Statistical Methods on Survival Data with Measurement Error

by

Ying Yan

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2014

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In survival data analysis, covariates are often subject to measurement error. A naive analysis with measurement error ignored commonly leads to biased estimation of parameters of survival models. Measurement error also causes efficiency loss for detecting possible association between risk factors and time to event. Furthermore, it induces difficulty on model building and model checking, because the presence of measurement error frequently masks true underlying patterns of data.

Although there has been a large body of literature to handle error-prone survival data since the paper by Prentice (1982), many important issues still remain unexplored in this area. This thesis focuses on several important issues of survival analysis with covariate measurement error.

One problem that has received little attention is on misspecification of measurement error models. In this thesis, we investigate this important problem with the attention particularly paid to error-contaminated survival data under the Cox model. In particular, we conduct bias analysis which offers a way to unify many existing methods of survival data with measurement error, and study the impact of misspecifying the error models in survival data analysis. A simple expression is obtained to quantify the bias of "working" estimators derived under misspecified error models. Consistent estimators under general error models are derived based on this simple expression. Furthermore, we study hypothesis testing with both model misspecification and measurement error present.

A second problem of our interest is about the validity of survival model assumptions when measurement error is involved. In the literature, a large number of methods have been developed to correct for measurement error effects, and these methods basically assume the survival model to be the Cox model. When the Cox model or the error model assumptions fail to hold, existing methods would break down. In this thesis, we address the issue of checking the Cox model assumptions with measurement error. We propose valid goodness of fit tests for survival data with covariate measurement error. This research offers us an important addition to the literature of survival data with measurement error.

Our third topic concerns survival data analysis under additive hazards models with covariate measurement error. The additive hazards model is a useful and important alternative to the Cox model. However, this model is relatively less studied for situations where covariates are measured with error. In this thesis, we make important contributions to this topic. Specifically, we explore asymptotic bias induced from ignoring measurement error. A number of inference methods are developed to correct for error effects. The validity of the proposed methods is justified both theoretically and empirically. We investigate issues of model checking and model misspecification as well.

In many studies, collection of data often includes a large number of variables in which many of them are unimportant in explaining survival of an individual. An important task is thus to identify relevant risk factors which are linked to the hazards of subjects. Although there is work on variable selection for survival data analysis, the available methods typically require all variables be precisely measured. This requirement is, however, often infeasible. More challengingly, in some studies, the dimension of the risk factors can be quite large or even much larger than the size of subjects. Our fourth topic concerns about estimation and variable selection for survival data with high dimensional mismeasured covariates. We propose corrected penalized methods. Our methods can adjust for measurement error effects, and perform estimation and variable selection simultaneously. Our research on this topic closes multiple gaps among the areas of survival analysis, measurement error and variable selection.

Acknowledgements

Foremost, I would like to express my deepest gratitude to my Ph.D. supervisor Dr. Grace Y. Yi, for her guidance throughout my Ph.D. program, for her kindness, and for her insight of statistics. I am so grateful to her for bringing me to the field of biostatistics. She stimulated my curiosity, opened my mind, and always encouraged me to explore unknown territories of statistics. Without her continuous support in the past four years, it would not have been possible for me to write this thesis. I am fortunate that I have had a great supervisor.

I thank Dr. Richard Cook, Dr. Jerry Lawless, Dr. Mei-Cheng Wang (Johns Hopkins University), and Dr. Sharon Kirkpatrick (School of Public Health and Health Systems at the University of Waterloo) for serving as my committee members.

I also wish to thank the department faculty members and staff. Special thanks go to Mary Lou Dufton, for her willingness to help and excellent administrative support.

I owe many thanks to my classmates and friends, who have made my life as a Ph.D. student colorful and enjoyable.

I would like to express my heartfelt thanks to my family, who are always standing by me, and motivating me to pursue a Ph.D. degree.

Dedication

To Min, for her love, understanding, endless support and encouragement.

Contents

List of Tables **xviii**

List of Figures **xx**

1 Introduction and Overview **1**

1.1 Survival Data Analysis 2

1.1.1 Assumptions and Notation 3

1.1.2 Survival Models and Inference Functions 4

1.1.3 Model Misspecification 8

1.1.4 Model Checking 10

1.2 Measurement Error Models 10

1.2.1 Measurement Error Models 11

1.2.2 Error Mechanisms 12

1.2.3 Data Sources 13

1.3 Impact of Measurement Error on Parameter Estimation 14

1.4 Review of Existing Methods on Survival Data with Measurement Error . . 18

1.4.1 Regression Calibration and Simulation Extrapolation 18

1.4.2 Likelihood-Based Methods 20
1.4.3 Estimating Equation Methods 22
1.4.4 Additional Literature Review 24
1.5 Outline of the Thesis 26

2 Corrected Profile Likelihood for Cox Model with Covariate Measurement Error 29

2.1 Introduction 29
2.2 Notation and model setup 31
2.2.1 Cox model 32
2.2.2 Measurement error models 34
2.3 Brief review of existing methods 36
2.3.1 Corrected score approach by Nakamura (1992) and Song and Huang (2005) 36
2.3.2 Semiparametric regression method by Hu and Lin (2002,2004) . . . 37
2.3.3 Nonparametric correction method by Huang and Wang (2000) . . . 38
2.3.4 Corrected likelihood method by Yi and Lawless (2007) 39
2.4 Corrected profile likelihood approach 39
2.4.1 Method for Scenario A 39
2.4.2 Method for Scenario B 43
2.4.3 Method for Scenario C 44
2.4.4 Application to Berkson error model 46
2.5 Empirical studies 47
2.5.1 General error model 47

2.5.2 Application to Berkson error model 49

2.5.3 An example 50

2.6 Discussion 52

3 Analysis of Survival Data with Covariate Error under Possibly Misspecified Error Models 69

3.1 Introduction 69

3.2 Notation and Model Setup 70

3.2.1 Cox Model 71

3.2.2 Measurement Error in Survival Data 71

3.3 Hypothesis Testing under Correctly Specified Measurement Error Model . 73

3.4 Inference under Misspecified Measurement Error Model 78

3.4.1 Estimation 79

3.4.2 Hypothesis Testing 81

3.5 Numerical Studies 83

3.5.1 Parameter Estimation 83

3.5.2 Hypothesis Testing 85

3.6 An example 86

3.7 Discussion 87

4 Model Checking for the Cox Model with Measurement Error 125

4.1 Introduction 125

4.2 Cox Model and Model Checking 126

4.3 Measurement Error Models and Estimation 128

4.3.1 Scenario 1 : Additive Error Model with Replicates 128
4.3.2 Scenario 2 : Additive Error Model with Known Parameters 130
4.4 Model Checking Procedures with Error in Covariates 131
4.4.1 Model Checking under Measurement Error Scenario 1 131
4.4.2 Model Checking under Measurement Error Scenario 2 134
4.5 Numerical Studies 137
4.5.1 Simulation Studies 137
4.5.2 An Example 138
4.6 Discussion 139

5 A Class of Functional Methods for Error-Contaminated Survival Data under Additive Hazards Models with Replicate Measurements 151

5.1 Introduction 151
5.2 Notation and Model Setup 153
5.2.1 Additive Hazards Model 153
5.2.2 Estimation in the Absence of Measurement Error 154
5.2.3 Measurement Error Model 155
5.3 Asymptotic Bias Analysis 156
5.4 Corrected Pseudo Score Approach 158
5.5 Estimating Equation Approach 162
5.6 Empirical Studies 164
5.6.1 Design of Simulation 164
5.6.2 Performance of Estimators 165

5.6.3 Impact of the Number of Replicates 165
5.6.4 Results on Cumulative Hazard Function 166
5.7 Model Misspecification and Model Checking 167
5.7.1 Model Misspecification 167
5.7.2 Model Checking 168
5.8 ACTG 175 Study 172
5.9 Extension and Discussion 173

6 Estimation and Variable Selection for High Dimensional Additive Hazards Regression with Covariate Measurement Error 205

6.1 Introduction 205
6.2 Notation and Model Setup 208
6.2.1 Penalized Methods 208
6.2.2 Penalties 209
6.2.3 Measurement Error Model 210
6.3 Corrected Penalized Methods for High Dimensional Scenario 211
6.4 Corrected Penalized Methods for Ultra-high Dimensional Scenario 214
6.5 Numerical Studies 216
6.5.1 Simulation Studies 216
6.5.2 Real Data Analysis 218
6.6 Discussion 220

7 Summary and Discussion 229

References 231

List of Tables

2.1 Simulation results under Scenario A 64

2.2 Simulation results under Scenario B 65

2.3 Simulation results under Scenario C 66

2.4 Simulation results under the Berkson error model 67

3.1 Simulation results with misspecified error variance 117

3.2 Simulation results for different misspecified error distributions 124

5.1 Comparison of the performance of various estimators 201

5.2 Performance of estimators when $n_i = 1$ for some subjects 202

5.3 Empirical coverage rate (in percent) of confidence bands with nominal level 0.95 203

5.4 Empirical size and empirical power of the proposed test statistic 203

5.5 Analyse of the ACTG 175 dataset using different methods 204

6.1 Simulation results for Scenario 1: $n = 200, p = 50$; values inside the brackets are standard deviations 227

6.2 Simulation results for Scenario 2: $n = 400, p = 800$ 227

6.3 Parameter estimation ($\times 1000$). YanYi in the second column indicates the corrected estimator by Yan and Yi (2014b); values inside the brackets in the second column are the difference between the naive estimator by Lin and Ying (1994) and the corrected estimator by Yan and Yi (2014b); values inside the brackets in the third to sixth columns are the difference between the naive penalized estimator by Lin and Lv (2013) and the proposed estimator; #S is the number of nonzero estimated parameters. 228

List of Figures

1.1 An illustration of the impact of measurement error on parameter estimation . . 15

2.1 Point estimation of β and the corresponding confidence interval by the corrected profile likelihood method under the error model (2.11). Vertical lines show the confidence intervals by the estimator of Hu and Lin (2004) under the error model (2.13), which is a special case of (2.11). 63

3.1 ARE of the corrected score and log rank tests compared to the true score test 112

3.2 Case 1: True error model is $W_i = X_i + K\epsilon_i$. 113

3.3 Case 2: True error model is $W_i = \gamma_x X_i + \epsilon_i$. 114

3.4 Power plot under correctly specified error model 115

3.5 Power plot under misspecified error model 116

3.6 Case S.1: True error model is $W_i = X_i + K\epsilon_i$. 118

3.7 Case S.2: True model is $W_i = \gamma_x X_i + \epsilon_i$. 119

3.8 Case S.3: True model is $W_i = \gamma_x X_i + K\epsilon_i$. 120

3.9 Case S.4: True model is the Berkson error model $X_i = W_i + \epsilon_i$. 121

3.10 Case S.5: The error distribution is misspecified. 122

3.11 Case S.6: The error distribution is misspecified. 123

5.1 Asymptotic relative bias of naive estimator $\hat{\beta}_{nv}$ with different replicates numbers n_i 200

Chapter 1

Introduction and Overview

Survival data analysis deals with time from a time origin to occurrence of some event (or endpoint). The time origin can be the date of birth, the disease onset, and the time of entry to a randomized clinical trial. Examples of endpoints include time to death, time to disease relapse, and time to failure of some component of a machine. A major goal of the statistical analysis of survival data aims at assessing the association of the failure time and risk factors (e.g., Kalbfleisch and Prentice 2002; Lawless 2003). For example, detection of treatment effects on survival is a main concern in many randomized clinical trials, and this is often accomplished by building an appropriate survival model which adjusts for treatment and other risk factors, such as age, sex and measure of blood pressure. Survival analysis provides tools to describe the trend of the risk of failure, estimate the frequency of occurrence of events, and predict the chance of failure given specific levels of risk factors.

In clinical trials and observational studies, some risk factors (or covariates) are often measured with error. Examples of error-prone risk factors are the CD4 lymphocyte counts in AIDS studies, blood pressure in coronary heart disease studies, and energy intake in nutrition studies. A naive analysis with measurement error ignored commonly leads to biased estimation of parameters of survival models as well as efficiency loss for detecting possible association of risk factors and time to event. In the following sections, we discuss these issues and review some literature on this topic. A less explored problem concerns model building and model checking. Model building based on the mismeasured version of the true data, though relatively straightforward, is lack of interpretability and of little interest; on the other hand, model building based on the true but unobserved data is more difficult, because measurement error tends to mask the pattern of the data. There is little literature, if any, that provides valid model checking techniques applicable to survival models with covariate measurement error. Most existing model checking techniques require covariates to be precisely measured. Naively applying those checking procedures with measurement error ignored is generally not feasible to evaluate survival model assumptions.

Although many problems remain unexplored, in this thesis, we focus on several particular topics, such as developing useful tools to handle survival models with covariate measurement error, studying the impact of measurement error on inference as well as the impact of misspecifying measurement error models, and providing appropriate procedures to check survival models with covariate measurement error. Our discussion also provides insights into the connections of several existing methods for survival data with mismeasured covariates. Before we present our work, we provide an overview of survival data analysis with measurement error in the remainder of this chapter.

In Section 1.1, we review statistical analysis of survival data and several important survival models, and describe some model checking techniques. In Section 1.2, we introduce measurement error models and measurement error mechanisms. In Section 1.3, we study the impact of measurement error on parameter estimation. In Section 1.4, we give a general review of existing methods that handle survival models with covariate measurement error. We conclude this chapter with an outline of subsequent chapters.

1.1 Survival Data Analysis

In Section 1.1.1, we define basic notation and introduce the concept of censorship. In Section 1.1.2, we introduce several important survival models. In Section 1.1.3, we describe the consequence of misspecifying the survival models, thus suggesting the necessity of model checking. In Section 1.1.4, we briefly review some model checking techniques for survival models.

1.1.1 Assumptions and Notation

A unique feature of survival data is that the observations of failure time may be incomplete for various reasons: (i). subjects may survive during the study period or may be lost to followup; (ii). we may know that the subject fails prior to some time point (or during some time period), but the exact failure time is unknown; (iii). a subject who is eligible for the study may not survive prior to the beginning of the study and thus never enters the study. The incomplete patterns in (i), (ii), and (iii) are called *right censoring*, *left censoring* (or *interval censoring*), and *left truncation*, respectively. We refer to Lawless (2003, Ch 2) for a detailed discussion of censoring and truncation. In this thesis, we focus on right censored data.

For $i = 1, \cdots, n$, let T_i be the failure time and C_i be the right censoring time. Let $Z_i(t)$ be a vector of *external* covariates (Kalbfleisch and Prentice 2002) for subject i. The $\{T_i, C_i, Z_i(t)\}$ are assumed to be independent, $i = 1, \cdots, n$. Suppose all the individuals are observed over a common time interval $[0, \tau]$, where $0 < \tau < \infty$. Let $S_i = \min(T_i, C_i)$, and $\delta_i = I(T_i \leq C_i)$. Let $N_i(t) = I(S_i \leq t, \delta_i = 1)$ be the number of observed failures for the ith subject up to and including time t, and $Y_i(t) = I(S_i \geq t)$ indicate whether the ith subject is at risk of failure at time t^-. Let $\mathcal{F}_t = \sigma\{N_i(s), Y_i(s^+), Z_i(s), 0 \leq s \leq t, i = 1, \cdots, n\}$ be the $\sigma-$ field generated by the observed event and covariates histories prior to time t for all subjects.

The right censoring scheme is called *independent* if for any time point t,

$$\lim_{\Delta t \to 0} \frac{\Pr\{t \leq T_i < t + \Delta t | T_i \geq t, C_i \geq t, Z_i(t)\}}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Pr\{t \leq T_i < t + \Delta t | T_i \geq t, Z_i(t)\}}{\Delta t}.$$

Several special cases of independent censoring includes: (i). *random censoring*, where C_i and T_i are independent given $Z_i(t)$; (ii). *type I censoring*, where $C_i \equiv c$ is a constant; and (iii). *type II censoring*, where the study is stopped when a given number of failures are observed.

Let $\lambda(t; Z_i(t))$ be the *hazard function* for subject i with covariates $Z_i(t)$, given by

$$\lambda(t; Z_i(t)) = \lim_{\Delta t \to 0} \frac{\Pr\{t \leq T_i < t + \Delta t | T_i \geq t, Z_i(t)\}}{\Delta t}.$$

Thus $\lambda(t; Z_i(t))dt = \Pr\{d\tilde{N}_i(t) = 1 | T_i \geq t, Z_i(t)\}$, where $\tilde{N}_i(t) = I(T_i \leq t)$, and $dA(t)$ represents $A(t^- + dt) - A(t^-)$ for a process $A(t)$. Since $\{dN_i(t) = 1\} = \{d\tilde{N}_i(t) = 1, Y_i(t) = 1\} = \{t \leq T_i < t + dt, Y_i(t) = 1\}$, the independent censoring assumption is equivalent to $\Pr\{dN_i(t) = 1 | \mathcal{F}_{t-}\} = Y_i(t)\lambda(t; Z_i(t))dt$ for each time t. Throughout this thesis, we assume the independent censoring mechanism. We also assume *noninformative* censoring in the sense of Lawless (2003, Ch 2.2.2).

1.1.2 Survival Models and Inference Functions

The hazard function $\lambda(t; Z_i(t))$ is often modeled to feature the relationship between the survival time and covariates. In this section, we introduce several important models.

Cox Model

The *Cox model* (Cox 1972) specifies that covariates have multiplicative effects on the hazard function. A most appeal of such models is that the baseline hazard function can be left unspecified when basing inference about covariate effects on the *partial likelihood* (Cox 1975). To be specific, the Cox model assumes that the hazard function of T_i is related to $Z_i(\cdot)$ through

$$\lambda(t; Z_i(t)) = \lambda_0(t) \exp(Z_i^T(t)\beta), \tag{1.1}$$

where $\lambda_0(\cdot)$ is the baseline hazard function, and β is the regression parameter. Let $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ be the *baseline cumulative hazard function*. Inference about the regression parameter β is typically based on the partial likelihood:

$$L_p(\beta) = \prod_{i=1}^{n} \left[\frac{\exp\{Z_i^T(s_i)\beta\}}{\sum_{\{j:s_j \geq s_i\}} \exp\{Z_j^T(s_i)\beta\}} \right]^{\delta_i}. \tag{1.2}$$

Maximizing $L_p(\beta)$ with respect to β leads to the partial likelihood estimator $\hat{\beta}$ of β. Alternatively, $\hat{\beta}$ can be obtained by solving the partial score function

$$\begin{aligned} U_p(\beta) &= \sum_{i=1}^{n} \delta_i \left[Z_i(s_i) - \frac{\sum_{\{j:s_j \geq s_i\}} Z_j(s_i) \exp\{Z_j^T(s_i)\beta\}}{\sum_{\{j:s_j \geq s_i\}} \exp\{Z_j^T(s_i)\beta\}} \right] \\ &= \sum_{i=1}^{n} \int_0^{\tau} \left[Z_i(t) - \frac{\sum_{j=1}^{n} Y_j(t) Z_j(t) \exp\{Z_j^T(t)\beta\}}{\sum_{j=1}^{n} Y_j(t) \exp\{Z_j^T(t)\beta\}} \right] dN_i(t). \end{aligned} \tag{1.3}$$

The partial likelihood method is advantageous in that the baseline hazard function $\lambda_0(t)$ is left unspecified, thus protecting us from obtaining invalid results about β when $\lambda_0(t)$ is mismodeled.

Additive Hazards Model

The *additive hazards model* (Breslow and Day 1980; Cox and Oakes 1984; Lin and Ying 1994) assumes that covariates act on the hazard function via an additive form:

$$\lambda(t; Z_i(t)) = \lambda_0(t) + Z_i^T(t)\beta,$$

where $\lambda_0(t)$ is the baseline hazard function, and β is the regression parameter.

Estimation of β can be carried out using the pseudo score function (Lin and Ying 1994):

$$U(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \{Z_i(t) - \bar{Z}(t)\} dM_i(t),$$

where $\bar{Z}(t) = \sum_{i=1}^{n} Y_i(t) Z_i(t) / \sum_{i=1}^{n} Y_i(t)$, and $M_i(t; \beta, \Lambda_0) = N_i(t) - \int_0^t Y_i(u)\{d\Lambda_0(u) + \beta^T Z_i(u) du\}$. Note that $M_i(t; \beta, \Lambda_0)$ is an $\mathcal{F}_t$-adapted martingale. By that $E[U(\beta_0)] = 0$, where β_0 is the true value of β, solving $U(\beta) = 0$ leads to a consistent estimator $\hat{\beta}$ of β, given by

$$\hat{\beta} = \left[\sum_{i=1}^{n} \int_0^{\tau} Y_i(t)\{Z_i(t) - \bar{Z}(t)\}^{\otimes 2} dt \right]^{-1} \left[\sum_{i=1}^{n} \int_0^{\tau} \{Z_i(t) - \bar{Z}(t)\} dN_i(t) \right].$$

Since $E[M_i(t; \beta_0, \Lambda_0)]$=0 by martingale properties (e.g., Kalbfleisch and Prentice 2002), a Breslow-type cumulative hazard estimator is obtained by solving

$$\sum_{i=1}^{n} M_i(t; \beta, \Lambda_0) = 0.$$

That is,

$$\hat{\Lambda}(t; \hat{\beta}) = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{j=1}^n Y_j(u)} - \int_0^t \frac{\sum_{i=1}^n Y_i(u) Z_i^T(u) \hat{\beta} du}{\sum_{j=1}^n Y_j(u)}.$$

Accelerated Failure Time Model

The *accelerated failure time model* (AFT) (Cox 1972) assumes that covariates have multiplicative effects on T_i, or equivalently, additive effects on the natural log of T_i, named $Y_i = \log(T_i)$:

$$Y_i = \beta_0 + Z_i^T \beta_1 + \sigma e_i,$$

where covariates Z_i are time-invariant. The distribution of the error term e_i, or equivalently the error density $f_0(e)$, can be modelled either parametrically or nonparametrically. When $f_0(e)$ is modelled parametrically, inference on $\beta = (\beta_0^T, \beta_1^T)^T$ can be based on the parametric likelihood function

$$L(\beta) = \prod_{i=1}^{n} \left[\frac{1}{\sigma} f_0 \left(\frac{y_i - \beta_0 - Z_i^T \beta_1}{\sigma} \right) \right]^{\delta_i} \left[S_0 \left(\frac{y_i - \beta_0 - Z_i^T \beta_1}{\sigma} \right) \right]^{1-\delta_i},$$

where $S_0(e) = \int_e^{\infty} f_0(u) du$. Several classes of classic parametric survival models, including exponential model, Weibull model, log-normal model and log-logistic model, can be regarded as special cases of parametric accelerated failure time models. Lawless (2003, Ch 5) provides a comprehensive treatment of parametric accelerated failure time models.

When $f_0(e)$ is modelled nonparametrically, the accelerated failure time model is a semi-parametric alternative to the Cox model, and inferences become more difficult. We refer the readers to Kalbfleisch and Prentice (2002, Ch 7) and Lawless (2003, Ch 8) for details.

Proportional Odds Model and Transformation Model

The three survival models we introduced above are perhaps the most important and popular choices in survival data analysis. These models have different advantages and strengthes, and corresponding parameter interpretations could be substantially different. Recently, the *proportional odds model* (Pettitt 1982; Bennett 1983) and its natural generalization, the *transformation model* (Cuzick 1988; Cheng, Wei and Ying 1995), started to attract increased attention. The proportional odds model assumes that

$$-\text{logit}\{S(t; Z_i)\} = h_0(t) + Z_i^T\beta,$$

where $\text{logit}(x) = \log\{x/(1-x)\}$, $S(t; Z_i) = \Pr(T_i \geq t|Z_i) = \exp\{-\int_0^t \lambda(t; Z_i)dt\}$ is the *survivor function*, $h_0(t)$ is an unspecified increasing function, and Z_i is time-independent. The proportional odds model is a special case of the parametric transformation model

$$g\{S(t; Z_i)\} = h_0(t) + Z_i^T\beta,$$

where $g(\cdot)$ is a known increasing function. Note that the Cox model can be rewritten as

$$\log[-\log\{S(t; Z_i)\}] = h_0(t) + Z_i^T\beta,$$

where $h_0(t) = \log\{\Lambda_0(t)\}$, and thus the Cox model with time independent covariates is also a special case of the parametric transformation model. When the function $g(\cdot)$ is left unspecified, then the transformation model is equivalent to

$$h_0(t) = Z_i^T\beta + \epsilon_i,$$

where ϵ_i is a random error with an unknown distribution. Inference procedures for this linear transformation is referred to Cheng, Wei and Ying (1995), among others.

While we have discussed several survival models, in this thesis, we will mainly focus on the Cox model and the additive hazards model. In Chapters 2, 3 and 4, we consider Cox models, and in Chapters 5 and 6, we restrict our attention to additive hazards models.

1.1.3 Model Misspecification

The seminal paper by White (1982) originally studied the impact of model misspecification on maximum likelihood estimators (MLE) of parametric models. The working MLE of a possibly misspecified parametric model converges to a certain limit that minimizes the *Kullback-Leibler information criterion* (Kullback and Leibler 1951), which measures the “distance” between the true model and a parametric working family. If the true model is contained in the parametric working family, then the limit of the working MLE is identical to the true value of the parameter of the true model; otherwise, this limit differs from the true parameter of the underlying model, and thus the working MLE is not a consistent estimator. The degree of bias depends on the “distance” between the parametric working family and the true model.

When full distributional models are not available, inferences are often based on marginal features, such as mean and variance, of distributions. In such situations, unbiased estimating functions are often invoked. Yi and Reid (2010) studied misspecified estimating functions with finite dimensional parameters. They showed how to derive a consistent estimator of the true parameter based on biased estimating functions.

Now we turn to the issues of misspecifying the Cox model. The Cox model is a semiparametric model, thus the theory of White (1982) is not readily applied. Solomon (1984), O’Neill (1986), and Struthers and Kalbfleisch (1986) provided semiparametric generalization of the theory of White (1982) from parametric models to the Cox model. They investigated the effect of wrongly using the Cox model, when the true model may be others (e.g., the accelerated failure time model). They found that the relative importance of the covariates is unchanged even when the Cox model is wrongly used, but the parameter estimation is considerably biased. Lin and Wei (1989) extended the results of Struthers and

Kalbfleisch (1986); they established asymptotic normality of the pseudo partial likelihood estimator and provided a robust variance estimator. Li and Ryan (2004) argued that the naive inference procedure of the Cox model with measurement error can be viewed as a type of model misspecification.

Model misspecification also has a significant impact on hypothesis testing. Lagakos (1988a, 1988b) investigated the phenomenon of loss of efficiency of score test induced by model misspecification. Kong and Slud (1997) proposed a robust log-rank test, and DiRienzo and Lagakos (2001a, 2001b) investigated the bias of score tests under misspecified Cox models and provided modifications. Xu and O'Quigley (2000), Boyd, Kittelson and Gillen (2012), and Hattori and Henmi (2012) proposed interpretable estimators for treatment effects even when the Cox model is misspecified.

Sometimes some important covariates may be omitted or have misspecified functional forms. Struthers and Kalbfleisch (1986) showed that simply ignoring a covariate in the Cox model biases estimation of the corresponding covariate coefficient. However, the naive test of no effect of risk factors on survival is still valid under some conditions of covariates and censoring. The impact of omitting covariates was also studied by Gail, Wieand and Piantadosi (1984), Lagakos and Schoenfeld (1984), Morgan (1986), Bretagnolle and Huber-Carol (1988), Gail, Tan and Piantadosi (1988), Lin and Wei (1989), and Anderson and Fleming (1995). Gerds and Schumacher (2001) investigated the problem of misspecifying the functional forms of the covariates.

Studies of model misspecification are not restricted to the Cox model. Hattori (2006, 2012) studied the impact of misspecifying the additive hazards model and the accelerated failure time model, respectively. In particular, he investigated the impact on tests of no treatment effects. Clegg et al. (2000), Kosorok, Lee and Fine (2004), Boher and Cook (2006), and Latouche et al. (2007) studied more complex event history models.

1.1.4 Model Checking

In this section, we restrict our attention to the Cox model. There have been a large body of model checking techniques for the Cox model since the early 1980's (Schoenfeld 1980, 1981; Wei 1984; Barlow and Prentice 1988; Therneau, Grambsch and Fleming 1990; Lin 1991; Lin and Wei 1991; Lin, Wei and Ying 1993; Grambsch and Therneau 1994; Grambsch, Therneau and Fleming 1995). Comprehensive reviews are referred to Therneau and Grambsch (2000) and Lawless (2003).

Lin, Wei and Ying (1993) proposed an omnibus test to check misspecification of the Cox model using martingale residuals. Similar test statistics could be used to check the proportional hazards assumption, the functional form of a covariate, and the exponential link function. The model checking procedures by Lin, Wei and Ying (1993) have been widely used and further developed by Spiekerman and Lin (1996) for marginal Cox models with multivariate failure times, by Lin and Spiekerman (1996) for parametric survival models, including accelerated failure time models, by Kim, Song and Lee (1998) for additive hazards models, by Yin (2007) for marginal additive hazards models with multivariate failure times, by Lin, Wei, and Ying (2002) for longitudinal data analysis, and by Pipper and Ritz (2007) for grouped data under the Cox model, among others.

1.2 Measurement Error Models

When covariates are measured with error, characterizing the association of mismeasured covariates and the underlying covariates is necessary for valid inference; different measurement error mechanisms require different adjustments of naive inference procedures to account for error effects. In this section, we introduce three widely used measurement error models, distinguish two measurement error mechanisms, and address four data sources that help build error models. For a comprehensive overview, we refer to Carroll et al. (2006).

1.2.1 Measurement Error Models

In the covariate vector $Z_i(t) = (X_i^T, V_i^T(t))^T$, we let X_i represent time-independent but error-prone covariates, and $V_i(t)$ be covariates that are precisely measured and possibly time-dependent. Suppose X_i is not observed, but its surrogate measurement W_i is collected.

Classical Additive Model

The *classical additive measurement error model* assumes that

$$W_i = X_i + \epsilon_i, \tag{1.4}$$

where the ϵ_i are independent and identically distributed (i.i.d.) with mean 0 and a positive-definite variance matrix Σ_0, and are independent of X_i. Often, a multivariate normal distribution is assumed for ϵ_i.

Berkson Model

The *Berkson measurement error model* has the form

$$X_i = W_i + \epsilon_i,$$

where the ϵ_i are independent and identically distributed (i.i.d.) with mean 0 and a positive-definite variance matrix, and are independent of W_i. Often, a multivariate normal distribution is assumed for ϵ_i.

Multiplicative Model

The *multiplicative measurement error model* is given by

$$W_i = X_i \epsilon_i,$$

where the ϵ_i are independent and identically distributed (i.i.d.) with mean 0 and a positive-definite variance matrix Σ_0, and are independent of X_i.

The classical additive model (1.4) is perhaps the most popular error model, especially in modelling covariate measurement error in survival data. However, it is important to note that the choice of an error model is determined by the data at hand. See Carroll et al. (2006) for details. When discrete variables are subject to error, it is usually called a *misclassification* problem, and error modelling strategies are different from the three models introduced. Buonaccorsi (2010) provided a detailed treatment for misclassification problems.

1.2.2 Error Mechanisms

Two error mechanisms: *nondifferential error mechanism* and *differential error mechanism* are often distinguished in survival analysis with covariate measurement error.

Nondifferential error mechanism occurs when W_i is independent of the underlying failure time T_i and censoring time C_i given the true covariates $(X_i^T, V_i^T(\cdot))^T$. Equivalently, the nondifferential error mechanism means that the distribution of T_i and C_i given X_i, W_i and $V_i(\cdot)$, does not depend on W_i:

$$f(t_i, c_i|x_i, w_i, v_i(\cdot)) = f(t_i, c_i|x_i, v_i(\cdot)), \tag{1.5}$$

where t_i, c_i, x_i, w_i and $v_i(\cdot)$ are realized values of T_i, C_i, X_i, W_i and $V_i(\cdot)$, respectively. Thus, the surrogate W_i is noninformative in that it does not contribute information about T_i and C_i if X_i were known. Measurement error is *differential* if (1.5) is not true. In this thesis, we assume the nondifferential error mechanism, as consistent with the treatment done by most authors.

1.2.3 Data Sources

We restrict our discussion to the classical additive error model (1.4). If the observed data only consist of $\{T_i, C_i, W_i, V_i(\cdot), i = 1, \cdots, n\}$ and the error distribution is unknown, then the error model (1.4) is not *identifiable* (Fuller 1987). Without additional data sources, one needs to make a distributional assumption on the error ϵ_i, and the corresponding parameters of this distribution require to be known or estimated by external data. The distributional assumption may be restrictive, and usually *sensitivity analysis* is required to assess the bias on estimation if the assumption is violated. In analysis with measurement error models, additional data sources are often needed. Here we briefly discuss common types of those data sources.

Replicated Measurements

Replication data assumes that X_i is repeatedly measured for n_i times ($n_i > 1$), resulting in the surrogates $W_{ir}, r = 1, \cdots, n_i$. The classical additive error model (1.4) becomes

$$W_{ir} = X_i + \epsilon_{ir},$$

where the ϵ_{ir} are independent and identically distributed (i.i.d.) with mean 0 and a positive-definite variance matrix Σ_0, and are independent of $N_i(\cdot), Y_i(\cdot)$, and $Z_i(\cdot)$, $i = 1, \cdots, n; r = 1, \cdots, n_i$. With the replicates W_{ir}, the covariance matrix Σ_0 of ϵ_{ir} can be consistently estimated by

$$\hat{\Sigma}_0 = \sum_{i=1}^{n} \sum_{r=1}^{n_i} (W_{ir} - \bar{W}_{i\cdot})^{\otimes 2} / \sum_{i=1}^{n} (n_i - 1),$$

where $a^{\otimes 2} = aa^T$ for a column vector a, and $\bar{W}_{i\cdot} = \sum_{r=1}^{n_i} W_{ir}/n_i$.

Validation Subsample

When a *validation subsample* is available, both measurements of X_i and W_i are available within this subsample. With validation data available, the measurement error problem can be treated as a *missing data* problem as well. The complete structure of measurement error can be examined by the validation sample, and thus measurement error is relatively easy to handle in this case. Validation data can be either *internal* or *external*. Distinctions of internal and external data may be found in Carroll et al. (2006, Ch 2.3).

Instrumental Variable

Sometimes a second measurement of X_i, say $\tilde{X}_i$, is available from another measurement method. This variable is called an *instrumental variable*. To be an instrument, $\tilde{X}_i$ needs to be uncorrelated with T_i, C_i, and W_i, given $X_i, V_i(\cdot)$. For the usage of the instrumental variable, we refer to Carroll et al. (2006, Ch 6).

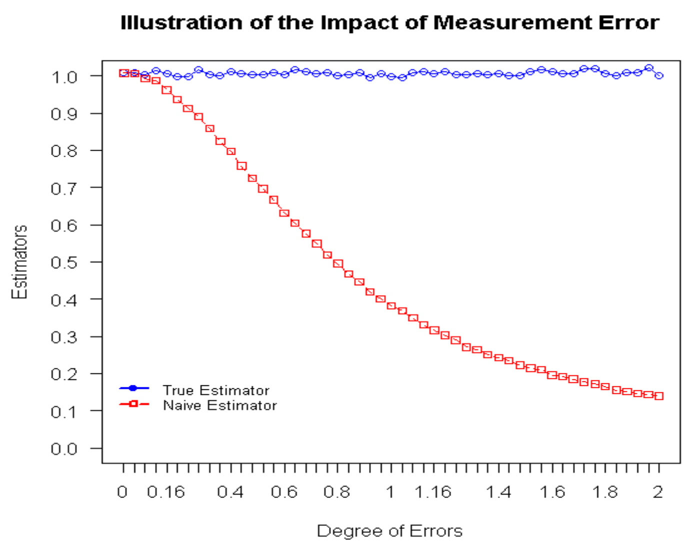
1.3 Impact of Measurement Error on Parameter Estimation

In the presence of measurement error, a *naive inference procedure* based on the partial likelihood function (1.2) with X_i replaced by its surrogate W_i leads to biased estimates of the regression parameters (Prentice 1982).

To demonstrate how measurement error may bias parameter estimation, we conduct a simple simulation study. We consider a Cox model with a scalar error-prone covariate X_i: $\lambda(t; X_i) = \lambda_0(t) \exp(X_i \beta)$, together with a classical error model $W_i = X_i + \epsilon_i$, where $X_i \sim N(0, 1)$, the true value of β is 1, and $\epsilon_i \sim N(0, \sigma^2)$. We refer the *true estimator* to be the partial likelihood estimator from the Cox model with true covariates X_i, and the *naive estimator* to be the partial likelihood estimator from the naive Cox model with X_i replaced by W_i. We vary σ within $[0, 2]$ to reflect different degrees of measurement error. For each value of σ, we simulate the data for 200 times, and record the empirical averages of the naive estimator and the true estimator, and plot them in Figure 1.1.

We see from Figure 1.1 that measurement error tends to bias the naive estimator to the null. The is the so-called *attenuation phenomenon*. The rate of the attenuation seems to be faster in small error cases relative to large error cases. When the error is large in that the *reliability ratio* (Carroll et al. 2006) $\sigma_x^2/(\sigma_x^2 + \sigma^2) < 0.5$ with σ_x^2 representing the variance of X_i, the bias of the naive estimator relative to the true estimator is over 60%.

Figure 1.1: An illustration of the impact of measurement error on parameter estimation



In the following, the attenuation phenomenon is confirmed theoretically. Under the Cox model (1.1) and the nondifferential measurement error assumption, the *induce hazard function* (Prentice 1982) based on the observed data is given by

$$
\begin{aligned}
\lambda(t;W_i) &= \lim_{\Delta t\to 0}\frac{\Pr(T_i\le t+\Delta t|T_i\ge t,W_i)}{\Delta t}\\
&= \lim_{\Delta t\to 0}\frac{E[\Pr(T_i\le t+\Delta t|T_i\ge t,X_i,W_i)|T_i\ge t,W_i]}{\Delta t}\\
&= \lim_{\Delta t\to 0}\frac{E[\Pr(T_i\le t+\Delta t|T_i\ge t,X_i)|T_i\ge t,W_i]}{\Delta t}\\
&= E\left[\lim_{\Delta t\to 0}\frac{\Pr(T_i\le t+\Delta t|T_i\ge t,X_i)}{\Delta t}\middle| T_i\ge t,W_i\right]\\
&= E\left[\lambda(t;X_i)|T_i\ge t,W_i\right]\\
&= \lambda_0(t)E[\exp(X_i^T\beta)|T_i\ge t,W_i]. \qquad (1.6)
\end{aligned}
$$

It follows that $\lambda(t;W_i)$ no longer has the proportional hazards structure due to the impact of measurement error. However, due to the complex form of (1.6), the consequence of ignoring measurement error is unclear: what would happen if one adopts the naive inference procedure? Now we discuss this point in detail.

Define the *naive estimator* $\hat{\beta}_{nv}$ as the maximizer of the partial likelihood (1.2) with X_i replaced by W_i, and the asymptotic limit of $\hat{\beta}_{nv}$ in probability as β^*. Hughes (1993, formula (8)) and Li and Ryan (2004) presented the relationship of β^* and β in different ways. However, it is difficult to quantify the difference between β^* and β. When some additional assumptions are made, the effect of measurement error on the relationship of β^* and β can be explicit or approximately explicit, as illustrated below.

Prentice (1982) made the *rare event assumption* that $\Pr(T_i\ge t)\approx 1$, and showed that under such an assumption, (1.6) becomes

$$\lambda(t;W_i)\approx\lambda_0(t)E[\exp(X_i^T\beta)|W_i]. \qquad (1.7)$$

Thus, assuming that the conditional distribution of X_i, given W_i, is a normal distribution with mean μ_w and variance Σ_w, one can approximately work out an explicit form of $\lambda(t;W_i)$:

$$\lambda(t;W_i)\approx\lambda_0(t)\exp(\mu_w^T\beta+\frac{1}{2}\beta^T\Sigma_w\beta). \qquad (1.8)$$

Note that μ_w and Σ_w are functions of W_i. Pepe, Self and Prentice (1989) showed that under the normality assumptions on both of X_i and ϵ_i (which imply the normality of X_i given W_i), $\mu_w = \mu_x + \Sigma_x(\Sigma_x + \Sigma_0)^{-1}(W_i - \mu_x)$ and Σ_w is deterministic, and thus (1.8) can be simplified as

$$\lambda(t; W_i) \approx \lambda_0^*(t) \exp\{W_i^T(\Sigma_x + \Sigma_0)^{-1}\Sigma_x\beta\}, \tag{1.9}$$

where Σ_x is the variance of X_i, and $\lambda_0^*(t)$ is an unknown deterministic function with an explicit form. See Augustin and Schwarz (2001) and Li and Ryan (2004) for details. Thus, the induced hazards function $\lambda(t; W_i)$ shares the proportional hazards structure approximately, just as that of $\lambda(t; X_i)$ in (1.1). However, the regression parameter of $\lambda(t; W_i)$ becomes $(\Sigma_x + \Sigma_0)^{-1}\Sigma_x\beta$, compared to that of $\lambda(t; X_i)$.

Hughes (1993) demonstrated that

$$\beta^* = (\Sigma_x + \Sigma_0)^{-1}\Sigma_x\beta \tag{1.10}$$

holds approximately in numerical studies. Li and Ryan (2004, Theorem 1) proved that $\hat{\beta}_{nv}$ is a consistent estimator of $\beta^* = (\Sigma_x+\Sigma_0)^{-1}\Sigma_x\beta$ as long as the rare event assumption holds. Consequently, the naive estimator $\hat{\beta}_{nv}$ based on the observed data is a biased estimator of β. In particular, when X_i is univariate, it follows immediately that measurement error biases $\hat{\beta}_{nv}$ towards to the null, and the level of deviation increases as the degree of measurement error increased. This is a theoretical justification for the attenuation phenomenon we observed from Figure 1.1. We note that when the dimension of X_i is great than one, or some accurately covariates V_i is available, complicated phenomena can happen. For instance, reverse-attenuation phenomena can arise (Jiang, Turnbull and Clark 1999; Li and Ryan 2004).

1.4 Review of Existing Methods on Survival Data with Measurement Error

Covariate measurement error has long been a concern in survival analysis, and it has attracted extensive research interest. Since Prentice (1982), a large number of inference methods have been developed to handle error-prone data. Although discussion on survival data with measurement error is not restricted to a single type of model, the Cox model has been the center of existing research. The impact of covariate error is well understood for the Cox model.

In Section 1.3, we demonstrated how measurement error biases estimation of regression parameters. In this section, we provide a review of valid inference methods to account for measurement error effects under various survival models. Specifically, in Section 1.4.1, we introduce the regression calibration approach, and illustrate how it is motivated from the bias analysis in Section 1.3. In Sections 1.4.2 and 1.4.3, we present the likelihood and score based approaches, respectively. In Section 1.4.4, we provide a comprehensive literature review.

1.4.1 Regression Calibration and Simulation Extrapolation

It follows immediately from the relationship (1.10) that

$$\hat{\beta}_{rc} = \hat{\Sigma}_x^{-1}(\hat{\Sigma}_x + \hat{\Sigma}_0)\hat{\beta}_{nv} \tag{1.11}$$

is a deattenuated estimator of β (Pepe, Self and Prentice 1989), where $\hat{\Sigma}_x$ and $\hat{\Sigma}_0$ are estimators of Σ_x and Σ_0 obtained from a validation sample or replication data or other data sources. This deattenuation procedure, however, requires the assumption that both of X_i and ϵ_i are normally distributed.

Without making a distributional assumption on X_i or ϵ_i, $\lambda(t; W_i)$ still has the approximate expression (1.7). Applying the first-order Taylor expansion (e.g., Liao et al. 2011) to the cumulant generating function of $X_i|W_i$, i.e., $\log(E[\exp(X_i^T\beta)|W_i])$, (1.7) becomes

$$\lambda(t;W_i) \approx \lambda_0^{**}(t)\exp(\mu_w^T\beta), \tag{1.12}$$

where $\mu_w = E[X_i|W_i]$ is the so-called *calibration function*, which plays a central role in the *regression calibration* approach. The idea of the regression calibration approach is to replace μ_w in (1.12) by its estimated version, say $\hat{\mu}_w$. A standard partial likelihood procedure is then carried out based on the working model (1.12), and the resultant working partial likelihood estimator is the so-called *regression calibration estimator*. A common choice of $\hat{\mu}_w$, derived by the best linear approximation approach (Carroll et al. 2006), has the expression $\hat{\mu}_w = \mu_x + \Sigma_x(\Sigma_x + \Sigma_0)^{-1}(W_i - \mu_x)$. Correspondingly, (1.12) with μ_w replaced by $\hat{\mu}_w$ is identical to (1.9), and the regression calibration estimator is identical to (1.11).

Clayton (1992), Wang et al. (1997), Xie, Wang and Prentice (2001), Liao et al. (2011), and Shaw and Prentice (2012) further extended the ordinary regression calibration approach for various contexts. In particular, Clayton (1992) and Xie, Wang and Prentice (2001) proposed the risk set calibration approach, which drops the rare event assumption by recalibrating within each risk set. Liao et al. (2011) extended the risk set calibration approach to the internal/external time-varying covariates situation, which is a computationally effective alternative to the common joint model strategies (e.g., Tsiatis and Davidian 2004). Shaw and Prentice (2012) developed the risk set calibration approach under a general measurement error model proposed by Prentice et al. (2002).

The *simulation-extrapolation* (SIMEX) approach (Cook and Stefanski 1994; Stefanski and Cook 1995; Carroll et al. 1996) is another popular approximate method that reduces bias in parameter estimation. Normality of the error term ϵ_i with known variance is typically assumed when applying the SIMEX approach, or this assumption could be removed if replication data is available. The general idea of SIMEX is to study the impact of various degrees of measurement error on the estimation procedure. To be specific, for fixed $\xi > 0$ and a specified large number B, create

$$W_i(\xi, b) = W_i + \sqrt{\xi}\epsilon_{ib}, \quad b = 1, \cdots, B,$$

where $\epsilon_{ib} \sim N(0, \Sigma_0)$ are generated independently and are independent of W_i. Thus the variance of $W_i(\xi, b)$ is inflated to $(1 + \xi)\Sigma_0$. Substitute $W_i(\xi, b)$ into (1.2), and denote by $\hat{\beta}_{nv}(\xi, b)$ the maximizer of the corresponding naive partial likelihood. Let $\hat{\beta}_{nv}(\xi) = B^{-1} \sum_{b=1}^{B} \hat{\beta}_{nv}(\xi, b)$. Repeat this procedure by varying ξ within a set of pre-determined values $0 < \xi_1 < \cdots < \xi_M$, then fit a regression model to the data $\{(\xi_r, \hat{\beta}_{nv}(\xi_r)), r = 1, \cdots, M\}$. The final step is to extrapolate the fit of the model to the case $\xi = -1$, that is, to track back the trend of the estimators under the M scenarios of different degrees of measurement error, and thus obtain an estimator $\hat{\beta}_{nv}(-1)$, called the SIMEX estimator, under the scenario of no measurement error. See Carroll et al. (2006, Ch 5) for a comprehensive introduction.

The regression calibration and SIMEX approaches are advantageous since they are easy to use and reduce bias induced by measurement error considerably. Both approaches are developed for various survival models; we defer the review to Section 1.4.4.

1.4.2 Likelihood-Based Methods

In this section, we first introduce a semiprametric likelihood method proposed by Hu, Tsiatis and Davidian (1998). Under the nondifferential error assumption and the measurement error model (1.4), Hu, Tsiatis and Davidian (1998) proposed the full likelihood based on the observed data:

$$L(\beta, \lambda_0) = \prod_{i=1}^{n} \left[\int \{\lambda_0(S_i) \exp(x^T \beta)\}^{\delta_i} \exp\left\{ - \int_0^{S_i} \lambda_0(u) \exp(x^T \beta) du \right\} f_{W|X}(W_i|x) f_X(x) dx \right], \tag{1.13}$$

where $f_{W|X}(w|x)$ is the conditional density of the surrogate W_i given X_i, and $f_X(x)$ is the marginal density of the true covariate X_i. Hu, Tsiatis and Davidian (1998) assumed that $\lambda_0(t)$ has point mass only at the death times, and thus they essentially used the profile likelihood approach (Murphy and van der Vaart 2000). In the absence of measurement error, we illustrated in Section 1.1.2 that maximizing the product of the integrand inside

(1.13) leads to a regression parameter estimator which is identical to the partial likelihood estimator, and a cumulative hazards estimator which is identical to the Breslow estimator.

The conditional density $f_{W|X}(w|x)$, or the density of ϵ_i, is assumed to be known. The covariate distribution structure $f_X(x)$ is treated by parametric, nonparametric and semiparametric methods. The EM algorithm is adopted to maximize this profile likelihood and the asymptotic variance of the regression parameter estimator is estimated by a special profile likelihood procedure. Dupuy (2005) provided a rigorous justification of this semiprametric likelihood approach using the modern semiparametric theory (Bickel et al. 1993) and the weak convergence theory (van der Vaart and Wellner 1996). Wen (2010) proposed an alternative full likelihood approach with a different factorization of the likelihood function.

Yi and Lawless (2007) did not use the factorization in (1.13). Instead, they extended the corrected likelihood method by Nakamura (1990) for generalized linear models. Let $\ell(\beta, \lambda_0)$ be the log likelihood based on the model linking the reponse with the true covariates. Yi and Lawless (2007) proposed to find a corrected log likelihood function $\ell^*(\beta, \lambda_0)$ based on the observed data such that

$$E_{W|X}[\ell^*(\beta, \lambda_0)] = \ell(\beta, \lambda_0), \tag{1.14}$$

where $E_{W|X}$ represents the expectation evaluated under the conditional distribution of W_i given X_i. Yi and Lawless (2007) imposed a piecewise constant structure on the baseline hazard function $\lambda_0(t)$. Thus, by the regular estimating function theory, solving the derivative of $\ell^*(\beta, \lambda_0)$ typically leads to a consistent regression parameter estimator and a consistent cumulative hazard estimator.

Yi and Lawless (2007) proposed to use

$$\ell^*(\beta, \lambda_0) = \sum_{i=1}^{n} \delta_i \{\log \lambda_0(S_i) + W_i^T \beta\} - \big(E[\exp(\epsilon_i^T \beta)]\big)^{-1} \int_0^{S_i} \lambda_0(u) \exp(W_i^T \beta) du,$$

which satisfies (1.14). The distribution of ϵ_i is assume to be known or well estimated, so that the expression of the moment generating function $E[\exp(\epsilon_i^T \beta)]$ is known.

Alternatively, Zucker (2005) adopted a pseudo partial likelihood approach. He constructed a pseudo partial likelihood by mimicking the standard partial likelihood procedure (Cox 1975) and replacing the hazard function (1.1) with the induced hazard function (1.6). This procedure gives the pseudo partial likelihood as

$$L_p^*(\beta, \lambda_0) = \prod_{i=1}^{n} \left[\frac{E[\exp(X_i^T \beta) | T_i \geq s_i, W_i]}{\sum_{\{j: s_j \geq s_i\}} E[\exp(X_j^T \beta) | T_j \geq s_i, W_j]} \right]^{\delta_i}, \tag{1.15}$$

where s_i is the observed version of S_i. Note that the pseudo partial likelihood involves the baseline hazard function λ_0 through the conditional expectation $E[\exp(X_i^T \beta)|T_i \geq t, W_i]$. Zucker (2005) adapted the profile likelihood approach (Murphy and van der Vaart 2000), and used an iterative procedure to estimate the cumulative induced hazard function consistently. The estimated cumulative induced hazard function is then plugged into (1.15), and maximized to obtain a consistent parameter estimator. A major advantage of this pseudo partial likelihood procedure is that it is not restricted to the classical measurement error model (1.4). However, the density of X_i given W_i needs to be known or properly estimated.

1.4.3 Estimating Equation Methods

Rather than finding a "corrected" log likelihood function $\ell^*(\beta, \lambda_0)$ to satisfy (1.14), Nakamura (1992) aimed at finding a corrected score function $U^*(\beta)$ based on the observed data such that

$$E_{W|X}[U^*(\beta)] = U(\beta), \tag{1.16}$$

where $U(\beta)$ is the partial score function defined in (1.3). The estimating function theory usually guarantees the solution of (1.3) is consistent of β, although extra care is needed since the baseline hazard $\lambda_0(t)$ is unspecified and thus semiparametric theory is needed.

By a first-order Taylor expansion, Nakamura (1992) showed that

$$U^*(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \left[W_i(t) - \frac{\sum_{j=1}^{n} Y_j(t) W_j(t) \exp\{W_j^T(t)\beta\}}{\sum_{j=1}^{n} \exp\{W_j^T(t)\beta\}} + \Sigma_0 \beta \right] dN_i(t) \tag{1.17}$$

satisfies (1.16) under the normality assumption of ϵ_i. Solving (1.17) leads to Nakamura (1992)'s corrected score estimator of β. In addition, he proposed a second-order Taylor series expansion to obtain an improved estimator. Kong and Gu (1999) justified the asymptotic property of the corrected score estimator by Nakamura (1992) using standard asymptotic arguments of Andersen and Gill (1982). Hu and Lin (2002, 2004), and Song and Huang (2005) further extended the corrected score methods to various settings, replacing the normality assumption of ϵ_i with the availability of replicated measurements or a validation sample. Augustin (2004) showed that the corrected score method by Nakamura (1992) could be viewed as a corrected Breslow likelihood method (Breslow 1972, 1974) by assuming a piecewise constant hazard function.

Assuming replicated measurements for X_i are available, Huang and Wang (2000) developed a nonparametric correction method. One drawback of this method is that it can not make use of subjects that have a single measurement. Song and Huang (2005) demonstrated that the nonparameric correction method by Huang and Wang (2000) could be viewed as an approximate version of the corrected score method, and improved the nonparameric correction method in that the information of subjects that have a single measurement is included in the inference procedure. Song and Huang (2005) also proposed the conditional score method by modifying the inference procedures developed by Tsiatis and Davidian (2001) and Song, Davidian and Tsiatis (2002), and showed that it had better performance than the corrected scored method in numerical studies.

Alternatively, the unbiased score functions by Buzas (1998) were constructed in a different way from (1.16), under the assumption that Σ_0 is known and the distribution of ϵ_i is symmetric but not necessarily normal. Zhou and Pepe (1995), Kong (1999), Zhou and Wang (2000), Wang and Pepe (2000), Chen (2002), and Li and Ryan (2006) developed other correction methods based on the score function of the Cox model.

1.4.4 Additional Literature Review

In this section, we consider various extensions of methodologies to handle more complex event history data, including recurrent event, multivariate failure time, and clustered data. Other survival models, including the additive hazards model and the accelarated failure time model, may better fit or explain the data at hand. In practise, the classical additive measurement error model (1.4) is assumed mainly because of its simplicity or the lack of data that can be used to construct a proper measurement error model.

There are some methods on additive hazards model with additive measurement error. Kulich and Lin (2000) used the corrected score approach with a complicated measurement error model when validation data are available. Sun, Zhang and Sun (2006), Sun and Zhou (2008), and Sun, Song and Mu (2012) extended the nonparametric correction method of Huang and Wang (2000), the bias-reduction method of Kong (1999), and the pseudo partial likelihood method of Zucker (2005), respectively. He, Yi and Xiong (2007) and Yi and He (2012) explored the SIMEX method under the accelarated failure time model and the proportional odds model, respectively. Cheng and Wang (2001) considered linear transformation models, which included the proportional odds model as a special case. Ma and Yin (2008) discussed the cure rate model.

In the presence of measurement error, Turnbull, Jiang and Clark (1997) and Jiang, Turnbull and Clark (1999) proposed parametric and semiparametric approaches to handle recurrent event data with random effects. Hu and Lin (2004) and Yi and Lawless (2007) extended their methods developed in the survival settings to the recurrent event cases.

In the context of clustered survival data, Li and Lin (2000) considered a frailty model with normal covariates and errors. They conducted bias analysis to reveal the relationship between the true parameter and the limit of the naive estimator. Li and Lin (2000) proposed a profile likelihood for parameter estimation. Li and Lin (2003) relaxed the distributional assumption on covariates while assuming the errors are normally distributed and replicated measurements are available. They adopted the SIMEX method without specifying the baseline hazard function.

Gorfine, Hsu and Prentice (2003) considered a setting for bivariate survival data. They showed that measurement error seriously biased the estimation of the dependence parameter when one adopted a naive full likelihood approach assuming a marginal Cox model, along with a Copula model to capture the dependence within paired correlated failure time data. They proposed a second-order Taylor expansion to correction the bias. Yi and He (2006) considered bivariate survival data under a marginal accelerated failure time models with covariate error.

For stratified data where different baseline hazards are assumed for different strata, Gorfine, Hsu and Prentice (2004) pointed out that direct extensions of the corrected score method and the nonparametric correction method of Huang and Wang (2000) by adopting the marginal modelling method (Wei, Lin and Weissfeld 1989) usually performed poorly. They modified the risk set calibration method of Xie, Wang and Prentice (2001) to obtain an estimator that has satisfactory numerical performance. In particular, they proposed an unbiased nonparametric corrected estimator by adopting the idea of Huang and Wang (2001) using weighted estimating equations. Greene and Cai (2004) extended the bias analysis technique of Hughes (1993) to stratified data, and proposed to use the SIMEX method to obtain bias-adjusted estimators.

Recently, there is increasing interest in flexible but more complex measurement error models, instead of the classical measurement error model (1.4). Examples can be found in Kulich and Lin (2000), Li and Ryan (2004), Wang (2008), and Shaw and Prentice (2012).

In addition, Kuchenhoff, Bender and Langner (2007) considered additive and multiplicative Berkon error models for parameter estimation under the Cox model. Liao et al. (2011) developed an extension of the additive Berkon error model with time varying covariates, and proposed a modified risk set calibration method. When the covariates are discrete, the additive error model (1.4) is generally not adequate. Zucker and Spiegelman (2004) constructed an estimator based on Kaplan-Meier estimators. Zucker and Spiegelman (2008) extended the corrected score approach to this setting.

Finally, we point out that some topics are not covered in our literature review here.

For example, we do not mention Bayesian analysis with measurement error (e.g., Cheng and Crainiceanu 2009).

1.5 Outline of the Thesis

In the previous two sections, we gave a review of bias analysis and existing methods for event history data with covariate measurement error. Although a large body of the literature is available on survival data analysis with covariate measurement error, many issues are overlooked or relatively less explored. In this thesis, we will look into several important problems.

Since Prentice (1982), a large number of inference methods have been developed to handle error-prone data that are modulated with the Cox model. However, similarity and difference among the existing methods are rarely studied. In Chapter 2, we propose the corrected profile likelihood approach, and show that some available methods can be unified within our inference framework. Our derivation of the corrected profile likelihood sheds light on understanding existing methods which are derived from different techniques which may be more mathematically involved. In addition, we use these results to construct consistent estimators under error models that are more general than the classical error model, a model frequently used in practice.

Hypothesis testing is rarely studied in the literature for survival models in the presence of covariate measurement error. Furthermore, no work seems available to study the impact of misspecifying error models. In Chapter 3, we investigate these important problems. We proposed corrected score and Wald tests under Cox models with mismeasured covariates and study their validity and efficiency properties. Results of the impact of misspecification on parameter estimation and hypothesis testing are provided.

Another issue that is ignored in the literature is the validity of the assumptions of the survival models with covariate measurement error. In Chapter 4, we address the issue of checking the Cox model assumptions with mismeasured covariates.

In contrast to proportional hazards models, additive hazards models offer a flexible tool to delineate survival processes. However, there is little research on measurement error effects under additive hazards models. In Chapter 5, we systematically investigate this important problem. New insights of measurement error effects are revealed, as opposed to well-documented results for proportional hazards models. In particular, we explore asymptotic bias of ignoring measurement error in the analysis. To correct for the induced bias, we develop a class of functional correction methods for error effects to exemplify the unique features of additive hazards models. The validity of the proposed methods is carefully examined, and issues of model checking and model misspecification are investigated. Theoretical results are rigorously established, complemented by numerical assessments.

In many clinical studies, high dimensional risk factors are collected and included in the data analysis procedure, and some risk factors may suffer from measurement error. There exists little work on variable selection and estimation for high dimensional survival models with measurement error. In Chapter 6, we propose corrected penalized methods to adjust for measurement error, and show that the proposed methods are suitable for estimation and variable selection theoretically. We illustrate their performance through simulation studies and real data analysis.

A summary of this thesis is included in Chapter 7.

Chapter 2

Corrected Profile Likelihood for Cox Model with Covariate Measurement Error

2.1 Introduction

The Cox proportional hazards model (Cox 1972) features that covariates have multiplicative effects on the hazard ratio, and leaves the temporal effects indicated by the baseline hazard function. This model is perhaps the most widely used model for survival data analysis. Inferences under this model have been commonly conducted based on the partial likelihood approach (Cox 1975) for which the baseline hazard function is left unattended to. An alternative approach, developed by Murphy and van der Vaart (2000), has received far less attention although it is more intuitive and convenient to estimate the regression parameters and the baseline hazard function simultaneously.

Although the Cox model has proven to be useful for survival analysis, inferences under this model are frequently challenged by the complexity of data. Covariate measurement error is a ubiquitous phenomenon occurring in clinical trials and observational studies. For example, in AIDS studies, CD4 lymphocyte counts are an important biomarker, but they are measured with unignorable error due to biological variability and measurement procedures. As illustrated by Prentice (1982), simply ignoring measurement error in covariates would normally result in substantially biased results. Consequently, many correction methods have been proposed to handle covariate measurement error under the Cox model. For instance, Prentice (1982), Pepe, Self and Prentice (1989), and Wang et al. (1997) proposed the regression calibration approach; Li and Lin (2003) explored the simulation extrapolation (SIMEX) approach; and Li and Ryan (2004) proposed a bias corrected estimator under the Cox model. These approaches can effectively reduce bias in many settings, although they cannot produce exactly consistent estimators.

In contrast, methods that yield consistent estimators have been developed, and they can be broadly classified into two categories: likelihood-based and score-based methods. For example, Hu, Tsiatis and Davidian (1998) developed a full likelihood approach under the assumption that measurement error is normal, and Augustin (2004) and Yi and Lawless (2007) extended the corrected likelihood approach of Nakamura (1990) to the Cox model. In particular, Yi and Lawless (2007) used a weakly parametric method to model the baseline hazard function. Likelihood-based methods typically require attention to handle the baseline hazard function when introducing corrections to adjust for measurement error effects. On the other hand, score-based methods focus on inducing corrections to the partial likelihood or partial likelihood score functions. Such methods are attractive in the sense that the baseline hazard function is left unattended to. Specifically, Zucker (2005) proposed a pseudo partial likelihood approach, and Nakamura (1992) developed a corrected score approach under the normality assumption of the errors. With replication data, Huang and Wang (2000) developed a nonparametric correction method to modify the partial score function. Hu and Lin (2002, 2004) developed semiparametric regression approaches for cases with replicated measurements or validation data. Song and Huang (2005) provided refinements of the method by Huang and Wang (2000) and proposed conditional score approaches. Li and Ryan (2006) proposed a imputation-based score approach.

Those existing methods are useful in correcting for measurement error effects for error-contaminated survival data. However, their derivations and theoretical properties are established very differently; many of them are considerably mathematically involved.

Those complex details somehow obscure the intrinsic connections among various methods. It is the goal of this chapter to explore this problem. We develop a general strategy to correct for covariate measurement error under the Cox model. Our method pertains, in principle, to the profiling method by Murphy and van der Vaart (2000) while the technical details are different. There are several important implications of our method. Notably, it supplies a unified framework into which many existing methods can be embedded. Moreover, in contrast to some existing methods that can only produce approximately consistent estimators, our method can yield exactly consistent estimators under general measurement error models, and asymptotic results of the resultant estimators are established rigorously. We also extend our results to the Berkson error model. To explore the problem in depth, we investigate the impact of model misspecification of the measurement error process; this research receives little attention in the literature. Our study uncovers interesting findings.

The remainder is organized as follows. In Section 2.2, we introduce the basic model setup. A brief review of some existing methods is included in Section 2.3. In Section 2.4, we propose the corrected profile likelihood approach under general regression measurement error models, and show that those existing methods can be unified by this approach. Simulation studies and a data analysis are reported in Section 2.5. Concluding discussion is provided in the last section.

2.2 Notation and model setup

Let T_i be the failure time, C_i be the censoring time, and Z_i be a vector of time-independent covariates, $i = 1, \cdots, n$. As common in practise, the independent censoring mechanism (Lawless 2003) is assumed. Suppose all the individuals are observed over a common time interval $[0, \tau]$, where $0 < \tau < \infty$. Let $S_i = \min(T_i, C_i)$, and $\delta_i = I(T_i \leq C_i)$. Let $N_i(t) = I(S_i \leq t, \delta_i = 1)$ be the number of observed failures for the ith subject up to and including time t, and $Y_i(t) = I(S_i \geq t)$ indicate whether the ith subject is at risk of failure at time t^-. Let $\mathcal{F}_t = \sigma\{N_i(s), Y_i(s^+), Z_i, 0 \leq s \leq t, i = 1, \cdots, n\}$ be the $\sigma-$ field generated by the observed event and covariates histories prior to time t for all subjects. Suppose T_i is continuous and there are no ties in the observed event times $s_1, \cdots, s_n$, i.e., the realized values of $S_1, \cdots, S_n$, respectively.

2.2.1 Cox model

We consider that the hazard function of T_i is related to Z_i through the Cox model (Cox 1972)

$$\lambda\{t; Z_i(t)\} = \lambda_0(t) \exp(Z_i^T \beta), \tag{2.1}$$

where $\lambda_0(\cdot)$ is the baseline hazard function, and β is a vector of unknown regression parameters. Let $\Lambda_0(t) = \int_0^t \lambda_0(u)du$. Inference about the regression parameter β is typically based on the partial likelihood (Cox 1975):

$$L_p(\beta) = \prod_{i=1}^{n} \left[\frac{\exp(Z_i^T \beta)}{\sum_{\{j: s_j \geq s_i\}} \exp(Z_j^T \beta)} \right]^{\delta_i}. \tag{2.2}$$

Maximizing $L_p(\beta)$ with respect to β leads to the partial likelihood estimator $\hat{\beta}$ of β. Alternatively, $\hat{\beta}$ can be obtained by solving the partial score function $U_p(\beta) = 0$, where

$$U_p(\beta) = \frac{\partial \log\{L_p(\beta)\}}{\partial \beta} = \sum_{i=1}^{n} \delta_i \left[Z_i - \frac{\sum_{\{j: s_j \geq s_i\}} Z_j \exp(Z_j^T \beta)}{\sum_{\{j: s_j \geq s_i\}} \exp(Z_j^T \beta)} \right]. \tag{2.3}$$

The partial likelihood method is advantageous in that the baseline hazard function $\lambda_0(t)$ is left unspecified, thus protecting us from obtaining invalid results about β when $\lambda_0(t)$ is mismodeled.

Alternatively, inference on β can be carried out using the profile likelihood approach developed by Murphy and van der Vaart (2000). The key idea is to restrict the baseline cumulative function $\Lambda_0(t)$ to the jumps only at $s_1, \cdots, s_n$, and to treat the sizes of these jumps, say $\Lambda_0\{s_1\}, \cdots, \Lambda_0\{s_n\}$, as unknown parameters, together with the parameter β. Then the full likelihood becomes

$$L(\beta, \Lambda_0) = \prod_{i=1}^{n} \left[\exp(Z_i^T\beta)\Lambda_0\{s_i\}\right]^{\delta_i} \exp\left[-\sum_{\{j:s_j \le s_i\}} \exp(Z_i^T\beta)\Lambda_0\{s_j\}\right], \qquad (2.4)$$

yielding the log likelihood

$$\ell(\beta, \Lambda_0) = \sum_{i=1}^{n} \left[\delta_i \left[Z_i^T\beta + \log\Lambda_0\{s_i\}\right] - \sum_{\{j:s_j \le s_i\}} \exp(Z_i^T\beta)\Lambda_0\{s_j\}\right]. \qquad (2.5)$$

It is immediate that the likelihood score functions are

$$\frac{\partial\ell(\beta, \Lambda_0)}{\partial\Lambda_0\{s_i\}} = \frac{\delta_i}{\Lambda_0\{s_i\}} - \sum_{\{j:s_j \ge s_i\}} \exp(Z_j^T\beta), \quad i = 1, \cdots, n, \qquad (2.6)$$

$$\text{and} \quad \frac{\partial\ell(\beta, \Lambda_0)}{\partial\beta} = \sum_{i=1}^{n} \left[\delta_i Z_i - \sum_{\{j:s_j \le s_i\}} Z_i \exp(Z_i^T\beta)\Lambda_0\{s_j\}\right]. \qquad (2.7)$$

For a given value of β, setting (2.6) to be zero gives the solution:

$$\hat{\Lambda}_0\{s_i\} = \frac{\delta_i}{\sum_{\{j:s_j \ge s_i\}} \exp(Z_j^T\beta)}, \quad i = 1, \cdots, n, \qquad (2.8)$$

which is identical to the usual Breslow estimator (Breslow 1972). Replacing $\Lambda_0\{s_i\}$ in (2.4) with the solution (2.8) yields the profile likelihood for the β parameter:

$$L_{prof}(\beta, \hat{\Lambda}_0) = \exp\left(-\sum_{i=1}^{n}\delta_i\right) \prod_{i=1}^{n} \left[\frac{\exp(Z_i^T\beta)}{\sum_{\{j:s_j \ge s_i\}} \exp(Z_j^T\beta)}\right]^{\delta_i}.$$

It is seen that this profile likelihood is identical to the partial likelihood (2.2) up to a constant. In addition, we note that the partial likelihood score function (2.3) is identical to what is resulted from plugging (2.8) into (2.7). Thus, the profile likelihood estimator of the regression coefficient β is the same as the partial likelihood estimator $\hat{\beta}$.

2.2.2 Measurement error models

Suppose that some covariates are subject to measurement error. We write Z_i as $Z_i = (X_i^T, V_i^T)^T$, where V_i is a subvector of precisely observed covariates, and X_i includes error-prone covariates. Suppose the dimension of V_i is q, and X_i is univariate for ease of exposition. Extensions to accommodating multiple dimensions of X_i are straightforward. Let W_i be a surrogate measurement of X_i. Write $\beta = (\beta_x, \beta_v^T)^T$ so that β_x and β_v correspond to X_i and V_i, respectively. We consider three useful scenarios of the measurement error process.

In Scenario A, the measurement error model is specified as

$$W_i = \gamma_0 + X_i\gamma_x + V_i^T\gamma_v + \epsilon_i, \quad i = 1, \cdots, n, \tag{2.9}$$

where the error terms $\epsilon_i, i = 1, \cdots, n$ are independent and identically distributed and are independent of $N_i(\cdot)$, $Y_i(\cdot)$, and Z_i, and the distribution of ϵ_i is assumed known. The regression coefficients γ_0, γ_x, and γ_v are assumed known in order to highlight the ideas of the inference methods accounting for error effects. Let $\eta_0(\beta_x) = E\{\exp(\epsilon_i\beta_x)\}$, $\eta_1(\beta_x) = E\{\epsilon_i \exp(\epsilon_i\beta_x)\}$, and $D(\beta_x) = \eta_0^{-1}(\beta_x)\eta_1(\beta_x)$. If ϵ_i is assumed to be normal with variance σ_0^2, then $D(\beta_x) = \sigma_0^2\beta_x$.

Scenario B relaxes the requirements of (2.9), but an external validation sample $\{(W_i, X_i, V_i) : i \in \mathcal{V}\}$ is available in addition to the data $\{(S_i, \delta_i, W_i, V_i) : i \in \mathcal{M}\}$ in the main study. Specifically, we assume

$$W_i = \gamma_0 + X_i\gamma_x + V_i^T\gamma_v + \epsilon_i, \quad i \in \mathcal{M} \cup \mathcal{V}, \tag{2.10}$$

where γ_0, γ_x, and γ_v are unknown regression coefficients, the error terms $\epsilon_i, i = 1, \cdots, n$ are independent and identically distributed and are independent of $N_i(\cdot)$, $Y_i(\cdot)$, and Z_i, and the distribution of ϵ_i is left unspecified. Let n and m be the size of $\mathcal{M}$ and $\mathcal{V}$, respectively. Assume that $\rho = \lim_{n\to\infty} m/n$ exists.

Scenario C describes a different situation where X_i is repeatedly measured n_i times by the surrogates $W_{ir}(r = 1, \cdots, n_i)$:

$$W_{ir} = \gamma_0 + X_i\gamma_x + V_i^T\gamma_v + \epsilon_{ir}, \quad i = 1, \cdots, n; r = 1, \cdots, n_i, \tag{2.11}$$

where the distribution of ϵ_{ir} is left unspecified, but we follow the convention to assume that the error terms ϵ_{ir} are independent and identically distributed with mean 0 and an unknown variance σ_0^2, and are independent of $N_i(\cdot)$, $Y_i(\cdot)$, and Z_i. The regression coefficients γ_0, γ_x and γ_v are assumed known as in Scenario A. With replicates W_{ir}, a consistent estimate of σ_0^2 is given by

$$\hat{\sigma}_0^2 = \frac{\sum_{i=1}^n \sum_{r=1}^{n_i}(W_{ir} - \bar{W}_{i\cdot})^2}{\sum_{i=1}^n (n_i - 1)},$$

where $\bar{W}_{i\cdot} = \sum_{r=1}^{n_i} W_{ir}/n_i$. Let $\eta_0(\beta_x) = E\{\exp(\epsilon_{ir}\beta_x)\}$, $\eta_1(\beta_x) = E\{\epsilon_{ir}\exp(\epsilon_{ir}\beta_x)\}$, and $D(\beta_x) = \eta_0^{-1}(\beta_x)\eta_1(\beta_x)$.

Let $\hat{Z}_i = (W_i, V_i^T)^T$ denote the observed covariates for Scenarios A and B, and let $\hat{Z}_{ir} = (W_{ir}, V_i^T)^T$ and $\hat{Z}_i = (\bar{W}_{i\cdot}, V_i^T)^T$ for Scenario C.

We comment that Scenario B imposes least assumptions on the error model (2.10). This gives us great flexibility to postulate the measurement error process. This flexibility is possible at the cost of requiring "better" data in the sense that a validation subsample is available. The availability of a validation subsample enables us to estimate the model parameters γ_0, γ_x, and γ_v, as well as understand the error distribution. When there are no validation data, model parameters may be nonidentifiable or inestimable. To overcome this problem, parameters γ_0, γ_x, and γ_v in Scenarios A and C are assumed known. This assumption may appear restrictive at the first sight, but it is useful in accommodating models which are widely used in the literature. For example, when $\gamma_0 = 0$, $\gamma_x = 1$, and $\gamma_v = 0$, then the error model (2.9) reduces to the widely used classical additive error model (Carroll et al. 2006)

$$W_i = X_i + \epsilon_i, \quad i = 1, \cdots, n. \tag{2.12}$$

When $\gamma_0 = 0$, $\gamma_x = 1$, and $\gamma_v = 0$, then model (2.11) reduces to the error model considered by many authors (e.g., Huang and Wang 2000)

$$W_{ir} = X_i + \epsilon_{ir}, \quad i = 1, \cdots, n; \quad r = 1, \cdots, n_i. \tag{2.13}$$

2.3 Brief review of existing methods

In this section, we briefly describe the methods proposed by Nakamura (1992), Huang and Wang (2000), Hu and Lin (2002, 2004), Song and Huang (2005), and Yi and Lawless (2007). These methods are developed under the Cox model with covariate measurement error modeled by the classical error models (2.12) or (2.13), which are special cases of Scenarios A-C.

2.3.1 Corrected score approach by Nakamura (1992) and Song and Huang (2005)

Under the Cox model (2.1) and the classical error model (2.12) in Scenario A, Nakamura (1992) applied the first-order Taylor expansion for the partial score function (2.2), and obtained that

$$E\left\{ \frac{\sum_{\{j:s_j\geq s_i\}} \hat{Z}_j \exp(\hat{Z}_j^T\beta)}{\sum_{\{j:s_j\geq s_i\}} \exp(\hat{Z}_j^T\beta)} \middle| \mathcal{F}_\tau \right\} \approx \frac{\sum_{\{j:s_j\geq s_i\}} Z_j \exp(Z_j^T\beta)}{\sum_{\{j:s_j\geq s_i\}} \exp(Z_j^T\beta)} + \begin{pmatrix} \sigma_0^2\beta_x \\ 0 \end{pmatrix}.$$

As a result, Nakamura (1992) proposed the so-called corrected score functions

$$U_{Naka}(\beta) = \sum_{i=1}^n \int_0^\tau \left\{ \hat{Z}_i(t) - \frac{\sum_{j=1}^n Y_j(t)\hat{Z}_i \exp(\hat{Z}_j^T\beta)}{\sum_{j=1}^n Y_j(t)\exp(\hat{Z}_j^T\beta)} + \begin{pmatrix} \sigma_0^2\beta_x \\ 0 \end{pmatrix} \right\} dN_i(t). \tag{2.14}$$

Nakamura (1992) showed that when $\sigma_0^2\beta_x^2$ is small, $E[U_{Naka}(\beta)|\mathcal{F}_\tau] \approx U_p(\beta)$, and thus $E[U_{Naka}(\beta)] \approx 0$, indicating that $U_{Naka}(\beta)$ are approximately unbiased. Solving (2.14) gives a consistent estimator of β.

With replicated data W_{ij} which are modeled by the additive error model (2.13) in Scenario C, Song and Huang (2005) extended the corrected score method of Nakamura (1992) and proposed the corrected score function

$$U_{SH}(\beta) = \sum_{i=1}^n \int_0^\tau \left\{ \hat{Z}_i(t) - \frac{\sum_{j=1}^n Y_j(t)n_j^{-1}\sum_{r=1}^{n_j} \hat{Z}_{jr}\exp(\hat{Z}_{jr}^T\beta)}{\sum_{j=1}^n Y_j(t)n_j^{-1}\sum_{r=1}^{n_j}\exp(\hat{Z}_{jr}^T\beta)} + \begin{pmatrix} \hat{D}_{SH}(\beta_x) \\ 0 \end{pmatrix} \right\} dN_i(t), \tag{2.15}$$

where

$$\hat{D}_{SH}(\beta_x) = \frac{\sum_{i=1}^{n} I(n_i > 1) n_i^{-1}(n_i - 1)^{-1} \sum_{r \neq s} (W_{ir} - W_{is}) \exp(W_{ir}\beta_x)}{\sum_{i=1}^{n} I(n_i > 1) n_i^{-1} \sum_{r=1}^{n_i} \exp(W_{ir}\beta_x)}$$

is a consistent estimator of $D(\beta_x)$. Solving (2.15) gives a consistent estimator of β.

2.3.2 Semiparametric regression method by Hu and Lin (2002,2004)

When there are replicated data W_{ij} and the additive error model (2.13) in Scenario C holds, Hu and Lin (2004) observed that

$$\begin{aligned} E\{\exp(\hat{Z}_{ir}^T\beta)|Z_i\} &= \eta_0(\beta_x)\exp(Z_i^T\beta), \\ \text{and} \quad E\{\hat{Z}_{ir}\exp(\hat{Z}_{ir}^T\beta)|Z_i\} &= \eta_0(\beta_x) Z_i \exp(Z_i^T\beta) + \exp(Z_i^T\beta)\begin{pmatrix} \eta_1(\beta_x) \\ 0 \end{pmatrix}. \end{aligned}$$

Assuming that the error distribution is symmetric, Hu and Lin (2004) proposed to estimate $\eta_0(\beta_x)$ and $\eta_1(\beta_x)$ by

$$\begin{aligned} \hat{\eta}_0(\beta_x) &= \left[\frac{\sum_{i=1}^{n} I(n_i > 1) n_i^{-1}(n_i - 1)^{-1} \sum_{r \neq s} \exp\{(W_{ir} - W_{is})\beta_x\}}{\sum_{i=1}^{n} I(n_i > 1)}\right]^{1/2}, \\ \text{and} \quad \hat{\eta}_1(\beta_x) &= \frac{\sum_{i=1}^{n} I(n_i > 1) n_i^{-1}(n_i - 1)^{-1} \sum_{r \neq s}(W_{ir} - W_{is}) \exp\{(W_{ir} - W_{is})\beta_x\}}{2\hat{\eta}_0(\beta_x)\sum_{i=1}^{n} I(n_i > 1)}, \end{aligned}$$

respectively. Consequently, Hu and Lin (2004) proposed the estimating functions

$$U_{HL}(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \left\{\hat{Z}_i - \frac{S_{HL}^{(1)}(\beta, t)}{S_{HL}^{(0)}(\beta, t)}\right\} dN_i(t), \tag{2.16}$$

where $S_{HL}^{(k)}(\beta, t) = n^{-1}\sum_{i=1}^{n} Y_i(t) R_{HL,i}^{(k)}(\beta), k = 0, 1$, $R_{HL,i}^{(0)}(\beta) = \hat{\eta}_0^{-1}(\beta_x) n_i^{-1} \sum_{r=1}^{n_i} \exp(\hat{Z}_{ir}^T\beta)$, $R_{HL,i}^{(1)}(\beta) = \hat{\eta}_0^{-1}(\beta_x) n_i^{-1} \sum_{r=1}^{n_i} \exp(\hat{Z}_{ir}^T\beta)\{\hat{Z}_{ir} - (\hat{D}_{HL}(\beta_x), 0^T)^T\}$, and $\hat{D}_{HL}(\beta_x) = \hat{\eta}_0^{-1}(\beta_x)\hat{\eta}_1(\beta_x)$. Solving (2.16) gives a consistent estimator of β.

Hu and Lin (2002) considered a validation subsample scenario that is slightly different from Scenario B. They assumed a validation sample $\{(S_i, \delta_i, W_i, X_i, V_i) : i \in \mathcal{V}\}$ is available in addition to the data in the main study. The measurement error model assumes the form (2.10) with $\gamma_0 = 0$, $\gamma_x = 1$, and $\gamma_v = 0$. Let $\xi_i = 1$ if the subject i is in $\mathcal{V}$, and 0 otherwise. Let $\hat{\eta}_{HL2,k}(\beta_x) = m^{-1}\sum_{i\in\mathcal{V}}\epsilon_i^k \exp(\beta_x\epsilon_i)$ be a consistent estimators of $\eta_k(\beta_x)$, $k = 0, 1$. Similar to Hu and Lin (2004), Hu and Lin (2002) developed the estimating functions

$$U_{HL2}(\beta) = \sum_{i\in\mathcal{M}\cup\mathcal{V}} \int_0^\tau \left\{ \xi_i Z_i + \bar{\xi}_i \hat{Z}_i - \frac{S_{HL2}^{(1)}(\beta,t)}{S_{HL2}^{(0)}(\beta,t)} \right\} dN_i(t), \tag{2.17}$$

where $\bar{\xi}_i = 1 - \xi_i$, $S_{HL2}^{(k)}(\beta,t) = (n+m)^{-1}\sum_{i\in\mathcal{M}\cup\mathcal{V}} Y_i(t) R_{HL2,i}^{(k)}(\beta), k = 0, 1,$

$$\begin{aligned}
R_{HL2,i}^{(0)}(\beta) &= \xi_i \exp(Z_i^T\beta) + \bar{\xi}_i \hat{\eta}_{HL2,0}^{-1}(\beta_x)\exp(\hat{Z}_i^T\beta), \\
R_{HL2,i}^{(1)}(\beta) &= \xi_i Z_i \exp(Z_i^T\beta) + \bar{\xi}_i \hat{\eta}_{HL2,0}^{-1}(\beta_x)\exp(\hat{Z}_i^T\beta)\{\hat{Z}_i - (\hat{D}_{HL2}(\beta_x), 0^T)^T\}, \\
\text{and} \quad \hat{D}_{HL2}(\beta_x) &= \hat{\eta}_{HL2,0}^{-1}(\beta_x)\hat{\eta}_{HL2,1}(\beta_x).
\end{aligned}$$

Solving (2.17) gives a consistent estimator of β, provided suitable regularity conditions.

2.3.3 Nonparametric correction method by Huang and Wang (2000)

When there are replicated data W_{ij} and the additive error model (2.13) under Scenario C holds, Huang and Wang (2000) observed that for any $r \neq s$,

$$E\left\{\hat{Z}_{ir}\exp(\hat{Z}_{is}^T\beta)|Z_i\right\} = Z_i \exp(Z_i^T\beta).$$

Using the empirical process techniques, Huang and Wang (2000) proposed the estimating functions

$$U_{HW}(\beta) = \sum_{i=1}^n \int_0^\tau \left[n_i^{-1}\sum_{r=1}^{n_i}\hat{Z}_{ir} - \frac{\sum_{i=1}^n Y_i(t) n_i^{-1}(n_i-1)^{-1}\sum_{r\neq s; n_i>1}\hat{Z}_{ir}\exp(\hat{Z}_{is}^T\beta)}{\sum_{i=1}^n Y_i(t) n_i^{-1}\sum_{r=1}^{n_i}\exp(\hat{Z}_{ir}^T\beta)} \right] dN_i(t). \tag{2.18}$$

Solving (2.18) gives a consistent estimator of β, provided certain regularity conditions are met.

2.3.4 Corrected likelihood method by Yi and Lawless (2007)

Under the additive error model (2.12) in Scenario A, Yi and Lawless (2007) extended the corrected likelihood method by Nakamura (1990) to the Cox model, and proposed the corrected log likelihood

$$\ell_c(\beta, \Lambda_0) = \sum_{i=1}^{n} \left[\delta_i \left\{ \hat{Z}_i^T \beta + \log \Lambda_0(s_i) \right\} - \eta_0^{-1}(\beta_x) \sum_{\{j: s_j \le s_i\}} \exp(\hat{Z}_i^T \beta) \Lambda_0(s_j) \right].$$

The corrected log likelihood $\ell_c(\beta, \Lambda_0)$ is unbiased for the log likelihood

$$\ell(\beta, \Lambda_0) = \sum_{i=1}^{n} \left[\delta_i \left\{ \hat{Z}_i^T \beta + \log \Lambda_0(s_i) \right\} - \sum_{\{j: s_j \le s_i\}} \exp(\hat{Z}_i^T \beta) \Lambda_0(s_j) \right],$$

in the sense that

$$E[\ell_c(\beta, \Lambda_0) | \mathcal{F}_\tau] = \ell(\beta, \Lambda_0). \tag{2.19}$$

With $\Lambda_0(t)$ modelled by a piecewise constant function, Yi and Lawless (2007) showed that under certain regularity conditions, maximizing $\ell_c(\beta, \Lambda_0)$ gives consistent estimators of β and Λ_0.

2.4 Corrected profile likelihood approach

Existing methods that handle survival data with measurement error are often developed under the classical error models (2.12), (2.13), or Scenario B with $\gamma_0 = 0$, $\gamma_x = 1$, and $\gamma_v = 0$. These models can be restrictive in application. In this section we consider broader classes of measurement error models which are specifically featured by Scenarios A, B, or C. We propose corrected profile likelihood methods to account for measurement error.

2.4.1 Method for Scenario A

We first consider the special error model (2.12) under Scenario A. Unlike Yi and Lawless (2007) who used a weakly parametric method to model the baseline hazard function, we take a different perspective to feature the baseline hazard function. Specifically, we adapt the profile likelihood method outlined in Section 2.2.1, with $\Lambda_0(s_i), i = 1, \cdots, n$ treated as parameters along with the regression parameter β. We propose the the corrected log profile likelihood

$$\ell_c(\beta, \Lambda_0) = \sum_{i=1}^{n} \left[\delta_i \left\{ \hat{Z}_i^T \beta + \log \Lambda_0\{s_i\} \right\} - \eta_0^{-1}(\beta_x) \sum_{\{j: s_j \le s_i\}} \exp(\hat{Z}_i^T \beta) \Lambda_0\{s_j\} \right], \tag{2.20}$$

where $\Lambda_0\{t\}$ is the size of the jump at t of $\Lambda_0(t)$, $t \in [0, \tau]$. It is clear that $\ell_c(\beta, \Lambda_0)$ is unbiased for the log likelihood $\ell(\beta, \Lambda_0)$ defined in (2.5) in the sense of (2.19). Correspondingly, the corrected profile score functions are given by

$$\frac{\partial \ell_c(\beta, \Lambda_0)}{\partial \Lambda_0\{s_i\}} = \frac{\delta_i}{\Lambda_0\{s_i\}} - \eta_0^{-1}(\beta_x) \sum_{\{j: s_j \ge s_i\}} \exp(\hat{Z}_j^T \beta), \quad i = 1, \cdots, n, \tag{2.21}$$

and $$\frac{\partial \ell_c(\beta, \Lambda_0)}{\partial \beta} = \sum_{i=1}^{n} \left[\delta_i \hat{Z}_i - \eta_0^{-1}(\beta_x) \sum_{\{j: s_j \le s_i\}} \left\{ \hat{Z}_i - \begin{pmatrix} D(\beta_x) \\ 0 \end{pmatrix} \right\} \exp(\hat{Z}_i^T \beta) \Lambda_0\{s_j\} \right]. \tag{2.22}$$

We calculate the solutions of (2.21) and (2.22) by following the standard profile likelihood procedure introduced in Section 2. To be specific, for fixed β, setting (2.21) equal to 0 gives a corrected estimator of the baseline hazard function

$$\hat{\Lambda}_c\{s_i\} = \frac{\delta_i}{\eta_0^{-1}(\beta_x) \sum_{\{j: s_j \ge s_i\}} \exp(\hat{Z}_j^T \beta)}, \quad i = 1, \cdots, n. \tag{2.23}$$

Plugging (2.23) into (2.22), we obtain the corrected profile score functions

$$U_c(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \left\{ \hat{Z}_i - \frac{S^{(1)}(\hat{Z}; \beta, t)}{S^{(0)}(\hat{Z}; \beta, t)} + \begin{pmatrix} D(\beta_x) \\ 0 \end{pmatrix} \right\} dN_i(t), \tag{2.24}$$

where $S^{(k)}(\hat{Z}; \beta, t) = n^{-1} \sum_{i=1}^{n} Y_i(t) \hat{Z}_i^k \exp(\hat{Z}_i^T \beta)$ for $k = 0$ and 1.

Solving (2.24) gives a corrected estimator, say $\hat{\beta}_c$, of β. We note that under the normal error assumption in (2.12), $\hat{\beta}_c$ is identical to the corrected score estimator proposed by

Nakamura (1992), who initially used the Taylor series expansion to correct for measurement error effect on the partial score functions. Therefore, our proposed corrected profile likelihood method bridges the corrected likelihood method by Yi and Lawless (2007) and the corrected score method by Nakamura (1992). Asymptotic results of Kong and Gu (1999) guarantee consistency of $\hat{\beta}_c$.

Now, we extend the corrected profile likelihood method to the general error model (2.9) in Scenario A. Our idea is to start with a convenient working model for the measurement process, and construct a working likelihood in combination with the survival model (2.1). Maximizing this working likelihood gives us a working estimator which will be used for developing a consistent estimator.

To be specific, we take the classical error model (2.12) in Scenario A to be a working measurement error model. Then mimicking the arguments of corrected profile likelihood from (2.20) to (2.24), we obtain working estimators. Let $\hat{\beta}_c$ and $\hat{\Lambda}_c(\cdot)$ denote the resulting working estimators of β and $\Lambda_0(\cdot)$, and $\beta_c = (\beta_{c,x}, \beta_{c,v}^T)^T$ and $\Lambda_c(\cdot)$ be the limit of $\hat{\beta}_c$ and $\hat{\Lambda}_c(\cdot)$ in probability, respectively. Let $\beta_0 = (\beta_{0,x}, \beta_{0,v}^T)^T$ denote the true value of β. In Appendix A2, we show the following result:

Lemma 1 *Under the regularity conditions R1-R6 listed in Appendix A1, and under Scenario A, we have*

$$\beta_{c,x} = \gamma_x^{-1}\beta_{0,x}, \quad \beta_{c,v} = \beta_{0,v} - \gamma_v\gamma_x^{-1}\beta_{0,x}, \tag{2.25}$$
$$\textit{and} \;\; \Lambda_c(\cdot) = \exp(-\gamma_0\gamma_x^{-1}\beta_{0,x})\Lambda_0(\cdot).$$

This result can then be used to construct a consistent estimator. Let

$$\hat{\beta}_{cc,x} = \gamma_x\hat{\beta}_{c,x}, \tag{2.26}$$
$$\hat{\beta}_{cc,v} = \hat{\beta}_{c,v} + \gamma_v\hat{\beta}_{c,x}, \tag{2.27}$$
$$\text{and} \;\; \hat{\Lambda}_{cc}(\cdot) = \exp(\gamma_0\hat{\beta}_{c,x})\hat{\Lambda}_c(\cdot).$$

Write $\hat{\beta}_{cc} = (\hat{\beta}_{cc,x}, \hat{\beta}_{cc,v}^T)^T$. The following theorem shows that $\hat{\beta}_{cc}$ and $\hat{\Lambda}_{cc}(\cdot)$ are consistent estimators of β_0 and $\Lambda_0(\cdot)$, respectively. Furthermore, the following theorem shows the asymptotic normality property of $\hat{\beta}_{cc}$ and the weak convergence property of $\hat{\Lambda}_{cc}(\cdot)$. The proof is outlined in Appendix A3.

Theorem 1 *Under the regularity conditions R1-R6 listed in Appendix A1, we obtain that*

(1). $\hat{\beta}_{cc}$ *and* $\hat{\Lambda}_{cc}(\cdot)$ *are consistent estimators of* β_0 *and* $\Lambda_0(\cdot)$*, respectively.*

(2).

$$n^{1/2}(\hat{\beta}_{cc} - \beta_0) \xrightarrow{d} N(0, \mathcal{D}^T\mathcal{I}^{-1T}\mathcal{J}\mathcal{I}^{-1}\mathcal{D}) \quad as \quad n \to \infty,$$

where

$$\mathcal{D} = \begin{pmatrix} \gamma_x & \gamma_v^T \\ 0 & I \end{pmatrix};$$

$$\mathcal{I} = \int_0^\tau \left[\frac{s^{(2)}(\hat{Z};\beta_c,t)}{s^{(0)}(\hat{Z};\beta_c,t)} - \left\{ \frac{s^{(1)}(\hat{Z};\beta_c,t)}{s^{(0)}(\hat{Z};\beta_c,t)} \right\}^{\otimes 2} - \begin{pmatrix} \frac{\partial D(\beta_x)}{\partial \beta_x^T}|_{\beta_x=\beta_{c,x}} & 0 \\ 0 & 0 \end{pmatrix} \right] dE\{N_i(t)\};$$

$$\begin{aligned}\mathcal{J} =& E\Bigg[\int_0^\tau \left\{ \hat{Z}_i - \frac{s^{(1)}(\hat{Z};\beta_c,t)}{s^{(0)}(\hat{Z};\beta_c,t)} + \begin{pmatrix} D(\beta_{c,x}) \\ 0 \end{pmatrix} \right\} dN_i(t) \\ &- \int_0^\tau \frac{Y_i(t)\exp(\hat{Z}_i^T\beta_c)}{s^{(0)}(\hat{Z};\beta_c,t)} \left\{ \hat{Z}_i - \frac{s^{(1)}(\hat{Z};\beta_c,t)}{s^{(0)}(\hat{Z};\beta_c,t)} \right\} dE\{N_i(t)\} \Bigg]^{\otimes 2};\end{aligned}$$

and $s^{(k)}(\hat{Z};\beta_c,t) = E\{Y_i(t)\hat{Z}_i^{\otimes k}\exp(\hat{Z}_i^T\beta_c)\}, k = 0, 1, 2.$

(3).

$$n^{1/2}\{\hat{\Lambda}_{cc}(t) - \Lambda_0(t)\} \rightsquigarrow \mathcal{G}(t) \quad in \quad l^\infty[0,\tau] \quad as \quad n \to \infty,$$

where $\rightsquigarrow$ *means weak convergence,* $l^\infty[0,\tau]$ *is the space of all bounded functions on* $[0,\tau]$ *(van der Vaart and Wellner 1996), and* $\mathcal{G}(t)$ *is a zero-mean Gaussian process* *with covariance function* $\Phi(s,t) = \exp(\gamma_0^2\beta_{c,x}^2)E\{\Psi_i(s)\Psi_i(t)\}$ *at* (s,t), *and*

$$\Psi_i(t) = -\int_0^t \eta_0(\beta_{c,x}) \left[\frac{s^{(1)}(\hat{Z};\beta_c,s)}{\{s^{(0)}(\hat{Z};\beta_c,s)\}^2} - \frac{1}{s^{(0)}(\hat{Z};\beta_c,s)} \begin{pmatrix} D(\beta_{c,x}) \\ 0 \end{pmatrix} \right]^T dE\{N_i(s)\} \times \mathcal{I}^{-1}$$
$$\times \left[\int_0^\tau \left\{ \hat{Z}_i - \frac{s^{(1)}(\hat{Z};\beta_c,t)}{s^{(0)}(\hat{Z};\beta_c,t)} + \begin{pmatrix} D(\beta_{c,x}) \\ 0 \end{pmatrix} \right\} dN_i(t) \right.$$
$$\left. - \int_0^\tau \frac{Y_i(t)\exp(\hat{Z}_i^T\beta_c)}{s^{(0)}(\hat{Z};\beta_c,t)} \left\{ \hat{Z}_i - \frac{s^{(1)}(\hat{Z};\beta_c,t)}{s^{(0)}(\hat{Z};\beta_c,t)} \right\} dE\{N_i(t)\} \right]$$
$$+ \int_0^\tau \frac{\eta_0(\beta_{c,x})}{s^{(0)}(\hat{Z};\beta_c,t)} \left\{ dN_i(t) - \eta_0^{-1}(\beta_{c,x})Y_i(t)\exp(\hat{Z}_i^T\beta_c)\exp(-\gamma_0\beta_{c,x})\lambda_0(t)dt \right\}.$$

2.4.2 Method for Scenario B

When the measurement error process is described by the error model (2.10) under Scenario B with $\gamma_0 = 0$, $\gamma_x = 1$, and $\gamma_v = 0$, we propose the corrected log profile likelihood

$$\ell_c(\beta,\Lambda_0) = \sum_{i\in\mathcal{M}} \left[\delta_i \left\{ \hat{Z}_i^T\beta + \log\Lambda_0\{s_i\} \right\} - \eta_0^{-1}(\beta_x) \sum_{\{j:s_j\le s_i\}} \exp(\hat{Z}_i^T\beta)\Lambda_0\{s_j\} \right],$$

which satisfies the property (2.19).

Mimicking the arguments in Section 2.4.1, we obtain the corrected profile score functions,

$$U_c(\beta) = \sum_{i\in\mathcal{M}} \left(\int_0^\tau \left\{ \hat{Z}_i - \frac{S^{(1)}(\hat{Z};\beta,t)}{S^{(0)}(\hat{Z};\beta,t)} + \begin{pmatrix} \hat{D}(\beta_x) \\ 0 \end{pmatrix} \right\} dN_i(t), \right.$$

where

$$\hat{D}(\beta_x) = \frac{\sum_{i\in\mathcal{V}} \epsilon_i \exp(\beta_x\epsilon_i)}{\sum_{i\in\mathcal{V}} \exp(\beta_x\epsilon_i)},$$

and $S^{(k)}(\hat{Z};\beta,t) = n^{-1}\sum_{i\in\mathcal{M}} Y_i(t)\hat{Z}_i^k \exp(\hat{Z}_i^T\beta)$ for $k = 0$ and 1. Let $\hat{\beta}_c$ be the solution of $U_c(\beta) = 0$, and let β_c denote the limit that $\hat{\beta}_c$ converges to in probability.

Next, we consider the general error model (2.10) in Scenario B where γ_0, γ_x, and γ_v are unknown. It is important to note that under this scenario, the Cox model and the error model are identifiable. With an external validation sample, the parameters $\gamma_0, \gamma_x, \gamma_v$ in the error model (2.10) can be consistently obtained by the least square method. Let $\hat{\gamma}_0, \hat{\gamma}_x, \hat{\gamma}_v$ denote the corresponding least square estimators for γ_0, γ_x, and γ_v, respectively. We define the corrected estimator $\hat{\beta}_{cc}$ through (2.26) and (2.27), with $\gamma_0, \gamma_x, \gamma_v$ replaced by $\hat{\gamma}_0, \hat{\gamma}_x, \hat{\gamma}_v$, respectively. The asymptotic properties of $\hat{\beta}_{cc}$ are given in the following theorem, and its proof is deferred to Appendix A4.

Theorem 2 *Under the regularity conditions R1-R6 listed in Appendix A1, we have*

(1). $\beta_{c,x} = \gamma_x^{-1}\beta_{0,x}$, *and* $\beta_{c,v} = \beta_{0,v} - \gamma_v\gamma_x^{-1}\beta_{0,x}$.

(2). $n^{1/2}(\hat{\beta}_{cc} - \beta_0) \xrightarrow{d} N(0, \mathcal{B})$, *as* $n \to \infty$, *where* $\mathcal{B}$ *is defined in Appendix A4.*

The estimate of the cumulative hazard function $\Lambda_0(t)$ has asymptotic results analogous to those in Lemma 1 and Theorem 1 under Scenario A.

2.4.3 Method for Scenario C

When the measurement error process is featured as in (2.13) under Scenario C, we propose the corrected log profile likelihood by mimicking (2.20):

$$\ell_c(\beta, \Lambda_0) = \sum_{i=1}^{n}\left[\delta_i\left\{\hat{Z}_i^T\beta + \log\Lambda_0\{s_i\}\right\} - \eta_0^{-1}(\beta_x)\sum_{\{j:s_j\le s_i\}} n_i^{-1}\sum_{r=1}^{n_i}\exp(\hat{Z}_{ir}^T\beta)\Lambda_0\{s_j\}\right],$$

which satisfies the property (2.19). Adapting the arguments in Scenario A, we obtain the corrected profile score functions

$$U_c(\beta) = \sum_{i=1}^{n}\int_0^{\tau}\left\{\hat{Z}_i - \frac{S_{re}^{(1)}(\hat{Z};\beta,t)}{S_{re}^{(0)}(\hat{Z};\beta,t)} + \begin{pmatrix}\hat{D}(\beta_x)\\ 0\end{pmatrix}\right\}dN_i(t), \tag{2.28}$$

where $S_{re}^{(k)}(\hat{Z};\beta,t) = n^{-1}\sum_{i=1}^{n} Y_i(t) n_i^{-1}\sum_{r=1}^{n_i}\hat{Z}_{ir}^k\exp(\hat{Z}_{ir}^T\beta)$ for $k=0$ and 1, and $\hat{D}(\beta_x)$ is a consistent estimate of $D(\beta_x)$ for a given β_x. Let $\hat{\beta}_c$ be the solution of (2.28).

Here we describe an expression of $\hat{D}(\beta_x)$. Note that

$$\begin{aligned} E\{W_{ir}\exp(W_{is}\beta_x)\} &= \eta_0(\beta_x)E\{X_i\exp(X_i\beta_x)\}, \\ E\{\exp(W_{ir}\beta_x)\} &= \eta_0(\beta_x)E\{\exp(X_i\beta_x)\}, \\ \text{and}\quad E\{W_{ir}\exp(W_{ir}\beta_x)\} &= \eta_0(\beta_x)E\{X_i\exp(X_i\beta_x)\} + \eta_1(\beta_x)E\{\exp(X_i\beta_x)\}, \end{aligned}$$

then $D(\beta_x) = \eta_0^{-1}(\beta_x)\eta_1(\beta_x)$ can be expressed as

$$\begin{aligned} D(\beta_x) &= \frac{\eta_0^{-1}(\beta_x)E\{W_{ir}\exp(W_{ir}\beta_x)\} - E\{X_i\exp(X_i\beta_x)\}}{E\{\exp(X_i\beta_x)\}} \\ &= \frac{E\{W_{ir}\exp(W_{ir}\beta_x)\} - E\{W_{ir}\exp(W_{is}\beta_x)\}}{E\{\exp(W_{ir}\beta_x)\}}. \end{aligned} \tag{2.29}$$

In view of (2.29), we can consistently estimate $D(\beta_x)$ by

$$\frac{\sum_{i=1}^n n_i^{-1}\sum_{r=1}^{n_i} W_{ir}\exp(W_{ir}\beta_x) - \sum_{i=1}^n I(n_i > 1)n_i^{-1}(n_i - 1)^{-1}\sum_{r\neq s} W_{ir}\exp(W_{is}\beta_x)}{\sum_{i=1}^n n_i^{-1}\sum_{r=1}^{n_i}\exp(W_{ir}\beta_x)}$$

for any fixed β_x. Alternatively, if $\hat{D}_{SH}(\beta_x)$ or $\hat{D}_{HL}(\beta_x)$ is respectively used for $\hat{D}(\beta_x)$ in (2.28), then the resulting estimator is identical to the corrected score estimator proposed by Song and Huang (2005) or the semiparametric regression estimator by Hu and Lin (2004), respectively.

Next, we consider the general error model (2.11) in Scenario C. Let $\hat{\beta}_{cc}$ be a corrected estimator defined through (2.26) and (2.27), and β_c be the limit to which $\hat{\beta}_c$ converges in probability. The following theorem descibes the asymptotic properties of $\hat{\beta}_{cc}$, and its proof is deferred to Appendix A5.

Theorem 3 *Under the regularity conditions R1-R6 listed in Appendix A1, we have*

(1). $\beta_{c,x} = \gamma_x^{-1}\beta_{0,x}$, *and* $\beta_{c,v} = \beta_{0,v} - \gamma_v\gamma_x^{-1}\beta_{0,x}$.

(2). $n^{1/2}(\hat{\beta}_{cc} - \beta_0) \xrightarrow{d} N(0, \mathcal{D}^T \mathcal{I}^{-1T} \mathcal{J}_2 \mathcal{I}^{-1} \mathcal{D}),$ *as* $n \to \infty$, *where*

$$\mathcal{J}_2 = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \Bigg[\int_0^\tau \left\{ \hat{Z}_i - \frac{s^{(1)}(\hat{Z}; \beta_c, t)}{s^{(0)}(\hat{Z}; \beta_c, t)} + \begin{pmatrix} D(\beta_{c,x}) \\ 0 \end{pmatrix} \right\} dN_i(t)$$
$$- \int_0^\tau \left[n_i^{-1} \sum_{r=1}^{n_i} \frac{Y_i(t) \exp(\hat{Z}_{ir}^T \beta_c)}{s^{(0)}(\hat{Z}; \beta_c, t)} \left\{ \hat{Z}_{ir} - \frac{s^{(1)}(\hat{Z}; \beta_c, t)}{s^{(0)}(\hat{Z}; \beta_c, t)} \right\} - \begin{pmatrix} D_i(\beta_{c,x}) \\ 0 \end{pmatrix} \right] dE\{N_i(t)\} \Bigg]^{\otimes 2},$$

$s^{(k)}(\hat{Z}; \beta_c, t) = E\{Y_i(t) \hat{Z}_{ir}^{\otimes k} \exp(\hat{Z}_{ir}^T \beta_c)\}, k = 0, 1, 2,$ *and* $D_i(\beta_{c,x})$ *is specified in Appendix A5.*

The estimate of the cumulative hazard function $\Lambda_0(t)$ has asymptotic results analogous to those in Lemma 1 and Theorem 1 under Scenario A.

2.4.4 Application to Berkson error model

The adjustment methods described for Scenarios A-C are potentially useful for other models as well. To see this, we consider the case that the true error model is the Berkson error model:

$$X_i = W_i + \epsilon_i, \tag{2.30}$$

where we assume that both of W_i and the error term ϵ_i are normal. ϵ_i is assumed to be independent of other variables and has mean zero and variance σ_0^2. The variance σ_0^2 is known or estimated from additional data sources. Let σ_w^2 be the variance of W_i.

Note that the error model (2.30) can be rewritten as

$$W_i = \gamma_0 + X_i \gamma_x + \epsilon_i^*, \tag{2.31}$$

where $\gamma_x = \sigma_w^2/(\sigma_w^2 + \sigma_0^2)$, $\gamma_0 = 1 - \gamma_x$, and ϵ_i^* is the error term with mean 0 and variance $\gamma_x \sigma_0^2$, and is independent of X_i. The model (2.31) can be equivalently expressed as

$$\gamma_x^{-1/2} W_i = \gamma_x^{-1/2} \gamma_0 + X_i \gamma_x^{1/2} + \gamma_x^{-1/2} \epsilon_i^*,$$

where the variance of $\gamma_x^{-1/2}\epsilon_i^*$ is now equal to that of ϵ_i.

We replace $\hat{Z}_i$ by $\hat{Z}_i^* = (\gamma_x^{-1/2} W_i, V_i^T)^T$ in the working likelihood (2.20), and obtain the working estimator $\hat{\beta}_c$ and $\hat{\Lambda}_c(\cdot)$ by maximizing the working likelihood. By Lemma 1 and Theorem 1, we obtain that

$$\begin{aligned} \hat{\beta}_{cc,x} = \gamma_x^{1/2}\hat{\beta}_{c,x}, \quad \hat{\beta}_{cc,v} = \hat{\beta}_{c,v}, \\ \text{and} \quad \hat{\Lambda}_{cc}(\cdot) = \exp(\gamma_0^{-1}\gamma_x^{1/2}\hat{\beta}_{c,x})\hat{\Lambda}_c(\cdot), \end{aligned}$$

are consistent estimators of $\beta_{0,x}$,$\beta_{0,v}$, and $\Lambda_0(\cdot)$, respectively.

We note that when the normality assumption of W_i or ϵ_i does not hold, then $E[\epsilon_i^*|X_i] = 0$ and $E[W_i|X_i] = \gamma_0 + X_i\gamma_x$ may no longer hold, and the resulting estimator $\hat{\beta}_{cc}$ could be biased.

2.5 Empirical studies

We conduct simulation studies to evaluate the finite sample performance of the proposed methods under Scenarios A-C. In addition, we uncover the impact of model misspecification of the measurement error process. We generate 1000 simulations for each parameter configuration, and let $Z_i = (X_i, V_i)^T$ be a 2×1 vector of covariates.

2.5.1 General error model

We consider two cases for covariates. In Case 1, the Z_i are uniformly simulated from the bivariate normal distribution:

$$Z_i \sim N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right\}, \quad i = 1, \cdots, n.$$

In Case 2, the X_i are uniformly simulated from the exponential distribution $X_i \sim \text{Exp}(1)$, and V_i follows a Bernoulli distribution with

$$\Pr(V_i = 1|X_i) = \frac{\exp(X_i)}{1 + \exp(X_i)}.$$

This gives that the correlation of X_i and V_i is about 0.27.

Survival times T_i are independently generated using the Cox model (2.1), where the true parameter $\beta = (1, 1)^T$, and the baseline hazard function $\lambda_0(t) = t$. About 30% of censoringship is generated. In particular, the censoring time C_i are generated from UNIF$[0, c]$, where c is set to be 5.4 and 2.05 for the first and second case, respectively.

We consider Scenarios A, B and C for the measurement error process. In Scenario A, we set $n = 200$, and generate $\epsilon_i \sim N(0, \sigma_0^2)$, where σ_0 is known. Take σ_0 to be 0.25 or 0.5 to represent different degrees of measurement error. Set $\gamma_0 = 0$, $\gamma_v = 1$ and let γ_x be 1.25 or 1.5. To estimate the β parameter, we consider three methods. The first method ignores measurement error by maximizing the partial likelihood (2.2) with Z_i replaced by $\hat{Z}_i$; let $\hat{\beta}_{nv}$ denote the resulting naive estimator. The second method comes from Nakamura (1992), and let $\hat{\beta}_c$ denote the resulting estimator. The third approach is the proposed method described in Section 2.4.1, and we let $\hat{\beta}_{cc}$ denote the corresponding estimator. For comparison, we also calculate $\hat{\beta}$ which maximizes (2.2) with the true covariate X_i treated available.

In Scenario B, we set $n = 300$, $m = 100$, and generate ϵ_i from the $N(0, \sigma_0^2)$ distribution, where σ_0 is set to be 0.25 or 0.5. $\hat{\beta}_c$ and $\hat{\beta}_{cc}$ are defined in Section 2.4.2. In Scenario C, we set $n = 200$, $n_i = 2$ for $i = 1, \cdots, n$, and generate ϵ_{ir} from the $N(0, \sigma_0^2)$ distribution, where σ_0 is set to be 0.25 or 0.5. $\hat{\beta}_c$ and $\hat{\beta}_{cc}$ are defined in Section 2.4.3, where we set $\hat{D}(\beta_x)$ to be $\hat{D}_{HL}(\beta_x)$. In both Scenarios B and C, we set $\gamma_0 = 0$, $\gamma_v = 1$ and $\gamma_x = 1.25$ or 1.5.

Tables 2.1, 2.2, and 2.3 present the empirical results for Scenarios A, B, and C, respectively, where we report the finite sample biases (Bias), the empirical variances (EVE), the average of the model-based variance estimates (MVE), the mean square errors (MSE), and the coverage rate of 95% confidence intervals.

It is clearly seen that both $\hat{\beta}_{nv}$ and $\hat{\beta}_c$ incur considerable biases, and the coverage rates for the 95% confidence intervals remarkably deviate from the nominal level. In some situations, $\hat{\beta}_c$ performs even worse than $\hat{\beta}_{nv}$. These demonstrates that ignoring measurement error would lead to biased results; secondly, the attempt to correct for error effects may be useless or even worse than not doing so if measurement error can not be reasonably captured.

On the other hand, our proposed estimator $\hat{\beta}_{cc}$ performs satisfactorily in all the situations. Finite sample biases are fairly small, and model-based variances estimates agree well with empirical variance. The coverage rates of the 95% confidence intervals are in good agreement with the nominal level. As expected, the estimate $\hat{\beta}$ performs best with the smallest finite sample biases and variance estimates. Reasonable agreement between the results for $\hat{\beta}_{cc}$ and $\hat{\beta}$ further suggests that our method performs reliably.

[Insert Tables 2.1, 2.2, 2.3 here!]

2.5.2 Application to Berkson error model

Different from the covariate generation discussed in Section 2.5.1, here we consider three cases for which surrogates W_i are first simulated, and then the true covariate X_i will be generated from the Berkson error model $X_i = W_i + \epsilon_i$, where ϵ_i follows a normal distribution $N(0, \sigma_0^2)$ with $\sigma_0 = 0.25$ or 0.5. We set $n = 200$.

In Case 1$'$, both of W_i and V_i are independently simulated from the standard normal distribution $N(0,1)$. In Cases 2$'$ and 3$'$, we generate W_i from the exponential distribution Exp(1), and use different ways to simulate the V_i covariate. In Case 2$'$, independent of W_i, we generate V_i from the standard normal distribution $N(0,1)$, while in Case 3$'$, conditional on X_i we generate V_i from a Bernoulli distribution $\Pr(V_i = 1|X_i) = \exp(X_i)/\{1+\exp(X_i)\}$.

Survival times and censoring times are generated using the same procedure as in Section 2.5.1. The only difference is to specify a different value for c to yield about 30%

censoringship for each when $\sigma_0 = 0.25$. In particular, set c to be 5.1, 3.1 and 2.1 for Cases $1'$, $2'$ and $3'$, respectively.

Table 2.4 records the simulation results. Same patterns as those in Section 2.5.1 are demonstrated here. In addition, we notice that with small measurement error, ignoring measurement error or misspecifying measurement error model does not incur noticeably bias. When measurement error becomes moderate, both $\hat{\beta}_{nv}$ and $\hat{\beta}_c$ incur considerable bias with useless confidence intervals produced. More strikingly, $\hat{\beta}_c$ even yields worse results than the naive estimator $\hat{\beta}_{nv}$, and this suggests that leaving measurement error unattended to is even better than attempting to correct it if there is not good knowledge about modeling the measurement error process. In comparison, the proposed corrected estimator $\hat{\beta}_{cc}$ successfully corrects the bias induced by measurement error. In all the cases, $\hat{\beta}_{cc}$ outperforms both $\hat{\beta}_{nv}$ and $\hat{\beta}_c$, even in Cases $2'$ and $3'$ where W_i is not normal.

[Insert Table 2.4 here!]

2.5.3 An example

We apply the proposed method to analyze the data of the AIDS Clinical Trials Group (ACTG) 175 (Hammer, et al. 1996). The ACTG 175 was a double-blind randomized clinical trial, comparing the effects of three HIV treatments for which three drugs were used in combination or alone: zidovudine, didanosine, and zalcitabine.

There were $n = 2139$ individuals in this study whose baseline measurements on CD4 were collected, ranging from 200 to 500 per cubic millimeter. Following the definition of Hammer et al. (1996), T_i is defined to be the time to the occurrence of the first event among the following events: (i) more than 50% decline of CD4 counts compared to the averaged baseline CD4 counts; (ii) disease progression to AIDS; or (iii) death. About 75.6% of outcome values are censored.

We are interested in studying the relationship how T_i is associated with baseline CD4 counts and how treatment may affect outcome variable T_i. We let V_i denote the treatment indicator for subject i, where $V_i = 1$ if a subject receive one of the three treatments, and 0 otherwise. We define X_i to be $\log(\text{CD4 counts} + 1)$, a usual normalization version of CD4 counts.

We employ the Cox model to feature the dependence of T_i on the covariates X_i and V_i:

$$\lambda(t) = \lambda_0(t) \exp(X_i \beta_x + V_i \beta_v),$$

where $\lambda_0(t)$ is the baseline hazard function, and $\beta = (\beta_x, \beta_v)^T$ is the regression parameter.

It is well known that CD4 counts are subject to measurement error due to biological variation and imprecise measurement procedures. The CD4 counts X_i for subject i are repeatedly measured twice, and their measurements are denoted by $W_{ir}, r = 1, 2$. It is not clear how exactly the true value X_i and the surrogates W_{ir} are linked. To assess how sensitive the analysis could be to various degrees of measurement error, we consider the model (2.11) in Scenario C.

With the replicate measurements W_{ir}, we can estimate the variance σ_0^2 of the error term ϵ_{ir}, but the parameters γ_0, γ_x, and γ_v are not identifiable and thus not estimable. To overcome the nonidentifiablity issue, we conduct sensitivity analyses by specifying the values of γ_0, γ_x, and γ_v to feature different measurement error situations.

We particularly consider two cases. For the first case, we set $\gamma_0 = 0$, $\gamma_v = 0$ and let γ_x vary from 0.5 to 1.5. For the second case, we set $\gamma_0 = 0$, $\gamma_x = 1$ and let γ_x vary from -0.5 to 0.5. In both cases, we are interested in evaluating the impact of measurement error on parameter estimation and associated confidence intervals for the response parameters. The estimator $\hat{\beta}_{cc}$ and its variance estimate are obtained by the procedure in Section 2.4.3, where we set $\hat{D}(\beta_x)$ to be $\hat{D}_{HL}(\beta_x)$. Results are reported in Figure 2.1. Although the covariate effects and the length of confidence intervals are differently estimated under different specification of measurement error models, all the analyses suggest significance of CD4 counts and treatment on affecting the outcome variable.

[Insert Figure 2.1 here!]

2.6 Discussion

Survival data with covariate measurement error have attracted extensive research attention. Many inference methods have been proposed for the Cox model with covariate measurement error using various techniques. In this chapter, we develop a new inference method to address covariate measurement error effects under the Cox model. The proposed corrected profile likelihood provides a simple and general way to correct for covariate measurement error, and it can accommodate many existing methods as special cases. Moreover, our method applies to a larger scope of measurement error scenarios than many available methods which focus on the classical additive error model or the Berkson error model. In addition, we study the impact of model misspecification of the measurement error process, and interesting findings are obtained. Attempting to correct for measurement error effects is not always rewarding; sometimes it can lead to more misleading results than ignoring measurement error. It is critical to correctly model the measurement error process, in order to develop valid inference methods.

Appendix

Appendix A1: Regularity conditions

R1. $\{N_i(\cdot), Y_i(\cdot), Z_i\}, i = 1, \cdots, n$ are independent and identically distributed.

R2. $\Pr\{Y_1(\tau) = 1\} > 0$.

R3. $\Lambda_0(\tau) < \infty$, and $\Lambda_0(t)$ is absolutely continuous over $[0, \tau]$.

R4. The parameter space for β is a compact subspace of the Euclidean space.

R5. $||E(Z_1^{\otimes 2})|| < \infty$, $||E(\epsilon_1^{\otimes 2})|| < \infty$, and $\log\{\eta_0(\beta_x)\}$ is twice continuously differentiable. Here for a matrix A, $||A|| = \sup_{i,j} |a_{ij}|$, where a_{ij} is the (i, j)th element of A.

R6. Condition D of Andersen and Gill (1982) holds.

Appendix A2: Proof of Lemma 1

Let $\beta^* = (\beta_x^*, \beta_v^{*T})^T$, $\beta_x^* = \gamma_x\beta_{c,x}$, $\beta_v^* = \beta_{c,v} + \gamma_v\beta_{c,x}$, and $\Lambda^*(\cdot) = \exp(\gamma_0\beta_{c,x})\Lambda_c(\cdot)$. Here, we need to prove $(\beta^*, \Lambda^*) = (\beta_0, \Lambda_0)$. Indeed, $\hat{\beta}_c$ is the root of

$$U_c(\beta) = \sum_{i=1}^n \int_0^\tau \left[\hat{Z}_i - \frac{\sum_{i=1}^n Y_i(t)\hat{Z}_i \exp(\hat{Z}_i^T\beta)}{\sum_{i=1}^n Y_i(t)\exp(\hat{Z}_i^T\beta)} + \begin{pmatrix} D(\beta_x) \\ 0 \end{pmatrix}\right] dN_i(t).$$

Furthermore, it is seen that

$$\hat{\Lambda}_c(t, \hat{\beta}_c) = \sum_{i=1}^n \int_0^t \frac{dN_i(s)}{\eta_0^{-1}(\hat{\beta}_{c,x})\sum_{i=1}^n Y_i(s)\exp(\hat{Z}_i^T\hat{\beta}_c)}.$$

By the Uniform Strong Laws of Large Numbers (Pollard, 1990), $n^{-1}U_c(\beta)$ converges almost surely to

$$\mathcal{U}_c(\beta) = \int_0^\tau \left[E\{\hat{Z}_i dN_i(t)\} - \frac{E\{Y_i(t)\hat{Z}_i\exp(\hat{Z}_i^T\beta)\}}{E\{Y_i(t)\exp(\hat{Z}_i^T\beta)\}} dE\{N_i(t)\} + \begin{pmatrix} D(\beta_x) \\ 0 \end{pmatrix} dE\{N_i(t)\}\right].$$

Thus, by convex analysis arguments (Rockafellar 1970, Theorem 10.8; Struthers and Kalbfleisch 1986; Lin et al. 2000, Appendix A.1), β_c is the unique root of $\mathcal{U}_c(\beta) = 0$ when the error degree is not too large. Write $\mathcal{U}_c(\beta) = (\mathcal{U}_{c,x}(\beta), \mathcal{U}_{c,v}^T(\beta))^T$, then we obtain that

$$\begin{aligned}
\mathcal{U}_{c,x}(\beta_c) = &\int_0^\tau E\{Y_i(t)W_i\lambda_0(t)\exp(X_i\beta_{0,x} + V_i^T\beta_{0,v})\}dt - \int_0^\tau \frac{E\{Y_i(t)W_i\exp(W_i\beta_{c,x} + V_i^T\beta_{c,v})\}}{E\{Y_i(t)\exp(W_i\beta_{c,x} + V_i^T\beta_{c,v})\}} \\
&\times E\{Y_i(t)\lambda_0(t)\exp(X_i\beta_{0,x} + V_i^T\beta_{0,v})\}dt + D(\beta_{c,x})\int_0^\tau E\{Y_i(t)\lambda_0(t)\exp(X_i\beta_{0,x} + V_i^T\beta_{0,v})\}dt \\
= &\int_0^\tau E\{Y_i(t)(X_i\gamma_x + V_i^T\gamma_v)\lambda_0(t)\exp(X_i\beta_{0,x} + V_i^T\beta_{0,v})\}dt \\
&- \int_0^\tau \frac{E[Y_i(t)(X_i\gamma_x + V_i^T\gamma_v)\exp\{X_i\gamma_x\beta_{c,x} + V_i^T(\beta_{c,v} + \gamma_v\beta_{c,x})\}]}{E[Y_i(t)\exp\{X_i\gamma_x\beta_{c,x} + V_i^T(\beta_{c,v} + \gamma_v\beta_{c,x})\}]} \\
&\times E\{Y_i(t)\lambda_0(t)\exp(X_i\beta_{0,x} + V_i^T\beta_{0,v})\}dt = 0,
\end{aligned}$$

and

$$
\begin{aligned}
&\mathcal{U}_{c,v}(\beta_c)\\
=&\int_0^\tau E\{Y_i(t)V_i\lambda_0(t)\exp(X_i\beta_{0,x}+V_i^T\beta_{0,v})\}dt\\
&-\int_0^\tau \frac{E\{Y_i(t)V_i\exp(W_i\beta_{c,x}+V_i^T\beta_{c,v})\}}{E\{Y_i(t)\exp(W_i\beta_{c,x}+V_i^T\beta_{c,v})\}}E\{Y_i(t)\lambda_0(t)\exp(X_i\beta_{0,x}+V_i^T\beta_{0,v})\}dt\\
=&\int_0^\tau E[Y_i(t)V_i\lambda_0(t)\exp(X_i\beta_{0,x}+V_i^T\beta_{0,v})]dt\\
&-\int_0^\tau \frac{E[Y_i(t)V_i\exp\{X_i\gamma_x\beta_{c,x}+V_i^T(\beta_{c,v}+\gamma_v\beta_{c,x})\}]}{E[Y_i(t)\exp\{X_i\gamma_x\beta_{c,x}+V_i^T(\beta_{c,v}+\gamma_v\beta_{c,x})\}]}E\{Y_i(t)\lambda_0(t)\exp(X_i\beta_{0,x}+V_i^T\beta_{0,v})\}dt\\
=&0.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\mathcal{U}_{c,x}(\beta_c)-\gamma_v^T\mathcal{U}_{c,v}(\beta_c)\\
=&\gamma_x\int_0^\tau E\{Y_i(t)X_i\lambda_0(t)\exp(X_i\beta_{0,x}+V_i^T\beta_{0,v})\}dt\\
&-\gamma_x\int_0^\tau \frac{E[Y_i(t)X_i\exp\{X_i\gamma_x\beta_{c,x}+V_i^T(\beta_{c,v}+\gamma_v\beta_{c,x})\}]}{E[Y_i(t)\exp\{X_i\gamma_x\beta_{c,x}+V_i^T(\beta_{c,v}+\gamma_v\beta_{c,x})\}]}E\{Y_i(t)\lambda_0(t)\exp(X_i\beta_{0,x}+V_i^T\beta_{0,v})\}dt\\
&=0.
\end{aligned}
$$

It follows that $g(\beta_0)-g(\beta^*)=0$, where

$$
g(\beta)=\int_0^\tau \frac{E\{Y_i(t)Z_i\exp(Z_i^T\beta)\}}{E\{Y_i(t)\exp(Z_i^T\beta)\}}E\{Y_i(t)\lambda_0(t)\exp(Z_i^T\beta_0)\}dt.
$$

Thus, both of β^* and β_0 are solutions of $g(\beta_0)-g(\beta)=0$.

To show $\beta^*=\beta_0$, it suffices to prove that $g(\beta_0)-g(\beta)=0$ has only a unique solution. To see this, let

$$
\begin{aligned}
f(\beta)=&\int_0^\tau E\{Y_i(t)Z_i^T\beta\lambda_0(t)\exp(Z_i^T\beta_0)\}dt\\
&-\int_0^\tau \log[E\{Y_i(t)\exp(Z_i^T\beta)\}]E\{Y_i(t)\lambda_0(t)\exp(Z_i^T\beta_0)\}dt.
\end{aligned}
$$

It follows that, by Cauchy-Schwarz inequality, the Hessian matrix of $f(\beta)$, i.e.,

$$\frac{\partial^2 f(\beta)}{\partial\beta\partial\beta^T} = -\int_0^\tau \left(\frac{E\{Y_i(t) Z_i^{\otimes 2} \exp(Z_i^T\beta)\}}{E\{Y_i(t)\exp(Z_i^T\beta)\}} - \left[\frac{E\{Y_i(t) Z_i \exp(Z_i^T\beta)\}}{E\{Y_i(t)\exp(Z_i^T\beta)\}} \right]^{\otimes 2} \right)$$
$$\times E\{Y_i(t)\lambda_0(t)\exp(Z_i^T\beta_0)\} dt$$

is always negative definite for all β in the parameter space. It follows that $f(\beta)$ is strictly concave (Boyd and Vandenberghe 2004) and thus has a unique maximum. Since the necessary and sufficient condition for β to be a maximum point of the strictly convex function $f(\beta)$ is that $\partial f(\beta)/\partial\beta = 0$ (Boyd and Vandenberghe 2004, (4.22)), thus setting $\partial f(\beta)/\partial\beta = 0$ has at most one solution. Note that $\partial f(\beta)/\partial\beta = g(\beta_0) - g(\beta)$, and we already showed that both β_0 and β^* are solutions of $g(\beta_0) - g(\beta) = 0$, thus $\beta^* = \beta_0$.

It remains to prove $\Lambda^*(\cdot) = \Lambda_0(\cdot)$. To see this, note that $\Lambda_c(t)$ is the limit to which $\hat{\Lambda}_c(t, \hat{\beta}_c)$ converges almost surely as $n \to \infty$. Equivalently, $\Lambda_c(t)$ can be regarded as the limit to which $\hat{\Lambda}_c(t, \beta_c)$ converges almost surely due to that $\hat{\beta}_c$ converges to β_c almost surely. By the Uniform Strong Law of Large Numbers, $\hat{\Lambda}_c(t, \beta_c)$ converges almost surely to

$$\int_0^t \frac{E\{dN_i(s)\}}{\eta_0^{-1}(\beta_{c,x}) E\{Y_i(s)\exp(\hat{Z}_i^T\beta_c)\}} = \int_0^t \frac{E\{Y_i(s)\exp(Z_i^T\beta_0)\lambda_0(s)\} ds}{\eta_0^{-1}(\beta_{c,x}) E\{Y_i(s)\exp(\hat{Z}_i^T\beta_c)\}} = \frac{\Lambda_0(t)}{\exp(\gamma_0\beta_{c,x})},$$

uniformly in t. Thus, $\Lambda_c(t) = \Lambda_0(t)/\exp(\gamma_0\beta_{c,x})$ for all t. It follows that that $\Lambda_0(t) = \exp(\gamma_0\beta_{c,x})\Lambda_c(t) = \Lambda^*(t)$ for all t, which completes the proof.

Appendix A3: Proof of Theorem 1

Proof of (1) of Theorem 1: The strong consistency of $\hat{\beta}_{cc}$ follows directly from Lemma 1 and the Slusky's Lemma (van der Vaart 1998). Similarly, by the fact that $\hat{\Lambda}_c(t, \hat{\beta}_c)$ converges to $\Lambda_c(t)$ almost surely uniformly in t as proved in Lemma 1, we obtain that $\hat{\Lambda}_{cc}(t, \hat{\beta}_c)$ converges to $\Lambda_0(t)$ almost surely uniformly in t.

Proof of (2) of Theorem 1: By the delta method (van der Vaart 1998) and the Lemma, we only need to show that

$$n^{1/2}(\hat{\beta}_c - \beta_c) \xrightarrow{d} N(0, \mathcal{I}^{-1T}\mathcal{J}\mathcal{I}^{-1}), \quad \text{as} \quad n \to \infty. \tag{2.32}$$

To show this, note that by the Taylor series expansion, we obtain that

$$n^{1/2}(\hat{\beta}_c - \beta_c) = \left\{ -n^{-1}\frac{\partial U_c(\beta)}{\partial \beta^T}|_{\beta=\tilde{\beta}} \right\}^{-1} n^{-1/2}U_c(\beta_c),$$

where $\tilde{\beta}$ is on the line segment between $\hat{\beta}_c$ and β_c, and $U_c(\beta)$ is defined in Lemma 1. By the Strong Law of Large Numbers, it follows that

$$-n^{-1}\frac{\partial U_c(\beta)}{\partial \beta^T}|_{\beta=\tilde{\beta}} \xrightarrow{a.s.} \mathcal{I}, \quad \text{as} \quad n \to \infty.$$

Now we write $U_c(\beta_c) = U_{c1}(\beta_c) + U_{c2}(\beta_c)$, where

$$U_{c1}(\beta_c) = \sum_{i=1}^{n} \int_0^{\tau} \left[\hat{Z}_i - \frac{\sum_{i=1}^{n} Y_i(t)\hat{Z}_i \exp(\hat{Z}_i^T \beta_c)}{\sum_{i=1}^{n} Y_i(t)\exp(\hat{Z}_i^T \beta_c)} \right] dN_i(t),$$

and $U_{c2}(\beta_c) = (D(\beta_{c,x}), 0^T)^T \sum_{i=1}^{n} N_i(\tau)$. By the proof of Theorem 2.1 of Lin and Wei (1989),

$$\begin{aligned} n^{-1/2}U_{c1}(\beta_c) = n^{-1/2} &\sum_{i=1}^{n} \int_0^{\tau} \left\{ \hat{Z}_i - \frac{s^{(1)}(\hat{Z};\beta_c,t)}{s^{(0)}(\hat{Z};\beta_c,t)} \right\} dN_i(t) \\ - n^{-1/2} &\sum_{i=1}^{n} \int_0^{\tau} \frac{Y_i(t)\exp(\hat{Z}_i^T\beta_c)}{s^{(0)}(\hat{Z};\beta_c,t)} \left\{ \hat{Z}_i - \frac{s^{(1)}(\hat{Z};\beta_c,t)}{s^{(0)}(\hat{Z};\beta_c,t)} \right\} dE\{N_i(t)\} + o_p(1). \end{aligned}$$

It follows that $n^{-1/2}U_c(\beta_c)$ is asymptotically equivalent to a sum of i.i.d. random vectors, and thus (2.32) is proved by using the Central Limit Theorem, and this completes the proof.

Proof of (3) of Theorem 1: By the functional delta method (van der Vaart 1998) and Lemma 1, we only need to show that

$$n^{1/2}\{\hat{\Lambda}_c(t) - \Lambda_c(t)\} \rightsquigarrow \mathcal{G}_c(t) \;\; \text{in} \;\; l^{\infty}[0,\tau] \;\; \text{as} \;\; n \to \infty \tag{2.33}$$

where $\mathcal{G}_c(t)$ is a zero-mean Gaussian process with covariance function $E\{\Psi_i(s)\Psi_i(t)\}$ at (s,t). To show this, write $n^{1/2}\{\hat{\Lambda}_c(t) - \Lambda_c(t)\} = A_1 + A_2$, where

$$A_1 = n^{1/2}\left\{\sum_{i=1}^{n}\int_0^t \frac{dN_i(s)}{\eta_0^{-1}(\beta_{c,x})\sum_{i=1}^{n} Y_i(s)\exp(\hat{Z}_i^T\beta_c)} - \Lambda_c(t)\right\}$$

$$\text{and}\quad A_2 = n^{1/2}\Bigg\{\sum_{i=1}^{n}\int_0^t \frac{dN_i(s)}{\eta_0^{-1}(\hat{\beta}_{c,x})\sum_{i=1}^{n} Y_i(s)\exp(\hat{Z}_i^T\hat{\beta}_c)} - \sum_{i=1}^{n}\int_0^t \frac{dN_i(s)}{\eta_0^{-1}(\beta_{c,x})\sum_{i=1}^{n} Y_i(s)\exp(\hat{Z}_i^T\beta_c)}\Bigg\}.$$

Similar to the argument of Lin et al. (2000, Appendix A.4), we obtain that A_1 is tight and

$$\begin{aligned} A_1 &= n^{-1/2}\left\{\sum_{i=1}^{n}\int_0^t \frac{dN_i(s) - \eta_0^{-1}(\beta_{c,x})Y_i(s)\exp(\hat{Z}_i^T\beta_c)d\Lambda_c(t)}{\eta_0^{-1}(\beta_{c,x})n^{-1}\sum_{i=1}^{n} Y_i(s)\exp(\hat{Z}_i^T\beta_c)}\right\} \\ &= n^{-1/2}\left\{\sum_{i=1}^{n}\int_0^t \frac{dN_i(s) - \eta_0^{-1}(\beta_{c,x})Y_i(s)\exp(\hat{Z}_i^T\beta_c)d\Lambda_c(t)}{\eta_0^{-1}(\beta_{c,x})s^{(0)}(\hat{Z};\beta_c,t)}\right\} + o_p(1) \end{aligned}$$

uniformly in t. By the Taylor series expansion, $A_2 = H^T(\tilde{\beta},t)n^{1/2}(\hat{\beta}_c - \beta_c)$, where $\tilde{\beta}$ is on the line segment between $\hat{\beta}_c$ and β_c, and

$$\begin{aligned} H(\beta,t) = &-\sum_{i=1}^{n}\int_0^t \eta_0(\beta_x)\frac{\sum_{i=1}^{n} Y_i(s)\hat{Z}_i\exp(\hat{Z}_i^T\beta)}{\{\sum_{i=1}^{n} Y_i(s)\exp(\hat{Z}_i^T\beta)\}^2}dN_i(s) \\ &+\sum_{i=1}^{n}\int_0^t \frac{\eta_0(\beta_x)}{\sum_{i=1}^{n} Y_i(s)\exp(\hat{Z}_i^T\beta)}\begin{pmatrix} D(\beta_x) \\ 0 \end{pmatrix} dN_i(s). \end{aligned}$$

Note that by the Uniform Strong Law of Large Numbers, we have

$$H(\tilde{\beta},t) \xrightarrow{a.s.} -\int_0^t \eta_0(\beta_{c,x})\left[\frac{s^{(1)}(\hat{Z};\beta_c,s)}{\{s^{(0)}(\hat{Z};\beta_c,s)\}^2} - \frac{1}{s^{(0)}(\hat{Z};\beta_c,s)}\begin{pmatrix} D(\beta_{c,x}) \\ 0 \end{pmatrix}\right] dE\{N_i(s)\}$$

uniformly in t. Furthermore, since

$$\begin{aligned} n^{1/2}(\hat{\beta}_c - \beta_c) = \mathcal{I}^{-1}n^{-1/2}\sum_{i=1}^{n}\Bigg(&\int_0^\tau \left\{\hat{Z}_i - \frac{s^{(1)}(\hat{Z};\beta_c,t)}{s^{(0)}(\hat{Z};\beta_c,t)} + \begin{pmatrix} D(\beta_{c,x}) \\ 0 \end{pmatrix}\right\} dN_i(t) \\ &- \int_0^\tau \frac{Y_i(t)\exp(\hat{Z}_i^T\beta_c)}{s^{(0)}(\hat{Z};\beta_c,t)}\left\{\hat{Z}_i - \frac{s^{(1)}(\hat{Z};\beta_c,t)}{s^{(0)}(\hat{Z};\beta_c,t)}\right\} dE\{N_i(t)\}\Bigg) + o_p(1), \end{aligned}$$

we thus obtain $n^{1/2}\{\hat{\Lambda}_c(t) - \Lambda_c(t)\} = A_1 + A_2 = n^{-1/2}\sum_{i=1}^n \Psi_i(t) + o_p(1)$. As a result, $n^{1/2}\{\hat{\Lambda}_c(t) - \Lambda_c(t)\}$ is tight and converges weakly to a mean-zero Gaussian process, and the covariance function is $E\{\Psi_i(s)\Psi_i(t)\}$ at (s,t). Thus, (2.33) holds. The proof is thus completed.

Appendix A4: Proof of Theorem 2

(1) in Theorem 2 can be proved by the arguments of Lemma 1. In the following, we prove (2) in Theorem 2. Write

$$
\begin{aligned}
& n^{-1/2}\sum_{i\in\mathcal{M}}\int_0^\tau \begin{pmatrix} \hat{D}(\beta_{c,x}) \\ 0 \end{pmatrix} dN_i(t) \\
=& \begin{pmatrix} \hat{D}(\beta_{c,x}) - D(\beta_{c,x}) \\ 0 \end{pmatrix} n^{1/2}\left[\sum_{i\in\mathcal{M}} N_i(\tau)/n - E\{N_i(\tau)\}\right] \\
& + n^{1/2}\begin{pmatrix} \hat{D}(\beta_{c,x}) - D(\beta_{c,x}) \\ 0 \end{pmatrix} E\{N_i(\tau)\} + \begin{pmatrix} D(\beta_{c,x}) \\ 0 \end{pmatrix} n^{1/2}\sum_{i\in\mathcal{M}} N_i(\tau)/n.
\end{aligned}
$$

Since the first term is of order $o_p(1)$, we have by Taylor series expansion that

$$
\begin{aligned}
& n^{-1/2}\sum_{i\in\mathcal{M}}\int_0^\tau \begin{pmatrix} \hat{D}(\beta_{c,x}) \\ 0 \end{pmatrix} dN_i(t) \\
=& n^{1/2}\begin{pmatrix} \hat{D}(\beta_{c,x}) - D(\beta_{c,x}) \\ 0 \end{pmatrix} E\{N_i(\tau)\} + \begin{pmatrix} D(\beta_{c,x}) \\ 0 \end{pmatrix} n^{1/2}\sum_{i\in\mathcal{M}} N_i(\tau)/n + o_p(1) \\
=& \frac{E\{N_i(\tau)\}}{\eta_0(\beta_{c,x})} n^{1/2}\begin{pmatrix} \sum_{i\in\mathcal{V}} \epsilon_i \exp(\beta_{c,x}\epsilon_i)/m - \eta_1(\beta_{c,x}) \\ 0 \end{pmatrix} - \frac{E\{N_i(\tau)\}\eta_1(\beta_{c,x})}{\eta_0^2(\beta_{c,x})} \\
& \times n^{1/2}\begin{pmatrix} \sum_{i\in\mathcal{V}} \exp(\beta_{c,x}\epsilon_i)/m - \eta_0(\beta_{c,x}) \\ 0 \end{pmatrix} + \begin{pmatrix} D(\beta_{c,x}) \\ 0 \end{pmatrix} n^{1/2}\sum_{i\in\mathcal{M}} N_i(\tau)/n + o_p(1) \\
=& \frac{E\{N_i(\tau)\}}{\eta_0(\beta_{c,x})}\frac{n^{1/2}}{m}\sum_{i\in\mathcal{V}}\begin{pmatrix} \epsilon_i \exp(\beta_{c,x}\epsilon_i) - D(\beta_{c,x})\exp(\beta_{c,x}\epsilon_i) \\ 0 \end{pmatrix} + n^{-1/2}\sum_{i\in\mathcal{M}}\int_0^\tau \begin{pmatrix} D(\beta_{c,x}) \\ 0 \end{pmatrix} dN_i(t) \\
& + o_p(1).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
n^{1/2}(\hat{\beta}_c - \beta_c) =& \mathcal{I}^{-1} n^{-1/2} \sum_{i \in \mathcal{M}} \Bigg(\int_0^\tau \left\{ \hat{Z}_i - \frac{s^{(1)}(\hat{Z}; \beta_c, t)}{s^{(0)}(\hat{Z}; \beta_c, t)} + \begin{pmatrix} D(\beta_{c,x}) \\ 0 \end{pmatrix} \right\} dN_i(t) \\
& - \int_0^\tau \left[\frac{Y_i(t) \exp(\hat{Z}_i^T \beta_c)}{s^{(0)}(\hat{Z}; \beta_c, t)} \left\{ \hat{Z}_i - \frac{s^{(1)}(\hat{Z}; \beta_c, t)}{s^{(0)}(\hat{Z}; \beta_c, t)} \right\} \right] dE\{N_i(t)\} \Bigg) \\
& + \mathcal{I}^{-1} n^{-1/2} \sum_{i \in \mathcal{M}} \int_0^\tau \begin{pmatrix} \hat{D}(\beta_{c,x}) - D(\beta_{c,x}) \\ 0 \end{pmatrix} dN_i(t) + o_p(1) \\
=& (n+m)^{-1/2} (1+\rho)^{1/2} \mathcal{I}^{-1} \sum_{i=1}^{n+m} \Bigg\{ \xi_i \Bigg(\int_0^\tau \left\{ \hat{Z}_i - \frac{s^{(1)}(\hat{Z}; \beta_c, t)}{s^{(0)}(\hat{Z}; \beta_c, t)} + \begin{pmatrix} D(\beta_{c,x}) \\ 0 \end{pmatrix} \right\} dN_i(t) \\
& - \int_0^\tau \left[\frac{Y_i(t) \exp(\hat{Z}_i^T \beta_c)}{s^{(0)}(\hat{Z}; \beta_c, t)} \left\{ \hat{Z}_i - \frac{s^{(1)}(\hat{Z}; \beta_c, t)}{s^{(0)}(\hat{Z}; \beta_c, t)} \right\} \right] dE\{N_i(t)\} \Bigg) \\
& + (1-\xi_i) \frac{E\{N_i(\tau)\}}{\eta_0(\beta_{c,x})} \frac{1}{\rho} \begin{pmatrix} \epsilon_i \exp(\beta_{c,x}\epsilon_i) - D(\beta_{c,x}) \exp(\beta_{c,x}\epsilon_i) \\ 0 \end{pmatrix} \Bigg\} + o_p(1) \\
\equiv& (n+m)^{-1/2} \sum_{i=1}^{n+m} A_{val,i} + o_p(1).
\end{aligned}
$$

Since $A_{val,1}, \cdots, A_{val,n+m}$ are i.i.d., by the Central Limit Theorem, we obtain that

$$n^{1/2}(\hat{\beta}_c - \beta_c) \xrightarrow{d} N(0, \mathcal{A}), \quad \text{as} \quad n \to \infty,$$

where $\mathcal{A} = E(A_{val,i}^{\otimes 2})$.

By the least square estimation,

$$
\begin{aligned}
n^{1/2}(\hat{\gamma}_z - \gamma_z) =& n^{1/2} \left\{ \sum_{i \in \mathcal{V}} (Z_i - \mu_z)(Z_i - \mu_z)^T \right\}^{-1} \sum_{i \in \mathcal{V}} (Z_i - \mu_z)(W_i - \mu_w) + o_p(1) \\
=& n^{1/2} m^{-1} \{Var(Z_i)\}^{-1} \sum_{i \in \mathcal{V}} (Z_i - \mu_z)(W_i - \mu_w) + o_p(1).
\end{aligned}
$$

where $\mu_z = (\mu_x, \mu_v^T)^T = (E(X_i), E^T(V_i))^T$, and $\mu_w = E(W_i)$. Therefore,

$$
\begin{aligned}
n^{1/2}(\hat{\beta}_{cc} - \beta_0) =& n^{1/2} \begin{pmatrix} \hat{\gamma}_x & 0 \\ \hat{\gamma}_v & I \end{pmatrix} \hat{\beta}_c - n^{1/2} \begin{pmatrix} \gamma_x & 0 \\ \gamma_v & I \end{pmatrix} \beta_c \\
=& \begin{pmatrix} \gamma_x & 0 \\ \gamma_v & I \end{pmatrix} n^{1/2}(\hat{\beta}_c - \beta_c) + n^{1/2} \left\{ \begin{pmatrix} \hat{\gamma}_x & 0 \\ \hat{\gamma}_v & I \end{pmatrix} - \begin{pmatrix} \gamma_x & 0 \\ \gamma_v & I \end{pmatrix} \right\} \hat{\beta}_c \\
=& \mathcal{D}^T n^{1/2}(\hat{\beta}_c - \beta_c) + n^{1/2}\hat{\beta}_{c,x}(\hat{\gamma}_z - \gamma_z) + o_p(1) \\
=& \mathcal{D}^T n^{1/2}(\hat{\beta}_c - \beta_c) + n^{1/2}\beta_{c,x}(\hat{\gamma}_z - \gamma_z) + o_p(1) \\
=& \mathcal{D}^T (n+m)^{-1/2} \sum_{i=1}^{n+m} A_{val,i} \\
& + n^{1/2} m^{-1} \beta_{c,x} \{Var(Z_i)\}^{-1} \sum_{i \in \mathcal{V}} (Z_i - \mu_z)(W_i - \mu_w) + o_p(1) \\
=& (n+m)^{-1/2} \sum_{i=1}^{n+m} \Bigg\{ \mathcal{D}^T A_{val,i} \\
& + (1 - \xi_i) \frac{(1+\rho)^{1/2}}{\rho} \beta_{c,x} \{Var(Z_i)\}^{-1} (Z_i - \mu_z)(W_i - \mu_w) \Bigg\} + o_p(1) \\
\equiv& (n+m)^{-1/2} \sum_{i=1}^{n+m} B_{val,i} + o_p(1).
\end{aligned}
$$

Since $B_{val,1}, \cdots, B_{val,n+m}$ are i.i.d., by the Central Limit Theorem, we obtain that

$$
n^{1/2}(\hat{\beta}_{cc} - \beta_0) \xrightarrow{d} N(0, \mathcal{B}), \quad \text{as} \quad n \to \infty,
$$

where $\mathcal{B} = E(B_{val,i}^{\otimes 2})$.

Appendix A5: Proof of Theorem 3

(1) in Theorem 3 is followed by the proof of Lemma 1. Next we prove (2). Similar to the proof of Theorem 1, we write $U_c(\beta_c) = U_{c1}(\beta_c) + U_{c2}(\beta_c)$, where

$$
U_{c1}(\beta_c) = \sum_{i=1}^{n} \int_0^{\tau} \left[\hat{Z}_i - \frac{S_{re}^{(1)}(\hat{Z}; \beta, t)}{S_{re}^{(0)}(\hat{Z}; \beta, t)} \right] dN_i(t),
$$

and $U_{c2}(\beta_c) = (\hat{D}(\beta_{c,x}), 0^T)^T \sum_{i=1}^n N_i(\tau)$. $U_{c1}(\beta_c)$ can be written a sum of independent terms as in the proof of Theorem 1. Next, we consider $U_{c2}(\beta_c)$. Since $\hat{D}(\beta_{c,x})$ is $\sqrt{n}$-consistent, we have

$$\begin{aligned} n^{-1/2}U_{c2}(\beta_c) &= \begin{pmatrix} D(\beta_{c,x}) \\ 0 \end{pmatrix} n^{-1/2}\sum_{i=1}^n N_i(\tau) + n^{1/2}\begin{pmatrix} \hat{D}(\beta_{c,x}) - D(\beta_{c,x}) \\ 0 \end{pmatrix} n^{-1}\sum_{i=1}^n N_i(\tau) \\ &= \begin{pmatrix} D(\beta_{c,x}) \\ 0 \end{pmatrix} n^{-1/2}\sum_{i=1}^n N_i(\tau) + \begin{pmatrix} D_i(\beta_{c,x}) \\ 0 \end{pmatrix} n^{-1}\sum_{i=1}^n N_i(\tau) + o_p(1). \end{aligned}$$

Since

$$n^{1/2}(\hat{\beta}_c - \beta_c) = \left\{ -n^{-1}\frac{\partial U_c(\beta)}{\partial \beta^T}|_{\beta=\tilde{\beta}} \right\}^{-1} n^{-1/2}U_c(\beta_c),$$

where $\tilde{\beta}$ is on the line segment between $\hat{\beta}_c$ and β_c. By the Strong Law of Large Numbers, we obtain

$$-n^{-1}\frac{\partial U_c(\beta)}{\partial \beta^T}|_{\beta=\tilde{\beta}} \xrightarrow{a.s.} \mathcal{I}, \quad \text{as} \quad n \to \infty.$$

Thus, by the Central Limit Theorem and Slusky's Theorem, Theorem 3 is proved.

Now we provide an example of $\hat{D}(\beta_x)$ that is $\sqrt{n}$-consistent of $D(\beta_x)$. Let

$$\begin{aligned} \hat{\eta}_0(\beta_x) &= \left[\frac{\sum_{i=1}^n I(n_i > 1) n_i^{-1}(n_i - 1)^{-1} \sum_{r \neq s} \exp\left\{(W_{ir} - W_{is})^T \beta_x\right\}}{\sum_{i=1}^n I(n_i > 1)} \right]^{1/2}, \\ \text{and} \quad \hat{\eta}_1(\beta_x) &= \frac{\sum_{i=1}^n I(n_i > 1) n_i^{-1}(n_i - 1)^{-1} \sum_{r \neq s}(W_{ir} - W_{is}) \exp\left\{(W_{ir} - W_{is})^T \beta_x\right\}}{2\hat{\eta}_0(\beta_x) \sum_{i=1}^n I(n_i > 1)}. \end{aligned}$$

Then $\hat{D}_{HL}(\beta_x) = \hat{\eta}_1(\beta_x)/\hat{\eta}_0(\beta_x)$, Hu and Lin (2004) showed that

$$\begin{aligned}
\hat{\eta}_0(\beta_x) - \eta_0(\beta_x) =& \{2\sum_{i=1}^{n} I(n_i > 1)\eta_0(\beta_x)\}^{-1} \\
& \times \sum_{i=1}^{n}\Big[I(n_i > 1)n_i^{-1}(n_i - 1)^{-1}\sum_{r \neq s}\exp\{(W_{ir} - W_{is})\beta_x\} \\
& - I(n_i > 1)\eta_0^2(\beta_x)\Big] + o_p(n^{-1/2}); \\
\hat{\eta}_1(\beta_x) - \eta_1(\beta_x) =& \{2\sum_{i=1}^{n} I(n_i > 1)\eta_0(\beta_x)\}^{-1} \\
& \times \sum_{i=1}^{n}\Big[I(n_i > 1)n_i^{-1}(n_i - 1)^{-1}\sum_{r \neq s}(W_{ir} - W_{is})\exp\{(W_{ir} - W_{is})\beta_x\} \\
& - I(n_i > 1)n_i^{-1}(n_i - 1)^{-1}\sum_{r \neq s}\exp\{(W_{ir} - W_{is})\beta_x\}\eta_0^{-1}(\beta_x)\eta_1(\beta_x) \\
& - I(n_i > 1)\eta_0(\beta_x)\eta_1(\beta_x)\Big] + o_p(n^{-1/2}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sqrt{n}\{\hat{D}_{HL}(\beta_x) - D(\beta_x)\} =& \sqrt{n}\left\{\frac{\hat{\eta}_1(\beta_x)}{\hat{\eta}_0(\beta_x)} - \frac{\eta_1(\beta_x)}{\eta_0(\beta_x)}\right\} \\
=& \frac{\sqrt{n}\{\hat{\eta}_1(\beta_x) - \eta_1(\beta_x)\}}{\eta_0(\beta_x)} - \frac{\eta_1(\beta_x)}{\eta_0^2(\beta_x)}\sqrt{n}\{\hat{\eta}_0(\beta_x) - \eta_0(\beta_x)\} + o_p(1) \\
=& \frac{1}{\sqrt{n}}\{2\eta_0^2(\beta_x)\sum_{i=1}^{n} I(n_i > 1)\}^{-1} \\
& \times \sum_{i=1}^{n}\Bigg(I(n_i > 1)n_i^{-1}(n_i - 1)^{-1}\Big[\sum_{r \neq s}(W_{ir} - W_{is})\exp\{(W_{ir} - W_{is})\beta_x\} \\
& - 2D(\beta_x)\sum_{r \neq s}\exp\{(W_{ir} - W_{is})\beta_x\}\Big]\Bigg) + o_p(1) \\
\equiv & \frac{1}{\sqrt{n}}\sum_{i=1}^{n} D_i(\beta_x) + o_p(1).
\end{aligned}$$

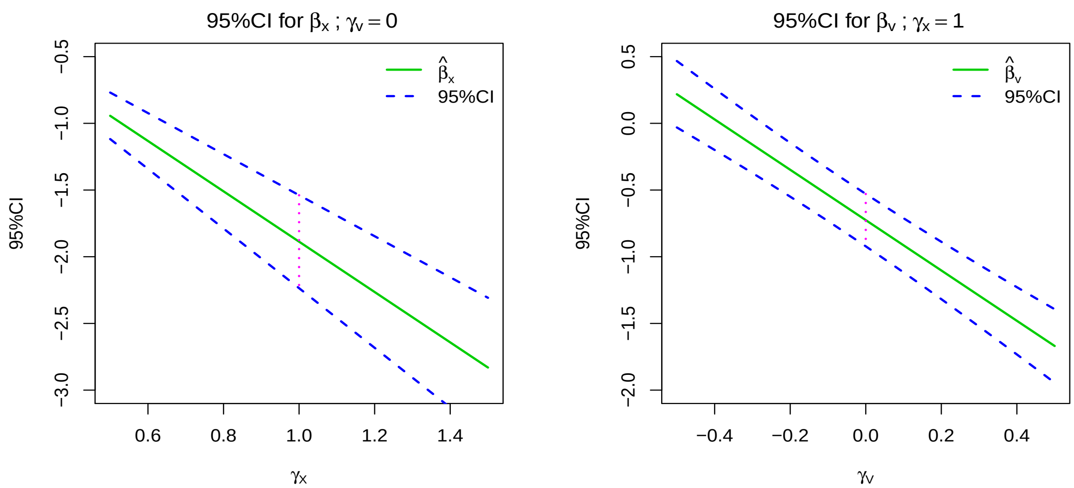


Figure 2.1: Point estimation of β and the corresponding confidence interval by the corrected profile likelihood method under the error model (2.11). Vertical lines show the confidence intervals by the estimator of Hu and Lin (2004) under the error model (2.13), which is a special case of (2.11).

Table 2.1: Simulation results under Scenario A

Case	γ_x	σ	Method	Estimation of β_x					Estimation of β_v				
				Bias[a]	EVE[b]	MVE[c]	MSE[d]	CP(%)[e]	Bias	EVE	MVE	MSE	CP(%)
Case 1	1.25	0.25	$\hat{\beta}_{nv}$	-0.250	0.009	0.008	0.072	23.0	-0.729	0.025	0.023	0.557	0.4
			$\hat{\beta}_c$	-0.188	0.012	0.010	0.048	51.7	-0.796	0.029	0.026	0.662	0.2
			$\hat{\beta}_{cc}$	0.015	0.019	0.016	0.020	94.2	0.016	0.014	0.014	0.014	94.5
			$\hat{\beta}$	0.005	0.015	0.014	0.015	94.6	0.015	0.013	0.014	0.014	94.7
		0.50	$\hat{\beta}_{nv}$	-0.377	0.008	0.007	0.150	1.5	-0.588	0.024	0.022	0.369	3.6
			$\hat{\beta}_c$	-0.158	0.024	0.019	0.049	67.4	-0.821	0.043	0.036	0.716	0.7
			$\hat{\beta}_{cc}$	0.052	0.038	0.029	0.040	94.5	0.021	0.017	0.017	0.018	93.7
			$\hat{\beta}$	0.005	0.015	0.014	0.015	94.6	0.015	0.013	0.014	0.014	94.7
	1.50	0.25	$\hat{\beta}_{nv}$	-0.362	0.007	0.006	0.138	1.1	-0.619	0.022	0.021	0.405	1.4
			$\hat{\beta}_c$	-0.326	0.008	0.007	0.114	4.8	-0.659	0.024	0.022	0.458	1.0
			$\hat{\beta}_{cc}$	0.012	0.018	0.015	0.018	94.1	0.015	0.014	0.014	0.014	94.3
			$\hat{\beta}$	0.005	0.015	0.014	0.015	94.6	0.015	0.013	0.014	0.014	94.7
		0.5	$\hat{\beta}_{nv}$	-0.442	0.006	0.005	0.202	0.2	-0.528	0.022	0.020	0.300	5.6
			$\hat{\beta}_c$	-0.310	0.013	0.010	0.109	17.8	-0.671	0.030	0.026	0.480	2.0
			$\hat{\beta}_{cc}$	0.035	0.028	0.023	0.030	94.2	0.019	0.016	0.016	0.016	94.1
			$\hat{\beta}$	0.005	0.015	0.014	0.015	94.6	0.015	0.013	0.014	0.014	94.7
Case 2	1.25	0.25	$\hat{\beta}_{nv}$	-0.230	0.007	0.006	0.059	21.3	-0.749	0.064	0.059	0.626	14.9
			$\hat{\beta}_c$	-0.178	0.008	0.008	0.040	45.0	-0.812	0.069	0.062	0.728	12.7
			$\hat{\beta}_{cc}$	0.027	0.013	0.012	0.014	94.7	0.009	0.056	0.051	0.056	93.2
			$\hat{\beta}$	0.019	0.011	0.010	0.011	95.3	0.008	0.052	0.049	0.052	94.1
		0.5	$\hat{\beta}_{nv}$	-0.343	0.006	0.006	0.123	1.6	-0.606	0.065	0.058	0.432	30.0
			$\hat{\beta}_c$	-0.156	0.015	0.013	0.039	62.2	-0.833	0.089	0.076	0.783	15.7
			$\hat{\beta}_{cc}$	0.055	0.023	0.020	0.026	94.4	0.011	0.067	0.058	0.068	93.5
			$\hat{\beta}$	0.019	0.011	0.010	0.011	95.3	0.008	0.052	0.049	0.052	94.1
	1.50	0.25	$\hat{\beta}_{nv}$	-0.347	0.005	0.005	0.125	0.3	-0.636	0.061	0.056	0.465	23.8
			$\hat{\beta}_c$	-0.317	0.006	0.005	0.106	3.1	-0.674	0.064	0.058	0.518	21.3
			$\hat{\beta}_{cc}$	0.025	0.013	0.012	0.013	94.9	0.009	0.055	0.050	0.055	93.6
			$\hat{\beta}$	0.019	0.011	0.010	0.011	95.3	0.008	0.052	0.049	0.052	94.1
		0.50	$\hat{\beta}_{nv}$	-0.417	0.004	0.004	0.178	0.1	-0.543	0.063	0.056	0.357	36.7
			$\hat{\beta}_c$	-0.305	0.008	0.007	0.101	10.7	-0.685	0.075	0.066	0.544	24.2
			$\hat{\beta}_{cc}$	0.042	0.018	0.016	0.020	94.5	0.010	0.062	0.055	0.062	93.1
			$\hat{\beta}$	0.019	0.011	0.010	0.011	95.3	0.008	0.052	0.049	0.052	94.1

[a] Bias: finite sample biases;
[b] EVE: empirical variances;
[c] MVE: average of the model-based variance estimates;
[d] MSE: mean square errors;
[e] MCP: model-based coverage probability.

Table 2.2: Simulation results under Scenario B

Case	γ_x	σ	Method	Estimation of β_x					Estimation of β_v				
				Bias	EVE	MVE	MSE	CP(%)	Bias	EVE	MVE	MSE	CP(%)
Case 1	1.25	0.25	$\hat{\beta}_{nv}$	-0.248	0.006	0.005	0.067	9.6	-0.741	0.016	0.015	0.565	0.0
			$\hat{\beta}_{c}$	-0.187	0.009	0.008	0.044	43.0	-0.806	0.020	0.018	0.669	0.0
			$\hat{\beta}_{cc}$	0.017	0.014	0.013	0.014	93.4	0.006	0.011	0.010	0.011	94.1
			$\hat{\beta}$	0.008	0.009	0.009	0.009	93.3	0.006	0.009	0.009	0.009	94.7
		0.50	$\hat{\beta}_{nv}$	-0.375	0.005	0.005	0.145	0.0	-0.600	0.016	0.015	0.376	0.6
			$\hat{\beta}_{c}$	-0.164	0.022	0.023	0.048	67.7	-0.825	0.036	0.035	0.716	0.3
			$\hat{\beta}_{cc}$	0.048	0.036	0.038	0.039	95.1	0.009	0.015	0.014	0.015	95.1
			$\hat{\beta}$	0.008	0.009	0.009	0.009	93.3	0.006	0.009	0.009	0.009	94.7
	1.5	0.25	$\hat{\beta}_{nv}$	-0.360	0.004	0.004	0.134	0.0	-0.630	0.014	0.014	0.412	0.1
			$\hat{\beta}_{c}$	-0.324	0.005	0.005	0.111	1.9	-0.669	0.016	0.015	0.464	0.1
			$\hat{\beta}_{cc}$	0.015	0.013	0.012	0.013	93.6	0.006	0.010	0.010	0.010	93.8
			$\hat{\beta}$	0.008	0.009	0.009	0.009	93.3	0.006	0.009	0.009	0.009	94.7
		0.50	$\hat{\beta}_{nv}$	-0.441	0.004	0.003	0.198	0.0	-0.540	0.014	0.013	0.306	0.9
			$\hat{\beta}_{c}$	-0.313	0.010	0.010	0.108	14.7	-0.678	0.023	0.021	0.483	0.2
			$\hat{\beta}_{cc}$	0.034	0.025	0.023	0.026	95.0	0.008	0.013	0.012	0.013	94.6
			$\hat{\beta}$	0.008	0.009	0.009	0.009	93.3	0.006	0.009	0.009	0.009	94.7
Case 2	1.25	0.25	$\hat{\beta}_{nv}$	-0.240	0.004	0.004	0.062	6.2	-0.732	0.040	0.039	0.575	5.9
			$\hat{\beta}_{c}$	-0.190	0.006	0.006	0.042	30.2	-0.794	0.043	0.042	0.673	3.5
			$\hat{\beta}_{cc}$	0.013	0.010	0.010	0.010	93.9	0.017	0.037	0.036	0.037	95.1
			$\hat{\beta}$	0.006	0.008	0.007	0.008	93.2	0.019	0.032	0.032	0.033	95.3
		0.50	$\hat{\beta}_{nv}$	-0.352	0.004	0.004	0.128	0.0	-0.591	0.040	0.039	0.390	16.5
			$\hat{\beta}_{c}$	-0.175	0.012	0.013	0.043	54.7	-0.811	0.057	0.056	0.714	5.6
			$\hat{\beta}_{cc}$	0.031	0.021	0.021	0.022	96.3	0.016	0.050	0.048	0.051	95.3
			$\hat{\beta}$	0.006	0.008	0.007	0.008	93.2	0.019	0.032	0.032	0.033	95.3
	1.50	0.25	$\hat{\beta}_{nv}$	-0.356	0.003	0.003	0.130	0.0	-0.619	0.038	0.037	0.422	12.1
			$\hat{\beta}_{c}$	-0.326	0.004	0.004	0.110	0.2	-0.657	0.039	0.039	0.472	9.9
			$\hat{\beta}_{cc}$	0.011	0.009	0.009	0.009	93.9	0.017	0.035	0.035	0.036	94.9
			$\hat{\beta}$	0.006	0.008	0.007	0.008	93.2	0.019	0.032	0.032	0.033	95.3
		0.50	$\hat{\beta}_{nv}$	-0.425	0.003	0.003	0.183	0.0	-0.529	0.039	0.037	0.319	23.1
			$\hat{\beta}_{c}$	-0.318	0.006	0.006	0.108	5.6	-0.667	0.047	0.046	0.492	13.5
			$\hat{\beta}_{cc}$	0.023	0.016	0.016	0.016	95.6	0.016	0.045	0.043	0.045	95.1
			$\hat{\beta}$	0.006	0.008	0.007	0.008	93.2	0.019	0.032	0.032	0.033	95.3

Table 2.3: Simulation results under Scenario C

Case	γ_x	σ	Method	Estimation of β_x					Estimation of β_v				
				Bias	EVE	MVE	MSE	CP(%)	Bias	EVE	MVE	MSE	CP(%)
Case 1	1.25	0.25	$\hat{\beta}_{nv}$	-0.223	0.009	0.008	0.058	32.4	-0.771	0.025	0.024	0.619	0.8
			$\hat{\beta}_c$	-0.191	0.010	0.010	0.046	49.7	-0.805	0.027	0.026	0.675	0.6
			$\hat{\beta}_{cc}$	0.012	0.016	0.015	0.016	94.8	0.004	0.015	0.014	0.015	94.7
			$\hat{\beta}$	0.008	0.014	0.014	0.014	94.4	0.003	0.015	0.014	0.015	94.2
		0.50	$\hat{\beta}_{nv}$	-0.298	0.008	0.008	0.097	9.6	-0.686	0.025	0.023	0.496	1.1
			$\hat{\beta}_c$	-0.177	0.015	0.015	0.047	60.0	-0.817	0.033	0.034	0.700	0.7
			$\hat{\beta}_{cc}$	0.029	0.024	0.023	0.025	95.3	0.007	0.017	0.017	0.017	95.0
			$\hat{\beta}$	0.008	0.014	0.014	0.014	94.4	0.003	0.015	0.014	0.015	94.2
	1.5	0.25	$\hat{\beta}_{nv}$	-0.345	0.006	0.006	0.125	1.6	-0.649	0.022	0.021	0.444	1.1
			$\hat{\beta}_c$	-0.326	0.007	0.007	0.113	3.8	-0.670	0.023	0.022	0.472	1.1
			$\hat{\beta}_{cc}$	0.010	0.015	0.015	0.015	94.7	0.004	0.015	0.014	0.015	94.7
			$\hat{\beta}$	0.008	0.014	0.014	0.014	94.4	0.003	0.015	0.014	0.015	94.2
		0.50	$\hat{\beta}_{nv}$	-0.391	0.006	0.005	0.159	0.2	-0.597	0.022	0.021	0.378	1.9
			$\hat{\beta}_c$	-0.319	0.009	0.008	0.111	9.7	-0.675	0.026	0.026	0.482	1.3
			$\hat{\beta}_{cc}$	0.021	0.020	0.019	0.020	94.8	0.006	0.016	0.016	0.016	95.1
			$\hat{\beta}$	0.008	0.014	0.014	0.014	94.4	0.003	0.015	0.014	0.015	94.2
Case 2	1.25	0.25	$\hat{\beta}_{nv}$	-0.217	0.007	0.006	0.054	24.8	-0.770	0.057	0.059	0.651	12.0
			$\hat{\beta}_c$	-0.192	0.008	0.007	0.044	38.0	-0.802	0.059	0.062	0.703	11.4
			$\hat{\beta}_{cc}$	0.011	0.012	0.012	0.012	94.9	0.006	0.049	0.051	0.049	94.8
			$\hat{\beta}$	0.008	0.011	0.010	0.011	94.2	0.008	0.047	0.049	0.047	95.5
		0.50	$\hat{\beta}_{nv}$	-0.280	0.006	0.006	0.085	7.4	-0.692	0.059	0.059	0.538	19.0
			$\hat{\beta}_c$	-0.181	0.011	0.010	0.044	50.3	-0.814	0.070	0.074	0.733	13.7
			$\hat{\beta}_{cc}$	0.023	0.017	0.016	0.017	95.0	0.004	0.055	0.059	0.055	94.3
			$\hat{\beta}$	0.008	0.011	0.010	0.011	94.2	0.008	0.047	0.049	0.047	95.5
	1.50	0.25	$\hat{\beta}_{nv}$	-0.342	0.005	0.005	0.122	0.6	-0.647	0.055	0.056	0.474	22.4
			$\hat{\beta}_c$	-0.327	0.005	0.005	0.112	1.2	-0.667	0.056	0.058	0.500	21.3
			$\hat{\beta}_{cc}$	0.010	0.012	0.011	0.012	94.7	0.006	0.049	0.050	0.049	94.9
			$\hat{\beta}$	0.008	0.011	0.010	0.011	94.2	0.008	0.047	0.049	0.047	95.5
		0.50	$\hat{\beta}_{nv}$	-0.380	0.005	0.004	0.149	0.3	-0.599	0.056	0.056	0.415	28.2
			$\hat{\beta}_c$	-0.322	0.006	0.006	0.110	4.1	-0.674	0.062	0.065	0.515	24.8
			$\hat{\beta}_{cc}$	0.017	0.015	0.014	0.015	95.2	0.005	0.053	0.055	0.053	94.5
			$\hat{\beta}$	0.008	0.011	0.010	0.011	94.2	0.008	0.047	0.049	0.047	95.5

Table 2.4: Simulation results under the Berkson error model

Case	σ	Method	Estimation of β_x					Estimation of β_v				
			Bias	EVE	MVE	MSE	CP(%)	Bias	EVE	MVE	MSE	CP(%)
Case 1′	0.25	$\hat{\beta}_{nv}$	-0.022	0.013	0.011	0.014	91.4	-0.020	0.013	0.011	0.013	92.4
		$\hat{\beta}_c$	0.093	0.020	0.016	0.029	88.5	0.021	0.015	0.013	0.016	92.7
		$\hat{\beta}_{cc}$	0.021	0.017	0.014	0.018	92.5	0.018	0.015	0.013	0.015	92.4
		$\hat{\beta}$	0.010	0.013	0.011	0.013	92.9	0.013	0.013	0.011	0.013	93.8
	0.50	$\hat{\beta}_{nv}$	-0.101	0.013	0.011	0.023	77.8	-0.100	0.013	0.011	0.023	79.0
		$\hat{\beta}_c$	0.522	0.109	0.286	0.382	81.0	0.079	0.178	0.077	0.184	94.0
		$\hat{\beta}_{cc}$	0.083	0.054	0.052	0.061	95.3	0.046	0.051	0.029	0.053	93.8
		$\hat{\beta}$	0.011	0.011	0.010	0.011	93.2	0.013	0.012	0.011	0.013	93.8
Case 2′	0.25	$\hat{\beta}_{nv}$	-0.012	0.011	0.010	0.011	92.7	-0.020	0.012	0.011	0.013	92.0
		$\hat{\beta}_c$	0.091	0.016	0.013	0.024	88.2	0.017	0.014	0.013	0.014	92.6
		$\hat{\beta}_{cc}$	0.017	0.013	0.011	0.014	92.6	0.015	0.014	0.012	0.014	92.7
		$\hat{\beta}$	0.015	0.010	0.009	0.011	93.0	0.012	0.012	0.011	0.012	93.4
	0.50	$\hat{\beta}_{nv}$	-0.083	0.011	0.009	0.018	83.2	-0.099	0.012	0.011	0.022	81.7
		$\hat{\beta}_c$	0.362	0.324	0.080	0.455	57.9	0.095	0.034	0.033	0.043	93.8
		$\hat{\beta}_{cc}$	0.008	0.029	0.020	0.029	90.7	0.033	0.022	0.019	0.023	93.8
		$\hat{\beta}$	0.014	0.010	0.009	0.010	92.8	0.012	0.012	0.011	0.012	93.1
Case 3′	0.25	$\hat{\beta}_{nv}$	-0.020	0.012	0.011	0.012	93.3	0.035	0.056	0.049	0.057	93.8
		$\hat{\beta}_c$	0.087	0.018	0.015	0.025	89.0	0.015	0.059	0.052	0.059	93.5
		$\hat{\beta}_{cc}$	0.015	0.015	0.013	0.015	93.9	0.016	0.059	0.052	0.059	93.5
		$\hat{\beta}$	0.014	0.011	0.010	0.011	94.3	0.013	0.055	0.049	0.055	93.8
	0.50	$\hat{\beta}_{nv}$	-0.106	0.012	0.010	0.023	78.0	0.094	0.057	0.050	0.066	91.6
		$\hat{\beta}_c$	0.322	0.547	0.103	0.650	72.2	0.005	0.095	0.079	0.095	92.4
		$\hat{\beta}_{cc}$	0.008	0.037	0.025	0.037	91.6	0.020	0.077	0.064	0.077	92.8
		$\hat{\beta}$	0.013	0.010	0.009	0.010	94.5	0.007	0.056	0.051	0.056	93.5

Chapter 3

Analysis of Survival Data with Covariate Error under Possibly Misspecified Error Models

3.1 Introduction

There are many well-known models for survival data analysis, including Cox proportional hazards models (Cox 1972), accelerated failure time models, and additive hazards models (Lin and Ying 1994). Although these models have been widely used for survival analysis, parameter estimation under these models are frequently challenged by mismeasurement of covariates. A well-known example is the CD4 lymphocyte counts in the AIDS studies, which are an important biomarker measured with considerable error (Hammer et al. 1996). Naively ignoring measurement error in covariates commonly leads to misleading results (Prentice 1982). Consequently, researchers proposed numerous methods to handle covariate measurement error, including the regression calibration approach (Prentice 1982; Pepe, Self and Prentice 1989; Wang et al. 1997), the likelihood based approaches (Hu, Tsiatis and Davidian 1998; Zucker 2005; Yi and Lawless 2007), and the score based approaches

(Nakamura 1992; Huang and Wang 2000; Hu and Lin 2004; Song and Huang 2005).

These methods are successful in correcting for measurement error effects; they focus on estimation of the parameters of survival models. In contrast to the large volume of estimation methods in the literature, hypothesis testing is rarely studied for survival models in the presence of covariate measurement error. In addition, when covariates are error-contaminated, usual model checking techniques (e.g., Therneau and Grambsch 2000; Lawless 2003) become invalid. It is desirable to develop valid testing procedures for a given survival model with error-prone covariates. Furthermore, it is interesting to understand the impact of model misspecification on inference of survival data with measurement error.

In this chapter, we explore these important problems. We first propose corrected score and Wald tests under Cox models with mismeasured covariates and study their validity and efficiency properties. Then we investigate the impact of model misspecification on parameter estimation and testing.

3.2 Notation and Model Setup

For subject i, $i = 1, \cdots, n$, let T_i be the failure time, C_i be the right censoring time, and $Z_i = (X_i^T, V_i^T)^T$ be a vector of r-dimensional time independent covariates. Here, the V_i are always observed, while the X_i are subject to measurement error. We assume $\{T_i, C_i, Z_i\}$ to be mutually independent, $i = 1, \cdots, n$. We assume that all subjects are under observation over a common time interval $[0, \tau]$, where τ is a positive constant. Define $S_i = \min(T_i, C_i)$, and $\delta_i = I(T_i \leq C_i)$. For $t \in (0, \tau]$, let $N_i(t) = I(S_i \leq t, \delta_i = 1)$ be a counting process, and $Y_i(t) = I(S_i \geq t)$ be an at-risk indicator. Throughout this article, we assume the conditional independent censoring mechanism, i.e., C_i and T_i are independent given Z_i.

3.2.1 Cox Model

The Cox model (Cox 1972) assumes that the failure time T_i is related to Z_i through the hazard function

$$\lambda(t; Z_i) = \lambda_0(t) \exp(Z_i^T \beta),$$

where $\lambda_0(\cdot)$ is the baseline hazard function, and β is the vector of regression parameters. Here we assume that the distribution of T_i is continuous.

Let $\beta_0 = (\beta_{0,x}^T, \beta_{0,v}^T)^T$ be the true value of the parameter, where $\beta_{0,x}$ and $\beta_{0,v}$ are the parameters corresponding to X_i and V_i, respectively. Inference about the regression parameter β, named $\hat{\beta}$, can be obtained by solving the partial score function (Cox 1975) $U(\beta) = 0$, where

$$U(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \left\{ Z_i - \frac{\sum_{j=1}^{n} Y_j(t) Z_j \exp(Z_j^T \beta)}{\sum_{j=1}^{n} Y_j(t) \exp(Z_j^T \beta)} \right\} dN_i(t). \tag{3.1}$$

3.2.2 Measurement Error in Survival Data

For $i = 1, \cdots, n$, let W_i denote the surrogate measurement of X_i. Let $\hat{Z}_i = (W_i^T, V_i^T)^T$ denote a measured version of Z_i. One may be tempted to ignore measurement error in X_i by using a partial score function based on the observed data:

$$U_{nv}(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \left\{ \hat{Z}_i - \frac{\sum_{j=1}^{n} Y_j(t) \hat{Z}_j \exp(\hat{Z}_j^T \beta)}{\sum_{j=1}^{n} Y_j(t) \exp(\hat{Z}_j^T \beta)} \right\} dN_i(t)$$

to proceed with the estimation of β. Let $\hat{\beta}_{nv}$ denote the root of $U_{nv}(\beta)$. It is shown that this estimator is often an inconsistent estimator of β (Struthers and Kalbfleisch 1986; Li and Ryan 2004).

To incorporate measurement error effects in inferential procedures, it is often necessary to model the measurement error process. We consider that the true covariates X_i and surrogate measurements W_i are featured by the classical additive measurement error model

(Carroll et al. 2006)

$$W_i = X_i + \epsilon_i, \tag{3.2}$$

where the ϵ_i are independent and identically distributed with mean 0 and a positive-definite variance matrix Σ_0. We assume that ϵ_i are independent of X_i, T_i, and C_i, and thus measurement error is nondifferential. Often, ϵ_i is assumed normally distributed (Carroll et al. 2006). In practise, the parameters of the error distribution are usually estimated through a validation subsample or replicated surrogate measurements (Yi and Lawless 2007).

We rewrite the classical measurement error model (3.2) as

$$\hat{Z}_i = Z_i + \tilde{\epsilon}_i,$$

where $\tilde{\epsilon}_i = (\epsilon_i^T, 0^T)^T$. Define

$$\begin{aligned} s^{(k)}(Z;\beta,t) =& E\{Y_i(t) Z_i^{\otimes k} \exp(Z_i^T \beta)\}, \\ \text{and} \quad s^{(k)}(\hat{Z};\beta,t) =& E\{Y_i(t) \hat{Z}_i^{\otimes k} \exp(\hat{Z}_i^T \beta)\}, k = 0,1,2. \end{aligned}$$

Let $D(\beta_x) = E\{\epsilon_i \exp(\epsilon_i^T \beta_x)\}/E\{\exp(\epsilon_i^T \beta_x)\}$.

Now, we describe the corrected score methods (Nakamura 1992; Hu and Lin 2002, 2004; Song and Huang 2005) which correct for measurement error effects. Define $U_0(\beta) = \sum_{i=1}^n \int_0^\tau \left(D(\beta_x)^T, 0^T\right)^T dN_i(t)$, and

$$U_c(\beta) = U_{nv}(\beta) + U_0(\beta). \tag{3.3}$$

Nakamura (1992), Hu and Lin (2002, 2004), and Song and Huang (2005) provided different ways to consistently estimate $D(\beta_x)$. Without loss of generality, we assume that $D(\beta_x)$ is known in this chapter to simplify our discussion. Solving $U_c(\beta) = 0$ gives an estimator of β; let $\hat{\beta}_c$ denote such an estimator. Kong and Gu (1999) showed the consistency and asymptotic normality of $\hat{\beta}_c$, provided certain regularity conditions hold.

3.3 Hypothesis Testing under Correctly Specified Measurement Error Model

In contrast to numerous estimation procedures proposed in the literature to correct for error effect, hypothesis testing procedures are rarely studied for survival models with mismeasured covariates. In this section, we propose a corrected score test and a corrected Wald test under the Cox model with covariate error, which are valid when the classical error model is correctly specified.

Write $\beta = (\beta^{+T}, \beta^{-T})^T$, where β^+ is a r^+ dimensional subvector of interest to be tested, and β^- is the r^- dimensional sub-vector. For simplicity of exposition, we let β^+ be the subvector that consists of the first r^+ elements of β, and β^- consists of the last r^- elements of β. We are interested in testing the null hypothesis:

$$H_0 : \beta^+ = \beta_0^+,$$

where β_0^+ is a given value.

Let Z_i^+ and Z_i^- be the subvectors of Z_i that correspond to β^+ and β^-, respectively. Let $\hat{D}^+$ and $\hat{D}^-$ be the sub-vectors of $\hat{D}$ corresponding to Z_i^+ and Z_i^-, and U_c^+ and U_c^- be the sub-vectors of U_c in (3.3) corresponding to Z_i^+ and Z_i^-, respectively. That is,

$$U_c^+(\beta) = \sum_{i=1}^n \int_0^\tau \left\{ \hat{Z}_i^+ - \frac{\sum_{j=1}^n Y_j(t) \hat{Z}_j^+ \exp(\hat{Z}_j^T \beta)}{\sum_{j=1}^n Y_j(t) \exp(\hat{Z}_j^T \beta)} + D^+(\beta_x) \right\} dN_i(t),$$

$$\text{and} \quad U_c^-(\beta) = \sum_{i=1}^n \int_0^\tau \left\{ \hat{Z}_i^- - \frac{\sum_{j=1}^n Y_j(t) \hat{Z}_j^- \exp(\hat{Z}_j^T \beta)}{\sum_{j=1}^n Y_j(t) \exp(\hat{Z}_j^T \beta)} + D^-(\beta_x) \right\} dN_i(t).$$

In addition, for any β in the parameter space, let

$$\mathcal{I}(\beta) = \int_0^\tau \left[\frac{s^{(2)}(\hat{Z};\beta,t)}{s^{(0)}(\hat{Z};\beta,t)} - \left\{ \frac{s^{(1)}(\hat{Z};\beta,t)}{s^{(0)}(\hat{Z};\beta,t)} \right\}^{\otimes 2} + \begin{pmatrix} \frac{\partial D(\beta_x)}{\partial \beta_x^T}|_{\beta_x=\beta_x} & 0 \\ 0 & 0 \end{pmatrix} \right] dE\{N_i(t)\},$$

$$J_i(\beta) = \int_0^\tau \left\{ \hat{Z}_i - \frac{s^{(1)}(\hat{Z};\beta,t)}{s^{(0)}(\hat{Z};\beta,t)} + D(\beta_x) \right\} dN_i(t)$$

$$- \int_0^\tau \frac{Y_i(t)\exp(\hat{Z}_i^T\beta)}{s^{(0)}(\hat{Z};\beta,t)} \left\{ \hat{Z}_i - \frac{s^{(1)}(\hat{Z};\beta,t)}{s^{(0)}(\hat{Z};\beta,t)} \right\} dE\{N_i(t)\},$$

and $\mathcal{J}(\beta) = E[\{J_i(\beta)\}^{\otimes 2}]$.

Write $\mathcal{I}(\beta)$ as

$$\mathcal{I}(\beta) = \begin{pmatrix} \mathcal{I}_{11}(\beta) & \mathcal{I}_{12}(\beta) \\ \mathcal{I}_{21}(\beta) & \mathcal{I}_{22}(\beta) \end{pmatrix},$$

where $\mathcal{I}_{11}(\beta)$ and $\mathcal{I}_{12}(\beta)$ are $r^+ \times r^+$ and $r^+ \times r^-$ submatrices of $\mathcal{I}(\beta)$.

We propose a corrected score test statistic as

$$T_{S,c} = n^{-1} U_c^{+T}(\beta_0^+, \tilde{\beta}_c^-) \hat{V}^{-1}(\beta_0^+, \tilde{\beta}_c^-) U_c^+(\beta_0^+, \tilde{\beta}_c^-),$$

where $\tilde{\beta}_c^-$ is the root of $U_c^-(\beta_0^+, \beta^-) = 0$, $\hat{V}(\beta)$ is a sample version of

$$\mathcal{V}(\beta) = \begin{pmatrix} I_{r^+}, & -\mathcal{I}_{12}(\beta)\mathcal{I}_{22}^{-1}(\beta) \end{pmatrix} \mathcal{J}(\beta) \begin{pmatrix} I_{r^+} \\ -\mathcal{I}_{22}^{-1}(\beta)\mathcal{I}_{21}(\beta) \end{pmatrix},$$

and I_{r^+} is a $r^+ \times r^+$ dimensional identity matrix.

Let $\mathcal{W}_{11}(\beta)$ be the upper left $r^+ \times r^+$ submatrix of $\mathcal{I}^{-1T}(\beta)\mathcal{J}(\beta)\mathcal{I}^{-1}(\beta)$, and $\hat{W}_{11}(\beta)$ be a sample version of $\mathcal{W}_{11}(\beta)$. Let $\hat{\beta}_c = (\hat{\beta}_c^{+T}, \hat{\beta}_c^{-T})^T$ be the solution of $U_c(\beta) = 0$. We propose the corrected Wald test statistic

$$T_{W,c} = n(\hat{\beta}_c^+ - \beta_0^+)^T \hat{W}_{11}^{-1}(\hat{\beta}_c)(\hat{\beta}_c^+ - \beta_0^+).$$

The following theorem establishes the asymptotic properties of the test statistics $T_{S,c}$ and $T_{W,c}$ under the null hypothesis H_0. The proof is deferred to Appendix A1.

Theorem 1 *Under mild regularity conditions, and under the null hypothesis H_0, we have that as $n \to \infty$,*

$$\begin{aligned} T_{S,c} &\xrightarrow{d} \chi^2_{r^+}, \\ \textit{and } T_{W,c} &\xrightarrow{d} \chi^2_{r^+}. \end{aligned}$$

Theorem 1 shows that the proposed test statistics $T_{S,c}$ and $T_{W,c}$ correctly adjust for the error effects, provide the basis of conducting inference of the regression parameters in the Cox model. Moreover, $T_{S,c}$ and $T_{W,c}$ have the same asymptotic distribution.

Let $P_0(T_{S,c} > \chi^2_{r^+,\alpha})$ and $P_0(T_{W,c} > \chi^2_{r^+,\alpha})$ be the power functions of $T_{S,c}$ and $T_{W,c}$, respectively, where $\alpha \in (0,1)$ is a constant, $\chi^2_{r^+,\alpha}$ is the upper α-quantile of the $\chi^2_{r^+}$ distribution, and P_0 is the probability measure under the null hypothesis. Theorem 1 implies that $\lim_{n\to\infty} P_0(T_{S,c} > \chi^2_{r^+,\alpha}) = \alpha$, suggesting a testing procedure based on $T_{S,c}$ for testing H_0: for a pre-specified size α, we reject the null hypothesis H_0 if $T_{S,c} > \chi^2_{r^+,\alpha}$. A testing procedure based on $T_{W,c}$ is similarly defined.

Next, we study the power of the proposed test statistics $T_{S,c}$ and $T_{W,c}$ for a sequence of local alternative hypotheses. Specifically, we consider a sequence of root n local alternatives

$$H_n : \beta_n^+ = \beta_0^+ + \frac{b}{\sqrt{n}}, \quad n = 1, 2, \cdots,$$

where b is a vector of non-zero constants. Let

$$\delta = b^T \left\{\mathcal{I}_{11}(\beta_0) - \mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0)\right\} \mathcal{V}^{-1}(\beta_0) \left\{\mathcal{I}_{11}(\beta_0) - \mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0)\right\} b.$$

The following theorem establishes the asymptotic properties of the test statistic $T_{S,c}$ and $T_{W,c}$ under the local alternative hypotheses H_n. The proof is deferred to Appendix A2.

Theorem 2 *Under mild regularity conditions, and under the sequence of local alternative hypotheses H_n, we have that as $n \to \infty$,*

$$\begin{aligned} T_{S,c} &\xrightarrow{d} \chi^2_{r^+}(\delta), \\ \textit{and } T_{W,c} &\xrightarrow{d} \chi^2_{r^+}(\delta), \end{aligned}$$

where $\chi^2_{r^+}(\delta)$ *is the noncentral Chi-squared distribution with the non-centrality parameter* δ*.*

Theorem 2 suggests that $T_{S,c}$ and $T_{W,c}$ have the same asymptotic behaviour under the alternative hypotheses. Therefore, in the following discussion, we focus on $T_{S,c}$ only. Let T_S and δ_0 be respectively defined as $T_{S,c}$ and δ with $\hat{Z}_i$ replaced by Z_i and $D(\beta_x)$ ignored. That is,

$$\begin{aligned} T_S =& n^{-1}U^{+T}(\beta_0^+, \tilde{\beta}^-)\hat{V}_0^{-1}(\beta_0^+, \tilde{\beta}^-)U^+(\beta_0^+, \tilde{\beta}_c^-), \\ \text{and} \quad \delta_0 =& b^T \left\{\mathcal{I}_{0,11}(\beta_0) - \mathcal{I}_{0,12}(\beta_0)\mathcal{I}_{0,22}^{-1}(\beta_0)\mathcal{I}_{0,21}(\beta_0)\right\} \mathcal{V}_0^{-1}(\beta_0) \\ & \times \left\{\mathcal{I}_{0,11}(\beta_0) - \mathcal{I}_{0,12}(\beta_0)\mathcal{I}_{0,22}^{-1}(\beta_0)\mathcal{I}_{0,21}(\beta_0)\right\} b, \end{aligned}$$

where

$$\begin{aligned} \mathcal{I}_0(\beta) =& \int_0^\tau \left[\frac{s^{(2)}(\hat{Z};\beta,t)}{s^{(0)}(\hat{Z};\beta,t)} - \left\{ \frac{s^{(1)}(\hat{Z};\beta,t)}{s^{(0)}(\hat{Z};\beta,t)} \right\}^{\otimes 2} \right] dE\{N_i(t)\} \\ \equiv& \begin{pmatrix} \mathcal{I}_{0,11}(\beta) & \mathcal{I}_{0,12}(\beta) \\ \mathcal{I}_{0,21}(\beta) & \mathcal{I}_{0,22}(\beta) \end{pmatrix}, \\ \mathcal{V}_0(\beta) =& \begin{pmatrix} I_{r^+}, & -\mathcal{I}_{0,12}(\beta)\mathcal{I}_{0,22}^{-1}(\beta) \end{pmatrix} \mathcal{I}_0(\beta) \begin{pmatrix} I_{r^+} \\ -\mathcal{I}_{0,22}^{-1}(\beta)\mathcal{I}_{0,21}(\beta) \end{pmatrix}, \end{aligned}$$

U^+ and U^- are the sub-vectors of U_c corresponding to Z_i^+ and Z_i^-, respectively, $\tilde{\beta}^-$ is the root of $U^-(\beta_0^+, \beta^-) = 0$, $\hat{V}_0(\beta)$ is a sample version of $\mathcal{V}_0(\beta)$, and $\mathcal{I}_{0,11}(\beta)$ and $\mathcal{I}_{0,12}(\beta)$ are $r^+ \times r^+$ and $r^+ \times r^-$ submatrices of $\mathcal{I}_0(\beta)$. Then following Theorem 2, we obtain that under H_n,

$$T_S \xrightarrow{d} \chi^2_{r^+}(\delta_0), \quad \text{as } n \to \infty.$$

Based on Theorem 2, we can compare the asymptotic relative efficiency (ARE) (Lehmann and Romano 2005) of the proposed test statistic $T_{S,c}$ relative to the true score test statistic T_S. For simplicity, we only consider the case where Z_i^+ is univariate. The proof is deferred to Appendix A3.

Corollary 1 *Under mild regularity conditions, and under the sequence of local alternative hypotheses H_n, we have that as $n \to \infty$, the ARE of $T_{S,c}$ relative to T_S is*

$$ARE(T_{S,c}; T_S) = \frac{\delta}{\delta_0} < 1.$$

Corollary 1 implies that the proposed test statistic $T_{S,c}$ incurs efficiency loss. This is due to that the underlying covariate process is not fully observed. To illustrate the degree of efficiency loss, we examine an example. In clinical trials, testing treatment effect is often of primary interest. Let V_i be 1 if the subject i is assigned to the treatment group, and 0 otherwise. In randomized trials, it is reasonable to assume that X_i is independent of V_i. In addition, we assume that C_i is independent of V_i. We show in Appendix A4 that Corollary 1 yields that

$$ARE(T_{S,c}; T_S) = \left(1 + \frac{Var\{\exp(\epsilon_i \beta_{0,x})\}}{[E\{\exp(\epsilon_i \beta_{0,x})\}]^2} \frac{E\left\{\int_0^\tau Y_i(t) \exp\left(X_i^T \beta_{0,x}\right) \lambda_0(t) dt\right\}^2}{E\left\{\int_0^\tau Y_i(t) \exp\left(X_i^T \beta_{0,x}\right) \lambda_0(t) dt\right\}}\right)^{-1}. \quad (3.4)$$

Furthermore, it is shown in Appendix A4 that

$$\left(1 + 2\frac{Var\{\exp(\epsilon_i \beta_{0,x})\}}{[E\{\exp(\epsilon_i \beta_{0,x})\}]^2}\right)^{-1} \leq ARE(T_{S,c}; T_S) < 1. \quad (3.5)$$

It is interesting to note that this lower bound of the ARE depends only on the measurement error magnitude and the covariate effect, but not on the baseline hazard and censoring information. As the ARE is smaller than 1, thus the corrected score test loses efficiency compared to the true score test. It can be shown that when the error is normally distributed with variance Σ_0, the efficiency loss of the corrected score test usually increases as the magnitude of measurement error Σ_0 increases. However, the corrected score test is typically more efficient than the log-rank test (e.g., Kong and Slud 1997). For illustration, we conduct a simple numerical study.

Suppose the failure times are generated from the Cox model $\lambda(t; X_i, V_i) = \lambda_0(t) \exp(X_i \beta_x + V_i \beta_v)$, where $\lambda_0(t) = t$, and the true parameter values are $(\beta_{0,x}, \beta_{0,v})^T = (1, 0)^T$. We are interested in testing the null hypothesis $H_0 : \beta_v = 0$. The univariate covariate X_i is standard normal, and is independent of the treatment indicator V_i. V_i follows the Bernoulli distribution Bernourlli(0.5). We observe W_i instead of X_i, where the error model $W_i = X_i + \epsilon_i$ is correctly specified, where ϵ_i are generated from $N(0, \sigma_0^2)$. We let σ_0 vary from 0 to 0.5 to represent different magnitudes of measurement error. The censoring times are generated from Unif$(0, a)$, and we choose $a = 5.4$ and ∞ to represent approximate 25% censoring rate and no censoring cases, respectively. We plot the ARE of the corrected score test relative to the true score test and its lower bound in Figure 3.1. For comparison, we also plot the ARE of the log rank test relative to the true score test.

The figure shows that the corrected score test is substantially more efficient than the log-rank test, even when measurement error is moderate or large. The ARE of the corrected score test is above the lower bound in (3.5) we derived, and this lower bound is rather sharp, especially for the case without censoring.

[Insert Figure 3.1 here!]

3.4 Inference under Misspecified Measurement Error Model

The consistency of the corrected score estimator $\hat{\beta}_c$ and our proposed testing statistics $T_{S,c}$ and $T_{W,c}$ rely on an important assumption that the measurement error model (3.2) correctly postulates the measurement error process. However, no model checking procedure is available for checking (3.2). It is thereby interesting to examine the impact of misspecification of the error model on point estimation and hypothesis testing about the regression parameter β.

For ease of exposition, we consider in this section that $\hat{\beta}_c$ is obtained by the method of Nakamura (1992), which is the solution of $U_c(\beta)$ in (3.3) with normally distributed measurement errors.

3.4.1 Estimation

Define $\tilde{U}_c(\beta) = \tilde{U}_{nv}(\beta) + \tilde{U}_0(\beta)$, where

$$\begin{aligned}
\tilde{U}_{nv}(\beta) = & \int_0^\tau E\left\{Y_i(t)\hat{Z}_i \exp\left(Z_i^T\beta_0\right)\right\}\lambda_0(t)dt \\
& - \int_0^\tau \frac{E\left\{Y_i(t)\hat{Z}_i \exp\left(\hat{Z}_i^T\beta\right)\right\}}{E\left\{Y_i(t)\exp\left(\hat{Z}_i^T\beta\right)\right\}} E\left\{Y_i(t)\lambda_0(t)\exp\left(Z_i^T\beta_0\right)\right\}dt, \\
\tilde{U}_0(\beta) = & \begin{pmatrix} \Sigma_0\beta_x \\ 0 \end{pmatrix} \int_0^\tau E\left\{Y_i(t)\exp\left(Z_i^T\beta_0\right)\right\}\lambda_0(t)dt,
\end{aligned}$$

and the expectation is taken under the true survival and error models.

Assume that $\tilde{U}_c(\beta) = 0$ has a unique solution. Let β_c be the solution of $\tilde{U}_c(\beta) = 0$. Let $\mathcal{I}(\beta)$ and $\mathcal{J}(\beta)$ be the defined as in Section 3.3, where the expectations are taken under the true models.

Suppose that the measurement error model (3.2) is misspecified, and the true error model that links W_i and X_i is unknown. The following theorem characterizes the asymptotic property of $\hat{\beta}_c$ under the misspecified error model (3.2). The proof is deferred to Appendix A5.

Theorem 3 *Under the regularity conditions, we have the following results:*

(1) $\hat{\beta}_c \xrightarrow{p} \beta_c$ *as* $n \to \infty$;

(2) $n^{1/2}(\hat{\beta}_c - \beta_c) \xrightarrow{d} N(0, \mathcal{I}^{-1T}(\beta_c)\mathcal{J}(\beta_c)\mathcal{I}^{-1}(\beta_c))$, *as* $n \to \infty$.

The implications of this theorem are two-fold. First, if the measurement error model is correctly specified, then β_0 is a solution of $\tilde{U}_c(\beta) = 0$; under the assumption of the unique root of $\tilde{U}_c(\beta)$, this implies the consistency of $\hat{\beta}_c$. Secondly, if misspecification of the error model occurs, this theorem offers us a tool to quantify the asymptotic bias incurred in the working estimator $\hat{\beta}_c$ under the misspecified measurement error model. Specifically, the difference between β_c and β_0 features consistency or inconsistency of the estimator $\hat{\beta}_c$. To find the relationship between β_c and β_0, it suffices to solve $\tilde{U}_c(\beta_c) = 0$, or equivalently,

$$\int_0^\tau E\left\{Y_i(t)\hat{Z}_i \exp\left(Z_i^T\beta_0\right)\right\}\lambda_0(t)dt - \int_0^\tau \frac{E\left\{Y_i(t)\hat{Z}_i \exp\left(\hat{Z}_i^T\beta_c\right)\right\}}{E\left\{Y_i(t)\exp\left(\hat{Z}_i^T\beta_c\right)\right\}} E\left\{Y_i(t)\lambda_0(t)\exp\left(Z_i^T\beta_0\right)\right\}dt$$

$$+\begin{pmatrix} \Sigma_0\beta_{c,x} \\ 0 \end{pmatrix}\int_0^\tau E\left\{Y_i(t)\exp\left(Z_i^T\beta_0\right)\right\}\lambda_0(t)dt = 0, \tag{3.6}$$

where the expectation is taken under the true survival and error models.

Equation (3.6) is usually complicated to evaluate, and there is generally no explicit relationship between β_0 and β_c. However, under certain special but useful scenarios of misspecification of the measurement error model, it is possible to gain interesting insights into the impact of misspecification of error models. Now we discuss several important situations. To highlight the key idea without introducing complex exposition, we consider the case that X_i is a univariate covariate. In the following, we examine only one scenario explicitly. Other scenarios and the proofs are included in Appendix S1 of the Supplementary Material.

Suppose we misspecify the measurement error model to be (3.2), the true error model is actually given by

$$W_i = \gamma_0 + \gamma_x X_i + \gamma_v^T V_i + \epsilon_i, \tag{3.7}$$

where γ_x and γ_v are regression coefficients, and ϵ_i is the error term with mean zero and is independent of other variables, and has the same distribution of the error term in the classical error model (3.2). Note that X_i and V_i can be correlated, and the distributional form of the error term ϵ_i is unknown.

This misspecification scenario includes several interesting cases. For example, it features the case that some covariates are omitted in the working error model, and the surrogate W_i does not have the same mean as X_i.

Using equation (3.6), we obtain the following relationship of β_c and β_0:

$$\beta_{c,x} = \gamma_x^{-1}\beta_{0,x}, \tag{3.8}$$

$$\text{and} \quad \beta_{c,v} = \beta_{0,v} - \gamma_v\gamma_x^{-1}\beta_{0,x}. \tag{3.9}$$

This analytic relationship indicates that $\hat{\beta}_c$ can be substantially biased when the true error model (3.7) is misspecified as the classical error model (3.2). The point estimates of $\beta_{c,x}$ and $\beta_{c,v}$ under the misspecified error model (3.2) can be either attenuated or inflated, depending on the coefficients γ_x and γ_v in the true error model (3.7).

3.4.2 Hypothesis Testing

In this subsection, we study the impact of misspecified error models on the proposed test statistics. As in Section 3.3, we are interested in testing the null hypothesis:

$$H_0 : \beta^+ = \beta_0^+.$$

Let the expectations in $\mathcal{I}(\beta)$ and $\mathcal{J}(\beta)$ in Section 3.3 be taken under the Cox model and the underlying unknown measurement error model. Let

$$T^*_{S,c} = n^{-1}U_c^{+T}(\beta_c^+, \tilde{\beta}_c^{*-})\hat{V}^{-1}(\beta_0^+, \tilde{\beta}_c^{*-})U_c^+(\beta_0^+, \tilde{\beta}_c^{*-}),$$

where $\tilde{\beta}_c^{*-}$ is the root of $U_c^-(\beta_c^+, \beta^-) = 0$. Let

$$T^*_{W,c} = n(\hat{\beta}_c^+ - \beta_c^+)^T\hat{W}_{11}^{-1}(\hat{\beta}_c)(\hat{\beta}_c^+ - \beta_c^+).$$

We show in the following theorem that the proposed corrected score and corrected Wald tests may become invalid. Similar to Theorem 1, we obtain the following result:

Theorem 4 *Under mild regularity conditions, and under the null hypotheses H_0, we have that as $n \to \infty$,*

$$T^*_{S,c} \xrightarrow{d} \chi^2_{r^+},$$

$$\textit{and} \quad T^*_{W,c} \xrightarrow{d} \chi^2_{r^+}.$$

As illustrated in Section 3.4, in many situations where the measurement error model is misspecified, $\beta_c^+ \neq \beta_0^+$. Therefore, it is often the case that $T_{S,c} \neq T_{S,c}^*$. It follows that the corrected score test based on $T_{S,c}$ is often invalid in the sense that $T_{S,c}$ is not Chi-Squared distributed asymptotically. Similarly, the corrected Wald test based on $T_{W,c}$ is often invalid.

However, in certain scenarios with misspecified measurement error models, the corrected score test and the corrected Wald test can still yield valid results. To see this, we consider the situation where $\beta_c^+ = \beta_0^+$. Under this case,

$$\hat{\beta}_c^+ \xrightarrow{p} \beta_0^+, \quad \text{as } n \to \infty. \tag{3.10}$$

Theorem 4 thus implies that $T_{S,c} \xrightarrow{d} \chi^2_{r^+}$, and $T_{W,c} \xrightarrow{d} \chi^2_{r^+}$, as $n \to \infty$, showing that the corrected score test based on $T_{S,c}$ and the corrected Wald test based on $T_{W,c}$ are asymptotically valid. When these tests are valid under misspecified error model, their efficiency property is similar to that described in Corollary 1.

In the following, we specify an important scenario where (3.10) is satisfied under a misspecified error model. The proof and other important scenarios are included in Appendix S2 of the Supplementary Material.

Suppose X_i is independent of V_i, and the censoring mechanism is noninformative. The underlying error model is unspecified. Let $\beta^+ = \beta_v$, and $\beta^- = \beta_x$. We are interested in testing the null hypothesis $H_0 : \beta^+ = 0$ (i.e., $\beta_v = 0$). Following Kong and Slud (1997), we assume that

$$\log P(C_i \geq t | X_i, V_i) = a(t, X_i) + b(t, V_i), \tag{3.11}$$

for some unknown positive deterministic functions $a(\cdot)$ and $b(\cdot)$. Note that the assumption (3.11) can be satisfied under various situations, including (i) C_i is independent of V_i; or (ii) C_i is independent of X_i; or (iii) Conditional on X_i, C_i is independent of V_i; or (iv) Conditional on V_i, C_i is independent of X_i; or (v) the randomized trial setting in Section 3.3.

Under H_0 and (3.11), we have $\beta_{c,v} = 0$, and thus (3.10) is satisfied. Therefore, the corrected score test based on $T_{S,c}$ and the corrected Wald test based on $T_{W,c}$ are valid in this scenario.

3.5 Numerical Studies

3.5.1 Parameter Estimation

In the subsection, we study the impact of misspecified error model on asymptotic bias of parameter estimation determined by (3.6). We consider the case that there are no V_i covariates, i.e., $Z_i = X_i$, and X_i is a univariate variable generated from $N(0, 1)$. Suppose the failure times T_i are generated from the Cox model $\lambda(t; X_i) = \lambda_0(t) \exp(X_i\beta_x)$, where $\lambda_0(t) = t$, and the true parameter value is $\beta_{0,x} = 1$. Suppose the censoring times C_i are simulated from Unif$(0, 5.4)$, leading to about 25% censoring rate. Suppose we incorrectly use the classical error model $W_i = X_i + \epsilon_i$ as a working model for featuring the measurement error process, where $\epsilon_i \sim N(0, \sigma_0^2)$ with given σ_0.

We use the relationship (3.6) to derive the limit β_c of the estimator $\hat{\beta}_c$. Similarly, we derive the limit β_{nv} of the naive estimator $\hat{\beta}_{nv}$. We are interested in the asymptotic bias of β_c relative to the true parameter value β_0, defined as $(\beta_c - \beta_0)/\beta_0$. As a comparison, we also plot the asymptotic relative bias of β_{nv} against β_0, defined as $(\beta_{nv} - \beta_0)/\beta_0$. In Figure 3.2, we consider Case 1 that the true error model is $W_i = X_i + K\epsilon_i$. In Figure 3.2, $K = 0.8, 0.9, 1.1, 1.2$, and σ varies from 0 to 0.5. In Figure 3.3, we consider Case 2 that the true model is $W_i = \gamma_0 + \gamma_x X_i + \epsilon_i$, where $\gamma_0 = 0$, $\gamma_x = 0.8, 0.9, 1.1, 1.2$, and σ varies from 0 to 0.5. In Appendix S3 in the Supplementary Material, we provide numerical studies for other misspecification scenarios.

Figures 3.2 and 3.3 reveal that the degree of asymptotic biases of the estimator $\hat{\beta}_c$ can be even worse than the naive estimator $\hat{\beta}_{nv}$ when measurement error model is misspecified. The asymptotic bias of $\hat{\beta}_c$ can be attenuated, inflated, or even constant when the degree of measurement error increases. This typically differs from the usual attenuation phenomenon we observed for the naive method in many settings.

[Insert Figures 3.2-3.3 here!]

Next, we study the finite sample biases of misspecifying the error model on parameter estimation of the Cox model. Additional simulation studies are summarized in Appendix S4 in the Supplementary Material.

Let the sample size $n = 200$ and generate 1000 simulations for each parameter configuration, and let $Z_i = (X_i, V_i)^T$ be a 2×1 vector of covariates. We consider two cases for covariates. In Case 1, the Z_i are bivariate normal, where both of X_i and V_i are standard normal and the correlation is 0.50. In the Case 2, the X_i are standard exponential, and V_i follow a Bernoulli distribution with

$$\Pr(V_i = 1|X_i) = \frac{\exp(X_i)}{1 + \exp(X_i)}.$$

The correlation of X_i and V_i is about 0.27. In both cases, we generate the surrogate W_i from error model $W_i = X_i + \epsilon_i$, where $\epsilon_i \sim N(0, \sigma^2)$ and is independent of $(X_i, V_i)^T$.

We consider three methods to estimate the parameter β . The first one is the naive method that ignores measurement error by solving the partial score function (3.1) with Z_i replaced by $\hat{Z}_i$; let $\hat{\beta}_{nv}$ denote the resulting naive estimator. The second one is the corrected score estimator $\hat{\beta}_c$. We also calculate $\hat{\beta}$ which solves (3.1) with the true covariate X_i treated available for comparison.

Let σ be 0.25 or 0.5 to represent different degrees of measurement error. In the following analysis we assume that σ is known. Suppose the failure times are generated from the Cox model $\lambda(t; X_i, V_i) = \lambda_0(t) \exp(X_i\beta_x + V_i\beta_v)$, where $\lambda_0(t) = t$, and the true parameter values are $(\beta_{0,x}, \beta_{0,v})^T = (1, 1)^T$. The censoring times C_i are generated from Unif$(0, c)$, where c

is set to be 5.4 and 2.05 for Cases 1 and 2, respectively, leading to about 30% censoring rates in both cases.

Table 3.1 presents the empirical results for the scenario that the true error model is $W_i = X_i + K\epsilon_i$, with K varying from $1/\sqrt{2}$ to $\sqrt{2}$, representing different degrees and directions of misspecification. We report the finite sample biases (Bias), the empirical variances (EVE), the average of the model-based variance estimates (MVE), the mean square errors (MSE), and the coverage rate of 95% confidence intervals.

We find in Table 3.1 that the finite sample biases of the estimator $\hat{\beta}_c$ can be even bigger than those of the naive estimator $\hat{\beta}_{nv}$. The direction of the biases of the estimator $\hat{\beta}_c$ resembles those of the limit β_c in Figure 3.2. Furthermore, the coverage rates of 95% confidence intervals produced by the estimator $\hat{\beta}_c$ can be quite poor.

[Insert Table 3.1 here!]

3.5.2 Hypothesis Testing

Set the sample size $n = 200$, and the number of simulation runs $m = 200$. The univariate covariate X_i are generated from standard normal distribution, and are independent of the treatment indicator V_i, where V_i are generated from Bernourlli(0.5). Suppose we correctly specify the Cox model $\lambda(t; X_i, V_i) = \lambda_0(t)\exp(X_i\beta_x + V_i\beta_v)$. The underlying baseline hazard function is $\lambda_0(t) = t$, the parameter value of $\beta_{0,x} = 1$, and the parameter value of $\beta_{0,v}$ varies from -1 to 1. The censoring time is uniformly distributed and the censoring rate is about 30%. We are interested in testing the null hypothesis $H_0 : \beta_v = 0$.

We observe the surrogate W_i instead of X_i, and suppose we specify the error model as $W_i = X_i + \epsilon_i$, where $\epsilon_i \sim N(0, \sigma_0^2)$ and $\sigma_0 = 0.25$ or 0.5 is assumed known. We consider two scenarios: In the first scenario, this error model is correctly specified; in the second scenario, this error model is misspecified, and the true error model is $W_i = X_i + K\epsilon_i$ with $K = 1.2$. In both scenarios, we report the empirical power of the corrected score test, the log rank test and the true score test.

In Figure 3.4, we study the first scenario that the error model is correctly specified. Measurement error tends to reduce the power of the corrected score test compared to the true score test. However, it is more powerful than the log rank test, which matches the observation in Section 3.3.

In Figure 3.5, we study the first scenario that the error model is misspecified. By the results in Section 3.4.2, the corrected score test is still valid in this scenario. Furthermore, the power of the corrected score test is still substantially higher than that of the log rank test in the presence of measurement error and model misspecification.

[Insert Figures 3.4-3.5 here!]

3.6 An example

We conduct data analysis of the AIDS Clinical Trials Group (ACTG) 175 (Hammer, et al. 1996) study. The ACTG 175 study is a double-blind randomized clinical trial that evaluated the effects of the HIV treatments for which three drugs were used in combination or alone: zidovudine, didanosine, and zalcitabine.

There were $n = 2139$ individuals in the study. The baseline measurements on CD4 were collected before randomization, ranging from 200 to 500 per cubic millimeter. Let V_i denote the treatment indicator for subject i, where $V_i = 1$ if a subject receive one of the three treatments, and 0 otherwise. T_i is defined to be the time to the occurrence of the first event among the following events: (i) more than 50% decline of CD4 counts compared to the averaged baseline CD4 counts; (ii) disease progression to AIDS; or (iii) death. About 75.6% of outcome values are censored. Let X_i be the normalization version of the true baseline CD4 counts: $\log(\text{CD4 counts} + 1)$. X_i were not observed in the study. The average baseline measurements W_i were observed instead.

We are interested in studying the relationship how T_i is associated with unobserved baseline CD4 counts X_i, and testing of treatment effect of the drugs on the event T_i is of particular interest. We employ the Cox model to feature the dependence of T_i on the covariates X_i and V_i:

$$\lambda(t) = \lambda_0(t) \exp(X_i \beta_x + V_i \beta_v),$$

where $\lambda_0(t)$ is the baseline hazard function, and $\beta = (\beta_x, \beta_v)^T$ is the regression parameter. Our interest in this example is $H_0 : \beta_v = 0$.

We employ the classical error model: $W_i = X_i + \epsilon_i$, where the error ϵ_i is assumed distributed as $N(0, \sigma_0^2)$. σ_0 can be consistently estimated by replicated measurements (Huang and Wang 2000).

We comment that in this example there is not enough knowledge what a reasonable error model would be. Unfortunately, there is lack of model checking techniques to check these models since X_i is unobserved. However, note that the correlation of W_i and V_i is weak (-0.025) due to randomization. Thus, the data structure is the same as the scenario described in Section 3.4.2. Therefore, the result in this scenario guarantees that if the error model is misspecified, the corrected score test is still valid, and it is usually more efficient than the log-rank test.

For comparison, we also carried out log rank test for this example. The p-values for both corrected score and log-rank tests are almost 0, indicating strong evidence to reject the null hypothesis.

3.7 Discussion

In this chapter, we propose the corrected score and Wald tests, and quantify the impact of measurement error on efficiency loss of the proposed tests. Furthermore, we explore the impact of model misspecification on the consistency of parameter estimation and validity and efficiency of hypothesis testing. We find that the impact is striking. In many situations, the effort of correcting error effects is not rewarding as the resulting corrected parameter estimator or testing procedure perform even worse than those of simply ignoring measurement error. Thus, developing goodness-of-fit test of the overall fit of the survival model and the error model is particularly important. This problem will be explored in Chapter 4. Finally, we note that under several important scenarios, our proposed corrected score and Wald tests are valid even when the error model is misspecified.

The discussion of misspecification is not restricted to score-based methods. One may apply the developments in this chapter to study the impact of misspecification on likelihood estimators (e.g., Hu, Tsiatis and Davidian 1998). We comment that the sample size required to achieve the prespecified size and power of the proposed tests can be calculated, for example, when testing treatment effects in the presence of mismeasured covariates. These extensions would be interesting to be further studied.

Appendix

Appendix A1

Proof of Theorem 1: In the following, we first show the asymptotic expansion of $n^{-1/2}U_c(\beta_0)$. Let

$$S^{(k)}(\hat{Z};\beta_0,t) \quad = \quad n^{-1}\sum_{i=1}^{n} Y_i(t)\hat{Z}_i^{\otimes k}\exp(\hat{Z}_i^T\beta_0),$$

where $k = 0, 1, 2$. Note that $U_c(\beta_0) = U_{nv}(\beta_0) + U_0(\beta_0)$, where

$$U_{nv}(\beta_0) = \sum_{i=1}^{n}\int_0^{\tau}\left\{\hat{Z}_i - \frac{S^{(1)}(\hat{Z};\beta_0,t)}{S^{(0)}(\hat{Z};\beta_0,t)}\right\} dN_i(t),$$

and

$$U_0(\beta_0) = \sum_{i=1}^{n}\int_0^{\tau}\begin{pmatrix} D(\beta_{0,x}) \\ 0 \end{pmatrix} dN_i(t).$$

Note that

$$
\begin{aligned}
n^{-1/2}U_{nv}(\beta_0) &= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{\hat{Z}_i - \frac{s^{(1)}(\hat{Z};\beta_0,t)}{s^{(0)}(\hat{Z};\beta_0,t)}\right\}dN_i(t) \\
&\quad -n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{\frac{S^{(1)}(\hat{Z};\beta_0,t)}{S^{(0)}(\hat{Z};\beta_0,t)} - \frac{s^{(1)}(\hat{Z};\beta_0,t)}{s^{(0)}(\hat{Z};\beta_0,t)}\right\}dN_i(t) \\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{\hat{Z}_i - \frac{s^{(1)}(\hat{Z};\beta_0,t)}{s^{(0)}(\hat{Z};\beta_0,t)}\right\}dN_i(t) \\
&\quad -\int_0^{\tau}\left\{\frac{S^{(1)}(\hat{Z};\beta_0,t)}{S^{(0)}(\hat{Z};\beta_0,t)} - \frac{s^{(1)}(\hat{Z};\beta_0,t)}{s^{(0)}(\hat{Z};\beta_0,t)}\right\}d\left(n^{1/2}\sum_{i=1}^{n}N_i(t)/n - E\{N_i(t)\}\right) \\
&\quad -\int_0^{\tau} n^{1/2}\left\{\frac{S^{(1)}(\hat{Z};\beta_0,t)}{S^{(0)}(\hat{Z};\beta_0,t)} - \frac{s^{(1)}(\hat{Z};\beta_0,t)}{s^{(0)}(\hat{Z};\beta_0,t)}\right\}dE\{N_i(t)\} \\
&\equiv A_1 - A_2 - A_3.
\end{aligned}
$$

Note that by Weak Law of Large Numbers, as $n \to \infty$,

$$
S^{(k)}(\hat{Z};\beta_0,t) \xrightarrow{p} s^{(k)}(\hat{Z};\beta_0,t), \quad k = 0, 1, 2.
$$

Thus,

$$
\frac{S^{(1)}(\hat{Z};\beta_0,t)}{S^{(0)}(\hat{Z};\beta_0,t)} - \frac{s^{(1)}(\hat{Z};\beta_0,t)}{s^{(0)}(\hat{Z};\beta_0,t)} = o_p(1).
$$

Besides, $n^{1/2}\sum_{i=1}^{n} N_i(t)/n - E\{N_i(t)\}$ converges weakly to a mean-zero Gaussian process. Therefore,

$$
A_2 = o_p(1).
$$

Note that by Taylor series expansion,

$$
\begin{aligned}
\frac{S^{(1)}(\hat{Z};\beta_0,t)}{S^{(0)}(\hat{Z};\beta_0,t)} - \frac{s^{(1)}(\hat{Z};\beta_0,t)}{s^{(0)}(\hat{Z};\beta_0,t)} &= \frac{1}{s^{(0)}(Z;\beta_0,t)}\left\{S^{(1)}(\hat{Z};\beta_0,t) - s^{(1)}(Z;\beta_0,t)\right\} \\
&\quad - \frac{s^{(1)}(Z;\beta_0,t)}{\{s^{(0)}(Z;\beta_0,t)\}^2}\left\{S^{(0)}(\hat{Z};\beta_0,t) - s^{(0)}(Z;\beta_0,t)\right\} + o_p(n^{-1/2}) \\
&= \frac{S^{(1)}(\hat{Z};\beta_0,t)}{s^{(0)}(Z;\beta_0,t)} - \frac{s^{(1)}(Z;\beta_0,t)}{\{s^{(0)}(Z;\beta_0,t)\}^2}S^{(0)}(\hat{Z};\beta_0,t) + o_p(n^{-1/2}).
\end{aligned}
$$

Hence,

$$A_3 = \int_0^\tau n^{1/2}\left\{\frac{S^{(1)}(\hat{Z};\beta_0,t)}{s^{(0)}(Z;\beta_0,t)} - \frac{s^{(1)}(Z;\beta_0,t)}{\{s^{(0)}(Z;\beta_0,t)\}^2}S^{(0)}(\hat{Z};\beta_0,t)\right\}dE\{N_i(t)\} + o_p(1).$$

Thus,

$$\begin{aligned}
n^{-1/2}U_c(\beta_0) &= n^{-1/2}U_{c1}(\beta_0) + n^{-1/2}U_{c1}(\beta_0) \\
&= (A_1 - A_2 - A_3) + n^{-1/2}U_{c1}(\beta_0) \\
&= n^{-1/2}\sum_{i=1}^n \int_0^\tau \left\{\hat{Z}_i - \frac{s^{(1)}(\hat{Z};\beta_0,t)}{s^{(0)}(\hat{Z};\beta_0,t)}\right\}dN_i(t) \\
&\quad - \int_0^\tau n^{1/2}\left\{\frac{S^{(1)}(\hat{Z};\beta_0,t)}{s^{(0)}(Z;\beta_0,t)} - \frac{s^{(1)}(Z;\beta_0,t)}{\{s^{(0)}(Z;\beta_0,t)\}^2}S^{(0)}(\hat{Z};\beta_0,t)\right\}dE\{N_i(t)\} \\
&\quad + \sum_{i=1}^n \int_0^\tau \begin{pmatrix} D(\beta_{0,x}) \\ 0 \end{pmatrix} dN_i(t) + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^n \int_0^\tau \left\{\hat{Z}_i - \frac{s^{(1)}(\hat{Z};\beta_0,t)}{s^{(0)}(\hat{Z};\beta_0,t)} + \begin{pmatrix} D(\beta_{0,x}) \\ 0 \end{pmatrix}\right\}dN_i(t) \\
&\quad - n^{-1/2}\sum_{i=1}^n \int_0^\tau \frac{Y_i(t)\exp(\hat{Z}_i^T\beta_0)}{s^{(0)}(\hat{Z};\beta_0,t)}\left\{\hat{Z}_i - \frac{s^{(1)}(\hat{Z};\beta_0,t)}{s^{(0)}(\hat{Z};\beta_0,t)}\right\}dE\{N_i(t)\} + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^n J_i(\beta_0) + o_p(1).
\end{aligned}$$

Note that $J_i(\beta_0)$ are independent mean-zero terms, and that $Var(J_i(\beta_0)) = \mathcal{J}(\beta_0)$. Therefore, as $n \to \infty$,

$$n^{-1/2}U_c(\beta_0) \xrightarrow{d} N(0, \mathcal{J}(\beta_0)).$$

Now, we investigate the asymptotic property of $-n^{-1}\frac{\partial U_c(\beta)}{\partial \beta^T}|_{\beta=\beta_0}$. Note that

$$\begin{aligned}
& -n^{-1}\frac{\partial U_c(\beta)}{\partial \beta^T}|_{\beta=\beta_0} \\
= \; & n^{-1}\sum_{i=1}^n \int_0^\tau \left[\frac{S^{(2)}(\hat{Z};\beta_0,t)}{S^{(0)}(\hat{Z};\beta_0,t)} - \left\{\frac{S^{(1)}(\hat{Z};\beta_0,t)}{S^{(0)}(\hat{Z};\beta_0,t)}\right\}^{\otimes 2} + \begin{pmatrix} \frac{\partial D(\beta_x)}{\partial \beta_x^T}|_{\beta_x=\beta_{0,x}} & 0 \\ 0 & 0 \end{pmatrix}\right]dN_i(t).
\end{aligned}$$

Thus, by Uniform Weak Law of Large Numbers (Pollard 1990), we have

$$\begin{aligned}
& -n^{-1}\frac{\partial U_c(\beta)}{\partial \beta}|_{\beta=\beta_0} \\
\xrightarrow{p} \ & \int_0^\tau \left[\frac{s^{(2)}(\hat{Z};\beta_0,t)}{s^{(0)}(\hat{Z};\beta_0,t)} - \left\{ \frac{s^{(1)}(\hat{Z};\beta_0,t)}{s^{(0)}(\hat{Z};\beta_0,t)} \right\}^{\otimes 2} + \begin{pmatrix} \frac{\partial D(\beta_x)}{\partial \beta_x^T}|_{\beta_x=\beta_{0,x}} & 0 \\ 0 & 0 \end{pmatrix} \right] dE\{N_i(t)\} \\
= \ & \mathcal{I}(\beta_0),
\end{aligned}$$

as $n \to \infty$.

By Taylor series expansion,

$$n^{1/2}(\tilde{\beta}_c^- - \beta_0^-) = \left\{ -n^{-1}\frac{\partial U_c^-(\beta_0^+, \beta^-)}{\partial \beta^{-T}}|_{\beta^-=\beta_0^-} \right\}^{-1} n^{-1/2}U_c^-(\beta_0) + o_p(1).$$

Thus, we obtain by Taylor series expansion that

$$\begin{aligned}
n^{-1/2}U_c^+(\beta_0^+, \tilde{\beta}_c^-) &= n^{-1/2}U_c^+(\beta_0) + \left\{ n^{-1}\frac{\partial U_c^+(\beta_0^+, \beta^-)}{\partial \beta^{-T}}|_{\beta^-=\beta_0^-} \right\} n^{1/2}(\tilde{\beta}_c^- - \beta_0^-) + o_p(1) \\
&= n^{-1/2}U_c^+(\beta_0) + \left\{ n^{-1}\frac{\partial U_c^+(\beta_0^+, \beta^-)}{\partial \beta^{-T}}|_{\beta^-=\beta_0^-} \right\} \\
&\quad \times \left\{ -n^{-1}\frac{\partial U_c^-(\beta_0^+, \beta^-)}{\partial \beta^{-T}}|_{\beta^-=\beta_0^-} \right\}^{-1} n^{-1/2}U_c^-(\beta_0) + o_p(1) \\
&= \begin{pmatrix} I_{r^+}, & -\frac{\partial U_c^+(\beta_0^+, \beta^-)}{\partial \beta^{-T}}|_{\beta^-=\beta_0^-} \left\{ \frac{\partial U_c^-(\beta_0^+, \beta^-)}{\partial \beta^{-T}}|_{\beta^-=\beta_0^-} \right\}^{-1} \end{pmatrix} \\
&\quad \times \begin{pmatrix} n^{-1/2}U_c^+(\beta_0) \\ n^{-1/2}U_c^-(\beta_0) \end{pmatrix} + o_p(1) \\
&= \begin{pmatrix} I_{r^+}, & -\mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0) \end{pmatrix} n^{-1/2}U_c(\beta_0) + o_p(1) \\
&\xrightarrow{d} N\left(0, \begin{pmatrix} I_{r^+}, & -\mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0) \end{pmatrix} \mathcal{J}(\beta_0) \begin{pmatrix} I_{r^+} \\ -\mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0) \end{pmatrix} \right) \\
&= N(0, \mathcal{V}(\beta_0)).
\end{aligned}$$

It follows that as $n \to \infty$,

$$T_{S,c} \xrightarrow{d} \chi^2_{r^+}.$$

Next we show the asymptotic property of $T_{W,c}$. Note that,

$$\begin{aligned} n^{1/2}(\hat{\beta}_c - \beta_0) &= \left\{ -n^{-1}\frac{\partial U_c(\beta)}{\partial \beta^T}|_{\beta=\beta_0} \right\}^{-1} n^{-1/2}U_c(\beta_0) + o_p(1) \\ &= \left\{ \mathcal{I}^{-1T}(\beta_0) \right\} n^{-1/2}U_c(\beta_0) + o_p(1) \\ &\xrightarrow{d} N(0, \mathcal{I}^{-1}(\beta_0)\mathcal{J}(\beta_0)\mathcal{I}^{-1T}(\beta_0)), \text{ as } n \to \infty. \end{aligned}$$

Thus,

$$n^{1/2}(\hat{\beta}_c - \beta_0) \xrightarrow{d} N(0, \mathcal{I}^{-1T}(\beta_0)\mathcal{J}(\beta_0)\mathcal{I}^{-1}(\beta_0)),$$

as $n \to \infty$. Therefore,

$$n^{1/2}(\hat{\beta}_c^+ - \beta_0^+) \xrightarrow{d} N(0, \mathcal{W}_{11}(\beta_0)),$$

as $n \to \infty$. Thus, as $n \to \infty$,

$$T_{W,c} \xrightarrow{d} \chi^2_{r^+}.$$

Thus, Theorem 1 is proved.

Appendix A2

Proof of Theorem 2: Let P_n denote the probability measure under the alternative hypothesis H_n, and let P_0 denote the probability measure under the null hypothesis H_0. Note that a random vector $A = o_{p_n}(1)$ is equivalent to $A = o_p(1)$.

First, we derive the asymptotic distribution of $T_{S,c}$ under the alternative hypothesis H_n. Let β_n^- be the solution of

$$E_n\{U_c^-(\beta_0^+, \beta^-)\} = 0,$$

where the expectation is taken under H_n. Note that

$$E_n\{U_c(\beta_0^+ + b/\sqrt{n}, \beta_0^-)\} = 0,$$

and $\beta_n^- \to \beta_0^-$ as $n \to \infty$. Then under the alternative hypothesis H_n, we have by Taylor series expansion that,

$$\begin{aligned} 0 &= n^{-1/2} E_n\{U_c^-(\beta_0^+, \beta_n^-)\} \\ &= n^{-1/2} E_n\{U_c^-(\beta_0^+, \beta_0^-)\} + \left[n^{-1} \frac{\partial E_n\{U_c^-(\beta_0^+, \beta^-)\}}{\partial \beta^{-T}} |_{\beta^- = \beta_0^-} \right] \sqrt{n}(\beta_n^- - \beta_0^-) + o_p(1). \end{aligned}$$

Therefore,

$$\begin{aligned} &\sqrt{n}(\beta_n^- - \beta_0^-) \\ =& \left[-n^{-1} \frac{\partial E_n\{U_c^-(\beta_0^+, \beta^-)\}}{\partial \beta^{-T}} |_{\beta^- = \beta_0^-} \right]^{-1} n^{-1/2} E_n\{U_c^-(\beta_0^+, \beta_0^-)\} + o_p(1) \\ =& \left[-n^{-1} \frac{\partial E_n\{U_c^-(\beta_0^+, \beta^-)\}}{\partial \beta^{-T}} |_{\beta^- = \beta_0^-} \right]^{-1} \Bigg(n^{-1/2} E_n\{U_c^-(\beta_0^+ + b/\sqrt{n}, \beta_0^-)\} \\ &+ \left[n^{-1} \frac{\partial E_n\{U_c^-(\beta^+, \beta_0^-)\}}{\partial \beta^{+T}} |_{\beta^+ = \beta_0^+ + b/\sqrt{n}} \right] \sqrt{n}\{\beta_0^+ - (\beta_0^+ + b/\sqrt{n})\} + o_p(1) \Bigg) + o_p(1) \\ =& \left[-n^{-1} \frac{\partial E_n\{U_c^-(\beta_0^+, \beta^-)\}}{\partial \beta^{-T}} |_{\beta^- = \beta_0^-} \right]^{-1} \\ &\times \left[-n^{-1} \frac{\partial E_n\{U_c^-(\beta^+, \beta_0^-)\}}{\partial \beta^{+T}} |_{\beta^+ = \beta_0^+ + b/\sqrt{n}} \right] b + o_p(1) \\ =& \mathcal{I}_{22}^{-1}(\beta_0) \mathcal{I}_{21}(\beta_0) b + o_p(1). \end{aligned}$$

Furthermore,

$$
\begin{aligned}
& n^{-1/2}E_n\{U_c^+(\beta_0^+,\beta_n^-)\} \\
=& n^{-1/2}E_n\{U_c^+(\beta_0^+ + b/\sqrt{n},\beta_n^-)\} \\
& + \left[n^{-1}\frac{\partial E_n\{U_c^+(\beta^+,\beta_n^-)\}}{\partial \beta^{+T}}|_{\beta^+=\beta_0^+ + b/\sqrt{n}} \right] \sqrt{n}\{\beta_0^+ - (\beta_0^+ + b/\sqrt{n})\} + o_p(1) \\
=& n^{-1/2}E_n\{U_c^+(\beta_0^+ + b/\sqrt{n},\beta_0^-)\} \\
& + \left[n^{-1}\frac{\partial E_n\{U_c^+(\beta_0^+ + b/\sqrt{n},\beta^-)\}}{\partial \beta^{-T}}|_{\beta^-=\beta_0^-} \right] \sqrt{n}(\beta_n^- - \beta_0^-) \\
& + \left[-n^{-1}\frac{\partial E_n\{U_c^+(\beta^+,\beta_n^-)\}}{\partial \beta^{+T}}|_{\beta^+=\beta_0^+ + b/\sqrt{n}} \right] b + o_p(1) \\
=& \left[n^{-1}\frac{\partial E_n\{U_c^+(\beta_0^+ + b/\sqrt{n},\beta^-)\}}{\partial \beta^{-T}}|_{\beta^-=\beta_0^-} \right] \mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}^{-1}(\beta_0)b \\
& + \left[-n^{-1}\frac{\partial E_n\{U_c^+(\beta^+,\beta_n^-)\}}{\partial \beta^{+T}}|_{\beta^+=\beta_0^+ + b/\sqrt{n}} \right] b + o_p(1) \\
=& -\{\mathcal{I}_{11}(\beta_0) - \mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0)\}b + o_p(1).
\end{aligned}
$$

Together with the fact that $E_n\{U_c^-(\beta_0^+,\beta_n^-)\} = 0$, we have

$$
n^{-1/2}E_n\{U_c(\beta_0^+,\beta_n^-)\} = \begin{pmatrix} -\{\mathcal{I}_{11}(\beta_0) - \mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0)\}b \\ 0 \end{pmatrix} + o_p(1).
$$

Following the arguments in Appendix A2 , we have that under the alternative hypothesis H_n,

$$
n^{-1/2}U_c^+(\beta_0^+,\tilde{\beta}_c^-) = \begin{pmatrix} 1_{r^+}, & -\mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0) \end{pmatrix} n^{-1/2}U_c(\beta_0) + o_p(1).
$$

Thus,

$$
\begin{aligned}
& n^{-1/2}U_c^+(\beta_0^+, \tilde{\beta}_c^-) \\
=& \left(\; I_{r^+}, \quad -\mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0) \; \right) \\
& \times \left(n^{-1/2}U_c(\beta_0^+, \beta_n^-) + \left\{ -n^{-1}\frac{\partial U_c(\beta_0^+, \beta^-)}{\partial \beta^{-T}}|_{\beta^-=\beta_n^-} \right\} \sqrt{n}(\beta_n^- - \beta_0^-) \right) + o_p(1) \\
=& \left(\; I_{r^+}, \quad -\mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0) \; \right) \Bigg(n^{-1/2}U_c(\beta_0^+, \beta_n^-) \\
& + \left\{ -n^{-1}\frac{\partial U_c(\beta_0^+, \beta^-)}{\partial \beta^{-T}}|_{\beta^-=\beta_n^-} \right\} \{\mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0)b + o_p(1)\} \Bigg) + o_p(1) \\
=& \left(\; I_{r^+}, \quad -\mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0) \; \right) n^{-1/2}U_c(\beta_0^+, \beta_n^-) \\
& + \left(\; I_{r^+}, \quad -\mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0) \; \right) \begin{pmatrix} -n^{-1}\frac{\partial U_c^+(\beta_0^+, \beta^-)}{\partial \beta^{-T}}|_{\beta^-=\beta_n^-} \\ -n^{-1}\frac{\partial U_c^-(\beta_0^+, \beta^-)}{\partial \beta^{-T}}|_{\beta^-=\beta_n^-} \end{pmatrix} \mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0)b + o_p(1) \\
=& \left(\; I_{r^+}, \quad -\mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0) \; \right) n^{-1/2}U_c(\beta_0^+, \beta_n^-) \\
& + \left(\; I_{r^+}, \quad -\mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0) \; \right) \begin{pmatrix} \mathcal{I}_{12}(\beta_0) \\ \mathcal{I}_{22}(\beta_0) \end{pmatrix} \mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0)b + o_p(1) \\
=& \left(\; I_{r^+}, \quad -\mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0) \; \right) n^{-1/2}U_c(\beta_0^+, \beta_n^-) + o_p(1) \\
=& \left(\; I_{r^+}, \quad -\mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0) \; \right) n^{-1/2}[U_c(\beta_0^+, \beta_n^-) - E_n\{U_c(\beta_0^+, \beta_n^-)\}] \\
& + \left(\; I_{r^+}, \quad -\mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0) \; \right) \begin{pmatrix} -\{\mathcal{I}_{11}(\beta_0) - \mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0)\}b \\ 0 \end{pmatrix} + o_p(1) \\
=& \left(\; I_{r^+}, \quad -\mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0) \; \right) n^{-1/2}[U_c(\beta_0^+, \beta_n^-) - E_n\{U_c(\beta_0^+, \beta_n^-)\}] \\
& - \{\mathcal{I}_{11}(\beta_0) - \mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0)\}b + o_p(1) \\
\xrightarrow{d}& N\left(-\{\mathcal{I}_{11}(\beta_0) - \mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0)\}b, \quad \mathcal{V}(\beta_0) \right), \quad \text{as } n \to \infty.
\end{aligned}
$$

It follows that

$$\begin{aligned}
T_{S,c} =& n^{-1} U_c^{+T}(\beta_0^+, \tilde{\beta}_c^-) \left\{ \hat{V}^{-1}(\beta_0^+, \tilde{\beta}_c^-) \right\} U_c^+(\beta_0^+, \tilde{\beta}_c^-) \\
=& n^{-1} U_c^{+T}(\beta_0^+, \tilde{\beta}_c^-) \mathcal{V}^{-1}(\beta_0) U_c^+(\beta_0^+, \tilde{\beta}_c^-) + o_p(1) \\
\xrightarrow{d}& \chi^2_{r^+}(\delta), \quad \text{as } n \to \infty,
\end{aligned}$$

where

$$\delta = \frac{1}{2} b^T \left\{ \mathcal{I}_{11}(\beta_0) - \mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0) \right\} \mathcal{V}^{-1}(\beta_0) \left\{ \mathcal{I}_{11}(\beta_0) - \mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0) \right\} b.$$

Next, we derive the asymptotic distribution of $T_{W,c}$ under the alternative hypothesis H_n. Note that under mild regularity conditions, $\hat{\beta}_c$ is a regular estimator (Tsiatis 2006, p. 27). Thus,

$$\sqrt{n}\{\hat{\beta}_c^+ - (\beta_0^+ + b/\sqrt{n})\} \xrightarrow{d} N(0, \mathcal{W}_{11}(\beta_0)), \quad \text{as } n \to \infty.$$

Therefore,

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_c^+ - \beta_0^+) =& \sqrt{n}\{\hat{\beta}_c^+ - (\beta_0^+ + b/\sqrt{n})\} + b \\
\xrightarrow{d}& N(b, \mathcal{W}_{11}(\beta_0)), \quad \text{as } n \to \infty.
\end{aligned}$$

It follows that

$$\begin{aligned}
T_{W,c} =& n(\hat{\beta}_c^+ - \hat{\beta}_0^+)^T \left\{ \hat{W}_{11}^{-1}(\hat{\beta}_c) \right\} (\hat{\beta}_c^+ - \hat{\beta}_0^+) \\
=& n(\hat{\beta}_c^+ - \hat{\beta}_0^+)^T \left\{ \mathcal{W}_{11}^{-1}(\beta_0) \right\} (\hat{\beta}_c^+ - \hat{\beta}_0^+) + o_p(1) \\
\xrightarrow{d}& \chi^2_{r^+}(\delta^*), \quad \text{as } n \to \infty,
\end{aligned}$$

where

$$\delta^* = \frac{1}{2} b^T \mathcal{W}_{11}^{-1}(\beta_0) b.$$

Following the arguments in Appendix A of Ma et al. (2011), we can show that $\delta^* = \delta$, and thus

$$T_{W,c} \xrightarrow{d} \chi^2_{r^+}(\delta), \quad \text{as } n \to \infty.$$

Thus, Theorem 2 is proved.

Appendix A3

Proof of Corollary 1: Let P_n be the probability measure under H_n, and let $P_n(T_{S,c} > \chi^2_{r^+,\alpha})$ be the power function. By the proof of Appendix A2,

$$n^{-1/2}U_c^+(\beta_0^+, \tilde{\beta}_c^-) \xrightarrow{d} N\left(-\{\mathcal{I}_{11}(\beta_0) - \mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0)\}b, \quad \mathcal{V}(\beta_0)\right), \quad \text{as } n \to \infty.$$

Therefore,

$$\begin{aligned}
&\lim_{n\to\infty} P_n(T_{S,c} > \chi^2_{r^+,\alpha}) \\
=&\lim_{n\to\infty} P_n\Big(n^{-1}[U_c^+(\beta_0^+, \tilde{\beta}_c^-) + \sqrt{n}\{\mathcal{I}_{11}(\beta_0) - \mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0)\}b]^2/\mathcal{V}(\beta_0) \\
&> \chi^2_{r^+,\alpha} + \delta\Big) \\
=&1 - \Psi(\chi^2_{r^+,\alpha} + \delta),
\end{aligned}$$

where Ψ is the cumulative distribution function of the $\chi^2_{r^+}$ distribution. Let ρ be a constant satisfying $0 < \alpha < \rho < 1$. Therefore, the limiting power of the corrected score test against H_n is ρ if and only if

$$\chi^2_{r^+,\alpha} + \delta = \chi^2_{r^+,\rho},$$

or equivalently,

$$\{\mathcal{I}_{11}(\beta_0) - \mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0)\}^2 b^2/\mathcal{V}(\beta_0) = \chi^2_{r^+,\rho} - \chi^2_{r^+,\alpha}.$$

Since $b = \lim_{n\to\infty} \sqrt{n}(\beta_n^+ - \beta_0^+)$, the limiting power of the corrected score test against H_n is ρ if and only if the sample size of the corrected score test satisfies

$$n_{S,c} \sim \frac{(\chi^2_{r^+,\rho} - \chi^2_{r^+,\alpha})\mathcal{V}(\beta_0)}{\{\mathcal{I}_{11}(\beta_0) - \mathcal{I}_{12}(\beta_0)\mathcal{I}_{22}^{-1}(\beta_0)\mathcal{I}_{21}(\beta_0)\}^2(\beta_n^+ - \beta_0^+)^2} = \frac{(\chi^2_{r^+,\rho} - \chi^2_{r^+,\alpha})b^2}{\delta(\beta_n^+ - \beta_0^+)^2},$$

where $A \sim B$ means $\lim_{n\to\infty} \frac{A}{B} = 1$. Similarly, the limiting power of the partial score test against H_n is ρ if and only if the sample size of the partial score test satisfies

$$n_S \sim \frac{(\chi^2_{r^+,\rho} - \chi^2_{r^+,\alpha})b^2}{\delta_0(\beta_n^+ - \beta_0^+)^2}.$$

Thus, the ARE of the corrected score test relative to the partial score test is

$$\lim_{n\to\infty} \frac{n_S}{n_{S,c}} = \frac{\delta}{\delta_0}.$$

It remains to prove that $\delta < \delta_0$. Let $\mathcal{V}_0(\beta_0)$ be $\mathcal{V}(\beta_0)$ with $\hat{Z}_i$ replaced by Z_i, and $D(\beta_{0,x})$ ignored. Note that $\mathcal{V}_0(\beta_0)$ is the asymptotic variance of the partial score estimator. Therefore, $\mathcal{V}_0(\beta_0) < \mathcal{V}(\beta_0)$. Thus,

$$\frac{\delta}{\delta_0} = \frac{\mathcal{V}_0(\beta_0)}{\mathcal{V}(\beta_0)} < 1.$$

Thus, Corollary 1 is proved.

Appendix A4

Note that

$$\mathcal{I}(\beta) = \int_0^\tau \left[\frac{s^{(2)}(Z;\beta,t)}{s^{(0)}(Z;\beta,t)} - \left\{ \frac{s^{(1)}(Z;\beta,t)}{s^{(0)}(Z;\beta,t)} \right\}^{\otimes 2} \right] dE\{N_i(t)\}.$$

Note also that in this randomized trial setting, $\beta^+ = \beta_v$, $D^+(\beta_x) = 0$, and $\hat{Z}_i^+ = V_i$. By assumption, $V_i \perp\!\!\!\perp C_i$, and under the null hypothesis H_0, $V_i \perp\!\!\!\perp T_i$, then we obtain that

$$\frac{E\{Y_i(t)V_i\}}{E\{Y_i(t)\}} = E(V_i),$$

and

$$\begin{aligned}
\mathcal{I}_{12}(\beta_0) = \mathcal{I}_{21}^T(\beta_0) &= \int_0^\tau \Bigg(\frac{E\{Y_i(t)X_i^T V_i \exp(X_i^T\beta_{0,x})\}}{E\{Y_i(t)\exp(X_i^T\beta_{0,x})\}} \\
&\quad - \frac{E\{Y_i(t)V_i \exp(X_i^T\beta_{0,x})\}E\{Y_i(t)X_i^T \exp(X_i^T\beta_{0,x})\}}{[E\{Y_i(t)\exp(X_i^T\beta_{0,x})\}]^2} \Bigg) dE\{N_i(t)\} \\
&= \int_0^\tau \Bigg(\frac{E(V_i)E\{Y_i(t)X_i^T \exp(X_i^T\beta_{0,x})\}}{E\{Y_i(t)\exp(X_i^T\beta_{0,x})\}} \\
&\quad - \frac{E(V_i)E\{Y_i(t)\exp(X_i^T\beta_{0,x})\}E\{Y_i(t)X_i^T \exp(X_i^T\beta_{0,x})\}}{[E\{Y_i(t)\exp(X_i^T\beta_{0,x})\}]^2} \Bigg) dE\{N_i(t)\} \\
&= 0.
\end{aligned}$$

Let $\mathcal{J}_{11}(\beta)$ be the upper left first element of $\mathcal{J}(\beta)$. Then

$$
\begin{aligned}
\mathcal{V}(\beta_0) =& \begin{pmatrix} 1, & 0 \end{pmatrix} \mathcal{J}(\beta_0) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
=& \mathcal{J}_{11}(\beta_0) \\
=& E\Bigg(\Bigg[\int_0^\tau \left\{ V_i - \frac{E\{Y_i(t) V_i \exp(W_i^T \beta_{0,x})\}}{E\{Y_i(t) \exp(W_i^T \beta_{0,x})\}} \right\} dN_i(t) \\
& - \int_0^\tau \frac{Y_i(t) \exp(W_i^T \beta_{0,x})}{E\{Y_i(t) \exp(W_i^T \beta_{0,x})\}} \left\{ V_i - \frac{E\{Y_i(t) V_i \exp(W_i^T \beta_{0,x})\}}{E\{Y_i(t) \exp(W_i^T \beta_{0,x})\}} \right\} dE\{N_i(t)\} \Bigg]^2 \Bigg) \\
=& E\left[\int_0^\tau \{V_i - E(V_i)\} \left\{ dN_i(t) - \frac{Y_i(t) \exp(W_i^T \beta_{0,x})}{E\{\exp(\epsilon_i^T \beta_{0,x})\}} \lambda_0(t) dt \right\} \right]^2 .
\end{aligned}
$$

Note that

$$
N_i(t) - \int_0^t Y_i(u) \exp(X_i^T \beta_{0,x}) \lambda_0(u) du
$$

is a martingale (Kalbfleisch and Prentice 2002). By martingale properties,

$$
\begin{aligned}
& E\left[\int_0^\tau \{V_i - E(V_i)\} \left\{ dN_i(t) - Y_i(t) \exp(X_i^T \beta_{0,x}) \lambda_0(t) dt \right\} \right]^2 \\
=& E\left[\int_0^\tau \{V_i - E(V_i)\}^2 Y_i(t) \exp(X_i^T \beta_{0,x}) \lambda_0(t) dt \right] .
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathcal{V}(\beta_0) =& E\left[\int_0^\tau \{V_i - E(V_i)\}\left\{dN_i(t) - \frac{Y_i(t)\exp(W_i^T\beta_{0,x})}{E\{\exp(\epsilon_i^T\beta_{0,x})\}}\lambda_0(t)dt\right\}\right]^2 \\
=& E\Bigg(\int_0^\tau \{V_i - E(V_i)\}\Bigg[\left\{dN_i(t) - Y_i(t)\exp(X_i^T\beta_{0,x})\lambda_0(t)dt\right\} \\
&+ \left\{Y_i(t)\exp(X_i^T\beta_{0,x})\lambda_0(t)dt - \frac{Y_i(t)\exp(W_i^T\beta_{0,x})}{E\{\exp(\epsilon_i^T\beta_{0,x})\}}\lambda_0(t)dt\right\}\Bigg]\Bigg)^2 \\
=& E\left[\int_0^\tau \{V_i - E(V_i)\}\left\{dN_i(t) - Y_i(t)\exp(X_i^T\beta_{0,x})\lambda_0(t)dt\right\}\right]^2 \\
&+ E\left[\int_0^\tau \{V_i - E(V_i)\}\left\{Y_i(t)\exp(X_i^T\beta_{0,x})\lambda_0(t)dt - \frac{Y_i(t)\exp(W_i^T\beta_{0,x})}{E\{\exp(\epsilon_i^T\beta_{0,x})\}}\lambda_0(t)dt\right\}\right]^2 \\
&+ 2E\Bigg(\int_0^\tau \{V_i - E(V_i)\}\left\{dN_i(t) - Y_i(t)\exp(X_i^T\beta_{0,x})\lambda_0(t)dt\right\} \\
&\times \int_0^\tau \{V_i - E(V_i)\}\left\{Y_i(t)\exp(X_i^T\beta_{0,x})\lambda_0(t)dt - \frac{Y_i(t)\exp(W_i^T\beta_{0,x})}{E\{\exp(\epsilon_i^T\beta_{0,x})\}}\lambda_0(t)dt\right\}\Bigg) \\
=& E\left[\int_0^\tau \{V_i - E(V_i)\}^2 Y_i(t)\exp(X_i^T\beta_{0,x})\lambda_0(t)dt\right] \\
&+ E\left[\int_0^\tau \{V_i - E(V_i)\}\left\{Y_i(t)\exp(X_i^T\beta_{0,x})\lambda_0(t)dt - \frac{Y_i(t)\exp(W_i^T\beta_{0,x})}{E\{\exp(\epsilon_i^T\beta_{0,x})\}}\lambda_0(t)dt\right\}\right]^2 \\
=& Var(V_i)E\left\{\int_0^\tau Y_i(t)\exp(X_i^T\beta_{0,x})\lambda_0(t)dt\right\} \\
&+ \frac{Var\{\exp(\epsilon_i\beta_{0,x})\}}{[E\{\exp(\epsilon_i\beta_{0,x})\}]^2}Var(V_i)E\left\{\int_0^\tau Y_i(t)\exp(X_i^T\beta_{0,x})\lambda_0(t)dt\right\}^2,
\end{aligned}
$$

and $\delta = b^2\mathcal{I}_{11}^2(\beta_0)/\mathcal{V}(\beta_0)$. Let $\mathcal{V}_0(\beta_0)$ be $\mathcal{V}(\beta_0)$ with $\hat{Z}_i$ replaced by Z_i, and $D(\beta_{0,x})$ ignored. Similarly we have

$$
\mathcal{V}_0(\beta_0) = Var(V_i)E\left\{\int_0^\tau Y_i(t)\exp(X_i^T\beta_{0,x})\lambda_0(t)dt\right\},
$$

and $\delta_0 = b^2\mathcal{I}_{11}^2(\beta_0)/\mathcal{V}_0(\beta_0)$. Therefore,

$$
ARE(T_{S,c};T_S) = \frac{\delta}{\delta_0} = \left(1 + \frac{Var\{\exp(\epsilon_i\beta_{0,x})\}}{[E\{\exp(\epsilon_i\beta_{0,x})\}]^2}\frac{E\left\{\int_0^\tau Y_i(t)\exp\left(X_i^T\beta_{0,x}\right)\lambda_0(t)dt\right\}^2}{E\left\{\int_0^\tau Y_i(t)\exp\left(X_i^T\beta_{0,x}\right)\lambda_0(t)dt\right\}}\right)^{-1},
$$

and thus (3.4) is proved. It remains to prove

$$\left(1+2\frac{Var\{\exp\left(\epsilon_i\beta_{0,x}\right)\}}{[E\{\exp\left(\epsilon_i\beta_{0,x}\right)\}]^2}\right)^{-1}\le ARE(T_{S,c};T_S),$$

or equivalently,

$$\frac{E\left\{\int_0^\tau Y_i(t)\exp\left(X_i^T\beta_{0,x}\right)\lambda_0(t)dt\right\}^2}{E\left\{\int_0^\tau Y_i(t)\exp\left(X_i^T\beta_{0,x}\right)\lambda_0(t)dt\right\}}\le 2.$$

Let $\Lambda(t) = \int_0^t \lambda_0(u)du$ be the cumulative hazard function. Let $f_{C,X}(\cdot)$ denote the unknown joint density function of the censoring time C_i and the covariates X_i, respectively. Let $a(t)=\Lambda_0(t)\exp(x^T\beta_{0,x})$. Then

$$\begin{aligned}
&E\left\{\int_0^\tau Y_i(t)\exp\left(X_i^T\beta_{0,x}\right)\lambda_0(t)dt\right\}^2\\
=&E\left\{\exp\left(2X_i^T\beta_{0,x}\right)\Lambda_0^2(\min(T_i,C_i))\right\}\\
=&\int_{-\infty}^{\infty}\int_0^\tau\int_0^c \exp\left(2x^T\beta_{0,x}\right)\Lambda_0^2(t)\lambda_0(t)\exp(x^T\beta_{0,x})\exp\{-\Lambda_0(t)\exp(x^T\beta_{0,x})\}f_{C,X}(c,x)dtdcdx\\
&+\int_{-\infty}^{\infty}\int_0^\tau\int_c^\infty \exp\left(2x^T\beta_{0,x}\right)\Lambda_0^2(c)\lambda_0(t)\exp(x^T\beta_{0,x})\exp\{-\Lambda_0(t)\exp(x^T\beta_{0,x})\}f_{C,X}(c,x)dtdcdx\\
=&\int_{-\infty}^{\infty}\int_0^\tau\int_0^{a(c)} y^2\exp(-y)dyf_{C,X}(c,x)dcdx+\int_{-\infty}^{\infty}\int_0^\tau\int_{a(c)}^\infty a^2(c)\exp(-y)dyf_{C,X}(c,x)dcdx\\
=&\int_{-\infty}^{\infty}\int_0^\tau\left\{2-2\exp(-a(c))-2a(c)\exp(-a(c))-a^2(c)\exp(-a(c))\right\}f_{C,X}(c,x)dcdx\\
&+\int_{-\infty}^{\infty}\int_0^\tau a^2(c)\exp(-a(c))f_{C,X}(c,x)dcdx\\
=&\int_{-\infty}^{\infty}\int_0^\tau 2\left\{1-\exp(-a(c))-a(c)\exp(-a(c))\right\}f_{C,X}(c,x)dcdx.
\end{aligned}$$

Similarly,

$$
\begin{aligned}
& E\left\{\int_0^\tau Y_i(t)\exp\left(X_i^T\beta_{0,x}\right)\lambda_0(t)dt\right\} \\
=& E\left\{\exp\left(X_i^T\beta_{0,x}\right)\Lambda_0(\min(T_i,C_i))\right\} \\
=& \int_{-\infty}^{\infty}\int_0^\tau\int_0^{a(c)} y\exp(-y)dy f_{C,X}(c,x)dcdx + \int_{-\infty}^{\infty}\int_0^\tau\int_{a(c)}^{\infty} a(c)\exp(-y)dy f_{C,X}(c,x)dcdx \\
=& \int_{-\infty}^{\infty}\int_0^\tau \left\{1-\exp(-a(c)) - a(c)\exp(-a(c))\right\} f_{C,X}(c,x)dcdx \\
& + \int_{-\infty}^{\infty}\int_0^\tau a(c)\exp(-a(c)) f_{C,X}(c,x)dcdx \\
=& \int_{-\infty}^{\infty}\int_0^\tau \left\{1-\exp(-a(c))\right\} f_{C,X}(c,x)dcdx.
\end{aligned}
$$

Therefore, by the fact that $a(t)$ is a increasing function, we have

$$
E\left\{\int_0^\tau Y_i(t)\exp\left(X_i^T\beta_{0,x}\right)\lambda_0(t)dt\right\}^2 \le 2E\left\{\int_0^\tau Y_i(t)\exp\left(X_i^T\beta_{0,x}\right)\lambda_0(t)dt\right\}.
$$

Appendix A5

Proof of Theorem 3: By Taylor series expansion,

$$
n^{1/2}(\hat\beta_c - \beta_c) = \left\{-n^{-1}\frac{\partial U_c(\beta)}{\partial\beta^T}|_{\beta=\beta_c}\right\}^{-1} n^{-1/2}U_c(\beta_c) + o_p(1).
$$

By following the arguments in Appendix A1, we have

$$
\begin{aligned}
n^{-1/2}U_c(\beta_c) &= n^{-1/2}\sum_{i=1}^n\int_0^\tau\left\{\hat Z_i - \frac{s^{(1)}(\hat Z;\beta_c,t)}{s^{(0)}(\hat Z;\beta_c,t)} + \begin{pmatrix} D(\beta_{c,x}) \\ 0 \end{pmatrix}\right\}dN_i(t) \\
& -n^{-1/2}\sum_{i=1}^n\int_0^\tau \frac{Y_i(t)\exp(\hat Z_i^T\beta_c)}{s^{(0)}(\hat Z;\beta_c,t)}\left\{\hat Z_i - \frac{s^{(1)}(\hat Z;\beta_c,t)}{s^{(0)}(\hat Z;\beta_c,t)}\right\}dE\{N_i(t)\} + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^n J_i(\beta_c) + o_p(1).
\end{aligned}
$$

Therefore, as $n \to \infty$,

$$n^{-1/2}U_c(\beta_c) \xrightarrow{d} N(0, \mathcal{J}(\beta_c)).$$

Note that

$$\begin{aligned}
& -n^{-1}\frac{\partial U_c(\beta)}{\partial \beta^T}|_{\beta=\beta_c} \\
= \;& n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left[\frac{S^{(2)}(\hat{Z};\beta_c,t)}{S^{(0)}(\hat{Z};\beta_c,t)} - \left\{\frac{S^{(1)}(\hat{Z};\beta_c,t)}{S^{(0)}(\hat{Z};\beta_c,t)}\right\}^{\otimes 2} + \begin{pmatrix} \frac{\partial D(\beta_x)}{\partial \beta_x^T}|_{\beta_x=\beta_{c,x}} & 0 \\ 0 & 0 \end{pmatrix}\right] dN_i(t) \\
\xrightarrow{p} \;& \int_0^{\tau}\left[\frac{s^{(2)}(\hat{Z};\beta_c,t)}{s^{(0)}(\hat{Z};\beta_c,t)} - \left\{\frac{s^{(1)}(\hat{Z};\beta_c,t)}{s^{(0)}(\hat{Z};\beta_c,t)}\right\}^{\otimes 2} + \begin{pmatrix} \frac{\partial D(\beta_x)}{\partial \beta_x^T}|_{\beta_x=\beta_{c,x}} & 0 \\ 0 & 0 \end{pmatrix}\right] dE\{N_i(t)\} \\
= \;& \mathcal{I}(\beta_c),
\end{aligned}$$

as $n \to \infty$. Therefore,

$$\begin{aligned}
n^{1/2}(\hat{\beta}_c - \beta_c) &= \left\{-n^{-1}\frac{\partial U_c(\beta)}{\partial \beta^T}|_{\beta=\beta_c}\right\}^{-1} n^{-1/2}U_c(\beta_c) + o_p(1) \\
&= \left\{\mathcal{I}^{-1T}(\beta_c)\right\} n^{-1/2}U_c(\beta_c) + o_p(1) \\
&\xrightarrow{d} N(0, \mathcal{I}^{-1T}(\beta_c)\mathcal{J}(\beta_c)\mathcal{I}^{-1}(\beta_c)), \text{ as } n \to \infty.
\end{aligned}$$

Thus, Theorem 3 is proved.

The derivations above require certain regularity conditions, including those listed in Appendix A1 of Chapter 2.

Supplementary Material

Appendix S1

We denote the scenario described in Section 4.1 as Scenario 1, where the true error model (3.7) is misspecified. We then consider three more scenarios, where $Z_i = X_i$, and X_i is a univariate covariate.

Scenario 2: classical additive error model with different degree of errors

Suppose we misspecify the measurement error model to be (3.2), but the true error model is

$$W_i = X_i + K\epsilon_i, \tag{3.12}$$

where $K \neq 1$ is a constant, and ϵ_i has the same distribution of the error term in the classical error model (3.2).

We obtain the following results: (i). If $\beta_{0,x} = 0$, then $\beta_{c,x} = 0$. (ii). If $\beta_{0,x} \neq 0$ and $K > 1$, then $|\beta_{c,x}| < |\beta_{0,x}|$, and $\beta_{c,x}$ and $\beta_{0,x}$ have the same sign; the degree of attenuation of $\beta_{c,x}$ increases as the degree of measurement error increases, or K increases. (iii). If $\beta_{0,x} \neq 0$ and $K < 1$ and K is close to 1, then $|\beta_{c,x}| > |\beta_{0,x}|$, and $\beta_{c,x}$ and $\beta_{0,x}$ have the same sign.

Scenario 3: Berkson error model

Suppose that we misspecify the measurement error model to be (3.2) with normally distributed covariate and error, but the true error model is the Berkson error model:

$$X_i = W_i + \epsilon_i. \tag{3.13}$$

The Berkson error model can be rewritten as

$$W_i = \gamma_0 + X_i\gamma_x + \epsilon_i^*,$$

where $\gamma_x = \Sigma_w/(\Sigma_w + \Sigma_0)$ with Σ_w being the variance of W_i, $\gamma_0 = 1 - \gamma_x$, and ϵ_i^* is the error term with mean 0 and variance $\gamma_x\Sigma_0$, and is independent of X_i. When Σ_0 is relatively small compared to Σ_w, we obtain that $|\beta_{c,x}| > \gamma_x^{-1}|\beta_{0,x}| > |\beta_{0,x}|$, and $\beta_{c,x}$ and $\beta_{0,x}$ have the same sign. Furthermore, the degree of inflation of β_c increases as the degree of measurement error increases.

Scenario 4: classical additive error model with different error distributions

Suppose that the true error model is a classical additive error model

$$W_i = X_i + \epsilon_i^*, \tag{3.14}$$

where the distribution of ϵ_i^* and that of ϵ_i from the misspecified error model (3.2) are different. That is, the scenario includes Scenario 2 as a special case. It is generally difficult to sort out the relationship between β_c and β_0 deterministically. Numerical approximations are often invoked to study the impact of misspecifying the error distribution on response parameter estimation. In our simulation studies, we specifically consider the impact of misspecifying the error distribution as normal, while the true distribution is a uniform, a logistic, a mixture normal, or an exponential distribution.

Now we justify the claims in Scenarios 1-3. Let

$$\begin{aligned}\tilde{U}(\beta) = &\int_0^\tau E\left\{Y_i(t)Z_i \exp\left(Z_i^T\beta_0\right)\right\}\lambda_0(t)dt \\ &- \int_0^\tau \frac{E\left\{Y_i(t)Z_i\exp\left(Z_i^T\beta\right)\right\}}{E\left\{Y_i(t)\exp\left(Z_i^T\beta\right)\right\}} E\left\{Y_i(t)\lambda_0(t)\exp\left(Z_i^T\beta_0\right)\right\}dt.\end{aligned}$$

Note that β_0 is the unique solution of $\tilde{U}(\beta) = 0$.

We first consider Scenario 1. Under the underlying error model (3.2), we have

$$\begin{aligned}&\tilde{U}_c(\beta) \\ = &\int_0^\tau E\left\{Y_i(t)\hat{Z}_i\exp\left(Z_i^T\beta_0\right)\right\}\lambda_0(t)dt - \int_0^\tau \frac{E\left\{Y_i(t)\hat{Z}_i\exp\left(\hat{Z}_i^T\beta\right)\right\}}{E\left\{Y_i(t)\exp\left(\hat{Z}_i^T\beta\right)\right\}} E\left\{Y_i(t)\lambda_0(t)\exp\left(Z_i^T\beta_0\right)\right\}dt \\ &+ \begin{pmatrix}\Sigma_0\beta_x \\ 0\end{pmatrix}\int_0^\tau E\left\{Y_i(t)\exp\left(Z_i^T\beta_0\right)\right\}\lambda_0(t)dt \\ = &\begin{pmatrix}\gamma_x & \gamma_v^T \\ 0 & I\end{pmatrix}\tilde{U}(\gamma_x\beta_x, \beta_v + \gamma_v\beta_x).\end{aligned}$$

Since $\tilde{U}_c(\beta_c) = 0$, we have $\tilde{U}(\gamma_x\beta_{c,x}, \beta_{c,v} + \gamma_v\beta_{c,x}) = 0$. By the uniqueness of the solution of $\tilde{U}(\beta) = 0$, we obtain that $\beta_{0,x} = \gamma_x\beta_{c,x}$, and $\beta_{0,v} = \beta_{c,v} + \gamma_v\beta_{c,x}$.

Second, we consider Scenario 2. Under the underlying error model (3.12), we have

$$\tilde{U}_c(\beta_x) = \tilde{U}(\beta_x) + (1-K)\Sigma_0\beta_x \int_0^\tau E\left\{Y_i(t)\exp\left(X_i\beta_{0,x}\right)\right\}\lambda_0(t)dt.$$

It is obvious that when $\beta_{0,x} = 0$, we have $\beta_{c,x} = 0$. Now suppose $K > 1$. It is straightforward that $\frac{\partial \tilde{U}_c(\beta_x)}{\partial \beta_x} < 0$. Thus $\tilde{U}_c(\beta_x)$ is a decreasing function of β_x, and there is unique solution of $\tilde{U}_c(\beta_x) = 0$. Observe that

$$\tilde{U}_c(\beta_{0,x}) = (1-K)\Sigma_0\beta_{0,x} \int_0^\tau E\left\{Y_i(t)\exp\left(X_i\beta_{0,x}\right)\right\}\lambda_0(t)dt.$$

Therefore, when $\beta_{0,x} > 0$, then $\tilde{U}_c(\beta_{0,x}) < 0$, and thus $\beta_{c,x} < \beta_{0,x}$; when $\beta_{0,x} < 0$, then $\tilde{U}_c(\beta_{0,x}) > 0$, and thus $\beta_{c,x} > \beta_{0,x}$. For the case where $K < 1$ and K is sufficiently close to 1,$\tilde{U}_c(\beta_x)$ is a decreasing function of β_x, and there is unique solution of $\tilde{U}_c(\beta_x) = 0$. When $\beta_{0,x} > 0$, then $\tilde{U}_c(\beta_{0,x}) > 0$, and thus $\beta_{c,x} > \beta_{0,x}$; when $\beta_{0,x} < 0$, then $\tilde{U}_c(\beta_{0,x}) < 0$, and thus $\beta_{c,x} < \beta_{0,x}$.

Next, we consider Scenario 3. Under the underlying error model (3.13), we have

$$\tilde{U}_c(\beta_x) = \gamma_x\tilde{U}(\gamma_x\beta_x) + (1-\gamma_x)\Sigma_0\beta_x \int_0^\tau E\left\{Y_i(t)\exp\left(X_i\beta_{0,x}\right)\right\}\lambda_0(t)dt,$$

where $\gamma_x = \Sigma_w/(\Sigma_w + \Sigma_0)$. Under the assumption that β_c exist and is unique, and when Σ_0 is relatively small, we have that $\tilde{U}_c(\beta_x)$ is a decreasing function of β_x. When $\beta_{0,x} > 0$, then $\tilde{U}_c(\gamma_x^{-1}\beta_{0,x}) > 0$, and thus $\beta_{c,x} > \gamma_x^{-1}\beta_{0,x}$; when $\beta_{0,x} < 0$, then $\tilde{U}_c(\gamma_x^{-1}\beta_{0,x}) < 0$, and thus $\beta_{c,x} < \gamma_x^{-1}\beta_{0,x}$.

Appendix S2

We denote the scenario described in Section 4.1 as Scenario 1′. We then consider three more scenarios.

Scenario $2'$: Suppose that W_i is linked with X_i and V_i through one of the four error models (3.7), (3.12), (3.13), and (3.14) specified in Scenarios 1-4 in Section 3.4.1. We are interested in testing the null hypothesis $H_0 : \beta_x = 0$. Under H_0, we have $\beta_{c,x} = 0$, and thus (3.10) is satisfied. Therefore, the corrected score test and the corrected Wald test are valid.

Scenario $3'$: Suppose $X_i \perp\!\!\!\perp V_i$, $C_i \perp\!\!\!\perp X_i$, and the error mechanism is noninformative. The underlying error model is unspecified. We are interested in testing the null hypothesis $H_0 : \beta_x = 0$. Under H_0, we have $\beta_{c,x} = 0$, and thus (3.10) is satisfied.

Scenario $4'$: Consider the setting in Scenario $1'$, except that we now are interested in testing the null hypothesis that the failure time T_i does not depend on V_i. That is

$$H_0^* : T_i \perp\!\!\!\perp V_i.$$

Suppose that both the Cox model and the error model are misspecified. Then under the null hypothesis H_0^*,

$$\begin{aligned} T_{S,c} &\overset{d}{\to} \chi^2_{r^+}, \\ \text{and } \; T_{W,c} &\overset{d}{\to} \chi^2_{r^+}, \end{aligned}$$

as $n \to \infty$. That is the corrected score test and the corrected Wald test are asymptotically valid.

Now we justify the claims in Scenarios $1' - 4'$. Define $\tilde{U}(\beta)$ as in Appendix A6. Let $\tilde{U}_c(\beta) = (\tilde{U}_c^{+T}(\beta), \tilde{U}_c^{-T}(\beta))^T$, where $\tilde{U}_c^{+}(\beta)$ and $\tilde{U}_c^{-}(\beta)$ correspond to β^+ and β^-, respectively. We need to verify that for Scenarios $1'$ to $3'$, (3.10) is satisfied.

First, we consider Scenario $1'$. In this scenario $\beta_v = \beta^+$, and under H_0, $\beta_{0,v} = \beta_0^+ = 0$.

Note that

$$
\begin{aligned}
& E\left\{Y_i(t)V_i \exp\left(X_i^T\beta_x\right)\right\} \\
= \; & E_{X_i,V_i}[E_{T_i,C_i|X_i,V_i}\left\{I(T_i \ge t)I(C_i \ge t)V_i \exp\left(X_i^T\beta_x\right)\right\}] \\
= \; & E_{X_i,V_i}[E_{T_i|X_i,V_i}\left\{I(T_i \ge t)\right\} E_{C_i|X_i,V_i}\left\{I(C_i \ge t)\right\} V_i \exp\left(X_i^T\beta_x\right)] \\
= \; & E_{X_i,V_i}[E_{T_i|X_i}\left\{I(T_i \ge t)\right\} \exp\{a(t,X_i)\} \exp\{b(t,V_i)\} V_i \exp\left(X_i^T\beta_x\right)] \\
= \; & E_{X_i}[E_{T_i|X_i}\left\{I(T_i \ge t)\right\} \exp\{a(t,X_i)\} \exp\left(X_i^T\beta_x\right)] E_{V_i}[\exp\{b(t,V_i)\}V_i].
\end{aligned}
$$

We can obtain similar expressions for $E\left\{Y_i(t)\exp\left(X_i^T\beta_x\right)\right\}$, $E\left\{Y_i(t)V_i\exp\left(W_i^T\beta_x\right)\right\}$, and $E\left\{Y_i(t)\exp\left(W_i^T\beta_x\right)\right\}$, respectively.

Therefore,

$$
\begin{aligned}
& \tilde{U}_c^+(0,\beta_{c,x}) \\
= \; & \int_0^\tau E\left\{Y_i(t)V_i\exp\left(X_i^T\beta_{0,x}\right)\right\}\lambda_0(t)dt \\
& -\int_0^\tau \frac{E\left\{Y_i(t)V_i\exp\left(W_i^T\beta_{c,x}\right)\right\}}{E\left\{Y_i(t)\exp\left(W_i^T\beta_{c,x}\right)\right\}} E\left\{Y_i(t)\lambda_0(t)\exp\left(X_i^T\beta_{0,x}\right)\right\}dt \\
= \; & \int_0^\tau E_{X_i}[E_{T_i|X_i}\left\{I(T_i \ge t)\right\}\exp\{a(t,X_i)\}\exp\left(X_i^T\beta_{0,x}\right)]E_{V_i}[\exp\{b(t,V_i)\}V_i]\lambda_0(t)dt \\
& -\int_0^\tau \frac{E_{X_i}[E_{T_i|X_i}\left\{I(T_i \ge t)\right\}\exp\{a(t,X_i)\}\exp\left(W_i^T\beta_{c,x}\right)]E_{V_i}[\exp\{b(t,V_i)\}V_i]}{E_{X_i}[E_{T_i|X_i}\left\{I(T_i \ge t)\right\}\exp\{a(t,X_i)\}\exp\left(W_i^T\beta_{c,x}\right)]E_{V_i}[\exp\{b(t,V_i)\}]} \\
\times \; & E_{X_i}[E_{T_i|X_i}\left\{I(T_i \ge t)\right\}\exp\{a(t,X_i)\}\exp\left(X_i^T\beta_{0,x}\right)]E_{V_i}[\exp\{b(t,V_i)\}]\lambda_0(t)dt \\
= \; & 0.
\end{aligned}
$$

Therefore, $\beta_{c,v} = 0$.

We now consider Scenario $2'$. In this scenario, $\beta^+ = \beta_x$, $D^+(\beta_x) = \Sigma_0\beta_x$, and $\beta_0^+ = 0$. We first assume the underlying error model is (3.7). We obtain from (3.8) that under H_0, we have $\beta_{c,x} = 0 = \beta_0^+$, and thus (3.10) is satisfied.

Next, we assume the underlying error model is (3.12) or (3.13). By the uniqueness assumption of the solution of $\tilde{U}_c(\beta) = 0$ and the derivation in Appendix A6, we obtain that when $\beta_{0,x} = 0$, then $\beta_{c,x} = 0$. Thus (3.10) is satisfied.

Now, we assume the underlying error model is (3.14). We have

$$\tilde{U}_c(\beta_x) = \tilde{U}(\beta_x) + \left(\Sigma_0\beta_x - \frac{E\{\epsilon_i^* \exp(\epsilon_i^*\beta_x)\}}{E\{\exp(\epsilon_i^*\beta_x)\}}\right)\int_0^\tau E\left\{Y_i(t)\exp\left(X_i\beta_{0,x}\right)\right\}\lambda_0(t)dt.$$

Therefore, $\tilde{U}_c(0) = 0$. By the uniqueness assumption of the solution of $\tilde{U}_c(\beta) = 0$, we have $\beta_{c,x} = 0$. Thus (3.10) is satisfied.

Next, we consider Scenario $3'$. In this scenario $\beta_x = \beta^+$, and under H_0, $\beta_{0,x} = \beta_0^+ = 0$. Under the assumption that β_c exist and is unique, we only need to verify that $\tilde{U}_c^+(0, \beta_{c,v}) = 0$. Note that under H_0,

$$\frac{E\{Y_i(t)W_i \exp(V_i^T\beta_v)\}}{E\{Y_i(t)\exp(V_i^T\beta_v)\}} = E(W_i),$$

and thus $\tilde{U}_c^+(0, \beta_{c,v}) = 0$. Therefore, $\beta_{c,x} = 0$.

Finally, we consider Scenario $4'$. Let $\lambda_i(t)$ be the underlying hazard function for subject i conditional of the covariates., Note that under H_0, $\lambda_i(t)$ is independent of V_i, but it may be associated with X_i. Similar to Scenario $3'$, We only need to show that

$$\tilde{U}_c^+(0, \beta_{c,x}) = \int_0^\tau E[Y_i(t)V_i\lambda_i(t)]dt - \int_0^\tau \frac{E[Y_i(t)V_i\exp(W_i^T\beta_{c,x})]}{E[Y_i(t)\exp(W_i^T\beta_{c,x})]}E[Y_i(t)\lambda_i(t)]dt = 0,$$

where $\lambda_i(t)$ is the underlying hazard function for subject i conditional of the covariates., Note that under H_0, $\lambda_i(t)$ is independent of V_i, but it may be associated with X_i. Thus, we write $\lambda_i(t; X_i)$ to represent $\lambda_i(t)$.

Note that

$$\begin{aligned}
\tilde{U}_c^+(0, \beta_{c,x}) &= \int_0^\tau E_{X_i}[E_{T_i|X_i}\{I(T_i \geq t)\}\exp\{a(t, X_i)\}\lambda_i(t)]E_{V_i}[\exp\{b(t, V_i)\}V_i]dt \\
&\quad - \int_0^\tau \frac{E_{X_i}[E_{T_i|X_i}\{I(T_i \geq t)\}\exp\{a(t, X_i)\}\exp\left(W_i^T\beta_{c,x}\right)]E_{V_i}[\exp\{b(t, V_i)\}V_i]}{E_{X_i}[E_{T_i|X_i}\{I(T_i \geq t)\}\exp\{a(t, X_i)\}\exp\left(W_i^T\beta_{c,x}\right)]E_{V_i}[\exp\{b(t, V_i)\}]} \\
&\times \quad E_{X_i}[E_{T_i|X_i}\{I(T_i \geq t)\}\exp\{a(t, X_i)\}\lambda_i(t)]E_{V_i}[\exp\{b(t, V_i)\}]dt \\
&= 0.
\end{aligned}$$

Therefore, $\beta_{c,v} = 0$.

Appendix S3

The setting of the survival model and working error model here is the same as in Section 3.5.1. In Figure 3.6, we consider Case S.1 that the true error model is $W_i = X_i + K\epsilon_i$. In Figure 3.6, σ_0 is fixed as 0.25 or 0.50, respectively, but K varies from $1/\sqrt{2}$ to $\sqrt{2}$, indicating that the variance of the error term is misspecified from twice of the true one to a half of the true one. In Figure 3.7, we consider Case S.2 that the true model is $W_i = \gamma_x X_i + \epsilon_i$, where σ is fixed as 0.25 or 0.50, respectively, but γ_x changes from $1/\sqrt{2}$ to $\sqrt{2}$. In Figure 3.8, we consider Case S.3 that the true model is $W_i = \gamma_x X_i + K\epsilon_i$, where $(K, \gamma_x) = (0.8, 0.8), (0.8, 1.2), (1.2, 0.8)$ or $(1.2, 1.2)$, and σ varies from 0 to 0.5. In Figure 3.9, we consider Case S.4 the true model is the Berkson error model $X_i = W_i + \epsilon_i$, and σ varies from 0 to 0.5. In Figure 3.10, we consider Case S.5 that the error distribution $N(0, \sigma_0^2)$ of ϵ_i is misspecified, and the true distribution is respectively the uniform distribution Unif$(-\sqrt{3}\sigma_0, \sqrt{3}\sigma_0)$, exponential distribution Exp$(1/\sigma_0)$, and logistic distribution Logistic$(0, \sqrt{3}\sigma_0/\pi)$. The parameters in these distributions are chosen to satisfy that the error variance is the same as the misspecified normal error variance. We also consider that the true error distribution is the mixture normal distribution $0.5N(0, \sigma_0^2) + 0.5N(0, \sigma_0^2/4)$. In Figure 3.11, We consider Case S.6 that the true error distribution is $pN(0, \sigma_0^2) + (1-p)N(0, 4\sigma_0^2)$ with $\sigma = 0.25$ or $pN(0, \sigma_0^2) + (1-p)N(0, \sigma_0^2/4)$ with $\sigma_0 = 0.50$, where p varies from 0 to 1. All the patterns revealed by the theoretical findings in Section 3.5.1 and Appendix S1.

[Insert Figures 3.6-3.11 here!]

Appendix S4

In Table 3.2, we consider the scenario that the error distribution is misspecified as $N(0, \sigma_0^2)$, whereas the true one is one of the uniform distribution Unif$(-\sqrt{3}\sigma_0, \sqrt{3}\sigma_0)$, exponential distribution Exp$(1/\sigma_0)$, logistic distribution Logistic$(0, \sqrt{3}\sigma_0/\pi)$, or the mixture normal distribution $0.5N(0, \sigma_0^2) + 0.5N(0, 4\sigma_0^2)$ with $\sigma_0 = 0.25$ or $0.5N(0, \sigma_0^2) + 0.5N(0, \sigma_0^2/4)$ with

$\sigma_0 = 0.50$. The data generating process of the true covariates, survival time, and censoring time are the same as those in Table 3.1.

Analogous to the phenomenon of asymptotic biases displayed in Appendix S1, we find in Table 3.2 that the finite sample biases of the estimator $\hat{\beta}_c$ can be even bigger than those of the naive estimator $\hat{\beta}_{nv}$. Furthermore, the coverage rates of 95% confidence intervals produced by the estimator $\hat{\beta}_c$ can be poorer than those of the naive estimator $\hat{\beta}_{nv}$.

[Insert Tables 3.2 here!]

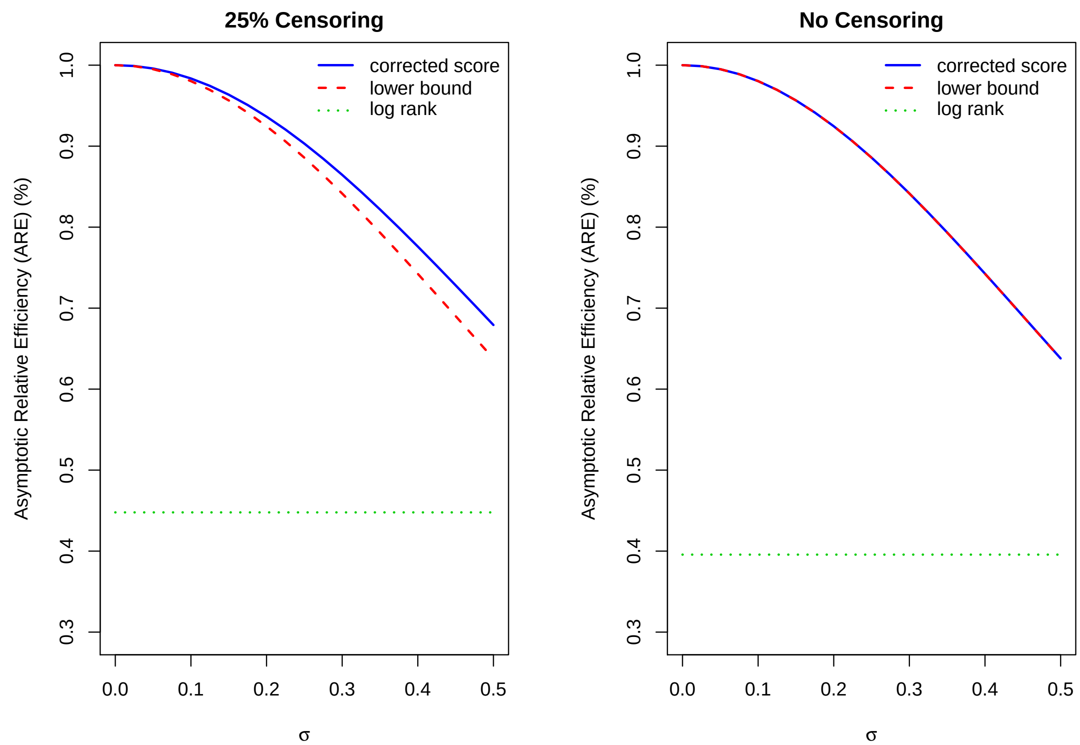


Figure 3.1: ARE of the corrected score and log rank tests compared to the true score test

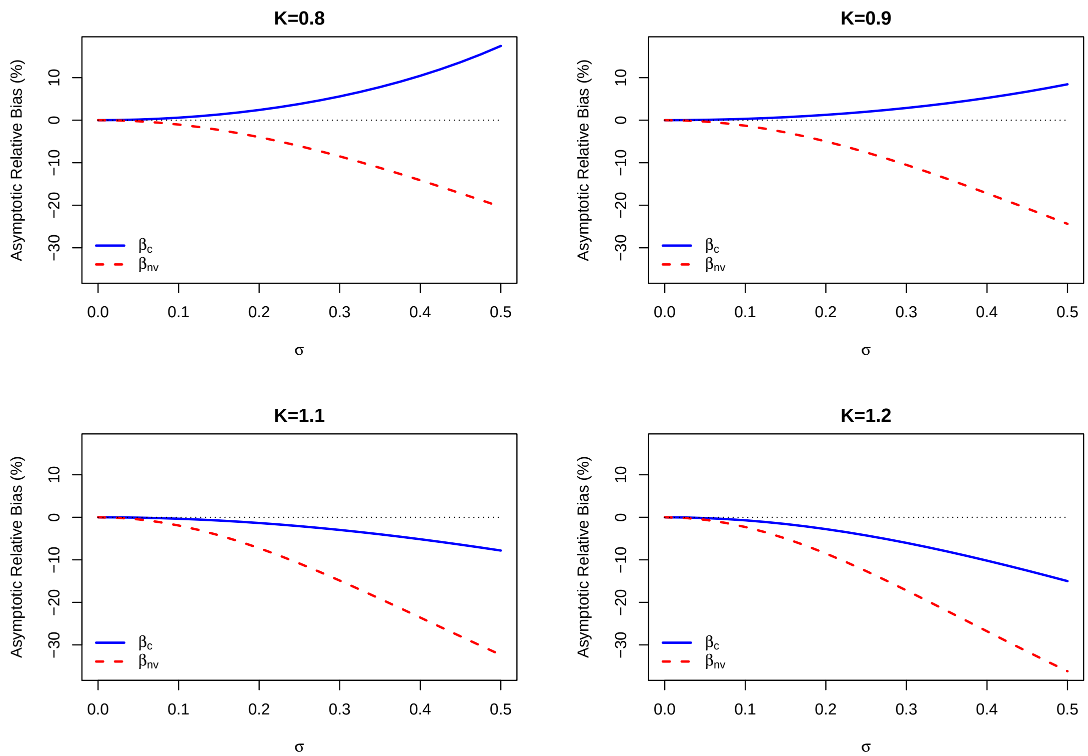


Figure 3.2: Case 1: True error model is $W_i = X_i + K\epsilon_i$.

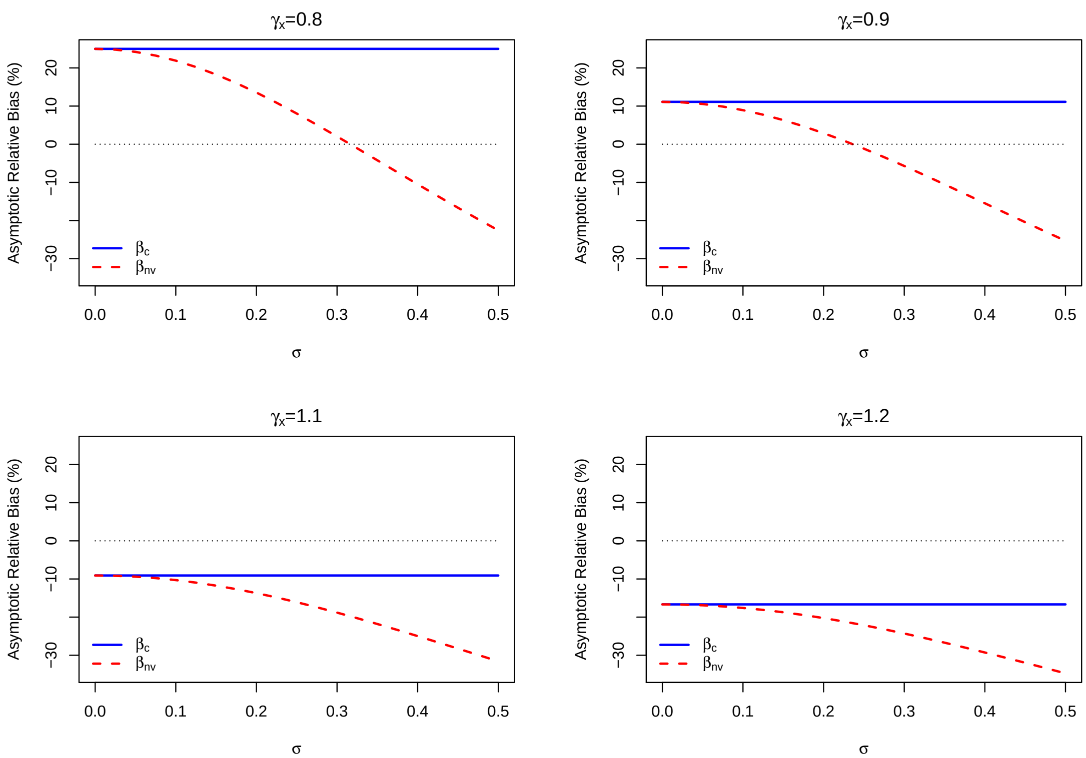


Figure 3.3: Case 2: True error model is $W_i = \gamma_x X_i + \epsilon_i$.

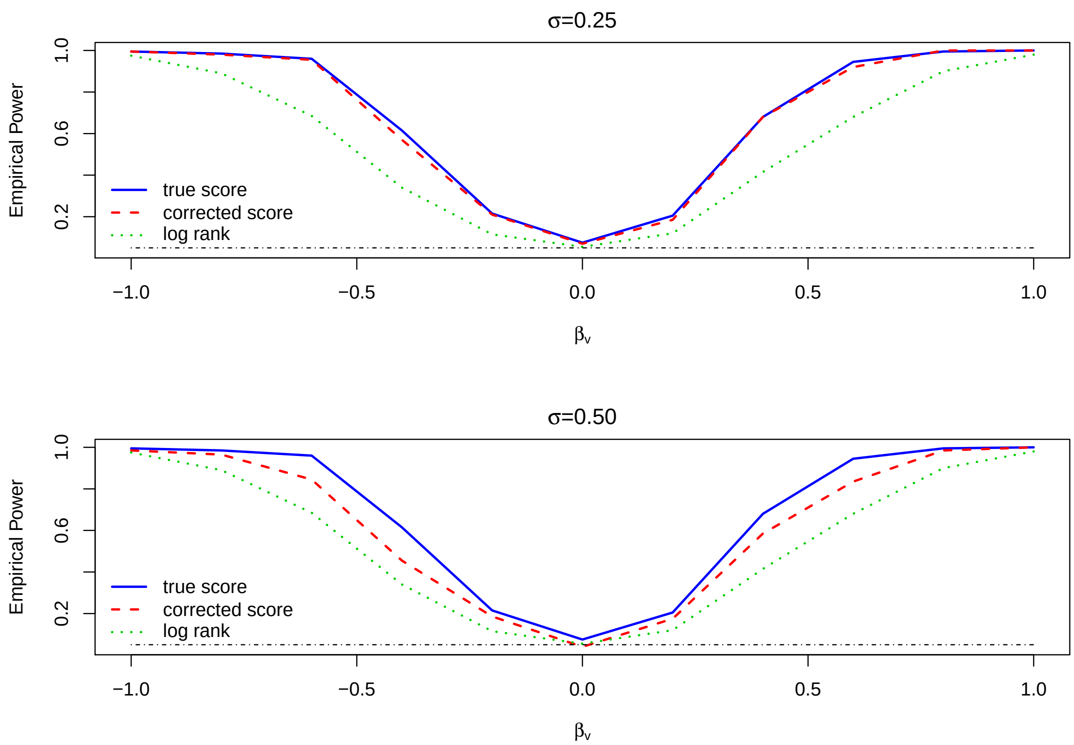


Figure 3.4: Power plot under correctly specified error model

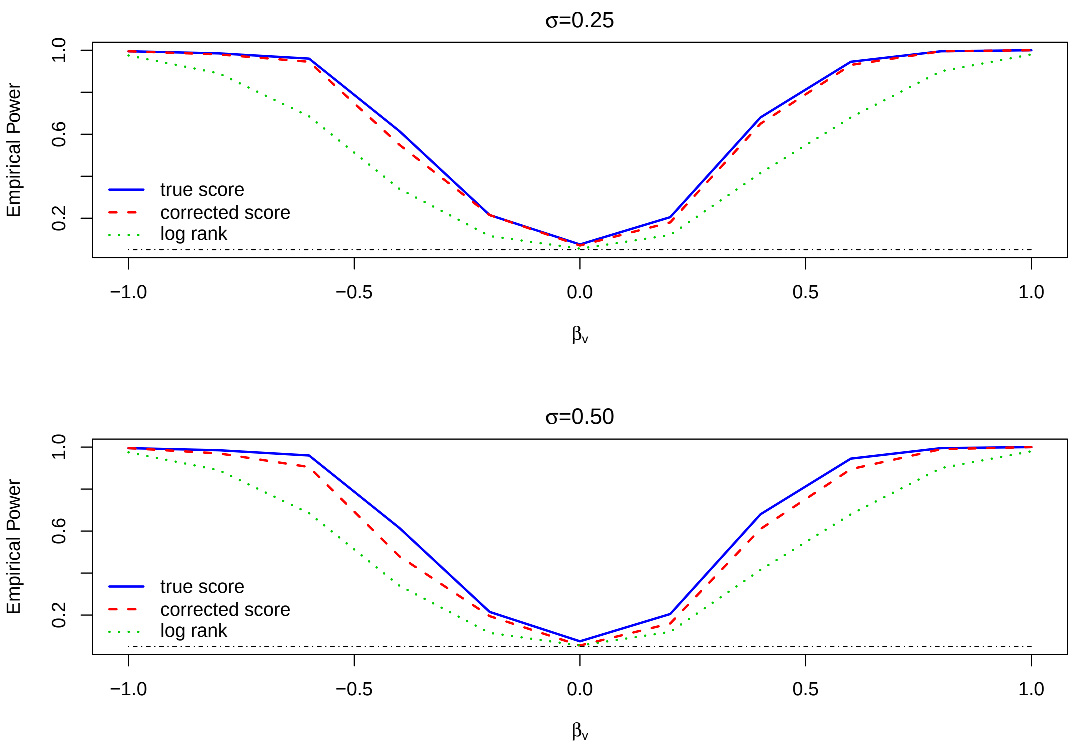


Figure 3.5: Power plot under misspecified error model

Table 3.1: Simulation results with misspecified error variance

Case	K	σ	Method	Estimation of β_x					Estimation of β_v				
				Bias[a]	EVE[b]	MVE[c]	MSE[d]	CP(%)[e]	Bias	EVE	MVE	MSE	CP(%)
Case 1	$1/\sqrt{2}$	0.25	$\hat{\beta}_{nv}$	-0.049	0.015	0.013	0.017	89.8	-0.019	0.013	0.014	0.014	94.1
			$\hat{\beta}_c$	0.084	0.024	0.020	0.031	91.4	0.012	0.015	0.015	0.015	94.6
		0.50	$\hat{\beta}_{nv}$	-0.181	0.013	0.011	0.046	56.6	0.031	0.014	0.014	0.015	93.5
			$\hat{\beta}_c$	0.445	0.106	0.193	0.304	94.9	0.097	0.064	0.049	0.073	93.7
	1.0	0.25	$\hat{\beta}_{nv}$	-0.098	0.014	0.012	0.024	81.2	0.023	0.014	0.014	0.014	93.9
			$\hat{\beta}_c$	0.020	0.022	0.018	0.022	94.1	0.017	0.015	0.015	0.015	94.5
		0.50	$\hat{\beta}_{nv}$	-0.308	0.011	0.009	0.106	15.0	0.045	0.014	0.014	0.016	93.0
			$\hat{\beta}_c$	0.089	0.060	0.054	0.068	96.8	0.035	0.021	0.020	0.022	93.3
	$\sqrt{2}$	0.25	$\hat{\beta}_{nv}$	-0.181	0.013	0.011	0.046	56.6	0.031	0.014	0.014	0.015	93.5
			$\hat{\beta}_c$	-0.086	0.019	0.016	0.026	83.8	0.025	0.015	0.015	0.015	93.8
		0.50	$\hat{\beta}_{nv}$	-0.471	0.008	0.007	0.230	0.1	0.065	0.015	0.014	0.019	90.5
			$\hat{\beta}_c$	-0.278	0.023	0.018	0.100	41.9	0.051	0.017	0.016	0.020	92.2
Case 2	$1/\sqrt{2}$	0.25	$\hat{\beta}_{nv}$	-0.025	0.010	0.010	0.011	93.8	0.018	0.054	0.049	0.054	93.8
			$\hat{\beta}_c$	0.081	0.016	0.014	0.022	90.5	-0.001	0.057	0.052	0.057	93.0
		0.50	$\hat{\beta}_{nv}$	-0.141	0.009	0.009	0.029	64.9	0.044	0.056	0.050	0.058	93.1
			$\hat{\beta}_c$	0.356	0.075	0.093	0.202	80.5	-0.028	0.092	0.079	0.093	92.5
	1.0	0.25	$\hat{\beta}_{nv}$	-0.067	0.010	0.010	0.014	87.9	0.027	0.055	0.049	0.055	93.3
			$\hat{\beta}_c$	0.031	0.015	0.014	0.016	94.4	0.010	0.058	0.052	0.058	92.9
		0.50	$\hat{\beta}_{nv}$	-0.263	0.009	0.009	0.078	19.2	0.069	0.057	0.050	0.062	92.8
			$\hat{\beta}_c$	0.084	0.040	0.032	0.047	96.4	0.013	0.077	0.067	0.077	93.3
	$\sqrt{2}$	0.25	$\hat{\beta}_{nv}$	-0.141	0.009	0.009	0.029	64.9	0.044	0.056	0.050	0.058	93.1
			$\hat{\beta}_c$	-0.057	0.013	0.012	0.016	89.4	0.028	0.059	0.052	0.060	92.8
		0.50	$\hat{\beta}_{nv}$	-0.432	0.008	0.008	0.195	0.4	0.104	0.057	0.050	0.068	91.8
			$\hat{\beta}_c$	-0.237	0.018	0.016	0.074	47.2	0.072	0.069	0.059	0.074	92.3

[a] Bias: finite sample biases;
[b] EVE: empirical variances;
[c] MVE: average of the model-based variance estimates;
[d] MSE: mean square errors;
[e] MCP: model-based coverage probability.

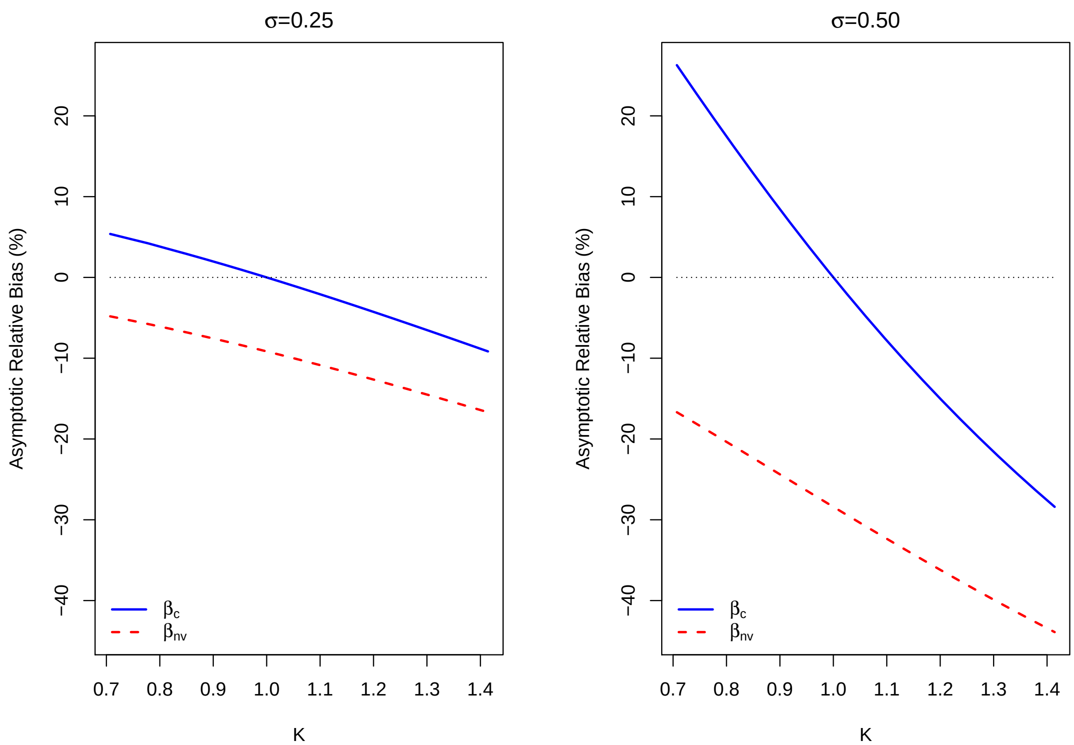


Figure 3.6: Case S.1: True error model is $W_i = X_i + K\epsilon_i$.

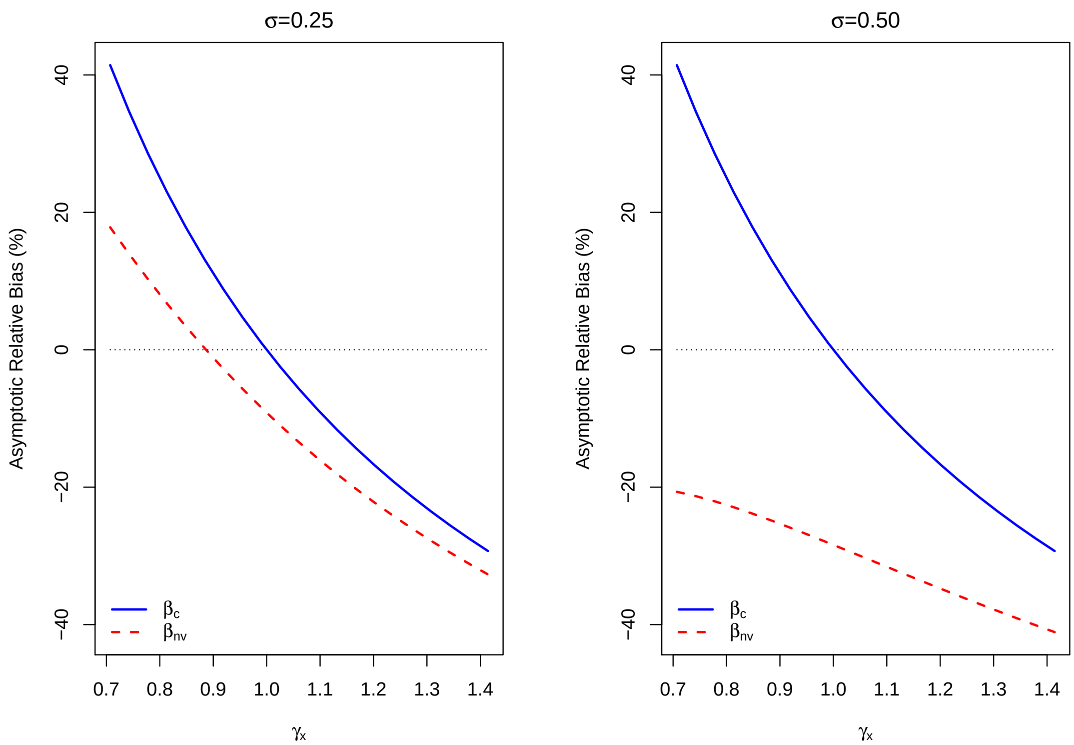


Figure 3.7: Case S.2: True model is $W_i = \gamma_x X_i + \epsilon_i$.

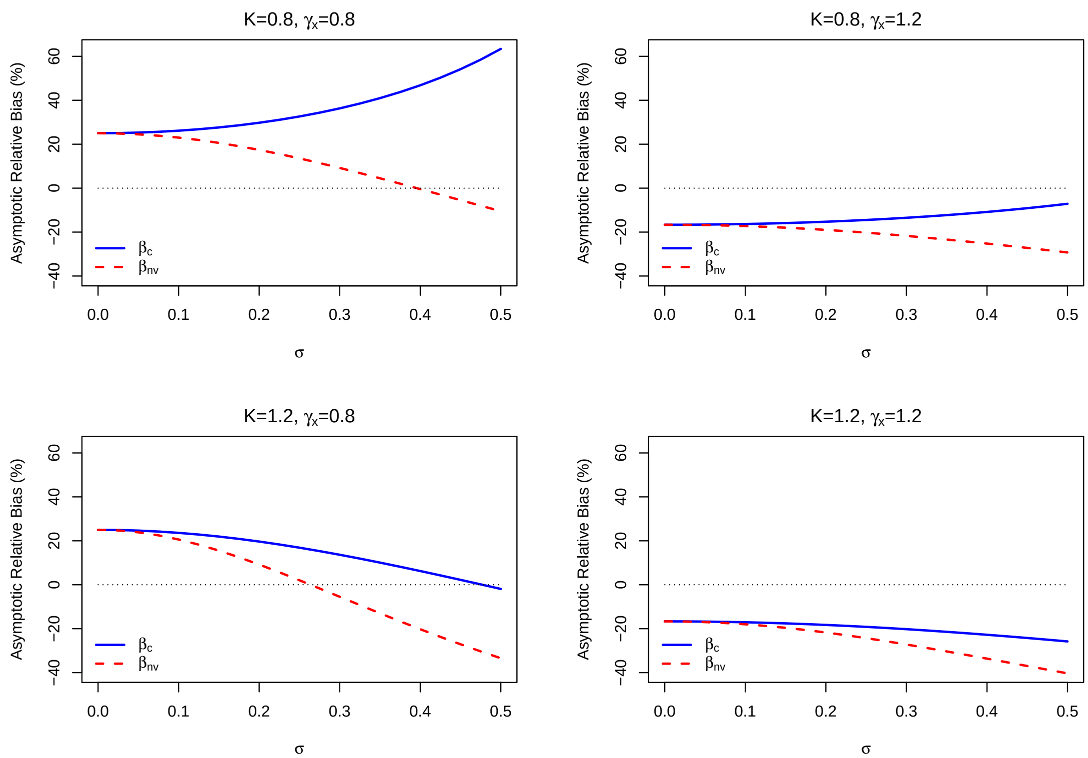


Figure 3.8: Case S.3: True model is $W_i = \gamma_x X_i + K\epsilon_i$.

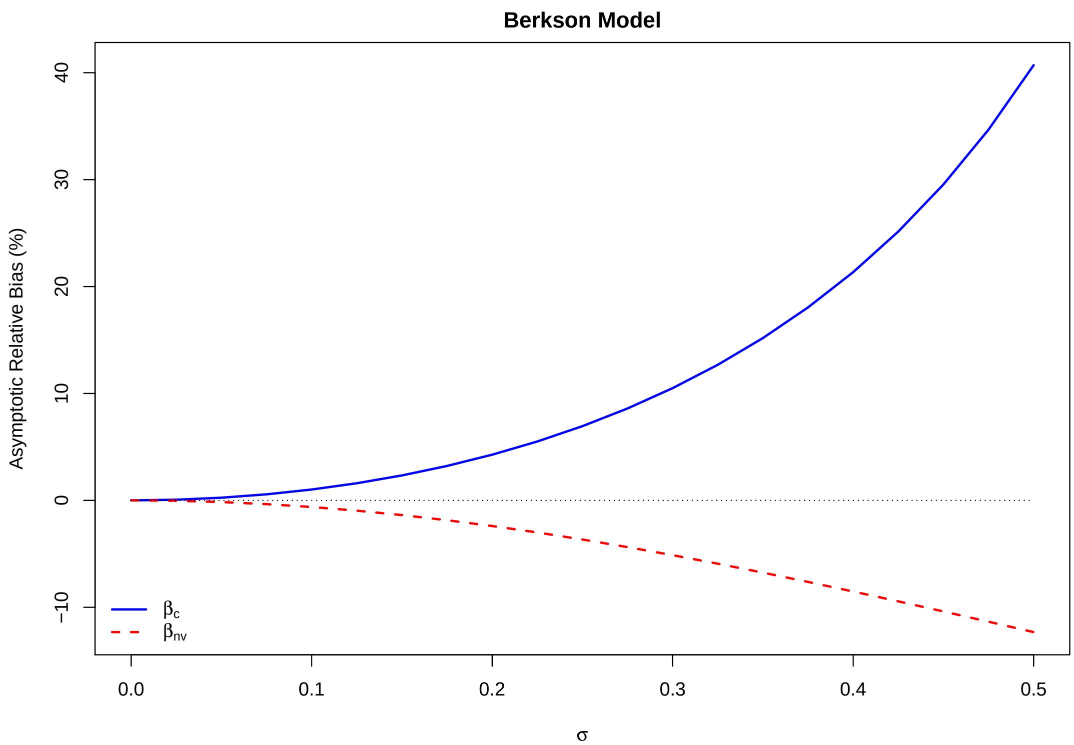


Figure 3.9: Case S.4: True model is the Berkson error model $X_i = W_i + \epsilon_i$.

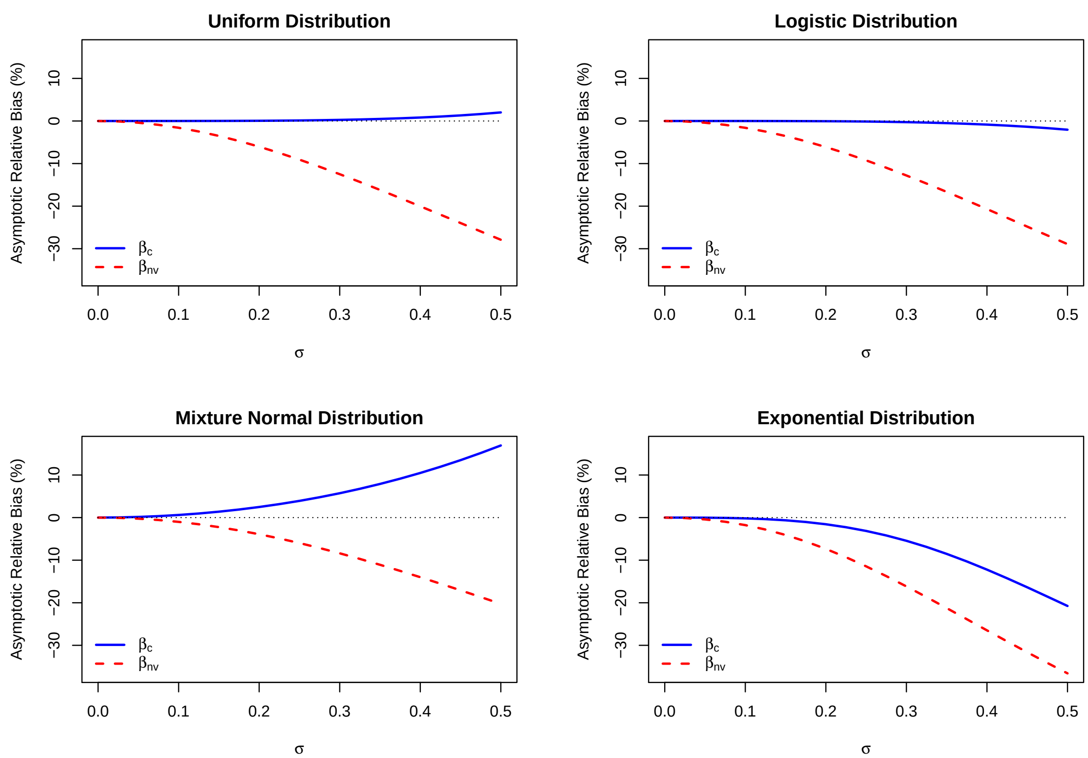


Figure 3.10: Case S.5: The error distribution is misspecified.

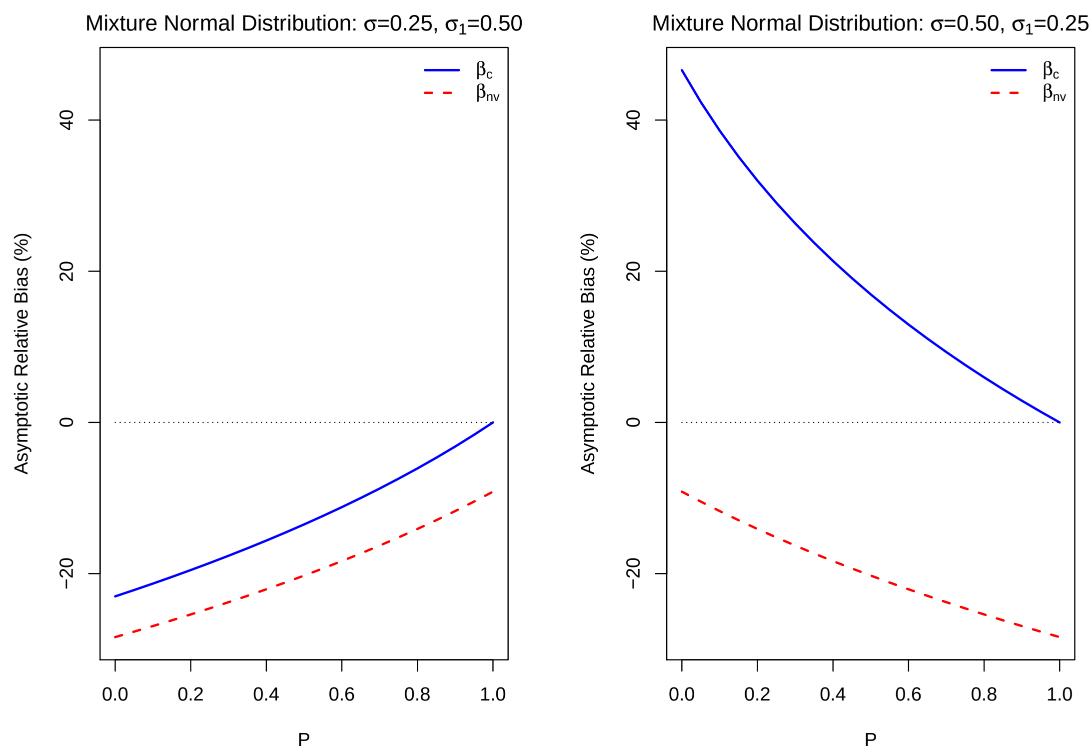


Figure 3.11: Case S.6: The error distribution is misspecified.

Table 3.2: Simulation results for different misspecified error distributions

Case	Distribution	σ	Method	Estimation of β_x					Estimation of β_v				
				Bias	EVE	MVE	MSE	CP(%)	Bias	EVE	MVE	MSE	CP(%)
Case 1	Uniform	0.25	$\hat{\beta}_{nv}$	-0.089	0.014	0.012	0.022	84.8	0.016	0.015	0.014	0.015	93.8
			$\hat{\beta}_c$	0.030	0.021	0.018	0.022	93.1	0.009	0.016	0.015	0.016	94.1
	Uniform	0.50	$\hat{\beta}_{nv}$	-0.298	0.010	0.009	0.099	14.9	0.038	0.016	0.014	0.017	92.7
			$\hat{\beta}_c$	0.114	0.060	0.060	0.073	96.7	0.023	0.031	0.021	0.032	93.2
	Logistic	0.25	$\hat{\beta}_{nv}$	-0.089	0.014	0.012	0.022	84.7	0.016	0.015	0.014	0.015	93.6
			$\hat{\beta}_c$	0.030	0.021	0.018	0.022	93.6	0.009	0.016	0.015	0.016	94.1
	Logistic	0.50	$\hat{\beta}_{nv}$	-0.301	0.011	0.010	0.102	17.2	0.039	0.016	0.014	0.017	92.8
			$\hat{\beta}_c$	0.094	0.064	0.059	0.073	96.8	0.023	0.023	0.020	0.023	92.9
	Mixture Normal	0.25	$\hat{\beta}_{nv}$	-0.211	0.012	0.011	0.057	47.2	0.034	0.015	0.014	0.016	93.5
			$\hat{\beta}_c$	-0.126	0.017	0.015	0.033	77.5	0.029	0.016	0.014	0.017	94.0
	Mixture Normal	0.50	$\hat{\beta}_{nv}$	-0.211	0.012	0.011	0.056	47.4	0.036	0.015	0.014	0.016	93.2
			$\hat{\beta}_c$	0.348	0.103	0.140	0.224	95.9	0.061	0.036	0.033	0.040	94.4
	Exponential	0.25	$\hat{\beta}_{nv}$	-0.115	0.013	0.013	0.026	80.4	0.025	0.015	0.014	0.015	93.8
			$\hat{\beta}_c$	-0.066	0.020	0.018	0.020	93.5	0.019	0.016	0.015	0.016	94.0
	Exponential	0.50	$\hat{\beta}_{nv}$	-0.365	0.013	0.010	0.147	8.4	0.054	0.015	0.014	0.018	91.0
			$\hat{\beta}_c$	-0.096	0.061	0.041	0.070	78.0	0.048	0.021	0.018	0.023	91.5
Case 2	Uniform	0.25	$\hat{\beta}_{nv}$	-0.079	0.010	0.010	0.016	84.7	0.026	0.050	0.049	0.051	94.7
			$\hat{\beta}_c$	0.017	0.015	0.013	0.015	94.3	0.008	0.053	0.052	0.053	94.3
	Uniform	0.50	$\hat{\beta}_{nv}$	-0.269	0.009	0.008	0.081	16.6	0.066	0.051	0.050	0.055	94.1
			$\hat{\beta}_c$	0.079	0.035	0.030	0.041	96.1	0.007	0.068	0.065	0.068	94.4
	Logistic	0.25	$\hat{\beta}_{nv}$	-0.080	0.010	0.010	0.017	85.0	0.026	0.051	0.049	0.051	94.2
			$\hat{\beta}_c$	0.016	0.015	0.013	0.015	94.2	0.008	0.054	0.052	0.054	94.2
	Logistic	0.50	$\hat{\beta}_{nv}$	-0.274	0.010	0.009	0.085	18.1	0.068	0.052	0.050	0.057	94.3
			$\hat{\beta}_c$	0.059	0.040	0.031	0.044	94.1	0.013	0.072	0.065	0.073	94.3
	Mixture Normal	0.25	$\hat{\beta}_{nv}$	-0.188	0.010	0.009	0.045	48.2	0.035	0.053	0.050	0.055	93.8
			$\hat{\beta}_c$	-0.112	0.013	0.012	0.026	77.7	0.021	0.057	0.052	0.057	93.9
	Mixture Normal	0.50	$\hat{\beta}_{nv}$	-0.189	0.010	0.009	0.046	47.9	0.035	0.055	0.050	0.056	93.8
			$\hat{\beta}_c$	0.242	0.059	0.055	0.118	89.2	-0.039	0.090	0.073	0.092	92.4
	Exponential	0.25	$\hat{\beta}_{nv}$	-0.092	0.012	0.010	0.020	80.4	0.033	0.056	0.050	0.057	92.9
			$\hat{\beta}_c$	-0.001	0.016	0.014	0.016	92.4	0.015	0.060	0.053	0.060	92.7
	Exponential	0.50	$\hat{\beta}_{nv}$	-0.335	0.014	0.010	0.126	11.0	0.087	0.058	0.052	0.065	92.4
			$\hat{\beta}_c$	-0.096	0.044	0.024	0.053	77.5	0.051	0.080	0.065	0.082	92.6

Chapter 4

Model Checking for the Cox Model with Measurement Error

4.1 Introduction

In the past fourty years, the Cox model (Cox 1972) has been widely adopted to analyze survival data with survival endpoints and correctly measured covariates. However, in many studies, survival analysis is frequently challenged by mismeasurement of covariates. Examples include the CD4 lymphocyte counts in the AIDS studies (Hammer et al. 1996) and the forced expiratoryexpiratory volume (FEV) in the randomized trial conducted to evaluate the effect of rhDNase (Fuchs et al. 1994). Prentice (1982) showed that simply ignoring measurement error in covariates leads to misleading results. Consequently, researchers proposed numerous methods to handle covariate measurement error, including Nakamura (1992), Hu, Tsiatis and Davidian (1998), Huang and Wang (2000), Hu and Lin (2002, 2004), and Song and Huang (2005), Zucker (2005), and Yi and Lawless (2007). These methods are successful in correcting for measurement error effects, and are commonly adopted in practise. They assumed a classical measurement error model (Carroll et al. 2006) that features the relationship between the underlying correct covariates and their observed surrogate measurements. However, standard model checking techniques (Therneau and Grambsch 2000; Lawless 2003) can not be directly applied to check either the survival model or the error model assumptions. Surprisingly, to the best of our knowledge, there is little work to check the fit of the survival model and the measurement error model.

In this chapter, we aim at developing valid goodness-of-fit tests based on the observed data. These tests can be used to check the overall fit of the survival model and the measurement error model simultaneously. In particular, we consider two commonly used measurement error scenarios, and propose model checking procedures under these two scenarios, respectively.

In Section 4.2, we introduce notations and existing model checking methods in survival data analysis in the absence of covariate error. In Section 4.3, we describe several scenarios of measurement error. In Section 4.4, we propose model checking procedures in the presence of covariate error, and show that they are valid to check the Cox model and the error model. in Section 4.5, we report simulation studies and provide an example. Concluding discussion is provided in the last section.

4.2 Cox Model and Model Checking

For subject i, let T_i be the failure time, C_i be the right censoring time, and Z_i be the vector of covariate, $i = 1, \cdots, n$. We assume that all subjects are under observation over a common time interval $[0, \tau]$, where τ is a positive constant, and the $\{T_i, C_i, Z_i\}$ are mutually independent. Define $S_i = \min(T_i, C_i)$, and $\delta_i = I(T_i \leq C_i)$. For $t \in (0, \tau]$, let $N_i(t) = I(S_i \leq t, \delta_i = 1)$ be the counting process, and $Y_i(t) = I(S_i \geq t)$ be the at-risk indicator. Throughout this article, we assume the conditional independent censoring mechanism, i.e., C_i and T_i are independent given Z_i.

The Cox model (Cox 1972) assumes that the failure time T_i is related to Z_i through the hazard function

$$\lambda(t; Z_i) = \lambda_0(t) \exp(Z_i^T \beta),$$

where $\lambda_0(\cdot)$ is the baseline hazard function, and β is the vector of regression parameters. Here we assume that the distribution of T_i is continuous.

Let β_0 be the true value of the parameter. Let

$$M_i(t;\beta_0,\Lambda_0) = N_i(t) - \int_0^t Y_i(u)\exp(Z_i^T\beta_0)\lambda_0(u)du$$

be the martingale with respective to the filtration $\mathcal{F}_t = \sigma\{N_i(u), Y_i(u), Z_i; 0 \le u \le t, i = 1,\cdots,n\}$. The regression parameter β can be estimated by solving the partial score function

$$U(\beta) = 0,$$

where

$$U(\beta) = \sum_{i=1}^n \int_0^\tau \left\{ Z_i - \frac{\sum_{j=1}^n Y_j(t) Z_j \exp(Z_j^T\beta)}{\sum_{j=1}^n Y_j(t)\exp(Z_j^T\beta)} \right\} dN_i(t).$$

Let $\tilde{\beta}$ denote the resulting estimator of β. The cumulative hazard function $\Lambda_0(t)$ can be estimated by the Breslow estimator:

$$\tilde{\Lambda}_0(t) = \sum_{i=1}^n \int_0^t \frac{dN_i(u)}{\sum_{i=1}^n Y_i(u)\exp(Z_i^T\tilde{\beta})}.$$

Let

$$M_i(t;\tilde{\beta},\tilde{\Lambda}_0) = N_i(t) - \int_0^t Y_i(u)\exp(Z_i^T\tilde{\beta})d\tilde{\Lambda}_0(u) \tag{4.1}$$

be the martingale residual that represents the difference of observed number and expected number of events for the ith subject. Martingale residuals are demonstrated to be informative about model misspecification and they have been served as building blocks of constructing model checking procedures for the Cox model (Barlow and Prentice 1988; Therneau, Grambsch and Fleming 1990; Lin, Wei and Ying 1993; Spierkerman and Lin 1998; Lin et al. 2000). Let

$$W(t,z) = \sum_{i=1}^n I(Z_i \le z) M_i(t;\tilde{\beta},\tilde{\Lambda}_0), \tag{4.2}$$

where $I(\cdot)$ is the indicator function, and $Z_i \leq z$ means that every component of Z_i is no larger than the corresponding component of z. Under the null hypothesis that the Cox model is correctly specified, Lin, Wei and Ying (1993) showed that $n^{-1/2}W(t,z)$ converges weakly to a mean-zero Gaussian process as $n \to \infty$. Consequently, they proposed an omnibus goodness-of-fit test statistic

$$\sup_{t,z} |n^{-1/2}W(t,z)|. \tag{4.3}$$

Lin, Wei and Ying (1993) proposed a testing procedure based on $\sup_{t,z} |n^{-1/2}W(t,z)|$ for testing the null hypothesis H_0: the Cox model is correctly specified. For a pre-specified size α, we reject H_0 if $\sup_{t,z} |n^{-1/2}W(t,z)| > w_\alpha$, where w_α is the estimated upper α-quantile of the distribution of $\sup_{t,z} |n^{-1/2}W(t,z)|$.

4.3 Measurement Error Models and Estimation

We consider the situation where some covariates are subject to measurement error. Write $Z_i = (X_i^T, V_i^T)^T$, where X_i is a $p \times 1$ vector of error-prone covariates, and V_i is a $q \times 1$ vector of precisely measured covariates. Let W_i be a surrogate measurement of X_i. Write $\beta = (\beta_x^T, \beta_v^T)^T$ so that β_x and β_v correspond to X_i and V_i, respectively. Let $\beta_0 = (\beta_{0x}^T, \beta_{0v}^T)^T$, where $\beta_{0,x}$ and $\beta_{0,v}$ are the parameters corresponding to X_i and V_i, respectively. We consider two scenarios of the measurement error process. Both of them have been widely adopted in the literature (Carroll et al. 2006).

4.3.1 Scenario 1 : Additive Error Model with Replicates

First, we describe a situation where X_i is repeatedly measured n_i times by the surrogates $W_{ir}(r = 1, \cdots, n_i)$:

$$W_{ir} = X_i + \epsilon_{ir}, \quad i = 1, \cdots, n; \quad r = 1, \cdots, n_i, \tag{4.4}$$

where the error terms ϵ_{ir} are independent and identically distributed with mean 0 and an unknown covariance matrix Σ_0, and are independent of $N_i(\cdot)$, $Y_i(\cdot)$, and Z_i. The distribution of ϵ_{ir} is left unspecified. Assume that $n_i > 1, i = 1, \cdots, n$. With replicates W_{ir}, a consistent estimate of Σ_0 is given by

$$\hat{\Sigma}_0 = \frac{\sum_{i=1}^n \sum_{r=1}^{n_i} (W_{ir} - \bar{W}_{i\cdot})^{\otimes 2}}{\sum_{i=1}^n (n_i - 1)},$$

where $a^{\otimes 2} = aa^T$ for a column vector a, and $\bar{W}_{i\cdot} = \sum_{r=1}^{n_i} W_{ir}/n_i$. Let $\hat{Z}_{ir} = (W_{ir}^T, V_i^T)^T$ and $\hat{Z}_i = (\bar{W}_{i\cdot}^T, V_i^T)^T$ for Scenario 1.

Under this circumstance of the measurement error process, estimation of β can be proceeded as follows. Let $S_{nc}^{(1)}(\hat{Z}; \beta, t) = n^{-1} \sum_{i=1}^n Y_i(t) \{n_i^{-1}(n_i - 1)^{-1} \sum_{r \neq s} \hat{Z}_{ir} \exp(\hat{Z}_{is}^T \beta)\}$, and $S_{nc}^{(0)}(\hat{Z}; \beta, t) = n^{-1} \sum_{i=1}^n Y_i(t) \{n_i^{-1} \sum_{r=1}^{n_i} \exp(\hat{Z}_{ir}^T \beta)\}$. Define the nonparametric correction function (Huang and Wang 2000) as

$$U_{nc}(\beta) = \sum_{i=1}^n \int_0^\tau \left\{ \hat{Z}_i - \frac{S_{nc}^{(1)}(\hat{Z}; \beta, t)}{S_{nc}^{(0)}(\hat{Z}; \beta, t)} \right\} dN_i(t),$$

Then solving $U_{nc}(\beta) = 0$ gives the nonparametric correction estimator (Huang and Wang 2000) of β; let $\hat{\beta}_{nc}$ denote such an estimator.

Define $S_{nc}^{(2)}(\hat{Z}; \beta, t) = n^{-1} \sum_{i=1}^n Y_i(t) \{n_i^{-1}(n_i - 1)^{-1} \sum_{r \neq s} \hat{Z}_{ir} \hat{Z}_{is}^T \exp(\hat{Z}_{is}^T \beta)\}$. Let $s_{nc}^{(k)}(\hat{Z}; \beta, t) = E\{S_{nc}^{(k)}(\hat{Z}; \beta, t)\}$, where $k = 0, 1, 2$. Let

$$\mathcal{I}_{nc}(\beta) = \int_0^\tau \left[\frac{s_{nc}^{(2)}(\hat{Z}; \beta, t)}{s_{nc}^{(0)}(\hat{Z}; \beta, t)} - \left\{ \frac{s_{nc}^{(1)}(\hat{Z}; \beta, t)}{s_{nc}^{(0)}(\hat{Z}; \beta, t)} \right\}^{\otimes 2} \right] dE\{N_i(t)\},$$

and

$$J_{nc,i}(\beta) = \int_0^\tau \left\{ \hat{Z}_i - \frac{s_{nc}^{(1)}(\hat{Z}; \beta, t)}{s_{nc}^{(0)}(\hat{Z}; \beta, t)} \right\} dN_i(t) - \int_0^\tau \left[\frac{Y_i(t) \{n_i^{-1}(n_i - 1)^{-1} \sum_{r \neq s} \hat{Z}_{ir} \exp(\hat{Z}_{is}^T \beta)\}}{s_{nc}^{(0)}(\hat{Z}; \beta, t)} - \frac{Y_i(t) \{n_i^{-1} \sum_{r=1}^{n_i} \exp(\hat{Z}_{ir}^T \beta)\} s_{nc}^{(1)}(\hat{Z}; \beta, t)}{\{s_{nc}^{(0)}(\hat{Z}; \beta, t)\}^2} \right] dE\{N_i(t)\}.$$

Define $\mathcal{J}_{nc}(\beta) = \lim_{n \to \infty} n^{-1} \sum_{i=1}^n E[\{J_{nc,i}(\beta)\}^{\otimes 2}]$. Huang and Wang (2000) showed that $\sqrt{n}(\hat{\beta}_{nc} - \beta_0)$ is asymptotically normal, with mean zero and covariance matrix $\mathcal{I}_{nc}^{-1}(\beta_0) \mathcal{J}_{nc}(\beta_0) \mathcal{I}_{nc}^{-1}(\beta_0)$.

4.3.2 Scenario 2 : Additive Error Model with Known Parameters

In Scenario 2, the measurement error model is given by

$$W_i = X_i + \epsilon_i, \quad i = 1, \cdots, n, \tag{4.5}$$

where the error terms $\epsilon_i, i = 1, \cdots, n$ are independent and identically distributed and are independent of $N_i(\cdot)$, $Y_i(\cdot)$, and Z_i, and $\epsilon_i \sim N(0, \Sigma_0)$ with the covariance matrix Σ_0 assumed known or consistently estimated for a priori study. Let $D(\beta_x) = \Sigma_0\beta_x$, and $\Sigma_1 = \text{diag}(\Sigma_0, 0)$ be a $(p+q) \times (p+q)$ matrix. Let $\hat{Z}_i = (W_i^T, V_i^T)^T$ denote the observed covariates for Scenario 2.

Under this circumstance of the measurement error process, estimation of β can be proceeded as follows. Let $S_c^{(k)}(\hat{Z};\beta,t) = n^{-1}\sum_{i=1}^n Y_i(t)\hat{Z}_i^{\otimes k}\exp(\hat{Z}_i^T\beta)$, and let $s_c^{(k)}(\hat{Z};\beta,t) = E\{S_c^{(k)}(\hat{Z};\beta,t)\}$, where $k = 0, 1, 2$. Define the corrected score function (Nakamura 1992) as

$$U_c(\beta) = \sum_{i=1}^n \int_0^\tau \left\{ \hat{Z}_i - \frac{S_c^{(1)}(\hat{Z};\beta,t)}{S_c^{(0)}(\hat{Z};\beta,t)} + \Sigma_1\beta \right\} dN_i(t).$$

Then solving $U_c(\beta) = 0$ gives the corrected score estimator (Nakamura 1992) of β; let $\hat{\beta}_c$ denote such an estimator.

Let

$$\mathcal{I}_c(\beta) = \int_0^\tau \left[\frac{s_c^{(2)}(\hat{Z};\beta,t)}{s_c^{(0)}(\hat{Z};\beta,t)} - \left\{ \frac{s_c^{(1)}(\hat{Z};\beta,t)}{s_c^{(0)}(\hat{Z};\beta,t)} \right\}^{\otimes 2} - \Sigma_1 \right] dE\{N_i(t)\},$$

$$\text{and} \quad J_{c,i}(\beta) = \int_0^\tau \left\{ \hat{Z}_i - \frac{s_c^{(1)}(\hat{Z};\beta,t)}{s_c^{(0)}(\hat{Z};\beta,t)} + \Sigma_1\beta \right\} dN_i(t) - \int_0^\tau \frac{Y_i(t)\exp(\hat{Z}_i^T\beta)}{s_c^{(0)}(\hat{Z};\beta,t)} \left\{ \hat{Z}_i - \frac{s_c^{(1)}(\hat{Z};\beta,t)}{s_c^{(0)}(\hat{Z};\beta,t)} \right\} dE\{N_i(t)\}.$$

Define $\mathcal{J}_c(\beta) = E[\{J_{c,i}(\beta)\}^{\otimes 2}]$. Kong and Gu (1999) showed that $\sqrt{n}(\hat{\beta}_c - \beta_0)$ is asymptotically normal, with mean zero and covariance matrix $\mathcal{I}_c^{-1}(\beta_0)\mathcal{J}_c(\beta_0)\mathcal{I}_c^{-1}(\beta_0)$.

4.4 Model Checking Procedures with Error in Covariates

In the presence of covariate measurement error, the martingale residual $M_i(t;\hat{\beta},\hat{\Lambda}_0)$ calculated from (4.1) is no longer available. Thus, the goodness-of-fit test based on the test statistic (4.3) cannot be applied to check the validity of the Cox model. In this section we develop a model checking procedure to assess whether or not a given model is the Cox model in the presence of covariate measurement error.

Let

$$W(t,z;\beta)=\sum_{i=1}^{n}\int_0^t\left\{I(Z_i\le z)-\frac{\sum_{j=1}^{n}Y_j(u)I(Z_j\le z)\exp(Z_j^T\beta)}{\sum_{j=1}^{n}Y_j(u)\exp(Z_j^T\beta)}\right\}dN_i(u).$$

Note that the process $W(t,z)$ defined in (4.2) is equivalent to $W(t,z;\tilde{\beta})$.

The basic idea is to construct a function, say $\hat{W}(t,z;\beta)$, of parameter β, time t, and covariate value z, based on the surrogate measurements $\hat{Z}_i$, such that

$$\sup_{\beta\in\mathcal{B},t\in[0,\tau],z\in\mathcal{Z}}\left\{n^{-1}|\hat{W}(t,z;\beta)-W(t,z;\beta)|\right\}\xrightarrow{a.s.}0, \tag{4.6}$$

as $n\to\infty$, where $\mathcal{B}$ is the parameter space, and $\mathcal{Z}$ is the covariate space. In the following two subsections, we describe methods of constructing $\hat{W}(t,z;\beta)$ for Scenarios 1 and 2, respectively. Let $\hat{\beta}$ represent a $\sqrt{n}$-consistent estimator of β based on the observed data. Using the property (4.6), we show that the processes $n^{-1/2}\hat{W}(t,z;\hat{\beta})$ under Scenarios 1 and 2 converge weakly to a mean-zero Gaussian process asymptotically. These results provide the basis of developing goodness-of-fit test procedures.

4.4.1 Model Checking under Measurement Error Scenario 1

We define the stochastic process

$$\hat{W}_{nc}(t,z;\beta)=\sum_{i=1}^{n}\int_0^t\left\{n_i^{-1}\sum_{r=1}^{n_i}I(\hat{Z}_{ir}\le z)-\frac{V_{nc}^{(1)}(\hat{Z};\beta,u,z)}{S_{nc}^{(0)}(\hat{Z};\beta,u)}\right\}dN_i(u),$$

where $V_{nc}^{(1)}(\hat{Z};\beta,u,z) = n^{-1}\sum_{i=1}^{n} Y_i(u)\{n_i^{-1}(n_i-1)^{-1}\sum_{r\neq s} I(\hat{Z}_{ir} \le z)\exp(\hat{Z}_{is}^T\beta)\}$. Let $V_{nc}^{(2)}(\hat{Z};\beta,u,z) = n^{-1}\sum_{i=1}^{n} Y_i(u)\{n_i^{-1}(n_i-1)^{-1}\sum_{r\neq s} I(\hat{Z}_{ir} \le z)\hat{Z}_{is}\exp(\hat{Z}_{is}^T\beta)\}$. Define $v_{nc}^{(k)}(\hat{Z};\beta,u,z) = E\{V_{nc}^{(k)}(\hat{Z};\beta,u,z)\}, k=1,2$.

In Appendix A1, we verify that $\hat{W}_{nc}(t,z;\beta)$ satisfies the property (4.6). Furthermore, $n^{-1/2}\hat{W}_{nc}(t,z;\hat{\beta}_{nc})$ is asymptotically equivalent to $n^{-1/2}\tilde{W}_{nc}(t,z)$, where

$$\begin{aligned}
\tilde{W}_{nc}(t,z) = &\sum_{i=1}^{n}\Bigg[\int_0^t\Bigg\{n_i^{-1}\sum_{r=1}^{n_i} I(\hat{Z}_{ir}\le z) - \frac{v_{nc}^{(1)}(\hat{Z};\beta_0,u,z)}{s_{nc}^{(0)}(\hat{Z};\beta_0,u)}\Bigg\}dN_i(u)\\
&-\int_0^t\Bigg[\frac{Y_i(u)\{n_i^{-1}(n_i-1)^{-1}\sum_{r\neq s} I(\hat{Z}_{ir}\le z)\exp(\hat{Z}_{is}^T\beta_0)\}}{s_{nc}^{(0)}(\hat{Z};\beta_0,u)}\\
&-\frac{Y_i(u)\{n_i^{-1}\sum_{r=1}^{n_i}\exp(\hat{Z}_{ir}^T\beta_0)\}v_{nc}^{(1)}(\hat{Z};\beta_0,u,z)}{\{s_{nc}^{(0)}(\hat{Z};\beta_0,u)\}^2}\Bigg]dE\{N_i(u)\}\\
&-\mathcal{H}_{nc}(t,z;\beta_0)\mathcal{I}_{nc}^{-1}(\beta_0)J_{nc,i}(\beta_0)\Bigg]\\
\equiv &\sum_{i=1}^{n}\tilde{W}_{nc,i}(t,z),
\end{aligned}$$

and

$$\mathcal{H}_{nc}(t,z;\beta) = \int_0^t\left[\frac{v_{nc}^{(2)T}(\hat{Z};\beta,u,z)}{s_{nc}^{(0)}(\hat{Z};\beta,u)} - \frac{v_{nc}^{(1)}(\hat{Z};\beta,u,z)s_{nc}^{(1)T}(\hat{Z};\beta,u)}{\{s_{nc}^{(0)}(\hat{Z};\beta,u)\}^2}\right]dE\{N_i(u)\}.$$

The proof is deferred to Appendix A2.

Note that $\tilde{W}_{nc}(t,z)$ is a sum of zero-mean independent random variables $\tilde{W}_{nc,i}(t,z)$. In Appendix A3, we establish the weak convergence property of $n^{-1/2}\tilde{W}_{nc}(t,z)$ and $n^{-1/2}\hat{W}_{nc}(t,z;\hat{\beta}_{nc})$, summarized in the following theorem.

Theorem 1 *Under Regularity Conditions, we have*

$$n^{-1/2}\tilde{W}_{nc}(t,z) \rightsquigarrow \mathcal{G}_{nc}(t,z) \quad in \quad l^\infty([0,\tau]\times\mathcal{R}^{(p+q)}) \quad as \quad n\to\infty,$$

where $\rightsquigarrow$ means weak convergence, $l^\infty[0,\tau]$ is the space of all bounded functions on $[0,\tau]$ (van der Vaart and Wellner 1996), and $\mathcal{G}_{nc}(t,z)$ is a zero-mean Gaussian process with *covariance function*

$$\Phi_{nc}(t_1, t_2, z_1, z_2) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} E[\tilde{W}_{nc,i}(t_1, z_1)\tilde{W}_{nc,i}(t_2, z_2)]$$

for time points t_1 *and* t_2 *and real values* z_1 *and* z_2. *Furthermore,*

$$n^{-1/2}\hat{W}_{nc}(t, z; \hat{\beta}_{nc}) \rightsquigarrow \mathcal{G}_{nc}(t, z) \quad in \quad l^{\infty}([0, \tau] \times \mathcal{R}^{(p+q)}) \quad as \quad n \to \infty.$$

Theorem 1 provides a basis for the subsequent development of goodness-of-fit test. It says that if the Cox model and the additive error model are both correctly specified, $n^{-1/2}\hat{W}_{nc}(t, z; \hat{\beta}_{nc})$ would fluctuate around zero randomly provided regularity conditions hold. This motivates us to propose an overall goodness-of-fit test statistic

$$S_{nc} = n^{-1/2} \sup_{t,z} |\hat{W}_{nc}(t, z; \hat{\beta}_{nc})|.$$

An abnormally large value of S_{nc} indicates that the Cox model and/or the error model are incorrectly specified. In Appendix A5, we investigate the consistency properties of the proposed test based on S_{nc}.

However, due to the complex structure of $\mathcal{G}_{nc}(t, z)$, the limiting distributions of S_{nc} and $n^{-1/2}\hat{W}_{nc}(t, z; \hat{\beta}_{nc})$ are difficult to evaluate. Thus, the p-value of S_{nc} is difficult to obtain. To overcome this difficulty, we adopt a resampling procedure. Let $\tilde{W}^{S}_{nc,i}(t, z)$ be the estimated version of $\tilde{W}_{nc,i}(t, z)$, where $s^{(k)}_{nc}(\hat{Z}; \beta_0, t)$ is replaced by $S^{(k)}_{nc}(\hat{Z}; \hat{\beta}_{nc}, t), k = 0, 1, 2,$ $v^{(k)}_{nc}(\hat{Z}; \beta_0, t)$ is replaced by $V^{(k)}_{nc}(\hat{Z}; \hat{\beta}_{nc}, t), k = 1, 2,$ $E\{N_i(t)\}$ is replaced by $n^{-1}\sum_{i=1}^{n} N_i(t)$, and β_0 is replaced by $\hat{\beta}_{nc}$. Let

$$\tilde{W}^{S}_{nc}(t, z) = \sum_{i=1}^{n} \tilde{W}^{S}_{nc,i}(t, z)\xi_i,$$

where $\{\xi_i, i = 1, \cdots, n\}$ are i.i.d. standard normal variables. It is shown in Appendix A4 that conditional on the observed data $\{N_i(t), Y_i(t), W_{ir}, V_i, i = 1, \cdots, n, r = 1, \cdots, n_i\}$,

$n^{-1/2}\tilde{W}_{nc}^{S}(t,z)$ converges weakly to $\mathcal{G}_{nc}(t,z)$. Correspondingly, conditional on the data, the limiting distribution of $S_{nc}^{S} = n^{-1/2}\sup_{t,z}|\tilde{W}_{nc}^{S}(t,z)|$ is the same as that of S_{nc}.

Therefore, to approximate the limit distribution of $n^{-1/2}\hat{W}_{nc}(t,z;\hat{\beta}_{nc})$, we simulate a number of realizations of $\tilde{W}_{nc}^{S}(t,z)$ by generating sets of i.i.d. standard normal variables $\{\xi_i, i = 1,\cdots,n\}$ for N times, where N is a large number, say $N = 200$, while holding the observed data fixed. Correspondingly, we have N independent replicates of S_{nc}^{S}, which have the same limiting distribution as that of S_{nc}. Therefore, the p-value $\Pr(S_{nc} \geq s)$ can be estimated by $\Pr(S_{nc}^{S} \geq s)$ based on the N replicates of S_{nc}^{S}.

4.4.2 Model Checking under Measurement Error Scenario 2

We define the stochastic process

$$\hat{W}_c(t,z;\beta) = \sum_{i=1}^{n}\int_0^t \left\{ I(\hat{Z}_i \leq z - \Sigma_1\beta) - \frac{V_c^{(1)}(\hat{Z};\beta,u)}{S_c^{(0)}(\hat{Z};\beta,u)} \right\} dN_i(u),$$

where $V_c^{(1)}(\hat{Z};\beta,u) = n^{-1}\sum_{i=1}^{n} Y_i(u)I(\hat{Z}_i \leq z)\exp(\hat{Z}_i^T\beta)$. Let

$$V_c^{(2)}(\hat{Z};\beta,u) = n^{-1}\sum_{i=1}^{n} Y_i(u)I(\hat{Z}_i \leq z)\hat{Z}_i\exp(\hat{Z}_i^T\beta),$$

and $v_c^{(k)}(\hat{Z};\beta,u) = E\{V_c^{(k)}(\hat{Z};\beta,u)\}$, $k = 1,2$.

We verify in Appendix A1 that $\hat{W}_c(t,z;\beta)$ satisfies the property (4.6). Write $I(\hat{Z}_i \leq z) = I(W_i \leq x)I(V_i \leq v)$, where $z = (x, v^T)^T$. Let $F_\epsilon(\cdot)$ be the cumulative distribution function of ϵ_i. In Appendix A2, we show that $n^{-1/2}\hat{W}_c(t,z;\hat{\beta}_c)$ is asymptotically equivalent to $n^{-1/2}\tilde{W}_c(t,z)$, where

$$
\begin{aligned}
&\tilde{W}_c(t,z)\\
=&\sum_{i=1}^{n}\Bigg[\int_0^t \left\{ I(W_i \le x - \Sigma_0\beta_{0,x}) I(V_i \le v) - \frac{v_c^{(1)}(\hat{Z};\beta_0,u)}{s_c^{(0)}(\hat{Z};\beta,u)} \right\} dN_i(u)\\
&- \int_0^t \left\{ \frac{Y_i(u) I(\hat{Z}_i \le z)\exp(\hat{Z}_i^T\beta_0)}{s_c^{(0)}(\hat{Z};\beta_0,u)} - \frac{Y_i(u)\exp(\hat{Z}_i^T\beta_0) v_c^{(1)}(\hat{Z};\beta_0,u)}{\{s_c^{(0)}(\hat{Z};\beta_0,u)\}^2} \right\} dE\{N_i(u)\}\\
&- \{\mathcal{H}_{c1}(t,z;\beta_0) - \mathcal{H}_{c2}(t,z;\beta_{0,x})\}\mathcal{I}_c^{-1}(\beta_0)\mathcal{J}_{c,i}(\beta_0)\Bigg]\\
&\equiv \tilde{W}_{c,i}(t,z).
\end{aligned}
$$

Here,

$$
\mathcal{H}_{c1}(t,z;\beta) = \int_0^t \left[\frac{v_c^{(2)}(\hat{Z};\beta,u)}{s_c^{(0)}(\hat{Z};\beta,u)} - \frac{v_c^{(1)}(\hat{Z};\beta,u) s_c^{(1)T}(\hat{Z};\beta,u)}{\{s_c^{(0)}(\hat{Z};\beta,u)\}^2} \right] dE\{N_i(u)\},
$$

$$
\text{and}\ \ \mathcal{H}_{c2}(t,z;\beta_x) = \left(E\left[\frac{\partial F_\epsilon(x - \Sigma_0\beta_x - X_i)}{\partial \beta_x^T} I(V_i \le v) N_i(t) \right], 0^T \right).
$$

Note that $\tilde{W}_c(t,z)$ is a sum of zero-mean independent random variables $\tilde{W}_{c,i}(t,z)$. In Appendix A3, we establish the weak convergence properties of $n^{-1/2}\tilde{W}_c(t,z)$ and $n^{-1/2}\hat{W}_c(t,z;\hat{\beta}_c)$, summarized in the following theorem.

Theorem 2 *Under Regularity Conditions, we have*

$$
n^{-1/2}\tilde{W}_c(t,z) \rightsquigarrow \mathcal{G}_c(t,z) \quad in \quad l^\infty([0,\tau]\times\mathcal{R}^{(p+q)}) \quad as \quad n\to\infty,
$$

where $\mathcal{G}_c(t,z)$ is a zero-mean Gaussian process with covariance function

$$
\Phi_c(t_1,t_2,z_1,z_2) = E[\tilde{W}_{c,i}(t_1,z_1)\tilde{W}_{c,i}(t_2,z_2)]
$$

for time points t_1 and t_2 and real values z_1 and z_2. Furthermore,

$$
n^{-1/2}\hat{W}_c(t,z;\hat{\beta}_c) \rightsquigarrow \mathcal{G}_c(t,z) \quad in \quad l^\infty([0,\tau]\times\mathcal{R}^{(p+q)}) \quad as \quad n\to\infty.
$$

We propose an overall goodness-of-fit test statistic

$$S_c = n^{-1/2} \sup_{t,z} |\hat{W}_c(t, z; \hat{\beta}_c)|.$$

The p-value of S_c is difficult to obtain directly by Theorem 2, and thus we adopt a resampling procedure. In Appendix A5, we investigate the consistency properties of the proposed test based on S_c.

However, the term $\partial F_\epsilon(x - \Sigma_0\beta_x - X_i)/\partial\beta_x^T$ in $\hat{W}_c(t, z; \beta)$ is difficult to estimate as it involves the unobserved variable X_i. In the following, we focus on the case that W_i is univariate, as is common in many studies. Let $\Sigma_0 = \sigma_0^2$, where σ_0 is the standard error of the measurement error. Then,

$$\frac{\partial F_\epsilon(x - \Sigma_0\beta_x - X_i)}{\partial\beta_x^T} = -\frac{\sigma_0}{\sqrt{2\pi}} A(X_i; \beta_x),$$

where

$$A(X_i; \beta_x) = \exp\left\{-\frac{(x - \sigma_0^2\beta_x - X_i)^2}{2\sigma_0^2}\right\}.$$

To make $\tilde{W}_{c,i}^S(t, z)$ computable based on the observed data, it is tempting to construct a function, say $\hat{A}(W_i; \beta_x)$, based on W_i, such that

$$E\{\hat{A}(W_i; \beta_x)|X_i\} = A(X_i; \beta_x),$$

where the conditional expectation is taken under the null hypothesis. Such $\hat{A}(W_i; \beta_x)$ can be served as an accurate approximation of $A(X_i; \beta_x)$ (Stefanski, 1989; Novick and Stefanski, 2002).

Let $\tilde{W}_{c,i}^S(t, z)$ be the estimated version of $\tilde{W}_{c,i}(t, z)$, where $s_c^{(k)}(\hat{Z}; \beta_0, t)$ is replaced by $S_c^{(k)}(\hat{Z}; \hat{\beta}_c, t), k = 0, 1, 2,$ $v_c^{(k)}(\hat{Z}; \beta_0, t)$ is replaced by $V_c^{(k)}(\hat{Z}; \hat{\beta}_c, t), k = 1, 2,$ $E\{N_i(t)\}$ is replaced by $n^{-1}\sum_{i=1}^n N_i(t)$, and β_0 is replaced by $\hat{\beta}_c$. Let

$$\tilde{W}_c^S(t, z) = \sum_{i=1}^n \tilde{W}_{c,i}^S(t, z)\xi_i,$$

where $\{\xi_i, i = 1, \cdots, n\}$ are i.i.d. standard normal variables. Analogous to Scenario 1, conditional on the observed data $\{N_i(t), Y_i(t), W_i, V_i, i = 1, \cdots, n\}$, $n^{-1/2}\tilde{W}_c^S(t, z)$ converges weakly to $\mathcal{G}_c(t, z)$, and the proof is sketched in Appendix A4. Correspondingly, conditional on the data, the limiting distribution of $S_c^S = n^{-1/2} \sup_{t,z} |\tilde{W}_c^S(t, z)|$ is the same as that of S_c. Therefore, the p-value $\Pr(S_c \geq s)$ can be estimated based on replicates of S_c^S.

The disscussion above assumes $\hat{A}(W_i; \beta_x)$ is available. However, by the arguments of Stefanski (1989), there does not exist $\hat{A}(W_i; \beta_x)$ that is unbiased of $A(X_i; \beta_x)$. To circumvent the difficulty, we use

$$\hat{A}(W_i; \beta_x) = A(W_i; \beta_x) \left\{ \frac{3}{2} - \frac{(x - \sigma_0^2\beta_x - W_i)^2}{2\sigma_0^2} \right\},$$

in the above arguments of estimating the p-value $\Pr(S_c \geq s)$. Following the arguments of Stefanski (1989), $\hat{A}(W_i; \beta_x)$ provides a reasonable approximation of $A(X_i; \beta_x)$.

4.5 Numerical Studies

4.5.1 Simulation Studies

We numerically assess the performance of the proposed test statistic S_{nc}. We consider the sample size $n = 100$, the number of replicates $N = 200$ in the resampling procedure, and generate 500 simulations for each parameter configuration. The true covariates X_i are generated from the standard exponential distribution EXP(1), and V_i are generated from

$$\Pr(V_i = 1|X_i) = \frac{\exp(X_i)}{1 + \exp(X_i)}.$$

First, we evaluate the empirical size of the tests. The null hypothesis is that both of the Cox model and the additive error model are correctly specified. Survival times are independently generated from the Cox model, where we take the baseline hazard function to be $\lambda_0(t) = \alpha\gamma t^{\gamma-1}$, with $\alpha = 0.5$, and $\gamma = 2$. The true values of β_x and β_v are set to be $(\beta_x, \beta_v) = (1, 1)$. Censoring times C_i are generated from uniform distribution UNIF$[0, C]$ where C is set to be 2.05, so that approximately 30% censoring is produced. The error model (4.4) in Scenario 1 is used to generate W_{ir} where $\epsilon_{ir} \sim N(0, \sigma_0^2)$ for $r = 1, 2, i = 1, \cdots, n$. We consider settings with $\sigma_0 = 0.25$ or 0.50 to indicate different degrees of measurement error.

We generate $N = 200$ sets of i.i.d. standard normal variables $\{\xi_i, i = 1, \cdots, n\}$, and then calculate $N = 200$ copies of S_{nc}^S, say $\{S_{nc,k}^S, k = 1, \cdots, N\}$. Empirical quantiles of S_{nc} can then be obtained based on $\{S_{nc,k}^S, k = 1, \cdots, N\}$. The nominal level is set to be 0.05.

The results for the empirical size of S_{nc} are around 0.075 for both settings with $\sigma_0 = 0.25$ or 0.50, which are close to the nominal level.

4.5.2 An Example

We apply the proposed methods to analyze the data arising from the AIDS Clinical Trials Group (ACTG) 175 study (Hammer, et al. 1996). The ACTG 175 study is a double-blind randomized clinical trial that evaluated the effects of HIV treatments. In this example, we are interested in evaluating how treatments are associated with the survival time T_i, which is defined to be the time to the occurrence of one of the events that CD4 counts decrease at least 50%, or disease progression to AIDS, or death. We consider a subset of $n = 344$ subjects in the study, who did not receive non-zidovudine antiretroviral therapy prior to initiation of study treatment, not use zidovudine in the 30 days prior to treatment initiation, and not have one of the major risk factors: homosexuality, injetion-drug use, and hemophilia before treatment.

Let V_i be the treatment assignment indicator for subject i, where $V_i = 1$ if a subject received the zidovudine only treatment, and 0 otherwise. In the ACTG 175 study, the baseline measurements on CD4 were collected before randomization, ranging from 200 to 500 per cubic millimeter. Let X_i be the normalization version of the true baseline CD4 counts: $\log(\text{CD4 counts} + 1)$, which was not observed in the study. Two replicated baseline measurements of CD4 counts, denoted by W_{i1} and W_{i2}, are available. An additive measurement error model is specified to link the underlying transformed CD4 counts with its surrogate measurements:

$$W_{ir} = X_i + \epsilon_{ir},$$

where $r = 1, 2$ for $i = 1, \cdots, 344$. We employ the Cox model to feature the dependence of T_i on the covariates X_i and V_i:

$$\lambda(t; Z_i) = \lambda_0(t) \exp(X_i \beta_x + V_i \beta_v).$$

We apply the proposed test based on the test statistic S_{nc} to the dataset. The p-value of the model test is 0.10, suggesting some evidence against the Cox model or the error model.

4.6 Discussion

The Cox model has been widely used in practise, and there are numerous methods that successfully correct measurement error effect under the Cox model since Prentice (1982). However, all these methods presume that both Cox model and error model are correctly specified. There is little work on model checking procedure of the Cox model and the error model. In this chapter, we develop two goodness-of-fit tests to fill the gap. Theoretical results are established together with some numerical studies. More simulation studies will be conducted to assess the finite sample performance of the proposed tests under a broader range of settings.

In this chapter, we assumed time-independent covariates for simplicity. It is interesting to extend our tests to incorporate time-dependent covariates. Furthermore, the proposed tests can be potentially adjusted analogue to Lin, Wei and Ying (1993), so that we may check specific assumptions of the Cox model (e.g.. the proportional hazards assumption).

Appendix

Appendix A1

We first prove $\hat{W}_{nc}(t, z; \beta)$ satisfies the property (4.6). The proof in Appendix 1 of Lin, Wei and Ying (1993) implied that under regularity conditions,

$$n^{-1} W(t, z; \beta) = o_{a.s.}(1),$$

uniformly in t, z, and β. Therefore, it remains to prove that

$$n^{-1} \hat{W}_{nc}(t, z; \beta) = o_{a.s.}(1),$$

uniformly in t, z, and β.

By the the Strong Uniform Law of Large Numbers (SULLN) (Pollard 1990, p.41), we obtain that

$$\begin{aligned} n^{-1} \sum_{i=1}^{n} n_i^{-1} \sum_{r=1}^{n_i} I(\hat{Z}_{ir} \le z) N_i(t) =& E\{I(\hat{Z}_{ir} \le z) N_i(t)\} + o_{a.s.}(1), \\ V_{nc}^{(1)}(\hat{Z}; \beta, t, z) =& v_{nc}^{(1)}(\hat{Z}; \beta, t, z) + o_{a.s.}(1), \\ S_{nc}^{(0)}(\hat{Z}; \beta, t) =& s_{nc}^{(0)}(\hat{Z}; \beta, t) + o_{a.s.}(1), \\ \text{and} \quad n^{-1} \sum_{i=1}^{n} N_i(t) =& E\{N_i(t)\} + o_{a.s.}(1), \end{aligned}$$

uniformly in t, z, and β. Let $\tilde{\epsilon}_{ir} = (\epsilon_{ir}^T, 0^T)^T$, so that $\tilde{\epsilon}_{ir}$ has the same dimension as Z_i.

Therefore,

$$
\begin{aligned}
n^{-1}\hat{W}_{nc}(t,z;\beta) =& E\{I(\hat{Z}_{ir} \le z)N_i(t)\} - \int_0^t \frac{v_{nc}^{(1)}(\hat{Z};\beta,u,z)}{s_{nc}^{(0)}(\hat{Z};\beta,u)} dE\{N_i(u)\} + o_{a.s.}(1) \\
=& E\{I(\tilde{\epsilon}_{ir} \le z - Z_i)N_i(t)\} \\
& - \int_0^t \frac{E\{Y_i(u)\exp(Z_i^T\beta)I(\tilde{\epsilon}_{ir} \le z - Z_i)\exp(\tilde{\epsilon}_{is}^T\beta)\}}{E\{Y_i(u)\exp(Z_i^T\beta)\}E\{\exp(\tilde{\epsilon}_{ir}^T\beta)\}} E\{Y_i(u)\exp(Z_i^T\beta)\}\lambda_0(u)du \\
& + o_{a.s.}(1) \\
=& E\{P(\tilde{\epsilon}_{ir} \le z - Z_i|Z_i)N_i(t)\} \\
& - \int_0^t \frac{E\{Y_i(u)\exp(Z_i^T\beta)P(\tilde{\epsilon}_{ir} \le z - Z_i|Z_i)\exp(\tilde{\epsilon}_{is}^T\beta)\}}{E\{\exp(\tilde{\epsilon}_{ir}^T\beta)\}} \lambda_0(u)du + o_{a.s.}(1) \\
=& E\left\{P(\epsilon_{ir} \le z - Z_i|Z_i)\int_0^t Y_i(u)\exp(Z_i^T\beta)\lambda_0(u)du\right\} \\
& - \int_0^t \frac{E\{Y_i(u)\exp(Z_i^T\beta)P(\tilde{\epsilon}_{ir} \le z - Z_i|Z_i)\}E\{\exp(\tilde{\epsilon}_{is}^T\beta)\}}{E\{\exp(\tilde{\epsilon}_{ir}^T\beta)\}} \lambda_0(u)du + o_{a.s.}(1) \\
=& E\left\{\int_0^t Y_i(u)\exp(Z_i^T\beta)P(\tilde{\epsilon}_{ir} \le z - Z_i|Z_i)\lambda_0(u)du\right\} \\
& - \int_0^t E\{Y_i(u)\exp(Z_i^T\beta)P(\tilde{\epsilon}_{ir} \le z - Z_i|Z_i)\lambda_0(u)\}du + o_{a.s.}(1),
\end{aligned}
$$

uniformly in t, z, and β. Under certain regularity conditions, the expectation and integration are exchangeable by Fubini's Theorem. It follows that $n^{-1}\hat{W}_{nc}(t,z;\beta) = o_{a.s.}(1)$ uniformly in t, z, and β. Therefore, $\hat{W}_{nc}(t,z;\beta)$ satisfies the property (4.6).

Now, we prove $\hat{W}_c(t,z;\beta)$ satisfies the property (4.6). Similar to the above arguments, we only need to show that

$$
n^{-1}\hat{W}_c(t,z;\beta) = o_{a.s.}(1),
$$

uniformly in t, z, and β.

Note that uniformly in t, z, and β,

$$
\begin{aligned}
n^{-1}\hat{W}_c(t,z;\beta) =& E\{I(\hat{Z}_i \le z - \Sigma_1\beta)N_i(t)\} - \int_0^t \frac{v_c^{(1)}(\hat{Z};\beta,u,z)}{s_c^{(0)}(\hat{Z};\beta,u)} dE\{N_i(u)\} + o_{a.s.}(1) \\
=& E\{I(\epsilon_i \le x - \Sigma_0\beta_x - X_i)I(V_i \le v)N_i(t)\} \\
&- \int_0^t \frac{E\{Y_i(u)\exp(Z_i^T\beta)I(V_i \le v)I(\epsilon_i \le x - X_i)\exp(\epsilon_i^T\beta_x)\}}{E\{\exp(\epsilon_i^T\beta_x)\}} \lambda_0(u)du + o_{a.s.}(1).
\end{aligned}
$$

Note also that

$$
\begin{aligned}
& E\{I(\epsilon_i \le x - \Sigma_0\beta_x - X_i)I(V_i \le v)N_i(t)\} \\
=& E\{F_\epsilon(x - \Sigma_0\beta_x - X_i)I(V_i \le v)N_i(t)\} \\
=& E\left\{\int_0^t Y_i(u)F_\epsilon(x - \Sigma_0\beta_x - X_i)I(V_i \le v)\exp(Z_i^T\beta)\lambda_0(u)du\right\}.
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
& E\{Y_i(u)\exp(Z_i^T\beta)I(V_i \le v)I(\epsilon_i \le x - X_i)\exp(\epsilon_i^T\beta_x)\} \\
=& E\left\{Y_i(u)\exp(Z_i^T\beta)I(V_i \le v)\int_{-\infty}^{x-X_i}(2\pi)^{-\frac{p}{2}}|\Sigma_0|^{-\frac{1}{2}}\exp(-\frac{1}{2}\epsilon^T\Sigma_0^{-1}\epsilon)\exp(\epsilon^T\beta_x)d\epsilon\right\} \\
=& E\Bigg[Y_i(u)\exp(Z_i^T\beta)I(V_i \le v) \\
&\times \int_{-\infty}^{x-X_i}(2\pi)^{-\frac{p}{2}}|\Sigma_0|^{-\frac{1}{2}}\exp(\frac{1}{2}\beta_x^T\Sigma_0^{-1}\beta_x)\exp\left\{-\frac{1}{2}(\epsilon - \Sigma_0\beta_x)^T\Sigma_0^{-1}(\epsilon - \Sigma_0^T\beta_x)\right\}d\epsilon\Bigg] \\
=& E\left\{Y_i(u)\exp(Z_i^T\beta)I(V_i \le v)\int_{-\infty}^{x-X_i-\Sigma_0\beta_x}(2\pi)^{-\frac{p}{2}}|\Sigma_0|^{-\frac{1}{2}}\exp(\frac{1}{2}\beta_x^T\Sigma_0^{-1}\beta_x)\exp(-\frac{1}{2}\epsilon^T\Sigma_0^{-1}\epsilon)d\epsilon\right\} \\
=& \exp(\frac{1}{2}\beta_x^T\Sigma_0^{-1}\beta_x)E\left\{Y_i(u)F_\epsilon(x - \Sigma_0\beta_x - X_i)I(V_i \le v)\exp(Z_i^T\beta)\right\},
\end{aligned}
$$

and

$$
E\{\exp(\epsilon_i^T\beta_x)\} = \exp(\frac{1}{2}\beta_x^T\Sigma_0^{-1}\beta_x).
$$

It follows that

$$
\begin{aligned}
&n^{-1}\hat{W}_c(t,z;\beta)\\
=&E\left\{\int_0^t Y_i(u)F_\epsilon(x-\Sigma_0\beta_x-X_i)I(V_i\le v)\exp(Z_i^T\beta)\lambda_0(u)du\right\}\\
&-\int_0^t \frac{\exp(\frac{1}{2}\beta_x^T\Sigma_0^{-1}\beta_x)E\left\{Y_i(u)F_\epsilon(x-\Sigma_0\beta_x-X_i)I(V_i\le v)\exp(Z_i^T\beta)\right\}}{\exp(\frac{1}{2}\beta_x^T\Sigma_0^{-1}\beta_x)}\lambda_0(u)du+o_{a.s.}(1)\\
=&E\left\{\int_0^t Y_i(u)F_\epsilon(x-\Sigma_0\beta_x-X_i)I(V_i\le v)\exp(Z_i^T\beta)\lambda_0(u)du\right\}\\
&-\int_0^t E\left\{Y_i(u)F_\epsilon(x-\Sigma_0\beta_x-X_i)I(V_i\le v)\exp(Z_i^T\beta)\lambda_0(u)\right\}du+o_{a.s.}(1),
\end{aligned}
$$

uniformly in t, z, and β. Under certain regularity conditions, the expectation and integration are exchangeable by Fubini's Theorem. It follows that $n^{-1}\hat{W}_c(t,z;\beta) = o_{a.s.}(1)$ uniformly in t, z, and β. Therefore, $\hat{W}_c(t,z;\beta)$ satisfies the property (4.6).

Appendix A2

We first consider $n^{-1/2}\hat{W}_{nc}(t,z;\hat{\beta}_{nc})$. Note that we have

$$
\begin{aligned}
&n^{-1/2}\hat{W}_{nc}(t,z;\beta_0)\\
=&n^{-1/2}\sum_{i=1}^n\int_0^t\left\{n_i^{-1}\sum_{r=1}^{n_i}I(\hat{Z}_{ir}\le z)-\frac{v_{nc}^{(1)}(\hat{Z};\beta_0,u,z)}{s_{nc}^{(0)}(\hat{Z};\beta_0,u)}\right\}dN_i(u)\\
&+n^{1/2}\int_0^t\left\{\frac{v_{nc}^{(1)}(\hat{Z};\beta_0,u,z)}{s_{nc}^{(0)}(\hat{Z};\beta_0,u)}-\frac{V_{nc}^{(1)}(\hat{Z};\beta_0,u,z)}{S_{nc}^{(0)}(\hat{Z};\beta_0,u)}\right\}dE\{N_i(u)\}\\
&+\int_0^t\left\{\frac{v_{nc}^{(1)}(\hat{Z};\beta_0,u,z)}{s_{nc}^{(0)}(\hat{Z};\beta_0,u)}-\frac{V_{nc}^{(1)}(\hat{Z};\beta_0,u,z)}{S_{nc}^{(0)}(\hat{Z};\beta_0,u)}\right\}d\left(n^{1/2}\left[\sum_{i=1}^n N_i(u)/n-E\{N_i(u)\}\right]\right).
\end{aligned}
$$

By SULLN,

$$
\int_0^t\left\{\frac{v_{nc}^{(1)}(\hat{Z};\beta_0,u,z)}{s_{nc}^{(0)}(\hat{Z};\beta_0,u)}-\frac{V_{nc}^{(1)}(\hat{Z};\beta_0,u,z)}{S_{nc}^{(0)}(\hat{Z};\beta_0,u)}\right\}d\left(n^{1/2}\left[\sum_{i=1}^n N_i(u)/n-E\{N_i(u)\}\right]\right)=o_{a.s.}(1)
$$

uniformly in t and z. By Taylor series expansion,

$$
\begin{aligned}
& n^{1/2} \int_0^t \left\{ \frac{v_{nc}^{(1)}(\hat{Z};\beta_0,u,z)}{s_{nc}^{(0)}(\hat{Z};\beta_0,u)} - \frac{V_{nc}^{(1)}(\hat{Z};\beta_0,u,z)}{S_{nc}^{(0)}(\hat{Z};\beta_0,u)} \right\} dE\{N_i(u)\} \\
&= - n^{-1/2} \int_0^t \Bigg[\frac{Y_i(u)\{n_i^{-1}(n_i-1)^{-1}\sum_{r\neq s} I(\hat{Z}_{ir}\le z)\exp(\hat{Z}_{is}^T\beta_0)\}}{s_{nc}^{(0)}(\hat{Z};\beta_0,u)} \\
&\quad - \frac{Y_i(u)\{n_i^{-1}\sum_{r=1}^{n_i}\exp(\hat{Z}_{ir}^T\beta_0)\}v_{nc}^{(1)}(\hat{Z};\beta_0,u,z)}{\{s_{nc}^{(0)}(\hat{Z};\beta_0,u)\}^2} \Bigg] dE\{N_i(u)\} + o_p(1).
\end{aligned}
$$

Furthermore, Huang and Wang (2000) showed that

$$
\sqrt{n}(\hat{\beta}_{nc}-\beta_0) = n^{-1/2}\sum_{i=1}^n \mathcal{I}_{nc}^{-1}(\beta_0)J_{nc,i}(\beta_0) + o_p(1).
$$

By SULLN, we have

$$
n^{-1}\frac{\partial \hat{W}_{nc}(t,z;\beta_0)}{\partial\beta^T}\Bigg|_{\beta=\beta_0} = -\mathcal{H}_{nc}(t,z;\beta_0) + o_{a.s.}(1),
$$

uniformly in t and z. Therefore, by Taylor series expansion, we have

$$
\begin{aligned}
n^{-1/2}\hat{W}_{nc}(t,z;\hat{\beta}_{nc}) &= n^{-1/2}\hat{W}_{nc}(t,z;\beta_0) + n^{-1}\frac{\partial \hat{W}_{nc}(t,z;\beta)}{\partial\beta^T}\Bigg|_{\beta=\beta_0}\sqrt{n}(\hat{\beta}_{nc}-\beta_0) + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^n \tilde{W}_{nc,i}(t,z) + o_p(1).
\end{aligned}
$$

Thus, $n^{-1/2}\hat{W}_{nc}(t,z;\hat{\beta}_{nc})$ is asymptotically equivalent to $n^{-1/2}\tilde{W}_{nc}(t,z)$.

Now we consider $n^{-1/2}\hat{W}_c(t,z;\hat{\beta}_c)$. Note that the class of indicator functions is a Donsker class (van der Vaart and Wellner 1996), we have by Theorem 2.1 of van der

Vaart and Wellner (2007) that

$$
\begin{aligned}
& I(W_i \le x - \Sigma_0 \hat{\beta}_{c,x}) I(V_i \le v) N_i(t) \\
= & I(W_i \le x - \Sigma_0 \beta_{0,x}) I(V_i \le v) N_i(t) + E\{I(W_i \le x - \Sigma_0 \hat{\beta}_{c,x}) I(V_i \le v) N_i(t)\} \\
& - E\{I(W_i \le x - \Sigma_0 \beta_{0,x}) I(V_i \le v) N_i(t)\} + o_p(\frac{1}{\sqrt{n}}) \\
= & I(W_i \le x - \Sigma_0 \beta_{0,x}) I(V_i \le v) N_i(t) + E\{F(x - \Sigma_0 \hat{\beta}_{c,x} - X_i) I(V_i \le v) N_i(t)\} \\
& - E\{F(x - \Sigma_0 \beta_{0,x} - X_i) I(V_i \le v) N_i(t)\} + o_p(\frac{1}{\sqrt{n}}) \\
= & I(W_i \le x - \Sigma_0 \beta_{0,x}) I(V_i \le v) N_i(t) + \mathcal{H}_{c2}(t, z; \beta_{0,x})(\hat{\beta}_c - \beta_0) + o_p(\frac{1}{\sqrt{n}}).
\end{aligned}
$$

Similar to the arguments that proved the asymptotically equivalence property of $n^{-1/2}\hat{W}_{nc}(t, z; \hat{\beta}_{nc})$, we have

$$
\begin{aligned}
& n^{-1/2}\hat{W}_c(t, z; \beta_0) \\
= & \sum_{i=1}^n \int_0^t \left\{ I(W_i \le x - \Sigma_0 \beta_{0,x}) I(V_i \le v) - \frac{v_c^{(1)}(\hat{Z}; \beta_0, u)}{s_c^{(0)}(\hat{Z}; \beta, u)} \right\} dN_i(u) \\
& - \int_0^t \left\{ \frac{Y_i(u) I(\hat{Z}_i \le z) \exp(\hat{Z}_i^T \beta_0)}{s_c^{(0)}(\hat{Z}; \beta_0, u)} - \frac{Y_i(u) \exp(\hat{Z}_i^T \beta_0) v_c^{(1)}(\hat{Z}; \beta_0, u)}{\{s_c^{(0)}(\hat{Z}; \beta_0, u)\}^2} \right\} dE\{N_i(u)\} + o_p(1).
\end{aligned}
$$

Furthermore, Kong and Gu (1999) showed that

$$
\sqrt{n}(\hat{\beta}_c - \beta_0) = n^{-1/2} \sum_{i=1}^n \mathcal{I}_c^{-1}(\beta_0) J_{c,i}(\beta_0) + o_p(1).
$$

By SULLN, we have

$$
n^{-1} \frac{\partial \hat{W}_c(t, z; \beta_0)}{\partial \beta^T} \bigg|_{\beta = \beta_0} = -\mathcal{H}_{c1}(t, z; \beta_0) + o_{a.s.}(1),
$$

uniformly in t and z. Therefore, we have

$$
\begin{aligned}
n^{-1/2}\hat{W}_c(t,z;\hat{\beta}_c) &= n^{-1/2}\hat{W}_c(t,z;\beta_0) + n^{-1}\frac{\partial \hat{W}_c(t,z;\beta)}{\partial \beta^T}\bigg|_{\beta=\beta_0}\sqrt{n}(\hat{\beta}_c-\beta_0) \\
&\quad + \mathcal{H}_{c2}(t,z;\beta_{0,x})\sqrt{n}(\hat{\beta}_c-\beta_0) + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^{n}\tilde{W}_{c,i}(t,z) + o_p(1).
\end{aligned}
$$

Thus, $n^{-1/2}\hat{W}_c(t,z;\hat{\beta}_c)$ is asymptotically equivalent to $n^{-1/2}\tilde{W}_c(t,z)$.

Appendix A3

We first show the tightness of $n^{-1/2}\tilde{W}_{nc}(t,z)$. Note that

$$
n^{-1/2}\hat{W}_{nc}(t,z;\hat{\beta}_{nc}) = n^{-1/2}\hat{W}_{nc}(t,z;\beta_0) + n^{-1}\frac{\partial \hat{W}_{nc}(t,z;\beta)}{\partial \beta^T}\bigg|_{\beta=\beta^*}\sqrt{n}(\hat{\beta}_{nc}-\beta_0),
$$

where β^* is in the line segment of $\hat{\beta}_{nc}$ and β_0. The second term in the above equation is tight since $n^{-1}\frac{\partial \hat{W}_{nc}(t,z;\beta_0)}{\partial \beta^T}\Big|_{\beta=\beta^*}$ converges almost surely to $-\mathcal{H}_{nc}(t,z;\beta_0)$ uniformly, and $\sqrt{n}(\hat{\beta}_{nc}-\beta_0)$ converges in distribution. Therefore, we only need to show the tightness of

$$
n^{-1/2}\hat{W}_{nc}(t,z;\beta_0) = n^{-1/2}\sum_{i=1}^{n}\int_0^t\left\{n_i^{-1}\sum_{r=1}^{n_i} I(\hat{Z}_{ir}\le z) - \frac{V_{nc}^{(1)}(\hat{Z};\beta_0,u,z)}{S_{nc}^{(0)}(\hat{Z};\beta_0,u)}\right\}dN_i(u).
$$

Note that $n^{-1/2}\sum_{i=1}^{n} n_i^{-1}\sum_{r=1}^{n_i} I(\hat{Z}_{ir}\le z)N_i(t)$ is sum of monotone functions, and thus is manageable (Pollard 1990). Therefore, the first term is tight.

By the Lindeberg-Feller Central Limit Theorem, we obtain that $n^{-1/2}\tilde{W}_{nc}(t,z)$ is asymptotically normal with mean zero, and covariance $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E[\tilde{W}_{nc,i}^2(t,z)]$. Therefore, $n^{-1/2}\tilde{W}_{nc}(t,z)$ converges weakly to $\mathcal{G}_{nc}(t,z)$.

By the asymptotic equivalence property of $n^{-1/2}\hat{W}_{nc}(t,z;\hat{\beta}_{nc})$ proved in A2, we have that $n^{-1/2}\hat{W}_{nc}(t,z;\hat{\beta}_{nc})$ also converges weakly to $\mathcal{G}_{nc}(t,z)$.

Appendix A4

We prove the weak convergence property of $\tilde{W}^S_{nc}(t,z)$ in the following, and the weak convergence property of $\tilde{W}^S_c(t,z)$ can be proved similarly.

Let $\tilde{W}^s_{nc}(t,z) = \sum_{i=1}^n \tilde{W}_{nc,i}(t,z)\xi_i$. By Theorem 1, $n^{-1/2}\tilde{W}_{nc}(t,z) = n^{-1/2}\sum_{i=1}^n \tilde{W}_{nc,i}(t,z)$ converges weakly to $\mathcal{G}_{nc}(t,z)$ unconditionally. Since the weak convergence of $n^{-1/2}\sum_{i=1}^n \tilde{W}_{nc,i}(t,z)$ implies that the Donsker condition (van der Vaart and Wellner 1996, Theorem 2.9.6) holds, it then follows from the conditional multiplier Central Limit Theorem (van der Vaart and Wellner 1996, Sec. 2.9) that $n^{-1/2}\tilde{W}^s_{nc}(t,z)$ converges weakly to $\mathcal{G}_{nc}(t,z)$ in probability conditional on the data. Thus, by Lemma 1 of Pipper and Ritz (2007), it suffices to show that $\sup_{t,z}\left\{n^{-1/2}|\tilde{W}^S_{nc}(t,z) - \tilde{W}^s_{nc}(t,z)|\right\} \xrightarrow{p} 0$. This can be proved analogue to the arguments in the Appendices of Spiekerman and Lin (1998).

Appendix A5

<u>*Consistency of S_{nc}*</u>: First, we study the scenario that the measurement error model is correctly specified, whereas the Cox model is misspecified. We need to show that when the error model (4.4) is correctly specified, the S_{nc} supremum test is consistent against the alternative hypothesis that there does not exist a constant vector β_0 and a function $\lambda_0(\cdot)$, such that the hazard function has the form

$$\lambda(t;z) = \lambda_0(t)\exp(z^T\beta_0).$$

for almost all $t \in [0,\tau]$, and z in the support of Z_i. Note that under this alternative, we have $\hat{\beta}_{nc} \xrightarrow{a.s.} \beta^*$, and that

$$\sum_{i=1}^n \int_0^t \frac{dN_i(u)}{\sum_{i=1}^n Y_i(u)\exp(Z_i^T\beta^*)} \xrightarrow{a.s.} \int_0^t \lambda_0^*(u)du,$$

as $n \to \infty$ (Lin and Wei 1989). Let $\lambda(t;Z_i)$ be the hazard function under the alternative. Let E_a denote the expectation taken under this general alternative. Let $H(\cdot)$ be the distribution function of Z_i. Let $\tilde{\epsilon}_{ir}$ be defined as in Appendix A1.

$$
\begin{aligned}
&n^{-1}\hat{W}_{nc}(t,z;\hat{\beta}_{nc})\\
=&n^{-1}\hat{W}_{nc}(t,z;\beta^*)+o_{a.s.}(1)\\
=&E_a\{I(\tilde{\epsilon}_{ir}\leq z-Z_i)N_i(t)\}\\
&-\int_0^t \frac{E_a\{Y_i(u)\exp(Z_i^T\beta^*)I(\tilde{\epsilon}_{ir}\leq z-Z_i)\exp(\tilde{\epsilon}_{is}^T\beta^*)\}}{E_a\{Y_i(u)\exp(Z_i^T\beta^*)\}E_a\{\exp(\tilde{\epsilon}_{ir}^T\beta^*)\}}dE_a\{N_i(u)\}+o_{a.s.}(1)\\
=&E_a\{P(\tilde{\epsilon}_{ir}\leq z-Z_i|Z_i)N_i(t)\}\\
&-\int_0^t \frac{E_a\{Y_i(u)\exp(Z_i^T\beta^*)P(\tilde{\epsilon}_{ir}\leq z-Z_i|Z_i)\exp(\tilde{\epsilon}_{is}^T\beta^*)\}}{E_a\{\exp(\tilde{\epsilon}_{ir}^T\beta^*)\}}\lambda_0^*(u)du+o_{a.s.}(1)\\
=&E_a\{P(\tilde{\epsilon}_{ir}\leq z-Z_i|Z_i)N_i(t)\}\\
&-\int_0^t E_a\{Y_i(u)\exp(Z_i^T\beta^*)P(\tilde{\epsilon}_{ir}\leq z-Z_i|Z_i)\}\lambda_0^*(u)du+o_{a.s.}(1)\\
=&\int_0^t\int_{-\infty}^{\infty}P(\epsilon_{ir}\leq z-a|a)E\{Y_i(u)|a\}\Big[\lambda(u;a)-\lambda_0^*(u)\exp(a^T\beta^*)\Big]dH(a)du+o_{a.s.}(1).
\end{aligned}
$$

Under the alternative, there usually exists some t and z, such that

$$\lambda(t;z)\neq\lambda_0^*(t)\exp(z^T\beta^*).$$

Thus, $n^{-1}\hat{W}_{nc}(t,z;\hat{\beta}_{nc})$ is nonzero for some t and z. Therefore, the S_{nc} test is usually consistent against the alternative.

Next, we study the scenario that the Cox model is correctly specified, whereas the measurement error model (4.4) is misspecified, and the underlying true error model has the following form

$$W_{ir}=g(Z_i)+\epsilon_{ir}^*,\quad i=1,\cdots,n;\quad r=1,\cdots,n_i,$$

where the deterministic function $g(\cdot)$ satisfies that $g(z)-x$ is not a constant vector for some value $z=(x^T,v^T)^T$ in the support of Z_i; the error terms ϵ_{ir}^* are independent and identically distributed with mean 0 and an unknown covariance matrix Σ_0, and are independent of $N_i(\cdot)$, $Y_i(\cdot)$, and Z_i. Furthermore, the X_i component of the true parameter $\beta_{0,x}\neq 0$.

Under this alternative, let β^* and $\lambda_0^*(t)$, E_a be defined analogue to those in the previous arguments. Write $\beta^* = (\beta_x^{*T}, \beta_v^{*T})^T$. Let $\tilde{\epsilon}_{ir}^* = (\epsilon_{ir}^T, 0^T)^T$. Let $\tilde{g}(\cdot) = (\tilde{g}^T(\cdot), 0^T)^T$. Since the Cox model is correctly specified, we have

$$
\begin{aligned}
& n^{-1}\hat{W}_{nc}(t, z; \hat{\beta}_{nc}) \\
=& n^{-1}\hat{W}_{nc}(t, z; \beta^*) + o_{a.s.}(1) \\
=& E_a\{I(\tilde{\epsilon}_{ir}^* \le z - \tilde{g}(Z_i))N_i(t)\} \\
& - \int_0^t \frac{E_a\{Y_i(u)\exp(\tilde{g}(Z_i)^T\beta^*)I(\tilde{\epsilon}_{ir}^* \le z - g(Z_i))\exp(\tilde{\epsilon}_{is}^{*T}\beta^*)\}}{E_a\{Y_i(u)\exp(\tilde{g}(Z_i)^T\beta^*)\}E_a\{\exp(\tilde{\epsilon}_{ir}^{*T}\beta^*)\}} dE_a\{N_i(u)\} + o_{a.s.}(1) \\
=& E_a\{P(\tilde{\epsilon}_{ir} \le z - \tilde{g}(Z_i)|Z_i)N_i(t)\} \\
& - \int_0^t \frac{E_a\{Y_i(u)\exp(\tilde{g}(Z_i)^T\beta^*)P(\tilde{\epsilon}_{ir} \le z - \tilde{g}(Z_i)|Z_i)\}}{E_a\{Y_i(u)\exp(\tilde{g}(Z_i)^T\beta^*)\}} E_a\{Y_i(u)\exp(Z_i^T\beta^*)\}\lambda_0(u)du \\
& + o_{a.s.}(1).
\end{aligned}
$$

Thus, $n^{-1}\hat{W}_{nc}(t, z; \hat{\beta}_{nc})$ is nonzero for some t and z. Therefore, the S_{nc} test is usually consistent against the alternative.

<u>*Consistency of S_c*</u>: Now we show that the S_c test is usually consistent against the alternative when the error model is correctly specified. Under this alternative, we have $\hat{\beta}_c \overset{a.s.}{\to} \beta^*$, and that

$$
\sum_{i=1}^n \int_0^t \frac{dN_i(u)}{\sum_{i=1}^n Y_i(u)\exp(Z_i^T\beta^*)} \overset{a.s.}{\to} \int_0^t \lambda_0^*(u)du,
$$

as $n \to \infty$. Let E_a and $\lambda(t; z)$ be defined as before. Write $\beta^* = (\beta_x^{*T}, \beta_v^{*T})^T$. Note that

$$
\begin{aligned}
& n^{-1}\hat{W}_c(t, z; \hat{\beta}_c) \\
=& n^{-1}\hat{W}_c(t, z; \beta^*) + o_{a.s.}(1) \\
=& E_a\{I(\epsilon_i \le x - \Sigma_0\beta_x^* - X_i)I(V_i \le v)N_i(t)\} \\
& - \int_0^t \frac{E_a\{Y_i(u)\exp(Z_i^T\beta^*)I(V_i \le v)I(\epsilon_i \le x - X_i)\exp(\epsilon_i^T\beta_x^*)\}}{E_a\{\exp(\epsilon_i^T\beta_x^*)\}} \lambda_0^*(u)du + o_{a.s.}(1).
\end{aligned}
$$

Note also that

$$
\begin{aligned}
&E_a\{I(\epsilon_i \leq x - \Sigma_0\beta_x^* - X_i)I(V_i \leq v)N_i(t)\} \\
=&E_a\{F_\epsilon(x - \Sigma_0\beta_x^* - X_i)I(V_i \leq v)N_i(t)\} \\
=&\int_0^t \int_{-\infty}^z E\{Y_i(u)|a_x\}F_\epsilon(x - \Sigma_0\beta_x^* - a_x)I(a_v \leq v)\lambda(u;a)dH(a)du,
\end{aligned}
$$

where we write $a = (a_x^T, a_v^T)^T$. Furthermore, analogue to the proof in Appendix A1, we have

$$
\begin{aligned}
&E_a\{Y_i(u)\exp(Z_i^T\beta^*)I(V_i \leq v)I(\epsilon_i \leq x - X_i)\exp(\epsilon_i^T\beta_x^*)\} \\
=&E_a\left\{Y_i(u)\exp(Z_i^T\beta^*)I(V_i \leq v)\int_{-\infty}^{x-X_i-\Sigma_0\beta_x^*}(2\pi)^{-\frac{p}{2}}|\Sigma_0|^{-\frac{1}{2}}\exp(\frac{1}{2}\beta_x^{*T}\Sigma_0^{-1}\beta_x^*)\exp(-\frac{1}{2}\epsilon^T\Sigma_0^{-1}\epsilon)d\epsilon\right\} \\
=&\exp(\frac{1}{2}\beta_x^{*T}\Sigma_0^{-1}\beta_x^*)E_a\left\{Y_i(u)F_\epsilon(x - \Sigma_0\beta_x^* - X_i)I(V_i \leq v)\exp(Z_i^T\beta^*)\right\} \\
=&\exp(\frac{1}{2}\beta_x^{*T}\Sigma_0^{-1}\beta_x^*)\int_{-\infty}^z E\{Y_i(u)|a_x\}F_\epsilon(x - \Sigma_0\beta_x^* - a_x)I(a_v \leq v)\exp(a^T\beta^*)dH(a).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&n^{-1}\hat{W}_c(t, z; \hat{\beta}_c) \\
=&\int_0^t \int_{-\infty}^z E\{Y_i(u)|a_x\}F_\epsilon(x - \Sigma_0\beta_x^* - a_x)I(a_v \leq v)\lambda(u;a)dH(a)du \\
&-\int_0^t \int_{-\infty}^z E\{Y_i(u)|a_x\}F_\epsilon(x - \Sigma_0\beta_x^* - a_x)I(a_v \leq v)\exp(a^T\beta^*)\lambda_0^*(u)dH(a)du + o_{a.s.}(1) \\
=&\int_0^t \int_{-\infty}^z E\{Y_i(u)|a_x\}F_\epsilon(x - \Sigma_0\beta_x^* - a_x)I(a_v \leq v)\left\{\lambda(u;a) - \lambda_0^*(u)\exp(a^T\beta^*)\right\}dH(a)du \\
&+ o_{a.s.}(1).
\end{aligned}
$$

Under the alternative, there usually exists some t and z, such that

$$\lambda(t;z) \neq \lambda_0^*(t)\exp(z^T\beta^*).$$

Thus, $n^{-1}\hat{W}_c(t, z; \hat{\beta}_c)$ is nonzero for some t and z. Therefore, the S_c test is usually consistent against the alternative.

Chapter 5

A Class of Functional Methods for Error-Contaminated Survival Data under Additive Hazards Models with Replicate Measurements

5.1 Introduction

Covariate measurement error has long been a concern in survival analysis, and it has attracted extensive research interest. Since Prentice (1982), a large number of inference methods have been developed to handle error-prone data (e.g., Nakamura 1992, Buzas 1998, Hu, Tsiatis and Davidian 1998, Huang and Wang 2000, Li and Lin 2003, Hu and Lin 2004, Song and Huang 2005, Yi and Lawless 2007, and Zucker and Spiegelman 2008). Although discussion on survival data with measurement error is not restricted to a single type of model, proportional hazards models have been the center of existing research. The impact of covariate error is well understood for proportional hazards models.

Proportional hazards models (Cox 1972) specify that covariates have multiplicative effects on the hazard ratio; a most appeal of such models is that baseline hazard functions can be left unspecified when conducting inference about covariate effects based on partial likelihood functions. In contrast to proportional hazards models, additive hazards models offer a flexible tool to delineate survival data (Breslow and Day 1980, Cox and Oakes 1984). Lin and Ying (1994) developed an inference method for covariate effects based on pseudo-score functions, and a key catch of this method is that the baseline hazard function is left unmodeled. Furthermore, this method allows a close form of the estimator of regression parameters.

Relative to a large body of literature on proportional hazards models with covariate measurement error, there is little research on measurement error effects under additive hazards models, although several authors investigated this problem. Sun, Zhang and Sun (2006) considered additive hazards models for the case with replicates of mismeasured covariates, and justified asymptotic results using empirical processes theory. Kulich and Lin (2000) proposed an unbiased corrected pseudo score approach for the case that a validation sample is available. However, a number of important questions remain unexplored. For instance, as indicated by the work for proportional hazards models, many correction methods can be developed to account for error effects. Are there any intrinsic connections among those methods? How do we assess the validity of the proposed methods which essentially rely on a correct model specification? Does measurement error in covariates have the same effects on additive hazards models as those for proportional hazards models? Can we reveal new insights by exemplifying the unique features of additive hazards models?

In this chapter we examine these important questions. In particular, we explore asymptotic bias induced in the naive analysis with measurement error ignored. To correct for the induced bias, we develop a class of correction methods to exemplify the unique features of additive hazards models. Our methods do not impose any distributional assumptions on the true covariates, thus appealing in protecting us from the risk of misspecifying the covariate model. The validity of the proposed methods is carefully examined, and we investigate issues of model checking and model misspecification. Theoretical results are rigorously established, and are complemented with various numerical assessments. In addition, different from the most work which assumes classical additive error models with error distributions specified, in this chapter we relax the requirement of specifying a full distributional assumption for error terms. With availability of replicated measurements, we consider a flexible model for measurement error processes which assumes only an additive structure. Moreover, we employ the functional modeling approach for which the distribution of the true covariates is left unmodeled.

The remainder is organized as follows. In Section 5.2, we introduce the basic model setup and estimation in the absence of measurement error. In Section 5.3, we conduct a bias study for the naive estimator which ignores covariate measurement error. In Section 5.4, we propose an approach based on pseudo score functions, to deal with survival data with replicates of mismeasured covariates. Asymptotic results are established. In Section 5.5, we propose an estimating equation based method. In Section 5.6, numerical studies for the estimators are provided. In Section 5.7, we study the impact of model misspecification and propose a goodness of fit test statistic. In Section 5.8, a real data example is provided. Concluding discussion is provided in the last section.

5.2 Notation and Model Setup

5.2.1 Additive Hazards Model

For $i = 1, \cdots, n$, let T_i be the failure time, C_i be the censoring time, and $Z_i(t) = (X_i^T, V_i^T(t))^T$ be a vector of covariates, where X_i is a $p \times 1$ vector of time-independent but error-prone covariates, and $V_i(t)$ is a $q \times 1$ vector of covariates that are precisely measured and possibly time-dependent. As common in practise, $V_i(t)$ are assumed to be external covariates (Kalbfleisch and Prentice 2002, p.197). We consider that the hazard function of T_i is related to $Z_i(\cdot)$ through the additive hazards model

$$\lambda(t; Z_i(t)) = \lambda_0(t) + \beta^T Z_i(t) = \lambda_0(t) + \beta_x^T X_i + \beta_v^T V_i(t), \tag{5.1}$$

where $\lambda_0(\cdot)$ is an unspecified baseline hazard function, and $\beta = (\beta_x^T, \beta_v^T)^T$ is a vector of unknown regression parameters. Let $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ be the baseline cumulative hazard function. We assume that the failure time T_i is continuous and $\Lambda_0(t)$ is absolutely continuous. T_i and C_i are assumed to be conditionally independent given $Z_i(t)$.

Suppose individuals are observed over a common time interval $[0, \tau]$, where τ is a positive constant. Let $S_i = \min(T_i, C_i, \tau)$, $\delta_i = I(T_i \leq \min\{C_i, \tau\})$, $N_i(t) = I(S_i \leq t, \delta_i = 1)$, and $Y_i(t) = I(S_i \geq t)$.

5.2.2 Estimation in the Absence of Measurement Error

If X_i were precisely measured, then estimation of β can be carried out using the pseudo score functions proposed by Lin and Ying (1994):

$$U(\beta) = \sum_{i=1}^{n} \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} d\{N_i(t) - Y_i(t)\beta^T Z_i(t)dt\}, \tag{5.2}$$

where $\bar{Z}(t) = \sum_{i=1}^n Y_i(t)Z_i(t)/\sum_{i=1}^n Y_i(t)$. Solving $U(\beta) = 0$ gives an estimator of β, which has an explicit form given by

$$\hat{\beta} = \left[\sum_{i=1}^{n} \int_0^\tau Y_i(t)\{Z_i(t) - \bar{Z}(t)\}^{\otimes 2} dt\right]^{-1} \left[\sum_{i=1}^{n} \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} dN_i(t)\right]. \tag{5.3}$$

This estimator is consistent, provided certain regularity conditions hold. Indeed, $U(\beta)$ can be equivalently written as

$$U(\beta) = \sum_{i=1}^{n} \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} dM_i(t; \beta, \Lambda_0),$$

where $M_i(t; \beta, \Lambda_0) = N_i(t) - \int_0^t Y_i(u)\{d\Lambda_0(u) + \beta^T Z_i(u)du\}$. Let $\mathcal{F}_t = \sigma\{N_i(s), Y_i(s), Z_i(s), 0 \leq s \leq t, i = 1, \cdots, n\}$ be the $\sigma-$ field generated by the event, covariates, and observation histories prior to time t for all subjects. Then $M_i(t; \beta, \Lambda_0)$ is an $\mathcal{F}_t$- adapted martingale (e.g., Kalbfleisch and Prentice 2002, Sec. 5.3). Consequently, $E\{U(\beta)\} = 0$, i.e., $U(\beta)$

are unbiased estimating functions of β. By estimating function theory, under regularity conditions, solving $U(\beta) = 0$ leads to a consistent estimator of β (e.g., Yanagimoto and Yamamoto 1991).

Noting that $E\{M_i(t; \beta, \Lambda_0)\}$=0 by the martingale property of $M_i(t; \beta, \Lambda_0)$, we estimate the baseline cumulative hazard function by solving $\sum_{i=1}^{n} M_i(t; \beta, \Lambda_0) = 0$. That is, $\Lambda_0(t)$ is estimated by

$$\hat{\Lambda}_0(t; \hat{\beta}) = \int_0^t \frac{\sum_{i=1}^{n} dN_i(u)}{\sum_{j=1}^{n} Y_j(u)} - \int_0^t \frac{\sum_{i=1}^{n} Y_i(u)\hat{\beta}^T Z_i(u)du}{\sum_{j=1}^{n} Y_j(u)}.$$

5.2.3 Measurement Error Model

Suppose X_i is repeatedly measured n_i times, resulting in the surrogates $W_{ir}, r = 1, \cdots, n_i$. Given $\mathcal{F}_t$ for any time t, we assume that

$$W_{ir} = X_i + \epsilon_{ir}, \tag{5.4}$$

where the ϵ_{ir} are independent and identically distributed (i.i.d.) with mean 0 and a positive-definite variance matrix Σ_0, $i = 1, \cdots, n; r = 1, \cdots, n_i$. This assumption says that given the true covariates $Z_i(t)$ at any time t, T_i and C_i are independent of surrogate measurements W_{ij}. This assumption is analogous to the usual nondifferential error mechanism for uncensored data (Carroll et al. 2006, p.36).

Let $\bar{W}_{i\cdot} = \sum_{r=1}^{n_i} W_{ir}/n_i$, and $\hat{Z}_i(t) = (\bar{W}_{i\cdot}^T, V_i^T(t))^T$. Then $E\{\hat{Z}_i(t)|\mathcal{F}_t\} = Z_i(t)$, and $E\{\hat{Z}_i^{\otimes 2}(t)|\mathcal{F}_t\} = Z_i^{\otimes 2}(t) + \Sigma_1/n_i$, where $a^{\otimes 2} = aa^T$ for a column vector a, $\Sigma_1 = \text{diag}(\Sigma_0, 0_q)$, 0_q is the $q \times q$ matrix of elements 0, and q is the dimension of $V_i(t)$. With the replicates W_{ir}, we estimate the covariance matrix Σ_0 by

$$\hat{\Sigma}_0 = \sum_{i=1}^{n} \sum_{r=1}^{n_i} (W_{ir} - \bar{W}_{i\cdot})^{\otimes 2} / \sum_{i=1}^{n} (n_i - 1).$$

Let $\hat{\Sigma}_1 = \text{diag}(\hat{\Sigma}_0, 0_q)$, then $E(\hat{\Sigma}_1|\mathcal{F}_t) = \Sigma_1$ for any time t.

5.3 Asymptotic Bias Analysis

We investigate measurement error effects on the structure of the hazard function. We derive the hazard function based on the observed covariates $(\bar{W}_{i\cdot}^T, V_i^T(t))^T$, and let $\lambda^*(t; \bar{W}_{i\cdot}, V_i(t))$ denote this hazard function. With the assumption made on the measurement error process,

$$\begin{aligned}\lambda^*(t; \bar{W}_{i\cdot}, V_i(t)) &= E\{\lambda(t; X_i, V_i(t))|T_i \geq t, \bar{W}_{i\cdot}, V_i(t)\} \\ &= \lambda_0(t) + \beta_x^T E\{X_i|T_i \geq t, \bar{W}_{i\cdot}, V_i(t)\} + \beta_v^T V_i(t). \qquad (5.5)\end{aligned}$$

The expression (5.5) indicates that the hazard function for the observed covariates retains the additive structure while the risk difference has a more complicated form than (5.1). Since the conditional expectation $E\{X_i|T_i \geq t, \bar{W}_{i\cdot}, V_i(t)\}$ generally differs from $\bar{W}_{i\cdot}$, (5.5) suggests that the naive analysis with $\bar{W}_{i\cdot}$ replacing X_i would lead to biased results.

We now quantify the asymptotic bias resulted from the naive analysis. Let $\tilde{Z}(t) = \sum_{i=1}^n Y_i(t)\hat{Z}_i(t)/\sum_{i=1}^n Y_i(t)$. Define

$$U_{nv}(\beta) = \sum_{i=1}^n \int_0^\tau \{\hat{Z}_i(t) - \tilde{Z}(t)\}\{dN_i(t) - Y_i(t)\hat{Z}_i^T(t)\beta dt\}.$$

That is, $U_{nv}(\beta)$ is the naive pseudo score function that is obtained from replacing $E\{X_i|T_i \geq t, \bar{W}_{i\cdot}, V_i(t)\}$ with $\bar{W}_{i\cdot}$ in (5.5), and then applying the pseudo score function form (5.2) to the observed data. Let $\hat{\beta}_{nv}$ be the solution of $U_{nv}(\beta) = 0$.

Let $\rho_0 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n n_i^{-1}$, and $e(t) = E\{Y_i(t)Z_i(t)\}/E\{Y_i(t)\}$. Define $B_1 = \int_0^\tau E\left[Y_i(t)\{Z_i(t) - e(t)\}^{\otimes 2}\right] dt$, and $B_2 = \rho_0\Sigma_1 \int_0^\tau E\{Y_i(t)\}dt$. Following the discussion of Yi and Reid (2010), we can show that under certain regularity conditions, $\hat{\beta}_{nv}$ converges in probability to a limit, say β_{nv}^*, as $n \to \infty$. We can further characterize the relationship between β and β_{nv}^*, given by

$$\beta_{nv}^* = (B_1 + B_2)^{-1} B_1 \beta. \qquad (5.6)$$

The details are presented in Appendix A2 of the Supplementary Material.

It is immediate that from (5.6), if $\|\beta\| = 0$, then $\|\beta^*_{nv}\| = 0$, where $\|\cdot\|$ is the Euclidean norm. If $Z_i(t)$ contains only a univariate X_i, then $|\beta^*_{nv}| < |\beta|$, suggesting an attenuated measurement error effect. If X_i and $V_i(t)$ are univariates and are independent, and either $V_i(t)$ or X_i are independent of the followup process, then $|\beta^*_{nv,x}| < |\beta_x|$ and $\beta^*_{nv,v} = \beta_v$, where β_x and β_v (or $\beta^*_{nv,x}$ and $\beta^*_{nv,v}$) are components of β (or β^*_{nv}) corresponding to the covariates X_i and $V_i(t)$, respectively. The justifications are provided in Appendix A2 of the Supplementary Material.

In the following, we numerically evaluate the asymptotic bias of the naive estimator with measurement error ignored in estimation procedures. Suppose the failure times T_i are generated from the additive hazards model

$$\lambda(t; X_i) = \lambda_0(t) + X_i\beta,$$

where the baseline hazard function is set as $\lambda_0(t) = 1$, X_i is a univariate variable generated from $UNIF(-1, 1)$, and the true parameter value is set as $\beta = 1$. The censoring times C_i are simulated from $UNIF(0, 4.2)$, leading to about 30% censoring rate. The error model (5.4) is used to generate W_{ir} where $\epsilon_{ir} \sim N(0, \sigma^2)$ for $r = 1, \cdots, n_i, i = 1, \cdots, n$. We consider settings where σ^2 varies from 0 to $Var(X_i)$ (which is 1/3), and $n_i = 1, 2, 4, 8$. We are interested in the asymptotic bias of β^*_{nv} relative to the true parameter value β, defined as $(\beta^*_{nv} - \beta)/\beta$. In Figure 5.1, we plot the asymptotic relative bias of β^*_{nv} against σ.

It is seen that the naive estimator is attenuated to the null as the degree of measurement error becomes large. Furthermore, the degree of attenuation decreases when the number of replicated surrogate measurements increases. These results confirm the above theoretical findings.

[Insert Figure 5.1 here!]

5.4 Corrected Pseudo Score Approach

As shown in Section 5.3, the naive analysis with measurement error ignored yields biased estimation of β. We now develop an inference method for β with measurement error effects taken into account. The idea is to find sensible estimating functions of β which satisfy two key conditions: (1) estimating functions must be computable in the sense of being expressed in terms of the observed data and parameters, and (2) estimating functions are unbiased. By estimating function theory, solving the resulting estimating equations leads to a consistent estimator of β if suitable regularity conditions hold.

Using the pseudo score functions (5.2) with X_i replaced by $\bar{W}_{i\cdot}$ gives us computable estimating functions, $U_{nv}(\beta)$, of β. But as implied by the discussion in Section 5.3, these estimating functions $U_{nv}(\beta)$ are not unbiased. As suggested by Yi and Reid (2010), a quick remedy to fixing this is to modify $U_{nv}(\beta)$ by subtracting their expectation $E\{U_{nv}(\beta)\}$ so that the resulting estimating functions, $U_{nv}(\beta)-E\{U_{nv}(\beta)\}$, are unbiased. However, evaluation of $E\{U_{nv}(\beta)\}$ is generally complicated due to the involvement of the joint distribution of the survival, censoring, and covariate processes, thus making the modified estimating functions $U_{nv}(\beta) - E\{U_{nv}(\beta)\}$ unappealing. To get around this problem, we alternatively evaluate the conditional expectation of $U_{nv}(\beta)$, given $\mathcal{F}_\tau$. As shown in Appendix A3 of the Supplementary Material, $E\{U_{nv}(\beta)|\mathcal{F}_\tau\} = U(\beta) - \int_0^\tau \left\{1 - \frac{1}{\sum_{j=1}^n Y_j(t)}\right\} \sum_{i=1}^n \{Y_i(t)\Sigma_1\beta/n_i\}dt$. This identity motivates us to consider corrected pseudo score functions:

$$\tilde{U}_c(\beta) = U_{nv}(\beta) + \int_0^\tau \left\{1 - \frac{1}{\sum_{j=1}^n Y_j(t)}\right\} \sum_{i=1}^n \{Y_i(t)\Sigma_1\beta/n_i\}\, dt.$$

By that $E\{U(\beta)\} = 0$, we obtain that $E\{\tilde{U}_c(\beta)\} = 0$, implying that $\tilde{U}_c(\beta)$ are unbiased estimating functions.

To use the corrected pseudo score function $\tilde{U}_c(\beta)$ to estimate β, we need to replace Σ_1 with its consistent estimate $\hat{\Sigma}_1$, and let $U_c(\beta)$ denote the resultant estimating functions. One might expect that the substitution of $\hat{\Sigma}_1$ for Σ_1 would break down the unbiasedness of $\tilde{U}_c(\beta)$, but this is not the case here. Because $E(\hat{\Sigma}_1|\mathcal{F}_t) = \Sigma_1$ for any time t, it follows that

$$
\begin{aligned}
E\{U_c(\beta)\} &= E\{U_{nv}(\beta)\} + E\left(\int_0^\tau E\left[\left\{1-\frac{1}{\sum_{j=1}^n Y_j(t)}\right\}\sum_{i=1}^n \left\{Y_i(t)\hat{\Sigma}_1\beta/n_i\right\} dt \middle| \mathcal{F}_t\right]\right)\\
&= E\{U_{nv}(\beta)\} + E\left[\int_0^\tau \left\{1-\frac{1}{\sum_{j=1}^n Y_j(t)}\right\}\sum_{i=1}^n \{Y_i(t)\Sigma_1\beta/n_i\} dt\right]\\
&= E\{\tilde{U}_c(\beta)\}.
\end{aligned}
$$

Therefore, $U_c(\beta)$ are unbiased estimating functions due to that $E\{\tilde{U}_c(\beta)\} = 0$.

Let $\hat{\beta}_c$ be the solution to the equations $U_c(\beta) = 0$. It is seen that

$$
\begin{aligned}
\hat{\beta}_c &= \left[\sum_{i=1}^n \int_0^\tau Y_i(t)\left\{\hat{Z}_i(t)-\tilde{Z}(t)\right\}^{\otimes 2} dt - \int_0^\tau \left\{1-\frac{1}{\sum_{j=1}^n Y_j(t)}\right\}\sum_{i=1}^n \left\{Y_i(t)\hat{\Sigma}_1/n_i\right\} dt\right]^{-1}\\
&\quad\times\left[\sum_{i=1}^n \int_0^\tau \left\{\hat{Z}_i(t)-\tilde{Z}(t)\right\} dN_i(t)\right].
\end{aligned}
$$

We comment that numerically, $\hat{\beta}_c$ performs stably. This can be explained by that the inverse matrix (scaled by n^{-1}) in $\hat{\beta}_c$ converges almost surely to a positive definite matrix under mild regularity conditions, thus singularity does not occur in the asymptotic sense. Details on this point are included in Lemma A.1 of Appendix A4 in the Supplementary Material.

Next, we discuss estimation of the baseline cumulative hazard function $\Lambda_0(t)$. Let $\tilde{M}_i(t;\beta,\Lambda_0) = N_i(t) - \int_0^t Y_i(u)\{d\Lambda_0(u) + \beta^T\hat{Z}_i(u)du\}$. After some algebra, $E\{\tilde{M}_i(t;\beta,\Lambda_0)\} = 0$. Solving $\sum_{i=1}^n \tilde{M}_i(t;\beta,\Lambda_0) = 0$ (Lin and Ying 1994) leads to an estimator of $\Lambda_0(t)$, say $\hat{\Lambda}_0(t;\hat{\beta}_c)$, given by

$$
\hat{\Lambda}_0(t;\hat{\beta}_c) = \int_0^t \left\{\frac{\sum_{i=1}^n dN_i(u)}{\sum_{j=1}^n Y_j(u)}\right\} - \int_0^t \hat{\beta}_c^T\tilde{Z}(u)du.
$$

To ensure monotonicity, we propose to use $\tilde{\Lambda}_0(t;\hat{\beta}_c) = \max_{0\le s\le t}\hat{\Lambda}_0(s;\hat{\beta}_c)$ to estimate $\Lambda_0(t)$ as in Lin and Ying (1994). Asymptotic properties of $\hat{\beta}_c$ and $\hat{\Lambda}_0(t;\hat{\beta}_c)$ are summarized in the following theorems, whose proofs are included in Appendices A4 and A5 of the Supplementary Material.

Let $\rho_1 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}(n_i - 1)$, and $\Sigma_2 = \text{diag}(\rho_0\rho_1^{-1}\sum_{r=1}^{n_i}(W_{ir} - \bar{W}_{i\cdot})^{\otimes 2}E(S_i), 0)$ be the dimension $(p+q)\times(p+q)$ block diagonal matrix. Define

$$\mathcal{D}_c = \lim_{n\to\infty} n^{-1}E\left[\sum_{i=1}^{n}\int_0^\tau Y_i(t)\left\{\hat{Z}_i(t) - e(t)\right\}^{\otimes 2}dt\right] - \rho_0 E(S_i)\Sigma_1,$$

$$\text{and}\ \ \Sigma_c = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}E\left[\int_0^\tau \left\{\hat{Z}_i(t) - e(t)\right\}d\tilde{M}_i(t;\beta,\Lambda_0) + \Sigma_2\beta - \rho_0 E(S_i)\Sigma_1\beta + \frac{S_i\Sigma_1\beta}{n_i}\right]^{\otimes 2}.$$

Theorem 1 *Under Regularity Conditions R1-R8 listed in Appendix A1 of the Supplementary Material, we have*

$$n^{1/2}(\hat{\beta}_c - \beta) \xrightarrow{d} N(0, \mathcal{D}_c^{-1}\Sigma_c\mathcal{D}_c^{-T}), \quad as \ \ n\to\infty. \tag{5.7}$$

Theorem 2 *Under Regularity Conditions R1-R8 listed in Appendix A1 of the Supplementary Material, we have*

$$n^{1/2}\{\hat{\Lambda}_0(t;\hat{\beta}_c) - \Lambda_0(t)\} \rightsquigarrow \mathcal{G}(t) \quad in \ \ l^\infty[0,\tau] \quad as \ \ n\to\infty, \tag{5.8}$$

where $\rightsquigarrow$ means weak convergence, $l^\infty[0,\tau]$ is the space of all bounded functions on $[0,\tau]$ (van der Vaart and Wellner 1996), $\mathcal{G}(t)$ is a zero-mean Gaussian process with covariance function $\Phi(s,t) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}E[\Psi_i(s)\Psi_i(t)]$ for time points s and t, and

$$\Psi_i(t) = \int_0^t \frac{d\tilde{M}_i(u;\beta,\Lambda_0)}{E[Y_i(u)]} - \int_0^t e^T(u)du\mathcal{D}_c^{-1}\left[\int_0^\tau \left\{\hat{Z}_i(t) - e(t)\right\}d\tilde{M}_i(t;\beta,\Lambda_0)\right.$$
$$\left. + \Sigma_2\beta - \rho_0 E(S_i)\Sigma_1\beta + \frac{S_i\Sigma_1\beta}{n_i}\right].$$

We comment that as seen from the proofs of Theorems 1 and 2, the term $\Sigma_2\beta - \rho_0 E(S_i)\Sigma_1\beta$ in Σ_c and $\Psi_i(t)$ can be interpreted as the substitution effect of replacing Σ_1 with its consistent estimate $\hat{\Sigma}_1$. If there are no replicate measurements, i.e., $n_i = 1, i = 1, \cdots, n$, and Σ_1 is simply known, then the asymptotic results of $\hat{\beta}_c$ and $\hat{\Lambda}_0(t; \hat{\beta}_c)$ are given by Theorems 1 and 2 with the term $\Sigma_2\beta - \rho_0 E(S_i)\Sigma_1\beta$ removed from Σ_c and $\Psi_i(t)$. More details are included in Appendix A6 of the Supplementary Material.

Theorem 2 implies that $\Pr\{\sup_{0\le t\le\tau} n^{1/2}|\hat{\Lambda}_0(t; \hat{\beta}_c) - \Lambda_0(t)| \le x\} \to \Pr\{\sup_{0\le t\le\tau} |\mathcal{G}(t)| \le x\}$ as $n \to \infty$ for any $x \ge 0$. It is difficult to use this result to conduct inference about $\Lambda_0(t)$ due to that the Gaussian process $\mathcal{G}(t)$ does not have the independent increment property and has a complex form.

To get around this difficulty, we suggest using resampling techniques to construct confidence bands for survival curves. Let $\hat{\Psi}_i(t)$ be $\Psi_i(t)$ with $\beta, \Lambda_0(t), \Sigma_1, E\{Y_i(t)\}, e(t)$ replaced with $\hat{\beta}_c, \hat{\Lambda}_0(t; \hat{\beta}_c), \hat{\Sigma}_1, \bar{Y}.(t), \tilde{Z}(t)$, respectively. Define

$$\hat{W}_n(t) = n^{-1/2} \sum_{i=1}^{n} \xi_i \hat{\Psi}_i(t),$$

where $\xi_i, i = 1, \cdots, n$ are i.i.d. standard normal random variables, and are independent of the data.

Theorem 3 *Under Regularity Conditions R1-R8 listed in Appendix A1 of the Supplementary Material, when conditional on the observed data* $\{N_i(t), Y_i(t), \hat{Z}_{ir}(t), t \in [0, \tau], i = 1, \cdots, n, r = 1, \cdots, n_i\}$, $\hat{W}_n(t)$ *converges weakly to* $\mathcal{G}(t)$ *in* $l^\infty[0, \tau]$ *in probability as* $n \to \infty$.

The proof of Theorem 3 is deferred to Appendix A7 of the Supplementary Material. This theorem suggests that we can legitimately use the distribution of $\sup_{0\le t\le\tau} |\hat{W}_n(t)|$ to approximate that of $\sup_{0\le t\le\tau} |\mathcal{G}(t)|$, and thus that of $\sup_{0\le t\le\tau} n^{1/2}|\hat{\Lambda}_0(t; \hat{\beta}_c) - \Lambda_0(t)|$.

To construct an approximate $(1 - \alpha)$ confidence band $(\hat{\Lambda}_0(t; \hat{\beta}_c) - n^{-1/2}q_\alpha, \hat{\Lambda}_0(t; \hat{\beta}_c) + n^{-1/2}q_\alpha)$ for $\Lambda_0(t)$ over $[0, \tau]$, we first repeatedly generate a set of $\{\xi_i, i = 1, \cdots, n\}$ independently from the standard normal distribution for a large number of times, say 1000, and calculate $\sup_{0\leq t\leq\tau}|\hat{W}_n(t)|$ for each time. Thus, we obtain an empirical quantile $\tilde{q}_\alpha$, and replace q_α with $\tilde{q}_\alpha$ to obtain an approximate $(1-\alpha)$ confidence band of $\Lambda_0(t)$ as $(\hat{\Lambda}_0(t;\hat{\beta}_c) - n^{-1/2}\tilde{q}_\alpha, \hat{\Lambda}_0(t;\hat{\beta}_c) + n^{-1/2}\tilde{q}_\alpha)$.

5.5 Estimating Equation Approach

Instead of focusing on the pseudo score function $U(\beta)$ alone as in the previous section, we now jointly look at unbiased estimating equations for β and $\Lambda_0(\cdot)$. Our starting point is the fact that $M_i(t;\beta,\Lambda_0)$ is an $\mathcal{F}_t$ - adapted martingale, which implies that $M_i(t;\beta,\Lambda_0)$ is a mean-zero process with

$$E\{dM_i(t;\beta,\Lambda_0)|\mathcal{F}_{t^-}\} = 0 \tag{5.9}$$

for all $0 \leq t \leq \tau$. Since $Z_i(t)$ is external, we have $E\{Z_i(t)dM_i(t;\beta,\Lambda_0)|\mathcal{F}_{t^-}\} = 0$, and furthermore

$$E\left\{\int_0^t Z_i(u)dM_i(u;\beta,\Lambda_0)\right\} = 0. \tag{5.10}$$

These results suggest that $dM_i(t;\beta,\Lambda_0)$ and $Z_i(t)dM_i(t;\beta,\Lambda_0)$ can be used to construct unbiased estimating functions for $\Lambda_0(\cdot)$ and β if X_i were error-free. As $M_i(t;\beta,\Lambda_0)$ contains unobserved covariates X_i, it is tempting to substitute $Z_i(t)$ with observed $\hat{Z}_i(t)$. By the error model (5.4), it is easily seen that this replacement does not change the property (5.9), but it breaks down (5.10). That is, $E\{\tilde{M}_i(t;\beta,\Lambda_0)\} = 0$, but $E\{\hat{Z}_i(t)d\tilde{M}_i(t;\beta,\Lambda_0)|\mathcal{F}_{t^-}\} \neq 0$. In fact,

$$E\{\hat{Z}_i(t)d\tilde{M}_i(t;\beta,\Lambda_0)|\mathcal{F}_{t^-}\} = Z_i(t)dM_i(t;\beta,\Lambda_0) - Y_i(t)\Sigma_1\beta/n_i dt.$$

Hence, we construct two sets of unbiased estimating equations:

$$U_{e1}(t;\beta,\Lambda_0) = \sum_{i=1}^n \tilde{M}_i(t,\beta,\Lambda_0) = 0; \tag{5.11}$$

$$U_{e2}(t;\beta,\Lambda_0) = \sum_{i=1}^n \int_0^t \hat{Z}_i(u)d\tilde{M}_i(u;\beta,\Lambda_0) + \sum_{i=1}^n \int_0^t Y_i(u)\hat{\Sigma}_1\beta/n_i du = 0. \tag{5.12}$$

Now, we need to investigate whether (5.11) and (5.12) are adequate for estimating β (a finite dimensional parameter) and $\Lambda_0(t)$ (a function). Since the function $\Lambda_0(t)$ can be regarded as an infinite dimensional parameter, the usual estimating equation theory does not guarantee that solving (5.11) and (5.12) simultaneously leads to appropriate estimators. For example, given an arbitrary estimator of $\Lambda_0(t)$, say $\hat{\Lambda}_0(t)$, which satisfies both (5.11) and (5.12), $\hat{\Lambda}_0(t) + C$ would also satisfy (5.11) and (5.12) for any constant C, yielding an unidentifiability issue. To resolve this problem, we adopt an *ad hoc* procedure, which shares the same spirit as that of Lin and Ying (1994), and thus identifiability can be achieved.

Note that (5.11) leads to $\sum_{i=1}^n Y_i(t) d\Lambda_0(t) = \sum_{i=1}^n dN_i(t) - \sum_{i=1}^n Y_i(t)\hat{Z}_i^T(t)\beta dt$. Hence, given any fixed β, we estimate $\Lambda_0(t)$ by $\hat{\Lambda}_0(t;\beta) = \int_0^t \sum_{i=1}^n dN_i(u) / \sum_{j=1}^n Y_j(u) - \int_0^t \tilde{Z}^T(u)\beta du$. Substituting $\hat{\Lambda}_0(t;\beta)$ into (5.12), we obtain

$$\hat{\beta}_e = \left[\sum_{i=1}^n \int_0^\tau Y_i(t)\left\{\hat{Z}_i(t) - \tilde{Z}(t)\right\}^{\otimes 2} dt - \sum_{i=1}^n \int_0^\tau Y_i(t)\hat{\Sigma}_1/n_i dt\right]^{-1} \left[\sum_{i=1}^n \int_0^\tau \left\{\hat{Z}_i(t) - \tilde{Z}(t)\right\} dN_i(t)\right]$$

as an estimator of β, where we set $t = \tau$ to fully use the whole data set.

Plugging $\hat{\beta}_e$ back into $\hat{\Lambda}_0(t;\beta)$, we obtain an estimator for the baseline cumulative hazard function

$$\hat{\Lambda}_0(t;\hat{\beta}_e) = \int_0^t \left\{\frac{\sum_{i=1}^n dN_i(u)}{\sum_{j=1}^n Y_j(u)}\right\} - \int_0^t \hat{\beta}_e^T \tilde{Z}(u) du.$$

It is interesting to note that $\hat{\beta}_e$ differs from $\hat{\beta}_c$ by a factor $\{1 - 1/\sum_{j=1}^n Y_j(t)\}$, and $\hat{\Lambda}_0(t;\hat{\beta}_e)$ and $\hat{\Lambda}_0(t;\hat{\beta}_c)$ assume the same form but with a different estimator $\hat{\beta}_e$ or $\hat{\beta}_c$. In the following corollary, we show that asymptotically $\hat{\beta}_e$ behaves the same as $\hat{\beta}_c$, and $\hat{\Lambda}_0(t;\hat{\beta}_e)$ behaves the same as $\hat{\Lambda}_0(t;\hat{\beta}_c)$. The proof is sketched in Appendix A8 of the Supplementary Material.

Corollary 1 *Under Regularity Conditions R1-R8 listed in Appendix A1 of the Supplementary Material, we have*

$$\begin{aligned} n^{1/2}(\hat{\beta}_e - \beta) \quad &\xrightarrow{d} \quad N(0, \mathcal{D}_c^{-1}\Sigma_c\mathcal{D}_c^{-T}), \quad as \quad n \to \infty, \\ and \quad n^{1/2}\{\hat{\Lambda}_0(t;\hat{\beta}_e) - \Lambda_0(t)\} \quad &\rightsquigarrow \quad \mathcal{G}(t) \quad in \quad l^\infty[0,\tau] \quad as \quad n \to \infty, \end{aligned}$$

where $\mathcal{G}(t)$ is the Gaussian process defined in Theorem 2.

5.6 Empirical Studies

We conduct various simulation studies to evaluate the finite sample performance of the proposed estimators. In particular, we contrast our proposed estimators $\hat{\beta}_c$ and $\hat{\beta}_e$ to the naive estimator $\hat{\beta}_{nv}$, the regression calibration estimator $\hat{\beta}_{rc}$ (Prentice 1982), and the estimator by Sun, Zhang and Sun (2006) which is denoted as $\hat{\beta}_{szs}$.

5.6.1 Design of Simulation

We consider $n = 200$ and generate 1000 simulations for each parameter configuration. We examine three scenarios for the bivariate time-independent covariates $Z_i = (X_i, V_i)^T$. In Scenario 1, the covariates X_i and V_i are independently generated, where $X_i \sim UNIF(-1, 1)$, and V_i is a binary variable taking value 1 or 0 each with probability 0.5. Scenarios 2 and 3 correspond to that covariates X_i and V_i are correlated. In Scenario 2, $Z_i = (X_i, V_i)^T$ is uniformly generated from the triangular $\{(X_i, V_i) : -1 \le X_i \le 1, -1 \le V_i \le 1, X_i + V_i \le 0\}$, while in Scenario 3, $X_i \sim EXP(1)$ and

$$\Pr(V_i = 1|X_i) = \frac{\exp(X_i)}{1 + \exp(X_i)}.$$

Survival times are independently generated using the additive hazards model (5.1), where we take the baseline hazard function to be $\lambda_0(t) = \alpha\gamma t^{\gamma-1}$, and we consider $\alpha = \gamma = 1$ for Scenarios 1 and 2 and $\alpha = 0.5, \gamma = 2$ for Scenario 3, respectively. The true values of β_x and β_v are set to be $(\beta_x, \beta_v) = (1, 0)$ for Scenario 1, and $(0.5, 0.5)$ for Scenarios 2 and 3, respectively. Censoring times C_i are generated from uniform distribution $UNIF(0, C)$ where C is set as 4.6 for Scenario 1, 4.7 for Scenario 2, and 2.7 for Scenario 3, respectively. Roughly, 30% censoring percentages are produced for each scenario. The error model (5.4) is used to generate W_{ir} where $\epsilon_{ir} \sim N(0, \sigma^2)$ for $r = 1, \cdots, n_i, i = 1, \cdots, n$. We consider settings with $\sigma = 0.25$ or 0.75, and $n_i = 2$.

5.6.2 Performance of Estimators

In Table 5.1, we report the finite sample biases (Bias), the empirical variances (EVE), the average of the model-based variance estimates (MVE), the mean square errors (MSE), the coverage rate of 95% confidence intervals, calculated by $\hat{\beta}_A \pm 1.96\sqrt{\mathrm{Var}(\hat{\beta}_A)}$, where $\mathrm{Var}(\hat{\beta}_A)$ is the estimated variances, and the subscript A refers to nv, szs, rc, c and e accordingly.

It is seen that $\hat{\beta}_{nv}$ is always biased toward 0, with an increasing magnitude as measurement error becomes more substantial. These findings confirm the theoretical result revealed by the bias analysis in Section 5.3. The regression calibration estimator $\hat{\beta}_{rc}$ only partially remove the bias induced from measurement error, and its variance estimate deviates from the empirical variance in some settings. The two proposed estimates $\hat{\beta}_c$ and $\hat{\beta}_e$ have small finite sample biases. Their variance estimates agree reasonably well with the empirical variances, and the coverage rates agree well with the nominal level 95%. In contrast, when the measurement error is large, the variance of $\hat{\beta}_{szs}$ is considerably larger than those of $\hat{\beta}_c$ and $\hat{\beta}_e$, and the model based variance estimates of $\hat{\beta}_{szs}$ tend to deviate from the empirical variance estimates with much larger magnitudes. Finally, we comment that the estimator $\hat{\beta}_{szs}$ tends to behave less stably than the proposed estimators $\hat{\beta}_c$ and $\hat{\beta}_e$, and the regression calibration estimator $\hat{\beta}_{rc}$. In our simulations, about 1% of divergence occurs for the estimator $\hat{\beta}_{szs}$ when measurement error is large, whereas only 0.5% of divergence occurs for $\hat{\beta}_c$, $\hat{\beta}_e$ and $\hat{\beta}_{rc}$.

[Insert Table 5.1 here!]

5.6.3 Impact of the Number of Replicates

We now further evaluate the performance of the estimators for situations that some subjects may not have replicates W_{ij}. Specifically, settings of different replicate numbers n_i are considered for Scenario 3 described above. In Setting I, 150 out of $n = 200$ subjects are randomly selected to have two measurements, and the rest have a single measurement;

whereas in Setting II, 100 out of $n = 200$ subjects are randomly selected to have two measurements, and the rest have a single measurement. We further consider two settings for which we use a probability mechanism to decide whether or not a subject has a single measurement. That is, we treat n_i as a random variable taking value 1 or 2. Specifically, in Setting III, we assume that $\Pr(n_i = 1) = 0.8$ if $T_i \leq$ median of all T_i, and $\Pr(n_i = 1) = 0.2$ otherwise; in Setting IV, $\Pr(n_i = 1) = 0.2$ if $T_i \leq$ median of all T_i, and $\Pr(n_i = 1) = 0.8$ otherwise. Simulation results are summarized in Table 5.2. The primary finding is that the estimator $\hat{\beta}_{szs}$ is not appropriate when the number of measurements depends on the underlying event failure time. The results show that when there is a portion of subjects that have a single measurement, $\hat{\beta}_c$ and $\hat{\beta}_e$ have smaller variances than $\hat{\beta}_{szs}$. This demonstrates that $\hat{\beta}_c$ and $\hat{\beta}_e$ can effectively use information from subjects that have only a single measurement.

Finally, we consider Setting V where all subjects have only one single measurement, and the error variance is known to be 0.25^2 or 0.75^2. The estimator $\hat{\beta}_{szs}$ does not work for this setting as it is developed only for the case where all subjects must have replicated measurements for X_i. However, our estimators $\hat{\beta}_c$ and $\hat{\beta}_e$ can handle this scenario, and the simulation results show that they have satisfactory performance.

[Insert Table 5.2 here!]

5.6.4 Results on Cumulative Hazard Function

In Table 5.3, we use the procedure described in Section 5.4 to construct confidence bands of the baseline hazard function. Here, we consider only Scenario 1. For each simulation run, we independently generate standard normal variables $\xi_i, i = 1, \cdots, n$, and we repeat this procedure for 1000 times; we calculate $\hat{W}_n(t)$ each time and thus obtain the empirical upper 0.05-quantile $\tilde{q}_{0.05}$. In the total number of 1000 simulation runs, we record the number of cases that $\sup_{t\in[0,\tau]} \sqrt{n}|\hat{\Lambda}_0(t;\hat{\beta}_c) - \Lambda_0(t)|$ is less than $\tilde{q}_{0.05}$, and produce the empirical coverage rate accordingly. We repeat the above procedure for $\hat{\Lambda}_0(t;\hat{\beta}_c)$ described in Section

5.4, and the naive cumulative hazard estimator based on Lin and Ying (1994), and further modify these two estimators by the procedure of Hall and Wellner (1980). Simulation results reveal that naively ignoring measurement error could result in low coverage rates, especially when measurement error is large. The corrected methods greatly outperform the naive method.

[Insert Table 5.3 here!]

5.7 Model Misspecification and Model Checking

5.7.1 Model Misspecification

In the preceding sections we explore various methods to correct for bias induced by measurement error. The validity of the proposed methods relies on the additive hazards model structure for survival data. An important concern therefore arises: what if the true hazard function $\lambda(t; Z_i(t))$ is not of the additive hazards structure (5.1), but we incorrectly assume model form (5.1) to fit data. In this subsection, we investigate this problem.

Suppose the true model is given by the Cox model

$$\lambda(t; Z_i(t)) = \lambda_{cox}(t) \exp\{\alpha^T Z_i(t)\}, \tag{5.13}$$

but we incorrectly use the additive hazards model (5.1) to fit the data, where $\lambda_{cox}(t)$ is the true baseline hazard function, and α represents the true covariate effects.

Let β_c^* be the asymptotic limit of $\hat{\beta}_c$ developed in Section 5.4. Then following Hattori (2006) and Yi and Reid (2010), we show that β_c^* is given by

$$\beta_c^* = \left(\int_0^\tau E_{true} \left[Y_i(t) \{Z_i(t) - e_{true}(t)\}^{\otimes 2} \right] dt \right)^{-1} \int_0^\tau E_{true} \left[Y_i(t) \{Z_i(t) - e_{true}(t)\} d\Lambda(t; Z_i(t)) \right], \tag{5.14}$$

where E_{true} represents the expectation taken under the true model (5.13) with cumulative hazard function $\Lambda(t; Z_i(t)) = \int_0^t \lambda(u; Z_i(u))du$, and $e_{true}(t) = E_{true}\{Y_i(t)Z_i(t)\}/E_{true}\{Y_i(t)\}$.

It is difficult to see how β_c^* differs from α based on (5.14). To gain an understanding of the relationship between β_c^* and α, we consider an approximation of (5.14) for the situation with small $|\alpha^T Z_i(t)|$. Using the Taylor series expansion $\exp\{\alpha^T Z_i(t)\} \approx 1 + \alpha^T Z_i(t)$, we approximate the true hazard function (5.13) with an additive form:

$$\lambda(t; Z_i(t)) \approx \lambda_{cox}(t)\{1 + \alpha^T Z_i(t)\}.$$

As a result,

$$\begin{aligned} \beta_c^* &\approx \left(\int_0^\tau E_{true}\left[Y_i(t)\{Z_i(t) - e_{true}(t)\}^{\otimes 2}\right] dt \right)^{-1} \\ &\quad \times \int_0^\tau E_{true}\left[Y_i(t)\{Z_i(t) - e_{true}(t)\}\lambda_{cox}(t)\{1 + \alpha^T Z_i(t)\}dt\right] \\ &= R\alpha, \end{aligned} \tag{5.15}$$

where

$$R = \left(\int_0^\tau E_{true}\left[Y_i(t)\{Z_i(t) - e_{true}(t)\}^{\otimes 2}\right] dt \right)^{-1} \int_0^\tau E_{true}\left[Y_i(t)\{Z_i(t) - e_{true}(t)\}^{\otimes 2}\right] \lambda_{cox}(t)dt.$$

Expression (5.15) approximately quantifies the asymptotic bias of using the estimator $\hat{\beta}_c$ under the misspecified model (5.1) to estimate the true covariate effects α. It is seen that the estimated covariate effects β_c^* approximately differ from the true covariate effects α by a product R of two nonnegative definite matrices. The factor R depends on both survival and censoring processes. Although the estimated covariate effects β_c^* and the true covariate effects α are different in general, they tend to have the same sign when the covariate is univariate. In a special situation that there is no covariate effect, the estimated effect β_c^* is close to zero.

5.7.2 Model Checking

In the above subsection, it is seen that using the developed methods can yield biased estimates if the true covariate effects do not act additively on the hazard function. Thus, it is important to develop a model checking procedure for additive hazards models.

Let $\hat{\Sigma}_c$ be the empirical counterpart of Σ_c defined in Theorem 1, $(\hat{\Sigma}_c^{-1})_{jj}$ be the jth diagonal element of $\hat{\Sigma}_c^{-1}$, and $\Sigma_2(t) = \mathrm{diag}(\rho_0\rho_1^{-1}\sum_{r=1}^{n_i}(W_{ir} - \bar{W}_{i\cdot})^{\otimes 2}E\{\min(S_i,t)\}, 0)$ be the block diagonal matrix. Define

$$U_c(\hat{\beta}_c, t) = \sum_{i=1}^{n}\left[\int_0^t \left\{\hat{Z}_i(u) - \tilde{Z}(u)\right\}\left\{dN_i(u) - Y_i(u)\hat{\beta}_c^T\hat{Z}_i(u)du\right\} + \int_0^t Y_i(u)\hat{\Sigma}_1\hat{\beta}_c/n_i du\right],$$

$$A_i(t) = \int_0^t \left\{\hat{Z}_i(u) - e(u)\right\} d\tilde{M}_i(u;\beta,\Lambda_0) + \Sigma_2(t)\beta - \rho_0 E\{\min(S_i,t)\}\Sigma_1\beta + \frac{\min(S_i,t)\Sigma_1\beta}{n_i},$$

and $\mathcal{D}_c(t) = \lim_{n\to\infty} n^{-1}E\left[\sum_{i=1}^{n}\int_0^t Y_i(u)\left\{\hat{Z}_i(u) - e(u)\right\}^{\otimes 2}du\right] - \rho_0 E\{\min(S_i,t)\}\Sigma_1.$

The following lemma describes the asymptotic behavior of $n^{-1/2}U_c(\hat{\beta}_c, t)$. The proof is included in Appendix A9 of the Supplementary Material.

Lemma 1 *Under Regularity Conditions R1-R8 listed in Appendix A1 of the Supplementary Material, we have*

$$n^{-1/2}U_c(\hat{\beta}_c, t) \rightsquigarrow \mathcal{G}_2(t) \quad in \quad l^\infty[0,\tau] \quad as \quad n \to \infty,$$

where $\mathcal{G}_2(t)$ is a zero-mean Gaussian process with covariance function $\Phi_2(s,t) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E[\Psi_{2,i}(s)\Psi_{2,i}(t)]$ for time points s and t, and $\Psi_{2,i}(t) = A_i(t) - \mathcal{D}_c(t)\mathcal{D}_c^{-1}(\tau)A_i(\tau)$.

Lemma 1 provides a basis for the subsequent development of goodness-of-fit test. It says that if the additive hazards model and the additive error model are both correctly specified, $n^{-1/2}U_c(\hat{\beta}_c, t)$ would fluctuate around zero randomly provided regularity conditions hold. This motivates us to propose an overall goodness-of-fit test statistic

$$S_c = \sup_{t\in[0,\tau]}\sum_{j=1}^{p+q}(\hat{\Sigma}_c^{-1})_{jj}^{1/2}|n^{-1/2}U_{c,(j)}(\hat{\beta}_c, t)|,$$

where $U_{c,(j)}(\hat{\beta}_c, t)$ is the jth component of $U_c(\hat{\beta}_c, t)$.

In the absence of measurement error, S_c reduces to the overall goodness-of-fit test statistic by Kim, Song and Lee (1998), which is a generalization of the test statistic for checking the Cox model assumption proposed by Lin, Wei and Ying (1993).

The asymptotic distribution of S_c is difficult to be identified due to the complexity of the limit process $\mathcal{G}_2(t)$ associated with $n^{-1/2}U_c(\hat{\beta}_c, t)$. However, an abnormally large value of S_c can indicate that the additive hazards model and/or the error model are incorrectly specified.

Now we describe an implementation procedure using the resampling techniques similar to those in Section 5.4. We generate a set of $\{\xi_i, i = 1, \cdots, n\}$ independently from the standard normal distribution, and calculate

$$\hat{S}_c = \sup_{t\in[0,\tau]} \sum_{j=1}^{p+q} (\hat{\Sigma}_c^{-1})_{jj}^{1/2} |n^{-1/2}\hat{U}_{c,(j)}(\hat{\beta}_c, t)|,$$

where $\hat{U}_{c,(j)}(\hat{\beta}_c, t)$ is the jth component of $\hat{U}_c(\hat{\beta}_c, t) = \sum_{i=1}^{n} \xi_i \left\{ \hat{A}_i(t) - \hat{D}_c(t)\hat{D}_c^{-1}(\tau)\hat{A}_i(\tau) \right\}$, and $\hat{A}_i(t)$ and $\hat{D}_c(t)$ are the empirical versions of $A_i(t)$ and $\mathcal{D}_c(t)$, respectively. Then $\hat{S}_c$ can be used to assess goodness-of-fit because it mimics the behaviour of S_c asymptotically, as indicated below. The proof is sketched in Appendix A10 of the Supplementary Material.

Theorem 4 *Assume Regularity Conditions R1-R8 listed in Appendix A1 of the Supplementary Material. Then conditional on the observed data* $\{N_i(t), Y_i(t), \hat{Z}_{ir}(t), t \in [0, \tau], i = 1, \cdots, n, r = 1, \cdots, n_i\}$, $n^{-1/2}\hat{U}_c(\hat{\beta}_c, t)$ *converges weakly to* $\mathcal{G}_2(t)$ *in* $l^\infty[0, \tau]$ *in probability as* $n \to \infty$, *where* $\mathcal{G}_2(t)$ *is the Gaussian process defined in Lemma 1.*

Theorem 4 also offers a justification to empirically evaluate the power of using S_c for model checking. Specifically, we generate sets of i.i.d. standard normal variables $\{\xi_{i,k}, i = 1, \cdots, n\}$ for N times, where N is a large number, say $N = 1000$. Then we calculate N copies of $\hat{S}_c$, say $\{\hat{S}_{c,k}, k = 1, \cdots, N\}$. Empirical quantiles of S_c can then be obtained based on the $\hat{S}_{c,k}$.

Now we numerically assess the performance of the proposed test statistic S_c. First, we evaluate the empirical size of the test. We take the setting of Scenario 1 to generate the data. We consider two cases with no censoring or 30% censoring percentage.

The results for the empirical size of the corrected goodness-of-test statistic S_c are summarized in Table 5.4, where the null hypothesis is that both of the additive hazards model and the additive error model are correctly specified. For comparison purposes, we also consider the naive goodness-of-test statistic S_{nv} by naively applying the method of Kim, Song and Lee (1998) with the difference between X_i and $\bar{W}_{i\cdot}$ ignored, and the "true" goodness-of-test statistic, named S_{true}, obtained by applying the method of Kim, Song and Lee (1998) to the true covariate measurements.

It is observed that in the presence of censoring, the test size of S_{nv} is close to the nominal level. However, when there is no censoring, the naive test statistic S_{nv} yields test sizes which completely deviate from the nominal size 0.05. In contrast, the proposed statistic S_c produces test sizes that are fairly close to the nominal level in all cases, and its performance is similar to the true goodness-of-test statistic S_{true}.

Next, we evaluate the power of the proposed test statistic. We generate the survival times from the Cox model $\lambda(t|Z_i) = \lambda_0(t)\exp(X_i\alpha_x + V_i\alpha_v)$ with $\lambda_0(t) = t$ and $(\alpha_x, \alpha_v)^T = (1, 0)^T$. The covariates X_i and V_i are generated as in Scenario 1 in Section 5.6.1. The error model (5.4) is used to generate W_{ir} where $\epsilon_{ir} \sim N(0, \sigma^2)$ for $r = 1, \cdots, n_i, i = 1, \cdots, n$. We consider settings with $\sigma = 0.25$ or 0.75, and $n_i = 2$. By taking C_i to be ∞ or generating C_i from $UNIF(0, 4.6)$, we obtain two censoring scenarios: no censoring and 30% censoring, respectively. The results are summarized in Table 5.4. It is seen that the power of the proposed test statistic S_c is fairly satisfactory, although the power would decrease when the degree of measurement error increases.

[Insert Table 5.4 here!]

5.8 ACTG 175 Study

We apply the proposed methods to analyze the data arising from the AIDS Clinical Trials Group (ACTG) 175 study (Hammer, et al., 1996). The ACTG 175 study is a double-blind randomized clinical trial that evaluated the effects of the four types of HIV treatments: zidovudine only, zidovudine and didanosine, zidovudine and zalcitabine, and didanosine only. In this example, we are interested in evaluating how different treatments are associated with the survival time T_i, which is defined to be the time to the occurrence of one of the events that CD4 counts decrease at least 50%, or disease progression to AIDS, or death. We consider a subset of $n = 2139$ subjects in this study. About 75.6% of the outcome values are censored.

Let V_i be the treatment assignment indicator for subject i, where $V_i = 1$ if a subject received the zidovudine only treatment, and 0 otherwise. In the ACTG 175 study, the baseline measurements on CD4 were collected before randomization, ranging from 200 to 500 per cubic millimeter. Let X_i be the normalization version of the true baseline CD4 counts: log(CD4 counts + 1), which was not observed in the study. Two replicated baseline measurements of CD4 counts, denoted by W_{i1} and W_{i2}, after the same transformation as for X_i, were observed for 2095 subjects, while the other 44 subjects were measured once for the CD4 counts at the baseline. An additive measurement error model is specified to link the underlying transformed CD4 counts with its surrogate measurements:

$$W_{ir} = X_i + \epsilon_{ir},$$

where $r = 1, 2$ for $i = 1, \cdots, 2095$, and $r = 1$ for $i = 2096, \cdots, 2139$. Here, no specific distributional assumption is made for the errors ϵ_{ir} except that the ϵ_{ir} are assumed to be independent and identically distributed with mean zero and variance Σ_0. With the replicates, we estimate the error variance as $\hat{\Sigma}_0 = 0.035$, and the variance of X_i as $\hat{\Sigma}_{xx} = 0.079$. These estimates give the reliability ratio $\hat{\Sigma}_{xx}/(\hat{\Sigma}_{xx} + \hat{\Sigma}_0) = 69.3\%$, indicating a considerable degree of measurement error in this study.

We employ the additive hazards model to feature the dependence of T_i on the covariates

X_i and V_i:

$$\lambda(t; Z_i) = \lambda_0(t) + X_i\beta_x + V_i\beta_v,$$

where $\lambda_0(t)$ is the unspecified baseline hazard function, and $\beta = (\beta_x, \beta_v)^T$ is the regression parameter.

We apply the methods considered in Section 5.6 to analyze the data: the data subsets with replicates and the entire data set. The analysis results are shown in Table 5.5. The naive estimate of β_x is smaller than those obtained from the other methods, while the naive estimate of β_v is similar to those produced by the other methods. All the consistent methods and the regression calibration method produce similar results. Although estimates of β_x and β_v differ from method to method, all the results suggest that both CD4 counts and treatment are statistically significant.

We also apply the proposed test statistic S_c to the ACTG 175 Study dataset. The p-value of the model test is 0.859, suggesting no evidence against the additive hazards model or the additive error model.

[Insert Table 5.5 here!]

5.9 Extension and Discussion

In this chapter, we make a number of contributions on additive hazards models with measurement error. Our bias analysis and regression calibration method fill up gaps in the literature. We propose several consistent and easily implemented estimators to correct for measurement error effects, and our methods are robust to possible misspecification for the distribution of the true covariates. Our methods embrace limited existing work, such as Sun, Zhang and Sun (2006) as a special case. Furthermore, our comprehensive development includes investigation of the impact of model misspecification of the survival process and construction of a test statistic for model checking. We rigorously establish asymptotic properties for the proposed estimators. Extensive numerical studies demonstrate satisfactory performance of our methods.

Our methods here are explicitly developed for the additive error model (5.4). In fact, our methods can be modified to accommodate more general error models. For instance, consider a regression measurement error model

$$W_{ir} = \gamma_0 + \gamma_x X_i + \gamma_v V_i + \epsilon_{ir}, \quad r = 1, \cdots, n_i, \tag{5.16}$$

where the error terms ϵ_{ir} are i.i.d. with mean 0 and a positive-definite variance matrix Σ_0, and are independent of $N_i(\cdot), Y_i(\cdot)$, and $Z_i(\cdot)$, $i = 1, \cdots, n; r = 1, \cdots, n_i$. Here, γ_x is a $p \times p$ matrix, and γ_v is a $p \times q$ matrix. Model (5.16) accommodates a wide class of error models, including the classical additive model (5.4) if we set $\gamma_0 = 0$, and $\gamma_x = 1$ with a zero vector γ_v. In the following, we consider a special case that $p = q = 1$.

Let $\hat{X}_{g,i} = (\bar{W}_{i\cdot} - \gamma_v V_i - \gamma_0)/\gamma_x$, then replacing X_i with $\hat{X}_{g,i}$ in (5.2), we obtain a corrected pseudo-score function

$$\begin{aligned} U_{gc}(\beta) &= \sum_{i=1}^{n} \int_0^{\tau} \left\{ \hat{Z}_{g,i}(t) - \tilde{Z}_g(t) \right\} \left\{ dN_i(t) - Y_i(t)\beta^T \hat{Z}_{g,i}(t) dt \right\} \\ &\quad + \int_0^{\tau} \left\{ 1 - \frac{1}{\sum_{j=1}^{n} Y_j(t)} \right\} \sum_{i=1}^{n} \left\{ Y_i(t) \hat{\Sigma}_1 \beta / (\gamma_x^2 n_i) \right\} dt, \end{aligned}$$

where $\hat{Z}_{g,i}(t) = (\hat{X}_{g,i}^T, V_i^T(t))^T$. Consequently, solving $U_{gc}(\beta) = 0$ gives an estimator of β:

$$\begin{aligned} \hat{\beta}_{gc} &= \left[\sum_{i=1}^{n} \int_0^{\tau} Y_i(t) \left\{ \hat{Z}_{g,i}(t) - \tilde{Z}_g(t) \right\}^{\otimes 2} dt - \int_0^{\tau} \left\{ 1 - \frac{1}{\sum_{j=1}^{n} Y_j(t)} \right\} \sum_{i=1}^{n} \left\{ Y_i(t) \hat{\Sigma}_1 / (\gamma_x^2 n_i) \right\} dt \right]^{-1} \\ &\quad \times \left(\sum_{i=1}^{n} \int_0^{\tau} \left\{ \hat{Z}_{g,i}(t) - \tilde{Z}_g(t) \right\} dN_i(t) \right), \end{aligned}$$

where $\tilde{Z}_g(t) = \sum_{i=1}^{n} Y_i(t) \hat{Z}_{g,i}(t) / \sum_{i=1}^{n} Y_i(t)$. Note that the derivation of $\hat{\beta}_{gc}$ is similar to that of $\hat{\beta}_c$ in Section 5.4. Similarly, we can construct other consistent estimator similar to $\hat{\beta}_e$. Furthermore, we can construct estimators of $\Lambda_0(t)$ similar to previous sections, which, however, further requires that γ_0 is known or estimated by a validation subsample.

Finally, we note that our methods are developed for error-contaminated survival data that are modulated by the additive hazards model (5.1). The additive hazards model (5.1) is a useful complement to the popularly-used proportional hazards model. This model allows for a simple procedure for conducting inference on the model parameter β whose estimator can be explicitly expressed. However, to ensure a legitimate hazard function, the linear term $\beta^T Z_i(t)$ in model (5.1) must be constrained to be nonnegative (Aalen, Borgan and Gjessing 2008, Sec. 4.2). To avoid this nonnegativity constraint, one may consider alternative forms of model (5.1). For example, one may replace the linear term $\beta^T Z_i(t)$ by an exponential form $\exp\{\beta^T Z_i(t)\}$. Alternative additive hazards models are discussed by Lin and Ying (1995, 1997). It would be interesting to modify our development here to other additive hazards models.

SUPPLEMENTARY MATERIAL

Appendix A1 of the Supplementary Material includes regularity conditions and the proofs of the theorems in the chapter. Appendix A2 of the Supplementary Material includes several more estimators are proposed and their theoretical properties are studied.

In the following derivations, we introduce some notations. For an $m \times 1$ vector $a = (a_1, a_2, \cdots, a_m)^T$, let $||a|| = (\sum a_i^2)^{1/2}$ denote the Euclidean norm for a. For a matrix A, define $||A|| = \max_{i,j} |a_{ij}|$, where a_{ij} is the (i, j)th element of A. For vector processes $A_n(t)$ and $A(t)$, $A_n(t)$ is said to converge almost surely to $A(t)$ uniformly in t if

$$\sup_{0 \le t \le \tau} ||A_n(t) - A(t)|| \overset{a.s.}{\to} 0, \text{as} \quad n \to \infty.$$

We define that a random matrix $A_n = o_{a.s.}(1)$ in the sense that $\Pr\{\lim_{n\to\infty} ||A_n|| = 0\} = 1$.

Appendix A1: Regularity Conditions

R1. $\{N_i(\cdot), Y_i(\cdot), Z_i(\cdot)\}, i = 1, \cdots, n$ are independent and identically distributed.

R2. $\Pr\{Y_i(\tau) = 1\} > 0$ for $i = 1, \cdots, n$.

R3. T_i and C_i are conditionally independent given $Z_i(t)$, $i = 1, \cdots, n$.

R4. $\sup_{t\in[0,\tau]} ||E\{Z_i^{\otimes 2}(t)\}|| < \infty$, $i = 1, \cdots, n$.

R5. Bounded variation condition: for $i = 1, \cdots, n$; $j = 1, \cdots, p + q$,

$$|Z_{ij}(0)| + \int_0^\tau |dZ_{ij}(u)| \leq K$$

holds almost surely for all the sample path, where K is a constant.

R6. All the $n_i (i = 1, \cdots, n)$ are bounded by a constant N_0, and $\lim_{n\to\infty} n^{-1} \sum_{i=1}^n I\{n_i = j\}$ exists, where $j = 1, \cdots, N_0$.

R7. $||E(\epsilon_{ir}^{\otimes 2})|| < \infty$, $i = 1, \cdots, n$; $r = 1, \cdots, n_i$.

R8. $\int_0^\tau E\left[Y_i(t)\{Z_i(t) - e(t)\}^{\otimes 2}\right] dt$ and Σ_c are positive definite, $i = 1, \cdots, n$.

These regularity conditions are imposed for the technical development. The conditions R1, R2, R3, R4, R5 and R8 are conventionally used for developing asymptotic theory in survival analysis, and they are analogous to those by, for example, Andersen and Gill (1982), Spiekerman and Lin (1998), Lin, Wei, Yang and Ying (2000), and Hu and Lin (2004). In particular, condition R1 assumes homogeneity among the subjects in the study. Condition R2 says that each subject in the study has a positive probability to be observed, and this condition ensures the denominator of $e(t) = E\{Y_i(t)Z_i(t)\}/E\{Y_i(t)\}$ to be bounded away from zero. Condition R3 is a common assumption for censoringship. Conditions R4 and R5 control the variability of the covariates in which condition R5 is a key assumption for using empirical process theory (e.g., Strong Uniform Law of Large Numbers by Pollard 1990). Condition R6 guarantees the existence of ρ_0 and ρ_1; imposing the upper bound for the n_i is often plausible as in practice, the n_i are usually not large. Condition R7 controls the variability of the error terms ϵ_{ir}, and condition R8 is needed for developing asymptotic normality of the proposed estimators.

Appendix A2: Proof of the relationship (5.6) and its consequence

Note that

$$\hat{\beta}_{nv} = \left[n^{-1} \sum_{i=1}^{n} \int_0^{\tau} Y_i(t) \left\{ \hat{Z}_i(t) - \tilde{Z}(t) \right\}^{\otimes 2} dt \right]^{-1} \left[n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \left\{ \hat{Z}_i(t) - \tilde{Z}(t) \right\} dN_i(t) \right].$$

Let

$$D_{nv} = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} Y_i(t) \left\{ \hat{Z}_i(t) - \tilde{Z}(t) \right\}^{\otimes 2} dt$$

denote the denominator of $\hat{\beta}_{nv}$. By the Strong Uniform Law of Large Numbers (USLLN) (Pollard 1990), it follows that $D_{nv} \to \mathcal{D}_{nv}$ almost surely as $n \to \infty$, where

$$\mathcal{D}_{nv} \;\; = \;\; E\left[\int_0^{\tau} Y_i(t) \left\{ Z_i(t) - e(t) \right\}^{\otimes 2} dt \right] + \int_0^{\tau} \rho_0 E\{Y_i(t)\} \Sigma_1 dt \triangleq B_1 + B_2.$$

Indeed, we have

$$\begin{aligned}
& n^{-1} \sum_{i=1}^{n} \int_0^{\tau} Y_i(t) \left\{ \hat{Z}_i(t) - \tilde{Z}(t) \right\}^{\otimes 2} dt \\
= \;& n^{-1} \sum_{i=1}^{n} \int_0^{\tau} Y_i(t) \left\{ \hat{Z}_i(t) - e(t) \right\}^{\otimes 2} dt + n^{-1} \sum_{i=1}^{n} \int_0^{\tau} Y_i(t) \left\{ e(t) - \tilde{Z}(t) \right\}^{\otimes 2} dt \\
& + n^{-1} \sum_{i=1}^{n} \int_0^{\tau} 2Y_i(t) \left\{ \hat{Z}_i(t) - e(t) \right\} \left\{ e(t) - \tilde{Z}(t) \right\}^{T} dt \\
= \;& n^{-1} \sum_{i=1}^{n} \int_0^{\tau} Y_i(t) \left\{ \hat{Z}_i(t) - e(t) \right\}^{\otimes 2} dt \\
& + \int_0^{\tau} \left(-n^{-1} \sum_{i=1}^{n} Y_i(t) e^{\otimes 2}(t) - n^{-1} \frac{\left\{ \sum_{j=1}^{n} Y_j(t) \hat{Z}_j(t) \right\}^{\otimes 2}}{\sum_{j=1}^{n} Y_j(t)} + 2n^{-1} \sum_{i=1}^{n} Y_i(t) \hat{Z}_i(t) e^{T}(t) \right) dt.
\end{aligned}$$

By Condition R5, each component of the vector $n^{-1} \sum_{i=1}^{n} Y_i(t) Z_i(t)$ is of bounded variation, so $n^{-1} \sum_{i=1}^{n} Y_i(t) Z_i(t)$ is the difference of two nondecreasing functions. By Lemma A.1 and A.2 of Bilias, Gu and Ying (1997), $n^{-1} \sum_{i=1}^{n} Y_i(t) Z_i(t)$ is manageable (Pollard 1990,

p.38), and $n^{-1}\sum_{i=1}^n Y_i(t)\hat{Z}_i(t)$ is manageable. Furthermore, Conditions R4, R6 and R7 justify that the envelope (Pollard 1990 p.19) F_i of $\{\hat{Z}_i(t), t \in [0,\tau]\}$ is finite and could only take at maximum N_0 possible values, and thus $\max_{i=1} E(||F_i^{\otimes 2}||) < \infty$. It follows that $\sum_{i=1}^{\infty} E(||F_i^{\otimes 2}||)/i^2 \leq \max_i E(||F_i^{\otimes 2}||)\sum_{i=1}^{\infty} 1/i^2 < \infty$. The two conditions of the Strong Uniform Law of Large Numbers (SULLN) (Pollard 1990, p.41) are thus verified. Consequently, we obtain that
$n^{-1}\sum_{i=1}^n Y_i(t)\hat{Z}_i(t) \overset{a.s.}{\to} E\{Y_i(t)Z_i(t)\}$ uniformly in t. Similarly, $n^{-1}\sum_{i=1}^n Y_i(t) \overset{a.s.}{\to} E\{Y_i(t)\}$ uniformly in t. By SULLN together with Condition R2, $\sum_{i=1}^n Y_i(t)\hat{Z}_i(t)/\sum_{i=1}^n Y_i(t) \overset{a.s.}{\to} e(t)$ uniformly in t. Thus, we obtain that

$$
\begin{aligned}
& \int_0^\tau \left(-n^{-1}\sum_{i=1}^n Y_i(t)e^{\otimes 2}(t) - n^{-1}\frac{\left\{\sum_{j=1}^n Y_j(t)\hat{Z}_j(t)\right\}^{\otimes 2}}{\sum_{j=1}^n Y_j(t)} + 2n^{-1}\sum_{i=1}^n Y_i(t)\hat{Z}_i(t)e^T(t) \right) dt \\
= & \int_0^\tau \left(-E\{Y_i(t)\}e^{\otimes 2}(t) - \frac{\left[E\{Y_j(t)\hat{Z}_j(t)\}\right]^{\otimes 2}}{E\{Y_j(t)\}} + 2E\{Y_i(t)\hat{Z}_i(t)\}e^T(t) \right) dt + o_{a.s.}(1) \\
= & \int_0^\tau \left(-E\{Y_i(t)\}e^{\otimes 2}(t) - \frac{[E\{Y_j(t)Z_j(t)\}]^{\otimes 2}}{E\{Y_j(t)\}} + 2E\{Y_i(t)Z_i(t)\}e^T(t) \right) dt + o_{a.s.}(1) \\
= & \int_0^\tau \left[-E\{Y_i(t)\}e^{\otimes 2}(t) - E\{Y_i(t)\}e^{\otimes 2}(t) + 2E\{Y_i(t)\}e^{\otimes 2}(t) \right] dt + o_{a.s.}(1) \\
= & \; o_{a.s.}(1),
\end{aligned}
$$

where the second last identity follows from the definition of $e(t)$. Let $\bar{\epsilon}_{i\cdot} = (n_i^{-1}\sum_{k=1}^{n_i}\epsilon_{ik}^T, 0^T)^T$, and $\rho_0 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n n_i^{-1}$, where the existence of the limit is ensured by the Regularity Conditions. By SULLN and observing that $E\{\hat{Z}_i^{\otimes 2}(t)\} = E\{Z_i^{\otimes 2}(t)\} + \Sigma_1/n_i$, we obtain that

$$\begin{aligned}
D_{nv} &= n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{\hat{Z}_i(t)-e(t)\right\}^{\otimes 2} dt + o_{a.s.}(1)\\
&= n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2} dt + n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\bar{\epsilon}_{i\cdot}^{\otimes 2} dt\\
&\quad + n^{-1}\sum_{i=1}^{n}\int_0^{\tau} 2Y_i(t)\left\{Z_i(t)-e(t)\right\}\bar{\epsilon}_{i\cdot}^{T} dt + o_{a.s.}(1)\\
&= n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2} dt + n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\bar{\epsilon}_{i\cdot}^{\otimes 2} dt\\
&\quad + \int_0^{\tau} 2E[Y_i(t)\left\{Z_i(t)-e(t)\right\}\bar{\epsilon}_{i\cdot}^{T}] dt + o_{a.s.}(1)\\
&= n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2} dt + n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\bar{\epsilon}_{i\cdot}^{\otimes 2} dt\\
&\quad + \int_0^{\tau} 2E[Y_i(t)\left\{Z_i(t)-e(t)\right\}E(\bar{\epsilon}_{i\cdot}^{T})] dt + o_{a.s.}(1)\\
&= n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2} dt + n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\bar{\epsilon}_{i\cdot}^{\otimes 2} dt + o_{a.s.}(1)\\
&= n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2} dt + \int_0^{\tau} \lim_{n\to\infty}\left[n^{-1}\sum_{i=1}^{n} E\{Y_i(t)\bar{\epsilon}_{i\cdot}^{\otimes 2}\}\right] dt + o_{a.s.}(1)\\
&= n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2} dt + \int_0^{\tau} E\{Y_i(t)\}\lim_{n\to\infty}\left\{n^{-1}\sum_{i=1}^{n} E(\bar{\epsilon}_{i\cdot}^{\otimes 2})\right\} dt + o_{a.s.}(1)\\
&= n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2} dt + \int_0^{\tau} E\{Y_i(t)\}\lim_{n\to\infty}\left\{n^{-1}\sum_{i=1}^{n} \Sigma_1/n_i\right\} dt + o_{a.s.}(1)\\
&= n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2} dt + \int_0^{\tau} \rho_0 E\{Y_i(t)\}\Sigma_1 dt + o_{a.s.}(1),
\end{aligned}$$

where the last two steps follow from the definition of Σ_1 and ρ_0. Thus, by the definition of $\mathcal{D}_{nv}$, we obtain that

$$D_{nv} = \mathcal{D}_{nv} + o_{a.s.}(1),$$

i.e., $D_{nv} \to \mathcal{D}_{nv}$ almost surely as $n \to \infty$.

Similarly, by SULLN we have $n^{-1}\sum_{i=1}^n N_i(t) \stackrel{a.s.}{\to} E\{N_i(t)\}$ uniformly in t. By Lemma 1 of Lin, Wei, Yang and Ying (2000), $n^{-1}\sum_{i=1}^n \int_0^\tau \tilde{Z}(t)dN_i(t) \stackrel{a.s.}{\to} \int_0^\tau e(t)E\{dN_i(t)\}$. Similarly, $n^{-1}\sum_{i=1}^n \int_0^\tau \hat{Z}_i(t)dN_i(t) \stackrel{a.s.}{\to} \int_0^\tau E\{Z_i(t)dN_i(t)\}$. Thus, we obtain that

$$\begin{aligned} n^{-1}\sum_{i=1}^n \int_0^\tau \left\{\hat{Z}_i(t)-\tilde{Z}(t)\right\} dN_i(t) &= n^{-1}\sum_{i=1}^n \int_0^\tau \left\{\hat{Z}_i(t)-e(t)\right\} dN_i(t) + o_{a.s.}(1) \\ &= E\left[\int_0^\tau \left\{\hat{Z}_i(t)-e(t)\right\} dN_i(t)\right] + o_{a.s.}(1) \\ &= E\left[\int_0^\tau \left\{Z_i(t)-e(t)\right\} dN_i(t)\right] + o_{a.s.}(1). \end{aligned}$$

As a result, we obtain that β^*_{nv}, the asymptotic limit of $\hat{\beta}_{nv}$, is given by

$$\beta^*_{nv} = \mathcal{D}_{nv}^{-1}E\left[\int_0^\tau \left\{Z_i(t)-e(t)\right\} dN_i(t)\right] = (B_1+B_2)^{-1}E\left[\int_0^\tau \left\{Z_i(t)-e(t)\right\} dN_i(t)\right]. \tag{5.17}$$

On the other hand, Lin and Ying (1994) showed that, in the absense of measurement error, the estimator

$$\hat{\beta} = \left[\sum_{i=1}^n \int_0^\tau Y_i(t)\{Z_i(t)-\bar{Z}(t)\}^{\otimes 2}dt\right]^{-1}\left[\sum_{i=1}^n \int_0^\tau \{Z_i(t)-\bar{Z}(t)\}dN_i(t)\right]$$

converges in probability to β. By analogy to the preceding arguments, we can show that

$$\beta = B_1^{-1}E\left[\int_0^\tau \left\{Z_i(t)-e(t)\right\} dN_i(t)\right]. \tag{5.18}$$

Comparison between (5.17) and (5.18) leads to the expression of β^*_{nv} and β, we obtain that $\beta^*_{nv} = (B_1+B_2)^{-1}B_1\beta$. The proof is then completed.

It is straightforward that if $\|\beta\| = 0$, then $\|\beta^*_{nv}\| = 0$. When $Z_i(t)$ is univariate, then $|(B_1+B_2)^{-1}B_1| < 1$, and it follows that $|\beta^*_{nv}| < |\beta|$. If X_i and $V_i(t)$ are univariates and are independent, and either $V_i(t)$ or X_i are independent of the followup process, then B_1 is a 2×2 diagonal matrix, i.e.,

$$B_1 = \begin{pmatrix} \int_0^\tau E\left[Y_i(t)\{X_i - e_1(t)\}^2\right] dt & 0 \\ 0 & \int_0^\tau E\left[Y_i(t)\{V_i(t) - e_2(t)\}^2\right] dt \end{pmatrix},$$

where $e_1(t)$ and $e_2(t)$ are the two components of the vector $e(t)$. Note that

$$B_2 = \begin{pmatrix} \rho_0 \Sigma_0 \int_0^\tau E\{Y_i(t)\} dt & 0 \\ 0 & 0 \end{pmatrix},$$

It follows that

$$(B_1 + B_2)^{-1} B_1 = \begin{pmatrix} \frac{\int_0^\tau E\left[Y_i(t)\{X_i - e_1(t)\}^2\right] dt}{\int_0^\tau E[Y_i(t)\{X_i - e_1(t)\}^2] dt + \rho_0 \Sigma_0 \int_0^\tau E\{Y_i(t)\} dt} & 0 \\ 0 & 1 \end{pmatrix}.$$

Therefore, the relationship (7) implies that $|\beta^*_{nv,x}| < |\beta_x|$ and $\beta^*_{nv,v} = \beta_v$.

Appendix A3: Derivation of $E\{U_{nv}(\beta)|\mathcal{F}_\tau\}$

$$
\begin{aligned}
& E\{U_{nv}(\beta)|\mathcal{F}_\tau\} \\
= \;& \sum_{i=1}^n \int_0^\tau [E\{\hat{Z}_i(t)|\mathcal{F}_\tau\} - E\{\tilde{Z}(t)|\mathcal{F}_\tau\}] dN_i(t) \\
& - \sum_{i=1}^n \int_0^\tau E\{\hat{Z}_i^{\otimes 2}(t)|\mathcal{F}_\tau\} Y_i(t)\beta dt + \sum_{i=1}^n \int_0^\tau E\{\tilde{Z}(t)\hat{Z}_i^T(t)|\mathcal{F}_\tau\} Y_i(t)\beta dt \\
= \;& \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} dN_i(t) - \sum_{i=1}^n \int_0^\tau \{Z_i^{\otimes 2}(t) + \Sigma_1/n_i\} Y_i(t)\beta dt \\
& + \sum_{i=1}^n \int_0^\tau \left\{ \frac{Y_i(t) E\{\hat{Z}_i^{\otimes 2}(t)|\mathcal{F}_\tau\} + \sum_{j\neq i} Y_j(t) E\{\hat{Z}_j(t)\hat{Z}_i^T(t)|\mathcal{F}_\tau\}}{\sum_{j=1}^n Y_j(t)} \right\} Y_i(t)\beta dt \\
= \;& \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} dN_i(t) - \sum_{i=1}^n \int_0^\tau \{Z_i^{\otimes 2}(t) + \Sigma_1/n_i\} Y_i(t)\beta dt \\
& + \sum_{i=1}^n \int_0^\tau \left\{ \frac{Y_i(t)\{Z_i^{\otimes 2}(t) + \Sigma_1/n_i\} + \sum_{j\neq i} Y_j(t) Z_j(t) Z_i^T(t)}{\sum_{j=1}^n Y_j(t)} \right\} Y_i(t)\beta dt \\
= \;& \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} dN_i(t) - \sum_{i=1}^n \int_0^\tau \{Z_i^{\otimes 2}(t) + \Sigma_1/n_i\} Y_i(t)\beta dt \\
& + \sum_{i=1}^n \int_0^\tau \left\{ \frac{Y_i(t)\Sigma_1/n_i + \sum_{j=1}^n Y_j(t) Z_j(t) Z_i^T(t)}{\sum_{j=1}^n Y_j(t)} \right\} Y_i(t)\beta dt \\
= \;& \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} dN_i(t) - \sum_{i=1}^n \int_0^\tau Z_i^{\otimes 2}(t) Y_i(t)\beta dt - \sum_{i=1}^n \int_0^\tau Y_i(t)\Sigma_1\beta/n_i dt \\
& + \sum_{i=1}^n \int_0^\tau \left\{ \frac{Y_i(t)\Sigma_1/n_i}{\sum_{j=1}^n Y_j(t)} \right\} Y_i(t)\beta dt + \sum_{i=1}^n \int_0^\tau \bar{Z}(t) Z_i^T(t) Y_i(t)\beta dt \\
= \;& U(\beta) - \int_0^\tau \left\{ 1 - \frac{1}{\sum_{j=1}^n Y_j(t)} \right\} \sum_{i=1}^n \{Y_i(t)\Sigma_1\beta/n_i\}\, dt.
\end{aligned}
$$

Appendix A4: Proof of Theorem 1

Recall that

$$\hat{\beta}_c = \left[n^{-1} \sum_{i=1}^{n} \int_0^\tau Y_i(t) \left\{ \hat{Z}_i(t) - \tilde{Z}(t) \right\}^{\otimes 2} dt - n^{-1} \int_0^\tau \left\{ 1 - \frac{1}{\sum_{j=1}^{n} Y_j(t)} \right\} \sum_{i=1}^{n} \left\{ Y_i(t) \hat{\Sigma}_1 / n_i \right\} dt \right]^{-1}$$
$$\times \left[n^{-1} \sum_{i=1}^{n} \int_0^\tau \left\{ \hat{Z}_i(t) - \tilde{Z}(t) \right\} dN_i(t) \right].$$

Let

$$D_c = n^{-1} \sum_{i=1}^{n} \int_0^\tau Y_i(t) \left\{ \hat{Z}_i(t) - \tilde{Z}(t) \right\}^{\otimes 2} dt - n^{-1} \int_0^\tau \left\{ 1 - \frac{1}{\sum_{j=1}^{n} Y_j(t)} \right\} \sum_{i=1}^{n} \left\{ Y_i(t) \hat{\Sigma}_1 / n_i \right\} dt$$

denote the denominator of $\hat{\beta}_c$. Let

$$\mathcal{D}_c = E \left[\int_0^\tau Y_i(t) \left\{ Z_i(t) - e(t) \right\}^{\otimes 2} dt \right].$$

We first prove the following lemmas:

Lemma A.1 *Under Regularity Conditions R1-R8, D_c converges to $\mathcal{D}_c$ almost surely as $n \to \infty$.*

Proof: The proof consists of two parts. In the first part, we examine the asymptotic behavior of the first term of D_c while in the second part we look at the second term of D_c.

Part 1:

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{\hat{Z}_i(t)-\tilde{Z}(t)\right\}^{\otimes 2} dt \\
= \; & n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{\hat{Z}_i(t)-e(t)\right\}^{\otimes 2} dt + n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{e(t)-\tilde{Z}(t)\right\}^{\otimes 2} dt \\
& + n^{-1}\sum_{i=1}^{n}\int_0^{\tau} 2Y_i(t)\left\{\hat{Z}_i(t)-e(t)\right\}\left\{e(t)-\tilde{Z}(t)\right\}^{T} dt \\
= \; & n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{\hat{Z}_i(t)-e(t)\right\}^{\otimes 2} dt \\
& + \int_0^{\tau}\left(-n^{-1}\sum_{i=1}^{n} Y_i(t)e^{\otimes 2}(t) - n^{-1}\frac{\left\{\sum_{j=1}^{n} Y_j(t)\hat{Z}_j(t)\right\}^{\otimes 2}}{\sum_{j=1}^{n} Y_j(t)} + 2n^{-1}\sum_{i=1}^{n} Y_i(t)\hat{Z}_i(t)e^{T}(t)\right) dt.
\end{aligned}
$$

Now we examine the second term of the expression above by applying the Strong Uniform Law of Large Numbers individually to each term and obtain

$$
\begin{aligned}
& \int_0^{\tau}\left(-n^{-1}\sum_{i=1}^{n} Y_i(t)e^{\otimes 2}(t) - n^{-1}\frac{\left\{\sum_{j=1}^{n} Y_j(t)\hat{Z}_j(t)\right\}^{\otimes 2}}{\sum_{j=1}^{n} Y_j(t)} + 2n^{-1}\sum_{i=1}^{n} Y_i(t)\hat{Z}_i(t)e^{T}(t)\right) dt \\
= \; & \int_0^{\tau}\left(-E\{Y_i(t)\}e^{\otimes 2}(t) - \frac{\left[E\{Y_j(t)\hat{Z}_j(t)\}\right]^{\otimes 2}}{E\{Y_j(t)\}} + 2E\{Y_i(t)\hat{Z}_i(t)\}e^{T}(t)\right) dt + o_{a.s.}(1) \\
= \; & \int_0^{\tau}\left(-E\{Y_i(t)\}e^{\otimes 2}(t) - \frac{[E\{Y_j(t)Z_j(t)\}]^{\otimes 2}}{E\{Y_j(t)\}} + 2E\{Y_i(t)Z_i(t)\}e^{T}(t)\right) dt + o_{a.s.}(1) \\
= \; & \int_0^{\tau}\left[-E\{Y_i(t)\}e^{\otimes 2}(t) - E\{Y_i(t)\}e^{\otimes 2}(t) + 2E\{Y_i(t)\}e^{\otimes 2}(t)\right] dt + o_{a.s.}(1) \\
= \; & o_{a.s.}(1),
\end{aligned}
$$

where the second last identity comes from the definition of $e(t)$.

Then it follows by the Strong Uniform Law of Large Numbers that

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\left\{\hat{Z}_i(t)-\tilde{Z}(t)\right\}^{\otimes 2}dt \\
= \;& n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\left\{\hat{Z}_i(t)-e(t)\right\}^{\otimes 2}dt + o_{a.s.}(1) \\
= \;& n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2}dt + n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\bar{\epsilon}_{i\cdot}^{\otimes 2}dt \\
& + n^{-1}\sum_{i=1}^{n}\int_0^\tau 2Y_i(t)\left\{Z_i(t)-e(t)\right\}\bar{\epsilon}_{i\cdot}^{T}dt + o_{a.s.}(1) \\
= \;& n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2}dt + n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\bar{\epsilon}_{i\cdot}^{\otimes 2}dt \\
& + \int_0^\tau 2E[Y_i(t)\left\{Z_i(t)-e(t)\right\}\bar{\epsilon}_{i\cdot}^{T}]dt + o_{a.s.}(1) \\
= \;& n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2}dt + n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\bar{\epsilon}_{i\cdot}^{\otimes 2}dt \\
& + \int_0^\tau 2E[Y_i(t)\left\{Z_i(t)-e(t)\right\}E(\bar{\epsilon}_{i\cdot}^{T})]dt + o_{a.s.}(1) \\
= \;& n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2}dt + n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\bar{\epsilon}_{i\cdot}^{\otimes 2}dt + o_{a.s.}(1) \\
= \;& n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2}dt + \int_0^\tau \lim_{n\to\infty}\left[n^{-1}\sum_{i=1}^{n}E\{Y_i(t)\bar{\epsilon}_{i\cdot}^{\otimes 2}\}\right]dt + o_{a.s.}(1) \\
= \;& n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2}dt + \int_0^\tau E\{Y_i(t)\}\lim_{n\to\infty}\left\{n^{-1}\sum_{i=1}^{n}E(\bar{\epsilon}_{i\cdot}^{\otimes 2})\right\}dt + o_{a.s.}(1) \\
= \;& n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2}dt + \int_0^\tau E\{Y_i(t)\}\lim_{n\to\infty}\left\{n^{-1}\sum_{i=1}^{n}\Sigma_1/n_i\right\}dt + o_{a.s.}(1) \\
= \;& n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2}dt + \int_0^\tau \rho_0 E\{Y_i(t)\}\Sigma_1 dt + o_{a.s.}(1),
\end{aligned}
$$

where the last two steps follow from the definition of Σ_1 and ρ_0.

Part 2: Now we examine the second term of D_c:

$$\begin{aligned}
& n^{-1}\int_0^\tau \left\{1-\frac{1}{\sum_{j=1}^n Y_j(t)}\right\}\sum_{i=1}^n \left\{Y_i(t)\hat{\Sigma}_1/n_i\right\}dt \\
= \; & n^{-1}\int_0^\tau \sum_{i=1}^n \left\{Y_i(t)\hat{\Sigma}_1/n_i\right\}dt - n^{-1}\hat{\Sigma}_1\int_0^\tau \frac{n^{-1}\sum_{i=1}^n Y_i(t)/n_i}{n^{-1}\sum_{i=1}^n Y_i(t)}dt \\
= \; & n^{-1}\int_0^\tau \sum_{i=1}^n \left\{Y_i(t)\hat{\Sigma}_1/n_i\right\}dt - n^{-1}\hat{\Sigma}_1\int_0^\tau \frac{\lim_{n\to\infty}\left[n^{-1}\sum_{i=1}^n E\{Y_i(t)\}/n_i\right]}{E\{Y_i(t)\}}dt + o_{a.s.}(1) \\
= \; & n^{-1}\int_0^\tau \sum_{i=1}^n \left\{Y_i(t)\hat{\Sigma}_1/n_i\right\}dt - n^{-1}\hat{\Sigma}_1\rho_0\tau + o_{a.s.}(1) \\
= \; & n^{-1}\int_0^\tau \sum_{i=1}^n \left\{Y_i(t)\hat{\Sigma}_1/n_i\right\}dt - n^{-1}\Sigma_1\rho_0\tau + o_{a.s.}(1) \\
= \; & n^{-1}\int_0^\tau \sum_{i=1}^n \left\{Y_i(t)\hat{\Sigma}_1/n_i\right\}dt + o_{a.s.}(1) \\
= \; & \int_0^\tau \rho_0 E\{Y_i(t)\}\Sigma_1 dt + o_{a.s.}(1).
\end{aligned}$$

As a result, combining parts 1 and 2 gives

$$\begin{aligned}
D_c &= n^{-1}\sum_{i=1}^n \int_0^\tau Y_i(t)\left\{Z_i(t)-e(t)\right\}^{\otimes 2}dt + \int_0^\tau \rho_0 E\{Y_i(t)\}\Sigma_1 dt - \int_0^\tau \rho_0 E\{Y_i(t)\}\Sigma_1 dt + o_{a.s.}(1) \\
&= \mathcal{D}_c + o_{a.s.}(1).
\end{aligned}$$

The proof of Lemma A.1 is now completed.

From the the proof of Lemma A.1, we obtain that the inverse matrix in $\hat{\beta}_c$, i.e., D_c, converges almost surely to a positive definite matrix under mild regularity conditions. Thus, the estimator $\hat{\beta}_c$ does not have the singularity and unstability issues.

Lemma A.2 *Under Regularity Conditions R1-R8, $\hat{\beta}_c$ converges to β almost surely as $n \to \infty$.*

Proof: Recall that we have proved in Appendix A2 that

$$n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left\{\hat{Z}_i(t)-\tilde{Z}(t)\right\}dN_i(t)=E\left[\int_0^{\tau}\{Z_i(t)-e(t)\}\,dN_i(t)\right]+o_{a.s.}(1).$$

Combined with Lemma A.1, we obtain that

$$\begin{aligned}\hat{\beta}_c &= D_c^{-1}n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left\{\hat{Z}_i(t)-\tilde{Z}(t)\right\}dN_i(t)\\ &= D_c^{-1}E\left[\int_0^{\tau}\{Z_i(t)-e(t)\}\,dN_i(t)\right]+o_{a.s.}(1)\\ &= \mathcal{D}_c^{-1}E\left[\int_0^{\tau}\{Z_i(t)-e(t)\}\,dN_i(t)\right]+o_{a.s.}(1)\\ &= \beta+o_{a.s.}(1).\end{aligned}$$

The proof of Lemma A.2 is now completed.

We now return to the proof of Theorem 1. Note that $U_c(\beta)=U_1-U_2$, where

$$\begin{aligned}U_1 &= \sum_{i=1}^{n}\int_0^{\tau}\left\{\hat{Z}_i(t)-\tilde{Z}(t)\right\}d\tilde{M}_i(t;\beta,\Lambda_0),\\ \text{and}\quad U_2 &= \int_0^{\tau}\left\{1-\frac{1}{\sum_{j=1}^{n}Y_j(t)}\right\}\sum_{i=1}^{n}\left\{Y_i(t)\hat{\Sigma}_1\beta/n_i\right\}dt.\end{aligned}$$

We now separately study the asymptotic expansion of $n^{-1/2}U_1$ and $n^{-1/2}U_2$.

By analogy to the proof of Theorem 1 of Kulich and Lin (2000), we have

$$n^{-1/2}U_1=n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{\hat{Z}_i(t)-e(t)\right\}d\tilde{M}_i(t;\beta,\Lambda_0)+o_p(1).$$

Note that

$$
\begin{aligned}
& n^{-1/2}U_2 \\
= \;& n^{-1/2}\int_0^\tau \sum_{i=1}^n \left\{Y_i(t)\hat{\Sigma}_1\beta/n_i\right\} dt - n^{-1/2}\int_0^\tau \frac{1}{\sum_{j=1}^n Y_j(t)} \sum_{i=1}^n \left\{Y_i(t)\hat{\Sigma}_1\beta/n_i\right\} dt \\
= \;& n^{-1/2}\int_0^\tau \sum_{i=1}^n \left\{Y_i(t)\hat{\Sigma}_1\beta/n_i\right\} dt + o_p(1) \\
= \;& n^{1/2}(\hat{\Sigma}_1 - \Sigma_1)\rho_0 E(S_i)\beta + n^{-1/2}\int_0^\tau \sum_{i=1}^n \{Y_i(t)\Sigma_1\beta/n_i\} dt + o_p(1) \\
= \;& n^{1/2}\frac{\sum_{i=1}^n \sum_{r=1}^{n_i}(W_{ir} - \bar{W}_{i\cdot})^{\otimes 2}}{\sum_{i=1}^n (n_i - 1)}\rho_0 E(S_i)\beta - n^{1/2}\Sigma_1\rho_0 E(S_i)\beta + n^{-1/2}\sum_{i=1}^n S_i\Sigma_1\beta/n_i + o_p(1) \\
= \;& n^{-1/2}\rho_0\rho_1^{-1}\sum_{i=1}^n \sum_{r=1}^{n_i}(W_{ir} - \bar{W}_{i\cdot})^{\otimes 2}E(S_i)\beta - n^{1/2}\Sigma_1\rho_0 E(S_i)\beta + n^{-1/2}\sum_{i=1}^n S_i\Sigma_1\beta/n_i + o_p(1),
\end{aligned}
$$

where the third identity comes from that

$$
n^{1/2}\left[\frac{\int_0^\tau \sum_{i=1}^n \{Y_i(t)/n_i\}\, dt}{n} - \rho_0 E(S_i)\right](\hat{\Sigma}_1 - \Sigma_1)\beta = o_p(1).
$$

Therefore,

$$
n^{-1/2}U_c(\beta) = n^{-1/2}\sum_{i=1}^n U_{c,i} + o_p(1),
$$

where

$$
U_{c,i} = \int_0^\tau \left\{\hat{Z}_i(t) - e(t)\right\} d\tilde{M}_i(t;\beta,\Lambda_0) + \rho_0\rho_1^{-1}\sum_{r=1}^{n_i}(W_{ir} - \bar{W}_{i\cdot})^{\otimes 2}E(S_i)\beta - \rho_0 E(S_i)\Sigma_1\beta + \frac{S_i\Sigma_1\beta}{n_i}.
$$

By the Taylor series expansion, $0 = n^{-1/2}U_c(\hat{\beta}_c) = n^{-1/2}U_c(\beta) + \left[n^{-1}\frac{\partial U_c(\beta)}{\partial\beta}\right] n^{1/2}(\hat{\beta}_c - \beta)$, we obtain that

$$
n^{1/2}(\hat{\beta}_c - \beta) = -\left[n^{-1}\frac{\partial U_c(\beta)}{\partial\beta}\right]^{-1} n^{-1/2}U_c(\beta).
$$

By the derivation in Appendix A2,

$$\begin{aligned}
-n^{-1}\frac{\partial U_c(\beta)}{\partial \beta} &= n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{\hat{Z}_i(t)-\tilde{Z}(t)\right\}^{\otimes 2} dt - n^{-1}\int_0^{\tau}\sum_{i=1}^{n}\left\{Y_i(t)\hat{\Sigma}_1/n_i\right\} dt \\
&\stackrel{a.s.}{\rightarrow} \lim_{n\to\infty} n^{-1}E\left[\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{\hat{Z}_i(t)-e(t)\right\}^{\otimes 2} dt\right] - \rho_0 E(S_i)\Sigma_1 \\
&= \mathcal{D}_c.
\end{aligned}$$

By Condition R6, $E(||n^{-1/2}U_{c,i}||^2)I(||n^{-1/2}U_{c,i}|| > \epsilon)$ can only take at most N_0 possible values for a given $\epsilon > 0$. Without loss of generality, suppose when $i = 1$, it achieves the maximum value. It follows from the Markov inequality that $\Pr\{||n^{-1/2}U_{c,1}|| > \epsilon\} \leq n^{-1}E(||U_{c,1}||^2)/\epsilon^2 \to 0$ as $n \to \infty$, and thus

$$\sum_{i=1}^{n} E(||n^{-1/2}U_{c,i}||^2)I\{||n^{-1/2}U_{c,i}|| > \epsilon\} \leq E(||U_{c,1}||^2)I\{||n^{-1/2}U_{c,1}|| > \epsilon\} \to 0 \quad \text{as} \quad n \to \infty,$$

suggesting that the Lindeberg condition (van der Vaart 1998, p.20) is satisfied.

By the multivariate Lindeberg-Feller Central Limit Theorem (van der Vaart 1998, p.20), we obtain that $n^{-1/2}U_c(\beta)$ is asymptotically normal with mean 0 and covariance matrix $\Sigma_c = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E(U_{c,i})^{\otimes 2}$. It follows that that $n^{1/2}(\hat{\beta}_c - \beta)$ is asymptotically normal with mean zero and covariance matrix $\mathcal{D}_c^{-1}\Sigma_c\mathcal{D}_c^{-T}$.

Appendix A5: Proof of Theorem 2

It can be calculated that

$$n^{1/2}\{\hat{\Lambda}_0(t;\hat{\beta}_c) - \Lambda_0(t)\} = n^{1/2}\int_0^t \frac{\sum_{i=1}^{n} d\tilde{M}_i(u;;\beta,\Lambda_0)}{\sum_{j=1}^{n} Y_j(u)} - \left\{\int_0^t \frac{\sum_{i=1}^{n} Y_i(u)\hat{Z}_i^T(u)}{\sum_{j=1}^{n} Y_j(u)} du\right\} n^{1/2}(\hat{\beta}_c - \beta).$$

By the arguments similar to those in Appendix A2, we obtain that $n^{1/2}\int_0^t \{\sum_{j=1}^{n} Y_j(u)\}^{-1}\sum_{i=1}^{n} d\tilde{M}_i(u;\beta,\Lambda_0)$ converges weakly to a Gaussian process, and thus is tight. $\int_0^t \{\sum_{j=1}^{n} Y_j(u)\}^{-1}\sum_{i=1}^{n} Y_i(u)\hat{Z}_i^T(u)du$ converges almost surely to $\int_0^t e^T(u)du$, and thus $\int_0^t \sum_{i=1}^n Y_i(u)\hat{Z}_i^T(u)/\sum_{j=1}^n Y_j(u)du$ is tight. Since tightness in $l^\infty[0,\tau]$ equipped with the uniform metric has the additivity property, and $n^{1/2}(\hat{\beta}_c - \beta)$ is asymptotic normal, we obtain that $n^{1/2}\{\hat{\Lambda}_0(t;\hat{\beta}_c) - \Lambda_0(t)\}$ is tight.

It follows from Appendix A4 that

$$n^{1/2}(\hat{\beta}_c - \beta) = n^{-1/2}\mathcal{D}_c^{-1}\sum_{i=1}^n\Bigg[\int_0^\tau \left\{\hat{Z}_i(t) - e(t)\right\} d\tilde{M}_i(t;\beta,\Lambda_0) \\ +\rho_0\rho_1^{-1}\sum_{r=1}^{n_i}(W_{ir} - \bar{W}_{i\cdot})^{\otimes 2}E(S_i)\beta - \rho_0 E(S_i)\Sigma_1\beta + \frac{S_i\Sigma_1\beta}{n_i}\Bigg] + o_p(1).$$

Thus, $n^{1/2}\left\{\hat{\Lambda}_0(t;\hat{\beta}_c) - \Lambda_0(t)\right\}$ is asymptotically equivalent to $n^{-1/2}\sum_{i=1}^n \Psi_i(t)$ uniformly in t. Similar to the proof of the asymptotic normality of $\hat{\beta}_c$, it is shown that $n^{-1/2}\sum_{i=1}^n \Psi_i(t)$ satisfies the Lindeberg condition, and thus the multivariate Lindeberg-Feller Central Limit Theorem applies to $n^{1/2}\left\{\hat{\Lambda}_0(t;\hat{\beta}_c) - \Lambda_0(t)\right\}$. Together with the fact that $n^{1/2}\{\hat{\Lambda}_0(t;\hat{\beta}_c) - \Lambda_0(t)\}$ is tight, the weak convergence result is proved.

Appendix A6: Asymptotic properties of $\hat{\beta}_c$ and $\hat{\Lambda}_0(t;\hat{\beta}_c)$ when Σ_1 is simply known

Now, we investigate the asymptotic property of $\hat{\beta}_c$ when Σ_1 is simply known. In this case, note that

$$\begin{aligned} n^{-1/2}U_2 &= n^{-1/2}\int_0^\tau \sum_{i=1}^n \{Y_i(t)\Sigma_1\beta/n_i\}\, dt \\ &= n^{-1/2}\sum_{i=1}^n S_i\Sigma_1\beta/n_i. \end{aligned}$$

It then follows from the proof of Theorem 1 that the asymptotic distribution of $\hat{\beta}_c$ is almost identical as the one in Theorem 1, with the only difference that $\Sigma_2\beta - \rho_0 E(S_i)\Sigma_1\beta$ is removed from Σ_c.

Now, we investigate the asymptotic property of $\hat{\Lambda}_0(t;\hat{\beta}_c)$ when Σ_1 is simply known. In this case, note that

$$n^{1/2}(\hat{\beta}_c - \beta) = n^{-1/2}\mathcal{D}_c^{-1}\sum_{i=1}^{n}\left[\int_0^\tau \left\{\hat{Z}_i(t) - e(t)\right\} d\tilde{M}_i(t;\beta,\Lambda_0) + \frac{S_i\Sigma_1\beta}{n_i}\right] + o_p(1).$$

It then follows from the proof of Theorem 2 that the $n^{1/2}\{\hat{\Lambda}_0(t;\hat{\beta}_c) - \Lambda_0(t)\}$ is tight, and the limiting process is almost identical as the one in Theorem 2, with the only difference that $\Sigma_2\beta - \rho_0 E(S_i)\Sigma_1\beta$ is removed from $\Psi_i(t)$.

Appendix A7: Proof of Theorem 3

Let $\tilde{W}_n(t) = n^{-1/2}\sum_{i=1}^n \xi_i\Psi_i(t)$. By the proof of the weak convergence of $n^{1/2}\{\hat{\Lambda}_0(t;\hat{\beta}_c) - \Lambda_0(t)\}$, $n^{-1/2}\sum_{i=1}^n \Psi_i(t)$ converges weakly to $\mathcal{G}(t)$ unconditionally. Since weak convergence of $n^{-1/2}\sum_{i=1}^n \Psi_i(t)$ implies that the Donsker condition (van der Vaart and Wellner 1996, Theorem 2.9.6) holds, it then follows from the conditional multiplier Central Limit Theorem (van der Vaart and Wellner 1996, Sec. 2.9) that $\tilde{W}_n(t)$ converges weakly to $\mathcal{G}(t)$ in probability conditional on the data. Thus, by Lemma 1 of Pipper and Ritz (2007), it suffices to show that $\sup_{t\in[0,\tau]}|\hat{W}_n(t) - \tilde{W}_n(t)| \xrightarrow{p} 0$.

Let $\hat{M}_i(t)$ be the empirical version of $\tilde{M}_i(t)$, $i = 1, \cdots, n$. Note that

$$\sup_{t\in[0,\tau]}|\hat{W}_n(t) - \tilde{W}_n(t)| \le \sup_{t\in[0,\tau]}|W_n^{(1)}(t)| + \sup_{t\in[0,\tau]}|W_n^{(2)}(t)| + \sup_{t\in[0,\tau]}|W_n^{(3)}(t)|,$$

where

$$
\begin{aligned}
W_n^{(1)}(t) &= n^{-1/2}\sum_{i=1}^{n}\xi_i\left\{\int_0^t \frac{d\hat{M}_i(u;\beta,\Lambda_0)}{n^{-1}\sum_{i=1}^n Y_i(u)} - \int_0^t \frac{d\tilde{M}_i(u;\beta,\Lambda_0)}{E[Y_i(u)]}\right\}, \\
W_n^{(2)}(t) &= \int_0^t \tilde{Z}^T(u)du \times \hat{\mathcal{D}}_c^{-1} \times n^{-1/2}\sum_{i=1}^{n}\xi_i\left[\int_0^\tau \left\{\hat{Z}_i(t)-\tilde{Z}(t)\right\} d\hat{M}_i(t;\beta,\Lambda_0)\right] \\
&\quad - \int_0^t e^T(u)du \times \mathcal{D}_c^{-1} \times n^{-1/2}\sum_{i=1}^{n}\xi_i\left[\int_0^\tau \left\{\hat{Z}_i(t)-e(t)\right\} d\tilde{M}_i(t;\beta,\Lambda_0)\right], \\
\text{and}\quad W_n^{(3)}(t) &= \int_0^t \tilde{Z}^T(u)du \times \hat{\mathcal{D}}_c^{-1} \times n^{-1/2}\sum_{i=1}^{n}\frac{\xi_i S_i \hat{\Sigma}_1\hat{\beta}_c}{n_i} \\
&\quad - \int_0^t e^T(u)du \times \mathcal{D}_c^{-1} \times n^{-1/2}\sum_{i=1}^{n}\frac{\xi_i S_i \Sigma_1 \beta_0}{n_i}.
\end{aligned}
$$

Employing the empirical process techniques used in Appendix A2, together with Lemma A.3 and Theorem 2 of Spiekerman and Lin (1998), we show that

$$
\sup_{t\in[0,\tau]} |W_n^{(j)}(t)| \xrightarrow{p} 0, j=1,2,3,
$$

and thus $\sup_{t\in[0,\tau]} |\hat{W}_n(t) - \tilde{W}_n(t)| \xrightarrow{p} 0$ holds.

Appendix A8: Proof of Corrolary 1

Note that the only difference between $\hat{\beta}_c$ and $\hat{\beta}_e$ is the term

$$
\int_0^\tau \left\{\frac{1}{\sum_{j=1}^n Y_j(t)}\right\}\sum_{i=1}^{n}\left\{Y_i(t)\hat{\Sigma}_1\beta/n_i\right\} dt
$$

in their denominators. By the arguments in Appendix A2, we have

$$
n^{-1/2}\int_0^\tau \left\{\frac{1}{\sum_{j=1}^n Y_j(t)}\right\}\sum_{i=1}^{n}\left\{Y_i(t)\hat{\Sigma}_1\beta/n_i\right\} dt = o_p(1).
$$

Therefore, $\hat{\beta}_e$ is asymptotically identical to $\hat{\beta}_c$, suggesting that the asymptotic normal distribution of $\hat{\beta}_e$ is identical to that of $\hat{\beta}_c$. Similarly, the limit process of $n^{1/2}\left\{\hat{\Lambda}_0(t;\hat{\beta}_e) - \Lambda_0(t)\right\}$ is identical to that of $n^{1/2}\left\{\hat{\Lambda}_0(t;\hat{\beta}_c) - \Lambda_0(t)\right\}$.

Appendix A9: Proof of Lemma 1

Using the proof of Theorem 1 and SULLN, we can show that

$$n^{-1/2}U_c(\beta, t) = n^{-1/2}\sum_{i=1}^{n} A_i(t) + o_p(1)$$

uniformly in t. Note that by the derivations in Appendix A4,

$$n^{1/2}(\hat{\beta}_c - \beta) = \mathcal{D}_c^{-1}(\tau)n^{-1/2}\sum_{i=1}^{n} A_i(\tau) + o_p(1).$$

Thus, applying the Taylor series expansion gives

$$\begin{aligned} n^{-1/2}U_c(\hat{\beta}_c, t) &= n^{-1/2}U_c(\beta, t) + \left[n^{-1}\frac{\partial U_c(\beta, t)}{\partial \beta}\right] n^{1/2}(\hat{\beta}_c - \beta) \\ &= n^{-1/2}\sum_{i=1}^{n} A_i(t) - D_c(t)\mathcal{D}_c^{-1}(\tau)n^{-1/2}\sum_{i=1}^{n} A_i(\tau) + o_p(1) \\ &= n^{-1/2}\sum_{i=1}^{n}\left\{A_i(t) - \mathcal{D}_c(t)\mathcal{D}_c^{-1}(\tau)A_i(\tau)\right\} + o_p(1) \end{aligned}$$

uniformly in t. The tightness of $n^{-1/2}\sum_{i=1}^{n}\{A_i(t) - \mathcal{D}_c(t)\mathcal{D}_c^{-1}(\tau)A_i(\tau)\}$ can be shown using the arguments similar to those in the proof of Theorem 2. The proof then follows.

Appendix A10: Proof of Theorem 4

The arguments are similar to those in Appendix A7 by adopting the conditional multiplier Central Limit Theorem, and thus omitted.

Appendix B1: Bias Corrected Estimators

It is useful to note that the relationship (5.6) between the limit of the naive estimator and the true value of the parameter provides a flexible way to construct a consistent estimator of

β. Indeed, using the inverse version of (5.6), we obtain a class of bias-corrected estimators for β as

$$\hat{\beta} = \hat{B}_1^{-1}(\hat{B}_1 + \hat{B}_2)\hat{\beta}_{nv}, \tag{5.19}$$

where $\hat{B}_1$ and $\hat{B}_2$ are any "reasonable" estimators of B_1 and B_2, respectively. By the slutsky Theorem, as long as $\hat{B}_1$ and $\hat{B}_2$ are consistent estimators of B_1 and B_2, respectively, the resulting estimator $\hat{\beta}$ is consistent (e.g., Yi and Reid 2010). Depending on the choices of consistent estimators of B_1 and B_2, various consistent estimators of β can be constructed. We comment that the nonparametric correction estimator by Sun, Zhang and Sun (2006) is a special case of $\hat{\beta}$ in (5.19).

We conclude this subsection with the development of a new estimator by using the result of (5.19) for the case with equal $n_i \geq 2$. Noting a key property that

$$E\left[\left\{\hat{Z}_{ir}(t) - \tilde{Z}_r(t)\right\}\left\{\hat{Z}_{is}(t) - \tilde{Z}_s(t)\right\}^T \middle| \mathcal{F}_\tau\right] = \{Z_i(t) - \bar{Z}(t)\}^{\otimes 2}$$

for any $1 \leq r \neq s \leq n_i$, we estimate B_1 and B_2 consistently by

$$\begin{aligned} \hat{B}_1 &= n^{-1}\sum_{i=1}^n \int_0^\tau \frac{Y_i(t)}{n_i(n_i-1)} \sum_{1\leq r\neq s\leq n_i} \left\{\hat{Z}_{ir}(t) - \tilde{Z}_r(t)\right\}\left\{\hat{Z}_{is}(t) - \tilde{Z}_s(t)\right\}^T dt \\ \text{and}\quad \hat{B}_2 &= n^{-1}\sum_{i=1}^n \int_0^\tau Y_i(t)\{\hat{Z}_i(t) - \tilde{Z}(t)\}^{\otimes 2} dt - \hat{B}_1, \end{aligned}$$

respectively. Then using (5.19) gives us a consistent estimator of β:

$$\begin{aligned} \hat{\beta} &= \left[\sum_{i=1}^n \int_0^\tau \frac{Y_i(t)}{n_i(n_i-1)} \sum_{1\leq r\neq s\leq n_i} \left\{\hat{Z}_{ir}(t) - \tilde{Z}_r(t)\right\}\left\{\hat{Z}_{is}(t) - \tilde{Z}_s(t)\right\}^T dt\right]^{-1} \\ &\quad \times \left[\sum_{i=1}^n \int_0^\tau \left\{\hat{Z}_i(t) - \tilde{Z}(t)\right\} dN_i(t)\right]. \end{aligned}$$

Simulation studies suggest that this new estimator performs similarly to the nonparametric correction estimator by Sun, Zhang and Sun (2006). In fact, these two estimators are asymptotically equivalent. Too see this, note that Note that

$$\sqrt{n}\left\{\tilde{Z}_r(t) - e(t)\right\}\left\{\tilde{Z}_s(t) - e(t)\right\}^T = o_p(1).$$

Thus,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \frac{Y_i(t)}{n_i(n_i-1)}\sum_{1\le r\ne s\le n_i}\left\{\hat{Z}_{ir}(t)-\tilde{Z}_r(t)\right\}\left\{\hat{Z}_{is}(t)-\tilde{Z}_s(t)\right\}^T dt$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \frac{Y_i(t)}{n_i(n_i-1)}\sum_{1\le r\ne s\le n_i}\hat{Z}_{ir}(t)\hat{Z}_{is}(t)^T dt - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\tilde{Z}(t)e(t)^T dt$$
$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau Y_i(t)e(t)\tilde{Z}(t)^T dt + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau Y_i(t)e^{\otimes 2}(t)dt + o_p(1).$$

Thus, it follows that $\hat{\beta}$ has the same asymptotic distribution as that of the nonparametric correction estimator of Sun, Zhang and Sun (2006).

Appendix B2: Regression Calibration Estimator

As the survival information on $T_i \ge t$ is required in (5.5), evaluation of the conditional expectation $E\{X_i|T_i \ge t, \bar{W}_{i\cdot}, V_i(t)\}$ is generally difficult, unless certain simplistic assumptions are imposed. Under the rare event assumption (Prentice 1982), for example, we write $E\{X_i|T_i \ge t, \bar{W}_{i\cdot}, V_i(t)\} \approx E\{X_i|\bar{W}_{i\cdot}, V_i(t)\}$. This approximation allows us to invoke the best linear approximation method (Carroll, et al. 2006):

$$\hat{E}\{X_i|\bar{W}_{i\cdot}, V_i(t)\} = \mu_x + (\Sigma_{xx}, \Sigma_{xv})\begin{bmatrix}\Sigma_{xx}+\Sigma_0/n_i & \Sigma_{xv}\\ \Sigma_{xv}^T & \Sigma_{vv}\end{bmatrix}^{-1}\begin{pmatrix}\bar{W}_{i\cdot}-\mu_x\\ V_i(t)-\mu_v\end{pmatrix},$$

where $\mu_x = E(X_i)$, $\mu_v = E\{V_i(t)\}$, $\Sigma_{xx} = \mathrm{Var}(X_i)$, $\Sigma_{vv} = \mathrm{Var}\{V_i(t)\}$, and $\Sigma_{xv} = \mathrm{Cov}\{X_i, V_i(t)\}$.

Let $\hat{X}_i^*(t)$ be the empirical version of $\hat{E}\{X_i|\bar{W}_{i\cdot}, V_i(t)\}$, with μ_x, μ_v, Σ_{xx}, Σ_{xv}, Σ_{vv} and Σ_0 replaced with their empirical estimates. Specifically, define

$$A_{rc,i} = \begin{bmatrix}A_{1,i} & A_{2,i}\\ A_{3,i} & A_{4,i}\end{bmatrix} = \begin{bmatrix}\Sigma_{xx} & \Sigma_{xv}\\ \Sigma_{xv}^T & \Sigma_{vv}\end{bmatrix}\begin{bmatrix}\Sigma_{xx}+\Sigma_0/n_i & \Sigma_{xv}\\ \Sigma_{xv}^T & \Sigma_{vv}\end{bmatrix}^{-1},$$

where $A_{1,i}$, $A_{2,i}$, $A_{3,i}$ and $A_{4,i}$ are $p \times p$, $p \times q$, $q \times p$, $q \times q$ matrices, respectively. After some matrix algebra, $A_{1,i} = (\Sigma_{xx} - \Sigma_{xv}\Sigma_{vv}^{-1}\Sigma_{xv}^T)(\Sigma_0/n_i + \Sigma_{xx} - \Sigma_{xv}\Sigma_{vv}^{-1}\Sigma_{xv}^T)^{-1}$, $A_{2,i} = (I - A_{1,i})\Sigma_{xv}\Sigma_{vv}^{-1}$, $A_{3,i} = O$, $A_{4,i} = I$.

We estimate Σ_{xx}, Σ_{xv}, Σ_{vv}, μ_x and $\mu_{v(t)}$ by $\hat{\Sigma}_{xx}$, $\hat{\Sigma}_{xv}$, $\hat{\Sigma}_{vv}$, $\bar{W}_{\cdot\cdot} = \sum_{i=1}^n \sum_{j=1}^{n_i} W_{ij} / \sum_{i=1}^n n_i$ and $\bar{V}_{\cdot}(t)$, respectively, where

$$\hat{\Sigma}_{vv} = (n-1)^{-1} \sum_{i=1}^n (V_i(t) - \bar{V}_{\cdot}(t))(V_i(t) - \bar{V}_{\cdot}(t))^T,$$

$$\hat{\Sigma}_{xv} = \frac{\sum_{i=1}^n n_i}{(\sum_{i=1}^n n_i)^2 - \sum_{i=1}^n n_i^2} \sum_{i=1}^n n_i(\bar{W}_{i\cdot} - \bar{W}_{\cdot\cdot})(V_i(t) - \bar{V}_{\cdot}(t))^T,$$

$$\text{and} \quad \hat{\Sigma}_{xx} = \frac{\sum_{i=1}^n n_i}{(\sum_{i=1}^n n_i)^2 - \sum_{i=1}^n n_i^2} \left[\sum_{i=1}^n n_i(\bar{W}_{i\cdot} - \bar{W}_{\cdot\cdot})(\bar{W}_{i\cdot} - \bar{W}_{\cdot\cdot})^T - (n-1)\hat{\Sigma}_0 \right].$$

Let $\hat{A}_{r,i}$ be $A_{r,i}$ with Σ_{xx}, Σ_{xv}, Σ_{vv} and Σ_0 replaced by their empirical estimates, $r = 1, 2$. Let $\hat{X}_i^*(t)$ denote $\hat{E}[X_i | \bar{W}_{i\cdot}, V_i(t)]$ with $(A_{1,i}, A_{2,i})$ replaced by $(\hat{A}_{1,i}, \hat{A}_{2,i})$, i.e., $\hat{X}_i^*(t) = \bar{W}_{\cdot\cdot} + \hat{A}_{1,i}(\bar{W}_{i\cdot} - \bar{W}_{\cdot\cdot}) + \hat{A}_{2,i}(V_i(t) - \bar{V}_{\cdot}(t))$. $\hat{A}_{rc,i}$ is $A_{rc,i}$ with $A_{1,i}$ and $A_{2,i}$ replaced by $\hat{A}_{1,i}$ and $\hat{A}_{2,i}$, respectively.

Under the rare event assumption, the comparision of the induced hazard function (5.5) to the true hazard function form (5.1) suggests that replacing $X_i(t)$ with $\hat{X}_i^*(t)$ in (5.3) can yield an approximately consistent estimator of β. We let $\hat{\beta}_{rc}$ denote this estimator and call it a regression calibration estimator as in Prentice (1982) for the proportional hazards models. Specifically,

$$\hat{\beta}_{rc} = \left[\sum_{i=1}^n \int_0^\tau Y_i(t) \left\{ \hat{Z}_{rc,i}^*(t) - \tilde{Z}_{rc}^*(t) \right\}^{\otimes 2} dt \right]^{-1} \left[\sum_{i=1}^n \int_0^\tau \left\{ \hat{Z}_{rc,i}^*(t) - \tilde{Z}_{rc}^*(t) \right\} dN_i(t) \right],$$

where $\hat{Z}_{rc,i}^*(t) = \left(\hat{X}_i^{*T}(t), V_i^T(t) \right)^T$ and $\tilde{Z}_{rc}^*(t) = \sum_{i=1}^n Y_i(t) \hat{Z}_{rc,i}^*(t) / \sum_{i=1}^n Y_i(t)$.

The regression calibration estimator $\hat{\beta}_{rc}$ is easy to calculate, and it is expected to outperform the naive estimator $\hat{\beta}_{nv}$, especially when the rare event assumption is feasible.

This estimator however can not completely remove the bias induced from covariate measurement error. It is thus interesting and important to study the asymptotic behaviour of $\hat{\beta}_{rc}$.

Now we investigate the properties of the regression calibration estimator $\hat{\beta}_{rc}$. Our exploration is conducted using the naive estimator $\hat{\beta}_{nv}$ as a reference. First we consider the circumstance where every subject has an equal number n_i of replicated measurements W_{ir}, i.e., all n_i are equal. In this case, there is a simple relationship between $\hat{\beta}_{rc}$ and $\hat{\beta}_{nv}$:

$$\hat{\beta}_{rc} = [\hat{A}_{rc,i}^{-1}]^T \hat{\beta}_{nv}, \tag{5.20}$$

where $\hat{A}_{rc,i}$ is the empirical version of $A_{rc,i}$, and

$$A_{rc,i} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xv} \\ \Sigma_{xv}^T & \Sigma_{vv} \end{bmatrix} \begin{bmatrix} \Sigma_{xx} + \Sigma_0/n_i & \Sigma_{xv} \\ \Sigma_{xv}^T & \Sigma_{vv} \end{bmatrix}^{-1}.$$

Furthermore, it can be shown that

$$\hat{\beta}_{rc} \xrightarrow{a.s.} \beta_{rc}^*, \quad \text{as} \quad n \to \infty, \tag{5.21}$$

where $\beta_{rc}^* = [A_{rc,i}^{-1}]^T \beta_{nv}^* = [A_{rc,i}^{-1}]^T (B_1 + B_2)^{-1} B_1 \beta$. Comparing (5.21) to (5.6) implies that the regression calibration estimator $\hat{\beta}_{rc}$ is not exactly but only an approximately consistent estimator of β, as its analogue for the proportional hazards model (Prentice 1982, Wang et al. 1997).

Under the more general situation where replicate numbers n_i are not necessarily equal, the preceding discussion can carry through with more notation involved. Here we end this subsection with the description of the asymptotic normality property of $\hat{\beta}_{rc}$.

Theorem B1 *Under Regularity Conditions R1-R8 listed in the Appendix A1, we have*

$$n^{1/2}(\hat{\beta}_{rc} - \beta_{rc}^*) \xrightarrow{d} N(0, \Sigma_{rc}^*), \quad as \quad n \to \infty,$$

where Σ_{rc}^* is specified in the following.

Proof: First we confine ourselves to the case that all n_i are equal. Extension to the general case is deferred until the end. Note that

$$A_{rc,1} = I_{(p+q)\times(p+q)} + \begin{bmatrix} -\Sigma_0/n_1 & 0_{p\times q} \\ 0_{q\times p} & 0_{q\times q} \end{bmatrix} \begin{bmatrix} \Sigma_{xx} + \Sigma_0/n_1 & \Sigma_{xv} \\ \Sigma_{xv}^T & \Sigma_{vv} \end{bmatrix}^{-1},$$

and thus the lower left $q \times p$ block and the lower right $q \times q$ block of $A_{rc,1}$ is $0_{q\times p}$ and $I_{q\times p}$, respectively. It follows that $\hat{Z}^*_{rc,i}(t) = \hat{A}_{rc,1}\hat{Z}_i(t) + (I_{(p+q)\times(p+q)} - \hat{A}_{rc,1})\hat{\mu}_z$. Thus,

$$\begin{aligned}
\hat{\beta}_{rc} &= \left[\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{\hat{Z}^*_{rc,i}(t) - \tilde{Z}^*_{rc}(t)\right\}^{\otimes 2} dt\right]^{-1} \left[\sum_{i=1}^{n}\int_0^{\tau}\left\{\hat{Z}^*_{rc,i}(t) - \tilde{Z}^*_{rc}(t)\right\} dN_i(t)\right] \\
&= \left[\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\hat{A}_{rc,i}\left\{\hat{Z}_i(t) - \tilde{Z}(t)\right\}^{\otimes 2} \hat{A}^T_{rc,i}dt\right]^{-1} \left[\sum_{i=1}^{n}\int_0^{\tau}\hat{A}_{rc,i}\left\{\hat{Z}_i(t) - \tilde{Z}(t)\right\} dN_i(t)\right] \\
&= \left[\hat{A}_{rc,1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{\hat{Z}_i(t) - \tilde{Z}(t)\right\}^{\otimes 2} dt\hat{A}^T_{rc,1}\right]^{-1} \left[\hat{A}_{rc,1}\sum_{i=1}^{n}\int_0^{\tau}\left\{\hat{Z}_i(t) - \tilde{Z}(t)\right\} dN_i(t)\right] \\
&= \hat{A}^{-1T}_{rc,1}\left[\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{\hat{Z}_i(t) - \tilde{Z}(t)\right\}^{\otimes 2} dt\right]^{-1} \hat{A}^{-1}_{rc,1}\hat{A}_{rc,1}\left[\sum_{i=1}^{n}\int_0^{\tau}\left\{\hat{Z}_i(t) - \tilde{Z}(t)\right\} dN_i(t)\right] \\
&= [\hat{A}^{-1}_{rc,1}]^T\hat{\beta}_{nv}.
\end{aligned}$$

It follows that $\hat{\beta}_{rc} \xrightarrow{a.s.} \beta^*_{rc} = [A^{-1}_{rc,1}]^T\beta^*_{nv} = [A^{-1}_{rc,1}]^T(B_1 + B_2)^{-1}B_1\beta$, as $n \to \infty$. Together with the fact that

$$\sqrt{n}([\hat{A}^{-1}_{rc,1}]^T - [A^{-1}_{rc,1}]^T)(\hat{\beta}_{nv} - \beta^*_{nv}) = o_p(1),$$

we obtain

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_{rc} - \beta^*_{rc}) &= \sqrt{n}([\hat{A}^{-1}_{rc,1}]^T\hat{\beta}_{nv} - [A^{-1}_{rc,1}]^T\beta^*_{nv}) \\
&= \sqrt{n}([A^{-1}_{rc,1}]^T\hat{\beta}_{nv} - [\hat{A}^{-1}_{rc,1}]^T\beta^*_{nv}) + o_p(1) \\
&= \sqrt{n}[A^{-1}_{rc,1}]^T(\hat{\beta}_{nv} - \beta^*_{nv}) - \sqrt{n}([\hat{A}^{-1}_{rc,1}]^T - [A^{-1}_{rc,1}]^T)\beta^*_{nv} + o_p(1). \quad (5.22)
\end{aligned}$$

The asymptotic expansion of $\sqrt{n}[A^{-1}_{rc,1}]^T(\hat{\beta}_{nv} - \beta^*_{nv})$ was obtained in the previous sections. Thus, to derive the asymptotic expansion of $\sqrt{n}(\hat{\beta}_{rc} - \beta^*_{rc})$, by (5.22) we only need to derive that of $\sqrt{n}([\hat{A}^{-1}_{rc,1}]^T - [A^{-1}_{rc,1}]^T)\beta^*_{nv}$. This can be done, in principle, by Taylor series expansion for example, but it is very complicated for the multivariate case. In the following, we first study the asymptotic expansion for the univariate case, and then the multivariate case.

Observe that by Taylor series expansion, we obtain

$$\begin{aligned}
\sqrt{n}([\hat{A}_{rc,1}^{-1}]^T - [A_{rc,1}^{-1}]^T) &= \sqrt{n}\left(\frac{\hat{\Sigma}_{xx} + \hat{\Sigma}_0/n_i}{\hat{\Sigma}_{xx}} - \frac{\Sigma_{xx} + \Sigma_0/n_i}{\Sigma_{xx}}\right) \\
&= \frac{\sqrt{n}\left(\hat{\Sigma}_{xx} + \hat{\Sigma}_0/n_i\right)}{\Sigma_{xx}} - \frac{\Sigma_{xx} + \Sigma_0/n_i}{\Sigma_{xx}^2}\sqrt{n}(\hat{\Sigma}_{xx} - \Sigma_{xx}) \\
&\quad - \frac{\sqrt{n}\left(\Sigma_{xx} + \Sigma_0/n_i\right)}{\Sigma_{xx}} + o_p(1) \\
&= \frac{\sqrt{n}\hat{\Sigma}_0/n_i}{\Sigma_{xx}} - \frac{\Sigma_0/n_i}{\Sigma_{xx}^2}\sqrt{n}\hat{\Sigma}_{xx} + o_p(1).
\end{aligned}$$

Note that

$$\begin{aligned}
\sqrt{n}\hat{\Sigma}_{xx} &= \sqrt{n}\left\{\frac{\sum_{i=1}^n(\bar{W}_{i\cdot} - \bar{W}_{\cdot\cdot})^2}{n-1}\right\} - \sqrt{n}\hat{\Sigma}_0/n_i \\
&= \sqrt{n}\left\{\frac{\sum_{i=1}^n(\bar{W}_{i\cdot} - \mu_x)^2}{n}\right\} - \sqrt{n}\left\{\frac{\sum_{i=1}^n\sum_{r=1}^{n_i}(W_{ir} - \bar{W}_{i\cdot})^{\otimes 2}}{n_i\sum_{i=1}^n(n_i - 1)}\right\} + o_p(1).
\end{aligned}$$

Plugging this result into (5.22), together with the asymptotic expansion of $\hat{\beta}_{nv}$ that we derived, and the fact that $\hat{\Sigma}_0/n_i$ is a sum of independent terms, we obtain that $\sqrt{n}(\hat{\beta}_{rc} - \beta_{rc}^*)$ is a sum of independent terms asymptotically, and thus $\hat{\beta}_{rc}$ is asymptotic normal with mean β_{rc}^* and variance Σ_{rc}^*, which is the expectation of the square of the independent term and is of a complicated form.

For the multivariate case, $\hat{\beta}_{rc}$ is still asymptotic normal with mean β_{rc}^* and variance of a complicated form. We suggest only use the first term of (5.22) to obtain the approximate variance of $\hat{\beta}_{rc}$: $\Sigma_{rc}^* \approx [A_{rc,1}^{-1}]^T \mathcal{D}_{nv}^{-1}\Sigma_{nv}\mathcal{D}_{nv}^{-T}[A_{rc,1}^{-1}]$, which can be consistently estimated by their empirical counterpart.

For the general case that the n_i are not necessarily equal, it can be shown that $\hat{\beta}_{rc}$ is still asymptotic normal, with a very complicated mean and variance. In this case, we suggest use the bootstrap method for variance estimate.

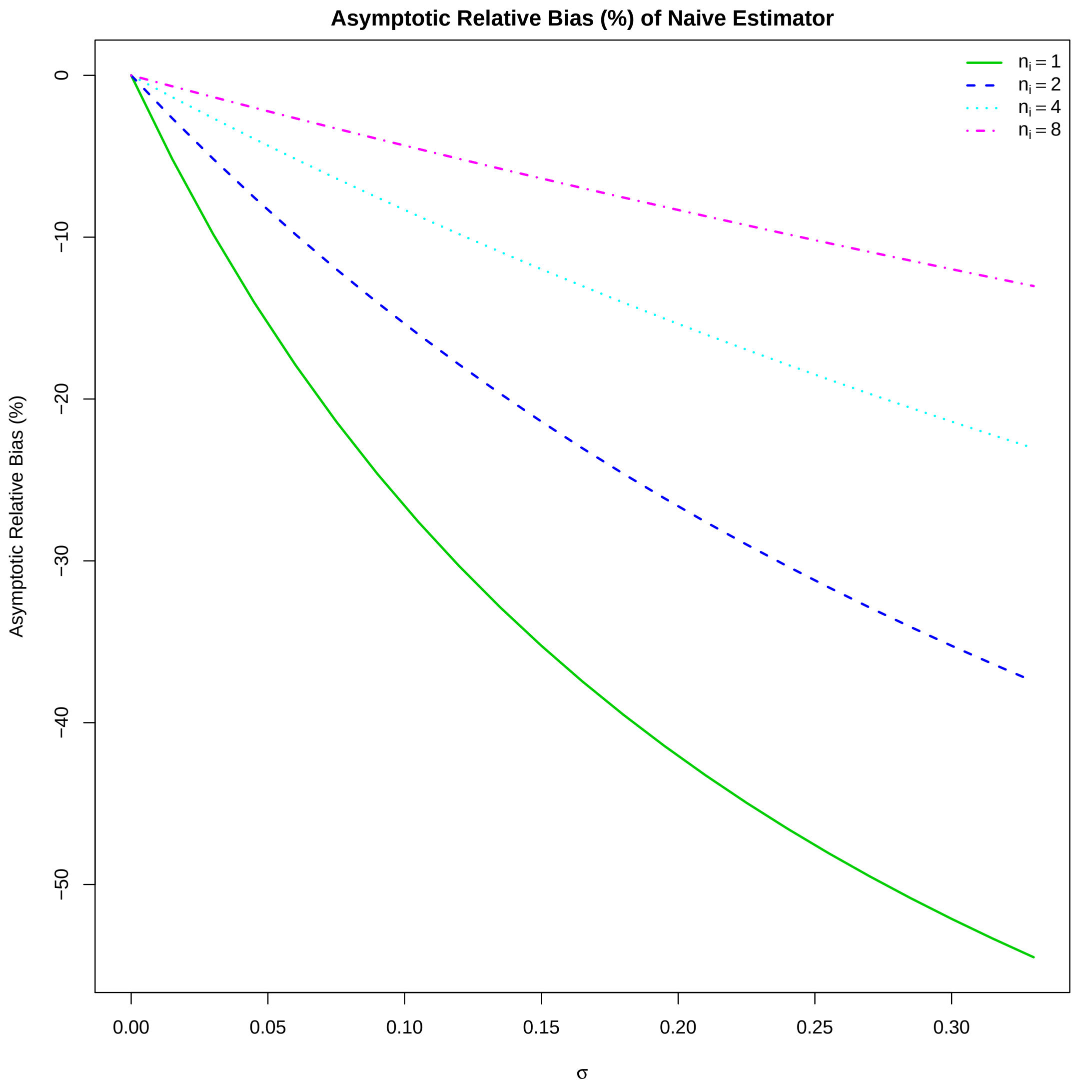


Figure 5.1: Asymptotic relative bias of naive estimator $\hat{\beta}_{nv}$ with different replicates numbers n_i

Table 5.1: Comparison of the performance of various estimators

Scenario	σ	Method	Estimating β_x					Estimating β_v				
			Bias[a]	EVE [b]	MVE[c]	MSE[d]	MCP(%)[e]	Bias	EVE	MVE	MSE	MCP(%)
Scenario 1	0.25	$\hat{\beta}_{nv}$	-0.088	0.018	0.017	0.026	85.6	-0.002	0.018	0.017	0.018	93.9
		$\hat{\beta}_{szs}$	0.023	0.026	0.024	0.027	93.1	-0.002	0.020	0.018	0.020	94.1
		$\hat{\beta}_{rc}$	-0.001	0.022	0.020	0.022	92.8	-0.002	0.018	0.017	0.018	93.9
		$\hat{\beta}_{c}$	0.019	0.025	0.023	0.026	93.5	-0.002	0.020	0.018	0.020	94.0
		$\hat{\beta}_{e}$	0.022	0.026	0.024	0.026	93.7	-0.002	0.020	0.018	0.020	94.0
	0.75	$\hat{\beta}_{nv}$	-0.498	0.009	0.008	0.257	0.2	-0.002	0.018	0.017	0.018	94.7
		$\hat{\beta}_{szs}$	0.159	0.222	0.271	0.247	95.0	-0.003	0.040	0.039	0.040	95.9
		$\hat{\beta}_{rc}$	-0.047	0.050	0.031	0.052	83.0	-0.003	0.021	0.017	0.021	92.8
		$\hat{\beta}_{c}$	0.098	0.145	0.162	0.155	95.4	-0.003	0.035	0.034	0.035	95.1
		$\hat{\beta}_{e}$	0.134	0.172	0.177	0.190	96.0	-0.003	0.037	0.036	0.037	95.1
Scenario 2	0.25	$\hat{\beta}_{nv}$	-0.056	0.014	0.013	0.017	89.7	-0.037	0.029	0.027	0.030	92.7
		$\hat{\beta}_{szs}$	0.013	0.019	0.017	0.019	93.1	0.016	0.032	0.030	0.032	93.7
		$\hat{\beta}_{rc}$	0.019	0.019	0.017	0.020	92.6	0.028	0.032	0.030	0.033	93.6
		$\hat{\beta}_{c}$	0.010	0.019	0.017	0.019	93.1	0.014	0.032	0.030	0.032	93.7
		$\hat{\beta}_{e}$	0.012	0.019	0.017	0.019	92.9	0.015	0.032	0.030	0.032	93.8
	0.75	$\hat{\beta}_{nv}$	-0.278	0.008	0.007	0.085	10.3	-0.207	0.027	0.025	0.070	68.2
		$\hat{\beta}_{szs}$	0.132	0.213	0.238	0.230	94.8	0.109	0.158	0.182	0.170	95.5
		$\hat{\beta}_{rc}$	0.103	0.098	0.060	0.108	89.3	0.125	0.120	0.063	0.135	88.6
		$\hat{\beta}_{c}$	0.077	0.114	0.132	0.120	94.5	0.066	0.099	0.108	0.104	94.4
		$\hat{\beta}_{e}$	0.102	0.146	0.142	0.157	95.3	0.087	0.119	0.131	0.127	94.6
Scenario 3	0.25	$\hat{\beta}_{nv}$	-0.003	0.027	0.024	0.027	93.3	0.002	0.042	0.040	0.042	93.8
		$\hat{\beta}_{szs}$	0.024	0.031	0.027	0.031	93.3	-0.010	0.042	0.040	0.042	93.4
		$\hat{\beta}_{rc}$	0.014	0.029	0.026	0.029	93.6	-0.009	0.042	0.040	0.042	93.3
		$\hat{\beta}_{c}$	0.024	0.031	0.027	0.031	93.3	-0.010	0.042	0.040	0.042	93.4
		$\hat{\beta}_{e}$	0.024	0.031	0.027	0.031	93.3	-0.010	0.042	0.040	0.042	93.4
	0.75	$\hat{\beta}_{nv}$	-0.144	0.015	0.015	0.036	74.4	0.062	0.041	0.040	0.045	92.9
		$\hat{\beta}_{szs}$	0.037	0.045	0.042	0.046	94.2	-0.015	0.047	0.046	0.047	94.4
		$\hat{\beta}_{rc}$	-0.031	0.026	0.025	0.027	92.4	-0.005	0.044	0.042	0.044	93.8
		$\hat{\beta}_{c}$	0.031	0.042	0.040	0.043	94.0	-0.013	0.046	0.045	0.046	94.4
		$\hat{\beta}_{e}$	0.037	0.043	0.041	0.044	94.2	-0.015	0.046	0.046	0.046	94.4

[a] Bias: finite sample biases; [b] EVE: empirical variances; [c] MVE: average of the model-based variance estimates; [d] MSE: mean square errors; [e] MCP: model-based coverage probability.

Table 5.2: Performance of estimators when $n_i = 1$ for some subjects

			Estimating β_x					Estimating β_v				
Setting	σ	Method	Bias	EVE	MVE	MSE	MCP(%)	Bias	EVE	MVE	MSE	MCP(%)
I	0.25	$\hat{\beta}_{nv}$	-0.006	0.025	0.023	0.025	92.4	0.011	0.041	0.040	0.041	94.2
		$\hat{\beta}_{szs}$	0.027	0.039	0.036	0.040	93.1	0.004	0.058	0.054	0.058	94.8
		$\hat{\beta}_{rc}$	0.015	0.027	0.025	0.028	93.7	-0.001	0.041	0.040	0.041	94.1
		$\hat{\beta}_{c}$	0.027	0.030	0.027	0.030	93.7	-0.003	0.041	0.041	0.041	93.9
		$\hat{\beta}_{e}$	0.027	0.030	0.028	0.031	93.7	-0.003	0.042	0.041	0.041	93.8
	0.75	$\hat{\beta}_{nv}$	-0.168	0.014	0.013	0.042	64.5	0.082	0.040	0.040	0.047	93.1
		$\hat{\beta}_{szs}$	0.047	0.061	0.059	0.063	94.3	-0.003	0.065	0.063	0.065	94.3
		$\hat{\beta}_{rc}$	-0.035	0.027	0.026	0.028	93.5	0.004	0.044	0.043	0.044	94.4
		$\hat{\beta}_{c}$	0.038	0.047	0.047	0.049	95.2	-0.006	0.048	0.048	0.048	94.3
		$\hat{\beta}_{e}$	0.045	0.050	0.049	0.052	95.2	-0.009	0.049	0.048	0.049	94.4
II	0.25	$\hat{\beta}_{nv}$	-0.013	0.028	0.023	0.028	90.3	0.029	0.039	0.040	0.040	94.6
		$\hat{\beta}_{szs}$	0.044	0.074	0.055	0.076	90.3	0.009	0.087	0.082	0.087	94.0
		$\hat{\beta}_{rc}$	0.013	0.031	0.026	0.031	91.8	0.013	0.039	0.041	0.039	94.9
		$\hat{\beta}_{c}$	0.027	0.034	0.028	0.035	92.1	0.012	0.040	0.041	0.040	95.0
		$\hat{\beta}_{e}$	0.027	0.034	0.028	0.035	92.0	0.012	0.040	0.041	0.040	95.0
	0.75	$\hat{\beta}_{nv}$	-0.192	0.014	0.012	0.050	55.4	0.106	0.038	0.040	0.050	91.4
		$\hat{\beta}_{szs}$	0.083	0.123	0.109	0.130	94.0	-0.008	0.109	0.104	0.109	94.8
		$\hat{\beta}_{rc}$	-0.040	0.031	0.027	0.032	90.6	0.015	0.043	0.044	0.044	94.8
		$\hat{\beta}_{c}$	0.047	0.065	0.069	0.067	93.9	0.002	0.049	0.054	0.049	95.4
		$\hat{\beta}_{e}$	0.056	0.069	0.077	0.072	94.1	-0.002	0.050	0.056	0.050	95.5
III	0.25	$\hat{\beta}_{nv}$	-0.023	0.024	0.023	0.025	92.6	0.003	0.038	0.040	0.038	95.5
		$\hat{\beta}_{szs}$	-0.168	0.030	0.027	0.058	73.7	-0.219	0.042	0.040	0.090	80.9
		$\hat{\beta}_{rc}$	0.003	0.028	0.026	0.027	93.3	-0.012	0.039	0.041	0.039	95.3
		$\hat{\beta}_{c}$	0.011	0.029	0.027	0.029	93.4	-0.012	0.039	0.041	0.039	95.3
		$\hat{\beta}_{e}$	0.012	0.029	0.027	0.029	93.4	-0.012	0.039	0.041	0.039	95.3
	0.75	$\hat{\beta}_{nv}$	-0.185	0.013	0.013	0.047	57.4	0.076	0.038	0.040	0.043	92.8
		$\hat{\beta}_{szs}$	-0.144	0.055	0.062	0.076	79.2	-0.227	0.048	0.047	0.099	82.6
		$\hat{\beta}_{rc}$	-0.032	0.031	0.028	0.032	92.3	-0.013	0.043	0.044	0.043	94.8
		$\hat{\beta}_{c}$	0.026	0.051	0.053	0.052	95.3	-0.018	0.047	0.050	0.047	95.8
		$\hat{\beta}_{e}$	0.033	0.054	0.056	0.055	95.5	-0.021	0.047	0.050	0.048	95.7
IV	0.25	$\hat{\beta}_{nv}$	-0.019	0.024	0.022	0.025	92.5	0.015	0.043	0.040	0.043	94.2
		$\hat{\beta}_{szs}$	0.109	0.114	0.089	0.126	91.3	0.265	0.253	0.223	0.323	87.9
		$\hat{\beta}_{rc}$	0.006	0.027	0.025	0.027	93.2	-0.001	0.044	0.040	0.044	94.3
		$\hat{\beta}_{c}$	0.024	0.031	0.028	0.031	94.0	-0.004	0.044	0.041	0.044	94.2
		$\hat{\beta}_{e}$	0.025	0.031	0.028	0.032	94.0	-0.004	0.044	0.041	0.044	94.2
	0.75	$\hat{\beta}_{nv}$	-0.202	0.011	0.011	0.052	49.0	0.020	0.015	0.018	0.018	90.5
		$\hat{\beta}_{szs}$	0.150	0.184	0.156	0.206	93.0	0.247	0.287	0.251	0.348	87.8
		$\hat{\beta}_{rc}$	-0.054	0.027	0.025	0.030	90.4	0.007	0.049	0.043	0.049	94.0
		$\hat{\beta}_{c}$	0.052	0.068	0.069	0.071	95.5	-0.013	0.058	0.053	0.059	94.4
		$\hat{\beta}_{e}$	0.063	0.075	0.075	0.018	95.7	-0.018	0.060	0.055	0.060	94.6
V[b]	0.25	$\hat{\beta}_{nv}$	-0.028	0.026	0.022	0.027	91.0	0.018	0.043	0.040	0.043	93.7
		$\hat{\beta}_{szs}$	-	-	-	-	-	-	-	-	-	-
		$\hat{\beta}_{rc}$	0.005	0.030	0.025	0.030	92.0	-0.002	0.044	0.041	0.044	94.0
		$\hat{\beta}_{c}$	0.022	0.035	0.028	0.035	92.1	-0.004	0.044	0.041	0.044	94.1
		$\hat{\beta}_{e}$	0.023	0.035	0.029	0.035	92.1	-0.005	0.044	0.041	0.044	94.1
	0.75	$\hat{\beta}_{nv}$	-0.230	0.012	0.010	0.065	34.5	0.107	0.042	0.040	0.054	90.6
		$\hat{\beta}_{szs}$	-	-	-	-	-	-	-	-	-	-
		$\hat{\beta}_{rc}$	-0.058	0.033	0.028	0.036	89.7	0.003	0.049	0.045	0.049	93.2
		$\hat{\beta}_{c}$	0.041	0.066	0.057	0.068	93.7	-0.010	0.057	0.054	0.057	93.9
		$\hat{\beta}_{e}$	0.052	0.067	0.061	0.070	94.0	-0.015	0.058	0.055	0.058	93.9

Table 5.3: Empirical coverage rate (in percent) of confidence bands with nominal level 0.95

σ	Methods	No censoring	30% censoring
0.25	Naive HW[a]	92.4	94.1
	Corrected HW[b]	93.5	94.0
0.75	Naive HW	67.7	91.8
	Corrected HW	95.8	94.3

[a] Naive Hall-Wellner band; [b] Corrected Hall-Wellner band.

Table 5.4: Empirical size and empirical power of the proposed test statistic

True Model	σ	Method	No Censoring	30% Censoring
Additive Hazards Model		S_{true}	0.057	0.054
	0.25	S_{nv}	0.158	0.056
		S_c	0.047	0.062
	0.75	S_{nv}	0.595	0.065
		S_c	0.047	0.052
Cox Model		S_{true}	0.920	0.791
	0.25	S_c	0.872	0.740
	0.75	S_c	0.462	0.402

Table 5.5: Analyse of the ACTG 175 dataset using different methods

Data	Method	log(CD4 counts + 1)			Treatment		
		EST[a]	MVE[b]	MCI[c]	EST[d]	MVE[e]	MCI[f]
Data Subsets with Replicates	$\hat{\beta}_{nv}$	-4.67	2.15	$(-5.58, -3.77)$	-2.12	1.18	$(-2.80, -1.45)$
	$\hat{\beta}_{szs}$	-5.76	3.36	$(-6.90, -4.63)$	-2.16	1.19	$(-2.84, -1.49)$
	$\hat{\beta}_{rc}$	-5.71	3.20	$(-6.82, -4.60)$	-2.14	1.18	$(-2.81, -1.47)$
	$\hat{\beta}_{c}$	-5.78	3.40	$(-6.93, -4.64)$	-2.16	1.19	$(-2.84, -1.49)$
	$\hat{\beta}_{e}$	-5.79	3.40	$(-6.93, -4.64)$	-2.16	1.19	$(-2.84, -1.49)$
Full Data	$\hat{\beta}_{nv}$	-4.72	2.13	$(-5.62, -3.81)$	-2.15	1.16	$(-2.81, -1.48)$
	$\hat{\beta}_{rc}$	-5.77	3.19	$(-6.88, -4.67)$	-2.16	1.16	$(-2.83, -1.50)$
	$\hat{\beta}_{c}$	-5.85	3.41	$(-7.00, -4.71)$	-2.18	1.17	$(-2.86, -1.51)$
	$\hat{\beta}_{e}$	-5.85	3.41	$(-7.00, -4.71)$	-2.18	1.17	$(-2.86, -1.51)$

[a] EST: estimates $\times 10^4$; [b] MVE: model-based variance estimates $\times 10^9$; [c] MCI: confidence intervals $\times 10^4$; [d] EST: estimates $\times 10^4$;
[e] MVE: model-based variance estimates $\times 10^9$; [f] MCI: confidence intervals $\times 10^4$.

Chapter 6

Estimation and Variable Selection for High Dimensional Additive Hazards Regression with Covariate Measurement Error

6.1 Introduction

In many clinical studies, a large number of variables are measured in addition to survival outcomes. To enhance model interpretation and identify important predictors, variable selection procedures are usually needed. Traditional selection methods, e.g., the best subset selection method, suffer from certain drawbacks, as discussed by Fan, Li and Li (2005). To overcome the drawbacks of the traditional selection methods, Tibshirani (1996) and Fan and Li (2001) proposed different types of penalized least square methods for linear models, named Lasso and SCAD, respectively. These methods were applied to the Cox model (Cox 1972) by Tibshirani (1997) and Fan and Li (2002), to the additive hazards model (Lin and Ying 1994) by Leng and Ma (2007), and to multivariate failure time data by Cai et al.

(2005). All these methods are developed for the scenario that the number of variables p is much smaller than the number of subjects n. In particular, Tibshirani (1997), Fan and Li (2002), and Leng and Ma (2007) imposed the assumption that p is fixed; Cai et al. (2005) considered the high dimensional scenario (Fan and Lv 2010) where p grows slowly with n in the sense that $p = o(n^{1/4})$.

In recent years, data collection and storage technologies develop rapidly, resulting in ultra-high dimensional (Fan and Lv 2010) data where the number of variables p can be much larger than the number of subjects n, meaning that p grows at a non-polynomial rate as n. For example, in the leukemia study reported by Golub et al. (1999), the expression levels of over 7000 genes were measured and recorded for around 70 leukemia patients. In the diffuse large-B-cell study reported by Resenwald et al. (2002), gene expressions measurements for over 7400 genes and survival information after chemotherapy were obtained for 240 patients. Detecting the link of gene expressions and the leukemia disease or survival outcome is challenging for ultra-high dimensional problems. To extract useful information from ultra-high dimensional data, statisticians have made great efforts in the past several years (e.g., Candes and Tao 2007; Bickel, Ritov and Tsybakov 2009; Bradic, Fan and Wang 2011; Buhlmann and van de Geer 2011; Fan and Lv 2011; Negahban et al. 2012; Fan, Xue and Zou 2014; Wang, Liu and Zhang 2014). When survival information is available in addition to ultra-high dimensional covariates, penalized methods have been proposed and their theoretical properties are well studied. Specifically, for high dimensional Cox models, Bradic, Fan and Jiang (2011) proposed a penalized partial likelihood method, and showed that the corresponding regression parameter estimator enjoys the strong oracle property (Bradic, Fan and Wang 2011), while Huang et al. (2013), Kong and Nan (2014), and Lemler (2012) showed oracle inequalities for the Cox regression with the Lasso penalty. For additive hazards models with high dimensional covariates, Lin and Lv (2013) showed that the penalized least square-type method with nonconcave penalties have the oracle property (Fan and Li 2001), while Gaiffas and Guilloux (2012) proposed a Lasso-based method and derived oracle inequalities.

All these methods assume that covariates are measured precisely. However, in many studies, survival analysis can be complicated by mismeasurement of covariates. For example, some clinical characteristics, e.g., blood pressure and CD4 counts, are measured with error. More challengingly, the mismeasured covariates can be high dimensional. For example, Chen, Dougherty and Bittner (1997) and Rocke and Durbin (2001) noted that the measurements of gene expression with cDNA are inaccurate, and the standard deviation of measurement error increases proportional to the expression level. High dimensionality of mismeasured covariates increases difficulty of statistical estimation and variable selection.

Liang and Li (2009) and Ma and Li (2010) studied penalized methods for parametric and semiparametric regression models with mismeasured covariates. However, their methods can not be directly extended to incorporate survival information. Furthermore, they did not consider the high dimensional or ultra-high dimensional scenario. Recently, some progresses have been made for ultra-high dimensional linear regression models and generalized linear models (McCullagh and Nelder 1989). In particular, Rosenbaum and T-sybarkov (2010, 2013) proposed the matrix uncertainty selectors and showed that they can be used to consistently estimate the regression parameters, and have the sparsity pattern recovery property; Loh and Wainwright (2012, 2013) developed regularized estimators and showed their statistical consistency and rate of convergence; Sorensen, Frigessi and Thoresen (2014) proposed the corrected Lasso method to adjust for measurement error. However, it is not clear how to extend these methods to incorporate survival information. Methods on estimation and variable selection with high dimensional or ultra-high dimensional mismeasured covariates are limited in survival analysis.

In this chapter, we propose corrected penalized methods to do variable selection and estimation for additive hazards models with mismeasured covariates for the high dimensional scenario where p grows slowly with n in the sense that $p = o(n^{1/4})$. The penalized methods are further extended to the ultra-high dimensional scenarios with $p \gg n$ that p grows exponentially with n. Specifically, In Section 6.2, we introduce notation and model setup. In Section 6.3, we study the impact of measurement error on estimation and variable selection, and propose the corrected penalized methods for the high dimensional scenario. In Section 6.4, we extend the proposed methods to the ultra-high dimensional scenario. In

Section 6.5, we conduct simulation studies and real data analysis. Concluding remarks are deferred to the last section.

6.2 Notation and Model Setup

For $i = 1, \cdots, n$, let T_i be the failure time, C_i be the censoring time, and $Z_i(t) = (X_i^T, V_i^T(t))^T$ be a vector of covariates, where X_i is a $p_1 \times 1$ vector of time-independent but error-prone covariates, and $V_i(t)$ is a $p_2 \times 1$ vector of covariates that are precisely measured and possibly time-dependent. As common in practise, $V_i(t)$ are assumed to be external covariates (Kalbfleisch and Prentice 2002, p.197). Let $p = p_1 + p_2$ be the number of parameters. Note that both of p_1 and p_2 can be high-dimensional.

We consider that the hazard function of T_i is related to $Z_i(\cdot)$ through the additive hazards model

$$\lambda(t; Z_i(t)) = \lambda_0(t) + \beta^T Z_i(t) = \lambda_0(t) + \beta_x^T X_i + \beta_v^T V_i(t),$$

where $\lambda_0(\cdot)$ is an unspecified baseline hazard function, and $\beta = (\beta_x^T, \beta_v^T)^T$ is a vector of unknown regression parameters. T_i and C_i are assumed to be conditionally independent given $Z_i(t)$. Suppose individuals are observed over a common time interval $[0, \tau]$, where τ is a positive constant. Let $S_i = \min(T_i, C_i, \tau)$, $\delta_i = I(T_i \leq \min\{C_i, \tau\})$, $N_i(t) = I(S_i \leq t, \delta_i = 1)$, and $Y_i(t) = I(S_i \geq t)$.

6.2.1 Penalized Methods

The pseudo score functions of Lin and Ying (1994) are defined as

$$U(\beta) = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} \{Z_i(t) - \bar{Z}(t)\}\{dN_i(t) - Y_i(t) Z_i^T(t) \beta dt\},$$

where $\bar{Z}(t) = \sum_{i=1}^n Y_i(t)Z_i(t)/\sum_{i=1}^n Y_i(t)$. Integrating $-U(\beta)$ with respective to β gives the loss function

$$L(\beta) = \frac{1}{2}\beta^T V\beta - b^T\beta,$$

where

$$\begin{aligned} V &= \frac{1}{n}\sum_{i=1}^n \int_0^\tau Y_i(t)\{Z_i(t) - \bar{Z}(t)\}^{\otimes 2} dt, \\ \text{and } \; b &= \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{Z_i(t) - \bar{Z}(t)\right\} dN_i(t). \end{aligned}$$

Leng and Ma (2007), Martinussen and Scheike (2009), and Lin and Lv (2013) proposed to estimate the regression parameter β via the penalized partial least square method:

$$\hat{\beta} \in \arg\min_{\beta\in\mathbb{R}^p} \left\{L(\beta) + \mathcal{R}_\lambda(\beta)\right\}, \tag{6.1}$$

where $\mathcal{R}_\lambda(\cdot)$ is a penalty function that depends on a tuning parameter $\lambda \geq 0$. This estimator does the variable selection automatically in the sense that $\hat{\beta}_j = 0$ for some j's. The estimation and variable selection properties of $\hat{\beta}$ with different penalties were studied by Leng and Ma (2007), Martinussen and Scheike (2009), and Lin and Lv (2013). They showed that different choices of the penalty function $\rho_\lambda(\cdot)$ lead to different estimation and variable selection results.

6.2.2 Penalties

In this chapter, we consider the following penalties. For convenience, we write

$$\mathcal{R}_\lambda(\beta) = \lambda\sum_{j=1}^p \rho_\lambda(|\beta_j|).$$

When $\lambda = 0$, then $\hat{\beta}$ reduces to the pseudo score estimator by Lin and Ying (1994). We write $\rho_\lambda(\cdot)$ as $\rho(\cdot)$ when it does not depend on λ. We consider the following penalties:

- Lasso (Tibshirani 1996) penalty: $\rho(t) = t, \quad t > 0$.
- SCAD (Fan and Li 2001) penalty:

$$\rho'_\lambda(t) = I(t \leq \lambda) + \frac{(a\lambda - t)_+}{(a-1)\lambda} I(t \geq \lambda), \quad t > 0,$$

  where $a > 2$ is a fixed parameter.
- MCP (Zhang 2010) penalty:

$$\rho'_\lambda(t) = \frac{(a\lambda - t)_+}{a\lambda}, \quad t > 0,$$

  where $a > 1$ is a fixed parameter.
- SICA (Lv and Fan 2009) penalty:

$$\rho(t) = \frac{(a+1)t}{a+t}, \quad t > 0,$$

  where $a > 0$ is a fixed parameter.

Note that the Lasso penalty is convex, whereas the SCAD, MCP, and SICA penalties are nonconvex.

6.2.3 Measurement Error Model

Suppose X_i is not available, and we observe its surrogate W_i instead, $i = 1, \cdots, n$. Suppose X_i and W_i are linked through the classical error model (Carroll et al. 2006)

$$W_i = X_i + \epsilon_i,$$

where the ϵ_i are independent and identically distributed (i.i.d.) with mean 0 and a positive-definite covariance matrix Σ_0. Σ_0 is assumed known. Assume that ϵ_i is independent of $\{T_i, C_i, Z_i(t)\}$. Let $\hat{Z}_i(t) = (W_i^T, V_i^T(t))^T$, and $\Sigma_1 = \text{diag}(\Sigma_0, 0_{p_2})$, where 0_{p_2} is the $p_2 \times p_2$ matrix of elements 0.

6.3 Corrected Penalized Methods for High Dimensional Scenario

In this section, we consider the high dimensional scenario (Fan and Lv 2010) where p grows slowly with n in the sense that $p = o(n^{1/4})$. In the presence of covariate error, the penalized estimator $\hat{\beta}$ defined in (6.1) is not available since X_i may not be observed. An intuitively appealing option is to directly replace $Z_i(t)$ with $\hat{Z}_i(t) = (\bar{W}_{i\cdot}^T, V_i^T(t))^T$ in $U(\beta)$ and obtain a naive pseudo score function

$$U_{nv}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{\hat{Z}_i(t) - \tilde{Z}(t)\}\{dN_i(t) - Y_i(t)\hat{Z}_i^T(t)\beta dt\},$$

where $\tilde{Z}(t) = \sum_{i=1}^{n} Y_i(t)\hat{Z}_i(t)/\sum_{i=1}^{n} Y_i(t)$. Then integrating $-U_{nv}(\beta)$ with respective to β gives the naive loss function

$$L_{nv}(\beta) = \frac{1}{2}\beta^T V_{nv}\beta - b_{nv}^T\beta,$$

where

$$\begin{aligned} V_{nv} &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\{\hat{Z}_i(t) - \tilde{Z}(t)\}^{\otimes 2}dt, \\ \text{and} \quad b &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left\{\hat{Z}_i(t) - \tilde{Z}(t)\right\} dN_i(t). \end{aligned}$$

The naive penalized estimator is defined as

$$\hat{\beta}_{nv} \in \arg\min_{\beta\in\mathbb{R}^p}\left\{L_{nv}(\beta) + \mathcal{R}_{\lambda}(\beta)\right\}.$$

We write λ as λ_n to emphasize its dependence on the sample size n.

Let $B_1 = \int_0^{\tau} Y_i(t)\{Z_i(t) - e(t)\}^{\otimes 2}dt$, and $B_2 = \Sigma_1\int_0^{\tau} E\{Y_i(t)\}dt$, where $e(t) = E\{Y_i(t)Z_i(t)\}/E\{Y_i(t)\}$. In the following lemma, we show that $\hat{\beta}_{nv}$ is not a consistent estimator of the parameter β_0 for all four penalties we discussed before. The proof is sketched in Appendix A1.

Lemma 1 *Assume $\lambda_n \to 0$ as $n \to \infty$. Then we have*

$$\hat{\beta}_{nv} \xrightarrow{p} (B_1 + B_2)^{-1} B_1 \beta_0, \quad as \quad n \to \infty.$$

Lemma 1 implies that it is infeasible to estimate the regression parameter by using the naive penalized method. To adjust for the bias of the naive penalized method, we propose the corrected penalized method.

Let $V_{c1} = n^{-1}(\sum_{i=1}^{n} S_i - \tau)\Sigma_1$, The corrected pseudo score functions of Yan and Yi (2014b) are defined as

$$U_c(\beta) = U_{nv}(\beta) + V_{c1}\beta.$$

We integrate $-U_c(\beta)$ with respective to β, and obtain the corrected loss function

$$L_c(\beta) = \frac{1}{2}\beta^T V_c \beta - b_c^T \beta,$$

where $V_c = V_{nv} - V_{c1}$, and $b_c = b_{nv}$.

We define the corrected penalized estimator as

$$\hat{\beta}_c \in \arg\min_{\beta \in \mathbb{R}^p} \left\{ L_c(\beta) + \mathcal{R}_\lambda(\beta) \right\}. \tag{6.2}$$

When $\lambda = 0$, then $\hat{\beta}_c$ reduces to the corrected pseudo score estimator by Yan and Yi (2014b). Now we establish the oracle property (Fan and Li 2001) of the proposed corrected penalized estimator $\hat{\beta}_c$.

Write $\beta_0 = (\beta_{0,1}, \cdots, \beta_{0,p})^T$. Let $A = \{j : \beta_{0,j} \neq 0\}$ be the index set which contains all nonzero components of β_0. Let s be the size of A. Note that s can depend on the sample size n. Let β_0^A be the subvector of β_0 which contains all nonzero components. Write $\beta_0^A = (\beta_{0,1}^A, \cdots, \beta_{0,s}^A)^T$. Let $\beta_0^{A^c}$ be the complement of β_0^A,

$$\begin{aligned}
a_n &= \max\{\lambda_n \rho'_{\lambda_n}(|\beta_{0,1}^A|), \cdots, \lambda_n \rho'_{\lambda_n}(|\beta_{0,s}^A|)\}, \\
b_n &= \max\{\lambda_n \rho''_{\lambda_n}(|\beta_{0,1}^A|), \cdots, \lambda_n \rho''_{\lambda_n}(|\beta_{0,s}^A|)\}, \\
d &= \{\lambda_n \rho'_{\lambda_n}(|\beta_{0,1}^A|)\mathrm{sgn}(\beta_{0,1}^A), \cdots, \lambda_n \rho'_{\lambda_n}(|\beta_{0,s}^A|)\mathrm{sgn}(\beta_{0,s}^A)\}^T, \\
\text{and } \Sigma_{\lambda_n} &= \mathrm{diag}\{\lambda_n \rho''_{\lambda_n}(|\beta_{0,1}^A|), \cdots, \lambda_n \rho''_{\lambda_n}(|\beta_{0,s}^A|)\},
\end{aligned}$$

where sgn$(\cdot)$ is the sign function such that sgn$(x) = 1$ if $x > 0$, sgn$(x) = 0$ if $x = 0$, and sgn$(x) = -1$ if $x < 0$.

The following theorem shows that there exists a local minimizer $\hat{\beta}_c$ which satisfies (6.2) so that its rate of convergence to β_0 is $O_p(\sqrt{p}(n^{-1/2} + a_n))$. The proof is sketched in Appendix A2.

Theorem 1 *If $a_n \to 0$, $b_n \to 0$, as $n \to \infty$, and $p = o(n^{1/4})$, then with probability approaching 1, there exists a local minimizer $\hat{\beta}_c$ in (6.2) such that*

$$\|\hat{\beta}_c - \beta_0\|_2 = O_p(\sqrt{p}(n^{-1/2} + a_n)).$$

In Theorem 1, p can be growing as n increases at a slow rate. Furthermore, if $a_n = O_p(n^{-1/2})$, and p is fixed, then $\|\hat{\beta}_c - \beta_0\|_2 = O_p(n^{-1/2})$, suggesting that $\hat{\beta}_c$ is a $\sqrt{n}$-consistent estimator of β.

Let $\hat{\beta}_c^A$ be the subvector of $\hat{\beta}_c$ which corresponds to β_0^A, and $\hat{\beta}_c^{A^c}$ be the complement of $\hat{\beta}_c^A$. Define

$$\begin{aligned}
\Sigma_c =& E\left[\int_0^\tau \left\{\hat{Z}_i(t) - e(t)\right\} dN_i(t) - \int_0^\tau Y_i(t)\left\{\hat{Z}_i(t) - e(t)\right\}^{\otimes 2} \beta_0 dt + \min(S_i, \tau)\Sigma_1\beta_0\right]^{\otimes 2};\\
\mathcal{D}_c =& E\left[\int_0^\tau Y_i(t)\left\{\hat{Z}_i(t) - e(t)\right\}^{\otimes 2} dt - \min(S_i, \tau)\Sigma_1\beta_0\right].
\end{aligned}$$

Let Σ_c^{AA} be the matrix formed by the components $\Sigma_{c,ij}$ of Σ_c where $i, j \in A$, and $\mathcal{D}_c^{AA}$ be defined similarly. The next theorem establishes the oracle property of $\hat{\beta}_c$. The proof is sketched in Appendix A3.

Theorem 2 *If $b_n \to 0$, $\lambda_n \to 0$, $\lambda_n\sqrt{n/p} \to \infty$ as $n \to \infty$, $a_n = O_p(n^{-1/2})$, and $p = o(n^{1/4})$, then with probability approaching 1, for any $p \times 1$ unit vector c_n, the local minimizer $\hat{\beta}_c$ in Theorem 1 has the following properties:*

(1). Sparsity: $\hat{\beta}_c^{A^c} = 0$;

(2). Asymptotic normality:

$$\sqrt{n} c_n^T (\Sigma_c^{AA})^{-1/2} (\mathcal{D}_c^{AA} + \Sigma_\lambda)(\hat{\beta}_c - \beta_0 - (\mathcal{D}_c^{AA} + \Sigma_\lambda)^{-1} d) \xrightarrow{d} N(0,1), \quad as \;\; n \to \infty.$$

6.4 Corrected Penalized Methods for Ultra-high Dimensional Scenario

In this section, we consider the ultra-high dimensional setting $n \ll p$ where p grows exponentially with n. More specifically, we consider the setting: $\frac{s \log p}{n} = O(1)$, where s is the number of non-zero parameters defined in Section 6.3. We define the regularized M-estimator of the regression parameter β as

$$\hat{\beta}_c \in \arg \min_{\|\beta\|_1 \le R} \left\{ L_c(\beta) + \mathcal{R}_\lambda(\beta) \right\}, \tag{6.3}$$

where R is a positive number.

It is important to note that $L_c(\beta)$ can be a nonconvex function. Indeed, V_{nv} has rank at most n, and V_{c1} has rank p_1. Therefore, the difference V_c can have a large number of negative eigenvalues and thus is nonconvex. Thus, unlike regularized convex optimization problems, which usually optimize the loss function plus the penalty over the entire Euclidean space, in our setting we impose a side constraint $\|\beta\|_1 \le R$ (or equivalently, $\beta \in \mathbb{B}_1(R)$, where $\mathbb{B}_1(R)$ is the l_1 ball in the p-dimensional space with center at the origin and radius R) to guarantee that there exists a global optimum (Loh and Wainwright 2013).

We first show that the defined loss function $L_c(\cdot)$ satisfies the restricted strong convexity (RSC) condition (Loh and Wainwright 2012, 2013; Negahban et al. 2012):

$$\langle \nabla L_c(\beta_0 + \triangle) - \nabla L_c(\beta_0), \triangle \rangle \ge \alpha \|\triangle\|_2^2 - \tau \frac{\log p}{n} \|\triangle\|_1^2, \quad \text{for all} \quad \|\triangle\|_2 \in \mathbb{R}^p,$$

where β_0 is the true value of the regression parameter, α is a positive universal constant, and τ is a nonnegative universal constant. Notation $\langle a, b \rangle$ represents the inner product of vectors a and b.

The RSC condition is a key property that guarantees statistical consistency and computational efficiency of regularized optimization problems for various statistical models. The RSC property has been shown to be satisfied for many model structures, including linear models, generalized linear models, and linear models with additive measurement error or missing data (Loh and Wainwright 2012, 2013).

However, verifying the RSC condition for our case is nontrivial since the additive hazards model is a nonlinear survival model that consists of the nonparametric baseline hazard function. Furthermore, verifying the RSC condition requires concentration results of certain random vectors and matrices that involves event and covariates information, which are nontrivial even in the case of no measurement error (Bradic, Fan and Jiang 2011; Huang et al. 2013; Lin and Lv 2013). We defer the proof of the RSC property of $L_c(\cdot)$ to the Appendix.

Due to the nonconvexity of the regularized optimization problem (6.3), the global minimum $\hat{\beta}_c$ is computationally intractable. Instead, we consider the statistical property of any local minima $\tilde{\beta}_c$ that satisfies the first-order optimality condition (Bertsekas 1999; Loh and Wainwright 2013):

$$\langle \nabla L_c(\tilde{\beta}_c) + \nabla \mathcal{R}_\lambda(\tilde{\beta}_c), \beta - \tilde{\beta}_c \rangle \geq 0, \quad \text{for all} \quad \beta \in \mathbb{B}_1(R).$$

Assume that $n \geq C_0 \max\{R^2, s\} \log p$ for some constant C_0. The following theorem provides the non-asymptotic bounds of the l_1 and l_2 error of $\hat{\beta}_c$. The proof is sketched in Appendix A4.

Theorem 3 *Under Regularity Conditions, with probability at least* $1 - C_1 \exp(-C_2 \log p)$, *and* λ *satisfies that*

$$C_3 \sqrt{\frac{\log p}{n}} \leq \lambda \leq \frac{C_4}{R},$$

where C_1, C_2, C_3, C_4 are constants, we have

$$\begin{aligned} \|\tilde{\beta}_c - \beta_0\|_2 &\leq \frac{K_1 \lambda \sqrt{s}}{\lambda_{\min}(\mathcal{V}_c) - 2\mu}, \\ \textit{and} \;\; \|\tilde{\beta}_c - \beta_0\|_1 &\leq \frac{K_2 \lambda s}{\lambda_{\min}(\mathcal{V}_c) - 2\mu}, \end{aligned}$$

where $\lambda_{\min}(\mathcal{V}_c)$ is defined in the Appendix, and μ is a positive number that depends on the penalty (Loh and Wainwright 2013).

6.5 Numerical Studies

The coordinate decent algorithm has been shown suitable for high dimensional and ultra-high dimensional data analysis (Friedman et al. 2007; Friedman, Hastie and Tibshirani 2010; Breheny and Huang 2011; Fan and Lv 2011). We adapt this algorithm to produce the solution path (Friedman et al. 2007) of the penalized estimators, and use ten-fold cross-validation to select the optimal λ and the corresponding estimators.

6.5.1 Simulation Studies

We conduct simulation studies to evaluate the finite sample performance of the proposed methods. We generate 100 simulations for each parameter configuration. We consider two scenarios. In Scenario 1, $n = 200$, and $p = 50$, representing the high dimensional case; in Scenario 2, $n = 400$, and $p = 800$, representing the ultra-high dimensional case that p is much larger than n.

The Z_i are simulated from the multivariate normal distribution: $Z_i \sim MVN(0_p, \Sigma_z)$, $i = 1, \cdots, n$. Here, the Σ_z is the Toeplitz matrix with the (j, k)th component of Σ_z given by $\rho^{|j-k|}$, $j, k = 1, \cdots, p$. We set $\rho = 0.5$. Survival times T_i are independently generated using the additive hazards model, where the first 15 components of true parameter β_0 is

$(v^T, v^T, v^T)^T$ with $v = (1, 0, -1, 0, 0)$, while the rest components of β_0 are all set as zero. Thus, $s = 6$. The baseline hazard function $\lambda_0(t) = 1$. We impose the constraint that the hazard function needs to be positive. The censoring time C_i is generated from UNIF$[0, c]$, where c is chosen to obtain about 25% censoring.

Suppose we do not observe the component X_i for all subjects. Instead, we observe the surrogate version W_i. In Scenario 1, the dimension of X_i is set to be $p_1 = 50$, while in scenario 2, $p_1 = 200$. We consider the classical additive error model for the measurement error process: $W_i = X_i + \epsilon_i$, where $\epsilon_i \sim MVN(0, \sigma_0^2 I_{p_1 \times p_1})$ with a given σ_0. Take σ_0 to be 0.1 or 0.2 to represent different degrees of measurement error.

To estimate the β parameter, we consider three methods. The first method is the proposed corrected penalized estimator $\hat{\beta}_c$ in (6.2). For comparison, we also consider two other methods. Specifically, the second method is the penalized estimator $\hat{\beta}$ by Lin and Lv (2013) based on the true covariates, and the third method is the oracle estimator based on the corrected pseudo score method by Yan and Yi (2014b) that knows the sparsity set A^c in advance.

We use four performance measures to compare these three estimators. l_2 ERROR is the l_2 estimation error $\|\tilde{\beta} - \beta_0\|_2$, where $\tilde{\beta}$ stands for the three estimators; l_1 ERROR is the l_1 estimation error $\|\tilde{\beta} - \beta_0\|_1$; #S is the total number of selected variables; #FN is the number of falsely excluded variables. We report the means and standard deviations of these measures for all three methods in Tables 6.1 and 6.2.

It is seen that the proposed corrected penalized methods performs satisfactorily compared to the penalized methods based on the true covariates. The penalized methods with nonconvex penalties outperform the corrected Lasso method. Mismeasurement of covariates tends to reduce the precision of estimation results.

[Insert Tables 6.1 and 6.2 here!]

6.5.2 Real Data Analysis

We conduct data analysis of the AIDS Clinical Trials Group (ACTG) 175 (Hammer, et al. 1996) study. The ACTG 175 study is a double-blind randomized clinical trial that evaluated the effects of the HIV treatments for which three drugs were used in combination or alone: zidovudine, didanosine, and zalcitabine.

In the ACTG 175 dataset described in the R package 'speff2trial' (Juraska 2010), there were $n = 2139$ individuals, with 26 observed variables for each individual. The names of this variables are age, wtkg, hemo, homo, drugs, karnof, oprior, z30, zprior, preanti, race, gender, str2, strat, symptom, treat, offtrt, cd40, cd420, cd496, r, cd80, cd820, cens, days, and arms. We refer the detailed description of these variables to Juraska (2010).

True values of the CD4 counts were not available due to biological variability. Instead, the baseline measurements on the CD4 counts were collected before randomization, ranging from 200 to 500 per cubic millimeter. The variable cd40 represents the average baseline measurement. The variable arms is the treatment arm indicator, where arms=0 for the zidovudine treatment, 1 for the zidovudine and didanosine treatment, 2 for the zidovudine and zalcitabine treatment, and 3 for the didanosine treatment, respectively. The failure time T_i is defined to be the time to the occurrence of the first event among the following events: (i) more than 50% decline of CD4 counts compared to the averaged baseline CD4 counts cd40; (ii) disease progression to AIDS; or (iii) death. About 75.6% of outcome values are censored. The variable days represents the censored failure time, named S_i.

Log transformation of the variables age, wtkg, karnof, preanti, cd40, cd420, cd80, cd820, and days were employed to normalize the data. We removed four variables zprior, treat, cd496, and r from our data analysis for the following reasons: zprior is the constant 1 for all subjects; the variable treat indicates whether or not the subject received the zidovudine treatment, which is overlapped with the the treatment indicator arms; the variable cd496 is missing for a large amount of subjects, and r is its missing indicator. As a result, we included p=20 covariates in the analysis.

Let X_i be the log transformation version of the true baseline CD4 counts, and W_i indicate the log transformation version of the surrogates cd40. We adopt the classical error model:

$$W_i = X_i + \epsilon_i,$$

where the variance of ϵ_i is estimated by the replicated baseline measurements.

Let Z_i represent the vector including X_i and 19 other precisely measured variables, and $\hat{Z}_i$ be its observed version, $i = 1, \cdots, n$. In this example, the dimension of error-prone covariates $p_1 = 1$, whereas the dimension of precisely measured covariates $p_2 = 19$. We employ the additive hazards model to feature the dependence of T_i on Z_i:

$$\lambda(t; Z_i) = \lambda_0(t) + Z_i^T \beta,$$

where estimation and variable selection of the parameter β are of interest.

We apply the proposed corrected penalized method with four different penalties. For comparison purposes, we also include three other methods: (1). the corrected pseudo score estimator by Yan and Yi (2014b) (without penalties) which adjusts for measurement error; (2). the naive pseudo score estimator by Lin and Ying (1994) (without penalties) which ignores measurement error; (3). the naive penalized method by Lin and Lv (2013) which ignores measurement error. The results are summarized in Table 6.3.

The results show that the proposed method with the Lasso penalty shrinks 6 out of 20 parameters to zero, the SCAD or MCP penalty sets 7 parameters to be zero, and the SICA penalty sets 9 parameters to be zero. In contrast, all the parameters estimated by the method by Yan and Yi (2014b) are nonzero. The naive methods of Lin and Ying (1994) and Lin and Lv (2013) result in inconsistent estimates of β. The naive method of Lin and Lv (2013) with SCAD or MCP penalty selects 1 additional variables compared to the proposed penalized method.

[Insert Table 6.3 here!]

6.6 Discussion

In this chapter, we propose corrected penalized methods for variable selection and estimation in the presence of covariate measurement error for survival data. Furthermore, we extend these methods to the ultra-high dimensional setting. The theoretical properties of these penalized estimators are studied for the cases where p grows slowly with n, or $p \gg n$. For the former case, we prove the oracle property of the proposed corrected penalized estimators with nonconvex penalties; for the latter case, we provide a sharp upper bound of estimation error of the proposed estimators. The finite sample performance of these estimators is evaluated by numerical studies.

Appendix

Let $Q_c(\beta) = L_c(\beta)+\mathcal{R}_\lambda(\beta)$. Let $S^{(k)}(t) = n^{-1}\sum_{i=1}^n Y_i(t)\hat{Z}_i^k(t)$ and $s^{(k)}(t) = E\{Y_i(t)\hat{Z}_i^k(t)\}, k = 0, 1, 2$. Let Ω_z be the event that $\max_{j=1}^p \sup_{t\in[0,\tau]} |\hat{Z}_j(t)| \le z$ for a fix $z > 0$. Let A_{jk} denote the element in the jth row and kth column of the matrix A. Let B_j denote the jth element of the vector B. To avoid confusion, we remove the dependence of the random terms on the subject index i. For example, we let $N(t)$ to denote $N_i(t)$. Let $\mathbb{P}_n$ denote the empirical measure, and P denote the probability measure. For example, $\mathbb{P}_n N(t) = n^{-1}\sum N(t)$, and $P\{N(t)\} = E\{N(t)\}$. Let $\tilde{M}(t) = N(t) - \int_0^t Y(u)\{d\Lambda_0(u) + \beta_0^T \hat{Z}(u)du\}$. Let $\mathcal{V}_c = E\int_0^\tau Y(t)\{Z(t) - e(t)\}^{\otimes 2}dt$. Let $\lambda_{\min}(A)$ denote the minimum eigenvalue of the square matrix A.

Appendix A1

Proof of Lemma 1: When $\lambda_n \to 0$, $\mathcal{R}_{\lambda_n}(\beta) \to 0$, as $n \to \infty$. The result of Lemma 1 then holds by Yan and Yi (2014b).

Appendix A2

Proof of Theorem 1: Let $\alpha_n = \sqrt{p}(n^{-1/2} + a_n)$. It suffices to show that for any given constant $\epsilon > 0$, there exists a sufficient large constant C, such that

$$P(\min_{\|u\|_2 = C} Q_c(\beta + \alpha_n u) > Q_c(\beta)) \geq 1 - \epsilon.$$

This inequality implies that there exists a local minimizer $\hat{\beta}_c$, such that $\|\hat{\beta}_c - \beta_0\|_2 = O_p(\alpha_n)$. The inequality can be proved by following the argument of Theorem 1 by Fan and Li (2001).

Appendix A3

Proof of Theorem 2: First of all, following Lemma A1 of Cai et al. (2005), we show that with probability approaching 1, for any β^A such that $\|\beta^A - \beta_0^A\|_2 = O_p(\sqrt{p/n})$ and any constant C, the following equality holds:

$$Q_c(\beta^A, 0) = \min_{\|\beta^{A_c}\|_2 \leq C\sqrt{p/n}} Q_c(\beta^A, \beta^{A_c}).$$

Therefore, sparsity of $\hat{\beta}_c$ follows. It remains to prove the asymptotic normality of $\hat{\beta}_c$, which follows by the Slusky's Theorem and the Central Limit Theorem.

Appendix A4

Proof of Theorem 3: The following lemmas and theorem are used to prove the upper bound in Theorem 3.

Lemma A1 *Under Regularity Conditions, there exists universal constants* $C, K > 0$, *such that*

$$\begin{aligned}
\Pr\left(\sup_{t\in[0,\tau]}|S^{(0)}(t)-s^{(0)}(t)|\geq Kn^{-1/2}(1+x)\right) &\leq \exp(-Cx^2),\\
\Pr\left(\sup_{t\in[0,\tau]}|S_j^{(1)}(t)-s_j^{(1)}(t)|\geq Kn^{-1/2}(1+x)\middle|\,\Omega_z\right) &\leq \exp(-Cx^2/z^2),\\
\textit{and}\quad \Pr\left(\sup_{t\in[0,\tau]}|S_{ij}^{(1)}(t)-s_{ij}^{(1)}(t)|\geq Kn^{-1/2}(1+x)\middle|\,\Omega_z\right) &\leq \exp(-Cx^2/z^4),
\end{aligned}$$

for all $x>0$, $i,j=1,\cdots,p$.

Proof. The first concentration inequality is (A.5) of Lemma A.2 Lin and Lv (2013). Now we prove the last two inequalities. Since $Z_j(\cdot), j=1,\cdots,p$ are of uniformly bounded variation, we have $\hat{Z}_j(\cdot), j=1,\cdots,p$ are also of uniformly bounded variation. Therefore, following the proof of Lemma A.2 Lin and Lv (2013), we obtain (6.4) and (6.4).

Lemma A2 *Under Regularity Conditions, there exists universal constants $C_1, C_2, K>0$, such that*

$$\Pr\left(|U_{c,j}(\beta_0)|\geq Kn^{-1/2}(1+x)\middle|\,\Omega_z\right)\leq C_1\exp\left(-C_2\frac{x^2\wedge n}{z^4}\right).$$

for all $x>0$, $j=1,\cdots,p$.

Proof. Note that

$$\begin{aligned}
U_{c,j}(\beta_0) &= \mathbb{P}_n\int_0^\tau \hat{Z}_j(t)d\tilde{M}(t)-\mathbb{P}_n\int_0^\tau \tilde{Z}_j(t)d\tilde{M}(t)+\mathbb{P}_n\left(S-\frac{\tau}{n}\right)(\Sigma_1\beta_0)_j\\
&\equiv T_1-T_2+T_3,
\end{aligned}$$

where $(\Sigma_1\beta_0)_j$ denotes the jth element of $\Sigma_1\beta_0$. Note that $\tilde{M}_i(t)$ is of bounded variation. Thus, $\int_0^\tau|d\tilde{M}_i(t)|\equiv M_0(z)\leq zM_1<\infty$. Thus, $|\int_0^\tau\hat{Z}_j(t)d\tilde{M}(t)|\leq\sup_{t\in[0,\tau]}|\hat{Z}_j(t)|\int_0^\tau|d\tilde{M}_i(t)|\leq z^2M_1$ conditional on the event Ω_z, where M_0 is depends linearly on z, and M_0 is a constant. Furthermore, T_1 is sum of i.i.d mean zero terms. Thus, by Hoeffding's Inequality (e.g., Buhlmann, P. and van de Geer 2011, Lem 14.11), we have $\Pr(|T_1| \geq n^{-1/2}x|\Omega_z) \leq \exp(-D_1x^2/z^4)$, where $D_1 > 0$ is a constant.

Note that T_2 and T_3 are sum of i.i.d. terms, but their mean are not zero. Indeed, the mean of $T_2 - T_3$ is zero. In the following, we bound the term $T_2 - T_3$. First, by Lemma A1, there exists some constant $\delta > 0$, the probability of the events $\sup_{t\in[0,\tau]} |S^{(0)}(t) - s^{(0)}(t)| \geq \delta$, or $\sup_{t\in[0,\tau]} |S_j^{(1)}(t) - s_j^{(1)}(t)| \geq \delta$ is bounded by $\exp(-D_2n/z^4)$ for some constant $D_2 > 0$. Thus, we only need to bound $T_2 - T_3$ conditional on the event Ω_0: $\sup_{t\in[0,\tau]} |S^{(0)}(t)-s^{(0)}(t)| \leq \delta$, and $\sup_{t\in[0,\tau]} |S_j^{(1)}(t)-s_j^{(1)}(t)| \leq \delta$. Let $\mathcal{G}_j = \{\int_0^\tau f(t)d\tilde{M}(t) + (S - \tau/n)(\Sigma_1\beta_0)_j : f$ is of bounded variation; $\sup_{t\in[0,\tau]} |f(t) - e_j(t)|$ is upper bounded$\}$. Let $G_j = \sup_{g\in\mathcal{G}_j} |(\mathbb{P}_n - P)g|$. Proceed as the arguments of Lin and Lv (2013, Lemma A.3), and observe that $(S - \tau/n)(\Sigma_1\beta_0)_j$ does not depend on t, we obtain that $PG_j \leq D_3n^{-1/2}$ for some constants $D_3 > 0$. Thus, by Massart's Inequality (Massart 2000; Buhlmann, P. and van de Geer 2011, Thm 14.2), we have $\Pr(|T_2 - T_3| \geq D_4n^{-1/2}(1 + x)|\Omega_z) \leq 2\exp(-D_5x^2/z^4)$, where $D_4, D_5 > 0$ are some constants. Thus, the result follows.

Lemma A3 *Under Regularity Conditions, there exists universal constants $C_1, C_2, K > 0$, such that*

$$\Pr\left(|V_{c,ij} - \mathcal{V}_{c,ij}| \geq Kn^{-1/2}(1+x)\middle|\,\Omega_z\right) \leq C_1 \exp\left(-C_2\frac{x^2 \wedge n}{z^4}\right).$$

for all $x > 0$, $j = 1, \cdots, p$.

Proof. Note that

$$\begin{aligned} V_{c,ij} &= \int_0^\tau \left\{S_{ij}^{(2)}(t) - \frac{S_i^{(1)}(t)S_j^{(1)}(t)}{S^{(0)}(t)}\right\}dt - \int_0^\tau \left\{S^{(0)}(t) - \frac{1}{n}\right\}\Sigma_{1,ij}dt, \\ \text{and } \mathcal{V}_{c,ij} &= \int_0^\tau \left\{s_{ij}^{(2)}(t) - \frac{s_i^{(1)}(t)s_j^{(1)}(t)}{s^{(0)}(t)}\right\}dt - \int_0^\tau s^{(0)}(t)\Sigma_{1,ij}dt. \end{aligned}$$

Therefore,

$$\begin{aligned} V_{c,ij} - \mathcal{V}_{c,ij} &= \int_0^\tau \{S_{ij}^{(2)}(t) - s_{ij}^{(2)}(t)\}dt + \int_0^\tau \left\{ \frac{S_i^{(1)}(t)S_j^{(1)}(t)}{S^{(0)}(t)} - \frac{s_i^{(1)}(t)s_j^{(1)}(t)}{s^{(0)}(t)} \right\} dt \\ &\quad - \int_0^\tau \left\{ S^{(0)}(t) - s^{(0)}(t) - \frac{1}{n} \right\} \Sigma_{1,ij} dt \\ &\equiv T_1 + T_2 - T_3. \end{aligned}$$

By Lemma A1, T_1 is bounded in the sense that $\Pr(|T_1| \geq D_1 n^{-1/2}(1+x)|\Omega_z) \leq \exp(-D_2 x^2/z^4)$, where $D_1, D_2 > 0$ are constants. T_2 is bounded by the arguments in Lemma A.4 of Lin and Lv (2013) such that $\Pr(|T_2| \geq D_3 n^{-1/2}(1+x)|\Omega_z) \leq \exp(-D_4 x^2/z^2)$, where $D_3, D_4 > 0$ are constants. By Lemma A1, T_3 is bounded so that $\Pr(|T_3| \geq D_5 n^{-1/2}(1+x)) \leq \exp(-D_6 x^2)$, where $D_5, D_6 > 0$ are constants. Combining these bounds, we obtain the result.

Theorem A1 *Under Regularity Conditions, and assume $n \gtrsim \log p$, $L_c(\cdot)$ satisfies the RSC condition with probability at least $1 - C_1 \exp(-C_2 \log p)$, where C_1, C_2 are constants.*

Proof. By Lemma A3, we have for any constant $x > 0$, there exists constants $D_1, D_2 > 0$, such that we have

$$\Pr\left(|V_{c,ij} - \mathcal{V}_{c,ij}| \geq x|\, \Omega_z\right) \leq D_1 \exp\left(-D_2 \frac{n(x^2 \wedge 1)}{z^4}\right).$$

Thus, let $x = \frac{\lambda_{\min}(\mathcal{V}_c)}{54}$, and assume $n \gtrsim \log p$, we have

$$\begin{aligned} \Pr\left(|v^T(V_c - \mathcal{V}_c)v| \geq \frac{\lambda_{\min}(\mathcal{V}_c)}{54} \middle| \Omega_z \right) &\leq \Pr\left(\|V_{c,ij} - \mathcal{V}_{c,ij}\|_{\max} \geq \frac{\lambda_{\min}(\mathcal{V}_c)}{54} \middle| \Omega_z \right) \\ &\leq \sum_{i,j} \Pr\left(|V_{c,ij} - \mathcal{V}_{c,ij}| \geq \frac{\lambda_{\min}(\mathcal{V}_c)}{54} \middle| \Omega_z \right) \\ &\leq p^2 D_1 \exp\left(-D_2 \frac{n(\frac{\lambda_{\min}^2(\mathcal{V}_c)}{54^2} \wedge 1)}{z^4} \right) \\ &\leq D_1 \exp\left(-D_3 \frac{\log p}{z^4} \right), \end{aligned}$$

where $D_3 > 0$ is a constant.

Together with the proof of Lemma 15 of Loh and Wainwright (2012), we have

$$\Pr\left(\sup_{v\in\mathbb{K}(2s)} |v^T(V_c - \mathcal{V}_c)v| \geq \frac{\lambda_{\min}(\mathcal{V}_c)}{54}\middle|\, \Omega_z\right) \leq C_1 \exp\left(-C_2\frac{n(\frac{\lambda^2_{\min}(\mathcal{V}_c)}{54^2}\wedge 1)}{z^4} + 2s\log p\right).$$

where $\mathbb{K}(2s) = \mathbb{B}_0(2s) \cap \mathbb{B}_2(1)$. Thus, by Lemma 12 of Loh and Wainwright (2012), we have that for probability at least $1 - C_1\exp(-C_2 \log p)$ with constants $C_1, C_2 > 0$,

$$|v^T(V_c - \mathcal{V}_c)v| \leq \frac{\lambda_{\min}(\mathcal{V}_c)}{2}(\|v\|_2^2 + \frac{1}{s}\|v\|_1^2),$$

holds any unit vector v, which leads to the RSC condition for $\alpha = \frac{\lambda_{\min}(\mathcal{V}_c)}{2}$, and $\tau = \frac{\lambda_{\min}(\mathcal{V}_c)}{2s}$, where s is chosen to be greater than 1. The proof is thus completed.

Lemma A4 *Under Regularity Conditions, and assume $n \gtrsim \log p$, for a constant $K > 0$, there exists universal constants $C_1, C_2 > 0$, such that for any fixed p-dimensional unit vector v (i.e., $\|v\|_2 = 1$), we have*

$$\Pr\left(|v^T(V_c - \mathcal{V}_c)v| \geq K\sqrt{\frac{\log p}{n}}\middle|\, \Omega_z\right) \leq C_1 \exp\left(-C_2\frac{\log p}{z^4}\right).$$

for all $x > 0$, $j = 1, \cdots, p$.

Proof. By Lemma A3, we have for any constant $x > 0$, there exists constants $D_1, D_2 > 0$, such that we have

$$\Pr\left(|V_{c,ij} - \mathcal{V}_{c,ij}| \geq x |\, \Omega_z\right) \leq D_1 \exp\left(-D_2\frac{n(x^2 \wedge 1)}{z^4}\right).$$

Thus, let $x = K\sqrt{\frac{\log p}{n}}$ where K is sufficiently large, and assume $n \gtrsim \log p$, we have

$$\begin{aligned}
\Pr\left(|v^T(V_c - \mathcal{V}_c)v| \geq K\sqrt{\frac{\log p}{n}} \middle| \Omega_z\right) &\leq \Pr\left(\|V_{c,ij} - \mathcal{V}_{c,ij}\|_{\max} \geq K\sqrt{\frac{\log p}{n}} \middle| \Omega_z\right) \\
&\leq \sum_{i,j} \Pr\left(|V_{c,ij} - \mathcal{V}_{c,ij}| \geq K\sqrt{\frac{\log p}{n}} \middle| \Omega_z\right) \\
&\leq p^2 D_1 \exp\left(-D_2 \frac{K \log p}{z^4}\right) \\
&\leq D_1 \exp\left(-D_3 \frac{\log p}{z^4}\right),
\end{aligned}$$

where $D_3 > 0$ is a constant. The proof is thus completed.

Based on the above lemmas, the proof of Theorem 3 follows by Lemma 1 of Loh and Wainwright (2012), and Corollary 1 of Loh and Wainwright (2013).

Table 6.1: Simulation results for Scenario 1: $n = 200, p = 50$; values inside the brackets are standard deviations

σ_0	Method	Penalty	l_2 ERROR	l_1 ERROR	#S	#FN
	Oracle		0.406(0.189)	0.869(0.450)	6.0(0.0)	0.0(0.0)
	$\hat{\beta}$	Lasso	1.104(0.308)	3.612(0.892)	21.6(4.8)	0.0(0.4)
		SCAD	0.505(0.283)	1.306(0.778)	10.8(2.9)	0.1(0.6)
		MCP	0.504(0.273)	1.262(0.738)	9.0(2.4)	0.1(0.6)
		SICA	0.485(0.226)	1.111(0.683)	6.7(2.0)	0.0(0.0)
0.1	$\hat{\beta}_c$	Lasso	1.055(0.254)	3.526(0.862)	21.1(4.2)	0.0(0.0)
		SCAD	0.560(0.325)	1.499(1.026)	11.4(2.9)	0.0(0.2)
		MCP	0.553(0.310)	1.390(0.840)	9.3(2.1)	0.0(0.4)
		SICA	0.514(0.265)	1.189(0.777)	7.0(2.5)	0.0(0.0)
0.2		Lasso	1.103(0.269)	3.625(0.870)	20.4(4.3)	0.0(0.0)
		SCAD	0.663(0.382)	1.765(1.003)	12.0(2.7)	0.1(0.4)
		MCP	0.693(0.435)	1.796(1.157)	9.9(2.2)	0.1(0.7)
		SICA	0.613(0.385)	1.435(1.065)	6.9(2.5)	0.1(0.5)

Table 6.2: Simulation results for Scenario 2: $n = 400, p = 800$

σ_0	Method	Penalty	l_2 ERROR	l_1 ERROR	#S	#FN
	Oracle		0.293(0.127)	0.624(0.298)	6.0(0.0)	0.0(0.0)
	$\hat{\beta}$	Lasso	1.574(0.284)	5.476(0.742)	50.7(14.6)	0.1(0.7)
		SCAD	0.389(0.131)	1.406(0.509)	27.5(9.5)	0.0(0.0)
		MCP	0.361(0.143)	1.061(0.525)	14.9(5.9)	0.0(0.0)
		SICA	0.318(0.168)	0.718(0.564)	6.7(4.7)	0.0(0.0)
0.1	$\hat{\beta}_c$	Lasso	1.572(0.281)	5.610(0.658)	53.6(14.7)	0.2(0.9)
		SCAD	0.408(0.152)	1.538(0.537)	30.2(8.8)	0.0(0.0)
		MCP	0.375(0.153)	1.418(0.501)	16.8(5.5)	0.0(0.0)
		SICA	0.304(0.164)	0.660(0.442)	6.4(2.6)	0.0(0.0)
0.2		Lasso	1.537(0.330)	5.603(0.659)	54.5(17.6)	0.3(1.1)
		SCAD	0.570(0.341)	2.227(0.967)	35.7(11.0)	0.1(0.7)
		MCP	0.474(0.194)	1.577(0.640)	20.6(6.0)	0.0(0.0)
		SICA	0.339(0.185)	0.753(0.510)	6.7(3.8)	0.0(0.0)

Table 6.3: Parameter estimation ($\times 1000$). YanYi in the second column indicates the corrected estimator by Yan and Yi (2014b); values inside the brackets in the second column are the difference between the naive estimator by Lin and Ying (1994) and the corrected estimator by Yan and Yi (2014b); values inside the brackets in the third to sixth columns are the difference between the naive penalized estimator by Lin and Lv (2013) and the proposed estimator; #S is the number of nonzero estimated parameters.

	Covariate	YanYi (LinYing)	Lasso	SCAD	MCP	SICA
	age	7.71(-0.54)	6.75(-0.25)	10.97(-0.54)	10.56(-0.57)	4.63(-0.18)
	wtkg	4.81(0.43)	3.28(0.32)	1.11(0.52)	1.11(0.76)	3.20(0.24)
	hemo	-4.97(-0.01)	-2.38(0.00)	-1.14(-0.31)	-1.83(-0.39)	0(0)
	homo	0.85(-0.05)	0(0)	0(0)	0(0)	0(0)
	drugs	-10.78(-0.07)	-8.06(-0.04)	-11.20(-0.01)	-11.11(-0.03)	-5.10(-0.03)
	karnof	-59.69(-0.41)	-58.86(-0.23)	-57.15(-0.32)	-57.10(-0.37)	-61.03(-0.17)
	oprior	-6.17(-0.02)	-5.31(0.05)	-8.91(-0.35)	-9.10(-0.19)	-1.77(0.04)
	z30	6.91(-0.33)	0.75(0.01)	0(0.08)	0(0.13)	0(0)
	preanti	5.45(0.05)	0(0)	0(0)	0(0)	0(0)
	race	-5.56(-0.02)	-3.89(0.00)	-3.64(-0.49)	-4.61(-0.29)	-1.71(0.00)
	gender	-1.04(-0.04)	0(0)	0(0)	0(0)	0(0)
	str2	-33.07(0.53)	0(0)	0(0)	0(0)	0(0)
	strat	-3.04(-0.25)	0(0)	0(0)	0(0)	0(0)
	symptom	12.73(-0.18)	10.65(-0.13)	12.84(-0.23)	12.72(-0.21)	8.82(-0.09)
	offtrt	14.55(-0.10)	12.07(-0.12)	14.77(-0.19)	14.71(-0.19)	9.48(-0.08)
	cd40	19.22(-6.48)	9.02(-3.30)	16.01(-4.59)	15.93(-4.58)	3.20(-2.40)
	cd420	-86.90(3.69)	-78.65(1.64)	-84.98(2.28)	-84.96(2.32)	-74.10(1.20)
	cd80	-7.67(2.99)	0(0)	0(0)	0(0)	0(0)
	cd820	23.42(-2.75)	13.93(-0.17)	17.56(-0.26)	17.53(-0.28)	11.02(-0.12)
	arms	-3.96(-0.07)	-1.55(-0.04)	-1.15(-0.27)	-1.57(-0.37)	0(0)
#S		20(0)	14(0)	13(1)	13(1)	11(0)

Chapter 7

Summary and Discussion

We conclude this thesis with a brief summary and discussion of further extensions.

In Chapter 2, we proposed a correct profile likelihood approach for the classical error model and general error model. The main part of this chapter formulates the paper Yan and Yi (2014a).

In Chapter 3, we studied the impact of misspecifying the error model on score-based estimation and hypothesis testing procedures on the Cox model with functional error models. It is of possible interest to extend the discussion to the full likelihood methods and structural measurement error models.

In Chapter 4, we proposed goodness-of-fit tests for checking the Cox model with covariate measurement error. More simulation studies will be conducted to confirm the theoretical justification of the proposed methods.

In Chapter 5, we proposed various estimation methods for the additive hazards model with covariate error effects accounted for, and studied the impact of ignoring measurement error. The material in this chapter formulates the papers Yan and Yi (2014b, c). Extensions to other additive hazards models (Aalen 1980, 1989; Mckeague and Sasieni 1994) would be an interesting topic to further explore.

In Chapter 6, we considered estimation and variable selection for high dimensional (and ultra-high dimensional) additive hazards model with covariate error through penalized methods. To the best of our knowledge, these methods are the first ones to address high dimensional problems for survival analysis with measurement error. In the future work, we may consider investigating the variable selection properties of ultra-high dimensional additive hazards model with covariate error. Furthermore, high dimensional (and ultra-high dimensional) Cox models with covariate error is a challenging research topic that deserves further research efforts.

Significant progress has been made in the area of survival models with covariate measurement error in the past thirty years. However, many important and interesting problems remain unexplored. In this thesis, we make several important contributions to this area.

References

Aalen, O. O. (1980). A model for non-parametric regression analysis of life times. In Klonecki, W., Kozek, A., and Rosinski, J., editors, *Mathematical Statistics and Probability Theory, Vol. 2 of Lecture Notes in Statistics*, pages 1–25. Springer-Verlag, New York.

Aalen, O. O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine*, 8:907–925.

Aalen, O. O., Borgan, O., and Gjessing, H. K. (2008). *Survival and Event History Analysis: A Process Point of View*. Springer, New York.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

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

Anderson, G. L. and Fleming, T. R. (1995). Model misspecification in proportional hazards regression. *Biometrika*, 82:527–541.

Augustin, T. (2004). An exact corrected log-likelihood function for Cox's proportional hazards model under measurement error and some extensions. *Scandinavian Journal of Statistics*, 31:43–50.

Augustin, T. and Schwarz, R. (2001). Cox's proportional hazards model under covariate measurement error - a review and comparison of methods. *Sonderforschungsbereich*, Discussion Paper 260.

Barlow, W. E. and Prentice, R. L. (1988). Residuals for relative risk regression. *Biometrika*, 75:65–74.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, 2:273–277.

Bertsekas, D. (1999). *Nonlinear Programming*. Athena Scientific, Massachusetts, 2nd edition.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Inference in Semiparametric Models*. Springer-Verlag, New York.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *The Annals of Statistics*, 37:1705–1732.

Bilias, Y., Gu, M., and Ying, Z. (1997). Towards a general asymptotic theory for Cox model with staggered entry. *The Annals of Statistics*, 25:662–682.

Boher, J. and Cook, R. J. (2006). Implications of model misspecification in robust tests for reccurrent events. *Lifetime Data Analysis*, 12:69–95.

Boyd, A. P., Kittelson, J. M., and Gillen, D. L. (2012). Estimation of treatment effect under non-proportional hazards and conditionally independent censoring. *Statistics in Medicine*, 31:3504–3515.

Boyd, S. and Vendenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge, UK.

Bradic, J., Fan, J., and Jiang, J. (2011a). Regularization for Cox's proportional hazards model with np-dimensionality. *The Annals of Statistics*, 39:3092–3120.

Bradic, J., Fan, J., and Wang, W. (2011b). Penalized composite quasi-likelihood for ultrahigh-dimensional variable selection. *Journal of Royal Statistical Society, Series B*, 73:325–349.

Breheny, P. and Huang, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Annals of Applied Statistics*, 5:232–253.

Breslow, D. E. and Day, N. E. (1980). *Statistical Models on Cancer Research, Vol. 1, The Design and Analysis of Case-Control Studies*. IARC, Lyon.

Breslow, N. E. (1972). Discussion of the paper by D. R. Cox. *Journal of the Royal Statistical Society, Series B*, 34:216–217.

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

Bretagnolle, J. and Huber-Carol, C. (1988). Effects of omitting covariates in Cox's model for survival data. *Scandinavian Journal of Statistics*, 15:125–138.

Buhlmann, P. and van de Geer, S. (2011). *Statistics for High-Dimensional Data*. Springer-verlag, Berlin Heidelberg.

Buonaccorsi, J. P. (2010). *Measurement Error and Misclassification: Models, Methods and Applications*. Chapman & Hall, Boca Raton.

Buzas, J. S. (1998). Unbiased scores in proportional hazards regression with covariate measurement error. *Journal of Statistical Planning and Inference*, 67:247–257.

Cai, J., Fan, J., Li, R., and Zhou, H. (2005). Variable selection for multivariate failure time data. *Biometrika*, 92:303–316.

Candes, E. and Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, 35:2313–2351.

Carroll, R. J., Kuchenhoff, H., Lombard, F., and Stefanski, L. A. (1996). Asymptotics for the simex estimator in nonlinear measurement error models. *Journal of the American Statistical Association*, 91:242–250.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman & Hall, Boca Raton, 2nd edition.

Chen, Y., Dougherty, E. R., and Bittner, M. L. (1997). Ratio-based decisions and the quantitative analysis of cdna microarray image. *Journal of Biomedical Optics*, 2:364–374.

Chen, Y. H. (2002). Cox regression in cohort studies with validation sampling. *Journal of the Royal Statistical Society, Series B*, 64:51–62.

Cheng, S. C. and Wang, N. (2001). Linear transformation models for failure time data with covariate measurement error. *Journal of the American Statistical Association*, 96:706–716.

Cheng, S. C., Wei, L. J., and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, 82:835–845.

Cheng, Y. J. and Crainiceanu, C. M. (2009). Cox models with smooth functional effect of covariates measured with error. *Journal of the American Statistical Association*, 104:1144–1154.

Clayton, D. G. (1992). Models for the analysis of cohort and case-control studies with inaccurately measured exposures. In Dwyer, J. H., Feinleib, M., Lipsert, P., and Hoffmeister, H., editors, *Statistical Models for Longitudinal Studies of Health*, pages 301–331. Oxford University Press, New York.

Clegg, L. X., Cai, J., Sen, P. K., and Kupper, L. L. (2000). Misspecification of marginal hazards models in multivariate failure time data. *Sankhya: The Indian Journal of Statistics, Series B*, 62:25–42.

Cook, J. R. and Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, 89:1314–1328.

Cook, R. J. and Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*. Springer-Verlag, New York.

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

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

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman & Hall, Boca Raton.

Cuzick, J. (1988). Rank regression. *The Annals of Statistics*, 16:1369–1389.

DiRienzo, A. G. and Lagakos, S. W. (2001a). Bias correction for score tests arising from misspecified proportional hazards regression models. *Biometrika*, 88:421–434.

DiRienzo, A. G. and Lagakos, S. W. (2001b). Effects of model misspecification on tests of no randomized treatment effect arising from Cox's proportional hazards model. *Journal of the Royal Statistical Society, Series B*, 63:745–757.

Dupuy, J. F. (2005). The proportional hazards model with covariate measurement error. *Journal of Statistical Planning and Inference*, 85:260–275.

Fan, J., Li, G., and Li, R. (2005). An overview of variable selection for survival analysis. In Fan, J. and Li, G., editors, *Contemporary Multivariate Analysis and Design of Experiment, In celebration of Prof. Kai-Tai Fang's 65th Birthday*, pages 315–336. World Scientific, Singapore.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96:1348–1360.

Fan, J. and Li, R. (2002). Variable selection for Coxs proportional hazards model and frality model. *The Annals of Statistics*, 30:74–99.

Fan, J. and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20:101–148.

Fan, J. and Lv, J. (2011). Non-concave penalized likelihood with np-dimensionality. *IEEE Transactions on Information Theory*, 57:5467–5484.

Fan, J., Xue, L., and Zou, H. (2014). Strong oracle optimality of folded concave penalized estimation. *The Annals of Statistics*, 42:819–849.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Springer-Verlag, New York.

Friedman, J., Hastie, T., Hoefling, H., and Tibshirani, R. (2007). Pathwise coordinate optimization. *Annals of Applied Statistics*, 1:302–332.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33:1–22.

Fuchs, H. J., Borowitz, D. S., Christiansen, D. H., Morris, E. M., Nash, M. L., Ramsey, B. W., Rosenstein, B. J., Smith, A. L., and Wohl, M. E. (1994). Effect of aerosolized recombinant human dnase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. *The New England Journal of Medicine*, 331:637–942.

Fuller, W. A. (1987). *Measurement Error Models*. Wiley, Hoboken.

Gaiffas, S. and Guilloux, A. (2012). High-dimensional additive hazards models and the lasso. *Electronic Journal of Statistics*, 6:522–546.

Gail, M. H., Tan, W. Y., and Piantadosi, S. (1988). Tests for no treatment effect in randomized clinical trials. *Biometrika*, 75:57–64.

Gail, M. H., Wieand, S., and Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*, 71:431–444.

Gerds, T. A. and Schumacher, M. (2001). On functional misspecification of covariates in the Cox regression model. *Biometrika*, 88:572–580.

Golub, T., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J., Coller, H., Loh, M. L., Downing, J., Caligiuri, M. A., and et al. (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, 286:531–536.

Gorfine, M., Hsu, L., and Prentice, R. L. (2003). Estimation of dependence between paired correlated failure times in the presence of covariate measurement error. *Journal of the Royal Statistical Society, Series B*, 65:643–661.

Gorfine, M., Hsu, L., and Prentice, R. L. (2004). Nonparametric correction for covariate measurement error in a stratified Cox model. *Biostatistics*, 5:75–87.

Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81:515–526.

Grambsch, P. M., Therneau, T. M., and Fleming, T. R. (1995). Diagnostic plots to reveal functional form for covariates in multiplicative intensity models. *Biometrics*, 51:1469–1482.

Greene, W. F. and Cai, J. (2004). Measurement error in covariates in the marginal hazards model for multivariate failure time data. *Biometrics*, 60:987–996.

Hall, W. J. and Wellner, J. A. (1980). Confidence bands for a survival curve from censored data. *Biometrika*, 67:133–143.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundaker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., Hirsch, M. S., and Merigan, T. C. (1996). A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter. *The New England Journal of Medicine*, 335:1081–1090.

Hattori, S. (2006). Some properties of misspecified additive hazards models. *Statistics & Probability Letters*, 76:1641–1646.

Hattori, S. (2012). Testing the no-treatment effect based on a possibly misspecified accelerated failure time model. *Statistics & Probability Letters*, 82:371–377.

Hattori, S. and Henmi, M. (2012). Estimation of treatment effects based on possibly misspecified Cox regression. *Lifetime Data Analysis*, 18:408–433.

He, W., Yi, G. Y., and Xiong, J. (2007). Accelerated failure time models with covariates subject to measurement error. *Statistics in Medicine*, 26:4817–4832.

Hu, C. and Lin, D. Y. (2002). Cox regression with covariate measurement error. *Scandinavian Journal of Statistics*, 29:637–655.

Hu, C. and Lin, D. Y. (2004). Semiparametric failure time regression with replicates of mismeasured covariates. *Journal of the American Statistical Association*, 99:105–118.

Hu, P., Tsiatis, A. A., and Davidian, M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics*, 54:1407–1419.

Huang, J., Sun, T., Ying, Z., Yu, Y., and Zhang, C.-H. (2013). Oracle inequalities for the lasso in the Cox model. *The Annals of Statistics*, 41:1142–1165.

Huang, Y. and Wang, C. Y. (2000). Cox regression with accurate covariates unascertainable: A nonparametric-correction approach. *Journal of the American Statistical Association*, 95:1209–1219.

Huang, Y. and Wang, C. Y. (2001). Consistent functional methods for logistic regression with errors in covariates. *Journal of the American Statistical Association*, 96:1469–1482.

Hughes, M. D. (1993). Regression dilution in the proportional hazards model. *Biometrics*, 49:1046–1066.

Jiang, W., Turnbull, B. W., and Clark, L. C. (1999). Semiparametric regression models for repeated events with random effects and measurement error. *Journal of the American Statistical Association*, 94:111–124.

Juraska, M. (2010). Package 'speff2trial'. *R software*.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, Hoboken, 2nd edition.

Kim, J., Song, M. S., and Lee, S. (1998). Goodness-of-fit tests for the additive risk model with ($p > 2$)-dimensional time-invariant covariates. *Lifetime Data Analysis*, 4:405–416.

Kong, F. H. (1999). Adjusting regression attenuation in the Cox proportional hazards model. *Journal of Statistical Planning and Inference*, 79:31–44.

Kong, F. H. and Gu, M. (1999). Consistent estimation in Cox proportional hazards model with covariate measurement errors. *Statistica Sinica*, 9:953–969.

Kong, F. H. and Slud, E. (1997). Robust covariate-adjusted logrank tests. *Biometrika*, 84:847–862.

Kong, S. and Nan, B. (2014). Non-asymptotic oracle inequalites for the high-dimensional Cox regression via lasso. *Statistica Sinica*, 24:25–42.

Kosorok, M. R., Lee, B. L., and Fine, J. P. (2004). Robust inference for univariate proportional hazards frailty regression models. *The Annals of Statistics*, 32:1448–1491.

Kuchenhoff, H., Bender, R., and Langner, I. (2007). Effect of berkson measurement error on parameter estimates in Cox regression models. *Lifetime Data Analysis*, 13:261–272.

Kulich, M. and Lin, D. Y. (2000). Additive hazards regression with covariate measurement error. *Journal of the American Statistical Association*, 95:238–248.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86.

Lagakos, S. W. (1988a). Effects of mismodelling and mismeasuring explanatory variables on tests of their association with a response variable. *Statistics in Medicine*, 7:257–274.

Lagakos, S. W. (1988b). The loss in efficiency from misspecifying covariates in proportional hazards regression models. *Biometrika*, 75:156–160.

Lagakos, S. W. and Schoenfeld, D. A. (1984). Properties of proportional-hazards score tests under misspecified regression models. *Biometrics*, 40:1037–1048.

Latouche, A., Boisson, V., Chevret, S., and Porcher, R. (2007). Misspecified regression model for the subdistribution hazard of a competing risk. *Statistics in Medicine*, 26:965–974.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. Wiley, Hoboken, 2nd edition.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypothesis*. Springer, New York, 3rd edition.

Lemler, S. (2012). Oracle inequalities for the lasso for the conditional hazard rate in a high-dimensional setting. *Arxiv*.

Leng, C. and Ma, S. (2007). Path consistent model selection in additive risk model via lasso. *Statistics in Medicine*, 26:3753–3770.

Li, Y. and Lin, X. (2000). Covariate measurement errors in frailty models for clustered survival data. *Biometrika*, 87:849–866.

Li, Y. and Lin, X. (2003). Functional inference in frailty measurement error models for clustered survival data using the simex approach. *Journal of the American Statistical Association*, 98:191–203.

Li, Y. and Ryan, L. (2004). Survival analysis with heterogeneous covariate measurement error. *Journal of the American Statistical Association*, 99:724–735.

Li, Y. and Ryan, L. (2006). Inference on survival data with covariate measurement error-an imputation-based approach. *Scandinavian Journal of Statistics*, 33:169–190.

Liang, H. and Li, R. (2009). Variable selection for partially linear models with measurement errors. *Journal of the American Statistical Association*, 104:234–248.

Liao, X., Zucker, D. M., Li, Y., and Spiegelman, D. (2011). Survival analysis with error-prone time-varying covariates, a risk set calibartion approach. *Biometrics*, 67:50–58.

Lin, D. Y. (1991). Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators. *Journal of the American Statistical Association*, 86:725–728.

Lin, D. Y. and Spiekerman, C. F. (1996). Model checking techniques for parametric regression with censored data. *Scandinavian Journal of Statistics*, 23:157–177.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association*, 84:1074–1078.

Lin, D. Y. and Wei, L. J. (1991). Goodness-of-fit tests for the general Cox regression model. *Statistica Sinica*, 1:1–17.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of Royal Statistical Society, Series B*, 62:711–730.

Lin, D. Y., Wei, L. J., and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80:557–572.

Lin, D. Y., Wei, L. J., and Ying, Z. (2002). Model checking techniques based on cumulative residuals. *Biometrics*, 58:1–12.

Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika*, 81:61–71.

Lin, D. Y. and Ying, Z. (1995). Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *The Annals of Statistics*, 22:1712–1734.

Lin, D. Y. and Ying, Z. (1997). Additive hazards regression models for survival data. In Lin, D. Y. and Fleming, T. R., editors, *Proceedings of the First Seattle Symposium in Biostatistics, Vol.123 of Lecture Notes in Statistics*, pages 185–198. Springer, New York.

Lin, W. and Lv, J. (2013). High-dimensional sparse additive hazards regression. *Journal of the American Statistical Association*, 108:247–264.

Loh, P.-L. and Wainwright, M. J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. *The Annals of Statistics*, 40:1637–1664.

Loh, P.-L. and Wainwright, M. J. (2013). Regularized m-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *Advances in Neural Information Processing Systems 26 (NIPS 2013).*

Lv, J. and Fan, Y. (2009). A unified approach to model selection and sparse recovery using regularized least squares. *The Annals of Statistics*, 37:3498–3528.

Ma, Y., Hart, J. D., Janicki, R., and Carroll, R. J. (2011). Local and omnibus goodness-of-fit tests in classical measurement error models. *Journal of the Royal Statistical Society, Series B*, 73:81–98.

Ma, Y. and Li, R. (2010). Variable selection in measurement error models. *Bernoulli*, 16:274–300.

Ma, Y. and Yin, G. (2008). Cure rate model with mismeasured covariates under transformation. *Journal of the American Statistical Association*, 103:743–756.

Martinussen, T. and Scheike, T. H. (2009). Covariate selection for the semiparametric additive risk model. *Scandinavian Journal of Statistics*, 36:602–619.

Massart, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *The Annals of Probability*, 28:863–884.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall, London, 2nd edition.

McKeague, I. W. and Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika*, 81:501–514.

Morgan, T. M. (1986). Omitting covariates from the proportional hazards model. *Biometrics*, 42:993–995.

Murphy, S. A. and van der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95:449–465.

Nakamura, T. (1990). Corrected score functions for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, 77:127–137.

Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics*, 48:829–838.

Negahban, S. N., Ravikumar, P., Wainwright, M. J., and Yu, B. (2012). A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. *Statistical Science*, 27:538–557.

Novick, S. J. and Stefanski, L. A. (2002). Corrected score estimation via complex variable simulation extrapolation. *Journal of the American Statistical Association*, 97:472–481.

O'Neill, T. J. (1986). Inconsistency of the misspecified proportional hazards model. *Statistics & Probability Letters*, 4:219–222.

Pepe, M. S., Self, S. G., and Prentice, R. L. (1989). Further results in covariate measurement errors in cohort studies with time to response data. *Statistics in Medicine*, 8:1167–1178.

Pettitt, A. N. (1982). Inference for the linear model using a likelihood based on ranks. *Journal of the Royal Statistical Society, Series B*, 44:234–243.

Pipper, C. B. and Ritz, C. (2007). Checking the grouped data version of the Cox model for interval-grouped survival data. *Scandinavian Journal of Statistics*, 34:405–418.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. IMS, Hayward.

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69:331–342.

Prentice, R. L., Sugar, E., Wang, C. Y., Neuhouser, M., and Patterson, R. (2002). Research strategies and the use of nutrient biomarkers in studies of diet and chronic disease. *Public Health Nutrition*, 5:977–984.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton.

Rocke, D. M. and Durbin, B. (2001). A model for measurement error for gene expression arrays. *Journal of Computational Biology*, 8:557–569.

Rosenbaum, M. and Tsybakov, A. B. (2010). Sparse recovery under matrix uncertainty. *The Annals of Statistics*, 38:2620–2651.

Rosenbaum, M. and Tsybakov, A. B. (2013). Improved matrix uncertainty selector. *IMS Collections: From Probability to Statistics and Back, High-Dimensional Models and Processes*, 9:276–290.

Rosenwald, A., Wright, G., Chan, W. C., Connors, J. M., Campo, E., Fisher, R. I., Gascoyne, R. D., Muller-Hermelink, H. K., Smeland, E. B., , and Staudt, L. M. (2002). The use of molecular profiling to predict survival after chemotherapy for diffuse large-b-cell lymphoma. *The New England Journal of Medicine*, 346:1937–1947.

Schoenfeld, D. (1980). Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika*, 67:145–153.

Schoenfeld, D. (1981). The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika*, 68:316–319.

Shaw, P. A. and Prentice, R. L. (2012). Hazard ratio estimation for biomarker-calibrated dietary exposures. *Biometrics*, 68:397–407.

Solomon, P. J. (1984). Effect of misspecification of regression models in the analysis of survival data. *Biometrika*, 71:291–298.

Song, X., Davidian, M., and Tsiatis, A. A. (2002). An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics*, 3:511–528.

Song, X. and Huang, Y. (2005). On corrected score approach for proportional hazards model with covariate measurement error. *Biometrics*, 61:702–714.

Sorensen, O., Frigessi, A., and Thoresen, M. (2014). Measurement error in lasso: Impact and correction. *Arxiv*.

Spiekerman, C. F. and Lin, D. Y. (1996). Checking the marginal Cox model for correlated failure time data. *Biometrika*, 83:143–156.

Spiekerman, C. F. and Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association*, 93:1164–1175.

Stefanski, L. A. (1989). Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Communication in Statistics - Theory and Methods*, 18:4335–4358.

Stefanski, L. A. and Cook, J. R. (1995). Simulation-extrapolation: The measurement error jackknife. *Journal of the American Statistical Association*, 90:1247–1256.

Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. *Biometrika*, 73:363–369.

Sun, L., Song, X., and Mu, X. (2012). Regression analysis for the additive hazards model with covariate errors. *Communications in Statistics - Theory and Methods*, 41:1911–1932.

Sun, L., Zhang, Z., and Sun, J. (2006). Additive hazards regression of failure time data with covariate measurement errors. *Statistica Neerlandica*, 60:497–509.

Sun, L. and Zhou, X. (2008). Inference in the additive risk model with time-varying covariates subject to measurement error. *Statistics & Probability Letters*, 78:2559–2566.

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

Therneau, T. M., Grambsch, P. M., and Fleming, T. R. (1990). Martingale based residuals for survival models. *Biometrika*, 77:147–160.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of Royal Statistical Society, Series B*, 58:267–288.

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Statistics in Medicine*, 16:385–395.

Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*. Springer-Verlag, New York.

Tsiatis, A. A. and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88:447–458.

Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica*, 14:809–834.

Turnbull, B. W., Jiang, W., and Clark, L. C. (1997). Regression models for recurrent event data: parametric random effects models with measurement error. *Statistics in Medicine*, 16:853–864.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, New York.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

van der Vaart, A. W. and Wellner, J. A. (2007). Empirical processes indexed by estimated functions. *Asymptotics: Particles, Processes and Inverse Problems, IMS Lecture Notes-Monograph Series*, 55:234–252.

Wang, C. Y. (2008). Non-parametric maximum likelihood estimation for Cox regression with subject-specific measurement error. *Scandinavian Journal of Statistics*, 35:613–628.

Wang, C. Y., Hsu, L., Feng, Z. D., and Prentice, R. L. (1997). Regression calibration in failure time regression. *Biometrics*, 53:131–145.

Wang, C. Y. and Pepe, M. S. (2000). Expected estimating equations to accommodate covariate measurement error. *Journal of the Royal Statistical Society, Series B*, 62:509–524.

Wang, Z., Liu, H., and Zhang, T. (2014). Optimal computational and statistical rates of convergence for sparse nonconvex learning problems. *The Annals of Statistics, accepted*.

Wei, L. J. (1984). Testing goodness of fit for proportional hazards model with censored observations. *Journal of the American Statistical Association*, 79:649–652.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, 84:1065–1073.

Wen, C. C. (2010). Semiparametric maximum likelihood estimation in Cox proportional hazards model with covariate measurement errors. *Metrika*, 72:199–217.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50:1–25.

Xie, S. X., Wang, C. Y., and Prentice, R. L. (2001). A risk set calibration method for failure time regression by using a covariate reliability sample. *Journal of the Royal Statistical Society, Series B*, 63:855–870.

Xu, R. and O'Quigley, J. (2000). Estimating average regression effect under non-proportional hazards. *Biostatistics*, 1:423–439.

Yan, Y. and Yi, G. Y. (2014a). A corrected profile likelihood method for survival data with covariate measurement error under the Cox model. *Canadian Journal of Statistics. Revised.*

Yan, Y. and Yi, G. Y. (2014b). A class of functional methods for error-contaminated survival data under additive hazards models with replicate measurements. *Journal of the American Statistical Association. Accepted with minor revision.*

Yan, Y. and Yi, G. Y. (2014c). Analysis of error-prone survival data under additive hazards models: Measurement error effects and adjustments. *Lifetime Data Analysis. Under revision.*

Yanagimoto, T. and Yamamoto, E. (1991). The role of unbiasedness in estimating equations. In Godambe, V. P., editor, *Estimating Functions*, pages 89–101. Oxford University Press, New York.

Yi, G. Y. and He, W. (2006). Methods for bivariate survival data with mismeasured covariates under an accelerated failure time model. *Communications in Statistics - Theory and Methods*, 35:1539–1554.

Yi, G. Y. and He, W. (2012). Bias analysis and the simulation-extrapolation method for survival data with covariate measurement error under parametric proportional odds models. *Biometrical Journal*, 54:343–360.

Yi, G. Y. and Lawless, J. F. (2007). A corrected likelihood method for the proportional hazards model with covariates subject to measurement error. *Journal of Statistical Planning and Inference*, 137:1816–1828.

Yi, G. Y. and Reid, N. (2010). A note on mis-specified estimating functions. *Statistica Sinica*, 20:1749–1769.

Yin, G. (2007). Model checking for additive hazards model with multivariate survival data. *Journal of Multivariate Analysis*, 98:1018–1032.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 28:894–942.

Zhou, H. and Pepe, M. S. (1995). Auxiliary covariate data in failure time regression. *Biometrika*, 82:139–149.

Zhou, H. and Wang, C. Y. (2000). Failure time regression with continuous covariates measured with error. *Journal of the Royal Statistical Society, Series B*, 62:657–665.

Zucker, D. M. (2005). A pseudo-partial likelihood method for semiparametric survival regression with covariate errors. *Journal of the American Statistical Association*, 100:1264–1277.

Zucker, D. M. and Spiegelman, D. (2004). Inference for the proportional hazards model with misclassified discrete-valued covariates. *Biometrics*, 60:324–334.

Zucker, D. M. and Spiegelman, D. (2008). Corrected score estimation in the proportional hazards model with misclassified discrete covariates. *Statistics in Medicine*, 27:1911–1933.