimator for β, not much is gained by using weighted estimating equations if within cluster correlation is moderate, or the censoring is heavy. The gains in efficiency tend to be larger for values of β closer to zero.

2.4 Modifications of Cai and Prentice approach

2.4.1 Modification of the weight matrix

Even though simulation results show that Cai and Prentice (1997) estimating equations can provide significant gains in efficiency in the case of low censoring rate and high correlation, the complexity, computational intensity, and lack of software for public use limit the practical use of the method.

Emura and Tsukuma (2003) modify the method of estimating weights for the Cai and Prentice estimating equations. Their estimated weight matrices are defined by

$$\mathbf{W}_k = \begin{pmatrix} 1 & \rho_{k12} & \cdots & \rho_{k1n} \\ \rho_{k12} & 1 & \cdots & \rho_{k2n} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{k1n} & \rho_{k2n} & \cdots & \rho_{knn} \end{pmatrix}^{-1},$$

where

$$\rho_{kij} = corr\{\hat{M}_{ki}(t_{ki}), \hat{M}_{kj}(t_{kj}) | Z_{ki}, Z_{kj}\}, \tag{2.7}$$

Z_{kj} is the value of the covariate, and $\hat{M}_{kj}(t_{kj})$ is the martingale for the jth subject in the kth cluster at observed time t_{kj}. The correlations used in weight matrices are simple correlation functions of martingales conditional on the combinations of covariates within a cluster. For example, if a covariate is binary taking a value of 0 or 1 and there are two respondents in each cluster, there are three combinations of covariate values, (0,0), (1,1) and (0,1). The (1,0) combination is equivalent to the (0,1) combination because the respondents have no ordering within clusters. A data set can be partitioned into 3 subsets according to the combinations of covariate values. One correlation of martingales is calculated for each subset. This is much easier conceptually, and much faster computationally than Cai and Prentice nonparametric approach. However, estimation of such correlations requires replication of covariate pairs. Even though values of continuous covariates and categorical covariates with multiple levels can be grouped to provide pseudo replication, the determination of groups may be quite arbitrary. Also generalizing the weighted estimating equations to cases where there are several covariates will result in many grouping categories with small sample sizes or coarse grouping categories. Therefore, we propose using

$$\rho = corr\{\hat{M}_{ki}(t_{ki}), \hat{M}_{kj}(t_{kj})\}, \tag{2.8}$$

to further simplify and generalize the procedure for estimating weight matrices, where

$$\hat{M}_{ki}(t_{ki}) = N_{ki}(t_{ki}) - \int_0^{t_{ki}} Y_{ki}(u) e^{\boldsymbol{\beta}'\mathbf{Z}_{ki}} d\hat{\Lambda}_0(u).$$

In (2.8), the correlation does not depend on covariate values. It is easier and faster to compute than the correlations defined by (2.7). More importantly, this approach is easily applied to cases with continuous covariates or any number of covariates with any number of levels. The solution to equation (2.5) using the correlation defined in (2.7) or (2.8) is denoted by $\hat{\beta}_c$.

2.4.2 Alternative weighted estimating equations

The partial likelihood equations for the Cox model with independent observations can be written in another form by rearranging the terms in equation (2.1). Note that

$$\begin{aligned}
\frac{\partial \log L(\beta)}{\partial \beta} &= \sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t \mathbf{Z}_{kj} dN_{kj}(u) \\
&\quad -\sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t \frac{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u)\mathbf{Z}_{lq}\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})}{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u)\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})} dN_{kj}(u) \\
&= \sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t \left\{ \mathbf{Z}_{kj} - \frac{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u)\mathbf{Z}_{lq}\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})}{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u)\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})} \right\} dN_{kj}(u) \\
&= \sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t (\mathbf{Z}_{kj} - E(u)) dN_{kj}(u) \\
&= 0 \qquad (2.9)
\end{aligned}$$

where

$$E(u) = \frac{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u)\mathbf{Z}_{lq}e^{\beta'\mathbf{Z}_{lq}}}{\sum_{l=1}^{K}\sum_{q=1} Y_{lq}(u)e^{\boldsymbol{\beta}'\mathbf{Z}_{lq}}}.$$

Fleming and Harrington (1991) page 150 show that

$$\sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t (\mathbf{Z}_{kj} - E(u))Y_{kj}(u)\exp(\boldsymbol{\beta}'\mathbf{Z}_{kj})d\Lambda_0(u) = 0,$$

and

$$dA_{kj}(u) = Y_{kj}(u)\exp(\boldsymbol{\beta}'\mathbf{Z}_{kj})d\Lambda_0(u).$$

It follows that

$$\sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t (\mathbf{Z}_{kj} - E(u))dA_{kj}(u) = 0.$$

Furthermore,

$$dM_{kj}(u) = dN_{kj}(u) - dA_{kj}(u),$$

and equation (2.9) can be written as

$$\begin{aligned}\frac{\partial \log L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \sum_{k=1}^{K}\sum_{j=1}^{n}\int_0^t (\mathbf{Z}_{kj} - E(u))dN_{kj}(u) \\ &= \sum_{k=1}^{K}\int_0^t (\mathbf{Z}_k - E(u))'(d\mathbf{N}_k(u) - d\mathbf{A}_k(u)) \\ &= \sum_{k=1}^{K}\int_0^t (\mathbf{Z}_k - E(u))'d\mathbf{M}_k(u) \\ &= 0 \end{aligned} \tag{2.10}$$

If $d\Lambda_0(u)$ is estimated by the Nelson-Aalen estimator, then

$$\hat{A}_{kj}(u) = \int_0^u Y_{kj}(s)\exp(\boldsymbol{\beta}'\mathbf{Z}_{kj})\frac{\sum_{l=1}^{K}\sum_{q=1}^{n} dN_{lq}(s)}{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(s)\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})}$$

and equation (2.10) becomes

$$\sum_{k=1}^{K}\int_0^t (\mathbf{Z}_k - E(u))'d\hat{\mathbf{M}}_k(u) = 0. \tag{2.11}$$

where $d\hat{\mathbf{M}}_k(u) = d\mathbf{N}_k(u) - d\hat{\mathbf{A}}_k(u)$.

As the weighted equations in (2.5) are obtained by inserting weights into (2.2), another set of weighted estimating equations is obtained by inserting weights into equation (2.11) to form

$$U(\boldsymbol{\beta}, t) = \sum_{k=1}^{K}\int_0^t (\mathbf{Z}_k - E(u))'\hat{\mathbf{W}}_k d\hat{\mathbf{M}}_k(u) = 0 \tag{2.12}$$

The solution of equation (2.12) is denoted by $\hat{\beta}_w$. Using the same weight matrix defined in (2.7) or (2.8), $\hat{\beta}_c$ and $\hat{\beta}_w$ were compared in simulation studies with respect to estimated bias and empirical variance.

Note that it is also seen in the literature that Cox proportional model is written in the form as

$$\lambda_{kj}(t) = Y_{kj}(t)\lambda_0(t)e^{\boldsymbol{\beta}'\mathbf{Z}_{kj}}.$$

Using this model, the partial likelihood score equation is of the form

$$\sum_{k=1}^{K} \int_0^t \{\mathbf{Y}_k(u)(\mathbf{Z}_k - E(u))\}' d\hat{\mathbf{M}}_k(u) = 0.$$

It is equivalent to equation (2.11). When weights are inserted, however,

$$\sum_{k=1}^{K} \int_0^t \{\mathbf{Y}_k(u)(\mathbf{Z}_k - E(u))\}' \hat{\mathbf{W}}_k d\hat{\mathbf{M}}_k(u) = 0 \tag{2.13}$$

is not equivalent to equation (2.12). Simulation results show that equation (2.13) produces highly biased estimators of β.

2.4.3 Illustration of the weighted estimating equations

To illustrate the construction of the estimating equations, consider the data from the following simple experiment. There are three clusters, and each provides a pair of responses. A cross indicates a failure event and a circle indicates a censoring event.

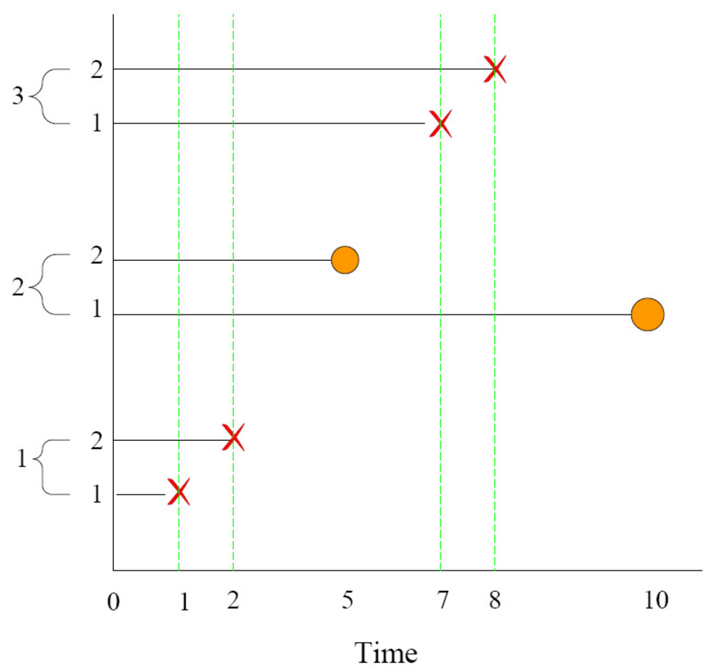

Figure 2.1 Illustration of weighted estimating equations

Only failure times (1, 2, 7, 8) are included in the equations. If we use a common baseline hazard function and a single bivariate covariate Z corresponding to the treatment effect involved, the following illustrates the contributions of the three pairs of observations to the estimating equation at times 1, 2, 7 and 8, respectively, where Z_{kj} is the treatment indicator for the jth subject in the kth cluster.

At time=1

$$\begin{pmatrix} Z_{11} - E(1) \\ Z_{12} - E(1) \end{pmatrix}' \hat{\mathbf{W}}_1 \begin{pmatrix} 1 - e^{\beta Z_{11}} d\Lambda_0(1) \\ 0 - e^{\beta Z_{12}} d\Lambda_0(1) \end{pmatrix}$$

$$\begin{pmatrix} Z_{21} - E(1) \\ Z_{22} - E(1) \end{pmatrix}' \hat{\mathbf{W}}_2 \begin{pmatrix} 0 - e^{\beta Z_{21}} d\Lambda_0(1) \\ 0 - e^{\beta Z_{22}} d\Lambda_0(1) \end{pmatrix}$$

$$\begin{pmatrix} Z_{31} - E(1) \\ Z_{32} - E(1) \end{pmatrix}' \hat{\mathbf{W}}_3 \begin{pmatrix} 0 - e^{\beta Z_{31}} d\Lambda_0(1) \\ 0 - e^{\beta Z_{32}} d\Lambda_0(1) \end{pmatrix}$$

At time=2

$$\begin{pmatrix} Z_{11} - E(2) \\ Z_{12} - E(2) \end{pmatrix}' \hat{\mathbf{W}}_1 \begin{pmatrix} 0 \\ 1 - e^{\beta Z_{12}} d\Lambda_0(2) \end{pmatrix}$$

$$\begin{pmatrix} Z_{21} - E(2) \\ Z_{22} - E(2) \end{pmatrix}' \hat{\mathbf{W}}_2 \begin{pmatrix} 0 - e^{\beta Z_{21}} d\Lambda_0(2) \\ 0 - e^{\beta Z_{22}} d\Lambda_0(2) \end{pmatrix}$$

$$\begin{pmatrix} Z_{31} - E(2) \\ Z_{32} - E(2) \end{pmatrix}' \hat{\mathbf{W}}_3 \begin{pmatrix} 0 - e^{\beta Z_{31}} d\Lambda_0(2) \\ 0 - e^{\beta Z_{32}} d\Lambda_0(2) \end{pmatrix}$$

At time=7

$$\begin{pmatrix} Z_{21} - E(7) \\ Z_{22} - E(7) \end{pmatrix}' \hat{\mathbf{W}}_2 \begin{pmatrix} 0 - e^{\beta Z_{21}} d\Lambda_0(7) \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} Z_{31} - E(7) \\ Z_{32} - E(7) \end{pmatrix}' \hat{\mathbf{W}}_3 \begin{pmatrix} 1 - e^{\beta Z_{31}} d\Lambda_0(7) \\ 0 - e^{\beta' Z_{32}} d\Lambda_0(7) \end{pmatrix}$$

At time=8

$$\begin{pmatrix} Z_{21} - E(8) \\ Z_{22} - E(8) \end{pmatrix}' \hat{\mathbf{W}}_2 \begin{pmatrix} 0 - e^{\beta Z_{21}} d\Lambda_0(8) \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} Z_{31} - E(8) \\ Z_{32} - E(8) \end{pmatrix}' \hat{\mathbf{W}}_3 \begin{pmatrix} 0 \\ 1 - e^{\beta Z_{32}} d\Lambda_0(8) \end{pmatrix}$$

where

$$E(u) = \frac{\sum_{l=1}^{K} \sum_{q=1}^{n} Y_{lq}(u) Z_{lq} \exp(\beta Z_{lq})}{\sum_{l=1}^{K} \sum_{q=1}^{n} Y_{lq}(u) \exp(\beta Z_{lq})},$$

$$d\hat{\Lambda}_0(u) = \frac{dN_{kj}(u)}{\sum_l \sum_q Y_{lq}(u) \exp(\beta Z_{lq})}$$

and $\hat{\mathbf{W}}_k$ is defined either in (2.7) or (2.8).

This can be easily generalized to the case with multiple covariates. Denote the ith covariate of the jth subject in the kth cluster by Z_{kji}, where $i = 1, ..., p$. Contributions to equation (2.12) at event time $t = 1$ become

$$\begin{pmatrix} Z_{111} - E_1(1) & Z_{121} - E_1(1) \\ \vdots & \vdots \\ Z_{11p} - E_p(1) & Z_{12p} - E_p(1) \end{pmatrix} \hat{\mathbf{W}}_1 \begin{pmatrix} 1 - e^{\boldsymbol{\beta}' \mathbf{Z}_{11}} d\Lambda_0(1) \\ 0 - e^{\boldsymbol{\beta}' \mathbf{Z}_{12}} d\Lambda_0(1) \end{pmatrix}$$

$$\begin{pmatrix} Z_{211} - E_1(1) & Z_{221} - E_1(1) \\ \vdots & \vdots \\ Z_{21p} - E_p(1) & Z_{22p} - E_p(1) \end{pmatrix} \hat{\mathbf{W}}_1 \begin{pmatrix} 1 - e^{\boldsymbol{\beta}' \mathbf{Z}_{21}} d\Lambda_0(1) \\ 0 - e^{\boldsymbol{\beta}' \mathbf{Z}_{22}} d\Lambda_0(1) \end{pmatrix}$$

$$\begin{pmatrix} Z_{311} - E_1(1) & Z_{321} - E_1(1) \\ \vdots & \vdots \\ Z_{31p} - E_p(1) & Z_{32p} - E_p(1) \end{pmatrix} \hat{\mathbf{W}}_1 \begin{pmatrix} 1 - e^{\boldsymbol{\beta}'\mathbf{Z}_{31}} d\Lambda_0(1) \\ 0 - e^{\boldsymbol{\beta}'\mathbf{Z}_{32}} d\Lambda_0(1) \end{pmatrix}$$

where $\mathbf{Z}_{kj}$ is a column vector of (Z_{kj1}, Z_{kj2}) and

$$E_1(u) = \frac{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u) Z_{lq1} \exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})}{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u) \exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})}$$

$$E_p(u) = \frac{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u) Z_{lqp} \exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})}{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u) \exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})}$$

2.5 Description of simulation studies

2.5.1 Simulation studies

In the following simulation studies we assess the behavior of the unweighted estimator, $\hat{\beta}_u$, the weighted estimators, $\hat{\beta}_w$ provided by equation (2.12), and $\hat{\beta}_c$ provided by equation (2.5). We compare the performance of these estimators using weight matrices defined in (2.7) and (2.8), with different treatment assignments, censoring types, censoring rates and levels of dependence.

Cai and Prentice (1997) only consider a random treatment assignment in which the treatment assignment is independently determined for each individual in a cluster. For clusters of size 2, the individuals can both receive the treatment with probability 0.25, both receive the control with probability 0.25, or one receives the treatment and the other one receives the control with probability 0.5. In addition to this, we also investigate a balanced random treatment assignment, in which one individual in the cluster is randomly selected to receive the treatment, and the other member of the cluster is assigned to the control.

All the simulation studies in this section are restricted to the case where there is a single binary covariate, corresponding to assignment to treatment or control, with clusters of size 2. 500 data sets are simulated for each simulation study.

2.5.2 Simulation of correlated survival data

The procedures of generating failure times and censoring times follow the method used by Cai and Prentice (1997). Using the Copula model examined by Clayton and Cuzick (1985), a bivariate survival distribution is defined as

$$Pr(T_{k1} > t_{k1}, T_{k2} > t_{k2} | Z_{k1}, Z_{k2}) = \{1 - F_1(t_{k1}; Z_{k1})^{-1/\theta} + F_2(t_{k2}; Z_{k2})^{-1/\theta}\}^{-\theta},$$

where F_1 and F_2 are specified univariate distribution functions. Bivariate exponential failure times (T_{k1}, T_{k2}) were generated from independent uniform variables u_{k1}, u_{k2} via

$$T_{k2} = -\log(1 - \mu_{k2})e^{-\boldsymbol{\beta}'\mathbf{Z}_{k2}}$$

and

$$T_{k1} = \theta \log\{(1 - a) + a(1 - \mu_{k1})^{-(1+\theta)^{-1}}\}e^{-\boldsymbol{\beta}'\mathbf{Z}_{k2}},$$

with $a = (1 - \mu_{k2})^{-\theta^{-1}}$. Then, the marginal probability density function of T_{kj} is

$$f(t_{kj}) = e^{\boldsymbol{\beta}'\mathbf{Z}_{kj}} \exp(-t_{kj}e^{\boldsymbol{\beta}'\mathbf{Z}_{kj}}), \text{ where } j = 1, 2.$$

Joint exponential survival times can be generated from a multivariate version of this construction as

$$T_{k1} = -log(1 - u_1)e^{-\boldsymbol{\beta}'\mathbf{Z}_{k1}}$$

$$T_{km} = \theta log\{(m-1) - \sum_{i=1}^{m-1} a_{ki} + (\sum_{i=1}^{m-1} a_{ki} - (m-2))(1 - u_{km})^{-\frac{1}{\theta+m-1}}\}e^{-\boldsymbol{\beta}'\mathbf{Z}_{km}}$$

where $a_{kl} = e^{\frac{1}{\theta}t_{kl}e^{\boldsymbol{\beta}'\mathbf{Z}_{kl}}}$, $l = 1, ..., m-1$, and $m = 2, ..., n$. Then, the marginal probability density function of T_{kj} is

$$f(t_{kj}) = e^{\boldsymbol{\beta}'\mathbf{Z}_{kj}} \exp(-t_{kj}e^{\boldsymbol{\beta}'\mathbf{Z}_{kj}}), \text{ where } j = 1, ..., n.$$

Z_{kj} is a scalar when there is only one covariate. In the case of a balanced randomized design for clusters of size 2, Z_{k1} is 1 and Z_{k2} is 0 for every cluster. For a completely randomized design, Z_{kj} is a binary variate that has value 1 with probability π_k, where $k = 1, ..., K$. In this study we use $\pi = 0.5$ for all k. Then, (Z_{k1}, Z_{k2}) are obtained from independent uniform variates (u_3, u_4) using the transformation

$$Z_{k1} = \begin{cases} 1 & \text{if } u_3 \leq \pi \\ 0 & \text{otherwise} \end{cases}$$

and

$$Z_{k2} = \begin{cases} 1 & \text{if } u_4 \leq \pi \\ 0 & \text{otherwise} \end{cases}$$

Correlations among failure times for different θ values are shown in Table 2.1.

θ	correlation
$\rightarrow 0$	$\rightarrow 1$
.25	.937
.80	.712
1.50	.512
$\rightarrow \infty$	$\rightarrow 0$

Table 2.1 Correlation among failure times for different θ values

2.5.3 Simulation of censoring times

In order to maintain similarities of simulation procedures used by Cai and Prentice, we adopted their method of generating independent random censoring time. Independent random censoring times C^*_{kj} were simulated from an exponential distribution with density

$$f_{c^*_{kj}}(x) = \alpha e^{-\alpha x}.$$

Values of (C^*_{k1}, C^*_{k2}) can be obtained by

$$C^*_{k1} = -\log(u_5)/\alpha,$$

$$C^*_{k2} = -\log(u_6)/\alpha,$$

where u_5 and u_6 were independently generated from uniform (0,1) distributions.

The censoring we use here is a combination of random censoring and Type I censoring (censoring at a fixed time), whichever happens first. Assume that the inspection is terminated for any individual who survives through time b, so the interval of interest for each individual is $[0, b]$. Then the censoring times C_{k1} and C_{k2} are

$$C_{k1} = b \wedge C^*_{k1},$$

$$C_{k2} = b \wedge C^*_{k2}.$$

It is assumed that any (C^*_{k1}, C^*_{k2}) are independent of any (T_{k1}, T_{k2}). Let P_b denote the censoring percentage. Then

$$\begin{aligned}
P_b &= P(C \leq T | T \leq b) + P(T > b) \\
&= \int_0^b \int_0^t f_T(t) f_C(c) dc dt + \int_b^\infty f_T(t) dt \\
&= \int_0^b \int_0^t e^{\beta Z} e^{-te^{\beta Z}} \alpha e^{-\alpha c} dc dt + \int_b^\infty e^{\beta Z} e^{-te^{\beta Z}} dt \\
&= \frac{e^{-b(e^{\beta z}+\alpha)} + \alpha e^{-\beta Z}}{1 + \alpha e^{\beta Z}}
\end{aligned}$$

This result can be used to derive a value of α for any desired termination time b. Assume the time interval we are interested in is $[0, 5]$. Table 2.2 gives values of α for specific values of P_b, Z, and β used in the simulation study.

For the cases considered in this simulation study, less than 0.1 percent of the survival times exceed $b = 5$, and most of the censoring is random censoring. Table 2.7 shows the effects of type of censoring on the behavior of the estimators.

		P_5		
β	Z	10%	20%	50%
0.693	0	.107	.247	1
	1	.222	.500	2
0	0	.107	.247	1
	1	.107	.247	1
-0.5	0	.107	.247	1
	1	.409	.133	.604

Table 2.2 The values of α used to simulate censoring times for values of β, Z, and P_5 used in this study

2.6 Simulation study

All the simulation results shown here are for cases for which the cluster size is two and there is only one binary covariate, taking a value of 1 or 0. Estimators of the regression parameters are denoted by $\hat{\beta}_u$, $\hat{\beta}_c$ and $\hat{\beta}_w$ for estimating equations (2.2), (2.5) and (2.12), respectively. The estimated bias and empirical variances of these estimators are compared. Each simulation result displayed in this section is based on 500 simulated data sets. In the following tables, estimated bias is the average of 500 simulated estimates minus the true coefficient value, and an empirical variance is also calculated from each sample of 500 estimates.

2.6.1 Comparison of weight matrices

First consider the situation with no censoring. The correlation given by (2.7) may depend on the covariate values, e.g.

$$\rho_{k11} = corr(\hat{M}_{ki}(t_{ki}), \hat{M}_{kj}(t_{kj})|Z_{ki} = 1, Z_{kj} = 1)$$
$$\rho_{k10} = corr(\hat{M}_{ki}(t_{ki}), \hat{M}_{kj}(t_{kj})|Z_{ki} = 1, Z_{kj} = 0)$$
$$\rho_{k00} = corr(\hat{M}_{ki}(t_{ki}), \hat{M}_{kj}(t_{kj})|Z_{ki} = 0, Z_{kj} = 0)$$

The correlation is estimated by Pearson correlation among martingales. The correlations defined in (2.8) are easier to estimate and generalize to any type or any number of covariates. But (2.8) could provide less efficient regression parameter estimators because weights are evaluated by averaging across covariate values. Simulation results for these two sets of correlations are shown in Tables 2.3, 2.4 and 2.5.

In all the tables, the true coefficient is $\beta = 0.693$. In Table 2.3, the bivariate exponential failure times are generated with association parameter $\theta = 0.25$ for all covariate pairs. The corresponding correlation is 0.937. The estimated bias and efficiency of both $\hat{\beta}_c$ and $\hat{\beta}_w$ using the two types of correlations are similar when failure times within clusters have about same level of association for all covariate combinations. The estimated correlations conditional on covariate pairs are 0.913 for covariate pair (1,1), 0.916 for covariate pair (1,0), and 0.907 for covariate pair (0,0). The estimated correlation evaluated by averaging across the covariate values is 0.912. Both $\hat{\beta}_c$ and $\hat{\beta}_w$ are similar with correlation estimated by either way.

	correlations calculated conditional on covariate pairs		correlation evaluated by averaging across covariate values	
	estimated bias	empirical variance	estimated bias	empirical variance
$\hat{\beta}_u$	0.0164	0.0226	0.0164	0.0226
$\hat{\beta}_c$	0.0191	0.0107	0.0120	0.0101
$\hat{\beta}_w$	0.0182	0.0108	0.0120	0.0101

Table 2.3 Simulation result for $\hat{\beta}_u$, $\hat{\beta}_c$, and $\hat{\beta}_w$ for 500 simulated samples of 100 clusters of size 2 with $\beta = 0.693$, $\theta = 0.25$ corresponding to correlation 0.937, and no censoring.

In Table 2.4, the bivariate exponential failure times are generated with association parameters $\theta = 0.25$ for covariate pairs (0,0) or (1,1), and 1.5 for covariate pairs (0,1) or (1,0). The corresponding correlations are 0.937 and 0.512 respectively. The estimated correlations conditional on covariate pairs are 0.913 for covariate pair (1,1), 0.907 for covariate pair (0,0), and 0.468 for covariate pairs (1,0) or (0,1). The estimated correlation evaluated by averaging across the covariate values is 0.688. Neither $\hat{\beta}_c$ nor $\hat{\beta}_w$ loses much efficiency when weights are obtained from correlations averaging across the covariate values.

In Table 2.5, the bivariate exponential failure times are generated with association parameters $\theta = 0.25$ for covariate pairs (0,0) or (1,1), and 10 for covariate pairs (0,1) or (1,0). The corresponding correlations are 0.937 and 0.098 respectively. The estimated correlations conditional on covariate pairs are 0.913 for covariate pair (1,1), 0.906 for covariate pair (0,0), and 0.088 for covariate pairs (1,0) or (0,1). The estimated correlation evaluated by averaging across the covariate values is 0.490. In this case, for two types of

	correlations calculated conditional on covariate pairs		correlation evaluated by averaging across covariate values	
	estimated bias	empirical variance	estimated bias	empirical variance
$\hat{\beta}_u$	0.0220	0.0279	0.0220	0.0279
$\hat{\beta}_c$	0.0201	0.0198	0.0184	0.0204
$\hat{\beta}_w$	0.0197	0.0196	0.0184	0.0204

Table 2.4 Simulation result for $\hat{\beta}_u$, $\hat{\beta}_c$, and $\hat{\beta}_w$ for 500 simulated samples of 100 clusters of size 2 with $\beta = 0.693$, $\theta = 0.25$ corresponding to correlation 0.937 for covariate pairs (0,0) and (1,1), $\theta = 1.5$ corresponding to correlation 0.512 for covariate pairs (1,0) and (0,1) and no censoring.

correlation estimation, neither $\hat{\beta}_c$ nor $\hat{\beta}_w$ has much gain in efficiency relative to $\hat{\beta}_u$.

	correlations calculated conditional on covariate pairs		correlation evaluated by averaging across covariate values	
	estimated bias	empirical variance	estimated bias	empirical variance
$\hat{\beta}_u$	0.0127	0.0333	0.0127	0.0333
$\hat{\beta}_c$	0.0220	0.0328	0.0183	0.0331
$\hat{\beta}_w$	0.0228	0.0327	0.0183	0.0331

Table 2.5 Simulation result for $\hat{\beta}_u$, $\hat{\beta}_c$, and $\hat{\beta}_w$ for 500 simulated samples of 100 clusters of size 2 with $\beta = 0.693$, $\theta = 0.25$ corresponding to correlation 0.937 for covariate pairs (0,0) and (1,1), $\theta = 10$ corresponding to correlation 0.098 for covariate pairs (1,0) and (0,1) and no censoring.

Table 2.4 shows that if correlation for some covariate pairs are very strong (the correlation is around 0.9), and at moderate level for the others (the correlation is around 0.5), there is no efficiency loss by using correlation averaging across the covariate values. Table 2.5 shows that if correlation for some covariate pairs are very strong (the correlation is around 0.9), and for others are very low (correlation around 0.1), neither estimators from two types of weights have much gains in efficiency.

In medical research, models with correlations that depend on covariate values are rarely considered. Based on this and the results in Tables 2.4 and 2.5, we see no practical need to further consider the models. Consequently, we only consider models with correlation that do not depend on covariates.

It can be seen from the tables that $\hat{\beta}_w$ and $\hat{\beta}_c$ are identical when a common correlation is used in the weight matrices. In that case, the solutions to equations (2.5) and (2.12) are mathematically equivalent. At any event time u,

$$\begin{aligned}
& \sum_{k=1}^{K}\sum_{j=1}^{n} d\hat{M}_{kj}(u) \\
= \; & \sum_{k=1}^{K}\sum_{j=1}^{n} dN_{kj}(u) - \sum_{k=1}^{K}\sum_{j=1}^{n} d\hat{A}_{kj}(u) \\
= \; & \sum_{k=1}^{K}\sum_{j=1}^{n} N_{kj}(u) - \sum_{k=1}^{K}\sum_{j=1}^{n} Y_{kj}(u)\exp(\boldsymbol{\beta}'\mathbf{Z}_{kj}) \frac{\sum_{l=1}^{K}\sum_{q=1}^{n} dN_{lq}(u)}{\sum_{l=1}^{K}\sum_{q=1}^{n} Y_{lq}(u)\exp(\boldsymbol{\beta}'\mathbf{Z}_{lq})} \\
= \; & 0,
\end{aligned}$$

Then considering the simplest case of clusters of size 2, the difference between equations (2.5) and (2.12) for clusters of size 2 is

$$\begin{aligned}
& \frac{1}{1-\rho^2}\sum_{k=1}^{K}\int_0^t \begin{pmatrix} E(u) & E(u) \end{pmatrix} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} d\hat{M}_{k1}(u) \\ d\hat{M}_{k2}(u) \end{pmatrix} \\
= \; & \frac{1}{1-\rho^2}\int_0^t \sum_{k=1}^{K} \Big\{(1-\rho)E(u)(d\hat{M}_{k1}(u) + d\hat{M}_{k2}(u))\Big\} = 0.
\end{aligned}$$

Therefore, in that case equations (2.12) and (2.5) are identical at any value of β. Similarly, for clusters of size 3 with exchangeable correlation structure, denote the determinant of the correlation matrix as $|A|$, then the difference between equations (2.5) and

(2.12) is

$$\frac{1}{|A|}\sum_{k=1}^{K}\int_0^t \begin{pmatrix} E(u) & E(u) & E(u) \end{pmatrix} \begin{pmatrix} 1-\rho^2 & \rho^2-\rho & \rho^2-\rho \\ \rho^2-\rho & 1-\rho^2 & \rho^2-\rho \\ \rho^2-\rho & \rho^2-\rho & 1-\rho^2 \end{pmatrix} \begin{pmatrix} d\hat{M}_{k1}(u) \\ d\hat{M}_{k2}(u) \\ d\hat{M}_{k3}(u) \end{pmatrix}$$

$$= \frac{1}{|A|}\sum_{k=1}^{K}\int_0^t \left\{(1-\rho)^2 E(u)(d\hat{M}_{k1}(u) + d\hat{M}_{k2}(u) + d\hat{M}_{k3}(u))\right\}$$

$$= \frac{1}{|A|}\int_0^t \sum_{k=1}^{K} \left\{(1-\rho)^2 E(u)(d\hat{M}_{k1}(u) + d\hat{M}_{k2}(u) + d\hat{M}_{k3}(u))\right\} = 0.$$

This can be generalized to clusters of size n for exchangeable correlation structure. The difference between equations (2.5) and (2.12) becomes

$$1/|A|\int_0^t (1-\rho)^{n-1}(E(u))\sum_{k=1}^{K}\sum_{j=1}^{n} d\hat{M}_{kn}(u),$$

which is equal to 0. Equations (2.12) and (2.5) are not necessarily equivalent when the weights are allowed to depend on the covariate values for clusters of size two. However, as shown in Tables 2.3, 2.4 and 2.5, the efficiency of estimators with two type of correlations are close.

2.6.2 Balanced and unbalanced designs

In Tables 2.3, 2.4 and 2.5, a completely randomized design is used. This means that the two individuals in a pair can both receive the treatment with probability π^2, or both get the control with probability $(1-\pi)^2$, or one receives the treatment and one receives the control with probability $2\pi(1-\pi)$. In our simulations we only use $\pi = 0.5$. If the design is a balanced randomized design, then the values of $\hat{\beta}_u$, $\hat{\beta}_w$, and $\hat{\beta}_c$ are identical. In that case, one of subjects in the pair receives the treatment, and the other one receives the control. The result is illustrated in the Figures 2.2 and 2.3.

Figures 2.2 and 2.3 are results for a single simulated data set with 100 clusters of size 2. There is no censoring considered. β is the true value of coefficient, and θ is the association parameter indicating the degree of correlation of the within cluster failure times. The horizontal axis is the value of estimated coefficient in the equation (2.12). Each curve is the squared value of the estimating equation for one simulation data set. Different curves correspond to different values of the constant correlation ρ, and the weight matrix is the inverse of the martingale residual correlation matrix. Therefore $\rho = 0$ corresponds to the unweighted equation (2.11). The value of the estimated coefficient at the minimum of the curve is the solution of the estimating equations $\hat{\beta}_u$ or $\hat{\beta}_w$, depending on the correlation used in the weight matrices.

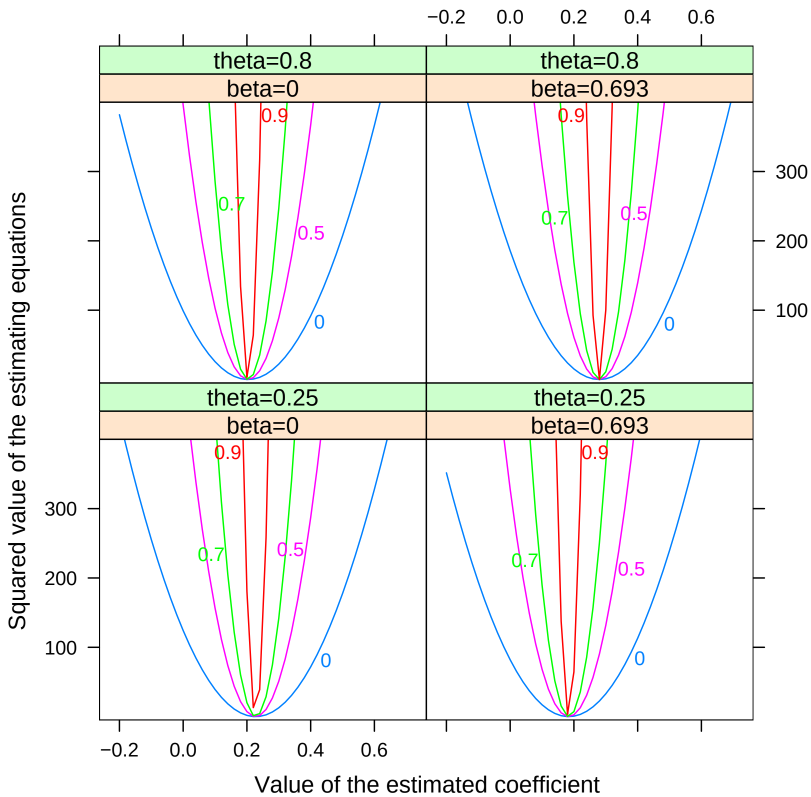

Figure 2.2 Squared values of weighted estimating equations from a single data set with 100 clusters of size 2 using a balanced randomized design. The correlations corresponding to the lines from inside to outside are 0.9, 0.7, 0.5 and 0.

From the plots, it can be seen that the solutions to equations (2.11) and (2.12)

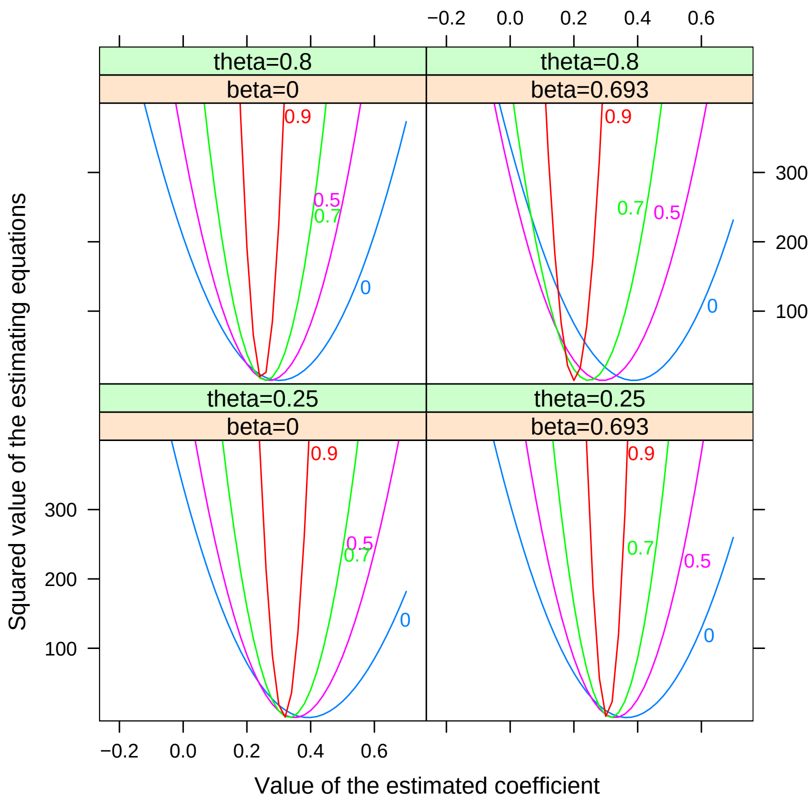

Figure 2.3 Squared values of weighted estimating equations from a single data set with 100 clusters of size 2 using a completely randomized design. The correlations corresponding to the lines from inside to outside are 0.9, 0.7, 0.5 and 0.

are not necessarily the same for completely randomized designs. When a balanced randomization is used, equations (2.11) and (2.12) always yield the same solution, and equation (2.12) always provides the same solution no matter what value of correlation is used in the weight matrix. This equivalence can be shown with simple algebra.

Equation (2.11) can be written as

$$\begin{aligned} U(\beta) &= \sum_{k=1}^{K} \int_0^t (Z_{k1} - E(u)\ Z_{k2} - E(u)) \begin{pmatrix} d\hat{M}_{k1}(u) \\ d\hat{M}_{k2}(u) \end{pmatrix} \\ &= \sum_{k=1}^{K} \int_0^t (Z_{k1} - E(u))\, d\hat{M}_{k1}(u) + (Z_{k2} - E(u))\, d\hat{M}_{k2}(u) \\ &= \sum_{k=1}^{K} \int_0^t d\hat{M}_{k1}(u) - E(u) d\hat{M}_{k1}(u) - E(u) d\hat{M}_{k2}(u), \end{aligned}$$

when $(Z_{k1}, Z_{k2}) = (1, 0)$. Similarly, weighted estimating equation (2.12) can be written as

$$\begin{aligned} U'(\beta) &= \frac{1}{1-\rho^2} \sum_{k=1}^{K} \int_0^t (Z_{k1} - E(u)\ Z_{k2} - E(u)) \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} d\hat{M}_{k1}(u) \\ d\hat{M}_{k2}(u) \end{pmatrix} \\ &= \frac{1}{1-\rho^2} \sum_{k=1}^{K} \int_0^t (Z_{k1} - E(u))\, d\hat{M}_{k1}(u) + (Z_{k2} - E(u))\, d\hat{M}_{k2}(u) \\ &\quad -\rho\, (Z_{k1} - E(u))\, d\hat{M}_{k2}(u) - \rho\, (Z_{k2} - E(u))\, d\hat{M}_{k1}(u) \\ &= \frac{1}{1-\rho^2} \sum_{k=1}^{K} \int_0^t d\hat{M}_{k1}(u) - E(u) d\hat{M}_{k1}(u) - E(u) d\hat{M}_{k2}(u) \\ &\quad -\rho d\hat{M}_{k2}(u) + \rho E(u) d\hat{M}_{k1}(u) + \rho E(u) d\hat{M}_{k2}(u) \end{aligned}$$

At the solution to equation (2.11), $\hat{\beta}_u$,

$$U(\hat{\beta}_u) = \sum_{k=1}^{K} \int_0^t d\hat{M}_{k1}(u) - \hat{E}(u) d\hat{M}_{k1}(u) - \hat{E}(u) d\hat{M}_{k2}(u) = 0.$$

Thus

$$U'(\hat{\beta}_u) = -\frac{\rho}{1-\rho^2} \left\{ \sum_{k=1}^{K} d\hat{M}_{k2}(u) + \rho \hat{E}(u) d\hat{M}_{k1}(u) + \rho \hat{E}(u) d\hat{M}_{k2}(u) \right\}.$$

Since $\sum_{k=1}^{K}\int_0^t d\hat{M}_{k1}(u) + d\hat{M}_{k2}(u) = \int_0^t \sum_{k=1}^{K} d\hat{M}_{k1}(u) + d\hat{M}_{k2}(u) = 0$,

$$\begin{aligned}
U'(\hat{\beta}_u) &= -\frac{\rho}{1-\rho^2}\left\{\sum_{k=1}^{K}\int_0^t d\hat{M}_{k2}(u) + \hat{E}(u)d\hat{M}_{k1}(u) + \hat{E}(u)d\hat{M}_{k2}(u)\right\} \\
&= -\frac{\rho}{1-\rho^2}\left\{\sum_{k=1}^{K}\int_0^t d\hat{M}_{k2}(u) + \hat{E}(u)d\hat{M}_{k1}(u) + \hat{E}(u)d\hat{M}_{k2}(u)\right\} \\
&\quad -\frac{\rho}{1-\rho^2}\left\{\sum_{k=1}^{K}\int_0^t d\hat{M}_{k1}(u) - \hat{E}(u)d\hat{M}_{k1}(u) - \hat{E}(u)d\hat{M}_{k2}(u)\right\} \\
&= -\frac{\rho}{1-\rho^2}\left\{\sum_{k=1}^{K}\int_0^t d\hat{M}_{k1}(u) +)d\hat{M}_{k2}(u)\right\} \\
&= 0
\end{aligned}$$

Therefore, equations (2.11) and (2.12) yield the same solutions in the balanced randomization case, regardless the value of ρ used in the common weight matrix.

Table 2.6 shows simulation results for $\hat{\beta}_u$ when a balanced randomized design is used. There are 100 clusters of size 2 for each simulated data set, with $\beta = 0.693$ and $\theta = 0.25$. When the study is ended at time 5 time units, the censoring rate with combination of random and Type I censoring, P_5 changes from 0 to 10%. Relative to the simulation results for a completely randomized design shown in Table 2.3, these simulation results show that the empirical variance of $\hat{\beta}_u$ for a balanced randomized design is about 20% of the empirical variance of $\hat{\beta}_u$ for a completely randomized design. It is preferable to use balanced randomized designs when possible, and for those cases $\hat{\beta}_u$ is equivalent to $\hat{\beta}_c$ and $\hat{\beta}_w$ if exchangeable correlation is used.

P_5	estimated bias	empirical variance
0	0.0157	0.0058
10%	0.0123	0.0067

Table 2.6 Simulation results for $\hat{\beta}_u$ for 500 simulated samples of 100 clusters of size 2 with $\beta = 0.693$, $\theta = 0.25$, and a balanced randomized design.

2.6.3 Comparison of estimators from unweighted and weighted estimating equations with different censoring types

In this section, we compare the effect of the censoring type on the behavior of estimators. For a certain value of censoring rate, 500 estimates from 500 simulated data sets with the combination of random and Type I censoring described in Section 2.5.2 are produced. Another 500 estimates from 500 simulated data sets with only Type II censoring are also produced. The two sets of 500 estimates are compared with respect to estimated bias and empirical variance. $SRE(\hat{\beta}_u|\hat{\beta}_w)$ is the empirical variance of $\hat{\beta}_w$ divided by the empirical variance of $\hat{\beta}_u$ (not shown) from the same 500 simulated data sets. $SRE(\hat{\beta}_u|\hat{\beta}_w)$ larger than one indicates that $\hat{\beta}_w$ has gains in efficiency relative to $\hat{\beta}_u$. In the table, the true coefficient β assumes value 0.693, and the association parameter θ assumes values 0.25, 0.80 and 1.50, corresponding to correlations 0.937, 0.712, and 0.512. P_5 is the censoring rate with combination of random and Type I censoring given that the study is ended at 5 time units. P is the censoring rate with Type II censoring. P_5 or P is set at different rates changing from 10% to 50%. Both estimators have smaller variance as within cluster correlation is stronger or the censoring rate is lower. The biases of the two estimators are similar. The results show that $\hat{\beta}_w$ for data with Type II censoring only tends to have smaller variance and higher gains in efficiency relative to $\hat{\beta}_u$ than $\hat{\beta}_w$ for data with combination of random and Type I censoring. The distribution of censoring rate also has effects on the relative efficiency. Figure 2.4 shows the histogram of censoring rates of 500 data sets of size 2 when the combination of random and Type I censoring is used for $\beta = 0.693$, $\theta = 0.25$ and the designed censoring rate is 10%. The average of censoring rates from 500 simulated data sets is 9.96% with standard error of 0.022. The relative efficiency for data with combination of random and Type I censoring is affected by the variation of the censoring rates.

		$\hat{\beta}_w$ with combination of random and Type I censoring			$\hat{\beta}_w$ with Type II censoring only		
P_5 or P	θ	estimated bias	empirical variance	$SRE(\hat{\beta}_u \mid \hat{\beta}_w)$	estimated bias	empirical variance	$SRE(\hat{\beta}_u \mid \hat{\beta}_w)$
10%	0.25	0.0083	0.0148	1.761	0.0119	0.0118	2.615
	0.80	0.0060	0.0215	1.214	0.0069	0.0192	1.469
	1.50	0.0070	0.0243	1.069	0.0068	0.0200	1.197
20%	0.25	0.0090	0.0177	1.659	0.0118	0.0144	2.302
	0.80	0.0027	0.0247	1.189	0.0035	0.0204	1.404
	1.50	0.0024	0.0285	1.041	0.0043	0.0240	1.143
50%	0.25	0.0016	0.0359	1.254	0.0000	0.0292	1.546
	0.80	-0.0044	0.0423	1.048	-0.0062	0.0410	1.110
	1.50	0.0044	0.0439	1.027	0.0056	0.0430	1.026

Table 2.7 Simulation results for $\hat{\beta}_w$ with combination of random and Type I censoring and $\hat{\beta}_w$ with Type II censoring only for 500 simulated samples of 100 clusters of size 2 with $\beta = 0.693$.

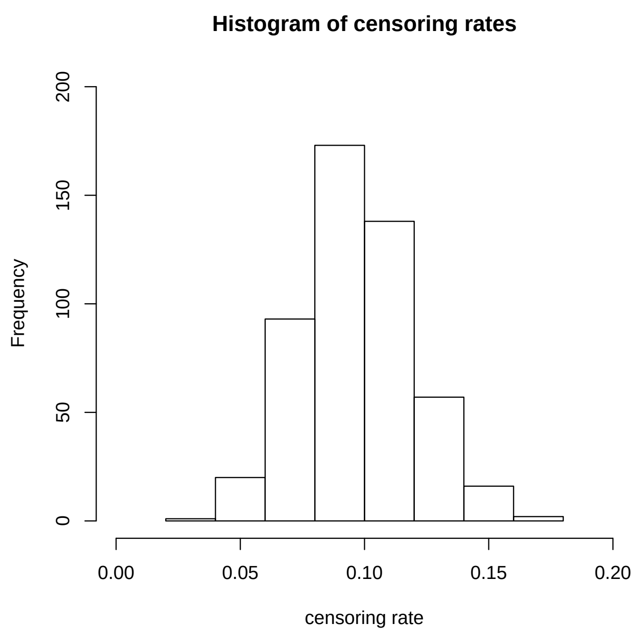

Figure 2.4 Histogram of censoring rates when combination of random and Type I censoring is used for 500 simulated samples of 100 clusters of size 2 with $\beta = 0.693$, designed censoring rate $P_5 = 10\%$, and $\theta = 0.25$ corresponding to correlation 0.937.

2.6.4 Comparison of estimators from unweighted and weighted estimating equations under a completely randomized design

In this section, stimulation studies are used to assess the behavior of the estimators from unweighted and weighted estimating equations. From the results in Section 2.6.1, we know that using correlations of martingales conditional on the covariate values defined in (2.7) does not improve the efficiency of the estimators much. In addition, in practical use there is rarely a case that correlations are considered depending on the covariate values. Therefore, all the simulation results shown here use the common correlation defined in (2.8) in the weight matrix. Only a completely randomized design is considered. Since $\hat{\beta}_c$ and $\hat{\beta}_w$ are the same in this case, only $\hat{\beta}_w$ is shown. The estimators are evaluated with respect to estimated bias and empirical variances.

The combinations of parameters used in the simulation studies in this section are as follows:

parameters	function	possible values
β	true coefficient	0.693, 0, -0.5
P_5	censoring rate (%)	0, 10, 20, 50
θ	association parameter	0.25, 0.8, 1.5
K	number of clusters	100
n	cluster size	2
N	number of simulated data sets	500

Tables 2.8, 2.9, 2.10 and 2.11 display simulation results. There are 500 datasets generated in each simulation study, thus 500 estimates were produced. In Table 2.8, the estimated bias is the difference of the average of the 500 estimates and the true parameter value β. Empirical variances are calculated using a sample of 500 estimates. $SRE(\hat{\beta}_u|\hat{\beta}_w)$ is the empirical variance of $\hat{\beta}_u$ divided by empirical variance of $\hat{\beta}_w$. A value of $SRE(\hat{\beta}_u|\hat{\beta}_w)$ larger than one indicates that the estimator produced by equations (2.12) has smaller variance than the IWM estimator. Then we say that there is a gain in efficiency of $\hat{\beta}_w$ relative to $\hat{\beta}_u$. In the table, the true coefficient β assumes values

0.693, 0 and -0.5, and the association parameter θ assumes values 0.25, 0.80 and 1.50, corresponding to correlations 0.937, 0.712, and 0.512. P_5 is the censoring rate with combination of random and Type I censoring given that the study is ended at 5 time units. P_5 is set at different rates changing from 0 to 50%.

For all the parameter settings, the estimated biases of $\hat{\beta}_u$ and $\hat{\beta}_w$ are both close to 0. There is no obvious advantage of using one method over the other to reduce the bias. The gains in efficiency provided by $\hat{\beta}_w$ are greater for stronger within cluster correlation, lower level of censoring and parameter values closer to 0. When the censoring rate is high, 50%, gains in efficiency are less than 4 percent for within cluster correlation of 0.512, and less than 42 percent for within cluster correlation of 0.937. When there is no censoring, gains in efficiency are as high at 6.961. At each level of censoring and for each parameter value, the efficiency gains decrease as the correlation level decreases, with the most prounced patterns when there is no censoring and when the true parameter is zero.

The values of regression coefficients 0.693 or -0.5 are larger values than are often seen in medical research. We would expect gains in efficiency closer to the relative gains shown in Table 2.8 for $\beta = 0$ when within cluster correlation is strong and the censoring rate is moderate. Proper weighting could double the efficiency of estimators for regression coefficients when the Cox model is applied to clustered survival data.

β	P_5	θ	$\hat{\beta}_u$ estimated bias	$\hat{\beta}_u$ empirical variance	$\hat{\beta}_w$ estimated bias	$\hat{\beta}_w$ empirical variance	$SRE(\hat{\beta}_u \mid \hat{\beta}_w)$
0.693	0	0.25	0.0187	0.0290	0.0210	0.0107	2.707
		0.80	0.0154	0.0267	0.0178	0.0165	1.613
		1.50	0.0166	0.0259	0.0151	0.0209	1.239
	10%	0.25	0.0165	0.0262	0.0083	0.0148	1.761
		0.80	0.0209	0.0264	0.0060	0.0215	1.214
		1.50	0.0082	0.0260	0.0070	0.0243	1.069
	20%	0.25	0.0115	0.0294	0.0090	0.0177	1.659
		0.80	0.0070	0.0294	0.0027	0.0247	1.189
		1.50	-0.0110	0.0297	0.0024	0.0285	1.041
	50%	0.25	0.0314	0.0451	0.0106	0.0359	1.254
		0.80	-0.0046	0.0444	0.0044	0.0423	1.048
		1.50	0.0041	0.0451	0.0044	0.0439	1.027
0	0	0.25	-0.0102	0.0233	0.0008	0.0033	6.961
		0.80	-0.0140	0.0223	0.0021	0.0112	1.983
		1.50	0.0035	0.0223	0.0027	0.0168	1.327
	10%	0.25	-0.0002	0.0262	-0.0110	0.0082	2.705
		0.80	0.0019	0.0264	-0.0076	0.0162	1.417
		1.50	0.0023	0.0260	-0.0013	0.0206	1.138
	20%	0.25	-0.0020	0.0254	-0.0077	0.0116	2.187
		0.80	0.0092	0.0260	-0.0105	0.0199	1.307
		1.50	0.0128	0.0272	-0.0074	0.0247	1.101
	50%	0.25	-0.0104	0.0407	-0.0058	0.0287	1.419
		0.80	-0.0037	0.0413	-0.0063	0.0372	1.108
		1.50	0.0029	0.0420	-0.0061	0.0404	1.037
-0.5	0	0.25	-0.0132	0.0235	-0.0144	0.0059	3.925
		0.80	-0.0188	0.0227	-0.0071	0.0126	1.808
		1.50	0.0002	0.0227	-0.0050	0.0177	1.280
	10%	0.25	-0.0015	0.0290	-0.0024	0.0173	1.671
		0.80	-0.0090	0.0282	-0.0154	0.0241	1.171
		1.50	0.0092	0.0285	-0.0064	0.0273	1.045
	20%	0.25	-0.0016	0.0283	0.0046	0.0184	1.536
		0.80	-0.0157	0.0284	0.0408	0.0250	1.134
		1.50	-0.0052	0.0291	0.0062	0.0303	0.960
	50%	0.25	-0.0242	0.0429	-0.0158	0.0319	1.342
		0.80	-0.0126	0.0448	-0.0115	0.0407	1.100
		1.50	-0.0159	0.0449	0.0062	0.0467	0.959

Table 2.8 Simulation results for $\hat{\beta}_u$ and $\hat{\beta}_w$ for 500 simulated samples with 100 clusters of size 2. $\theta = 0.25$ corresponding to correlation 0.913, $\theta = 0.25$ corresponding to correlation 0.712, and $\theta = 1.5$ corresponding to correlation 0.512.

Empirical variances are the same as those in Tables 2.8 and 2.9. The column labeled average robust variance contains the average of 500 robust variance estimates for the IWM estimator. The variance ratio for $\hat{\beta}_u$ is the average robust estimate of variance of $\hat{\beta}_u$ divided by the empirical variance of $\hat{\beta}_u$. A value of the variance ratio of $\hat{\beta}_u$ greater than one indicates that the robust variance overestimates the true variance, and a value of the variance ratio of $\hat{\beta}_u$ less than one indicates that the robust variance underestimates the true variance. The variance of $\hat{\beta}_w$ is obtained by a bootstrap resampling method. For clustered data, a bootstrap method resamples clusters instead of individual subjects. For each of the 500 simulated data sets, a set of 100 bootstrap samples was taken to produce a bootstrap variance estimate for $\hat{\beta}_w$. The average bootstrap variance is the average of the 500 bootstrap variances of $\hat{\beta}_w$ obtained from the 500 simulated datasets. The variance ratio for $\hat{\beta}_w$ is the average bootstrap variance of $\hat{\beta}_w$ divided by the empirical variance $\hat{\beta}_w$. A value of the variance ratio for $\hat{\beta}_w$ greater than one indicates that the bootstrap variance overestimates the variance, and a value of the variance ratio for $\hat{\beta}_w$ less than 1 indicates that the bootstrap variance underestimates the variance. From the results provided in Table 2.9, all values of the variance ratios for $\hat{\beta}_w$ are close to one. The robust variance estimator for $\hat{\beta}_u$ tends to underestimate the variance of $\hat{\beta}_u$, with larger biases for lower censoring rates. Overall the bootstrap method provides a good estimate for the small sample variance of $\hat{\beta}_w$.

β	P_5	θ	$\hat{\beta}_u$ empirical variance	$\hat{\beta}_u$ average robust variance	$\hat{\beta}_u$ variance ratio	$\hat{\beta}_w$ empirical variance	$\hat{\beta}_w$ average bootstrap variance	$\hat{\beta}_w$ variance ratio
0.693	0	0.25	0.0290	0.0231	0.796	0.0107	0.0112	1.053
		0.80	0.0267	0.0228	0.853	0.0165	0.0166	1.008
		1.50	0.0259	0.0224	0.864	0.0209	0.0204	0.977
	10%	0.25	0.0262	0.0256	0.977	0.0148	0.0143	0.968
		0.80	0.0264	0.0250	0.946	0.0215	0.0204	0.950
		1.50	0.0260	0.0248	0.953	0.0243	0.0233	0.961
	20%	0.25	0.0294	0.0287	0.976	0.0177	0.0187	1.062
		0.80	0.0294	0.0280	0.952	0.0247	0.0249	1.011
		1.50	0.0297	0.0278	0.936	0.0285	0.0282	0.991
	50%	0.25	0.0451	0.0454	1.006	0.0359	0.0390	1.087
		0.80	0.0444	0.0444	1.000	0.0423	0.0443	1.048
		1.50	0.0451	0.0443	0.982	0.0439	0.0459	1.046
0	0	0.25	0.0233	0.0197	0.845	0.0033	0.0045	1.366
		0.80	0.0223	0.0206	0.923	0.0112	0.0125	1.124
		1.50	0.0223	0.0201	0.901	0.0168	0.0171	1.019
	10%	0.25	0.0262	0.0256	0.977	0.0082	0.0082	1.001
		0.80	0.0264	0.0250	0.946	0.0162	0.0158	0.980
		1.50	0.0260	0.0248	0.953	0.0206	0.0199	0.967
	20%	0.25	0.0254	0.0251	0.988	0.0116	0.0124	1.072
		0.80	0.0260	0.0252	0.969	0.0199	0.0199	1.005
		1.50	0.0272	0.0252	0.926	0.0247	0.0224	0.955
	50%	0.25	0.0407	0.0399	0.980	0.0287	0.0301	1.052
		0.80	0.0413	0.0401	0.970	0.0372	0.0382	1.027
		1.50	0.0420	0.0412	0.980	0.0404	0.0412	1.022
-0.5	0	0.25	0.0235	0.0214	0.910	0.0059	0.0070	1.197
		0.80	0.0227	0.0212	0.933	0.0126	0.0140	1.112
		1.50	0.0227	0.0212	0.933	0.0177	0.0183	1.034
	10%	0.25	0.0290	0.0289	0.996	0.0173	0.0176	1.022
		0.80	0.0282	0.0286	1.014	0.0241	0.0244	1.013
		1.50	0.0285	0.0285	1.000	0.0273	0.0274	1.004
	20%	0.25	0.0283	0.0271	0.957	0.0184	0.0181	0.986
		0.80	0.0284	0.0268	0.943	0.0250	0.0219	0.938
		1.50	0.0291	0.0268	0.920	0.0303	0.0241	0.924
	50%	0.25	0.0429	0.0430	1.002	0.0319	0.0335	1.053
		0.80	0.0448	0.0427	0.953	0.0407	0.0413	1.016
		1.50	0.0449	0.0437	0.973	0.0467	0.0429	0.920

Table 2.9 Simulation results for variances estimates for 500 simulated samples with 100 clusters of size 2. The number of bootstrap samples is 100 for each of 500 simulated sets. The variance ratio for $\hat{\beta}_u$ is the average robust variance estimate divided by the empirical variance, and the variance ratio for $\hat{\beta}_w$ is the average bootstrap variance divided by the empirical variance.

Tables 2.10 and 2.11 show coverage rates and average lengths of nominal 95% confidence intervals of the regression coefficients. The confidence intervals based on $\hat{\beta}_u$ are contructed as $\hat{\beta}_u \pm (1.96)S_{robust}$, where S_{robust} is the standard error of $\hat{\beta}_u$ obtained from the robust covariance procedure. Four bootstrap methods were used to construct bootstrap confidence intervals using $\hat{\beta}_w$, normal, quantile, basic and accelerated bias-corrected (BC_α). These methods are described by Davison and Hinkley (1997). The normal method is evaluated as $\hat{\beta}_w \pm (1.96)S_{boot}$, where S_{boot} is the bootstrapped standard error for $\hat{\beta}_w$. For b bootstrap samples, denote the bootstrap distribution of bootstrap estimates by $\hat{F}_b(x)$. The quantile confident interval method uses the $\alpha/2$ and $1-\alpha/2$ quantiles of $\hat{F}_b(x)$ as the endpoints of the $1-\alpha$ level confidence interval. The basic confident interval method uses the upper quantile of the bootstrap distribution to calculate the lower confidence bound, and the lower quantile to calculate the upper confidence bound. The endpoints using this method are $2\hat{\beta}_g - \hat{F}_b^{-1}(1-\alpha/2)$, and $2\hat{\beta}_g - \hat{F}_b^{-1}(\alpha/2)$. The accelerated bias-corrected method obtains end points by inverting percentiles of the bootstrap distribution after adjusting for bias and acceleration shift. Denote the acceleration constant by a, and the normal cdf by $\Phi(x)$. Define $Z_0 = \Phi^{-1}(\hat{F}_b(\hat{\beta}_w))$, and define $Z_{\alpha/2}$ as the $\alpha/2$-percentile of a standard normal distribution. Then, the α endpoints of the BC_a confidence interval are $\hat{F}_b^{-1}\left\{\Phi\left(Z_0 + \frac{Z_0 - Z_\alpha/2}{1-a(Z_0 - Z_{\alpha/2})}\right\}\right)$ and $\hat{F}_b^{-1}\left\{\Phi\left(Z_0 + \frac{Z_0 + Z_\alpha/2}{1-a(Z_0 + Z_{\alpha/2})}\right\}\right)$. Among all the methods, the normal method is the simplest. From the result in Table 2.10, all the methods provide coverage rates 95 percent. The simple normal method performs as well as the other bootstrap methods. From the results in Table 2.11, when the censoring rate is moderate, and the correlation is strong, the lengths of bootstrap confidence intervals based on $\hat{\beta}_w$ tend to be narrower than the lengths of the confidence intervals based on $\hat{\beta}_u$. Comparing all the bootstrap methods based on $\hat{\beta}_w$, the normal method tends to provide the most narrow confidence intervals.

β	P_5	θ	average coverage rate of confidence interval based on robust variance estimator for $\hat{\beta}_u$	average coverage rate of bootstrap confidence interval based on $\hat{\beta}_w$			
				normal	quantile	basic	BCa
0.693	0	0.25	0.920	0.946	0.928	0.938	0.916
		0.80	0.934	0.940	0.940	0.946	0.928
		1.50	0.934	0.940	0.940	0.959	0.932
	10%	0.25	0.948	0.934	0.930	0.946	0.920
		0.80	0.948	0.932	0.940	0.938	0.924
		1.50	0.956	0.936	0.938	0.948	0.932
	20%	0.25	0.940	0.946	0.944	0.954	0.928
		0.80	0.952	0.940	0.954	0.946	0.938
		1.50	0.954	0.936	0.956	0.950	0.946
	50%	0.25	0.948	0.962	0.948	0.966	0.954
		0.80	0.950	0.960	0.950	0.970	0.946
		1.50	0.954	0.954	0.954	0.960	0.946
0	0	0.25	0.932	0.982	0.960	0.986	0.974
		0.80	0.940	0.962	0.946	0.968	0.948
		1.50	0.928	0.946	0.936	0.954	0.934
	10%	0.25	0.948	0.954	0.952	0.962	0.952
		0.80	0.948	0.954	0.950	0.960	0.936
		1.50	0.956	0.942	0.946	0.962	0.938
	20%	0.25	0.954	0.966	0.962	0.964	0.958
		0.80	0.956	0.966	0.950	0.968	0.950
		1.50	0.948	0.960	0.950	0.966	0.952
	50%	0.25	0.942	0.958	0.958	0.968	0.948
		0.80	0.944	0.960	0.944	0.964	0.952
		1.50	0.948	0.950	0.942	0.968	0.944
-0.5	0	0.25	0.938	0.952	0.952	0.954	0.940
		0.80	0.942	0.952	0.950	0.962	0.944
		1.50	0.948	0.934	0.950	0.940	0.948
	10%	0.25	0.962	0.950	0.944	0.966	0.934
		0.80	0.944	0.962	0.946	0.962	0.948
		1.50	0.946	0.954	0.948	0.964	0.962
	20%	0.25	0.944	0.950	0.942	0.970	0.948
		0.80	0.946	0.946	0.936	0.968	0.942
		1.50	0.950	0.952	0.942	0.960	0.942
	50%	0.25	0.948	0.964	0.964	0.972	0.946
		0.80	0.928	0.948	0.940	0.958	0.948
		1.50	0.934	0.952	0.942	0.960	0.942

Table 2.10 Simulated coverage rates of nominal 95% confidence intervals for 500 simulated samples with 100 clusters of size 2. Each bootstrap confidence interval was obtained from 100 bootstrap samples.

β	P_5	θ	average length of confidence interval based on robust variance estimator for $\hat{\beta}_u$	average length of bootstrap confidence interval based on $\hat{\beta}_w$			
				normal	quantile	basic	BCa
0.693	0	0.25	0.595	0.414	0.419	0.419	0.408
		0.80	0.591	0.505	0.529	0.529	0.519
		1.50	0.586	0.559	0.587	0.587	0.583
	10%	0.25	0.627	0.468	0.492	0.492	0.480
		0.80	0.619	0.559	0.601	0.601	0.594
		1.50	0.617	0.598	0.632	0.632	0.620
	20%	0.25	0.664	0.536	0.553	0.553	0.538
		0.80	0.656	0.618	0.645	0.645	0.627
		1.50	0.653	0.658	0.687	0.687	0.681
	50%	0.25	0.835	0.774	0.785	0.785	0.770
		0.80	0.826	0.825	0.856	0.856	0.854
		1.50	0.825	0.839	0.862	0.862	0.858
0	0	0.25	0.550	0.262	0.275	0.275	0.274
		0.80	0.562	0.438	0.457	0.457	0.457
		1.50	0.555	0.509	0.532	0.532	0.533
	10%	0.25	0.627	0.350	0.367	0.367	0.367
		0.80	0.620	0.491	0.512	0.512	0.515
		1.50	0.617	0.550	0.572	0.572	0.575
	20%	0.25	0.621	0.432	0.453	0.453	0.455
		0.80	0.622	0.551	0.577	0.577	0.579
		1.50	0.622	0.599	0.624	0.624	0.626
	50%	0.25	0.783	0.675	0.708	0.708	0.706
		0.80	0.785	0.763	0.798	0.796	0.800
		1.50	0.795	0.793	0.829	0.829	0.832
-0.5	0	0.25	0.573	0.327	0.338	0.338	0.336
		0.80	0.570	0.463	0.484	0.484	0.480
		1.50	0.570	0.526	0.550	0.550	0.548
	10%	0.25	0.666	0.516	0.543	0.543	0.541
		0.80	0.662	0.608	0.634	0.634	0.631
		1.50	0.661	0.646	0.674	0.674	0.674
	20%	0.25	0.645	0.527	0.561	0.561	0.559
		0.80	0.642	0.570	0.596	0.596	0.593
		1.50	0.642	0.617	0.642	0.642	0.644
	50%	0.25	0.812	0.717	0.751	0.751	0.742
		0.80	0.810	0.796	0.813	0.813	0.809
		1.50	0.819	0.812	0.833	0.833	0.821

Table 2.11 Simulated lengths of confidence intervals for 500 simulated samples with 100 clusters of size 2. Each bootstrap confidence interval was obtained from 100 bootstrap samples.

2.7 Summary and discussion

We studied cases where the cluster size is two, and there is only one binary covariate. It was shown that weighted estimating equations provide no gain in efficiency if a balanced randomized design is used. When the treatment assignment is unbalanced within clusters, the estimators from partial likelihood score equations initially assuming independence within clusters lose efficiency. Using weighted estimating equations can greatly improve efficiency.

In the case we studied, if all pairs are assigned the same weight matrix, the weighted estimating equations (2.12) we proposed and the modified Cai and Prentice estimating equations (2.5) result in the identical regression estimators. If weights depending on the covariates in the pair are used, estimators from these two weighted estimating equations are only slightly different. Using weights depending on the combination of covariate pairs requires replication of covariate pairs. When there are multiple dichotomous covariates or multiple level for categorical covariates, the sample sizes available to estimate same correlations might be quite small. Also it is not straightforward to extend the model with correlations that dependent on covariate values to continuous covariates. We simulate a case where survival times are correlated with different dependence levels depending on the covariate pairs. The estimators do not lose efficiency by using correlation estimates evaluated by averaging across covariate values. In addition, examples of considering weights depending on the covariate values are rare in practical use. Thus the common weight approach was used in the rest of the simulation studies. For situations we considered, $\hat{\beta}_c$ and $\hat{\beta}_w$ are equivalent.

Table 2.7 shows the effect of censoring type on the behavior of the estimator. It appeared that the estimators from weighted estimating equations for data with only Type II censoring tend to have higher gains in efficiency than for data with combination of random and Type I censoring. But it is also caused by the difference in distributions of censoring rates using these two types of censoring in the simulation study.

The simulation results in Table 2.8 show if the design is completely randomized, using weighted estimating equation can greatly improve the efficiency of the regression coefficient estimates in the proportional hazards model. Gains in efficiency for $\hat{\beta}_w$ relative to $\hat{\beta}_u$ are largest when within cluster correlation between failure times is strong and the censoring rate is low. With moderate correlation or heavy censoring, the gain becomes smaller. Gains in efficiency are greater for parameter values closer to zero. Even for the regression parameter values 0.693 or -0.5 that larger than often seen in medical studies, gains in efficiency are observed with a high within cluster correlation and a low censoring rate. The results in Table 2.9 show that bootstrap variances tend to be close to empirical variance. All four bootstrap methods for constructing confidence intervals provided coverage rates close to the nominal 95% levels. When within cluster correlation is strong and the censoring rate is moderate, the lengths of 95% confidence intervals based on $\hat{\beta}_w$ are smaller than those based on $\hat{\beta}_u$. The normal method gives the shortest confidence intervals among the four bootstrap confidence interval procedures we studied.

CHAPTER 3. Generalized estimating equations

3.1 Introduction

In this chapter, a new set of estimating equations for clustered survival data is developed using a generalized estimating equation approach. Wedderburn (1974) proposed quasi-likelihood estimation. McCullagh and Nelder (1983) extended quasi-likelihood estimation using generalized linear models to handle a variety of discrete and continuous variables. Liang and Zeger (1986) applied a quasi-likelihood approach to derive generalized estimating equations for analyzing longitudinal data. To apply this approach, it is assumed that responses are sampled from an exponential family distribution, and the variance of each response is a known function of its expectation. Consistent estimators of regression parameters are obtained without specifying the joint distribution of dependent responses and correct correlation structure. To apply a generalized estimating equation approach to clustered survival data, we consider partial likelihood equations for the Cox model in the counting process context. In that form, counting process differentials can be approximated by Poisson random variables, which have mean and variance of corresponding compensator differentials. Therefore, an alternative set of estimating equations can be derived using a generalized estimating equation approach to estimate regression parameters in the Cox model for clustered survival data. Bootstrap resampling methods are used to estimate the variances of coefficients in the Cox model and construct confidence intervals. Simulation studies are provided to compare estimators obtained from partial likelihood score equations that ignore within cluster dependence to estimators obtained from this new set of estimating equations.

3.2 Generalized estimating equations

3.2.1 Derivation of estimating equations

Assume the marginal density of the event time distribution for the jth observation from the kth cluster, T_{kj}, is of the exponential form

$$\begin{aligned} f(t_{kj}) &= \exp[\{t_{kj}h(\boldsymbol{\beta}'\mathbf{Z}_{kj}) - a(h(\boldsymbol{\beta}'\mathbf{Z}_{kj})) + b(t_{kj})\}\phi] \\ &= \exp[\{t_{kj}\theta_{kj} - a(\theta_{kj}) + b(t_{kj})\}\phi], \end{aligned} \tag{3.1}$$

where $\theta_{kj} = h(\boldsymbol{\beta}'\mathbf{Z}_{kj})$, $k = 1, ..., K$, $j = 1, ..., n_k$, $\boldsymbol{\beta}$ is a $p \times 1$ coefficient vector, and ϕ is a scaler parameter. Then the first two moments for T_{kj} are

$$\begin{aligned} E(T_{kj}) &= a'(\theta_{kj}) = a'(h(\boldsymbol{\beta}'\mathbf{Z}_{kj})) \\ Var(T_{kj}) &= a''(\theta_{kj})/\phi = a''(h(\boldsymbol{\beta}'\mathbf{Z}_{kj}))/\phi \end{aligned}$$

Define $\eta_{kj} = \boldsymbol{\beta}'\mathbf{Z}_{kj}$. If the subjects within a cluster response independently, the regression parameters $\boldsymbol{\beta}$ can be estimated by the estimating equations as

$$\mathbf{U}_I(\boldsymbol{\beta}) = \sum_{k=1}^{K} \mathbf{Z}_k^T \Delta_k (\mathbf{T}_k - E(\mathbf{T}_k)) = 0, \tag{3.2}$$

where

$$\Delta_k = diag\left(\frac{d\theta_{kj}}{d\eta_{kj}}\right) = diag\left(h'(\boldsymbol{\beta}'\mathbf{Z}_{kj})\right)$$

is a $n_k \times n_k$ matrix. Define $\mathbf{Z}_k$ is a $n_k \times p$ covariate matrix, and define $\mathbf{T}_k$ is a $n_k \times 1$ response vector.

When the subjects within a cluster do not response independently, the generalized estimating equations (GEEs) can be used to estimate regression parameters. They are of the form

$$\mathbf{U}_G(\beta) = \sum_{k=1}^{K} \mathbf{D}_k^T \mathbf{V}_k^{-1} (\mathbf{T}_k - E(\mathbf{T}_k)) = 0, \tag{3.3}$$

where $\mathbf{D}_k = d\{a'_k(\theta)\}/d\beta$, and the jth row of $\mathbf{D}_k$ corresponds to $\mathbf{D}_{kj} = dE(T_{kj})/d\beta = a''(\theta_{kj})h'(\beta'\mathbf{Z}_{kj})\mathbf{Z}_{kj}$. Then

$$\mathbf{D}_k = \begin{pmatrix} \mathbf{D}_{k1}^T \\ \mathbf{D}_{k2}^T \\ \vdots \\ \mathbf{D}_{kn_k}^T \end{pmatrix} = \begin{pmatrix} a''(\theta_{k1})h'(\beta'\mathbf{Z}_{k1})\mathbf{Z}_{k1}^T \\ a''(\theta_{k2})h'(\beta'\mathbf{Z}_{k2})\mathbf{Z}_{k2}^T \\ \vdots \\ a''(\theta_{kn_k})h'(\beta'\mathbf{Z}_{kn_k})\mathbf{Z}_{kn_k}^T \end{pmatrix}$$

$$= \begin{pmatrix} a''(\theta_{k1})h'(\beta'\mathbf{Z}_{k1})Z_{k1,1} & a''(\theta_{k1})h'(\beta'\mathbf{Z}_{k1})Z_{k1,2} & \cdots & a''(\theta_{k1})h'(\beta'\mathbf{Z}_{k1})Z_{k1,p} \\ a''(\theta_{k2})h'(\beta'\mathbf{Z}_{k2})Z_{k2,1} & a''(\theta_{k2})h'(\beta'\mathbf{Z}_{k2})Z_{k2,2} & \cdots & a''(\theta_{k2})h'(\beta'\mathbf{Z}_{k2})Z_{k2,p} \\ \vdots & \vdots & \vdots & \vdots \\ a''(\theta_{kn_k})h'(\beta'\mathbf{Z}_{kn_k})Z_{kn_k,1} & a''(\theta_{kn_k})h'(\beta'\mathbf{Z}_{kn_k})Z_{kn_k,2} & \cdots & a''(\theta_{kn_k})h'(\beta'\mathbf{Z}_{kn_k})Z_{kn_k,p} \end{pmatrix}$$

$$= \begin{pmatrix} a''(\theta_{k1}) & 0 & \cdots & 0 \\ 0 & a''(\theta_{k2}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a''(\theta_{kn_k}) \end{pmatrix} \begin{pmatrix} h'(\theta_{k1}) & 0 & \cdots & 0 \\ 0 & h'(\theta_{k2}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & h'(\theta_{kn_k}) \end{pmatrix} \begin{pmatrix} \mathbf{Z}_{k1}^T \\ \mathbf{Z}_{k2}^T \\ \vdots \\ \mathbf{Z}_{kn_k}^T \end{pmatrix}$$

Therefore, $\mathbf{D}_k = \mathbf{B}_k \mathbf{\Delta}_k \mathbf{Z}_k$ is a $n_k \times p$ matrix, where $B_k = diag(a''(\theta_{kj}))$.

Define $\mathbf{V}_k = \mathbf{B}_k^{\frac{1}{2}} \mathbf{R}_k(\boldsymbol{\rho}) \mathbf{B}_k^{\frac{1}{2}}/\phi$, where $\mathbf{R}_k(\boldsymbol{\rho})$ is the $n_k \times n_k$ correlation matrix of $\mathbf{T}_k$, and $\boldsymbol{\rho}$ is a vector of unknown parameters that defines the correlation matrix. $\mathbf{R}_k(\boldsymbol{\rho})$ is called a "working" correlation matrix because consistent estimators can be obtained even when $\mathbf{R}_k(\boldsymbol{\rho})$ is not correctly specified. $\mathbf{V}_k$ is a function of $\boldsymbol{\beta}$, ϕ and $\boldsymbol{\rho}$, where $\boldsymbol{\beta}$ is the parameter of interest, and $\boldsymbol{\rho}$ and ϕ are nuisance parameters. Then equation (3.3) becomes

$$\begin{aligned} \mathbf{U}_G(\beta) &= \sum_{k=1}^{K} \mathbf{D}_k^T \mathbf{V}_k^{-1} (\mathbf{T}_k - E(\mathbf{T}_k)) \\ &= \sum_{k=1}^{K} (\mathbf{B}_k \Delta_k \mathbf{Z}_k)^T \left[\frac{\mathbf{B}_k^{\frac{1}{2}} \mathbf{R}(\rho) \mathbf{B}_k^{\frac{1}{2}}}{\phi} \right]^{-1} (\mathbf{T}_k - E(\mathbf{T}_k)). \end{aligned} \tag{3.4}$$

Thus

$$
\begin{aligned}
&\mathbf{U}_G(\beta)\\
&= \sum_{i=k}^{K} \mathbf{D}_k^T \mathbf{V}_k^{-1} (\mathbf{T}_k - E(\mathbf{T}_k))\\
&= \sum_{i=k}^{K} \left[\begin{pmatrix} a''(\theta_{k1}) & 0 & \cdots & 0 \\ 0 & a''(\theta_{k2}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a''(\theta_{kn_k}) \end{pmatrix} \begin{pmatrix} h'(\theta_{k1}) & 0 & \cdots & 0 \\ 0 & h'(\theta_{k2}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & h'(\theta_{kn_k}) \end{pmatrix} \begin{pmatrix} \mathbf{Z}_{k1}^T \\ \mathbf{Z}_{k2}^T \\ \vdots \\ \mathbf{Z}_{kn_k}^T \end{pmatrix} \right]^T
\end{aligned}
$$

$$
\begin{pmatrix} \sqrt{a''(\theta_{k1})} & 0 & \cdots & 0 \\ 0 & \sqrt{a''(\theta_{k2})} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{a''(\theta_{kn_k})} \end{pmatrix}^{-1} \begin{pmatrix} r_{11}/\phi & r_{12}/\phi & \cdots & r_{1n_k}/\phi \\ r_{21}/\phi & r_{21}/\phi & \cdots & r_{2n_k}/\phi \\ \vdots & \vdots & \vdots & \vdots \\ r_{n_k1}/\phi & r_{n_k2}/\phi & \cdots & r_{n_kn_k}/\phi \end{pmatrix}^{-1}
$$

$$
\begin{pmatrix} \sqrt{a''(\theta_{k1})} & 0 & \cdots & 0 \\ 0 & \sqrt{a''(\theta_{k2})} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{a''(\theta_{kn_k})} \end{pmatrix}^{-1} \begin{pmatrix} T_{k1} - E(T_{k1}) \\ T_{k2} - E(T_{k2}) \\ \vdots \\ T_{kn_k} - E(T_{kn_k}) \end{pmatrix}
$$

When responses within clusters are all independent,

$$
\mathbf{R}_k(\boldsymbol{\rho}) \quad = \quad = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}
$$

Then equation (3.3) is equivalent to equation (3.2).

3.2.2 Solving the equations

The following steps can be used to solve the estimating equations.

Step 1 Set up the generalized estimating equation

$$\sum_{k=1}^{K} U_{Gk}\left[\boldsymbol{\beta}, \hat{\rho}(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta}))\right] = 0, \text{ where } U_{Gk} = \mathbf{D}_k^T \mathbf{V}_k^{-1}(\mathbf{T}_k - E(\mathbf{T}_k)).$$

Step 2 Update the estimate of $\boldsymbol{\beta}$ using current values of $\hat{\boldsymbol{\beta}}$, $\hat{\rho}$ and $\hat{\phi}$ with

$$\hat{\boldsymbol{\beta}}_{m+1} = \hat{\boldsymbol{\beta}}_m - \left[\sum_{k=1}^{K} \mathbf{D}'_k(\hat{\boldsymbol{\beta}}_m)\tilde{\mathbf{V}}_k^{-1}(\hat{\boldsymbol{\beta}}_m)\mathbf{D}_k(\hat{\boldsymbol{\beta}}_m)\right]^{-1} \left[\sum_{k=1}^{K} \mathbf{D}'_k(\hat{\boldsymbol{\beta}}_m)\tilde{\mathbf{V}}_k^{-1}(\hat{\boldsymbol{\beta}}_m)\mathbf{S}_k(\hat{\boldsymbol{\beta}}_m)\right],$$

where $\mathbf{S}_k = \mathbf{T}_k - E(\mathbf{T}_k)$.

Step 3 Given the current value of $\hat{\boldsymbol{\beta}}$, Pearson residuals are evaluated as

$$\hat{r}_{kj} = \left[\frac{t_{kj} - a'(\hat{\theta}_{kj})}{a''(\hat{\theta}_{kj})}\right]^{1/2},$$

where $\hat{\theta}_{kj}$ depends on the current value of $\hat{\boldsymbol{\beta}}$. The estimate of ϕ can be updated by $\hat{\phi}^{-1} = \sum_{k=1}^{K}\sum_{j=1}^{n_k} \frac{\hat{r}_{kj}^2}{N-p}$, where $N = \sum n_k$. Estimation of $\boldsymbol{\rho}$ depends upon the choice of the correlation structure $\mathbf{R}_k(\boldsymbol{\rho})$. The simplest case is that correlations are the same for any pair of observations. Then ρ becomes a scalar.

Step 4 Repeat the above steps, until the iterations converge.

3.2.3 Properties generalized estimating equation estimators

Under mild regularity conditions that the second order derivative of left hand side of equation (3.3) is a smooth function and the third order derivative exists, and that

1. $\hat{\rho}$ is $K^{\frac{1}{2}}$-consistent given $\boldsymbol{\beta}$ and ϕ;

2. $\hat{\phi}$ is $K^{\frac{1}{2}}$-consistent given $\boldsymbol{\beta}$;

3. $|\partial\hat{\rho}(\boldsymbol{\beta}, \phi)/\partial\phi| \leq H(Y, \boldsymbol{\beta})$ which is $O_p(1)$,

Liang and Zeger (1986) showed that $K^{1/2}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})$ coverge to a multivariate Gaussian random vector with zero mean and covariance matrix $\mathbf{V}_G$

$$\mathbf{V}_G = \lim_{K\to\infty} K \left(\sum_{k=1}^{K} \mathbf{D}_k^T \mathbf{V}_k^{-1} \mathbf{D}_k \right)^{-1} \left\{ \sum_{k=1}^{K} \mathbf{D}_k^T \mathbf{V}_k^{-1} cov(\mathbf{T}_k) \mathbf{V}_k^{-1} \mathbf{D}_k \right\} \left(\sum_{k=1}^{K} \mathbf{D}_k^T \mathbf{V}_k^{-1} \mathbf{D}_k \right)^{-1}$$

This covariate matrix is estimated by evaluating D_k and V_k at $\hat{\rho}$, $\hat{\phi}$ and $\hat{\beta}$ and directly estimating $cov(\mathbf{T}_k)$.

3.3 Introduction to counting process notation

Counting process analysis is based on a history of the process, often called the filtration, denoted $\{\mathcal{F}_t; t > 0\}$. A natural choice is the history of the process up to time t.

A counting process is a stochastic process $\{N(t) : t \geq 0\}$ adapted to a filtration $\{\mathcal{F}_t : t \geq 0\}$ with $N(0) = 0$ and $N(t) < \infty$ a.s., and for which the possible paths are with probability one right-continuous, piecewise constant, and have only jump discontinuities, with positive jumps of size 1.

In the counting process approach to survival analysis, the time and censoring indicator pair (T_i, δ_i) is replaced by the pair $(N_i(t), Y_i(t))$, where

$$N_i(t) = \text{the number of observed events in [0, t] for unit i}$$

and

$$Y_i(t) = \begin{cases} 1 & \text{if unit i is under observation and at risk at time } t \\ 0 & \text{otherwise} \end{cases}$$

According to the Doob-Meyer decomposition theorem, any counting process may be uniquely decomposed as the sum of a martingale and a predictable, right continuous process that is 0 at time 0, called the compensator. The counting process $N_i(t)$ has compensator $A_i(t) = \int_0^t Y_i(s)\lambda_i(s)ds$, where $\lambda_i(t)$ is the hazard function. Then

$$N_i(t) = A_i(t) + M_i(t).$$

where

$$M_i(t) = N_i(t) - A_i(t) = N_i(t) - \int_0^t Y_i(s)\lambda_i(s)ds$$

is a counting process martingale with respect to the history given above.

Let $dN_i(t)$, a counting process differential, denote the increment in N_i over the infinitesimal time interval $[t, t + dt]$. Then

$$dN_i(t) = \begin{cases} 1 & \text{if a failure occurs in } [t, t+dt] \\ 0 & \text{otherwise} \end{cases}$$

$\mathcal{F}_{t-}$ contains all the information on [0,t), and

$$E(dN_i(t)|\mathcal{F}_{t-}) = dA_i(t) = Y_i(t)\lambda_i(t)dt.$$

Martingale increments have mean 0, i.e.,

$$E(dM_i(t)|\mathcal{F}_{t-}) = 0 \text{ for any } t > 0.$$

3.4 Generalized estimating equations for clustered survival data

3.4.1 Derivation of the estimating equations

The main advantage of using the generalized estimating equation approach by Liang and Zeger is that consistent estimators can be obtained without specifying either the joint distribution or the correct correlation structure of responses within a cluster. In Chapter 2, it was shown that the partial likelihood for the Cox proportional hazards model that ignores the within cluster correlation can be expressed in counting process notation. In the expression of the partial likelihood, a counting process differential, $dN_{kj}(t)$ can be approximately considered as a Poisson random variable, which belongs to the exponential family. The expectation and variance of $dN_{kj}(t)$ is the compensator differential, $dA_{kj}(t)$. Therefore the generalized estimating equation approach can be used to take within cluster correlation into consideration.

In this chapter, we consider a special case with only one binary covariate. The more general situation of multiple covariates will be discussed in Chapter 4. The Poisson probability function for $dN_{kj}(t)$ is proportional to

$$\exp\{-dA_{kj}(t)\}\, dA_{kj}(t)^{dN_{kj}(t)} \tag{3.5}$$

Using the exponential family density from (3.1), it follows that (3.5) can be written as

$$\begin{aligned} & \exp\{-dA_{kj}(t)\}\exp\{dN_{kj}(t)\log[dA_{kj}(t)]\} \\ = \; & \exp\left\{dN_{kj}(t)\log[Y_{kj}(t)e^{\beta Z_{kj}}d\Lambda_0(t)] - Y_{kj}(t)e^{\beta Z_{kj}}d\Lambda_0(t))\right\} \end{aligned}$$

Let $\phi = 1$, and take

$$\theta_{kj} = h(\eta_{kj}) = \log[Y_{kj}(t)e^{\beta Z_{kj}}d\Lambda_0(t)], \text{ where } \eta_{kj} = \beta Z_{kj},$$

then

$$\frac{\partial h(\eta_{kj})}{\partial \eta_{kj}} = 1$$

$$a(\theta_{kj}) = Y_{kj}(t)e^{\beta Z_{kj}}d\Lambda_0(t) = e^{\theta_{kj}}$$

$$E(dN_{kj}(t)) = a'(\theta_{kj}) = e^{\theta_{kj}} = dA_{kj}(t) = Y_{kj}(t)\exp(\beta Z_{kj})d\Lambda_0(t)$$

$$Var(dN_{kj}) = a''(\theta_{kj}) = e^{\theta_{kj}} = dA_{kj}(t) = Y_{kj}(t)\exp(\beta Z_{kj})d\Lambda_0(t)$$

and

$$\frac{\partial dA_{kj}(t)}{\partial \beta} = Y_{kj}(t)Z_{kj}\exp(\beta Z_{kj})d\Lambda_0(t) = Z_{kj}dA_{kj}(t)$$

A set of equations corresponding to the generalized estimating equations in (3.3) is

$$
\begin{aligned}
\mathbf{U}(\beta) &= \sum_{k=1}^{K}\int_0^t \left(\frac{\partial d\mathbf{A}_k(u)}{\partial\beta}\right)' \mathbf{V}_k^{-1}\left(d\mathbf{N}_k(u) - d\mathbf{A}_k(u)\right) \\
&= \sum_{k=1}^{K}\int_0^t \left(\frac{\partial d\mathbf{A}_k(u)}{\partial\beta}\right)' \left[diag(d\mathbf{A}_k(u))^{1/2}R(\rho)diag(d\mathbf{A}_k(u))^{1/2}\right]^{-1} \\
&\quad (d\mathbf{N}_k(u) - d\mathbf{A}_k(u)) \\
&= \sum_{k=1}^{K}\int_0^t \mathbf{Z}_k' diag\left(d\mathbf{A}_k(u)\right)\left[diag(d\mathbf{A}_k(u))^{1/2}R(\rho)diag(d\mathbf{A}_k(u))^{1/2}\right]^{-1} \\
&\quad (d\mathbf{N}_k(u) - d\mathbf{A}_k(u))
\end{aligned}
$$

Since $d\Lambda_0(t)$ is unknown, the Nelson-Aalen estimator is used to estimate it as

$$
d\hat{\Lambda}_0(t) = \frac{\sum_l \sum_q dN_{lq}(t)}{\sum_l \sum_q Y_{lq}(t)e^{\beta Z_{kj}}}.
$$

Then $dA_{kj}(u)$ can be estimated by

$$
\begin{aligned}
d\hat{A}_{kj}(u) &= Y_{kj}(u)e^{\beta Z_{kj}}d\hat{\Lambda}_0(u) \\
&= Y_{kj}(u)e^{\beta Z_{kj}}\frac{\sum_l \sum_q dN_{lq}(u)}{\sum_l \sum_q Y_{lq}(u)e^{\beta Z_{lq}}}
\end{aligned}
$$

Therefore the proposed generalized estimating equation for clustered survival data is of the form

$$
\sum_{k=1}^{K}\int_0^t \mathbf{Z}_k' diag\left(d\hat{\mathbf{A}}_k(u)\right)\left[diag(d\hat{\mathbf{A}}_k(u))^{1/2}R(\rho)diag(d\hat{\mathbf{A}}_k(u))^{1/2}\right]^{-1} \left(d\mathbf{N}_k(u) - d\hat{\mathbf{A}}_k(u)\right) = 0 \qquad (3.6)
$$

The solution to equation (3.6) is defined as $\hat{\beta}_g$.

If all the subjects within a cluster respond independently, then

$$
\mathbf{R}_k(\rho) = I
$$

and equation (3.6) becomes

$$
\sum_{k=1}^{K}\int_0^t \mathbf{Z}_k'\left(d\mathbf{N}_k(u) - d\hat{\mathbf{A}}_k(u)\right) = 0, \qquad (3.7)
$$

which is the same as the partial likelihood score equations that ignore the within cluster correlation.

The simplest case of non-zero correlation is that $\mathbf{R}_k(\rho)$ is the same for all the clusters, and ρ is a scaler. Define $\mathbf{R}_k(\rho)$ as $\mathbf{R}(\rho)$, with

$$\mathbf{R}(\rho) = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \vdots & \vdots & \vdots \\ \rho & \cdots & 1 \end{pmatrix}$$

where $\rho = corr(M_{ki}(t_{ki}), M_{kj}(t_{kj}))$. The details for estimating ρ were discussed in Chapter 2.

3.4.2 Illustration of generalized estimating equations

To illustrate the construction of generalized estimating equations, consider the data from the following simple experiment. There are three clusters, and each has a pair of data. A cross indicates a failure event and a circle indicates a censoring event.

Only failure times (1, 2, 7, 8) are included in the equations. If we use a common baseline hazard function and a single bivariate covariate Z corresponding to the treatment effect involved, the following illustrates the contributions of the three pairs of observations to the estimating equation at times 1, 2, 7 and 8 respectively, where Z_{kj} is the treatment indicator for the jth subject in the kth cluster.

At time=1

$$\begin{pmatrix} D_{11}(1) \\ D_{12}(1) \end{pmatrix}' \hat{W}_1(1) \begin{pmatrix} 1 - e^{\beta Z_{11}} d\Lambda_0(1) \\ 0 - e^{\beta Z_{12}} d\Lambda_0(1) \end{pmatrix}$$

$$\begin{pmatrix} D_{21}(1) \\ D_{22}(1) \end{pmatrix}' \hat{W}_2(1) \begin{pmatrix} 0 - e^{\beta Z_{21}} d\Lambda_0(1) \\ 0 - e^{\beta Z_{22}} d\Lambda_0(1) \end{pmatrix}$$

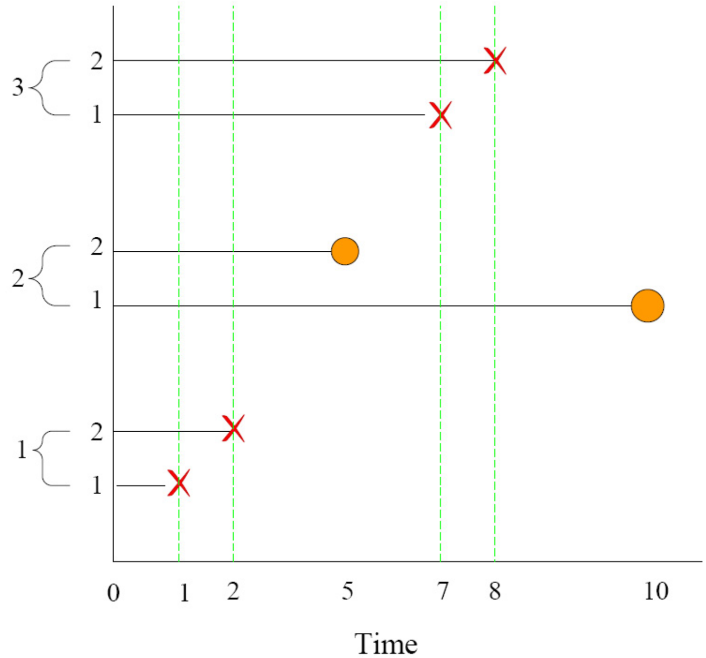

Figure 3.1 Illustration of generalized estimating equations

$$\begin{pmatrix} D_{31}(1) \\ D_{32}(1) \end{pmatrix}' \hat{W}_3(1) \begin{pmatrix} 0 - e^{\beta Z_{31}} d\Lambda_0(1) \\ 0 - e^{\beta Z_{32}} d\Lambda_0(1) \end{pmatrix}$$

At time=2

$$\begin{pmatrix} D_{11}(2) \\ D_{12}(2) \end{pmatrix}' \hat{W}_1(2) \begin{pmatrix} 0 \\ 1 - e^{\beta Z_{12}} d\Lambda_0(2) \end{pmatrix}$$

$$\begin{pmatrix} D_{21}(2) \\ D_{22}(2) \end{pmatrix}' \hat{W}_2(2) \begin{pmatrix} 0 - e^{\beta Z_{21}} d\Lambda_0(2) \\ 0 - e^{\beta Z_{22}} d\Lambda_0(2) \end{pmatrix}$$

$$\begin{pmatrix} D_{31}(2) \\ D_{32}(2) \end{pmatrix}' \hat{W}_3(2) \begin{pmatrix} 0 - e^{\beta Z_{31}} d\Lambda_0(2) \\ 0 - e^{\beta Z_{32}} d\Lambda_0(2) \end{pmatrix}$$

At time=7

$$\begin{pmatrix} D_{21}(7) \\ D_{22}(7) \end{pmatrix}' \hat{W}_2(7) \begin{pmatrix} 0 - e^{\beta Z_{21}} d\Lambda_0(7) \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} D_{31}(7) \\ D_{32}(7) \end{pmatrix}' \hat{W}_3(7) \begin{pmatrix} 1 - e^{\beta Z_{31}} d\Lambda_0(7) \\ 0 - e^{\beta' Z_{32}} d\Lambda_0(7) \end{pmatrix}$$

At time=8

$$\begin{pmatrix} D_{21}(8) \\ D_{22}(8) \end{pmatrix}' \hat{W}_2(8) \begin{pmatrix} 0 - e^{\beta Z_{21}} d\Lambda_0(8) \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} D_{31}(8) \\ D_{32}(8) \end{pmatrix}' \hat{W}_3(8) \begin{pmatrix} 0 \\ 1 - e^{\beta Z_{32}} d\Lambda_0(8) \end{pmatrix}$$

where

$$\begin{pmatrix} D_{k1}(u) \\ D_{k2}(u) \end{pmatrix} = \begin{pmatrix} Z_{k1} & Z_{k2} \end{pmatrix} \begin{pmatrix} dA_{k1}(u) & \\ & dA_{k2}(u) \end{pmatrix}$$

3.5 Simulation study

Simulation studies are used to compare the estimators using partial likelihood score equations ignoring the within cluster correlation and estimators using the proposed set of generalized estimating equations. The solution to the partial likelihood score equation, equation (3.7), is denoted by the IWM estimator, $\hat{\beta}_u$. The solution to the generalized estimating equations, equation (3.6), is denoted by the GEE estimator, $\hat{\beta}_g$.

The simulation studies were described in Chapter 2. The parameter values considered in the simulation studies in this section are listed in Table 3.1. For each simulation study, 500 simulated data sets containing 100 clusters of size 2 were used. Each of the 500 simulated data sets produces one value of $\hat{\beta}_u$ and one value of $\hat{\beta}_g$.

There are two designs considered in the simulation study. One design is a balanced randomized design for which one subject in each cluster is randomly assigned to the treatment and the other subject receives the control. The other design is a completely

parameters	function	possible values
β	true coefficient	0.693, 0, -0.5
P_5	censoring rate (%)	0, 10, 20, 50
θ	association parameter	0.25, 0.8, 1.5
K	number of clusters	100
n	cluster size	2
N	number of simulated data sets	500

Table 3.1 Parameters of simulation studies

randomized design for which the random treatment assignment is independently determined for each subject in a cluster. The two subjects can both receive the treatment, or both receive the control, or one receives the treatment and one receives the control. The bias and relative efficiency of $\hat{\beta}_u$ and $\hat{\beta}_g$ are examined for these two designs.

3.5.1 Balanced randomized design

Table 3.2 displays simulation results when a balanced randomized design is used. There are 500 datasets generated in each simulation study, thus 500 pairs of estimates are produced. In the table, the estimated bias is the difference between the average of 500 estimates and the true parameter value β. Empirical variances are calculated using a sample of 500 estimates. Simulation results show that $\hat{\beta}_u$ and $\hat{\beta}_g$ are identical in the case of a balanced randomized design. Therefore, only the results for $\hat{\beta}_u$ are shown. If the design is balanced, $\hat{\beta}_u$ which ignores within cluster dependence, does not lose efficiency. The empirical variance of $\hat{\beta}_u$ is much smaller than that for a completely randomized design. So it is preferable to used balanced randomized design when possible.

β	P_5	estimated bias	empirical variance
0.693	0	0.0097	0.0057
	10%	0.0139	0.0066
	20%	0.0150	0.0010
	50%	0.0098	0.0240

Table 3.2 Small sample simulation results for $\hat{\beta}_u$ for 500 simulated samples of 100 clusters of size 2 with $\theta = 0.25$ corresponding to correlation 0.937 using a balanced randomized design.

3.5.2 Completely randomized design

Table 3.3, Table 3.4, Table 3.5 and Table 3.6 display simulation results when treatments are assigned randomly within clusters. There are 500 datasets generated in each simulation study, thus 500 estimates were produced. In Table 3.3, the estimated bias is the difference between the average of 500 estimates and the true parameter value β. Empirical variances are calculated from the sample of 500 estimates. $SRE(\hat{\beta}_u|\hat{\beta}_g)$ is the empirical variance of $\hat{\beta}_u$ divided by the empirical variance of $\hat{\beta}_g$. A value of $SRE(\hat{\beta}_u|\hat{\beta}_g)$ larger than one indicates that the estimator produced by equations (3.6) has the smaller variance than the IWM estimator. Then we say that there is a gain in efficiency of $\hat{\beta}_g$ relative to $\hat{\beta}_u$. In the tables, the true coefficient β assumes values 0.693, 0 and -0.5, and the association parameter θ assumes values 0.25, 0.80 and 1.50, corresponding to correlations 0.937, 0.712, and 0.512. P_5 is the censoring rate with combination of random and Type I censoring given that the study is ended at 5 time units. P_5 is set at different rates changing from 0 to 50%. For all the parameter settings, the estimated biases of $\hat{\beta}_u$ and $\hat{\beta}_g$ are close to 0. There is no obvious advantage of using one method over the other to reduce bias. When the value of β is 0.693 and there is no censoring, $\hat{\beta}_g$ provides gains in efficiency for all three levels of association. The gains in efficiency provided by $\hat{\beta}_g$ is greater for stronger correlation. When the censoring rate is moderate, 10% and 20%, $\hat{\beta}_g$ provides gains in efficiency when the within cluster correlation is strong. When the censoring rate is high, 50%, a gain in efficiency only occurs when correlation is very strong. When the value of β is 0, gains in efficiency follow the trend for β equal to 0.693, but the gains in efficiency are even more pronouned, especially when there is no censoring, and the within cluster correlation is strong. When the value of β is -0.5, the $SRE(\hat{\beta}_u|\hat{\beta}_g)$ values are between those for β values of 0 and 0.693 for all censoring rates and correlation levels considered in this study.

β	P_5	θ	$\hat{\beta}_u$ estimated bias	$\hat{\beta}_u$ empirical variance	$\hat{\beta}_g$ estimated bias	$\hat{\beta}_g$ empirical variance	$SRE(\hat{\beta}_u \mid \hat{\beta}_g)$
0.693	0	0.25	0.0187	0.0290	0.0224	0.0104	2.788
		0.80	0.0154	0.0267	0.0125	0.0165	1.618
		1.50	0.0166	0.0259	0.0176	0.0206	1.257
	10%	0.25	0.0165	0.0262	0.0112	0.0146	1.794
		0.80	0.0209	0.0264	0.0150	0.0213	1.239
		1.50	0.0082	0.0260	0.0107	0.0239	1.088
	20%	0.25	0.0115	0.0294	0.0145	0.0175	1.680
		0.80	0.0070	0.0294	-0.0034	0.0242	1.215
		1.50	-0.0110	0.0297	-0.0095	0.0273	1.088
	50%	0.25	0.0314	0.0451	0.0149	0.0349	1.292
		0.80	-0.0046	0.0444	-0.0084	0.0421	1.055
		1.50	0.0041	0.0451	0.0038	0.0440	1.025
0	0	0.25	-0.0102	0.0233	0.0010	0.0033	7.061
		0.80	-0.0140	0.0223	-0.0089	0.0112	1.991
		1.50	0.0035	0.0223	0.0042	0.0168	1.327
	10%	0.25	-0.0002	0.0262	-0.0033	0.0938	2.794
		0.80	0.0019	0.0264	0.0022	0.0213	1.239
		1.50	0.0023	0.0260	0.0003	0.0239	1.087
	20%	0.25	-0.0020	0.0254	0.0019	0.0115	2.208
		0.80	0.0092	0.0260	0.0064	0.0198	1.313
		1.50	0.0128	0.0272	0.0129	0.0246	1.105
	50%	0.25	-0.0104	0.0407	-0.0091	0.0286	1.423
		0.80	-0.0037	0.0413	-0.0066	0.0372	1.110
		1.50	0.0029	0.0420	-0.0007	0.0403	1.042
-0.5	0	0.25	-0.0132	0.0235	-0.0143	0.0060	3.916
		0.80	-0.0188	0.0227	-0.0167	0.0127	1.787
		1.50	0.0002	0.0227	-0.0040	0.0177	1.282
	10%	0.25	-0.0015	0.0290	0.0000	0.0172	1.686
		0.80	-0.0090	0.0282	-0.0113	0.0239	1.180
		1.50	0.0092	0.0285	0.0058	0.0271	1.052
	20%	0.25	-0.0016	0.0283	-0.0048	0.0184	1.538
		0.80	-0.0157	0.0284	-0.0179	0.0250	1.136
		1.50	-0.0052	0.0291	-0.0076	0.0282	1.032
	50%	0.25	-0.0242	0.0429	-0.0193	0.0319	1.344
		0.80	-0.0126	0.0448	-0.0126	0.0407	1.100
		1.50	-0.0159	0.0449	-0.0149	0.0432	1.039

Table 3.3 Simulation results for $\hat{\beta}_u$ and $\hat{\beta}_g$ for 500 simulated samples with 100 clusters of size 2. $\theta = 0.25$ corresponding to correlation 0.937, $\theta = 0.80$ corresponding to correlation 0.712, and $\theta = 1.50$ corresponding to correlation 0.512.

Empirical variances in Table 3.4 are the same as those in Table 3.3. The column labelled average robust variance contains the average of 500 robust variance estimates of IWM estimator. The variance ratio for $\hat{\beta}_u$ is the average robust estimate variance of $\hat{\beta}_u$ divided by the empirical variance of $\hat{\beta}_u$. A value of the variance ratio of $\hat{\beta}_u$ greater than one indicates that robust variance overestimates the variance. And a value of the variance ratio of $\hat{\beta}_u$ less than one indicates that the robust variance underestimates the variance. The variance of $\hat{\beta}_g$ is obtained by a bootstrap resampling method. For clustered data, a bootstrap method resamples clusters instead of individual subjects. A set of 100 bootstrap samples was taken from each of the 500 simulated data sets to produce a bootstrap variance estimate. The average bootstrap variance is the average of the 500 bootstrap variance estimates of $\hat{\beta}_g$ obtained from the 500 datasets. The variance ratio for $\hat{\beta}_g$ is the average bootstrap variance of $\hat{\beta}_g$ divided by the empirical variance of $\hat{\beta}_g$. A value of the variance ratio for $\hat{\beta}_g$ greater than one indicates that the bootstrap variance overestimates the variance. A value of the variance ratio of $\hat{\beta}_g$ less than 1 indicates that the bootstrap variance underestimates the variance. From the results provided in Table 3.4, all values of the variance ratios are close to one. The robust variances estimator for $\hat{\beta}_u$ tends to underestimate the variances in some cases. Overall the bootstrap method provides an accurate estimate of the variance of $\hat{\beta}_g$.

			$\hat{\beta}_u$			$\hat{\beta}_g$		
β	P_5	θ	empirical variance	average robust variance	variance ratio	empirical variance	average bootstrap variance	variance ratio
0.693	0	0.25	0.0290	0.0231	0.796	0.0104	0.0102	0.980
		0.80	0.0267	0.0228	0.853	0.0165	0.0163	0.987
		1.50	0.0259	0.0224	0.864	0.0206	0.0201	0.975
	10%	0.25	0.0262	0.0256	0.977	0.0146	0.0139	0.952
		0.80	0.0264	0.0250	0.946	0.0213	0.0198	0.929
		1.50	0.0260	0.0248	0.953	0.0239	0.0231	0.966
	20%	0.25	0.0294	0.0287	0.976	0.0175	0.0182	1.040
		0.80	0.0294	0.0280	0.952	0.0242	0.0241	0.995
		1.50	0.0297	0.0278	0.936	0.0273	0.0271	0.993
	50%	0.25	0.0451	0.0454	1.006	0.0349	0.0380	1.088
		0.80	0.0444	0.0444	1.000	0.0421	0.0430	1.021
		1.50	0.0451	0.0443	0.982	0.0440	0.0442	1.004
0	0	0.25	0.0233	0.0197	0.845	0.0033	0.0039	1.181
		0.80	0.0223	0.0206	0.923	0.0112	0.0125	1.116
		1.50	0.0223	0.0201	0.901	0.0168	0.0171	1.017
	10%	0.25	0.0262	0.0256	0.977	0.0938	0.0090	0.960
		0.80	0.0264	0.0250	0.946	0.0213	0.0198	0.929
		1.50	0.0260	0.0248	0.953	0.0239	0.0231	0.966
	20%	0.25	0.0254	0.0251	0.988	0.0115	0.0124	1.078
		0.80	0.0260	0.0252	0.969	0.0198	0.0199	1.005
		1.50	0.0272	0.0252	0.926	0.0246	0.0235	0.955
	50%	0.25	0.0407	0.0399	0.980	0.0286	0.0301	1.052
		0.80	0.0413	0.0401	0.970	0.0372	0.0382	1.026
		1.50	0.0420	0.0412	0.980	0.0403	0.0405	1.004
-0.5	0	0.25	0.0235	0.0214	0.910	0.0060	0.0073	1.216
		0.80	0.0227	0.0212	0.933	0.0127	0.0142	1.118
		1.50	0.0227	0.0212	0.933	0.0177	0.0183	1.033
	10%	0.25	0.0290	0.0289	0.996	0.0172	0.0175	1.017
		0.80	0.0282	0.0286	1.014	0.0239	0.0242	1.012
		1.50	0.0285	0.0285	1.000	0.0271	0.0273	1.007
	20%	0.25	0.0283	0.0271	0.957	0.0184	0.0186	1.010
		0.80	0.0284	0.0268	0.943	0.0250	0.0234	0.936
		1.50	0.0291	0.0268	0.920	0.0282	0.0261	0.925
	50%	0.25	0.0429	0.0430	1.002	0.0319	0.0336	1.053
		0.80	0.0448	0.0427	0.953	0.0407	0.0412	1.012
		1.50	0.0449	0.0437	0.973	0.0432	0.0430	0.995

Table 3.4 Simulation results for 500 simulated samples with 100 clusters of size 2. The number of bootstrap samples is 100 for each of the 500 simulated sets. The variance ratio for $\hat{\beta}_u$ is the average robust variance estimate divided by the empirical variance, and the variance ratio for $\hat{\beta}_g$ is the average bootstrap variance divided by the empirical variance.

Tables 3.5 and 3.6 show coverage rates and coverage lengths for nominal 95% confidence intervals for β_u and β_g. The confidence intervals based on $\hat{\beta}_u$ are contructed as $\hat{\beta}_u \pm (1.96)S_{robust}$, where S_{robust} is the standard error of $\hat{\beta}_u$ obtained from the robust covariance procedure. There are four bootstrap methods used to construct bootstrap confidence intervals of β_g, normal, quantile, basic and accelerated bias-corrected (BC_α). They follow the notation in Davison and Hinkley (1997). The normal method is evaluated as $\hat{\beta}_g \pm (1.96)S_{boot}$, where S_{boot} is the bootstrapped standard error of $\hat{\beta}_g$. For b bootstrap samples, denote the bootstrap distribution of bootstrap estimates by $\hat{F}_b(x)$. The quantile confident interval method uses the $\alpha/2$ and $1-\alpha/2$ quantiles of $\hat{F}_b(x)$ as the endpoints of the $1-\alpha$ level confidence interval. The basic confident interval uses the upper quantile of a bootstrap distribution to calculate the lower confidence bound, and the lower quantile to calculate the upper confidence bound. The endpoints using this method are $2\hat{\beta}_g - \hat{F}_b^{-1}(1-\alpha/2)$, and $2\hat{\beta}_g - \hat{F}_b^{-1}(\alpha/2)$. The accelerated bias-corrected method obtains end points by inverting percentiles of the bootstrap distribution after adjusting for bias and acceleration shift. Denote the acceleration constant by a, and the normal cdf by $\Phi(x)$. Define $Z_0 = \Phi^{-1}(\hat{F}_b(\hat{\beta}_g))$, and define $Z_{\alpha/2}$ as the $\alpha/2$-percentile of a standard normal distribution. Then the α endpoints of the BC_a confidence interval are $\hat{F}_b^{-1}\left\{\Phi\left(Z_0 + \frac{Z_0 - Z_\alpha/2}{1-a(Z_0 - Z_{\alpha/2})}\right\}\right)$ and $\hat{F}_b^{-1}\left\{\Phi\left(Z_0 + \frac{Z_0 + Z_\alpha/2}{1-a(Z_0 + Z_{\alpha/2})}\right\}\right)$. Among all the methods, the normal method is the simplest. From the results in Table 3.5, all the methods provide coverage rates around 95 percent. The simple normal method performs as well as any of the other bootstrap methods. From the results in Table 3.6, when the censoring rate is moderate, and the correlation is strong, the lengths of bootstrap confidence intervals based on $\hat{\beta}_g$ tend to be narrower than the lengths of the confidence intervals based on $\hat{\beta}_u$. Comparing all the bootstrapping methods based on $\hat{\beta}_g$, the normal method tends to provide the most narrow confidence intervals.

β	P_5	θ	average coverage rate of confidence interval based on robust variance estimator for $\hat{\beta}_u$	average coverage rate of bootstrap confidence interval based on $\hat{\beta}_g$			
				normal	quantile	basic	BCa
0.693	0	0.25	0.920	0.940	0.930	0.924	0.916
		0.80	0.934	0.950	0.932	0.946	0.928
		1.50	0.934	0.946	0.944	0.946	0.940
	10%	0.25	0.948	0.926	0.938	0.940	0.920
		0.80	0.948	0.924	0.942	0.932	0.920
		1.50	0.956	0.936	0.940	0.950	0.936
	20%	0.25	0.940	0.942	0.952	0.952	0.932
		0.80	0.956	0.950	0.944	0.958	0.948
		1.50	0.954	0.940	0.956	0.954	0.946
	50%	0.25	0.948	0.960	0.948	0.966	0.952
		0.80	0.950	0.960	0.952	0.964	0.954
		1.50	0.954	0.956	0.950	0.962	0.950
0	0	0.25	0.932	0.982	0.960	0.986	0.980
		0.80	0.940	0.962	0.946	0.962	0.948
		1.50	0.928	0.946	0.936	0.956	0.940
	10%	0.25	0.948	0.926	0.938	0.940	0.920
		0.80	0.948	0.924	0.942	0.932	0.920
		1.50	0.956	0.936	0.940	0.950	0.936
	20%	0.25	0.954	0.966	0.962	0.976	0.958
		0.80	0.956	0.968	0.948	0.970	0.944
		1.50	0.948	0.962	0.950	0.970	0.958
	50%	0.25	0.942	0.960	0.956	0.974	0.946
		0.80	0.944	0.960	0.944	0.966	0.956
		1.50	0.948	0.950	0.942	0.966	0.942
-0.5	0	0.25	0.938	0.958	0.956	0.958	0.936
		0.80	0.942	0.946	0.946	0.960	0.946
		1.50	0.948	0.938	0.950	0.938	0.946
	10%	0.25	0.962	0.958	0.950	0.968	0.944
		0.80	0.944	0.958	0.946	0.956	0.952
		1.50	0.946	0.958	0.948	0.960	0.956
	20%	0.25	0.944	0.954	0.934	0.968	0.936
		0.80	0.946	0.954	0.928	0.966	0.948
		1.50	0.950	0.952	0.946	0.960	0.948
	50%	0.25	0.948	0.960	0.962	0.970	0.960
		0.80	0.928	0.950	0.942	0.960	0.946
		1.50	0.934	0.952	0.940	0.960	0.932

Table 3.5 Simulated coverage rates for nominal 95% confidence intervals for 500 simulated samples with 100 clusters of size 2. The number of bootstrap samples is 100 for each of the 500 simulated sets.

β	P_5	θ	average length of confidence interval based on robust variance estimator for $\hat{\beta}_u$	average length of bootstrap confidence interval based on $\hat{\beta}_g$			
				normal	quantile	basic	BCa
0.693	0	0.25	0.595	0.390	0.409	0.409	0.395
		0.80	0.591	0.495	0.518	0.518	0.513
		1.50	0.586	0.552	0.578	0.578	0.572
	10%	0.25	0.627	0.456	0.477	0.477	0.464
		0.80	0.619	0.547	0.571	0.571	0.566
		1.50	0.617	0.591	0.617	0.617	0.615
	20%	0.25	0.664	0.522	0.550	0.550	0.538
		0.80	0.656	0.604	0.631	0.631	0.629
		1.50	0.653	0.640	0.669	0.669	0.668
	50%	0.25	0.835	0.757	0.794	0.794	0.787
		0.80	0.826	0.816	0.853	0.853	0.855
		1.50	0.825	0.822	0.831	0.831	0.828
0	0	0.25	0.550	0.260	0.273	0.273	0.272
		0.80	0.562	0.435	0.455	0.455	0.456
		1.50	0.555	0.509	0.532	0.532	0.533
	10%	0.25	0.627	0.348	0.365	0.365	0.366
		0.80	0.620	0.489	0.511	0.511	0.513
		1.50	0.617	0.549	0.572	0.572	0.574
	20%	0.25	0.621	0.431	0.451	0.451	0.453
		0.80	0.622	0.549	0.575	0.575	0.578
		1.50	0.622	0.598	0.623	0.623	0.625
	50%	0.25	0.783	0.675	0.708	0.708	0.706
		0.80	0.785	0.762	0.796	0.796	0.800
		1.50	0.795	0.788	0.792	0.792	0.803
-0.5	0	0.25	0.573	0.329	0.346	0.346	0.332
		0.80	0.570	0.462	0.484	0.484	0.480
		1.50	0.570	0.527	0.550	0.550	0.548
	10%	0.25	0.666	0.514	0.540	0.540	0.538
		0.80	0.662	0.606	0.632	0.632	0.630
		1.50	0.661	0.644	0.658	0.658	0.653
	20%	0.25	0.645	0.469	0.490	0.490	0.483
		0.80	0.642	0.572	0.598	0.598	0.594
		1.50	0.642	0.618	0.634	0.634	0.632
	50%	0.25	0.812	0.713	0.752	0.752	0.751
		0.80	0.810	0.791	0.805	0.805	0.803
		1.50	0.819	0.812	0.815	0.815	0.807

Table 3.6 Simulated lengths of confidence intervals for 500 simulated samples with 100 clusters of size 2. The number of bootstrap samples is 100 for each of the 500 simulated sets.

3.6 Summary

In this chapter, a new set of estimating equations was developed using a generalized estimating equation approach. Simulation studies were used to compare the estimators provided by those new estimating equations to the estimators from the partial likelihood score equations that ignore the within cluster dependence.

Table 3.2 displays simulation results for balanced randomized designs. Those simulation results show that the partial likelihood score equations that ignore the within cluster dependence produced the same estimators as generalized estimating equations. There is no efficiency loss by ignoring the within cluster dependence.

Table 3.3, 3.4 and 3.5 display simulation results when the treatment effects are randomized within a cluster. From Table 3.3, it can be seen that when the correlation is high and the censoring rate is moderate, there are efficiency gains. The magnitudes of the gains differ from the values of the regression parameters. Efficiency gains tend to be greater for parameter values closer to zero. Results in Table 3.4 show that the bootstrap variances are close to the empirical variances. They provide reliable estimates of variances of $\hat{\beta}_g$. From Tables 3.5 and 3.6, all four bootstrap methods give similar results. The coverage rates are around 95%. When within cluster correlation is strong, and the censoing rate is moderate, the lengths of 95% confidence intervals of β_g are smaller than those of β_u. The normal method gives confidence intervals with shortest length among the four bootstrap confidence intervals we studied.

CHAPTER 4. Extended simulation results of generalized estimating equations

4.1 Introduction

In Chapter 3, we developed a generalized estimating equation approach to estimate the coefficients in the Cox proportional hazards model for clustered survival data. Simulation studies were used to assess the performance of the estimator for models with a single dichotomous covariate and 100 clusters of size 2. Simulation results show that when the censoring rate is moderate, and the within cluster correlation is high, the estimators from the generalized estimating equations can provide substantial gains in efficiency relative to the estimators obtained by ignoring within cluster dependence. In this chapter, additional simulation studies are examined to assess performance with clusters of varying numbers and sizes, a single continuous covariate, and multiple covariates. Different within cluster correlation structures are considered for cluster sizes greater than two. When a data set is generated assuming a first-order autoregressive correlation structure, the estimators using the partial likelihood score equations that ignore the within cluster dependence, using the generalized estimating equations with an exchangeable correlation structure, and using the generalized estimating equations with the correct correlation structure are studied.

4.2 General description of the simulation study

In the following sections except Section 4.7, the data generation follows the procedures described in Chapter 2. The solution to the partial likelihood score equations that ignore the within cluster dependence is denoted by the IMW estimator, $\hat{\beta}_u$. The solution to the generalized estimating equations is denoted by the GEE estimator, $\hat{\beta}_g$. There are 500 data sets generated in each simulation study, and each of the 500 simulated data sets produces one IMW estimator and one GEE estimator. The estimated bias and small sample relative efficiency are two criteria used to evaluate the estimators. The estimated bias is the difference of the average of 500 estimates and the true parameter value β. The empirical variances are calculated from the sample of 500 estimates. $SRE(\hat{\beta}_u|\hat{\beta}_g)$ is the empirical variance of $\hat{\beta}_u$ divided by the empirical variance of $\hat{\beta}_g$. A value of $SRE(\hat{\beta}_u|\hat{\beta}_g)$ larger than one indicates that the GEE estimator has the smaller variance than the IWM estimator. Then we say that there is a gain in efficiency of $\hat{\beta}_g$ relative to $\hat{\beta}_u$.

4.3 Effect of the number of clusters

In this section, the effect of the number of clusters is examined. Table 4.1 displays the simulation results of how the bias and small sample relative efficiency changes as the number of clusters changes. In the table, the number of clusters, K, changes from 20 to 100 with a single dichotomous covariate for a cluster size of 2 and an association parameter of 0.25. P_5 is the censoring rate with combination of random and Type I censoring given that the study is ended at 5 time units. P_5 is set at different rates changing from 0 to 50%. For all the parameter settings we studied, there are more cases for which $\hat{\beta}_g$ has smaller bias than $\hat{\beta}_u$, but there is no obvious advantage of using generalized estimating equations with respect to reducing bias. Bias tends to be larger for smaller numbers of cluster and values of β further away from zero. When K is fixed, the efficiency of the GEE estimator relative to the IWM estimator decreases as the censoring rate increases. There are larger gains in efficiency for the GEE estimator relative to the IWM estimator when $\beta = 0$ than when $\beta = 0.693$ for each value of K and each censoring rate considered. This is consistent with the simulation results presented in Chapter 3.

Figure 4.1 shows the variance changes of $\hat{\beta}_u$ and $\hat{\beta}_g$ as the number of clusters changes using clusters of size 2, and 500 simulated data sets for the true coefficient $\beta = 0.693$ and no censoring. The solid line indicates the simulated variance of $\hat{\beta}_u$ and the dashed line indicates the simulated variance of $\hat{\beta}_g$. The variances of both $\hat{\beta}_u$ and $\hat{\beta}_g$ decrease as K increases. When the number of clusters is smaller, increasing the number of clusters results in larger decrease in variance. For example, the decrease in variance from $K = 20$ to $K = 40$ is much larger than from $K = 80$ to $K = 100$. There is no obvious pattern of changes in relative efficiency as the number of clusters changes.

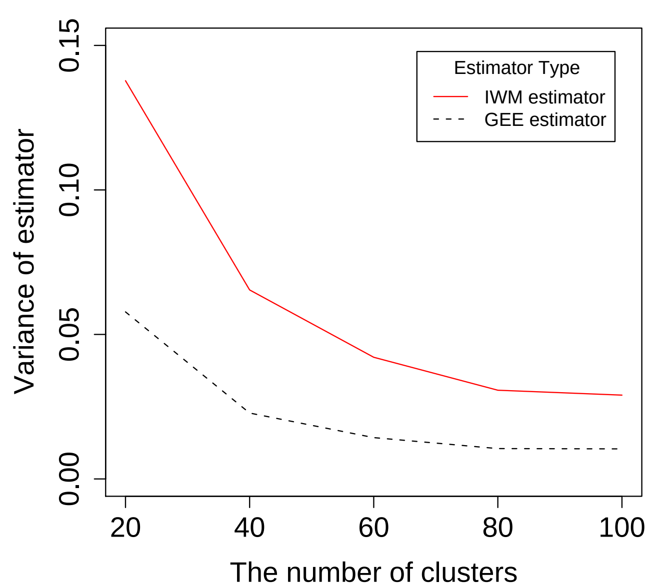

Figure 4.1 The variances of $\hat{\beta}_u$ and $\hat{\beta}_g$ for 500 simulated samples of the clusters of size 2 with $\beta = 0.693$, $\theta = 0.25$ corresponding to correlation 0.937 and no censoring.

β	P_5	K	$\hat{\beta}_u$ estimated bias	$\hat{\beta}_u$ empirical variance	$\hat{\beta}_g$ estimated bias	$\hat{\beta}_g$ empirical variance	$SRE(\hat{\beta}_u\|\hat{\beta}_g)$
0.693	0	20	0.1048	0.1378	0.0842	0.0579	2.380
		40	0.0324	0.0654	0.0370	0.0228	2.862
		60	0.0212	0.0421	0.0211	0.0143	2.941
		80	0.0230	0.0307	0.0185	0.0105	2.913
		100	0.0187	0.0290	0.0225	0.0104	2.784
	10%	20	0.0513	0.1475	0.0629	0.0747	1.973
		40	0.0313	0.0688	0.0326	0.0301	2.285
		60	0.0104	0.0486	0.0153	0.0212	2.291
		80	0.0207	0.0357	0.0122	0.0159	2.237
		100	0.0081	0.0262	0.0067	0.0146	1.782
	20%	20	0.0451	0.1726	0.0518	0.0949	1.817
		40	0.0268	0.0793	0.0310	0.0415	1.910
		60	0.0010	0.0528	0.0054	0.0271	1.944
		80	0.0170	0.0376	0.0063	0.0195	1.924
		100	0.0063	0.0294	0.0082	0.0175	1.681
	50%	20	0.0358	0.2665	0.0417	0.2225	1.197
		40	0.0306	0.1352	0.0263	0.0941	1.436
		60	-0.0273	0.0785	-0.0167	0.0582	1.348
		80	0.0135	0.0589	0.0019	0.0446	1.320
		100	0.0101	0.0451	0.0113	0.0349	1.291
0	0	20	0.0285	0.1225	0.0085	0.0348	3.513
		40	-0.0080	0.0541	-0.0061	0.0107	5.016
		60	-0.0002	0.0357	-0.0026	0.0066	5.374
		80	0.0066	0.0257	-0.0004	0.0045	5.702
		100	-0.0012	0.0233	0.0009	0.0033	7.032
	10%	20	-0.0134	0.1347	-0.0070	0.0612	2.198
		40	0.0038	0.0600	-0.0032	0.0218	2.747
		60	-0.0048	0.0426	-0.0059	0.0138	3.091
		80	0.0064	0.0283	-0.0022	0.0101	2.805
		100	-0.0077	0.0222	-0.0109	0.0081	2.732
	20%	20	-0.0186	0.1527	-0.0123	0.0816	1.870
		40	0.0058	0.0675	0.0066	0.0298	2.260
		60	-0.0121	0.0456	-0.0093	0.0199	2.292
		80	0.0048	0.0298	-0.0047	0.0134	2.225
		100	-0.0091	0.0254	-0.0077	0.0115	2.200
	50%	20	0.0032	0.2235	0.0011	0.1572	1.421
		40	0.0024	0.1163	-0.0003	0.0808	1.440
		60	-0.0309	0.0689	-0.0022	0.0471	1.462
		80	-0.0000	0.0486	-0.0099	0.0331	1.469
		100	-0.0049	0.0407	-0.0058	0.0286	1.421

Table 4.1 Simulation results for $\hat{\beta}_u$ and $\hat{\beta}_g$ for 500 samples of K clusters of size 2 with $\theta = 0.25$ corresponding to correlation 0.937.

Table 4.2 shows the behavior of estimated correlation among martingales $\hat{\rho}_M$ as the number of cluster, K changes. 500 simulated data sets with K clusters of size 2 were generated. In the table, K changes from 20 to 100 with a single dichotomous covariate for an association parameter $\theta = 0.25$, corresponding to correlation among failures time 0.937. P_5 is the censoring rate with combination of random and Type I censoring given that the study is ended at 5 time units. P_5 is set at different rates changing from 0 to 50%. Estimated correlation is the average of 500 estimates of the correlation. Empirical variances are calculated from a sample of 500 estimates. Simulation results show that the estimated correlations among martingales are lower than the correlations among failure times. The difference is smaller as the number of clusters is larger and the censoring rate is lower. The empirical variance of $\hat{\rho}_M$ is smaller as the number of clusters is larger.

β	P_5	K	estimated correlation	empirical variance
0.693	0	20	0.8409	0.0100
		40	0.8808	0.0026
		60	0.8959	0.0015
		80	0.9027	0.0010
		100	0.9172	0.0008
	10%	20	0.7830	0.0119
		40	0.8024	0.0064
		60	0.8134	0.0045
		80	0.8174	0.0033
		100	0.8352	0.0025
	20%	20	0.7211	0.0137
		40	0.7354	0.0074
		60	0.7454	0.0053
		80	0.7483	0.0037
		100	0.7516	0.0030
	50%	20	0.5457	0.0221
		40	0.5569	0.0116
		60	0.5649	0.0069
		80	0.5652	0.0059
		100	0.5704	0.0041
0	0	20	0.8391	0.0099
		40	0.8798	0.0027
		60	0.8964	0.0015
		80	0.9029	0.0010
		100	0.9175	0.0008
	10%	20	0.7740	0.0128
		40	0.7946	0.0068
		60	0.8019	0.0052
		80	0.8031	0.0039
		100	0.8124	0.0028
	20%	20	0.7122	0.0147
		40	0.7304	0.0076
		60	0.7405	0.0057
		80	0.7449	0.0038
		100	0.7508	0.0032
	50%	20	0.5412	0.0117
		40	0.5548	0.0111
		60	0.5557	0.0070
		80	0.5589	0.0059
		100	0.5612	0.0043

Table 4.2 Simulation results for estimated correlation among martingales for 500 simulated samples of K clusters of size 2 with $\theta = 0.25$, corresponding to correlation among failure times 0.937.

4.4 Effect of cluster size

In this section, we discuss how the cluster size, n, impacts efficiency. The case of 100 clusters of varying size, n, is considered with a single binary covariate with true coefficient $\beta = 0.693$. P_5 is the censoring rate with combination of random and Type I censoring given that the study is ended at 5 time units. P_5 is set at 0 and 20%. The cluster size n changes from 2 to 25. A multivariate version of the Copula model examined by Cai and Prentice (1997) is used. It was described in Section 2.5.2.

Table 4.3 shows that there is no obvious advantage of using generalized estimating equations with respect to reducing bias. When n is fixed, the efficiency of the GEE estimator relative to the IWM estimator decreases as the censoring rate increases. This is consistent with the effect of censoring rate obtained in simulation results in Chapter 3. The results also show as the cluster size increases, the efficiency of $\hat{\beta}_g$ relative to $\hat{\beta}_u$ first goes up and then goes down, but the maximum gain in efficiency is not as large for 20% censoring. The variances of both $\hat{\beta}_u$ and $\hat{\beta}_g$ decrease as the cluster size increases. When there is no censoring, for example, the decrease in variance from $n = 2$ to $n = 3$ is much larger for $\hat{\beta}_g$ than for $\hat{\beta}_u$. That results in a larger gain in efficiency for $\hat{\beta}_g$ at $n = 3$ than $n = 2$.

A balanced randomized design for cluster size of two means that one of the subjects in the cluster receives the treatment, and the other one receives the control. In Chapter 3, simulation results show that $\hat{\beta}_u$ and $\hat{\beta}_g$ are equivalent if a balanced randomized design is used. For cluster size of more than two, simulation results (not presented) showed that when there is a single binary covariate and all the clusters have the same assignment of covariate levels, $\hat{\beta}_u$ and $\hat{\beta}_g$ are the same if exchangeable correlation structure is used. Gains in efficiency of $\hat{\beta}_g$ relative to $\hat{\beta}_u$ occurs when random treatment assignment, obtained from "tossing a coin" for each individual subject is used. It produces unbalanced treatment assignments within a cluster. But within cluster imbalance tends to diminish for larger cluster sizes. Therefore the gains in efficiency tend to be smaller as the cluster size increases.

			$\hat{\beta}_u$		$\hat{\beta}_g$		
β	P_5	n	estimated bias	empirical variance	estimated bias	empirical variance	$SRE(\hat{\beta}_u\|\hat{\beta}_g)$
0.693	0	2	0.0187	0.0290	0.0224	0.0104	2.788
		3	0.0141	0.0203	0.0162	0.0064	3.181
		4	0.0171	0.0143	0.0209	0.0060	2.401
		5	0.0135	0.0110	0.0132	0.0050	2.192
		6	0.0153	0.0099	0.0150	0.0048	2.072
		7	0.0225	0.0098	0.0181	0.0049	2.004
		8	0.0163	0.0089	0.0194	0.0049	1.818
		9	0.0183	0.0088	0.0209	0.0048	1.833
		10	0.0239	0.0079	0.0240	0.0047	1.680
		25	0.0192	0.0055	0.0213	0.0042	1.298
	20%	2	0.0115	0.0294	0.0145	0.0175	1.680
		3	0.0137	0.0220	0.0113	0.0104	2.121
		4	0.0139	0.0167	0.0102	0.0085	1.962
		5	0.0034	0.0113	-0.0006	0.0070	1.918
		6	0.0103	0.0109	0.0083	0.0068	1.902
		7	0.0196	0.0117	0.0120	0.0064	1.816
		8	0.0094	0.0102	0.0081	0.0056	1.801
		9	0.0111	0.0097	0.0092	0.0055	1.755
		10	0.0138	0.0079	0.0081	0.0051	1.556
		25	0.0123	0.0055	0.0055	0.0050	1.117

Table 4.3 Simulation results for $\hat{\beta}_u$ and $\hat{\beta}_g$ for 500 simulated samples of 100 clusters of size n with $\theta = 0.25$.

Table 4.4 shows the behavior of estimated correlation among martingales $\hat{\rho}_M$ as cluster size, n changes. 500 simulated data sets with 100 clusters of size n were generated. In the table, n changes from 2 to 25 with a single dichotomous covariate for an association parameter $\theta = 0.25$, corresponding to correlation among failure times 0.937. P_5 is the censoring rate with combination of random and Type I censoring given that the study is ended at 5 time units. P_5 is set at different rates changing from 0 to 50%. Estimated correlation is the average of 500 estimates of the correlation. Empirical variances are calculated from a sample of 500 estimates. Simulation results show that the estimated correlations among martingales are lower than the correlations among failure times. The difference is smaller as the cluster size is larger and the censoring rate is lower. The empirical variances of $\hat{\rho}_M$ is smaller as the cluster size is larger.

β	P_5	n	estimated correlation	empirical variance
0.693	0	2	0.9120	0.0007
		3	0.9145	0.0004
		4	0.9164	0.0003
		5	0.9197	0.0003
		6	0.9199	0.0003
		7	0.9205	0.0003
		8	0.9205	0.0003
		9	0.9206	0.0003
		10	0.9207	0.0003
		25	0.9226	0.0002
	20%	2	0.7498	0.0035
		3	0.7540	0.0018
		4	0.7543	0.0013
		5	0.7540	0.0010
		6	0.7541	0.0009
		7	0.7545	0.0008
		8	0.7546	0.0007
		9	0.7562	0.0006
		10	0.7567	0.0005
		25	0.7568	0.0003

Table 4.4 Simulation results for estimated correlation among martingales for 500 simulated samples of 100 clusters of size n with $\theta = 0.25$, corresponding to correlation among failure times 0.937.

4.5 Continuous covariates

In the previous two sections, the covariate in the simulation results is a binary variable. In this section, a single continuous covariate is considered. The covariate values are randomly generated from either a beta(1,3) distribution and or a uniform(0.1) distribution. The data generation procedure follows the description of simulation study in Chapter 2. For the censoring time, only Type II censoring is used for simplicity.

4.5.1 Beta distribution

In this section, simulation results with a single continuous covariate randomly generated from a beta(1,3) distribution are displayed. Figure 4.2 shows the probability density function of this distribution.

The parameters used in the simulation studies in this section are listed in Table 4.5.1. For each simulation study, 500 simulated data sets with each set containing 100 clusters of size 2 are used. Four Type II censoring rates, three levels of within cluster association, and three different values of the regression coefficient are considered. Each dataset produces one value of $\hat{\beta}_u$ and one value of $\hat{\beta}_g$.

parameters	function	possible values
β	true coefficient	0.693, 0, -0.5
P	censoring rate (%)	0, 10, 20, 50
θ	association parameter	0.25, 0.8, 1.5
K	number of clusters	100
n	cluster size	2
N	number of simulated data sets	500

Table 4.5 Parameters of simulation studies

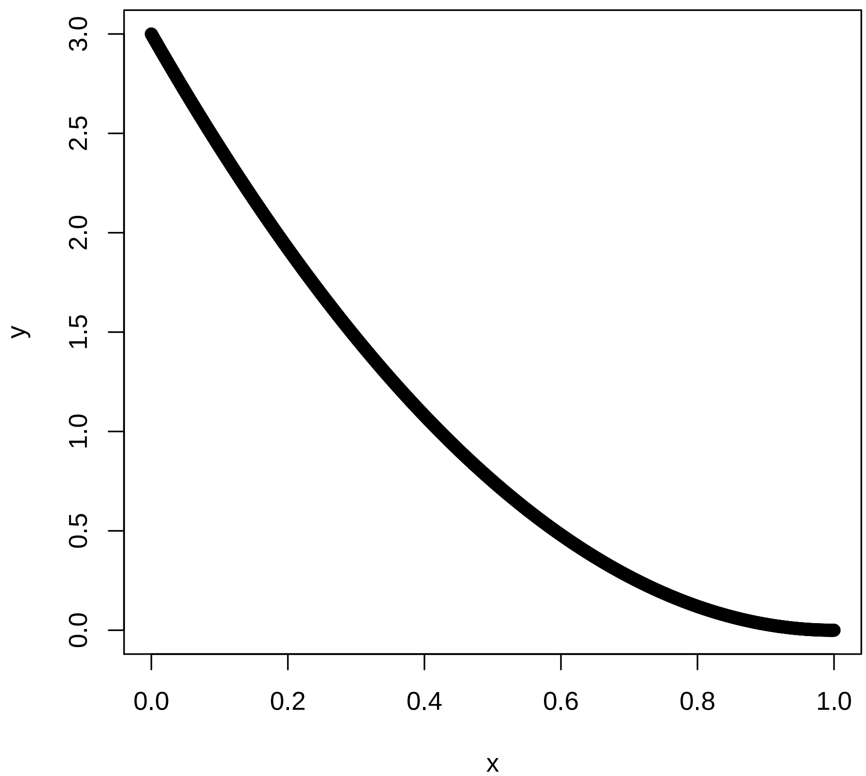

Figure 4.2 The probability density function of a beta(1,3) distribution

In Table 4.6, the true coefficient β assumes values 0.693, 0 and -0.5, and the association parameter θ assumes values 0.25, 0.80 and 1.50, corresponding to the within correlations 0.937, 0.712 and 0.512. P is the censoring rate with Type II censoring. P is set at different rates changing from 0 to 50%. There is no obvious advantage of using generalized estimating equations with respect to reducing the bias for any combination of parameter values, but $\hat{\beta}_g$ provides gains in efficiency relative to $\hat{\beta}_u$ for all the paramater settings we studied. The gains are more prounced for values of β closer to zero, especially when there is no censoring and the within cluster correlation is strong.

Compared to the simulation results with a single binary covariate in Chapter 3, there tend to be greater gains in efficiency when a continuous covariate is involved. As discussed earlier, if a balanced randomized design is used, $\hat{\beta}_u$ and $\hat{\beta}_g$ are equivalent. Gains in efficiency occur when there are unbalanced treatment assignments within a cluster. Within cluster imbalance tends to be greater for randomly generated value of a continuous covariate, since each subject is assigned to different covariate values.

β	P	θ	$\hat{\beta}_u$ estimated bias	$\hat{\beta}_u$ empirical variance	$\hat{\beta}_g$ estimated bias	$\hat{\beta}_g$ empirical variance	$SRE(\hat{\beta}_u \mid \hat{\beta}_g)$
0.693	0	0.25	0.0170	0.1450	0.0289	0.0275	5.280
		0.80	0.0029	0.1415	0.0009	0.0738	1.917
		1.50	−0.0004	0.1406	−0.0052	0.1037	1.356
	10%	0.25	0.0159	0.1479	0.0166	0.0334	4.422
		0.80	0.0095	0.1482	0.0033	0.0866	1.711
		1.50	0.0054	0.1518	−0.0001	0.1202	1.263
	20%	0.25	0.0114	0.1649	0.0047	0.0450	3.667
		0.80	0.0131	0.1624	0.0066	0.1044	1.555
		1.50	0.0091	0.1616	0.0036	0.1323	1.221
	50%	0.25	−0.0033	0.2520	0.0017	0.1142	2.208
		0.80	−0.0035	0.2555	−0.0056	0.1996	1.280
		1.50	−0.0007	0.2413	−0.0075	0.2169	1.112
0	0	0.25	−0.0076	0.1463	−0.0089	0.0220	6.654
		0.80	−0.0081	0.1406	−0.0099	0.0723	1.945
		1.50	−0.0089	0.1391	−0.0115	0.1024	1.358
	10%	0.25	−0.0019	0.1474	−0.0008	0.0239	6.166
		0.80	0.0034	0.1489	−0.0027	0.0835	1.783
		1.50	0.0025	0.1498	−0.0058	0.1177	1.273
	20%	0.25	0.0059	0.1726	0.0004	0.0404	4.272
		0.80	0.0041	0.1719	−0.0013	0.1046	1.644
		1.50	0.0080	0.1696	0.0015	0.1378	1.231
	50%	0.25	−0.0078	0.2672	0.0010	0.1143	2.339
		0.80	−0.0255	0.2696	−0.0231	0.1937	1.392
		1.50	−0.0236	0.2614	−0.0268	0.2125	1.230
-0.5	0	0.25	−0.0375	0.1536	−0.0474	0.0270	5.697
		0.80	−0.0254	0.1443	−0.0297	0.0770	1.874
		1.50	−0.0160	0.1423	−0.0178	0.1076	1.322
	10%	0.25	−0.0118	0.1530	−0.0181	0.0329	4.656
		0.80	−0.0010	0.1551	−0.0102	0.0896	1.731
		1.50	−0.0025	0.1541	−0.0106	0.1232	1.251
	20%	0.25	−0.0025	0.1756	−0.0075	0.0432	4.066
		0.80	0.0030	0.1750	−0.0029	0.1101	1.589
		1.50	0.0022	0.1759	−0.0037	0.1441	1.221
	50%	0.25	−0.0146	0.2867	0.0031	0.1287	2.228
		0.80	−0.0315	0.2836	−0.0256	0.2179	1.233
		1.50	−0.0297	0.2917	−0.0307	0.2472	1.074

Table 4.6 Simulation results for $\hat{\beta}_u$ and $\hat{\beta}_g$ from 500 simulated samples of 100 clusters of size 2 with a continuous covariate randomly generated from a beta(1,3) distribution.

Robust and bootstrap variance estimates are examined in Table 4.7. The average robust variance is the average of 500 robust variance estimates for the IWM estimator. The variance ratio for $\hat{\beta}_u$ is the average robust variance estimate divided by the empirical variance of $\hat{\beta}_u$. A value of the variance ratio of $\hat{\beta}_u$ greater than one indicates that the robust variance tends to overestimate the true variance of $\hat{\beta}_u$. A value of the variance ratio of $\hat{\beta}_u$ less than one indicates the robust variance estimate tends to underestimate the variance of $\hat{\beta}_u$. Variance estimates for $\hat{\beta}_g$ are obtained from a bootstrap resampling method. For clustered data, the bootstrap method resamples clusters instead of individual subjects. In this study, 100 bootstrap samples were taken from each of 500 simulated data sets to produce the bootstrap variance estimates. The average bootstrap variance is the average of 500 bootstrap variance estimates of $\hat{\beta}_g$. The variance ratio for $\hat{\beta}_g$ is the average bootstrap variance divided by the empirical variance of $\hat{\beta}_g$. A value of the variance ratio greater than one indicates that the bootstrap variance tends to overestimate the variance of $\hat{\beta}_g$. A value of the variance ratio $\hat{\beta}_g$ less than one indicates that the bootstrap variance tends to underestimate the variance of $\hat{\beta}_g$. The results provided in Table 4.7 suggest that robust variance estimator tends to slightly underestimate the variance of $\hat{\beta}_u$ and the bootstrap variance estimator tends to slightly overestimate the variance of $\hat{\beta}_g$.

β	P	θ	$\hat{\beta}_u$ empirical variance	$\hat{\beta}_u$ average robust variance	$\hat{\beta}_u$ variance ratio	$\hat{\beta}_g$ empirical variance	$\hat{\beta}_g$ average bootstrap variance	$\hat{\beta}_g$ variance ratio
0.693	0	0.25	0.1450	0.1314	0.906	0.0275	0.0300	1.092
		0.80	0.1415	0.1327	0.938	0.0738	0.0849	1.151
		1.50	0.1406	0.1335	0.950	0.1037	0.1160	1.119
	10%	0.25	0.1479	0.1459	0.986	0.0334	0.0367	1.099
		0.80	0.1482	0.1456	0.982	0.0866	0.0988	1.140
		1.50	0.1518	0.1457	0.960	0.1202	0.1302	1.083
	20%	0.25	0.1649	0.1627	0.987	0.0450	0.0457	1.016
		0.80	0.1624	0.1618	0.996	0.1044	0.1180	1.130
		1.50	0.1616	0.1614	0.999	0.1323	0.1482	1.120
	50%	0.25	0.2520	0.2528	1.003	0.1142	0.1307	1.145
		0.80	0.2555	0.2495	0.977	0.1996	0.2203	1.104
		1.50	0.2413	0.2482	1.028	0.2169	0.2482	1.144
0	0	0.25	0.1463	0.1295	0.885	0.0220	0.0229	1.042
		0.80	0.1406	0.1315	0.935	0.0723	0.0835	1.154
		1.50	0.1391	0.1329	0.956	0.1024	0.1145	1.118
	10%	0.25	0.1474	0.1479	1.004	0.0239	0.0250	1.050
		0.80	0.1489	0.1482	0.996	0.0835	0.0968	1.160
		1.50	0.1498	0.1481	0.989	0.1177	0.1302	1.107
	20%	0.25	0.1726	0.1663	0.964	0.0404	0.0462	1.114
		0.80	0.1719	0.1668	0.970	0.1046	0.1201	1.149
		1.50	0.1696	0.1661	0.980	0.1378	0.1516	1.100
	50%	0.25	0.2672	0.2694	1.008	0.1143	0.1366	1.195
		0.80	0.2696	0.2698	1.001	0.1973	0.2192	1.132
		1.50	0.2614	0.2694	1.031	0.2125	0.2416	1.137
-0.5	0	0.25	0.1536	0.1351	0.880	0.0270	0.0296	1.097
		0.80	0.1443	0.1353	0.938	0.0770	0.0875	1.136
		1.50	0.1423	0.1358	0.955	0.1076	0.1176	1.093
	10%	0.25	0.1530	0.1539	1.006	0.0329	0.0356	1.084
		0.80	0.1551	0.1528	0.985	0.0896	0.1028	1.148
		1.50	0.1541	0.1526	0.990	0.1232	0.1355	1.100
	20%	0.25	0.1756	0.1749	0.996	0.0432	0.0435	1.007
		0.80	0.1750	0.1735	0.991	0.1101	0.1269	1.152
		1.50	0.1759	0.1731	0.984	0.1441	0.1592	1.105
	50%	0.25	0.2867	0.2871	1.001	0.1287	0.1524	1.184
		0.80	0.2836	0.2883	1.017	0.2179	0.2449	1.124
		1.50	0.2917	0.2886	0.989	0.2472	0.2635	1.066

Table 4.7 Simulation results of variance estimates for 500 simulated samples of 100 clusters of size 2 with a single covariate randomly generated from a beta(1,3) distribution. The number of bootstrap samples is 100 for each simulated sets. The variance ratio for $\hat{\beta}_u$ is the average robust variance divided by the empirical variance, and the variance ratio for $\hat{\beta}_g$ is the average bootstrap variance divided by the empirical variance.

Table 4.8 and 4.9 show coverage rates for nominal 95% confidence intervals and the lengths of nominal 95% confidence intervals of β_u and β_g. The confidence intervals for β_u are contructed as $\hat{\beta}_u \pm (1.96)S_{robust}$, where S_{robust} is the standard error of $\hat{\beta}_u$ obtained from the robust variance procedure. There are four bootstrap methods used to construct bootstrap confidence intervals of β_g, normal, quantile, basic and accelerated bias-corrected (BC_α). These methods are described by Davison and Hinkley (1997). The normal method is evaluated as $\hat{\beta}_g \pm (1.96)S_{boot}$, where S_{boot} is the bootstrapped standard error for $\hat{\beta}_g$. For b bootstrap samples, denote the bootstrap distribution of bootstrap estimates by $\hat{F}_b(x)$. The quantile confidence interval method uses the $\alpha/2$ and $1-\alpha/2$ quantiles of $\hat{F}_b(x)$ as the limits of the $1-\alpha$ level confidence interval. The basic confident interval method uses the upper quantile of the bootstrap distribution to calculate the lower confidence bound, and the lower quantile to calculate the upper confidence bound. The lower and upper endpoints using this method are $2\hat{\beta}_g - \hat{F}_b^{-1}(1-\alpha/2)$, and $2\hat{\beta}_g - \hat{F}_b^{-1}(\alpha/2)$, respectively. The accelated bias-corrected method obtains end points by inverting percentiles of the bootstrap distribution after adjusting for bias and acceleration shift. Denote the acceleration constant by a, and the normal cdf by $\Phi(x)$. Define $Z_0 = \Phi^{-1}(\hat{F}_b(\hat{\beta}_g))$, and let $Z_{\alpha/2}$ denote the $\alpha/2$-percentile of a standard normal distribution. Then, the α endpoints of the BC_a confidence interval are $\hat{F}_b^{-1}\left\{\Phi\left(Z_0 + \frac{Z_0 - Z_\alpha/2}{1-a(Z_0 - Z_{\alpha/2})}\right\}\right)$ and $\hat{F}_b^{-1}\left\{\Phi\left(Z_0 + \frac{Z_0 + Z_\alpha/2}{1-a(Z_0 + Z_{\alpha/2})}\right\}\right)$, respectively. Among all the bootstrap methods, the normal method is the simplest. From the results in Table 4.8, all the methods provide confidence interval with coverage rates around 95 percent. The normal method performs as well as the other bootstrapping method. The results in Table 4.9 show that the lengths of the bootstrap confidence intervals using $\hat{\beta}_g$ tend to be smaller than the lengths of robust confidence intervals using $\hat{\beta}_u$. The differences in the lengths of the confidences intervals are greatest when the within cluster correlation is stronger and the censoring rate is lower. Comparing all the bootstrapping methods, the normal method gives the most narrow confidence intervals.

β	P	θ	average coverage rate of confidence interval based on robust variance estimator for $\hat{\beta}_u$	average coverage rate of bootstrap confidence interval based on $\hat{\beta}_g$			
				normal	quantile	basic	BCa
0.693	0	0.25	0.938	0.946	0.944	0.956	0.956
		0.80	0.936	0.960	0.952	0.964	0.946
		1.50	0.950	0.966	0.966	0.970	0.952
	10%	0.25	0.948	0.952	0.960	0.950	0.954
		0.80	0.946	0.938	0.944	0.948	0.946
		1.50	0.938	0.942	0.942	0.948	0.940
	20%	0.25	0.948	0.940	0.958	0.948	0.946
		0.80	0.952	0.948	0.944	0.948	0.948
		1.50	0.944	0.946	0.944	0.954	0.950
	50%	0.25	0.950	0.954	0.958	0.968	0.952
		0.80	0.940	0.964	0.950	0.974	0.952
		1.50	0.948	0.944	0.950	0.958	0.946
0	0	0.25	0.924	0.982	0.958	0.982	0.974
		0.80	0.936	0.966	0.954	0.976	0.950
		1.50	0.942	0.964	0.950	0.974	0.964
	10%	0.25	0.942	0.970	0.958	0.980	0.968
		0.80	0.948	0.950	0.940	0.952	0.942
		1.50	0.948	0.944	0.944	0.956	0.942
	20%	0.25	0.940	0.958	0.946	0.964	0.948
		0.80	0.934	0.954	0.948	0.964	0.956
		1.50	0.938	0.948	0.942	0.962	0.948
	50%	0.25	0.948	0.962	0.954	0.970	0.944
		0.80	0.944	0.952	0.952	0.960	0.944
		1.50	0.942	0.948	0.944	0.962	0.946
-0.5	0	0.25	0.922	0.972	0.932	0.980	0.950
		0.80	0.930	0.964	0.948	0.974	0.946
		1.50	0.942	0.950	0.944	0.972	0.944
	10%	0.25	0.948	0.958	0.944	0.964	0.950
		0.80	0.952	0.952	0.946	0.964	0.950
		1.50	0.946	0.956	0.942	0.960	0.940
	20%	0.25	0.954	0.966	0.962	0.976	0.960
		0.80	0.950	0.956	0.944	0.972	0.948
		1.50	0.944	0.958	0.952	0.966	0.956
	50%	0.25	0.946	0.960	0.948	0.962	0.954
		0.80	0.950	0.942	0.944	0.948	0.940
		1.50	0.938	0.954	0.944	0.956	0.948

Table 4.8 Simulated coverage rates for nominal 95% confidence intervals based on 500 simulated samples with 100 clusters of size 2. The number of bootstrap samples is 100 for each simulated set. Covariate values are randomly generated from a beta(1,3) distribution.

β	P	θ	average length of confidence interval based on robust variance estimator for $\hat{\beta}_u$	average length of bootstrap confidence interval based on $\hat{\beta}_g$			
				normal	quantile	basic	BCa
0.693	0	0.25	1.407	0.678	0.712	0.712	0.689
		0.80	1.414	1.127	1.180	1.180	1.172
		1.50	1.420	1.320	1.382	1.382	1.382
	10%	0.25	1.484	0.750	0.841	0.841	0.826
		0.80	1.484	1.213	1.276	1.276	1.274
		1.50	1.485	1.397	1.466	1.466	1.467
	20%	0.25	1.570	0.837	0.882	0.882	0.873
		0.80	1.567	1.329	1.396	1.396	1.401
		1.50	1.565	1.493	1.543	1.543	1.546
	50%	0.25	1.961	1.396	1.468	1.468	1.468
		0.80	1.948	1.839	1.911	1.911	1.923
		1.50	1.943	1.952	2.040	2.040	2.043
0	0	0.25	1.397	0.593	0.629	0.629	0.601
		0.80	1.408	1.117	1.174	1.174	1.170
		1.50	1.417	1.311	1.382	1.382	1.380
	10%	0.25	1.496	0.700	0.737	0.737	0.730
		0.80	1.498	1.202	1.263	1.263	1.267
		1.50	1.499	1.398	1.471	1.471	1.471
	20%	0.25	1.589	0.822	0.866	0.866	0.866
		0.80	1.592	1.341	1.406	1.406	1.415
		1.50	1.590	1.511	1.589	1.589	1.587
	50%	0.25	2.024	1.422	1.498	1.498	1.493
		0.80	2.026	1.835	1.904	1.904	1.909
		1.50	2.025	1.927	2.019	2.019	2.029
-0.5	0	0.25	1.428	0.674	0.719	0.719	0.694
		0.80	1.430	1.143	1.208	1.208	1.200
		1.50	1.433	1.329	1.406	1.406	1.409
	10%	0.25	1.528	0.739	0.778	0.778	0.759
		0.80	1.523	1.241	1.302	1.302	1.304
		1.50	1.522	1.428	1.503	1.503	1.505
	20%	0.25	1.631	0.817	0.859	0.859	0.852
		0.80	1.625	1.379	1.443	1.443	1.445
		1.50	1.623	1.548	1.626	1.626	1.633
	50%	0.25	2.090	1.505	1.583	1.583	1.577
		0.80	2.095	1.829	1.972	1.972	1.975
		1.50	2.096	1.949	2.094	2.094	2.098

Table 4.9 Simulated lengths of confidence intervals for 500 simulated samples with 100 clusters of size 2. The number of bootstrap samples is 100 for each simulated set. Covariate values are randomly generated from a beta(1,3) distribution.

4.5.2 Uniform distribution

In this section, the effects of the distribution of a continuous covariate are discussed. A single covariate is randomly generated from a uniform(0,1) distribution. Only Type II censoring is considered. In each case, 500 data sets are generated, each containing 100 clusters of size 2. The true regression coefficient is $\beta = 0.693$. The association parameter θ assumes values 0.25, 0.80 and 1.50, corresponding to correlations 0.937, 0.712, and 0.512 respectively. P is the censoring rate with Type II censoring. P is set at different rates changing from 0 to 50%. From the results presented in Tables 4.10 through 4.13, we see similar trends in the relative efficiency of $\hat{\beta}_u$ and $\hat{\beta}_g$ as the association parameter and the censoring rate change. When the within cluster correlation is high, and the censoring rate is moderate, there are gains in efficiency of $\hat{\beta}_g$ relative to $\hat{\beta}_u$, but the magnitudes of the gains are different from what occurs when the covariate values are generated from a bernoulli or a beta distributions. Gains in relative efficiency are not as large as in the corresponding case where the covariate values are generated from a beta(1,3) distribution, because covariates values generated from the beta(1,3) distribution tend to create more imbalance within clusters due to the skewed distribution. The bootstrap variance estimator performs well. Coverage rates for nominal 95% confidence intervals are about 95 percent. When the within cluster correlation is strong, and the censoring rate is moderate, the bootstrap confidence intervals based on $\hat{\beta}_g$ are narrower than robust confidence interval based on $\hat{\beta}_u$.

			$\hat{\beta}_u$		$\hat{\beta}_g$		
			estimated	empirical	estimated	empirical	
β	P	θ	bias	variance	bias	variance	$SRE(\hat{\beta}_u\|\hat{\beta}_g)$
0.693	0	0.25	0.0220	0.0684	0.0264	0.0155	4.402
		0.80	0.0155	0.0668	0.0112	0.0396	1.686
		1.50	0.0119	0.0663	0.0062	0.0551	1.204
	10%	0.25	0.0158	0.0694	0.0111	0.0219	3.167
		0.80	0.0172	0.0721	0.0112	0.0512	1.407
		1.50	0.0113	0.0752	0.0092	0.0670	1.121
	20%	0.25	0.0187	0.0759	0.0174	0.0288	2.637
		0.80	0.0192	0.0761	0.0175	0.0591	1.287
		1.50	0.0153	0.0810	0.0157	0.0775	1.044
	50%	0.25	0.0105	0.1221	0.0074	0.0750	1.628
		0.80	0.0219	0.1275	0.0249	0.1178	1.081
		1.50	0.0204	0.1286	0.0226	0.1285	1.001

Table 4.10 Simulation results for $\hat{\beta}_u$ and $\hat{\beta}_g$ for 500 simulated samples with 100 clusters of size 2 with a continuous covariate randomly generated from a uniform(0,1) distribution.

			$\hat{\beta}_u$			$\hat{\beta}_g$		
β	P	θ	empirical variance	average robust variance	variance ratio	empirical variance	average bootstrap variance	variance ratio
0.693	0	0.25	0.0684	0.0626	0.915	0.0155	0.0156	1.006
		0.80	0.0668	0.0622	0.931	0.0396	0.0413	1.042
		1.50	0.0663	0.0622	0.938	0.0551	0.0547	0.992
	10%	0.25	0.0694	0.0706	1.017	0.0219	0.0220	1.052
		0.80	0.0721	0.0698	0.968	0.0512	0.0475	0.927
		1.50	0.0752	0.0695	0.924	0.0670	0.0608	0.907
	20%	0.25	0.0759	0.0781	1.028	0.0288	0.0291	1.010
		0.80	0.0761	0.0775	1.018	0.0591	0.0581	0.983
		1.50	0.0810	0.0771	0.951	0.0775	0.0705	0.909
	50%	0.25	0.1221	0.1239	1.014	0.0750	0.0688	0.917
		0.80	0.1275	0.1226	0.961	0.1178	0.1099	0.932
		1.50	0.1286	0.1225	0.952	0.1291	0.1219	0.944

Table 4.11 Simulation results for variance estimates for 500 simulated samples with 100 clusters of size 2 with covariate values randomly generated from a uniform(0,1) distribution. The number of bootstrap samples is 100 for each simulated sample. The variance ratio for $\hat{\beta}_u$ is the average robust variance estimates divided by the empirical variance, and the variance ratio for $\hat{\beta}_g$ is the average bootstrap variance divided by the empirical variance.

β	P	θ	average coverage rate of confidence interval based on robust variance estimator for $\hat{\beta}_u$	average coverage rate of bootstrap confidence interval based on $\hat{\beta}_g$			
				normal	quantile	basic	BCa
0.693	0	0.25	0.938	0.954	0.948	0.964	0.928
		0.80	0.954	0.938	0.950	0.944	0.934
		1.50	0.944	0.928	0.942	0.948	0.938
	10%	0.25	0.958	0.950	0.936	0.948	0.932
		0.80	0.940	0.940	0.926	0.952	0.936
		1.50	0.932	0.940	0.922	0.944	0.930
	20%	0.25	0.944	0.946	0.938	0.948	0.940
		0.80	0.932	0.956	0.940	0.964	0.948
		1.50	0.930	0.948	0.936	0.946	0.940
	50%	0.25	0.952	0.932	0.936	0.936	0.932
		0.80	0.948	0.928	0.916	0.952	0.922
		1.50	0.936	0.946	0.926	0.954	0.932

Table 4.12 Simulated coverage rates for nominal 95% confidence intervals for 500 simulated samples with 100 clusters of size 2. The number of bootstrap samples is 100 for each simulated sample. Covariate values are randomly generated from a uniform(0,1) distribution.

β	P	θ	average length of confidence interval based on robust variance estimator for $\hat{\beta}_u$	average length of bootstrap confidence interval based on $\hat{\beta}_g$			
				normal	quantile	basic	BCa
0.693	0	0.25	0.973	0.489	0.513	0.513	0.509
		0.80	0.971	0.786	0.826	0.826	0.820
		1.50	0.972	0.908	0.950	0.950	0.946
	10%	0.25	1.035	0.574	0.600	0.600	0.595
		0.80	1.030	0.844	0.880	0.880	0.881
		1.50	1.028	0.958	1.000	1.000	1.001
	20%	0.25	1.091	0.660	0.689	0.689	0.690
		0.80	1.087	0.934	0.980	0.980	0.980
		1.50	1.085	1.033	1.081	1.081	1.084
	50%	0.25	1.376	1.014	1.063	1.063	1.067
		0.80	1.368	1.289	1.343	1.343	1.350
		1.50	1.368	1.360	1.366	1.366	1.368

Table 4.13 Simulated lengths of confidence intervals for 500 simulated samples with 100 clusters of size 2. The number of bootstrap samples is 100 for each simulated sample. Covariate values are randomly generated from a uniform(0,1) distribution.

4.6 Multiple covariates

4.6.1 Derivation of the estimating equations with multiple covariates

Frequently there are other covariates of interest in addition to the treatment effect. The estimating equations developed in Chapter 3 can be generalized to the case with multiple covariates easily. Now denote Z_{kji} as the ith covariate of jth subject in kth cluster, where $i = 1, ..., p$. $\mathbf{Z}_{kj}$ is a covariate vector of $(Z_{kj1}, ..., Z_{kjp})$. The same example used in Chapter 2 and 3 is used again to illustrate the construction of the estimating equations.

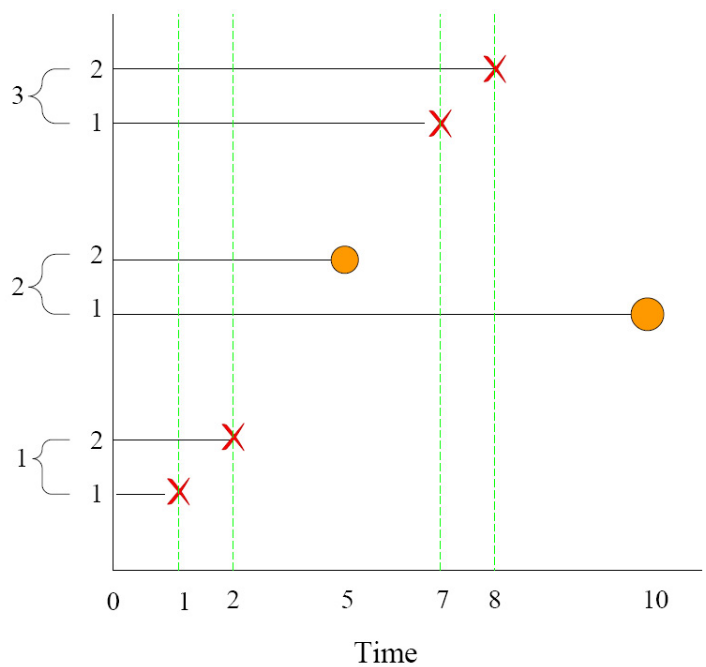

Figure 4.3 Illustration of generalized estimating equations

Only failure times are included in the equations. In this illustration, failures occur at times 1, 2, 7 and 8. The following illustrates the contribution of each subject to the estimating equations at time 1. Note that $D_{kji}(u) = Z_{kji}A_{kji}(u)$.

At time=1

$$\begin{pmatrix} D_{111}(1) & D_{121}(1) \\ \vdots & \vdots \\ D_{11p}(1) & D_{12p}(1) \end{pmatrix} \hat{W}_1(1) \begin{pmatrix} 1 - e^{\boldsymbol{\beta}'\mathbf{Z}_{11}} d\Lambda_0(1) \\ 0 - e^{\boldsymbol{\beta}'\mathbf{Z}_{12}} d\Lambda_0(1) \end{pmatrix}$$

$$\begin{pmatrix} D_{211}(1) & D_{221}(1) \\ \vdots & \vdots \\ D_{21p}(1) & D_{22p}(1) \end{pmatrix} \hat{W}_2(1) \begin{pmatrix} 0 - e^{\boldsymbol{\beta}'\mathbf{Z}_{21}} d\Lambda_0(1) \\ 0 - e^{\boldsymbol{\beta}'\mathbf{Z}_{22}} d\Lambda_0(1) \end{pmatrix}$$

$$\begin{pmatrix} D_{311}(1) & D_{321}(1) \\ \vdots & \vdots \\ D_{31p}(1) & D_{32p}(1) \end{pmatrix} \hat{W}_3(1) \begin{pmatrix} 0 - e^{\boldsymbol{\beta}'\mathbf{Z}_{31}} d\Lambda_0(1) \\ 0 - e^{\boldsymbol{\beta}'\mathbf{Z}_{32}} d\Lambda_0(1) \end{pmatrix}$$

Similarly, we can obtain the contribution to the estimating estimating equation of all the clusters at all the failure times. Summing the contributions over all the clusters and all failure times, the values of the estimating equations for a given value of $\hat{\beta}$ can be calculated.

4.6.2 Simulation results

Tables 4.14 and 4.15 display the results with two covariates for 500 simulated data sets with 100 clusters of size 2, where β_1 is the coefficient of a binary variable Z_1 taking a value 0 or 1, and β_2 is the coefficient of a continuous variable Z_2 randomly generated from a uniform(0,1) distribution. Z_1 and Z_2 are generated independently. The true values of the coefficients are $\beta_1 = 0.693$, and $\beta_2 = 0.05$. The association parameter θ assumes a value of 0.25, corresponding to a within cluster correlation 0.937. P is the censoring rate with Type II censoring. P is set changing from 0 to 20%. Let $\hat{\boldsymbol{\beta}}_u$ denote the vector consisting of $\hat{\beta}_{u1}$ and $\hat{\beta}_{u2}$, the estimators of the coefficients of binary and continous covariates using the partial likelihood score equations that ignore the within cluster correlation. Let $\hat{\boldsymbol{\beta}}_g$ denote the vector consisting of $\hat{\beta}_{g1}$ and $\hat{\beta}_{g2}$, the estimators of the coefficients of the binary and continous covariates from the generalized estimating equations..

In Table 4.14, a balanced randomized design is used to assign the binary effect. In this case, the estimator from generalized estimating equations $\hat{\beta}_{g1}$ does not provide any gain in efficient relative to the estimator from the partial likelihood score equations that ignore the within cluster correlation $\hat{\beta}_{u1}$. There is a gain in efficiency of $\hat{\beta}_{g2}$ relative to $\hat{\beta}_{u2}$. The gain is greater for no censoring than for 20% censoring rate. These results are consistent with the results for a single continous covariate. The bias of $\hat{\beta}_{g2}$ is about the half of the bias of $\hat{\beta}_{u2}$.

In Table 4.15, a completely randomized design is used to assign the levels of the binary variable. Then, both $\hat{\beta}_{g1}$ and $\hat{\beta}_{g2}$ provide gains in efficiency relative to $\hat{\beta}_{u1}$ and $\hat{\beta}_{u2}$. The gains are greater for no censoring than for 20% censoring. The bias of $\hat{\beta}_{g1}$ is smaller than the bias of $\hat{\beta}_{u1}$, and the bias of $\hat{\beta}_{g2}$ is about the half of the bias of $\hat{\beta}_{u2}$.

Comparing the result in Tables 4.14 and 4.15, the efficiency of the continuous variable coefficient estimator $\hat{\beta}_{g2}$ relative to $\hat{\beta}_{u2}$ is greater when a completely randomized design is used to assign the binary effect. The unbalanced assignment of the binary factor within a cluster increases the imbalance within a cluster, resulting in a large gain in efficiency for $\hat{\beta}_{g2}$ relative to $\hat{\beta}_{u2}$.

	$\boldsymbol{\beta}$	P	$\hat{\beta}_u$ estimated bias	$\hat{\beta}_u$ empirical variance	$\hat{\beta}_g$ estimated bias	$\hat{\beta}_g$ empirical variance	$SRE(\hat{\beta}_u \vert \hat{\beta}_g)$
β_1	0.693	0	0.0200	0.0061	0.0152	0.0060	1.020
β_2	0.050	0	-0.0013	0.0644	-0.0006	0.0184	3.485
β_1	0.693	20%	0.0190	0.0065	0.0172	0.0064	1.001
β_2	0.050	20%	0.0021	0.0721	0.0010	0.0356	2.025

Table 4.14 Simulation results for $\hat{\beta}_u$ and $\hat{\beta}_g$ of 500 simulated samples with 100 clusters of size 2, two covariates, and $\theta = 0.25$. The coefficient of the binary covariate is β_1, and a balanced randomized design is used to assign this effect. The coefficient of the continuous covariate is β_2, and the continuous covariate is randomly generated from a uniform(0,1) distribution.

	$\boldsymbol{\beta}$	P	$\hat{\beta}_u$ estimated bias	$\hat{\beta}_u$ empirical variance	$\hat{\beta}_g$ estimated bias	$\hat{\beta}_g$ empirical variance	$SRE(\hat{\beta}_u \vert \hat{\beta}_g)$
β_1	0.693	0	0.0290	0.0208	0.0205	0.0103	2.023
β_2	0.050	0	0.0026	0.0590	-0.0010	0.0125	4.690
β_1	0.693	20%	0.0300	0.0223	0.0232	0.0183	1.214
β_2	0.050	20%	0.0031	0.0612	0.0015	0.0228	2.683

Table 4.15 Simulation results for $\hat{\beta}_u$ and $\hat{\beta}_g$ for 500 simulated samples with 100 clusters of size 2, two covariates, and $\theta = 0.25$. The coefficient of the binary covariate is β_1, and a completely randomized design is used to assign this effect. The coefficient of the continuous covariate is β_2, and the continuous covariate is randomly generated from a uniform(0,1) distribution.

4.7 Discussion of correlation structure

4.7.1 Introduction

When cluster sizes are larger than 2, different correlation structures can be considered. In this section we discuss the effects of correlation structure on the estimators of regression coefficient for cases with varying cluster sizes. The exchangeable and first-order autoregressive correlation structures are considered. The exchangeable correlation structure assumes equal correlation between any two subjects in the same cluster. An autoregressive correlation structure may occur when responses within a cluster are taken over time. Responses sampled within a smaller time intervals may be more correlated than those within a larger time interval. The exchangeable correlation structure for cluster size of n is of the form

$$\begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \dots & \dots & \dots & \dots & \dots \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix}$$

The first-order autoregressive correlation structure for clusters of size n is of the form

$$\begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}$$

The performance of the estimators using the partial likelihood score equations that ignore within cluster dependence, the estimators using the generalized estimating equations with the exchangeable correlation structure, and estimators using the generalized estimating equations with the first-order autoregressive correlation structure is assessed by the simulation study. It is of interest to determine how much efficiency of the coefficient estimators is lost by using a simpler correlation structure. Therefore, data with within cluster correlation are generated with the first-order autoregressive correlation structure. Then the coefficient is estimated with exchangeable and first-order autoregressive correlation structures. Comparison of bias and efficiency are made. For the first-order autoregressive correlation structure, the first-order correlation ρ is estimated by averaging the first-order distance of martingales. For example, if data from clusters are arranged in the order of the level of assumed correlation from the strongest to the weakest. Then ρ is estimated by the average correlation of martingales between the first and second observations, second and third observations, third and four observations, etc.

In simulation studies by Cai and Prentice (1997), failure times were generated using the Copula model examined by Clayton and Cuzick (1985). To maintain the similarities with their simulation procedure, we used the same model in previous sections. Cai and Prentice only studied the case that failure times are correlated at the same dependence level. For clustered failure times that are correlated at different dependence level, e.g. a data set with first-order autoregressive correlation structure, a multivariate version of Copula model is complicated. For simplicity of data generation procedure, correlated survival times with first-order autoregressive correlation structure in this section are generated using a different procedure from the procedure described in Chapter 2.

The probability density function of marginal distribution of t_{kj} is

$$f(t_{kj}) = e^{\beta Z_{kj}} \exp(-t_{kj} e^{\beta Z_{kj}}).$$

Z_{kj} is a binary variable taking a value of 0 or 1. For completely randomized design, it is randomly assigned within clusters with probability 0.5. We first generate failure times with multivariate normal distribution with specified correlation structure. Then failure times from correlated uniform distribution are obtained by using CDF transformation. Finally, failure times from multivariate exponential distribution with a specified correlation structure can be obtained by inverse CDF method.

4.7.2 Simulation results

Tables 4.16 and 4.17 display simulation results for a single binary covariate for 500 simulated data sets with 100 clusters of size n. The true coefficient is $\beta = 0.693$, and the first-order correlation assumes a value of 0.9. It is the correlation among failure times of observation 1 and 2, observation 2 and 3, and etc. P is the censoring rate with Type II censoring. P is set changing from 0 to 20%. $\hat{\beta}_u$ is the estimator using the partial likelihood score equations that ignores within cluster dependence. $\hat{\beta}_{gexch}$ is the estimator using the generalized estimating equations with the exchangeable correlation structure. $\hat{\beta}_{gauto}$ is the estimator using the generalized estimating equations with the first-order autoregressive correlation structure. SRE of $\hat{\beta}_{gexch}$ is the empirical variance of $\hat{\beta}_u$ divided by the empirical variance of $\hat{\beta}_{gexch}$. SRE of $\hat{\beta}_{gauto}$ is the empirical variance of $\hat{\beta}_u$ divided by the empirical variance of $\hat{\beta}_{gauto}$. SRE of $\hat{\beta}_{gauto}$ vs $\hat{\beta}_{gexch}$ is the empirical variance of $\hat{\beta}_{gexch}$ divided by the empirical variance of $\hat{\beta}_{gauto}$.

In table 4.16, a balanced randomized design is used. For clusters of size n, the first two subjects within clusters are assigned to the treatment effect, and the rest of $n-2$ subjects are assigned to the control. In this case, if exchangeable correlation structure is used, generalized estimating equations produce the same results as partial likelihood score equations that ignore within cluster correlation (not shown). When first-order autoregressive correlation structure is used, there are gains in efficiency of $\hat{\beta}_{gauto}$ relative to $\hat{\beta}_u$. Gains in efficiency are greater for no censoring than for 20% censoring. Cluster size also affects gains in efficiency, but there is no obvious pattern.

		$\hat{\beta}_u$		$\hat{\beta}_{gauto}$		
P	n	estimated bias	empirical variance	estimated bias	empirical variance	SRE
0	3	0.0107	0.0074	0.0113	0.0070	1.046
	4	0.0139	0.0063	0.0117	0.0056	1.132
	5	0.0104	0.0062	0.0098	0.0050	1.249
	6	0.0073	0.0061	0.0068	0.0049	1.237
	7	0.0097	0.0060	0.0067	0.0042	1.458
	8	0.0071	0.0053	0.0091	0.0042	1.258
20%	3	0.0105	0.0082	0.0112	0.0081	1.014
	4	0.0095	0.0075	0.0121	0.0074	1.015
	5	0.0067	0.0074	0.0081	0.0069	1.085
	6	0.0136	0.0072	0.0091	0.0060	1.187
	7	0.0086	0.0067	0.0094	0.0059	1.125
	8	0.0102	0.0068	0.0111	0.0059	1.116

Table 4.16 Simulation results for $\hat{\beta}_u$, $\hat{\beta}_{gexch}$, and $\hat{\beta}_{gauto}$ for 500 simulated samples of 100 clusters of size 2 with $\beta = 0.693$ and $\rho = 0.9$. All clusters get the same treatment assignment.

In table 4.17, a completely randomized design is used. Treatment effects are randomly assigned within clusters. Both $\hat{\beta}_{gexch}$ and $\hat{\beta}_{gauto}$ can be more efficient than $\hat{\beta}_u$. Gains in efficiency are greater for no censoring than for 20% censoring. Comparing $\hat{\beta}_{gexch}$ and $\hat{\beta}_{gauto}$, $\hat{\beta}_{gauto}$ has greater efficiency gains relative to $\hat{\beta}_u$ than the gains of $\hat{\beta}_{gexch}$ for all the parameter settings we studied. When the cluster size increases, the difference between the efficiency gain of $\hat{\beta}_{gauto}$ and $\hat{\beta}_{gexch}$ becomes larger. The gains in efficiency of $\hat{\beta}_{gauto}$ are greater when a completely randomized design is used.

		$\hat{\beta}_u$		$\hat{\beta}_{gexch}$			$\hat{\beta}_{gauto}$			SRE of $\hat{\beta}_{gauto}$ vs $\hat{\beta}_{gexch}$
P	n	estimated bias	empirical variance	estimated bias	empirical variance	SRE	estimated bias	empirical variance	SRE	
0	3	0.0197	0.0179	0.0167	0.0068	2.594	0.0152	0.0065	2.750	1.060
	4	0.0062	0.0121	0.0121	0.0064	1.904	0.0133	0.0054	2.242	1.177
	5	0.0182	0.0116	0.0202	0.0062	1.878	0.0186	0.0052	2.235	1.190
	6	0.0123	0.0099	0.0153	0.0052	1.925	0.0133	0.0043	2.302	1.195
	7	0.0107	0.0092	0.0100	0.0042	2.190	0.0109	0.0035	2.612	1.197
	8	0.0163	0.0077	0.0117	0.0041	1.877	0.0123	0.0031	2.467	1.314
20%	3	0.0162	0.0184	0.0192	0.0090	2.050	0.0179	0.0083	2.219	1.082
	4	0.0103	0.0143	0.0091	0.0081	1.787	0.0090	0.0072	1.984	1.110
	5	0.0218	0.0119	0.0200	0.0075	1.593	0.0192	0.0066	1.801	1.130
	6	0.0104	0.0110	0.0098	0.0068	1.610	0.0108	0.0057	1.923	1.194
	7	0.0109	0.0098	0.0100	0.0067	1.461	0.0103	0.0056	1.750	1.198
	8	0.0152	0.0083	0.0158	0.0064	1.293	0.0142	0.0052	1.604	1.241

Table 4.17 Simulation results for $\hat{\beta}_u$, $\hat{\beta}_{gexch}$, and $\hat{\beta}_{gauto}$ for 500 simulated samples of 100 clusters of size 2 with $\beta = 0.693$ and $\rho = 0.9$. Treatments are assigned randomly within clusters.

4.8 Summary

In this chapter, we investigated several factors that affect the efficiency of regression parameters estimators obtained from generalized estimating equations, $\hat{\beta}_g$, relative to the efficiency of estimators obtained from the partial likelihood score equations that ignore within cluster correlation, $\hat{\beta}_u$. The factors considered include the number of clusters, cluster sizes, strength of within cluster correlation, censoring levels and the distribution of values of a single continuous covariate. We also generalized to cases with more than one covariate to assess the behaviors of the estimators. When the cluster size is more than two, the effect of correlation structure on the estimators was also examined.

The number of clusters influences relative efficiency, but there is no obvious pattern. Gains in efficiency tend to be smaller as the cluster size increases. When a single covariate is a continuous variable, $\hat{\beta}_g$ is more efficient than $\hat{\beta}_u$ for strong within correlations and moderate censoring rates. The relative efficiency of $\hat{\beta}_g$ increases as within cluster correlation increases and the censoring rate decreases. The magnitude of the gains in efficiency depends the true value of the regression parameter and the distribution of the covariate. Trends in relative efficiency are more pronounced for parameter values closer to zero. For the beta(1,3) and uniform(0,1) distributions we investigated, the gains in efficiency are larger with a continuous covariate than with a binary covariate for all three levels of association, and four different values of the censoring rate. The existence of a continuous variable increases the imbalance within clusters.

The bootstrapping resampling method was used to estimate variances of the generalized estimating equation estimators. The simulation results show that the bootstrap method provide reliable variance estimates. The coverage rates of nominal 95% confidence intervals are around 95%. Confidence intervals based on $\hat{\beta}_g$ are narrower than robust confidence intervals based on $\hat{\beta}_u$ when within cluster correlation is strong and the censoring rate is moderate. The simulation results for Cox model with one binary and one continuous covariate show similar trends in relative efficiency as levels of within cluster correlation and censoring rates change.

When cluster sizes are larger than two, only a single binary covariate case is considered. If all the clusters have the same covariate value assignment, generalized estimating equations and partial likelihood score equations that ignore within cluster correlation produce the same estimates if exchangeable correlation structure is used. However, when the covariate values are randomly assigned within clusters, the assumed correlation structure affects the gain in efficiency of $\hat{\beta}_g$ relative to $\hat{\beta}_u$. If the true correlation structure is first-order autoregressive, there are gains in efficiency even when the exchangeable correlation structure is used in the generalized estimating equations, but there are not as large as when the correct correlation structure is used. The difference in efficiency gains between the estimators using these two correlation structures is greater when the cluster size is larger.

CHAPTER 5. General summary

The analysis of clustered survival data is an important statistical question in medical research. A commonly used method obtains parameter estimates from the partial likelihood score equations based on a model that incorrectly assumes independent observations, i.e., the independent working model (IWM). A robust covariance estimator provides a consistent estimator of the covariance matrix of the parameter estimates. In the context of the Cox proportional hazards model, this estimation procedure is available in the coxph function in S-Plus, a function of the same name in the survival package for R, and in the PHREG procedure in SAS. The availability of the software has led to the wide use of this methodology for correlated survival data. The main deficiency of this methodology is the loss of efficiency when the within cluster correlation is strong. We investigated several methods to establish estimating equations by incorporating weight matrices to improve the efficiency of estimators of regression parameters.

Cai and Prentice (1997) proposed the weighted estimating equations by inserting weight matrices into partial likelihood score equations that ignore within cluster correlation. The method has not been widely adopted because of the complexity in estimating weight matrices, computationally intensity and unavailability in software. Therefore we suggested using simple correlation functions to estimate the correlations among martingales of observations within clusters. The weight matrices are the inverses of these correlation matrices. We studied the behavior of estimators using weights conditional on covariate pairs, and estimators using common weights evaluated by averaging across the covariate values. Simulation results show that the loss of efficiency by using common weights is small even when the survival times are correlated with different dependence levels. More importantly, common weighting can be extended to the cases with any type or any number of covariates. In practical examples, it is rare that the correlations are considered dependent on covariates. Therefore models with correlations conditional on covariate values were not further considered.

Another set of weighted estimating equations can be formed by inserting weight matrices into an equivalent form of partial likelihood score equations. $\hat{\beta}_u$, $\hat{\beta}_c$ and $\hat{\beta}_w$ are denoted as the solution to the partial likelihood score equations that ignore within cluster correlation, the Cai and Prentice estimating equations with modified weight matrices, and a new set of equation equations respectively. The behavior of the estimators was studied with respect to estimated bias and empirical variance.

In case with clusters of size two and a single binary covariate, when a balanced randomized design is used, $\hat{\beta}_u$, $\hat{\beta}_c$ and $\hat{\beta}_w$ are equivalent. There is no gain in efficiency by using weighted estimating equations. When a completely randomized design is implemented with a single binary covariate, $\hat{\beta}_c$ and $\hat{\beta}_w$ are equivalent. Simulation results show that there might be gains in efficiency of $\hat{\beta}_c$ or $\hat{\beta}_w$ relative to $\hat{\beta}_u$ by using weighted estimating equations. The gain is substantial for a low censoring rate, and high within cluster correlation. The gain is greater with smaller coefficient value. Bootstrapping method was used to estimate variances of estimated regression parameters. Simulation results showed that bootstrapping variances provide reliable estimates of variances for estimated regression parameters.

A new set of estimating equations using a generalized estimating equation approach by Liang and Zeger (1986) was also developed to improve the efficiency of the estimators of coefficients in Cox proportional hazard model. We consider partial likelihood equations for the Cox model in the counting process context. The counting process differentials can be approximated by Poisson random variables which have mean and variance of corresponding compensator differentials. Therefore generalized estimating equation approach could be applied to establish a new set of estimating equations for clustered survival data. The solution to this set of estimating equations is denoted by $\hat{\beta}_g$. We first considered the case of clusters of size two with a single binary covariate. The gains in efficiency of $\hat{\beta}_g$ relative to $\hat{\beta}_u$ are greater when within cluster correlation is stronger and the censoring rate is lower. The bootstrap variance estimates provide reasonable variance estimates of regression parameter estimates.

Additional simulation studies were examined to further investigate the behavior of $\hat{\beta}_g$. We discussed the effects of the number of clusters, cluster size, one single continuous covariate, multiple covariates, and correlation structures for cluster of size greater than two. The influence of the number of clusters on relative efficiency does not have obvious pattern. Cluster size also has influence on the gains in efficiency. Overall, the imbalance caused by randomized treatment assignment within clusters tends to diminish with larger cluster size. Thus lower gains in efficiency were observed with larger cluster size. For the continuous covariate, covariate values are randomly generated from a beta(1,3) or uniform(0,1) distribution. Conclusions with respect to gains in efficiency are similar to the binary covariate case. Compared to binary covariate, there tends to be a higher gain in efficiency when there is a continuous covariate. In addition, gains in efficiency of $\hat{\beta}_g$ for covariate values from a beta(1,3) distribution tends to have greater gains in efficiency than for covariate values from uniform(0,1) distribution. This is caused by the skewed distribution of beta(1,3). When there are two independent covariates, one binary and one continuous, both binary and continuous estimators have efficiency gain with a completely randomized design for the binary covariate when within cluster correlation is strong and the censoring rate is moderate. We also generated data assuming first-order autoregressive correlation structure. We found that high correlation leads to an increased gain in efficiency, even when the simple correlation matrix was used. When the correct correlation structure is used, there is more gain in efficiency.

The proposed estimating equations provide gains in efficiency of estimators relative to estimators from unweighted estimating equations when within cluster correlation is high and the censoring rate is low. Overall $\hat{\beta}_g$ tends to have slightly higher gains in efficiency relative to $\hat{\beta}_u$ than $\hat{\beta}_w$. Estimating equations discussed in Chapter 2 were set up from the partial likelihood perspective, whereas the estimating equations discussed in Chapter 3 were developed with generalized estimating equations framework. This allows people to investigate clustered survival data from different point of view. The asymptotic results of estimators from generalized estimating equations by Liang and Zeger (1986) might be further investigated to apply to clustered survival data.

We used bootstrap method to estimate the variances of the estimate of regression coefficients. Simulation results show that bootstrapping variances provide reliable estimates of the variances for parameter estimates. Binder (1992) proposed a method to estimate variance of weighted estimator using the idea of approximate jackknife variance estimates. More research could be done to compare the variance estimation of the two methods.

In this project, we discussed the Cox proportional hazards model using a common baseline hazard function. In some applications, it might be assumed that subjects within clusters have different baseline hazards. This type of proportional hazards model can be investigated. We used the exponential distribution as the marginal distribution of failure time. Some other marginal models can be simulated to examine relative efficiency of estimators produced by unweighted and weighted estimating equations. Also more complicated model with time dependent covariates can be studied. In addition a method to assess the selection of correlation structure needs to be developed.

All of the future work would require more software development. The package "survgee" developed by the author to implement the proposed methodology still needs a great deal more work. Right now the package can only handle the cases of clusters of equal sizes, and only allow exchangeable and first-order autoregressive correlation structure with time independent covariates. The more general correlation structure and covariate types need to be implemented, which will allow further investigation of the behaviors of the estimators.

APPENDIX Help file of the author defined package survgee

All the computations were implemented in an R package "survgee" developed by the author. The main fitting routine in the package is written in C. The development of the R package provides a simple way for users to apply the proposed methodology.

Description

The survgee package was developed to fit the weighted estimating equations presented in Chapter 2 or the generalized estimating equations in Chapter 3 and 4 to estimate coefficients in Cox proportional hazards models applied to clustered survival data. The Cox proportional hazards model considered is the common baseline hazard model that allows every subject in the study has the same common baseline hazard. This package allows single or multiple time independent covariates that are categorical or continuous, any number of clusters and cluster sizes as long as they are equal cluster sizes. The correlation structures can be exchangeable or first-order autoregressive. Also bootstrap method can be used to construct confidence intervals for the regression parameters using normal, basic, percentile and bias correction accelerated methods. The package cannot handle unequal cluster sizes, more general correlation structure or time dependent covariates.

Usage

```
survgee(formula, data, subset, eetype = "g", boot.ci = FALSE,
    boot.reps = 100, boot.type = c("norm", "basic", "perc", "bca"),
```

```
boot.conf = 0.95, prn.debug = c(FALSE, FALSE) )
```

Arguments

`formula` a formula object. If a formula object is supplied, it is must have a Surv object as the response with time and status in the form of Surv(time, status) on the left of the tilde operator and terms on the right. Besides the covariates, the terms must include cluster and subject specifications defined with cluster() and subject().

`data` a data frame that contains the data for the variables named in the formula.

`eetype` specify the type of the estimating equations. eetype="c", for the weighted estimating equations described in Chapter 2, eetype="g", for the generalized estimating equations described in Chapter 3 and 4. The default equation type is "g".

`corr.type` specify the type of correlation structure. corr.type="ex", for exchangeable correlation structure, corr.type="ar1" for first-order autoregressive correlation. The default correlation structure is exchangeable.

`boot.ci` a logical parameter. If TRUE a bootstrap variance estimate is computed. Default is FALSE.

`boot.reps` number of bootstrap samples. The default value is 100.

`boot.types` method used to construct a bootstrap confidence interval for regression parameters using normal, basic, quantile, and bias correction accelerated method. The default methods calculate bootstrap confidence interval using all four methods.

`boot.conf` a confidence level for constructing confidence intervals. The default value is 0.95.

`prn.debug` a logical vector of length 2. The default is a vector of c(FALSE, FALSE). If the first component is TRUE, print out the contribution of each cluster to the estimating equations at each failure time. If the second component is TRUE, print out the martingale residuals for each subject at the corresponding failure time.

Value

`times` the vector of times including failure and censoring times

`covar` a vector of covariates

`means` the mean object created by coxph.details in the survival package. It is a vector of length of number of failure times. It is the weighted mean at certain time t, E(t) defined in Chapter 2

`lp` a vector of linear predictors

`status` a vector of status of all times. status=1 for a failure time, status=0 for a censoring time

`nXvars` the number of covariates

`npergrp` a vector of cluster sizes

`maxGrpN` the maximum cluster size

`hz` the hazard increment created by coxph.details

`ngrps` the number of clusters

`grps` a vector of cluster numbers ordered by times

`subNo` a vector of subject numbers within clusters ordered by time `maxtime` a vector of maximum time rank within clusters

`cox likelihood` the estimated regression parameters from coxph

the squared values of the estimating equations. It should be close to zero at the solution of the estimating equations.

`martingale`

the matrix of martingale residuals

`corM` the estimated correlation matrix

`red` the estimated weight matrix. It is the inverse of the estimated correlation matrix

`coef` the estimating equations estimates of regression parameters. The first row is the estimated coefficient using weighted or generalized estimating equations depending on the choice of eetype. The second row is the coefficient estimate using the independent working model provided by coxph.

Examples

```
A data frame named dataall is of the form

group subNo z      obsT      status

1      1  0  0.999027682      1

2      1  0  0.302858919      1

3      1  0  0.367302563      1

4      1  0  1.306125223      1

5      1  0  0.702722135      1

....

1      2  0  0.927935066      1

2      2  0  0.348090959      1
```

```
3     2  1  0.146980806     1
4     2  0  1.675034510     1
5     2  1  0.209513154     1
...
```

It contains 100 clusters of size 2. The true coefficient value is 0.693, the association parameter of 0.25 corresponding to correlation 0.937. No censoring is considered. There is a single binary covariate randomized within clusters.

```
surv.fit<- survgee( Surv( obsT, status ) ~ z +  cluster(group) +
           subject(subNo) , data = dataall, corr.type=``ex'',
           eetype= ``g'', boot.ci=TRUE, boot.reps=100 )
```

Part of the output is presented

```
$corrM
          [,1]      [,2]
[1,] 1.0000000 0.9254744
[2,] 0.9254744 1.0000000

$red
          [,1]      [,2]
[1,]  6.968776 -6.449424
[2,] -6.449424  6.968776


$coef
             z
```

```
beta 0.9297654
     0.9074665


$formula
Surv(obsT, status) ~ z + cluster(group) + subject(subNo)


$boot.ci
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 100 bootstrap replicates


CALL :
boot.ci(boot.out = bootout, conf = boot.conf, type = boot.type)


Intervals :
Level      Normal              Basic
95%   ( 0.6666,  1.2171 )   ( 0.6426,  1.2031 )


Level     Percentile            BCa
95%   ( 0.6564,  1.2169 )   ( 0.7003,  1.3132 )


Bootstrap Statistics :
     original    bias        std. error
t1* 0.9297654 -0.01209328   0.1404333
```

BIBLIOGRAPHY

Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes: A larger sample study. *The Annals of Statistics*, 10:1100–1120.

Bailey, K. (1983). The asymptotic joint distribution of regression and survival parameter estimates in the cox model. *The Annals of Statistics*, 11:39–48.

Binder, D. (1992). Fitting cox's proportional hazards models from survey data. *Biometrika*, 79(1):139–147.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 30:89–99.

Cai, J. and Prentice, R. (1995). Estimating equations for hazard ratio parameters based on correlated failure data. *Biometrika*, 82:151–164.

Cai, J. and Prentice, R. (1997). Regression estimation using multivariate failure time data and a common baseline hazard function model. *Lifetime Data Analysis*, 3:197–213.

Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65:141–151.

Clayton, D. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *Journal of the Royal Statistical Society, Series B*, 148:82–117.

Cox, D. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society. Series B*, 34(2):187–220.

Cox, D. (1975). Partial likelihood. *Biometrika*, 62:269–276.

Davison, A. C. and Hinkley, D. (1997). *Bootstrap methods and their application*. Cambridge University Press, London.

Efron, B. (1977). The efficiency of cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72:557–565.

Emura, T. and Tsukuma (2003). Hypothesis testing based on the maximum of two statistics from weighted and unweighted estimating equations. *Technical Reports of Mathematical Sciences, Chiba University*, 19.

Fleming, T. and Harrington, D. (1991). *Counting Processes and Survival Analysis*. John Wiley, New York.

Gray, R. (2000). Estimation of regression parameters and the hazard function in transformed linear survival models. *Biometrics*, 56:571–576.

Gray, R. (2003). Weighted estimating equations for linear regression analysis of clustered failure time data. *Lifetime Data Analysis*, 9:123–138.

Hougaard, P. (1995). Frailty models for survival data. *Lifetime data analysis*, 1:255–273.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer, New York.

Huber, P. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium*, 221-233.

Huster, W., Brookmeyer, R., and Self, S. (1989). Modeling paired survival data with covariates. *Biometrics*, 45:145–156.

Kalbfleisch, J. D. and Prentice, R. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika*, 60:267–278.

Koehler, K. and Symanowski, J. (1995). Constructing multivariate distributions with specific marginal distributions. *Journal of Multivariate Analysis*, 55:261–282.

Lee, E., Wei, L., and Amato, D. (1992). *Cox-Type Regression Analysis for Large Numbers of Small Groups of Correlated Failure Time Observations*. Kluwer Academic Publishers.

Liang, K.-Y. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrics*, 73(1):13–22.

Lipsitz, S., Laird, N., and Harrington, D. (1990). Using the jackknife to estimate the variance of regression estimators from repeated meausres studies. *Communications in Statistics–Theory and Methods*, 19:821–845.

Loughin, T. and Koehler, K. (1997). Bootstrapping regression parameters in multivariate survival analysis. *Lifetime Data Analysis*, 3:157–177.

Louis, T. (1981). Nonparametric analysis of an accelerated failure time model. *Biometrika*, 68:381–390.

McCullagh, P. and Nelder, J. (1983). Quasi-likelihood functions. *The Annals of Statistics*, 11:59–67.

Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society, Series B*, 44(2):414–422.

Peto, R. (1972). Comment on "regression model and life tables". *Journal of the Royal Statistical Society, Series B*, 34:205.

Prentice, R. and Cai, J. (1992). Covariance and survival function estimation using censored multivariate failure time data. *Biometrika*, 79:495–512.

Prentice, R. and Gloeckler, L. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 34:57–67.

Rebolledo, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. verw. Gebiete*, 51:269–286.

Royall, R. M. (1986). Model robust confidence intervals using maximum likelihood estimators. *International statistical review*, 54:221–226.

Therneau, T. M. and Grambsch, P. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer, New York.

Tsiatis, A. (1981). A large sample study of Cox's regression model. *The Annals of Statistics*, 9:93–108.

Tsiatis, A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics*, 18:354–372.

Vaupel, J., Manton, K., and Stallard, E. (1979). The impact of heterogeneity in individual frailty and the dynamics of mortality. *Demography*, 16:439–447.

Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61:439–447.

Wei, L., Lin, D., 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.