STATISTICAL ESTIMATION AND INference FOR THE ASSOCIATIONS OF MULTIVARIATE RECURRENt EVENT PROCESSES

by

Peilin Chen

A dissertation submitted to the faculty of The University of North Carolina at Charlotte in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Applied Mathematics Charlotte 2019

Approved by:

___________________________________________________________
Dr. Yanqing Sun

___________________________________________________________
Dr. Shaoyu Li

___________________________________________________________
Dr. Qingning Zhou

___________________________________________________________
Dr. Aidong Lu
ABSTRACT

PEILIN CHEN. Statistical Estimation and Inference for the Associations of Multivariate Recurrent Event Processes. (Under the direction of DR. YANQING SUN)

In this dissertation, we aim to develop a brand new method with a two-stage procedure to investigate the association between multivariate recurrent event processes.

First, under the assumption of independent censoring, we model each recurrent event process marginally through a mean rate model. There are two popular mean rate assumptions - multiplicative or additive to an unspecified baseline rate function. The robust semi-parametric approaches can be applied to estimate the covariate effects as well as the baseline rate function.

Second, inspired by Kendall’s tau, we propose the rate ratio as an association measurement, which is the quotient of two conditional rates - the mean rate of two joint events over the marginal rates, both conditional on the covariates. Utilizing the information from the first stage, an unbiased and consistent estimator of the rate ratio is developed under the Generalized Estimation Equation method. The asymptotic properties of the rate ratio estimators are derived theoretically. Without modeling the joint events directly, the rate ratio can measure the association between two recurrent processes over time.

Since the rate ratio we proposed can be parametric, time and covariate dependent, it has a good interpretability. We developed a formal hypothesis testing procedure to validate the parametric assumption of the rate ratio. Simulation studies shows it is quite powerful under moderate to strong association.
ACKNOWLEDGMENTS

I would like to express my deep gratitude to my advisor Dr. Yanqing Sun, who not only guided me academically but also give me support mentally. Her rigid statistical thoughts and passionate hard working spirit have highly impressed and influenced me.

I also want to thank Dr. Shaoyu Li, Dr. Qingning Zhou and Dr. Aidong Lu for their constructive and thoughtful suggestions as well as serving in my dissertation committee. I would also like to thank Dr. Peter Gilbert at Fred Hutchinson Cancer Research Center and University of Washington for motivating discussions on the future applications to the Malaria vaccine study. This research was partially supported by the National Science Foundation grant DMS1513072 and NIAID NIH award number R37AI054165.

I am very grateful for the support of all the honorable professors in the Department of Applied Mathematics and Statistics, who provided me such unforgettable study experience.

My dissertation cannot be completed without the support from my family. Especially my dear husband Alexander Hohl, who always encourages and supports me with his endless love.
# TABLE OF CONTENTS

**LIST OF FIGURES** vii
**LIST OF TABLES** viii

**CHAPTER 1: INTRODUCTION** 1

1.1. Bivariate /Multivariate Recurrent Event Data 1
1.2. Modeling Recurrent Event Data 2
1.3. Modeling Multivariate Recurrent Event 4
1.4. Study of Associations 5

**CHAPTER 2: CONDITIONAL RATE RATIO AS ASSOCIATION MEASURE FOR MULTIVARIATE RECURRENT EVENT PROCESSES** 7

2.1. Preliminaries 7
2.2. Estimation and Inference Procedures 8

**CHAPTER 3: ESTIMATION AND INFERENCE OF THE RATE RATIO UNDER THE ADDITIVE MARGINAL MODEL** 10

3.1. Estimation by a two-stage approach 10
3.1.1. Review of the estimation of the marginal model 11
3.1.2. Estimation of the rate ratio 13
3.1.3. Simulation studies 14
3.2. Hypothesis testing of the rate ratio 27
3.2.1. Procedure description 27
3.2.2. Simulation studies 29
LIST OF FIGURES

FIGURE 1: Visualization of Piecewise Constant $\rho(s, t, \theta)$ (PWC) under the Additive Marginal Models. The variation of $\rho(s, t)$ between different pieces is growing from PWC1 to PWC4. 33

FIGURE 2: The contour plot of the Rate Ratio $\rho(s, t)$ under the additive marginal mean rate models. The x-axis and y-axis represents the observation time for type1 and type2 events. From upper left to lower right, the heterogeneity of $\rho(s, t)$ is increased. 34

FIGURE 3: Visualization of Piecewise Constant $\rho(s, t, \theta)$ (PWC) under the Additive Marginal Models. The variation of $\rho(s, t)$ between different pieces is growing from PWC1 to PWC4. 62

FIGURE 4: The contour plot of the Rate Ratio $\rho(s, t)$ under the Multiplicative Marginal Models. The x-axis and y-axis represents the observation time for type1 and type2 events. From upper left to lower right, the heterogeneity of $\rho(s, t)$ is increased. 63
### LIST OF TABLES

**TABLE 1**: Scenario I - $\rho(s,t,\theta) = \theta_0$. Estimation of coefficients in the marginal additive model. The Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Errors) and the Empirical Coverage Probability of 95% confidence interval (CP) of $(\beta_{01}, \beta_{02})$. Each entry is based on 1000 simulations.

**TABLE 2**: Scenario I - Estimation of $\rho(s,t,\theta) = \theta_0$. Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP (95% Coverage Probability) lists. Each entry is based on 1000 simulated datasets. The marginal models are additive and association come from the shared random effect.

**TABLE 3**: Scenario II - Estimation of $\rho(s,t,\theta) = 1 + \theta_0(-0.15t + 0.9)(-0.152 + 0.9)$. Summary of Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP for the parameter $\theta_0$ in $\rho(s,t,\theta)$. Each entry is based on 1000 simulations with correctly specified marginals and Rate Ratio form.

**TABLE 4**: Scenario III - Estimation of $\theta$’s in $\rho(\theta; Z_k) = \theta_1 I(Z_k = 1) + \theta_2 I(Z_k = 0)$, with true value $\theta_1 = 1.25$ and $\theta_2 = 1.75$. Summary of Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP. Each entry is based on 1000 simulations with correctly specified marginals and rate ratio form.

**TABLE 5**: Scenario IV - estimates $\theta_1, \theta_2$ in the underline models where $\rho(\theta, s, t|Z_k) = 1 + \theta_1(0.25 + 0.1Z_k)(0.25 + 0.25Z_k)$ and $\rho(\theta, s, t|Z_k) = 1 + \theta_2(0.5 + 0.1Z_k)(0.5 + 0.2Z_k)$. Summary of Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP of $\theta$ where each entry is based on 1000 simulations. The averaged observed events for type1 (2) event is 2.44(2.56)

**TABLE 6**: Summary of simulation settings under the piecewise constant rate ratio model with the corresponding $\rho$ values followed from Proposition 2.

**TABLE 7**: Simulation settings of the Time Varying Rate Ratio (TD models). From TD1 to TD4, the value of $\sigma^2/\mu^2$ is increasing and so is the association between the bivariate recurrent event processes.
TABLE 8: Observed sizes and powers of the test statistic T via the proposed model-checking procedure under $H_0 : \rho = 1$ vs $H_a : \rho(s, t, \theta) = \theta$ and $\theta > 1$, at significance level 0.05. The numbers in the parentheses represent the count for type 1 and type 2 event across the observation period. Each entry is calculated based on 1000 Gaussian multiplier samples with 1000 replicates.

TABLE 9: Observed sizes and powers of the test statistic T for the proposed model-checking procedure under $H_0 : \rho(\theta, s, t) = \theta$ (i.e. constant) vs $H_a : \rho$ is not constant, at 0.05 significance level. Each entry is calculated based on 1000 Gaussian multiplier samples with 1000 replicates.

TABLE 10: Scenario I: Numerical results for $(\beta_1, \beta_2)$ with true value equals (0.2, 0.4). Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP summarized for $(\beta_1, \beta_2)$. Each entry is based on 1000 simulated datasets under shared random effect model with Multiplicative marginals.

TABLE 11: Scenario I - Estimation of $\rho$ with summary of Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP. Each entry is based on 1000 simulations under shared random effect model with Multiplicative marginals

TABLE 12: Scenario II - Estimation of $\theta$ in $\rho(s, t, \theta) = 1 + \theta(-0.15t + 0.9)(-0.15s + 0.9)$. The summary of Bias, SEE (Standard Error of Estimates), ESE(Estimated Standard Error) and CP (Coverage Probability). The Marginal model is multiplicative and the parametric form of $\rho(s, t, \theta)$ is correctly specified. Each entry is based on 1000 simulations.

TABLE 13: Scenario III - Summary of Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP. The Rate Ratio is covariate dependent, where true values $\rho(\theta, s, t; Z_k) = \theta_1I(Z_k = 1) + \theta_2I(Z_k = 0)$, with true value $\theta_1 = 1.25$ and $\theta_2 = 1.75$. Each entry is based on 1000 simulations with correctly specified multiplicative marginals and Rate Ratio form.
TABLE 14: Scenario IV - estimates $\theta_1, \theta_2$ in the underline models where
\[ \rho(\theta, s, t|Z_k) = 1 + \theta_1 \frac{(0.25te^{0.12k_1})(0.25se^{0.22k_2})}{(0.25+0.25te^{0.12k_1})(0.25+0.25se^{0.22k_2})} \]
and \[ \rho(\theta, s, t|Z_k) = 1 + \theta_2 \frac{(0.5te^{0.12k_1})(0.5se^{0.22k_2})}{(0.25+0.5te^{0.12k_1})(0.25+0.5se^{0.22k_2})} \]. With the true values of $\theta_1, \theta_2$ equal to 0.25, 0.5, 0.75 and 1.00. Summary of Bias, SEE (Standard Error
of Estimates), ESE (Estimated Standard Error), CP of $\theta$ where each entry is based on 1000 simulations. The averaged observed events for type 1(2) event is 2.44(2.56)

TABLE 15: Summary of simulation settings under the PWC model with the corresponding $\rho$ values followed from Proposition 2. The Marginal model is multiplicative.

TABLE 16: Observed size of the test statistic $T$ for the proposed model-checking procedure under $H_0 : \rho(\theta, s, t) = \theta$ is parametric vs $H_a : \rho(\theta, s, t)$ is not parametric, at significance level 0.05. The numbers in the parentheses represent the average observed count of type 1 and type 2 event after censoring. Each entry is calculated based on 1000 Gaussian multiplier samples and 1000 replicates.

TABLE 17: Power of $H_0 : \rho(\theta, s, t) = \theta_0$ vs $H_a : \rho(\theta, s, t)$ is not parametric. The $H_a$ model has Piecewise Constant Rate Ratio (PWC model). Each entry is calculated based on 1000 Gaussian multiplier samples with 1000 replicates.

TABLE 18: Power of $H_0 : \rho(\theta, s, t) = \theta_0$ vs $H_a : \rho(\theta, s, t)$ is not parametric. The $H_a$ model is Time and Dependent (TD). Each entry is calculated based on 1000 Gaussian multiplier samples with 1000 replicates.

TABLE 19: Power of $H_0 : \rho(\theta, s, t) = \theta_0$ vs $H_a : \rho(\theta, s, t)$ is not parametric. The $H_a$ model is Time and Covariate Dependent (TCD). Each entry is calculated based on 1000 Gaussian multiplier samples and 1000 replicates.
CHAPTER 1: INTRODUCTION

This chapter aims to review related works and introduce the benefits and challenges of estimating the association between multivariate recurrent event processes. The structure of this chapter is as following. In section 1.1-1.2 we review the basic background for Recurrent Event Data and popular approaches to estimate the mean event rate or the intensity of Hazard. Literatures that focus on modeling multivariate Recurrent Event Data are discussed in Section 1.3.

1.1 Bivariate /Multivariate Recurrent Event Data

Recurrent events involve repeat occurrences of the same type of event over time, whereas a process that generate such data are called recurrent event process. Examples of recurrent events include multiple relapses from remission for leukemia patients, wild fires, and hurricanes. In Recent years, recurrent event data raises in many fields such as public health, business and industry, reliability, the social sciences, and insurance, and keep receiving fast growing attention. For instance, the tumor development time for 48 rats who were injected with a carcinogen represented Gail1980; the automobile warranty claims data for a specific car model considered by Lawless and Nadeau (1995).

Bivariate or multivariate recurrent event processes are often encountered in longitudinal data studies involving more than one type of event of interest. Unlike Life
Data which is valid to assume events are independent, recurrent event data are usually correlated because they represent the event time measured for the same subject over a time period.

1.2 Modeling Recurrent Event Data

Many statistical methods focus on modeling the rate or intensity of the event recurrence. Nelson (1988, 1995) proposed the nonparametric estimation of the mean function for general processes and Aalen (1978) studied the properties of the Nelson-Aalen estimate in the Poisson case. Early development was extended from survival analysis for the Cox Proportional hazards model (Cox, 1972a). Anderson and Gill (1982) introduced the semiparametric regression model for the rate functions and derived the asymptotic results based on the counting process theory.


Event rate models recently became more popular than the intensity based model because they are easier to interpret. Lin et al. (2000) compared the intensity and
rate based model. In their paper, $N^*(t)$ denotes the number of events occur over time $[0,t]$ and $Z(\cdot)$ is a $p$-dimensional covariate process, whereas $\mathcal{F}_t$ is the history of $\{N^*(s), Z(s) : 0 \leq s \leq t\}$ and $\lambda_Z(t)$ is the intensity of $N^*(t)$ associated with $\mathcal{F}_t$.

The Anderson-Gill intensity model

$$\lambda_Z(t) = e^{\beta^T Z(t) \lambda_0(t)} \quad (1.1)$$

is a special case under the assumptions that (a) $E[dN^*(t)|\mathcal{F}_t] = E[dN^*(t)|Z(t)]$ and (b) $E[dN^*(t)|Z(t)] = e^{\beta^T Z(t) \lambda_0(t)} dt$.

Lin (2000) proposed a mean rate model

$$E[dN^*(t)|Z(t)] = d\mu_Z(t) \quad (1.2)$$

without assumption (a), which is impractical to verify if the time-varying covariates adequately captured the dependence of the recurrent events. The regression coefficients in the mean event rate model nicely reflect covariate effects on the frequency.

Compared to the Anderson-Gill model (1.1), which is a special case of equation (1.2) by taking

$$d\mu_Z(t) = e^{\beta^T Z(t) \mu_0(t)},$$

$$d\mu_0(t) = \lambda_0(t) dt,$$

model (1.2) is more versatile.
1.3 Modeling Multivariate Recurrent Event

Here, we introduce the Random Effect Models for Multitype Events here. For more details consult Cook and Lawless (2007). Let $k$ index the subjects (or clusters) and $j$ index the event type. The event rate at time $t$ for events of type $j$ conditional on subject and type-specific positive random effect $r_{kj}$ is denoted by

$$
\lambda_{kj}(t|F_{kt}, r_{kj}) = \lim_{\Delta t \to 0^+} \frac{Pr(\Delta N_{kj}(t) = 1|F_{kt}, r_{kj})}{\Delta t} \quad (1.3)
$$

where $r_{kj}$ denotes the multivariate random effect. With multivariate random effects, it is often assumed that conditional on $r_{kj}$ and $F_{kt} = \{N_{kj}(s), Z_{kj}: 0 \leq s \leq t\}$, type $i$ and type $j$ event are independent if $i \neq j$, that is

$$
\lambda_{kj}(t|F_{kt}, r_{kj}) = r_{kj}\lambda_{kj}(t|F_{kt}) \quad (1.4)
$$

Random effect models are usually parameterized by assuming $r_{kj}$ comes from an underlying distribution $G(r_k; \phi)$ so that $E(r_{kj}) = 1$, $\text{var}(r_{kj}) = \phi_j$ and $\text{cov}(r_{kj}, r_{ij}) = \phi_{ki}$. The corresponding likelihood conditional on $r_{kj}$ is

$$
\prod_{j=1}^{J} \left\{ \prod_{l=1}^{r_{kj}} r_{kj}\lambda_{kj}(t_{kjl}|F_{kt}) \exp \left( -r_{kj} \int_{0}^{t_{kjl}} \lambda_{kj}(u|F_{kt}) \, du \right) \right\}, \quad (1.5)
$$

and the marginal likelihood for individual $k$ as

$$
\int \prod_{j=1}^{J} \left\{ \prod_{l=1}^{r_{kj}} r_{kj}\lambda_{kj}(t_{kjl}|F_{kt}) \exp \left( -r_{kj} \int_{0}^{t_{kjl}} \lambda_{kj}(u|F_{kt}) \, du \right) \right\} dG(r_k; \phi) \quad (1.6)
$$

Analogous to the derivation above, we obtain Mixed Poisson Models as well as their overall and marginal likelihood function by letting $\lambda_{kj}(t|F_{kt}) = \lambda_{kj}(t)$. Related estimation approaches have been developed such as Abu-Libdeh et al. (1990), Lawless...
and Nadeau (1995), Ng and Cook (1999) and Chen et al. (2005).

If the covariance or association parameters are not of interest, modeling multivariate recurrent event can be adapted from the analysis of univariate recurrent event under the working independence assumption. Schaubel and Cai (2004, 2005) developed the estimation and inference for marginal analysis for the Cox type model and H Scheike (2002) formulated a similar robust approach for the additive. Both of their work did not incorporate the association structure.

1.4 Study of Associations

Association measurement such as Kendall’s tau (Oakes, 1989), the correlation coefficient (Clayton, 1978), Cross Ratio (Anderson et al., 1992) and Odds Ratio (Scheike, 2012) are designed for Life Time data. These methods only considered first occurrence of each event type and are not suitable for censored recurrent event data. Most recently (Ning et al., 2015) proposed a time-dependent measure, termed the rate ratio as

\[ \rho(s, t) = \frac{\lambda_{1|2}(s|t)}{\lambda_1(s)}, \quad s, t \geq 0, \]  

(1.7)

where the conditional rate function is defined as

\[ \lambda_{1|2}(s|t) = \lim_{\Delta \to 0^+} \Pr\{N_1(s + \Delta) - N_1(s) > 0|N_2(t + \Delta) - N_2(t) > 0\}/\Delta \]  

(1.8)

to assess the local dependence between two types of recurrent event processes. A composite likelihood procedure was developed for model fitting and estimation. However, the composite likelihood based method lacks clear interpretation and is hard to con-
It is not clear how the method can be extended to a regression model of recurrent event processes for multiple types of events when the covariates are present. Here, we develop an alternative approach to model the rate ratio parametrically by a score function and provide a model checking procedure to test the parametric form of the rate ratio.
CHAPTER 2: CONDITIONAL RATE RATIO AS ASSOCIATION MEASURE FOR MULTIVARIATE RECURRENT EVENT PROCESSES

2.1 Preliminaries

Let \( N_{kj}^* \) be a counting process registering the number of event occurrences by time \( t \) for the \( j \)th subject in cluster \( k \) (or equivalently the type \( j \) event for subject \( k \)), for \( j = 1, 2 \) and \( k = 1, ..., N \). Suppose \((N_{k1}^*(s), N_{k2}^*(t)) \) are i.i.d. and let \( Z_{kj}(s), Z_{kj}(t) \) represents the associated covariate vector.

The event times for subjects within a cluster, which would be a family or a clinical center, or the sequentially observed times for a subject, are naturally correlated. Therefore we did not put any restriction here. The goal of this project is to characterize and model the association between the occurrences of events.

The marginal conditional rate function for \( N_{kj}^*(t) \) is defined by

\[
\mu_j(t | z_{kj}) = \lim_{dt \to 0^+} \frac{dN_{kj}^*(t) \mid Z_{kj} = z_{kj}}{dt}, \quad \text{for} \quad j = 1, 2.
\]

Let \( \mu_{2|1}(s,t; z_{k1}, z_{k2}) = E\{dN_{k2}^*(t) = 1 \mid dN_{k1}^*(s) = 1, Z_{k1} = z_{k1}, Z_{k2} = z_{k2}\} \). The conditional rate ratio is defined as

\[
\rho(s,t; z_1, z_2) = \frac{\mu_{2|1}(s,t; z_{k1}, z_{k2})}{\mu_2(t; z_{k2})}, \quad \text{for} \quad s, t \geq 0, \quad (2.1)
\]

which is a measure of how the occurrence of an event for subject 1 (or type 1 event) at time \( s \) modifies the likelihood of event occurrence for subject 2 in the same cluster.
(or type 2 event of the same subject) at time $t$. It is natural to see that \( \rho(s, t; z_{k1}, z_{k2}) \) measures the dependence of \( \{N_{k1}^*(\cdot), N_{k2}^*(\cdot)\} \) at time \((s, t)\). If the two processes are independent then \( \rho(s, t; z_{k1}, z_{k2}) = 1 \).

Under the definition of rate ratio,

\[
E\{dN_{k1}^*(s) dN_{k2}^*(t) \mid Z_{k1} = z_{k1}, Z_{k2} = z_{k2}\} = \rho(s, t; z_{k1}, z_{k2}) \mu_1(s; z_{k1}) \mu_2(t; z_{k2}) dsdt,
\]

(2.2)

where the marginal conditional rates \( \mu_1(t; z_{k1}) \) and \( \mu_2(t; z_{k2}) \) can be modeled, for example, by the semiparametric models such as the additive model of H Scheike (2002) and the multiplicative models of Lin et al. (2000). The association measure \( \rho(s, t\mid z_{i1}, z_{i2}) \) can be modeled through parametric or semiparametric models. Consequently, a two-stage estimating procedure can be adopted.

2.2 Estimation and Inference Procedures

Let \( Y_{kj}(t) = I(C_{kj} \geq t) \) be the at-risk process and \( N_{kj}(t) = \int_0^t Y_{kj}(u) dN_{kj}^*(u) \) be the observed recurrent process. Let \( \hat{\mu}_1(s; z_{k1}) \) and \( \hat{\mu}_2(t; z_{k2}) \) be the estimates of the marginal rates \( \mu_1(s; z_{k1}) \) and \( \mu_2(t; z_{k2}) \), respectively, which is considered as the first-stage estimation. There are a number of options to estimate the conditional rate ratio \( \rho(s, t; z_{i1}, z_{i2}) \) including nonparametric, parametric and semiparametric approaches, each with commonly known strengths and weaknesses. The nonparametric approach may suffer from the \textit{curse-of-dimensionality} while the parametric models can be misspecified. On the other hand, the association measure based on parametric models can be more interpretable.

Suppose that \( \rho(s, t, \theta; z_{i1}, z_{i2}), \theta \in \Theta, \) is a parametric model for \( \rho(s, t; z_{i1}, z_{i2}) \), where
\( \Theta \) is a dimensional compact set. The estimating equation for \( \theta \) can be constructed as

\[
U(\theta, \hat{\mu}_1(\cdot; z_{k1}), \hat{\mu}_2(\cdot; z_{k2})) = \sum_{k=1}^{N} \int_{0}^{T} \int_{0}^{T} \frac{\partial \rho(s,t,\theta; z_{k1}, z_{k2})}{\partial \theta} \left\{ dN_{k1}(s) dN_{k2}(t) \right\} - \rho(s, t, \theta; z_{k1}, z_{k2}) Y_{k1}(s) \hat{\mu}_1(s; z_{k1}) Y_{k2}(t) \hat{\mu}_2(t; z_{k2}) \ ds \ dt \}
\]

(2.3)

The model checking is an essential part of the parametric approach. We proposed a goodness-of-fit procedure to test the parametric form of the rate ratio base on the supremum test statistic given by

\[
T = \sup_{s,t \in [0, \tau]^2} \| V(s, t, \hat{\theta}, \hat{\mu}_1(\cdot; z_{k1}), \hat{\mu}_2(\cdot; z_{k2})) \|
\]

where

\[
V(s, t, \hat{\theta}, \hat{\mu}_1(\cdot; z_{k1}), \hat{\mu}_2(\cdot; z_{k2})) = N^{-1/2} \sum_{k=1}^{N} \int_{0}^{T} \int_{0}^{T} W_n(u, v) \frac{\partial \rho(u, v, \theta; z_{k1}, z_{k2})}{\partial \theta} \left\{ dN_{k1}(u) dN_{k2}(v) \right\} - \rho(u, v, \theta; z_{k1}, z_{k2}) Y_{k1}(u) \hat{\mu}_1(u; z_{k1}) Y_{k2}(v) \hat{\mu}_2(v; z_{k2}) \ du \ dv \}
\]

(2.4)

\( W_n(u, v) \) is prespecified weight function and \( \| \cdot \| \) is the Euclidean norm. The critical values can be approximated by implementing the Gaussian multiplier method (cf. Sun, Li and Gilbert (2016*)).
CHAPTER 3: ESTIMATION AND INFERENCE OF THE RATE RATIO UNDER
THE ADDITIVE MARGINAL MODEL

3.1 Estimation by a two-stage approach

We illustrate the two-stage approach described in Chapter 2 when the marginal
conditional rate model is additive. Let \( N_{kj}^*(t) \) follows the additive rates model

\[
E[dN_{kj}^*(t)|Z_{kj}(t)] = d\mu_j(t|Z_{kj}(t)),
\]
\[
d\mu_j(t|Z_{kj}) = d\mu_{0j}(t) + \beta_j^T Z_{kj}(t) \, dt, \quad k = 1, ..., N; j = 1, 2 \quad (3.1)
\]

where \( \mu_{0j}(t) \) is an unspecified baseline rate function and \( \beta_j \) an unknown \( p \)-dimensional
vector. We consider the parametric approach by assuming \( \rho(s, t, \theta; z_{k1}, z_{k2}) \), where \( \theta \)
is the \( q \)-dimensional parameter of interest.

In the following sections, we first review the estimation procedure of \( \beta_j \) and \( \mu_{0j}(t) \)
from the additive marginal mean rate model by adapting the method proposed by
H Scheike (2002). Then we develop the estimation procedures for parametric rate ratio
and investigate its asymptotic properties. A goodness-of-fit procedure is also proposed
to test the parametric assumption of the rate ratio. Lastly, we conduct simulations
to validate the estimation and inference procedures, with the results presented at the
end of this chapter.
3.1.1 Review of the estimation of the marginal model

We define a zero-mean stochastic process as

$$M_{kj}(t, \beta_j) = N_{kj}(t) - \int_0^t Y_{kj}(u)\{d\mu_{0j}(u) + \beta_j^T Z_{kj}(u) du\}. \quad (3.2)$$

Following the Generalized Estimating Equations proposed by (GEE; Liang and Zeger 1986), the estimating functions for $\mu_{0j}(t)$ and $\beta_j$ are as

$$\sum_{k=1}^N \int_0^t Y_{kj}(u) dM_{kj}(u; \beta_j) = 0, \quad 0 \leq t \leq \tau. \quad (3.3)$$

$$\sum_{k=1}^N \int_0^\tau Y_{kj}(u) Z_{kj}(u) dM_{kj}(u; \beta_j) = 0. \quad (3.4)$$

respectively. By solving (3.3), we obtain the $\hat{\mu}_{0j}(t; \beta_j)$ as an estimate of $\mu_{0j}(t)$, where

$$\hat{\mu}_{0j}(t; \beta_j) = \int_0^t \frac{\sum_{k=1}^N [dN_{kj}(u) - Y_{kj}(u)\beta_j^T Z_{kj}(u) du]}{\sum_{k=1}^N Y_{kj}(u)}. \quad (3.5)$$

With some simple algebra, equation (3.4) is equivalent to

$$L_j(\beta_j) = \sum_{k=1}^N \int_0^\tau \{Z_{kj}(u) - \bar{Z}_j(u)\} \left[ dN_{kj}(u) - Y_{kj}(u)\beta_j^T Z_{kj}(u) du \right],$$

where $\bar{Z}_j(t) = \frac{\sum_{k=1}^N Z_{kj}(t)Y_{kj}(t)}{\sum_{k=1}^N Y_{kj}(t)}$. Substituting $\hat{\mu}_{0j}(t; \beta_j)$ into equation (3.2) and solve equation (3.4) gives us the estimate of $\beta_j$ as

$$\hat{\beta}_j = \left[ \sum_{k=1}^N \int_0^\tau Y_{kj}(u)\{Z_{kj}(u) - \bar{Z}_j(u)\} \otimes^2 du \right]^{-1} \sum_{k=1}^N \int_0^\tau \{Z_{kj}(u) - \bar{Z}_j(u)\} dN_{kj}(u), \quad (3.6)$$

where $a \otimes^2 = aa^T$ for a vector $a$. Once $\hat{\beta}_j$ is obtained, $\mu_{0j}(t)$ can be estimated by $\hat{\mu}_{0j}(t; \hat{\beta}_j)$ from equation (3.5).
For convenience we summarize the estimation method of the additive marginal model developed by H Scheike (2002) here.

**Theorem 3.1 (H Scheike (2002) Theorem 1)** Under the regularity (C.1.)-(C.5.),  
\( \hat{\beta}_j \) converges almost surely to \( \beta_j \), and has the following asymptotic approximation  
\[
\sqrt{N}\{ \hat{\beta}_j - \beta_j \} = A_j^{-1/2} \sum_{k=1}^N \xi_{kj} + o_p(1)
\]
where  
\[
\xi_{kj} = \int_0^\tau \{ Z_{kj}(u) - \bar{Z}_j(u) \} \, dM_{kj}(u, \hat{\beta}_j) \quad \text{and} \quad \bar{Z}_j(t) = \lim_{N \to \infty} \bar{Z}_j(t).
\]
\( \sqrt{N}(\hat{\beta}_j-\beta_j) \) is asymptotically normal with mean zero and covariance matrix  
\[
A_j^{-1} \Sigma_j A_j^{-1},
\]
where  
\[
A_j = E\{ \int_0^\tau \{ Z_{kj}(u) - \bar{Z}_j(u) \}^\otimes 2 \, ds \},
\]
\[
\Sigma_j = E[\int_0^\tau \{ Z_{1j}(u) - \bar{Z}_j(u) \} \, dM_{1j}(u, \beta_j) \int_0^\tau \{ Z_{1j}(v) - \bar{Z}_j(v) \} \, dM_{1j}(v, \beta_j)].
\]
The asymptotic covariance matrix can by consistently estimated by  
\[
\hat{A}_j^{-1} \hat{\Sigma}_j \hat{A}_j^{-1},
\]
with  
\[
\hat{A}_j = N^{-1} \sum_{k=1}^N \int_0^\tau \{ Z_{kj}(u) - \bar{Z}_j(u) \}^\otimes 2 \, du,
\]
\[
\hat{\Sigma}_j = N^{-1} \sum_{k=1}^N \hat{\xi}_{kj}^\otimes 2,
\]
\[
\hat{\xi}_{kj} = \int_0^\tau \{ Z_{kj}(u) - \bar{Z}_j(u) \} \, d\hat{M}_{kj}(u; \hat{\beta}_j),
\]
\[
d\hat{M}_{kj}(t; \hat{\beta}_j) = dN_{kj}(t) - Y_{kj}(t) \{ d\hat{\mu}_{0j}(t) + \hat{\beta}_j^T Z_{kj}(t) \, dt \}.
\]

**Theorem 3.2 (H Scheike (2002) Theorem 2)** Under the regularity (C.1)-(C.5),  
\( \hat{\mu}_{0j}(t) \) converges almost surely to \( \mu_{0j}(t) \) uniformly in \( t \in [0, \tau] \).  
\( \sqrt{N}\{ \hat{\mu}_{0j}(t) - \mu_{0j}(t) \} \) converges weakly to a mean-zero Gaussian process with covariance function  
\[
\Gamma_j(s,t) = E[\phi_{kj}(s)\phi_{kj}(t)]
\] (3.7)
where
\[
\phi_{kj}(t) = \int_0^t \pi_j^{-1}(u)dM_{kj}(u; \beta_j) - H^T(t)A_j^{-1}\int_0^t \{Z_{kj}(u) - \bar{z}_j(u)\} dM_{kj}(u; \beta_j),
\] 
(3.8)

with \(H(t) = \int_0^t \bar{z}_j(u)du\), \(\bar{z}^T_j(t) = \lim_{N \to \infty} \bar{Z}^T_j(t)\) and \(\pi_j(t) = N^{-1} \lim_{N \to \infty} \sum_{k=1}^N Y_{kj}(t)\).

The consistent estimates of \(\Gamma(s, t)\) is denoted by \(\hat{\Gamma}_j(s, t) = N^{-1} \sum_{k=1}^N \hat{\phi}_{kj}(s)\hat{\phi}_{kj}(t)\), with \(\hat{\phi}_{kj}(t) = \int_0^t \hat{\pi}_j^{-1}(u)d\hat{M}_{kj}(u; \hat{\beta}_j) - \hat{H}^T(t)\hat{A}_j^{-1}\int_0^t \{\hat{Z}_{kj}(u) - \hat{Z}_j(u)\} d\hat{M}_{kj}(u; \hat{\beta}_j),\hat{\pi}_j(t) = N^{-1} \sum_{k=1}^N Y_{kj}(t)\) and \(\hat{H}(t) = \int_0^t \hat{Z}_j(u)du\).

3.1.2 Estimation of the rate ratio

The rate ratio can be estimated by equation (2.3), the realization of which under model (3.1) is
\[
U(\theta, \hat{\beta}_1, \hat{\beta}_2, \hat{\mu}_0(\cdot), \hat{\mu}_2(\cdot)) = \sum_{k=1}^N U_k(\theta, \hat{\beta}_1, \hat{\beta}_2, \hat{\mu}_0(\cdot), \hat{\mu}_2(\cdot)),
\] 
(3.9)

where
\[
U_k(\theta, \hat{\beta}_1, \hat{\beta}_2, \hat{\mu}_0(\cdot), \hat{\mu}_2(\cdot)) = \int_0^T \int_0^T \frac{\partial \rho(s, t; \theta; Z_{k1}, Z_{k2})}{\partial \theta} \left\{ dN_{k1}(s) dN_{k2}(t) - \rho(s, t; Z_{k1}, Z_{k2})Y_{k1}(s)[d\hat{\mu}_0(1) + \hat{\beta}_1 Z_{k1}(s) ds]Y_{k2}(t)[d\hat{\mu}_2(1) + \hat{\beta}_2 Z_{k2}(1) dt] \right\}.
\]

Denote \(\hat{\theta}\) the solution to \(U(\theta, \hat{\beta}_1, \hat{\beta}_2, \hat{\mu}_0(\cdot), \hat{\mu}_2(\cdot)) = 0\). We investigate the asymptotic properties of \(U(\hat{\theta}, \hat{\beta}_1, \hat{\beta}_2, \hat{\mu}_0(\cdot), \hat{\mu}_2(\cdot))\) and \(\hat{\theta}\) in Theorem 3.3 and 3.4 below.

**Theorem 3.3** \(N^{-1/2}\{U(\theta, \hat{\beta}_1, \hat{\beta}_2, \hat{\mu}_0(\cdot), \hat{\mu}_2(\cdot)) - U(\theta, \beta_1, \beta_2, \mu_0(\cdot), \mu_2(\cdot))\}\) converges to a mean-zero Gaussian process, with covariance
\[
\Omega = \lim_{N \to \infty} N^{-1} \sum_{k=1}^N \left\{ h_{1,N} \xi_{k1} A_1^{-1} + g_{1,N,k} + h_{2,N} \xi_{k2} A_2^{-1} + g_{2,N,k} \right\}^{\otimes 2}.
\]
The consistent estimates of $\Omega$ is

$$\hat{\Omega} = N^{-1} \sum_{k=1}^{N} \left\{ \hat{h}_{1,N} \hat{\xi}_{k1} \hat{A}_1^{-1} + \hat{g}_{1,N,k} + \hat{h}_{2,N} \hat{\xi}_{k2} \hat{A}_2^{-1} + \hat{g}_{2,N,k} \right\} \otimes^2,$$

where $\hat{h}_{j,N}, \hat{\xi}_{kj}, \hat{g}_{j,N,k}(s,t)$ $(j = 1, 2)$ are shown in the appendix.

**Theorem 3.4** $\sqrt{N}(\hat{\theta} - \theta)$ can be approximated by a mean zero Gaussian process

$$\sqrt{N}(\hat{\theta} - \theta) = N^{-1/2} [\mathcal{I}(\theta)]^{-1} \sum_{k=1}^{N} W_k(\theta) + o_p(1), \quad (3.10)$$

for which the formulae for $\mathcal{I}(\theta)$ and $W_k(\theta)$ are given in the appendix.

The variance of $\sqrt{N}(\hat{\theta} - \theta)$ can be estimated by $\hat{\Phi} = N^{-1} (\hat{\mathcal{I}})^{-1} \sum_{k=1}^{N} (\hat{W}_k)^{\otimes 2} (\hat{\mathcal{I}}^T)^{-1}$, where $\hat{\mathcal{I}}$ and $\hat{W}_k$ are the empirical counterparts of $\mathcal{I}(\theta)$ and $W_k(\theta)$.

### 3.1.3 Simulation studies

Before we conduct finite sample studies to investigate performance of the proposed estimation procedure, we want to show some examples that motivate us to model the rate ratio parametrically.

**Proposition 1** Under shared frailty model

$$d\mu_j(t) = R_k \cdot \{ d\mu_0(t) + \beta_j^T Z_{kj}(t) dt \}, \quad (3.11)$$

where $R_k$ is identically and independently distributed positive random variable, with $E(R_k) = \mu$ and $\text{var}(R_k) = \sigma^2$. The rate ratio only depends on the variance of frailty random variable and can be explicitly expressed as

$$\rho(s,t,\theta) = \rho = 1 + \frac{\sigma^2}{\mu^2}, \quad (3.12)$$
Proposition 2  Let $\tau$ be the maximum observation time and $c_0$ lies in the middle of $0$ and $\tau$. Suppose the shared frailty mean rate model for $N^*_{kj}(t)$ is

$$
d\mu_j(t|Z_{kj}(t), R_k(t)) = R_k(t)\{d\mu_{0j}(t) + \beta^T_j Z_{kj}(t) dt\}
$$

where $R_k(t) = I(t \leq c_0)R_{k0} + I(t > c_0)R_{k1}$.

Before we exam the rate ratio in this time varying additive mean rate model, we introduce the shifted gamma distribution. Define the probability density function of the shifted Gamma($a, b, \delta$) as

$$
f(x|a, b, \delta) = \frac{1}{\Gamma(a)b^a}(x - \delta)^{a-1}e^{-(x-\delta)/b}, x \in [\delta, \infty), \delta \geq 0
$$

for $x \in [\delta, \infty)$, $\delta \geq 0$ and here $\Gamma(\cdot)$ denotes the Gamma function. Let $X$ come from shifted Gamma($a, b, \delta$) then we have $E(X) = a \cdot b + \delta$ and $\text{var}(X) = a \cdot b^2$. As we can see when $\delta = 0$, the shifted Gamma distribution is reduced to the gamma distribution.

If $R_{k0}$ and $R_{k1}$ are independently from the corresponding shifted gamma distribution ($a_0, b_0, \delta_0$) and ($a_1, b_1, \delta_1$), then the rate ratio is piecewise constant:

$$
\rho(\theta, s \leq c_0, t \leq c_0) = 1 + \frac{a_0b_0^2}{(a_0b_0 + \delta_0)^2},
$$

$$
\rho(\theta, s > c_0, t > c_0) = 1 + \frac{a_1b_1^2}{(a_1b_1 + \delta_1)^2},
$$

$$
\rho(\theta, s \leq c_0, t > c_0) = \rho(\theta, s > c_0, t \leq c_0) = 1.
$$

Proposition 3  For $j = 1, 2$, denote $\tilde{\lambda}_j(t|z_j)$ the event rate of nonhomogeneous Pos-
sion Process \( \tilde{N}_j(t) \). Let \( N_0(t) \) be a nonhomogeneous Poisson process with event rate \( \lambda_0(t|z_j) \). Assume that \( \tilde{N}_j(t) \) and \( N_0(t) \) be mutually independent, i.e. for any \( u_1, u_2, ..., u_n \), the random vectors \( \{\tilde{N}_1(u_1), \tilde{N}_1(u_1), ..., \tilde{N}_1(u_n)\}, \{\tilde{N}_2(u_1), ..., \tilde{N}_2(u_n)\} \) and \( \{N_0(u_1), ..., N_0(u_n)\} \) are independent to each other.

Define the counting process \( N_j(t) \) as \( N_j(t) = \tilde{N}_j(t) + N_0(t) \) for \( j = 1, 2 \). Since \( N_j(t) \) is the summation of two independent Poisson processes, \( N_j(t) \) is also a Poisson process with rate \( \lambda_j(t|z_j) = \tilde{\lambda}_j(t|z_j) + \lambda_0(t|z_j) \).

Let \( \rho_0(s, t, \theta|z_1, z_2) \) and the \( \rho(s, t, \theta|z_1, z_2) \) be the rate ratio of \( \{N_0(s), N_0(t)\} \) and \( \{N_1(s), N_2(t)\} \) for \( s, t \geq 0 \), then we have \( \rho(\theta, s, t|z_1, z_2) \)

\[
\rho(\theta, s, t|z_1, z_2) = 1 + \frac{\{\rho_0(\theta, s, t|z_1, z_2) - 1\}\lambda_0(s|z_1)\lambda_0(t|z_2)}{\lambda_1(s|z_1)\lambda_2(t|z_2)}. \tag{3.16}
\]

The association is introduced by the shared counting process \( N_0(s) \) and \( N_0(t) \). If \( \rho_0(\theta, s, t|z_1, z_2) = 1, \rho(\theta, s, t|z_1, z_2) = 1 \), thus if \( \{N_0(s), N_0(t)\} \) is independent so is \( \{N_1(s), N_2(t)\} \).

We conduct simulation studies to evaluate the finite sample properties based on the guidance of Proposition 1, 2 and 3. Let \( \tau = 5 \), \( C_{kj} \) follows a uniform distribution on \([0, \tau]\), and covariates \( Z_{kj} \) are from a uniform\([1, 2]\) for \( j = 1, 2 \). The observed events for the \( j \)th type in cluster \( k \) would be all the event times that are smaller than \( C_{kj} \).

We consider I, II, III scenarios where the rate ratio is constant, time varying, and covariate dependent. Scenario IV is an extension from II and III, with the rate ratio depending on event time and covariates.
\[(I) \textbf{Constant } \rho(s, t, \theta) = \theta_0\]

Recall the shared frailty model in equation (3.11)

\[d\mu_j(t | R_k, Z_{kj}(t)) = R_k \cdot \{ d\mu_{0j}(t) + \beta_j Z_{kj}(t) \} \quad \text{for } j = 1, 2.\]

Let \( R_k \) follows i.i.d Gamma\((a, b)\) with \( E(R_k) = ab \) and \( \text{var}(R_k) = ab^2 \). By proposition 1, \( \rho(s, t, \theta) = \theta_0 \) where \( \theta_0 = 1 + ab^2/(ab)^2 = 1 + 1/a \).

Let \( \beta_1 = 0.5, \beta_2 = 1, \mu_{01}(t) = \mu_{02}(t) = 0.25t, 0.5t, t \). The averaged observed type 1(2) events after right censoring are 2.50(4.37), 3.13(5.02) and 4.37(6.26) respectively. To variate the strength of the association, we take \( R_k \) from the pairs of \((a, b)\) equal to \((4, 0.25), (2, 0.5), (1.33, 0.75)\) and \((1, 1)\) so that \( \theta_0 = 1.25, 1.5, 1.75 \) and 2 correspondingly.

By taking the expectation of \( R_k \) in equation (3.11), the mean event rate still follows model (3.1). In the first-stage, \( \hat{\beta}_j, \hat{\mu}_{0j}(t) \) are evaluated by equation (3.6) and (3.5). In general, the estimates of \( \beta_{0j} \) and \( \mu_{0j}(t) \) agree with the discussions in literatures. We show part of the numerical results for the first-stage estimates in Table 1, from which it is observed that \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) converges to the true values \( \beta_1 = 0.5 \) and \( \beta_2 = 1 \). The mean Estimated Standard Error of \( \beta_j \) (ESE) is very close to the Sample Standard Error of Estimates (SSE) and Empirical Coverage Probability (CP) is around to 0.95. We will skip the marginal model simulation result and focus on the estimation of the parameters in the rate ratio in the studies.

In the second-stage, \( \hat{\beta}_j, \hat{\mu}_{0j} \) for \( j = 1, 2 \) are plugged into equation (3.9) and the root is derived by the Newton-Raphson method. Convergence is achieved at the \( i \)th iteration if \( \frac{\theta^{(i)} - \theta^{(i-1)}}{\theta^{(i-1)}} < 10^{-5} \) or \( i > 50 \). In Table 2, the Bias is negligible for
all the cases and the Standard Error of Estimates (SEE) is close to the Estimated Standard Error (ESE). The 95% coverage probability (CP) is also around 0.95. Both SEE and ESE decrease with a larger sample size. It is also observed that the SEE and ESE increase when the association between the two processes becomes stronger (i.e. $\theta_0$ is larger) and such increment is slowly reduced by increasing the sample size. A possible interpretation is that for bivariate recurrent event processes, given the observed dataset with a fixed sample size, less information would be obtained if the two events are highly related. We might be able to adapt a weight function in the estimation equation (3.9) to improve the efficiency of this estimating procedure.

(II) Time Dependent Rate Ratio $\rho(\theta, s, t)$

For the $j$th individual in the $k$th cluster, let

$$N_{kj}(t) = \tilde{N}_{kj}(t) + N_{k0}(t), \quad \text{for } j = 1, 2$$

(3.17)

where $\{\tilde{N}_{k1}(\cdot), \tilde{N}_{k2}(\cdot), N_{k0}(\cdot)\}$ are independent Poisson process, conditional on covariates and frailty. Consider $E\{dN_{k0}(t) | z_j, R_k\} = R_k \cdot \lambda_{k0}(t | Z_{kj}(t) = z_j)dt$, where $\lambda_{k0}(t | Z_{kj}) dt = d\mu_0(t) + \beta_0 Z_{kj}(t)$, and $R_k$ is the frailty and variable is from a positive i.i.d Distribution. Let $E(R_k) = \mu_0$ and $\text{var}(R_k) = \sigma_0^2$ the rate ratio of $\{N_{k0}(s), N_{k0}(t)\}$ can be obtained from Proposition 1 as

$$\rho_0(s, t, \theta | z_1, z_2) = 1 + \sigma_0^2/\mu_0^2.$$  

Denote the mean rate for $\tilde{N}_{kj}(t)$ as $\tilde{\lambda}_{kj}(t | Z_{kj}(t) = z_{kj})$ (for $j = 1, 2$ and $t \in (0, \tau)$) and assume $\tilde{\lambda}_{kj}(t | Z_{kj}(t) = z_{kj}) = m_j(t) \lambda_{k0}(t | Z_{kj}(t) = z_{kj})$, with $m_j(u) \geq 0$. By equation (3.17), the mean rate of $N_{kj}(t)$ is $\lambda_{kj}(t | z_{kj}) = [1 + m_j(t)] \lambda_{k0}(t | z_{kj})$. Intuitively, it suggests that the mean rate of $\tilde{N}_{kj}(t)$ is proportional to that of the underline common counting process $N_{k0}(t)$. Especially, $m_j(t) = 0$ makes $N_{kj}(t)$ is
reduced to the shared frailty in equation (3.1).

Following Proposition 3, the rate ratio of \( \{N_{k1}(s), N_{k2}(t)\} \) can be expressed as

\[
\rho(\theta, s, t) = 1 + \theta_0 \times \frac{1}{(1 + m_1(t))(1 + m_2(s))},
\]

(3.18)

where \( \theta_0 = \frac{\sigma^2_0}{\mu_0^2} \). Let \( 1/(1 + m_1(t)) = -0.15t + 0.9 \), \( 1/(1 + m_2(s)) = -0.15s + 0.9 \), by equation (3.18) we obtain

\[
\rho(\theta, s, t) = 1 + \theta_0 \times (-0.15t + 0.9)(-0.15s + 0.9).
\]

(3.19)

Let \( R_k \) are i.i.d Gamma\((a, b)\) so that \( \mu_0 = ab \) and \( \sigma_0 = ab^2 \). We take \((a, b)\) as \((4, 0.25), (2, 0.5), (1, 1) \) and \((0.635, 1.6)\) and therefore the corresponding \( \theta_0 \) are 0.25, 0.5, 1 and 1.6. To generate moderate and frequent event observations, we take \( \beta_{01} = \beta_{02} = 0 \) and set \( \mu_{01}(t) = \mu_{02}(t) \) to be \( 0.25t, 0.5t, 0.75t \) and \( t \), which gives us averaged events count as 2.13, 4.17, 5.21 and 6.39 respectively.

The Bias of the estimates (Bias), the Estimated Standard Error (ESE), the Sample Standard Error of Estimates (SSE) and 95% Empirical Coverage Probability (CP) are calculated from 1000 simulated datasets with sample size \( N = 200, 500, 800 \). The bias of \( \theta_0 \) is low, the ESE is close to the SSE and the coverage probability is around 0.95. When the rate ratio of \( N_{k1}(\cdot) \) and \( N_{k2}(\cdot) \) become stronger, the ESE and SSE both increase, which is similar to the scenario I. For details, see Table 3.

(III) Covariate Dependent Rate Ratio \( \rho(\theta; Z_k) = \theta_1 I(Z_k = 1) + \theta_2 I(Z_k = 0) \)

Let \( Z_k \) be a cluster level binary covariate. Assume the counting process \( N_{kj}^*(t) \)
follows the shared frailty model

\[ E[dN^*_k(t) | Z_k, R_k] = R_k \{ d\mu_{0j}(t) + \beta_j Z_k(t) \, dt \}, \tag{3.20} \]

where \( E[R_k | Z_k] = \mu(Z_k) \) and \( \text{var}[R_k | Z_k] = \sigma^2(Z_k) \). Following Proposition 1, we obtain

\[ \rho(\theta; Z_k) = 1 + \frac{\sigma^2(Z_k)}{\mu^2(Z_k)}. \tag{3.21} \]

We take \( \beta_1 = 0.5, \beta_2 = 1, \mu_{01}(t) = 0.25t, 0.5t, 0.75t \). Let \( Z_k \) come from \( \text{Bernoulli}(p = 0.5) \), so that \( X_k \) has equal chance to be 0 or 1. We generate \( R_k \) from \( \text{Gamma}(4, 0.25) \) and \( \text{Gamma}(1.33, 0.75) \) for \( Z_k = 1 \) and \( Z_k = 0 \) respectively.

In equation (3.21), \( \rho(\theta; Z_k = 1) = 1.25, \rho(\theta; Z_k = 0) = 1.75 \) and therefore we rewrite the rate ratio as

\[ \rho(\theta; Z_k) = \theta_1 I(Z_k = 1) + \theta_2 I(Z_k = 0), \tag{3.22} \]

with \( \theta_1 = 1.25 \) and \( \theta_2 = 1.75 \). Under this setting, the averaged observed type 1(2) events after right censoring are 2.50(4.37), 3.13(5.02) and 4.37(6.26).

\textbf{(IV) Time and Covariate Dependent Rate Ratio}

Consider the bivariate counting processes \{\( N_{k1}(\cdot), N_{k2}(\cdot) \)} constructed by the summation of two independent Poisson processes \( \tilde{N}_{kj}(\cdot) \) and \( N_{k0}(\cdot) \), as described in Proposition 3. Denote \( \rho_0(\theta, s, t | z_1, z_2) \) and \( \rho(\theta, s, t | z_1, z_2) \) be the rate ratio of \( (N_{k0}(t), N_{k0}(s)) \) and \( (\{N_{k1}(s), N_{k2}(t)\}) \) respectively. Following from Proposition 3, we have

\[ \rho(\theta, s, t | z_1, z_2) = 1 + \frac{\{\rho_0(\theta, s, t | z_1, z_2) - 1\} \lambda_0(s | z_1) \lambda_0(t | z_2)}{\lambda_1(s | z_1) \lambda_2(t | z_2)}, \]

where \( \lambda_{k0}(s | z_1) \, ds, \lambda_j(s | z_1) \, ds \) are the conditional mean rate of \( N_{k0}(s) \) and \( N_{k1}(s) \),
whereas \( \lambda_{k0}(t|z_2) \, ds \), \( \lambda_s(t|z_1) \, dt \) are that of \( N_{k0}(t) \) and \( N_{k2}(t) \).

Let \( \lambda_{k0}(t|Z_k, R_k) = R_k(0.25 + \beta_{0j} Z_k) \) and \( \tilde{\lambda}_{kj}(t) = 0.5t \), where \( R_k \) is generated from i.i.d Gamma\((a, b)\) and \( Z_k \) is from Bernoulli\((0.5)\). Consider \((a, b)\) equal to \((4, 0.25)\), \((2, 0.5)\) and \((1.33, 0.75)\) such that \( \rho_0(\theta, s, t|z_1, z_2) = 1.25, 1.5 \) and \( 1.75 \). Let \( \beta_{01} = 0.1, \beta_{02} = 0.2 \). The rate ratio of \( N_{k1}(s) \) and \( N_{k2}(t) \) is time-varying and dependent on the covariate \( Z_{kj} \), where

\[
\rho(\theta, s, t|Z_k) = 1 + \theta \frac{(0.25 + 0.1Z_k)(0.25 + 0.2Z_k)}{(0.5t + 0.25 + 0.1Z_k)(0.5s + 0.25 + 0.2Z_k)}. \tag{3.23}
\]

with \( \theta = \frac{\sigma^2}{\mu^2} = 0.25, 0.5, 0.75 \) and 1.

To evaluate the influence of observed event frequency on the estimating procedure, we modified \( \lambda_{k0}(t|Z_k, R_k) = R_k(0.5 + \beta_{0j} Z_k) \) and kept all the other settings so that

\[
\rho(\theta, s, t|Z_k) = 1 + \theta \frac{(0.5 + 0.1Z_k)(0.5 + 0.2Z_k)}{(0.5t + 0.5 + 0.1Z_k)(0.5s + 0.5 + 0.2Z_k)}. \tag{3.24}
\]

1000 datasets are generated from the above settings. With the estimated \( \beta_{0j} \) and \( \mu_{0j}(t) \) plugged into equation (3.23), the estimates of \( \sigma^2/\mu^2 \) can be computed. The simulation result is summarized in Table 5. The bias is going to zero and the ESE is getting close to SSE as sample size increase. The coverage probability is getting around 95% for both \( \theta \).
Table 1: Scenario I - $\rho(s,t,\theta) = \theta_0$. Estimation of coefficients in the marginal additive model. The Bias, SEE(Standard Error of Estimates), ESE (Estimated Standard Errors) and the Empirical Coverage Probability of 95% confidence interval (CP) of $(\beta_{01}, \beta_{02})$. Each entry is based on 1000 simulations.

<table>
<thead>
<tr>
<th>$\mu_{0j}(t)$</th>
<th>$\theta_0$</th>
<th>N</th>
<th>Bias($\beta_{01}, \beta_{02}$)</th>
<th>SEE($\beta_{01}, \beta_{02}$)</th>
<th>ESE($\beta_{01}, \beta_{02}$)</th>
<th>CP($\beta_{01}, \beta_{02}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.75t</td>
<td>1</td>
<td>200</td>
<td>(0.0097, 0.0127)</td>
<td>(0.1871, 0.2223)</td>
<td>(0.1902, 0.2335)</td>
<td>(0.9560, 0.9590)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(0.0076, -0.0051)</td>
<td>(0.1183, 0.1464)</td>
<td>(0.1202, 0.1466)</td>
<td>(0.9500, 0.9500)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(-0.0001, -0.0028)</td>
<td>(0.0944, 0.1159)</td>
<td>(0.0947, 0.1163)</td>
<td>(0.9540, 0.9460)</td>
</tr>
<tr>
<td>1.25</td>
<td>200</td>
<td>(0.0126, 0.0145)</td>
<td>(0.2938, 0.3876)</td>
<td>(0.2830, 0.3936)</td>
<td>(0.9540, 0.9580)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(-0.0005, 0.0057)</td>
<td>(0.1746, 0.2579)</td>
<td>(0.1803, 0.2497)</td>
<td>(0.9600, 0.9470)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(0.0040, 0.0025)</td>
<td>(0.1478, 0.1976)</td>
<td>(0.1428, 0.1984)</td>
<td>(0.9450, 0.9540)</td>
</tr>
<tr>
<td>1.5</td>
<td>200</td>
<td>(0.0106, 0.0110)</td>
<td>(0.3490, 0.5128)</td>
<td>(0.3524, 0.5073)</td>
<td>(0.9560, 0.9440)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(0.0109, 0.0108)</td>
<td>(0.2337, 0.3232)</td>
<td>(0.2244, 0.3217)</td>
<td>(0.9440, 0.9610)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(0.0039, 0.0254)</td>
<td>(0.1826, 0.2631)</td>
<td>(0.1773, 0.2550)</td>
<td>(0.9470, 0.9460)</td>
</tr>
<tr>
<td>1.75</td>
<td>200</td>
<td>(0.0027, -0.0388)</td>
<td>(0.4227, 0.6018)</td>
<td>(0.4096, 0.5902)</td>
<td>(0.9540, 0.9490)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(0.0067, -0.0002)</td>
<td>(0.2623, 0.3876)</td>
<td>(0.2597, 0.3790)</td>
<td>(0.9570, 0.9480)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(-0.0019, 0.0056)</td>
<td>(0.2042, 0.2906)</td>
<td>(0.2072, 0.3024)</td>
<td>(0.9470, 0.9580)</td>
</tr>
<tr>
<td>t</td>
<td>1</td>
<td>200</td>
<td>(0.0091, -0.0117)</td>
<td>(0.2041, 0.2467)</td>
<td>(0.2048, 0.2433)</td>
<td>(0.9500, 0.9580)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(-0.0087, -0.0027)</td>
<td>(0.1265, 0.1538)</td>
<td>(0.1293, 0.1545)</td>
<td>(0.9480, 0.9530)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(0.0050, -0.0011)</td>
<td>(0.1009, 0.1246)</td>
<td>(0.1022, 0.1223)</td>
<td>(0.9500, 0.9600)</td>
</tr>
<tr>
<td>1.25</td>
<td>200</td>
<td>(-0.0110, -0.0028)</td>
<td>(0.3289, 0.4274)</td>
<td>(0.3202, 0.4282)</td>
<td>(0.9410, 0.9560)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(-0.0039, -0.0064)</td>
<td>(0.2022, 0.2755)</td>
<td>(0.2037, 0.2730)</td>
<td>(0.9460, 0.9550)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(-0.0013, 0.0005)</td>
<td>(0.1656, 0.2209)</td>
<td>(0.1611, 0.2164)</td>
<td>(0.9460, 0.9530)</td>
</tr>
<tr>
<td>1.5</td>
<td>200</td>
<td>(0.0295, -0.0237)</td>
<td>(0.4212, 0.5470)</td>
<td>(0.4010, 0.5476)</td>
<td>(0.9520, 0.9530)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(0.0150, 0.0041)</td>
<td>(0.2649, 0.3599)</td>
<td>(0.2566, 0.3519)</td>
<td>(0.9510, 0.9450)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(-0.0039, -0.0123)</td>
<td>(0.2067, 0.2745)</td>
<td>(0.2029, 0.2793)</td>
<td>(0.9540, 0.9580)</td>
</tr>
<tr>
<td>1.75</td>
<td>200</td>
<td>(0.0181, 0.0012)</td>
<td>(0.4745, 0.6696)</td>
<td>(0.4692, 0.6541)</td>
<td>(0.9480, 0.9500)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(0.0093, 0.0208)</td>
<td>(0.3089, 0.4311)</td>
<td>(0.3007, 0.4203)</td>
<td>(0.9500, 0.9450)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(0.0005, -0.0082)</td>
<td>(0.2418, 0.3357)</td>
<td>(0.2381, 0.3333)</td>
<td>(0.9570, 0.9530)</td>
</tr>
</tbody>
</table>
Table 2: Scenario I - Estimation of $\rho(s,t,\theta) = \theta_0$. Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP (95% Coverage Probability) lists. Each entry is based on 1000 simulated datasets. The marginal models are additive and association come from the shared random effect.

<table>
<thead>
<tr>
<th>$\mu_{0j}(t)$</th>
<th>$\theta_0$</th>
<th>$N$</th>
<th>Bias</th>
<th>SEE</th>
<th>ESE</th>
<th>CP</th>
<th>$\mu_{0j}(t)$</th>
<th>$\theta_0$</th>
<th>$N$</th>
<th>Bias</th>
<th>SEE</th>
<th>ESE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25t</td>
<td>1.25</td>
<td>200</td>
<td>-0.0031</td>
<td>0.0845</td>
<td>0.0818</td>
<td>0.9350</td>
<td>0.75t</td>
<td>1.25</td>
<td>200</td>
<td>-0.0011</td>
<td>0.0770</td>
<td>0.0726</td>
<td>0.9330</td>
</tr>
<tr>
<td>0.25t</td>
<td>3.75</td>
<td>500</td>
<td>0.0005</td>
<td>0.0567</td>
<td>0.0536</td>
<td>0.9420</td>
<td>0.75t</td>
<td>3.75</td>
<td>500</td>
<td>0.0011</td>
<td>0.0520</td>
<td>0.0476</td>
<td>0.9300</td>
</tr>
<tr>
<td>0.25t</td>
<td>6.25</td>
<td>800</td>
<td>0.0001</td>
<td>0.0432</td>
<td>0.0428</td>
<td>0.9520</td>
<td>0.75t</td>
<td>6.25</td>
<td>800</td>
<td>-0.0001</td>
<td>0.0395</td>
<td>0.0385</td>
<td>0.9420</td>
</tr>
<tr>
<td>1.50</td>
<td>200</td>
<td>0.0131</td>
<td>0.1392</td>
<td>0.1248</td>
<td>0.8960</td>
<td>1.50</td>
<td>200</td>
<td>0.0079</td>
<td>0.1280</td>
<td>0.1179</td>
<td>0.8970</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.50</td>
<td>500</td>
<td>0.0024</td>
<td>0.0896</td>
<td>0.0849</td>
<td>0.9200</td>
<td>1.50</td>
<td>500</td>
<td>0.0025</td>
<td>0.0851</td>
<td>0.0791</td>
<td>0.9290</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.50</td>
<td>800</td>
<td>0.0045</td>
<td>0.0672</td>
<td>0.068</td>
<td>0.9500</td>
<td>1.50</td>
<td>800</td>
<td>0.0041</td>
<td>0.0667</td>
<td>0.0636</td>
<td>0.9310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.75</td>
<td>200</td>
<td>0.0160</td>
<td>0.1926</td>
<td>0.1715</td>
<td>0.8930</td>
<td>1.75</td>
<td>200</td>
<td>0.0126</td>
<td>0.1978</td>
<td>0.1659</td>
<td>0.8850</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.75</td>
<td>500</td>
<td>0.0066</td>
<td>0.1311</td>
<td>0.1212</td>
<td>0.9130</td>
<td>1.75</td>
<td>500</td>
<td>0.0082</td>
<td>0.1212</td>
<td>0.1121</td>
<td>0.9160</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.75</td>
<td>800</td>
<td>-0.0049</td>
<td>0.1031</td>
<td>0.0962</td>
<td>0.9290</td>
<td>1.75</td>
<td>800</td>
<td>-0.0018</td>
<td>0.0954</td>
<td>0.0914</td>
<td>0.9260</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5t</td>
<td>1.25</td>
<td>200</td>
<td>-0.0016</td>
<td>0.0858</td>
<td>0.0762</td>
<td>0.9090</td>
<td>t</td>
<td>1.25</td>
<td>200</td>
<td>-0.0057</td>
<td>0.0747</td>
<td>0.0699</td>
<td>0.9230</td>
</tr>
<tr>
<td>0.5t</td>
<td>3.75</td>
<td>500</td>
<td>0.0019</td>
<td>0.0517</td>
<td>0.0500</td>
<td>0.9370</td>
<td>0.75t</td>
<td>3.75</td>
<td>500</td>
<td>0.0003</td>
<td>0.0472</td>
<td>0.0465</td>
<td>0.9410</td>
</tr>
<tr>
<td>0.5t</td>
<td>6.25</td>
<td>800</td>
<td>0.0009</td>
<td>0.0402</td>
<td>0.0402</td>
<td>0.9460</td>
<td>0.75t</td>
<td>6.25</td>
<td>800</td>
<td>0.0012</td>
<td>0.0379</td>
<td>0.0374</td>
<td>0.9460</td>
</tr>
<tr>
<td>1.5</td>
<td>200</td>
<td>0.0138</td>
<td>0.1361</td>
<td>0.1192</td>
<td>0.8830</td>
<td>1.5</td>
<td>200</td>
<td>0.0125</td>
<td>0.1235</td>
<td>0.1127</td>
<td>0.9130</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>500</td>
<td>0.0002</td>
<td>0.0862</td>
<td>0.0827</td>
<td>0.9310</td>
<td>1.5</td>
<td>500</td>
<td>0.0026</td>
<td>0.0847</td>
<td>0.0777</td>
<td>0.9050</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>800</td>
<td>-0.0014</td>
<td>0.0672</td>
<td>0.0652</td>
<td>0.9290</td>
<td>1.5</td>
<td>800</td>
<td>0.0052</td>
<td>0.0658</td>
<td>0.0621</td>
<td>0.9210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.75</td>
<td>200</td>
<td>0.0101</td>
<td>0.1916</td>
<td>0.1656</td>
<td>0.8840</td>
<td>1.75</td>
<td>200</td>
<td>0.0154</td>
<td>0.1742</td>
<td>0.1577</td>
<td>0.8840</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.75</td>
<td>500</td>
<td>0.0015</td>
<td>0.1224</td>
<td>0.1172</td>
<td>0.9160</td>
<td>1.75</td>
<td>500</td>
<td>0.0031</td>
<td>0.1207</td>
<td>0.1135</td>
<td>0.9250</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.75</td>
<td>800</td>
<td>-0.0047</td>
<td>0.1019</td>
<td>0.0933</td>
<td>0.9270</td>
<td>1.75</td>
<td>800</td>
<td>0.0018</td>
<td>0.0980</td>
<td>0.0912</td>
<td>0.9300</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Scenario II - Estimation of $\rho(s, t, \theta) = 1 + \theta_0(-0.15t + 0.9)(-0.152 + 0.9)$. Summary of Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP, for the parameter $\theta_0$ in $\rho(s, t, \theta)$. Each entry is based on 1000 simulations with correctly specified marginals and Rate Ratio form.

<table>
<thead>
<tr>
<th>$\mu_{0j}(t)$</th>
<th>$\theta_0$</th>
<th>N</th>
<th>Bias</th>
<th>SEE</th>
<th>ESE</th>
<th>CP</th>
<th>$\mu_{0j}(t)$</th>
<th>$\theta_0$</th>
<th>N</th>
<th>Bias</th>
<th>SEE</th>
<th>ESE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5t</td>
<td>0.25</td>
<td>200</td>
<td>0.0114</td>
<td>0.1793</td>
<td>0.1697</td>
<td>0.9410</td>
<td>t: 0.25</td>
<td>200</td>
<td>-0.0073</td>
<td>0.1205</td>
<td>0.1207</td>
<td>0.9450</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>500</td>
<td>0.0002</td>
<td>0.1134</td>
<td>0.1089</td>
<td>0.9440</td>
<td></td>
<td>500</td>
<td>0.0029</td>
<td>0.0760</td>
<td>0.0781</td>
<td>0.9520</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>800</td>
<td>0.0001</td>
<td>0.0891</td>
<td>0.0860</td>
<td>0.9520</td>
<td></td>
<td>800</td>
<td>0.0007</td>
<td>0.0613</td>
<td>0.0620</td>
<td>0.9490</td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>200</td>
<td>0.0053</td>
<td>0.2325</td>
<td>0.2162</td>
<td>0.9260</td>
<td>0.5</td>
<td>200</td>
<td>-0.0047</td>
<td>0.1832</td>
<td>0.1698</td>
<td>0.9160</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.0038</td>
<td>0.1427</td>
<td>0.1406</td>
<td>0.9310</td>
<td></td>
<td>500</td>
<td>-0.0036</td>
<td>0.1124</td>
<td>0.1099</td>
<td>0.9310</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.0026</td>
<td>0.1156</td>
<td>0.1118</td>
<td>0.9360</td>
<td></td>
<td>800</td>
<td>-0.0038</td>
<td>0.0904</td>
<td>0.0881</td>
<td>0.9290</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>200</td>
<td>-0.0096</td>
<td>0.3406</td>
<td>0.3295</td>
<td>0.9170</td>
<td>1</td>
<td>200</td>
<td>-0.0193</td>
<td>0.2988</td>
<td>0.2811</td>
<td>0.8960</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0066</td>
<td>0.2120</td>
<td>0.2180</td>
<td>0.9370</td>
<td></td>
<td>500</td>
<td>-0.0048</td>
<td>0.1779</td>
<td>0.1890</td>
<td>0.9490</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.0007</td>
<td>0.1697</td>
<td>0.1765</td>
<td>0.9440</td>
<td></td>
<td>800</td>
<td>-0.0031</td>
<td>0.1472</td>
<td>0.1518</td>
<td>0.9420</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.6</td>
<td>200</td>
<td>-0.0112</td>
<td>0.4884</td>
<td>0.4857</td>
<td>0.9150</td>
<td>1.6</td>
<td>200</td>
<td>-0.0101</td>
<td>0.4317</td>
<td>0.4388</td>
<td>0.9060</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0120</td>
<td>0.3338</td>
<td>0.3325</td>
<td>0.9140</td>
<td></td>
<td>500</td>
<td>-0.0065</td>
<td>0.3049</td>
<td>0.3030</td>
<td>0.9130</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>-0.0073</td>
<td>0.2697</td>
<td>0.2685</td>
<td>0.9240</td>
<td></td>
<td>800</td>
<td>-0.0085</td>
<td>0.2427</td>
<td>0.2452</td>
<td>0.9370</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>-0.0324</td>
<td>0.6739</td>
<td>0.6011</td>
<td>0.8740</td>
<td>2</td>
<td>200</td>
<td>-0.0270</td>
<td>0.6112</td>
<td>0.5526</td>
<td>0.8780</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.0166</td>
<td>0.4226</td>
<td>0.4164</td>
<td>0.9040</td>
<td></td>
<td>500</td>
<td>-0.0249</td>
<td>0.3853</td>
<td>0.3821</td>
<td>0.9100</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>-0.0059</td>
<td>0.3300</td>
<td>0.3431</td>
<td>0.9300</td>
<td></td>
<td>800</td>
<td>-0.0053</td>
<td>0.3039</td>
<td>0.3198</td>
<td>0.9390</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4: Scenario III - Estimation of $\theta$'s in $\rho(\theta; Z_k) = \theta_1 I(Z_k = 1) + \theta_2 I(Z_k = 0)$, with true value $\theta_1 = 1.25$ and $\theta_2 = 1.75$. Summary of Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP. Each entry is based on 1000 simulations with correctly specified marginals and rate ratio form.

<table>
<thead>
<tr>
<th>$\mu_{0j}(t)$</th>
<th>N</th>
<th>$\theta_1$</th>
<th></th>
<th></th>
<th></th>
<th>$\theta_2$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Bias</td>
<td>SEE</td>
<td>ESE</td>
<td>CP</td>
<td>Bias</td>
<td>SEE</td>
<td>ESE</td>
<td>CP</td>
</tr>
<tr>
<td>0.25$t$</td>
<td></td>
<td>-0.0094</td>
<td>0.1284</td>
<td>0.1177</td>
<td>0.9250</td>
<td>-0.0012</td>
<td>0.4543</td>
<td>0.3610</td>
<td>0.8860</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.0046</td>
<td>0.0817</td>
<td>0.0776</td>
<td>0.9280</td>
<td>-0.0033</td>
<td>0.2675</td>
<td>0.2476</td>
<td>0.9190</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0028</td>
<td>0.0653</td>
<td>0.0626</td>
<td>0.9430</td>
<td>-0.0043</td>
<td>0.2218</td>
<td>0.2016</td>
<td>0.9130</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.0010</td>
<td>0.0561</td>
<td>0.0545</td>
<td>0.9510</td>
<td>0.0151</td>
<td>0.1906</td>
<td>0.1798</td>
<td>0.9380</td>
</tr>
<tr>
<td></td>
<td>1100</td>
<td>0.0040</td>
<td>0.0561</td>
<td>0.0545</td>
<td>0.9510</td>
<td>0.0202</td>
<td>0.1906</td>
<td>0.1798</td>
<td>0.9380</td>
</tr>
<tr>
<td>0.50$t$</td>
<td></td>
<td>-0.0055</td>
<td>0.1193</td>
<td>0.1075</td>
<td>0.9190</td>
<td>-0.0160</td>
<td>0.3267</td>
<td>0.2703</td>
<td>0.8720</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.0046</td>
<td>0.0765</td>
<td>0.0715</td>
<td>0.9320</td>
<td>-0.0226</td>
<td>0.2118</td>
<td>0.1873</td>
<td>0.8950</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0013</td>
<td>0.0597</td>
<td>0.0583</td>
<td>0.9530</td>
<td>-0.0081</td>
<td>0.1674</td>
<td>0.1542</td>
<td>0.9180</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>-0.0003</td>
<td>0.0494</td>
<td>0.0502</td>
<td>0.9640</td>
<td>-0.0066</td>
<td>0.1408</td>
<td>0.1355</td>
<td>0.9240</td>
</tr>
<tr>
<td></td>
<td>1100</td>
<td>-0.0003</td>
<td>0.0494</td>
<td>0.0502</td>
<td>0.9640</td>
<td>-0.0066</td>
<td>0.1408</td>
<td>0.1355</td>
<td>0.9240</td>
</tr>
<tr>
<td>0.75$t$</td>
<td>500</td>
<td>-0.0103</td>
<td>0.1131</td>
<td>0.1011</td>
<td>0.8960</td>
<td>-0.0182</td>
<td>0.2826</td>
<td>0.2469</td>
<td>0.8820</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0020</td>
<td>0.0722</td>
<td>0.0689</td>
<td>0.9340</td>
<td>-0.0095</td>
<td>0.1933</td>
<td>0.1706</td>
<td>0.9110</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>-0.0009</td>
<td>0.0564</td>
<td>0.0552</td>
<td>0.9370</td>
<td>-0.0037</td>
<td>0.1512</td>
<td>0.1401</td>
<td>0.9200</td>
</tr>
<tr>
<td></td>
<td>1100</td>
<td>-0.0020</td>
<td>0.0486</td>
<td>0.0470</td>
<td>0.9420</td>
<td>0.0033</td>
<td>0.1251</td>
<td>0.1231</td>
<td>0.9330</td>
</tr>
</tbody>
</table>
Table 5: Scenario IV - estimates $\theta_1$, $\theta_2$ in the underline models where $\rho(\theta, s, t|Z_k) = 1 + \theta_1 \frac{(0.25+0.1Z_k)(0.25+0.2Z_k)}{(0.5+0.5+0.5Z_k)(0.5+0.5+0.5Z_k)}$ and $\rho(\theta, s, t|Z_k) = 1 + \theta_2 \frac{(0.25+0.1Z_k)(0.25+0.2Z_k)}{(0.5+0.5+0.5Z_k)(0.5+0.5+0.5Z_k)}$. Summary of Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP of $\theta$ where each entry is based on 1000 simulations. The averaged observed events for type1 (2) event is 2.44(2.56)

<table>
<thead>
<tr>
<th>$\theta_1$</th>
<th>N</th>
<th>Bias</th>
<th>SEE</th>
<th>ESE</th>
<th>CP</th>
<th>$\theta_2$</th>
<th>Bias</th>
<th>SEE</th>
<th>ESE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>200</td>
<td>-0.0017</td>
<td>0.4030</td>
<td>0.4401</td>
<td>0.9580</td>
<td>0.25</td>
<td>-0.0027</td>
<td>0.2143</td>
<td>0.2513</td>
<td>0.9700</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0221</td>
<td>0.2529</td>
<td>0.2840</td>
<td>0.9650</td>
<td></td>
<td>0.0005</td>
<td>0.1320</td>
<td>0.1625</td>
<td>0.9810</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.0014</td>
<td>0.1916</td>
<td>0.2255</td>
<td>0.9770</td>
<td></td>
<td>-0.0020</td>
<td>0.1078</td>
<td>0.1287</td>
<td>0.9790</td>
</tr>
<tr>
<td></td>
<td>1100</td>
<td>0.0038</td>
<td>0.1728</td>
<td>0.1933</td>
<td>0.9720</td>
<td></td>
<td>-0.0001</td>
<td>0.0897</td>
<td>0.1104</td>
<td>0.9820</td>
</tr>
<tr>
<td>0.50</td>
<td>200</td>
<td>-0.0177</td>
<td>0.4404</td>
<td>0.4948</td>
<td>0.9730</td>
<td>0.50</td>
<td>-0.0202</td>
<td>0.2703</td>
<td>0.3050</td>
<td>0.9560</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0178</td>
<td>0.2843</td>
<td>0.3201</td>
<td>0.9570</td>
<td></td>
<td>-0.0047</td>
<td>0.1703</td>
<td>0.2001</td>
<td>0.9670</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>-0.0001</td>
<td>0.2235</td>
<td>0.2570</td>
<td>0.9750</td>
<td></td>
<td>-0.0001</td>
<td>0.1337</td>
<td>0.1605</td>
<td>0.9820</td>
</tr>
<tr>
<td></td>
<td>1100</td>
<td>0.0026</td>
<td>0.2003</td>
<td>0.2197</td>
<td>0.9660</td>
<td></td>
<td>0.0014</td>
<td>0.1123</td>
<td>0.1378</td>
<td>0.9850</td>
</tr>
<tr>
<td>0.75</td>
<td>200</td>
<td>-0.0154</td>
<td>0.5316</td>
<td>0.5598</td>
<td>0.9510</td>
<td>0.75</td>
<td>-0.0438</td>
<td>0.3170</td>
<td>0.3607</td>
<td>0.9470</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0031</td>
<td>0.3184</td>
<td>0.3641</td>
<td>0.9680</td>
<td></td>
<td>-0.0079</td>
<td>0.2029</td>
<td>0.2443</td>
<td>0.9750</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>-0.0105</td>
<td>0.2513</td>
<td>0.2899</td>
<td>0.9720</td>
<td></td>
<td>0.0030</td>
<td>0.1554</td>
<td>0.1952</td>
<td>0.9830</td>
</tr>
<tr>
<td></td>
<td>1100</td>
<td>-0.0050</td>
<td>0.2156</td>
<td>0.2496</td>
<td>0.9720</td>
<td></td>
<td>-0.0029</td>
<td>0.1385</td>
<td>0.1671</td>
<td>0.9720</td>
</tr>
<tr>
<td>1.00</td>
<td>200</td>
<td>0.0008</td>
<td>0.5731</td>
<td>0.6270</td>
<td>0.9630</td>
<td>1.00</td>
<td>-0.0282</td>
<td>0.3863</td>
<td>0.4341</td>
<td>0.9530</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0002</td>
<td>0.3891</td>
<td>0.4154</td>
<td>0.9590</td>
<td></td>
<td>-0.0047</td>
<td>0.2473</td>
<td>0.2920</td>
<td>0.9590</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>-0.0007</td>
<td>0.2939</td>
<td>0.3308</td>
<td>0.9730</td>
<td></td>
<td>-0.0019</td>
<td>0.1955</td>
<td>0.2331</td>
<td>0.9740</td>
</tr>
<tr>
<td></td>
<td>1100</td>
<td>0.0010</td>
<td>0.2584</td>
<td>0.2834</td>
<td>0.9620</td>
<td></td>
<td>-0.0107</td>
<td>0.1614</td>
<td>0.1985</td>
<td>0.9740</td>
</tr>
</tbody>
</table>
3.2 Hypothesis testing of the rate ratio

Although the parametric rate ratio model has better interpretability than nonparametric ones, it might suffer from model misspecification and induce model bias. In this section, we aim at providing a goodness-of-fit procedure to test the parametric assumption of the rate ratio, i.e. $H_0 : \rho(s,t; \theta_1, z_2) = \theta_0$, under the additive marginal mean rate model. A finite sample study is also conducted to check the performance of the goodness-of-fit procedure.

3.2.1 Procedure description

The residual process followed by equation (2.4) under model (3.1) is defined as

$$V(s, t, \hat{\theta}, \hat{\beta}_1, \hat{\mu}_{01}(\cdot), \hat{\beta}_2, \hat{\mu}_{02}(\cdot))$$

$$= N^{-1/2} \sum_{k=1}^{N} \int_0^t \int_0^s W_N(u, v) \left. \frac{\partial \rho(u, v; \theta)}{\partial \theta} \right|_{\theta=\hat{\theta}} \left\{ dN_k(u) dN_k(v) \right\}$$

$$- \rho(u, v; \hat{\theta}) Y_k(u) \{ d\hat{\mu}_{01}(u) + \hat{\beta}_1^T Z_k(u) du \} Y_k(v) \{ d\hat{\mu}_{02}(v) + \hat{\beta}_2^T Z_k(v) dv \},$$

(3.25)

where $W_N(u, v)$ is a prespecified weight and for simplicity let $W_N(u, v) = 1$. With correctly specified marginal mean rate and $\rho(s, t; \theta_0; z_1, z_2)$, one would expect the value of equation (3.25) to fluctuate around zero at the any $(s, t) \in [0, \tau]^2$.

Let $T = \sup_{s,t \in [0, \tau]^2} \| V(s, t, \hat{\theta}, \hat{\beta}_1, \hat{\mu}_{01}(\cdot), \hat{\beta}_2, \hat{\mu}_{02}(\cdot)) \|$ be the supremum test statistic which measures the maximum observed residuals across the observable periods of type 1(2) events. A reasonable small $T$ value is expected from a fitting. Since the underlying distribution of $T$ is intractable, we apply the Gaussian multiplier method to approximate its empirical distribution.
The Gaussian multiplier method.

The first order Taylor expansion of equation (3.25) w.r.t $\theta$ is

$$
V(s, t, \hat{\theta}, \hat{\beta}_1, \hat{\mu}_{01}(\cdot), \hat{\beta}_2, \hat{\mu}_{02}(\cdot)) = V(s, t, \theta, \hat{\beta}_1, \hat{\mu}_{01}(\cdot), \hat{\beta}_2, \hat{\mu}_{02}(\cdot))
$$

$$
+ N^{-1/2} \frac{\partial V(s, t, \theta, \hat{\beta}_1, \hat{\mu}_{01}(\cdot), \hat{\beta}_2, \hat{\mu}_{02}(\cdot))}{\partial \theta} N^{1/2}(\hat{\theta} - \theta)
$$

$$
+ o_p(1),
$$

(3.26)

which can be further decomposed as

$$
V(s, t, \hat{\theta}, \hat{\beta}_1, \hat{\mu}_{01}(\cdot), \hat{\beta}_2, \hat{\mu}_{02}(\cdot)) = V(s, t, \theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot))
$$

$$
+ N^{-1/2} \sum_{k=1}^{N} \left\{ \Upsilon_{k1}(s, t, \theta) + \Upsilon_{k2}(s, t, \theta) + \zeta_{k1}(s, t, \theta) + \zeta_{k2}(s, t, \theta) \right\} + o_p(1),
$$

(3.27)

with details shown in Appendix C. Let $T^* = \sup_{s,t \in [0,\tau]} \| V^*(s, t) \|$ and

$$
V^*(s, t, \hat{\theta}) = \left\{ V(s, t, \hat{\theta}, \hat{\beta}_1, \hat{\mu}_{01}(\cdot), \hat{\beta}_2, \hat{\mu}_{02}(\cdot))
$$

$$
+ N^{-1/2} \sum_{k=1}^{N} \tilde{\Upsilon}_{k1}(s, t, \hat{\theta}) + \tilde{\Upsilon}_{k2}(s, t, \hat{\theta}) + \tilde{\zeta}_{k1}(s, t, \hat{\theta}) + \tilde{\zeta}_{k2}(s, t, \hat{\theta}) \right\} G_k,
$$

(3.28)

where $G = (G_1, G_2, G_3, ... G_N)$ is a vector of i.i.d standard normal random numbers.

Comparing equation (3.28) and (3.27), the multiplication of a Gaussian Random variable $G_k$ keeps $V^*(s, t, \hat{\theta})$ and $\tilde{V}(s, t, \hat{\theta}, \hat{\mu}_1(\cdot; Z_{k1}), \hat{\beta}_1, \hat{\mu}_{01}(\cdot), \hat{\beta}_2, \hat{\mu}_{02}(\cdot))$ sharing the same expectation and variance, thus $T^*$ and $T$ follows the same distribution.

The Gaussian Multiplier re-sampling method is summarized in Algorithm 1. In a single simulation, one Gaussian random vector $\{G_1, G_2, ..., G_N\}$ is generated and
$V^*(s,t,\hat{\theta})$ is calculated from equation (3.28). By taking the maximum of $V^*(s,t,\hat{\theta})$ across all the equally distanced grids, we have one sample from the $T^*$ distribution. Repeating this simulation procedure 1000 times allows us to obtain 1000 samples and therefore the empirical distribution of $T^*$. On the other hand, the supremum test statistic $T$ can be obtained by taking the maximum of equation (3.25). We consider the 95th percentile among the 1000 realizations of $T^*$ as the critical value ($C_{95}$) and would reject $H_0$ if $T > C_{95}$.

### 3.2.2 Simulation studies

In this section, we conduct simulation studies to investigate the performance of the proposed goodness-of-fit procedure.

For bivariate counting processes, we will firstly detect the existence of dependency. The null model is the independent bivariate counting processes and the the constant rate ratio model is treated as its alternative. Secondly, we propose the $H_0 : \rho(s, t, \theta; z_1, z_2) = \theta_0$ and Piecewise Constant (PWC), Time Dependent (TD), Time and Covariate Dependent (TCD) models as $H_a$ models. The size and power of the hypothesis test are also computed via Gaussian Multiplier Method.

#### 3.2.2.1 Testing for independence

The first hypothesis of interest is whether \( \{N_{k1}(\cdot)\} \) and \( \{N_{k2}(\cdot)\} \) are independent, which is equivalent to test $H_0 : \rho = 1$ vs $H_a : \rho \neq 1$. To investigate the size, events data are generated from an additive marginal model

$$d\mu_j(t; Z_{kj}(t)) = d\mu_{0j}(t) + \beta_j Z_{kj}(t).$$
Let $\tau = 5$, $\beta_{01} = 0.5$, $\beta_{02} = 1$, $C_{kj}$ follows Uniform$[0, \tau]$ and covariates $Z_{k1}, Z_{k2}$ are from a uniform distribution on $[1, 2]$. We take $\mu_{0j}(t) = 0.25t$, $0.5t$, $0.75t$, and $t$ which gives the average observed events counts range from 2.50 to 6.26. Datasets under $H_a : \rho(\theta, s, t) = \theta_0$ are generated from the shared frailty model

$$d\mu_j(t; Z_{kj}(t), R_k) = R_k\{d\mu_{0j}(t) + \beta_j Z_{kj}(t)\},$$

where $R_k$ from Gamma$(a, b)$ with $(a, b) = (4, 0.25), (2, 0.5), (1.33, 0.75), (1, 1)$. Thus $\theta_0$ in $H_a$ are equal to $1.25, 1.5, 1.75, 2$. We compare the supreme test statistic under $\rho(\theta, s, t) = \theta_0$ to the corresponding value obtained by assuming $\rho = 1$ and regard the rejection rate among 1000 simulations as the power of the test.

We only consider the case when $\mu_{0j} = 0.25t$, since it has the smallest number of observed events and other cases would have even more rejection, i.e. higher power. The empirical size (power) calculated as the rejection rate from 1000 simulated datasets under $H_0 : \rho = 1$ ($H_a : \rho(s, t, \theta; z_1, z_2) = \theta_0$).

Table 8 shows that the proposed testing procedure has size around its nominee value (5%). The test procedure is powerful at detecting the non-independent case with probability above 99%.

### 3.2.2.2 Testing for parametric form with constant rate ratio

We are also interested in testing the parametric assumption of the rate ratio, i.e. $H_0 : \rho(\theta, s, t) = \theta_0$. The null model is the shared frailty model in equation (3.11),

**Shared Frailty:**

$$E[dN^*_{kj}(u)|R_k, Z_{kj}(u)] = R_k\{d\mu_{0j}(u) + \beta_j Z_{kj}(u) du\}$$

from which $\rho(s, t, \theta) = \theta_0$ where $\theta_0 = 1 + \sigma^2/\mu^2$, $E[R_k] = \mu$ and var$[R_k] = \sigma^2$. 

From the first section in Table 9, we see the empirical size of the test under null model is bounded by its nominee value 0.05. Thus the hypothesis testing can control the probability of mistakenly reject $H_0 : \rho(s, t, \theta) = \theta_0$ under 0.05.

To investigate the power of the test, we propose three alternative models to introduce the time varying and covariate dependency cases: the Piecewise Constant Rate Ratio Model (PWC), the Time Dependent Rate Ratio Model (TD Model) and the Covariate Dependent Rate Ratio Model (CD Model). Alternative models and the corresponding performance are illustrated in the following sections.

(I) The piecewise constant rate ratio model - PWC Model

Described in equation (3.29), the random effect is time varying, which is a natural generalization of the shared frailty model

$$
PWC: d\mu_j(t|R_k(t), Z_{kj}(t)) = R_k(t)\{d\mu_{0j}(t) + \beta_j^T Z_{kj}(t) dt\}. \quad (3.29)$$

For simplicity, we consider $R_k(t)$ come from different distributions only when $t$ falls in non-overlapping intervals.

Let $\tau = 5$, $R_k(t) = I(t < 2.5)R_{k0} + I(t > 2.5)R_{k1}$, where $R_{k0}$ and $R_{k1}$ are independently from the shifted Gamma($a_0, b_0, \delta_0$) and Gamma($a_1, b_1, \delta_1$) respectively. The shifted Gamma Distribution with $(a, b, \delta)$ as shape, scale and shift parameters is introduced here to avoid rare event observations. We take $\mu_{01}(u) = \mu_{02}(u) = 0.125u^2$, $\beta_1 = 0.5$, $\beta_2 = 1$, and $Z_{k1}(u), Z_{k2}(u)$ from uniform[1, 2].

Table 6 summarizes the parameter settings and the corresponding Rate Ratio value. We see the variation of the association is increasing from PWC1 to PWC4 and one
can visualize the trend in Figure 1 as well.

Table 6: Summary of simulation settings under the piecewise constant rate ratio model with the corresponding $\rho$ values followed from Proposition 2.

<table>
<thead>
<tr>
<th>Settings</th>
<th>PWC1</th>
<th>PWC2</th>
<th>PWC3</th>
<th>PWC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{k0}$ : $(a_0, b_0, \delta_0)$</td>
<td>(0.25,1,0.75)</td>
<td>(0.5,1,0.5)</td>
<td>(0.25,2,0.5)</td>
<td>(0.25,2,0.5)</td>
</tr>
<tr>
<td>$R_{k1}$ : $(a_1, b_1, \delta_1)$</td>
<td>(0.25,1,0.75)</td>
<td>(0.25,1,0.75)</td>
<td>(0.5,1,0.5)</td>
<td>(0.25,1,0.75)</td>
</tr>
<tr>
<td>$\rho(s &lt; 2.5, t &lt; 2.5)$</td>
<td>1.25</td>
<td>1.5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$\rho(s &gt; 2.5, t &lt; 2.5)$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\rho(s &gt; 2.5, t &gt; 2.5)$</td>
<td>1.25</td>
<td>1.25</td>
<td>1.5</td>
<td>1.25</td>
</tr>
</tbody>
</table>

To evaluate the power of the test, first, we generate 1000 datasets and within each simulation, the rate ratio $\rho(\theta, s, t)$ is estimated under ($H_0 : \rho(s, t, \theta) = \theta_0$). The residual process and supreme statistic $T$ are computed and a rejection is made when $T > C_{95}$, where $C_{95}$ is the 95% percentile of Gaussian Multiplier samplers. The overall rejection rate among the 1000 datasets is considered as the empirical power of the hypothesis test. From Table 9, the power increases with the sample size and it is more likely to detect the divergence from $H_0$ when the association become stronger.

(II) Time dependent rate ratio model - TD model

Assuming \{${N}_{k1}(s), {N}_{k2}(t)$\} follows the Bivariate Counting processes below

\[
{N}_{k1}(s) = \bar{N}_{k1}(s) + {N}_{k0}(s),
\]

\[
{N}_{k2}(t) = \bar{N}_{k2}(t) + {N}_{k0}(t),
\]

where $\bar{N}_{k1}(\cdot)$, $\bar{N}_{k2}(\cdot)$ and $N_{k0}(\cdot)$ follow Poisson Processes and are also mutually independent.

Let $\lambda_{k0}(t|Z_{kj}, R_k) dt$ be the event rate of $N_{k0}(t)$ and $\lambda_{k0}(t|Z_{kj}, R_k) = R_k(d\mu_{0j}(t) + \beta_{0j}Z_{kj}(t)$, where $R_k$ is the frailty variable with mean $\mu$ and variance $\sigma^2$. For $j =$
Figure 1: Visualization of Piecewise Constant $\rho(s, t, \theta)$ (PWC) under the Additive Marginal Models. The variation of $\rho(s, t)$ between different pieces is growing from PWC1 to PWC4.

1, 2 and $t \in (0, \tau)$, assume $\tilde{\lambda}_{kj}(t|Z_{kj}(t)) = m_j(t)\lambda_{k0}(t|Z_{kj}(t))$, with a nonnegative multiplier function $m_j(t)$. Following simulation settings in equation 3.19 to generate data that share the rate ratio as

$$TD \text{ model: } \rho(\theta, s, t) = 1 + \theta_0 \times (-0.15s + 0.9)(-0.15t + 0.9).$$

(3.31)

where $\theta_0 = \frac{\sigma^2}{\mu^2}$ reflects the time varying component in $\rho(\theta, s, t)$ proportionally. To capture different time varying levels, we take $R_k$ from a shifted gamma distribution, with parameters $(a, b, \delta) = (0.25, 2, 0.5), (0.2, 3, 0.4), (0.25, 3, 0.25)$ and $(0.2, 4, 0.2)$ so that $\mu = 1$ and $\sigma^2 = 1, 1.8, 2.25$ and 3.2. Let $\beta_{01} = \beta_{02} = 0$, $\tau = 5$, $C_{kj}$ be uniform on $(0, \tau)$, and $Z_{k1}, Z_{k2}$ are i.i.d uniform$(1, 2)$. Simulation settings are summarized in
Table 7 and Figure 2.

Table 7: Simulation settings of the Time Varying Rate Ratio (TD models). From TD1 to TD4, the value of $\sigma^2/\mu^2$ is increasing and so is the association between the bivariate recurrent event processes.

<table>
<thead>
<tr>
<th>Settings</th>
<th>TD1</th>
<th>TD2</th>
<th>TD3</th>
<th>TD4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(\mu, \sigma^2)$</td>
<td>(1, 1)</td>
<td>(1, 1.8)</td>
<td>(1, 2.25)</td>
<td>(1, 3.2)</td>
</tr>
<tr>
<td>$\sigma^2/\mu^2$</td>
<td>1</td>
<td>1.8</td>
<td>2.25</td>
<td>3.2</td>
</tr>
</tbody>
</table>

Figure 2: The contour plot of the Rate Ratio $\rho(s, t)$ under the additive marginal mean rate models. The x-axis and y-axis represents the observation time for type1 and type2 events. From upper left to lower right, the heterogeneity of $\rho(s, t)$ is increased.

(a) TD1 contour plot
(b) TD2 contour plot
(c) TD3 contour plot
(d) TD4 contour plot

The variation of $\rho(\theta, s, t)$ is scaling up from TD1 to TD4, so does the empirical power of the test shown in Table 9. From our observation, the proposed model checking procedure performs well with a large sample size, especially when the Rate
Ratio is very time dependent.

(III) Time and Covariates Dependent Rate Ratio Model - TCD Model

Under the same framework of the TD Model, assume $N_{k0}(t)$ and $\tilde{N}_{kj}(t)$ are Poisson processes with rate conditional on covariates and unobservable frailty $R_k$ as

$$\lambda_{k0}(t|Z_{kj}, R_k) = R_k \{0.25 + \beta_{0j} Z_{kj}\} \text{ and } \tilde{\lambda}_{kj}(t) = t \text{ respectively.}$$

The conditional rate of $N_{kj}(t)$ equals to $\lambda_{kj}(t|Z_{kj}, R_k)$ where $\lambda_{kj}(t|Z_{kj}, R_k) = t + \{0.25 + \beta_{0j} Z_{kj}\}$. Let $\beta_{01} = 0.5$, $\beta_{02} = 1$, $Z_{kj}$ follow uniform$(1, 2)$. Take $R_k$ as i.i.d Gamma$(1/v, v)$ with $v = 0.5, 0.8, 1, 2$ so that $E(R_k) = 1$ and $\text{var}(R_k) = 0.5, 0.8, 1, 2$. Denoted by $\rho(\theta, s, t|Z_{k1}, Z_{k2})$ the rate ratio of $\{N_{k1}(s), N_{k2}(t)\}$, where

$$\rho(\theta, s, t|Z_{k1}, Z_{k2}) = 1 + \theta \frac{(0.25 + 0.5Z_{k1})(0.25 + Z_{k2})}{(t + 0.25 + 0.5Z_{k1})(s + 0.25 + Z_{k2})}. \quad (3.32)$$

is obtained by Proposition 3, with true $\theta$ equal to 0.5, 0.8, 1 and 2.

The average rejection of $H_0 : \rho(s, t, \theta) = \theta_0$ under equation (3.32) among 1000 are summarized in Table 9. The test is powerful at detecting violation of $H_0$ and the rejection rate of the test is consistently increase when the sample size changed from 200 to 800.
Algorithm 1 Gaussian Multiplier Method

For dataset $m = 1, 2, ..., \mathcal{M}$

1. Calculate $T$ by (3.2.1)

2. Consider a large integer $B$, say 1000. We generate a $B \times N$ matrix $G$ composed by i.i.d Standard Gaussian random numbers, so that each row is an $N$ dimensional vector:

\[
\begin{bmatrix}
G_{11} & G_{12} & G_{13} & \ldots & G_{1n} \\
G_{21} & G_{22} & G_{23} & \ldots & G_{2n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
G_{B1} & G_{B2} & G_{B3} & \ldots & G_{Bn}
\end{bmatrix}
\]

For each row, applying equation (3.28) to calculate the realization of $V^*(s,t)$ and $T^*$. Enumerate all the rows to get a list of $\{V^*_1(s,t), V^*_2(s,t), ..., V^*_B(s,t)\}$.

3. Denote the 95th percentile of $\{V^*_1(s,t), V^*_2(s,t), ..., V^*_B(s,t)\}$ to be $C_{95}$. We would reject $H_0$ if $T > C_{95}$ and fail to reject $H_0$ if $T < C_{95}$.

Calculate the percentage of rejections in a total of $\mathcal{M}$ datasets to find the size or the power of test statistic.
Table 8: Observed sizes and powers of the test statistic $T$ via the proposed model-checking procedure under $H_0 : \rho = 1$ vs $H_a : \rho(s, t, \theta) = \theta$ and $\theta > 1$, at significance level 0.05. The numbers in the parentheses represent the count for type 1 and type 2 event across the observation period. Each entry is calculated based on 1000 Gaussian multiplier samples with 1000 replicates.

<table>
<thead>
<tr>
<th>Event Count</th>
<th>$\rho$</th>
<th>$N=200$</th>
<th>$N=500$</th>
<th>$N=800$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2.50, 4.37)</td>
<td>1</td>
<td>0.043</td>
<td>0.052</td>
<td>0.051</td>
</tr>
<tr>
<td>(3.13, 5.02)</td>
<td>1</td>
<td>0.051</td>
<td>0.057</td>
<td>0.051</td>
</tr>
<tr>
<td>(3.76, 5.64)</td>
<td>1</td>
<td>0.043</td>
<td>0.053</td>
<td>0.041</td>
</tr>
<tr>
<td>(4.37, 6.26)</td>
<td>1</td>
<td>0.045</td>
<td>0.049</td>
<td>0.054</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Event Count</th>
<th>Power</th>
<th>$N=200$</th>
<th>$N=500$</th>
<th>$N=800$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2.50, 4.37)</td>
<td>1.25</td>
<td>0.995</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>1.5</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>1.75</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>
Table 9: Observed sizes and powers of the test statistic $T$ for the proposed model-checking procedure under $H_0: \rho(\theta, s, t) = \theta$ (i.e. constant) vs $H_a: \rho$ is not constant, at 0.05 significance level. Each entry is calculated based on 1000 Gaussian multiplier samples with 1000 replicates.

<table>
<thead>
<tr>
<th>Event Count</th>
<th>N</th>
<th>$\rho = 1.25$</th>
<th>$\rho = 1.5$</th>
<th>$\rho = 2$</th>
<th>$\rho = 2.25$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200</td>
<td>0.038</td>
<td>0.038</td>
<td>0.031</td>
<td>0.038</td>
</tr>
<tr>
<td>(3.50, 4.67)</td>
<td>500</td>
<td>0.057</td>
<td>0.037</td>
<td>0.032</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.042</td>
<td>0.042</td>
<td>0.051</td>
<td>0.046</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Event Count</th>
<th>N</th>
<th>PWC1</th>
<th>PWC2</th>
<th>PWC3</th>
<th>PWC4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200</td>
<td>0.173</td>
<td>0.579</td>
<td>0.638</td>
<td>0.755</td>
</tr>
<tr>
<td>(2.91, 4.80)</td>
<td>500</td>
<td>0.421</td>
<td>0.912</td>
<td>0.958</td>
<td>0.983</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.622</td>
<td>0.979</td>
<td>0.990</td>
<td>0.999</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Event Count</th>
<th>N</th>
<th>TD1</th>
<th>TD2</th>
<th>TD3</th>
<th>TD4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200</td>
<td>0.197</td>
<td>0.231</td>
<td>0.311</td>
<td>0.307</td>
</tr>
<tr>
<td>(4.27, 4.27)</td>
<td>500</td>
<td>0.556</td>
<td>0.621</td>
<td>0.760</td>
<td>0.738</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.773</td>
<td>0.821</td>
<td>0.887</td>
<td>0.894</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Event Count</th>
<th>N</th>
<th>TCD1</th>
<th>TCD2</th>
<th>TCD3</th>
<th>TCD4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200</td>
<td>0.250</td>
<td>0.336</td>
<td>0.405</td>
<td>0.455</td>
</tr>
<tr>
<td>(6.67, 8.54)</td>
<td>500</td>
<td>0.514</td>
<td>0.678</td>
<td>0.756</td>
<td>0.823</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.735</td>
<td>0.900</td>
<td>0.917</td>
<td>0.933</td>
</tr>
</tbody>
</table>
CHAPTER 4: ESTIMATION AND INference OF THE RATE RATIO UNDER THE MULTIPLICATIVE MARGINAL MODEL

4.1 Estimation by a two-stage approach

Additive and multiplicative mean rate models postulate a different relationship between the underline counting process and the covariates. The multiplicative model, also known as Cox model is popular due to its easy implementation and clear interpretation of the covariate effect. In this chapter, we develop the estimation procedure for the rate ratio under the multiplicative marginal event rate model.

Lin et al. (2000) proposed the mean rate of the counting process $N_{kj}^*(t)$ as

$$E[dN_{kj}^*(t)|Z_{kj}(t)] = d\mu_j(t; Z_{kj}(t)),$$
$$d\mu_j(t; Z_{kj}(t)) = e^{\beta_j^TZ_{kj}(t)} d\mu_{0j}(t),$$ (4.1)

where $\beta_j$ is a $p$-dimensional vector, $\mu_{0j}(t)$ is an unspecified baseline rate at time $t$. Assume $\rho(s, t, \theta; z_{K1}, z_{K2})$ is the rate ratio of $N_{k1}^*(t)$ and $N_{k2}^*(s)$. $\theta$ is the dependence parameter which can be approximated by solving the estimation equation (2.3), with the $\hat{\mu}_j(t)$ estimated by the method proposed by Lin et al. (2000). We adjust some notations from Chapter 3 with a superscription $c$ to represent estimators derived from model (4.1).
4.1.1 Review the estimation of the marginal model

Adapting from the approach of Lin et al. (2000), for type $j$ event we define

$$d\bar{N}_j(t) = \sum_{k=1}^{N} dN_{kj}(t),$$

$$M^c_{kj}(t) = N_{kj}(t) - \int_{0}^{t} Y_{kj}(u)e^{\beta^T Z_{kj}(u)} d\mu_{01}(u),$$

$$S^d_j(t, \beta) = N^{-1} \sum_{k=1}^{N} Y_{kj}(t) Z_{kj}^{\otimes d}(t)e^{\beta^T Z_{kj}(t)}, \quad d = 0, 1, 2 \quad (4.2)$$

where $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa^T$. Let $\tilde{Z}_j(t, \beta) = S^1_j(t, \beta)/S^0_j(t, \beta)$; $\tilde{z}_j(t, \beta)$, $s^d_j(t, \beta)$ be the limit of $\tilde{Z}_j(\beta, t)$ and $S^d_j(t, \beta)$ as $N \to \infty$ respectively.

Denote $\tilde{\beta}_j$ the solution to $L^c_j(\beta, \tau) = 0$, where $L^c_j(\beta, \tau) = \sum_{k=1}^{N} \int_{0}^{\tau} \{Z_{kj}(u) - \tilde{Z}_j(u, \beta)\} dN_{kj}(u)$ is the partial likelihood score function.

Under certain regularity conditions, $\tilde{\beta}_j$ converges almost surely to $\beta_j$ and $\sqrt{n}(\tilde{\beta}_j - \beta_j)$ has weak convergence to a zero-mean normal random vector with covariance matrix $\Gamma_j \equiv (A^c_j)^{-1}\Sigma^c_j(A^c_j)^{-1}$. When $\tilde{\beta}_j$ is available, the baseline function $\mu_{0j}(t)$ can be consistently estimated by the Aalen-Breslow type estimator

$$\bar{\mu}_{0j}(t, \tilde{\beta}_j) = \int_{0}^{t} \frac{d\bar{N}_j(u)}{N\tilde{s}^0_j(u, \tilde{\beta}_j)}, \quad t \in [0, \tau]. \quad (4.3)$$

We investigate the asymptotic properties of $\hat{\theta}$ under the assumption that the distribution functions of the $C_{kj}$ are independent from covariates and the counting process.

We recall Theorem 4.1, Theorem 4.2 due to Lin et al. (2000).

**Theorem 4.1** $\tilde{\beta}_j$ converges almost surely to $\beta_j$ and $\sqrt{N}(\tilde{\beta}_j - \beta_j)$ is asymptotically
normal with covariance matrix \((A_j^c)^{-1} \Sigma_j^c(A_j^c)^{-1}\), where

\[
A_j^c = E\left[ \int_0^\tau \{Z_{1j}(u) - \tilde{z}_j(u, \beta_j)\} \otimes^2 Y_{1j}(u) e^{\beta_j^T Z_{1j}(u)} d\mu_{0j}(u) \right],
\]

\[
\Sigma_j^c = E\left[ \int_0^\tau \{Z_{1j}(u) - \tilde{z}_j(u, \beta_j)\} dM_{i1}^c(u) \int_0^\tau \{Z_{1j}(v) - \tilde{z}_j(v, \beta_j)\} dM_{i1}^c(v) \right]. \tag{4.4}
\]

The asymptotic approximation of \(\tilde{\beta}_j\) is

\[
\sqrt{N}(\tilde{\beta}_j - \beta_j) = (A_j^c)^{-1} N^{-1/2} \sum_{k=1}^N \int_0^\tau \{Z_{kj}(u) - \tilde{z}_{kj}(u, \tilde{\beta}_j)\} dM_{kj}(u, \beta_j) + o_p(1), \tag{4.5}
\]

from which the covariance matrix can be consistently estimated by \(\tilde{A}_j^{-1} \tilde{\Sigma}_j \tilde{A}_j^{-1}\), with

\[
\tilde{A}_j = N^{-1} \sum_{k=1}^N \int_0^\tau \{Z_{kj}(u) - \tilde{Z}_{kj}(u, \tilde{\beta}_j)\} \otimes^2 Y_{kj}(u) e^{\beta_j^T Z_{kj}(u)} d\tilde{\mu}_{0j}(u),
\]

\[
\tilde{\Sigma}_j = N^{-1} \sum_{k=1}^N \tilde{\xi}_{kj}^2,
\]

\[
\tilde{\xi}_{kj} = \int_0^\tau \{Z_{kj}(u) - \tilde{Z}_{kj}(u, \tilde{\beta}_j)\} d\tilde{M}_{kj}(u),
\]

\[
\tilde{M}_{kj}(t) = N_{kj}(t) - \int_0^t Y_{kj}(u) e^{\beta_j^T Z_{kj}(u)} d\tilde{\mu}_{0j}(u).
\]

**Theorem 4.2** For \(j = 1, 2\), \(\tilde{\mu}_{0j}(t) \equiv \tilde{\mu}_{0j}(t, \tilde{\beta}_j)\) converges almost surely to \(\mu_{0j}(t)\) in \(t \in [0, \tau]\), and \(\sqrt{N}\{\tilde{\mu}_{0j}(t) - \mu_{0j}(t)\}\) converges weakly to a Gaussian process with mean zero and covariance function given by

\[
\Gamma_j^c(s, t) = E[\phi_{kj}^c(s) \phi_{kj}^c(t)] \quad \text{at} \quad (s, t),
\]

where

\[
\phi_{kj}^c(t) = \int_0^t \frac{dM_{kj}^c(u; \beta_j)}{s_{kj}^2(u, \beta_j)} - H^T(t; \beta_j)(A_j^c)^{-1} \int_0^\tau \{Z_{kj}(u) - \tilde{z}_{kj}(u, \beta_j)\} d\tilde{M}_{kj}(u), \quad k = 1, \ldots, N \tag{4.6}
\]
and

\[ H(t; \beta_j) = \int_0^t \tilde{z}_j(u, \beta_j) d\mu_0(u) \] (4.7)

The covariance function \( \Gamma_j^c(s, t) \) can be consistently estimated by

\[ \tilde{\Gamma}_j(s, t) = N^{-1} \sum_{k=1}^N \tilde{\phi}_{kj}(s)\tilde{\phi}_{kj}(t) \] (4.8)

where

\[ \tilde{\phi}_{kj}(t) = \int_0^t \frac{dM_{kj}(u)}{\tilde{S}_j^0(u, \beta_j)} - \tilde{H}_j^T(t; \bar{\beta}_j) \tilde{A}_j^{-1} \int_0^\tau \{ Z_{kj}(u) - \tilde{Z}_{kj}(u, \tilde{\beta}_j) \} d\tilde{M}_{kj}(u) \]

and

\[ \tilde{H}(t; \bar{\beta}_j) = \int_0^t \tilde{Z}_j^T(u, \bar{\beta}_j) \frac{d\tilde{N}_j(u)}{NS_j^0(u, \bar{\beta}_j)}. \]

4.1.2 Estimation of the rate ratio

In the second stage, the dependence parameter can be estimated by the root to the following estimation equation

\[ U^c\left( \theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot) \right) = \sum_{k=1}^N U_k^c\left( \theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot) \right), \] (4.9)

where

\[ U_k^c\left( \theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot) \right) = \left\{ \int_0^\tau \int_0^\tau \frac{\partial \rho(\theta, s, t)}{\partial \theta} \right. \]

\[ \left. \{ dN_{k1}(s) dN_{k2}(t) - \rho(\theta, s, t)Y_{k1}(s)e^{\beta_1^TZ_{k1}(s)} d\mu_{01}(s)Y_{k2}(t)e^{\beta_2^TZ_{k2}(s)} d\mu_{02}(t) \right\}. \] (4.10)
with $\beta_1, \beta_2, \mu_01(\cdot), \mu_02(\cdot)$ replaced by estimator $\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\mu}_0(\cdot), \tilde{\mu}_02(\cdot)$ from the first stage.

The resulting estimator $\tilde{\theta}$ does not have an explicit form. We adapt the asymptotic properties of $\beta_j$ and $\mu_j(\cdot)$ from Theorem 4.1 and Theorem 4.2 to show the weak convergence of $\tilde{\theta}$.

**Theorem 4.3** $N^{-1/2}\left\{ U^c\left( \theta, \tilde{\beta}_1, \tilde{\mu}_0(\cdot), \tilde{\beta}_2, \tilde{\mu}_02(\cdot) \right) - U^c\left( \theta, \beta_1, \mu_01(\cdot), \beta_2, \mu_02(\cdot) \right) \right\}$ follows a mean zero Gaussian process and has the following approximation

$$N^{-1/2}\left\{ U^c(\theta, \tilde{\beta}_1, d\mu_0(\cdot), \tilde{\beta}_2, d\mu_02(\cdot)) - U^c(\theta, \beta_1, d\mu_01(\cdot), \beta_2, d\mu_02(\cdot)) \right\}$$

$$= N^{-1/2} \sum_{k=1}^{N} \left\{ h_{1,N}^c(A_1)^{-1} \xi_{k1} + g_{1,N,k}^c + h_{2,N}^c(A_2)^{-1} \xi_{k2} + g_{2,N,k}^c \right\} + o_p(N^{-1/2}), \quad (4.11)$$

where

$$q_i^c(\theta, s, t) = -\rho(\theta, s, t) \frac{\partial \rho(\theta, s, t)}{\partial \theta} Y_{11}(s) e^{\beta_1^T Z_{11}(s)} Y_{12}(t) e^{\beta_2^T Z_{12}(t)}$$

$$h_{1,N}^c = N^{-1} \sum_{l=1}^{N} \int_0^T \int_0^T q_i^c(\theta, s, t) Z_{11}^T(s) d\mu_01(s) d\mu_02(t)$$

$$h_{2,N}^c = N^{-1} \sum_{l=1}^{N} \int_0^T \int_0^T q_i^c(\theta, s, t) Z_{12}^T(s) d\mu_02(t) d\mu_01(s)$$

$$g_{1,N,k}^c = N^{-1} \sum_{l=1}^{N} \int_0^T \int_0^T q_i^c(\theta, s, t) d\mu_02(t) d\phi_{k1}^c(s)$$

$$g_{2,N,k}^c = N^{-1} \sum_{l=1}^{N} \int_0^T \int_0^T q_i^c(\theta, s, t) d\mu_01(s) d\phi_{k2}^c(t). \quad (4.12)$$

The right hand side in equation (4.11) can be estimated by

$$N^{-1/2} \sum_{k=1}^{N} \left\{ \tilde{h}_{1,N} \tilde{A}_1^{-1} \tilde{\xi}_{k1} + \tilde{g}_{1,N,k} + \tilde{h}_{2,N} \tilde{A}_2^{-1} \tilde{\xi}_{k2} + \tilde{g}_{2,N,k} \right\},$$

where $\tilde{h}_{1,N}, \tilde{\xi}_{k1}, \tilde{h}_{2,N}, \tilde{\xi}_{k2}, \tilde{g}_{1,N,k}, \tilde{g}_{2,N,k}$ are obtained by plugging $\tilde{\beta}_j, \tilde{\theta}, \tilde{\mu}_0j(\cdot)$ and
\[ \hat{\phi}_{kj}(t) \] into equation (4.12).

**Theorem 4.4** We show in the Appendix that \( \sqrt{N}(\hat{\theta} - \theta) \) is asymptotically normal and has the following i.i.d. approximation:

\[
\sqrt{N}(\hat{\theta} - \theta) = N^{-1/2}\{I^c(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot))\}^{-1} \sum_{k=1}^{N} W^c_k(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) + o_p(1) \quad (4.13)
\]

where

\[
I^c(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) = -N^{-1} \sum_{k=1}^{N} \left\{ \frac{\partial U^c_k(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot))}{\partial \theta} \right\}^T, \quad (4.14)
\]

and

\[
W^c_k(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) = U^c_k(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) + \left\{ h_{1,N}^c(A_1^c)^{-1}\xi_{k1}^c + g_{1,N,k}^c + h_{2,N}^c(A_2^c)^{-1}\xi_{k2}^c + g_{2,N,k}^c \right\}.
\]

By the central limit theorem \( \sqrt{N}(\hat{\theta} - \theta) \) is asymptotically normal with mean 0 and variance which can be estimated by \( \tilde{\Phi} = N^{-1}(\tilde{I})^{-1}(\sum_{k=1}^{N} \tilde{W}_k^{\otimes 2})(\tilde{I}^T)^{-1} \), where \( \tilde{I} \) and \( \tilde{W}_k \) are the empirical counterparts of

\[
I^c(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot))
\]

\[
W^c_k(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot))
\]

respectively, obtained by substituting \( \tilde{\theta}, \tilde{\beta}_1, \tilde{\mu}_{01}(\cdot), \tilde{\beta}_2, \tilde{\mu}_{02}(\cdot) \) into equation (4.14) and (4.15).
4.1.3 Simulation studies

To evaluate the performance of the proposed method, we conduct a finite sample simulation study with some shared settings. The end of study time is set as $\tau = 4$, censoring time follows uniform(3, 4), and covariates $\{Z_{kj}\}$ for the two types of disease are generated from uniform(1, 2).

(I) Constant Rate Ratio

Under the shared random effect model, $E[dN^*_k(t)\mid R_k, Z_{kj}(t)] = R_k\{e^{\beta_j Z_{kj}(t)} d\mu_0(t)\}$, where $\{R_k\}$ is the cluster level random effect, and are assumed to be i.i.d from a positive distribution with mean $E(R_k) = 1$ and variance $\text{var}(R_k) = \sigma^2$. Proved in Proposition 1 that the Rate Ratio is reduced to $\rho(\theta) = 1 + \sigma^2$, which only related to the variance of random effect $R_k$.

Let $\beta_1 = 0.2$, $\beta_2 = 0.4$. Take $\mu_{01}(t) = \mu_{02}(t) =$ $0.125t^2$, $0.25t^2$, $0.375t^2$, and $0.5t^2$ such that the averaged observed type 1(2) events after right censoring are $2.06(2.84)$, $4.18(5.67)$, $6.25(8.48)$, $8.3(11.3)$ respectively. $R_k$ are independently simulated from a Gamma distribution with mean 1 and variance $\sigma^2 = 0.25, 0.5, 0.75$, which leads to $\rho = 1.25, 1.5, 1.75$.

In the first-stage, we estimate $\beta_1, \beta_2$ based marginal mean rate model (4.1). From the result in Table 10, $\hat{\beta}_1$ and $\hat{\beta}_2$ converges to true the values $\beta_1 = 0.2$ and $\beta_2 = 0.4$, and the ESE (Estimated Standard Error) is close to SEE (Standard Error of Estimates). The empirical coverage probability is close to its 95% nominee value.

In the second-stage, we substitute $\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\mu}_{01}(\cdot), \tilde{\mu}_{02}(\cdot)$ into the estimation equation (4.9) and obtain $\tilde{\theta}$ by solving $U(\theta, \tilde{\beta}_1, \tilde{\mu}_{01}(\cdot), \tilde{\beta}_2, \tilde{\mu}_{02}(\cdot)) = 0$. The average Bias, SEE
(Standard Error of Estimates), ESE (Estimated Standard Error), CP (coverage probability of the 95% confidence interval) of $\rho$ are summarized in Table 11, where each entry based on 1000 replicates.

$\tilde{\rho}$ is unbiased and the estimated standard error can be reduced by increasing the sample size. Similar to the estimation result shown in table 3, the standard error is underestimated which cause the coverage probability consistently slightly smaller than 95%, especially when the $\rho$ increases. One possible explanation is that the information gains from increasing the sample size is offset by the stronger association between two recurrent event processes. An extreme condition is that the two processes are identical, then we are actually observing and utilizing the information for a single process and therefore the rate ratio would be underestimated.

(II) Time varying Rate Ratio

Assume the counting process for $j$ th type event in cluster $K$ at time $u$ as

$$N_{kj}(t) = \tilde{N}_{kj}(t) + N_{k0}(t), \text{ for } j = 1, 2$$

where $\{\tilde{N}_{kj}(t)\}$, and $\{N_{k0}(t)\}$ are mutually independent. Denote $\rho_0(\theta, s, t)$ be the rate ratio of $N_{k0}(s)$ and $N_{k0}(t)$. By proposition 3, we have the rate ratio of $\{N_{k1}(s), N_{k2}(t)\}$ as

$$\rho(\theta, s, t | z_1(s), z_2(t)) = 1 + \{\rho_0(\theta, s, t) - 1\} \frac{\lambda_{k0}(s | z_1(s))\lambda_{k0}(t | z_2(t))}{\lambda_{k1}(s | z_1(s))\lambda_{k2}(t | z_2(t))},$$

where $E\{dN_{k0}(s) | z_1\} = \lambda_{k0}(s | z_1(s)) \, ds$, $E\{dN_{k1}(s) | z_1(s)\} = \lambda_{k1}(s | z_1(s)) \, ds$, while $E\{dN_{k2}(t) | z_2(t)\} = \lambda_{k2}(t | z_2(t)) \, dt$. It is straight forward to show $\lambda_{k1}(s | z_1(s)) = \lambda_{k0}(s | z_1(s)) + \tilde{\lambda}_{k1}(s | z_1(s))$, with $\tilde{\lambda}_{k1}(s | z_1(s))$ be the mean event rate for counting
process $\tilde{N}_{k1}(s)$. The same logic applies to type 2 event.

For simulation, we start with a simple model by letting $\tilde{\lambda}_{k1}(s|z_1(s)) = m(s)\lambda_{k0}(s|z_1(s))$ and $\tilde{\lambda}_{k1}(t|z_2(t)) = m(t)\lambda_{k0}(t|z_2(t))$, where $m(s), m(t) > 0$, for $s, t \in [0, \tau]$. Therefore the rate ratio would be

$$\rho(\theta, s, t) = 1 + \left\{ \rho_0(\theta, s, t) - 1 \right\} \frac{1}{(1 + m(s))(1 + m(t))}.$$ 

By specifying $m(\cdot)$, the rate ratio could be designed to be time varying under certain patterns. Here we let $1/(1+m(s)) = (-0.15s+0.9)$ and $1/(1+m(t)) = (-0.15t+0.9)$.

To specify the $\rho_0(\theta, s, t)$, we assume that

$$E[dN_{k0}^*(s)|R_k, Z_{k1}(s)] = R_k \cdot \{ e^{\beta_1^T Z_{k1}(s)} d\mu_{01}(s) \},$$

and

$$E[dN_{k0}^*(t)|R_k, Z_{k2}(t)] = R_k \cdot \{ e^{\beta_2^T Z_{k2}(t)} d\mu_{02}(t) \},$$

where $R_k$ is the cluster level random effect which is independent and identically from a positive distribution. The coefficient of covariates for type 1 and type 2 events are $\beta_1 = \beta_2 = 0$. $R_k$ are generated from Gamma distribution with mean 1 and variance 0.25, 0.5, 1, 1.5, and 2, and therefore $\rho_0(\theta, s, t) = 1.25, 1.5, 2$ and 2.5. The rate ratio $\rho(\theta, s, t)$ can be represented as

$$\rho(\theta, s, t) = 1 + \theta(-0.15s + 0.9)(-0.15t + 0.9),$$

with the parameter $\theta$ equal to 0.25, 0.5, 1 and 1.5.

A simulation study for the Rate Ratio with sample size $K = 200, 500, 800$ is summarized in Table 12, with each entry based on 1000 simulations. The estimator is
unbiased and the estimated standard error is very close to its true value, with coverage probability around 95%. The SSE and ESE is decreasing while increasing the sample size showing that the estimation procedure is more efficient with a large sample size. We observe consistently higher standard error when the association between bivariate recurrent processes increases.

(III) Covariate Dependent Rate Ratio

Let \( Z_{kj} = Z_k \) denote the cluster level covariates. Assume the shared Frailty model

\[
E[dN_{kj}^*(t)|Z_k, R_k] = R_k \cdot e^{\beta_{0j} Z_k(t)} d\mu_{0j}(t) \tag{4.17}
\]

where \( E[R_k|Z_k] = \mu(Z_k) \) and \( \text{var}[R_k|Z_k] = \sigma^2(Z_k) \). Following Proposition 1, \( \rho(s, t, \theta) = 1 + \frac{\sigma^2(Z_k)}{\mu^2(Z_k)} \). Denoted by the \( \theta_1 \) and \( \theta_2 \) the value of \( \rho(s, t, \theta) \) when \( Z_k = 1, 0 \), i.e.

\[
\rho(s, t, \theta) = \theta_1 I(Z_k = 1) + \theta_2 I(Z_k = 0). \tag{4.18}
\]

Let \( \beta_{01} = 0.2 \), \( \beta_{02}(t) = 0.4 \). We consider \( \mu_{0j}(t) = 0.125t^2, 0.25t^2 \) for moderately observed event process, whereas \( \mu_{0j}(t) = 0.375t^2 \) and \( 0.5t^2 \) stand for more frequently observed ones. \( Z_k \) from Bernoulli(\( p = 0.5 \)) and \( R_k \) from Gamma(\( 1/v_k, v_k \)) so that \( E[R_k] = 1 \) and \( \text{var}[R_k] = v_k \). To represent the weak and the strong association, consider \( v_k \) equal to 0.25 and 0.75 for \( Z_k = 1 \) and \( Z_k = 0 \) respectively which gives us \( \theta_1 = 1.25 \) and \( \theta_2 = 1.75 \) correspondingly.

Simulation result for sample size 200, 500, 800 and 1100, each with 1000 replicates are shown in Table 13. The estimator is unbiased and the ESE is close to SEE. The coverage probability is approaching to 0.95 when the sample size increases from 200 to 1100. The ESE and SEE of \( \theta_2 \) are consistently larger than that of \( \theta_1 \), even through
both are reduced in a larger sample size.

**IV) Time and Covariate Dependent Rate Ratio**

For \( j = 1, 2 \), we construct a bivariate counting process \( N_{kj} \) with \( N_{kj}(t) = \tilde{N}_{kj}(t) + N_{k0}(t) \). Let

\[
E\{dN_{k0}(t)|Z_{kj}, R_k\} = \lambda_{k0}(t|Z_{kj}, R_k) \, dt
\]

\[
E\{d\tilde{N}_{kj}(t)|Z_{kj}, R_k\} = \tilde{\lambda}_{kj}(t) \, dt
\]

where \( \lambda_{k0}(t|Z_{kj}, R_k) = R_k e^{\beta_0 j Z_{kj} 0.25 t} \) and \( \tilde{\lambda}_{kj}(t) = 0.25 \).

We take \( R_k \) from i.i.d Gamma\((a, b)\) with \((a, b)\) equal to \((4, 0.25), (2, 0.5), (1.33, 0.75)\) and \((1, 1)\) such that \( \rho_0(\theta, s, t) = 1.25, 1.5, 1.75 \) and \( 2 \). Let \( Z_k \) is from Bernoulli\((0.5)\), \( \beta_{01} = 0.1 \) and \( \beta_{02} = 0.2 \). By Proposition 3, the rate ratio of \( N_{k1}(s) \) and \( N_{k2}(t) \) is time-varying and dependent on the covariate \( Z_{kj} \) which is denoted by

\[
\rho(\theta, s, t|Z_{k1}, Z_{k2}) = 1 + \theta \frac{(0.25 t e^{0.1 Z_{k1}})(0.25 s e^{0.2 Z_{k2}})}{(0.25 + 0.25 t e^{0.1 Z_{k1}})(0.25 + 0.25 s e^{0.2 Z_{k2}})}, \quad (4.19)
\]

where \( \theta = \rho_0(\theta, s, t) - 1 = 0.25, 0.5, 0.75 \) and \( 1 \).

To evaluate the performance difference between moderate and high frequency event processes, we consider \( \lambda_{k0}(t|Z_{kj}, R_k) = R_k \cdot 0.5 e^{\beta_0 j Z_{kj}} \). While keeping other settings the same, the event process \( N_{kj}(t) \) would expect to have more observations than the previous setting and following equation (4.19) we have

\[
\rho(\theta, s, t|Z_k) = 1 + \theta \frac{(0.5 t e^{0.1 Z_k})(0.5 s e^{0.2 Z_k})}{(0.25 + 0.5 t e^{0.1 Z_k})(0.25 + 0.5 s e^{0.2 Z_k})}. \quad (4.20)
\]

The simulation result from Table 14 shows that the estimating procedure works well for both settings. The bias is going to zero and the ESE is getting close to SSE.
as sample size increase. The coverage probability is getting around 95% for both $\theta$. 
Table 10: Scenario I: Numerical results for \((\beta_1, \beta_2)\) with true value equals \((0.2, 0.4)\). Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP summarized for \((\beta_1, \beta_2)\). Each entry is based on 1000 simulated datasets under shared random effect model with Multiplicative marginals.

<table>
<thead>
<tr>
<th>(\mu_j(t))</th>
<th>(\rho)</th>
<th>(K)</th>
<th>bias of ((\hat{\beta}_1, \hat{\beta}_2))</th>
<th>SEE ((\hat{\beta}_1, \hat{\beta}_2))</th>
<th>ESE((\hat{\beta}_1, \hat{\beta}_2))</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25t²</td>
<td>1.25</td>
<td>200</td>
<td>(-0.0063, 0.0019)</td>
<td>(0.1679, 0.1569)</td>
<td>(0.1710, 0.1600)</td>
<td>(0.9620, 0.9550)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(0.0042, 0.0021)</td>
<td>(0.1080, 0.1037)</td>
<td>(0.1088, 0.1021)</td>
<td>(0.9480, 0.9490)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(0.0032, -0.0046)</td>
<td>(0.0871, 0.0822)</td>
<td>(0.0861, 0.0807)</td>
<td>(0.9480, 0.9420)</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>200</td>
<td>(0.0061, 0.0092)</td>
<td>(0.2139, 0.2072)</td>
<td>(0.2107, 0.2028)</td>
<td>(0.9460, 0.9410)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(-0.0075, 0.0092)</td>
<td>(0.1335, 0.1320)</td>
<td>(0.1341, 0.1286)</td>
<td>(0.9470, 0.9470)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(0.0029, 0.0059)</td>
<td>(0.1100, 0.1050)</td>
<td>(0.1061, 0.1020)</td>
<td>(0.9410, 0.9430)</td>
</tr>
<tr>
<td></td>
<td>1.75</td>
<td>200</td>
<td>(-0.0199, 0.0017)</td>
<td>(0.2457, 0.2394)</td>
<td>(0.2428, 0.2365)</td>
<td>(0.9430, 0.9500)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(-0.0066, 0.0016)</td>
<td>(0.1589, 0.1529)</td>
<td>(0.1545, 0.1507)</td>
<td>(0.9390, 0.9480)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(-0.0066, 0.0011)</td>
<td>(0.1236, 0.1192)</td>
<td>(0.1227, 0.1190)</td>
<td>(0.9500, 0.9560)</td>
</tr>
<tr>
<td>0.375t²</td>
<td>1.25</td>
<td>200</td>
<td>(0.0068, -0.0028)</td>
<td>(0.1565, 0.1571)</td>
<td>(0.1571, 0.1506)</td>
<td>(0.9540, 0.9420)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(-0.0039, 0.0042)</td>
<td>(0.0990, 0.0959)</td>
<td>(0.0996, 0.0950)</td>
<td>(0.9450, 0.9480)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(-0.0004, 0.0013)</td>
<td>(0.0816, 0.0751)</td>
<td>(0.0790, 0.0751)</td>
<td>(0.9470, 0.9550)</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>200</td>
<td>(0.0071, 0.0017)</td>
<td>(0.2106, 0.1958)</td>
<td>(0.1987, 0.1936)</td>
<td>(0.9350, 0.9500)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(0.0012, 0.0011)</td>
<td>(0.1237, 0.1316)</td>
<td>(0.1267, 0.1228)</td>
<td>(0.9600, 0.9460)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(0.0023, -0.0018)</td>
<td>(0.1041, 0.0966)</td>
<td>(0.1005, 0.0975)</td>
<td>(0.9450, 0.9600)</td>
</tr>
<tr>
<td></td>
<td>1.75</td>
<td>200</td>
<td>(-0.0046, 0.0038)</td>
<td>(0.2314, 0.2376)</td>
<td>(0.2318, 0.2279)</td>
<td>(0.9510, 0.9320)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(-0.0051, 0.0033)</td>
<td>(0.1499, 0.1514)</td>
<td>(0.1482, 0.1456)</td>
<td>(0.9460, 0.9370)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(-0.0002, -0.0054)</td>
<td>(0.1213, 0.1199)</td>
<td>(0.1174, 0.1156)</td>
<td>(0.9460, 0.9480)</td>
</tr>
<tr>
<td>0.5t²</td>
<td>1.25</td>
<td>200</td>
<td>(-0.0028, -0.0030)</td>
<td>(0.1531, 0.1462)</td>
<td>(0.1489, 0.1433)</td>
<td>(0.9460, 0.9480)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(0.0005, 0.0007)</td>
<td>(0.0943, 0.0941)</td>
<td>(0.0945, 0.0912)</td>
<td>(0.9480, 0.9410)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(0.0021, 0.0028)</td>
<td>(0.0748, 0.0710)</td>
<td>(0.0751, 0.0722)</td>
<td>(0.9490, 0.9510)</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>200</td>
<td>(0.0095, -0.0104)</td>
<td>(0.1940, 0.1907)</td>
<td>(0.1929, 0.1884)</td>
<td>(0.9510, 0.9460)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(0.0009, -0.0004)</td>
<td>(0.1239, 0.1244)</td>
<td>(0.1229, 0.1199)</td>
<td>(0.9470, 0.9390)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(-0.0015, 0.0003)</td>
<td>(0.0940, 0.0941)</td>
<td>(0.0975, 0.0953)</td>
<td>(0.9600, 0.9540)</td>
</tr>
<tr>
<td></td>
<td>1.75</td>
<td>200</td>
<td>(-0.0074, 0.0031)</td>
<td>(0.2372, 0.2272)</td>
<td>(0.2270, 0.2247)</td>
<td>(0.9340, 0.9460)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>(0.0023, 0.0046)</td>
<td>(0.1473, 0.1486)</td>
<td>(0.1456, 0.1438)</td>
<td>(0.9400, 0.9400)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>(0.0063, 0.0013)</td>
<td>(0.1119, 0.1109)</td>
<td>(0.1152, 0.1136)</td>
<td>(0.9550, 0.9560)</td>
</tr>
</tbody>
</table>
Table 11: Scenario I - Estimation of $\rho$ with summary of Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP. Each entry is based on 1000 simulations under shared random effect model with Multiplicative marginals

<table>
<thead>
<tr>
<th>$\mu_0(t)$</th>
<th>$\rho$</th>
<th>$K$</th>
<th>Bias</th>
<th>SEE</th>
<th>ESE</th>
<th>CP</th>
<th>$\mu_0(t)$</th>
<th>$\rho$</th>
<th>Bias</th>
<th>SEE</th>
<th>ESE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.125$t^2$</td>
<td>1.25</td>
<td>200</td>
<td>-0.0041</td>
<td>0.0911</td>
<td>0.0878</td>
<td>0.9210</td>
<td>0.375$t^2$</td>
<td>1.25</td>
<td>200</td>
<td>-0.0010</td>
<td>0.0411</td>
<td>0.0386</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td></td>
<td>0.0004</td>
<td>0.0588</td>
<td>0.0572</td>
<td>0.9420</td>
<td></td>
<td>500</td>
<td></td>
<td>-0.0003</td>
<td>0.0257</td>
<td>0.0252</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td></td>
<td>-0.0002</td>
<td>0.0444</td>
<td>0.0450</td>
<td>0.9530</td>
<td></td>
<td>800</td>
<td></td>
<td>-0.0002</td>
<td>0.0206</td>
<td>0.0199</td>
</tr>
<tr>
<td>1.50</td>
<td>200</td>
<td></td>
<td>-0.0055</td>
<td>0.1273</td>
<td>0.1199</td>
<td>0.9190</td>
<td></td>
<td>1.50</td>
<td>200</td>
<td>-0.0045</td>
<td>0.0742</td>
<td>0.0702</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td></td>
<td>-0.0046</td>
<td>0.0816</td>
<td>0.0789</td>
<td>0.9290</td>
<td></td>
<td>500</td>
<td></td>
<td>-0.0031</td>
<td>0.0493</td>
<td>0.0459</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td></td>
<td>-0.0028</td>
<td>0.0646</td>
<td>0.0630</td>
<td>0.9380</td>
<td></td>
<td>800</td>
<td></td>
<td>-0.0010</td>
<td>0.0388</td>
<td>0.0370</td>
</tr>
<tr>
<td>1.75</td>
<td>200</td>
<td></td>
<td>-0.0082</td>
<td>0.1593</td>
<td>0.1556</td>
<td>0.9170</td>
<td></td>
<td>1.75</td>
<td>200</td>
<td>-0.0094</td>
<td>0.1081</td>
<td>0.1025</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td></td>
<td>-0.0007</td>
<td>0.1109</td>
<td>0.1042</td>
<td>0.9260</td>
<td></td>
<td>500</td>
<td></td>
<td>-0.0051</td>
<td>0.0733</td>
<td>0.0697</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td></td>
<td>-0.0056</td>
<td>0.0838</td>
<td>0.0825</td>
<td>0.9410</td>
<td></td>
<td>800</td>
<td></td>
<td>-0.0014</td>
<td>0.0572</td>
<td>0.0560</td>
</tr>
<tr>
<td>0.25$t^2$</td>
<td>1.25</td>
<td>200</td>
<td>-0.0022</td>
<td>0.0443</td>
<td>0.0432</td>
<td>0.9220</td>
<td>0.5$t^2$</td>
<td>1.25</td>
<td>200</td>
<td>-0.0006</td>
<td>0.0370</td>
<td>0.0362</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td></td>
<td>-0.0005</td>
<td>0.0285</td>
<td>0.0278</td>
<td>0.9370</td>
<td></td>
<td>500</td>
<td></td>
<td>-0.0004</td>
<td>0.0235</td>
<td>0.0234</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td></td>
<td>-0.0009</td>
<td>0.0228</td>
<td>0.0223</td>
<td>0.9440</td>
<td></td>
<td>800</td>
<td></td>
<td>-0.0010</td>
<td>0.0192</td>
<td>0.0186</td>
</tr>
<tr>
<td>1.50</td>
<td>200</td>
<td></td>
<td>0.0002</td>
<td>0.0850</td>
<td>0.0759</td>
<td>0.9100</td>
<td></td>
<td>1.50</td>
<td>200</td>
<td>-0.0040</td>
<td>0.0725</td>
<td>0.0673</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td></td>
<td>0.0015</td>
<td>0.0512</td>
<td>0.0492</td>
<td>0.9440</td>
<td></td>
<td>500</td>
<td></td>
<td>-0.0015</td>
<td>0.0454</td>
<td>0.0446</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td></td>
<td>0.0008</td>
<td>0.0382</td>
<td>0.0391</td>
<td>0.9480</td>
<td></td>
<td>800</td>
<td></td>
<td>0.0013</td>
<td>0.0376</td>
<td>0.0362</td>
</tr>
<tr>
<td>1.75</td>
<td>200</td>
<td></td>
<td>-0.0082</td>
<td>0.1212</td>
<td>0.1078</td>
<td>0.8970</td>
<td></td>
<td>1.75</td>
<td>200</td>
<td>-0.0057</td>
<td>0.1088</td>
<td>0.1020</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td></td>
<td>-0.0063</td>
<td>0.0743</td>
<td>0.0715</td>
<td>0.9230</td>
<td></td>
<td>500</td>
<td></td>
<td>0.0005</td>
<td>0.0715</td>
<td>0.0686</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td></td>
<td>-0.0020</td>
<td>0.0613</td>
<td>0.0589</td>
<td>0.9340</td>
<td></td>
<td>800</td>
<td></td>
<td>0.0008</td>
<td>0.0554</td>
<td>0.0556</td>
</tr>
</tbody>
</table>
Table 12: Scenario II - Estimation of $\theta$ in $\rho(s, t, \theta) = 1 + \theta(-0.15t + 0.9)(-0.15s + 0.9)$. The summary of Bias, SEE (Standard Error of Estimates), ESE(Estimated Standard Error) and CP (Coverage Probability). The Marginal model is multiplicative and the parametric form of $\rho(s, t, \theta)$ is correctly specified. Each entry is based on 1000 simulations.

<table>
<thead>
<tr>
<th>$\mu_{0j}(t)$</th>
<th>$\theta$</th>
<th>K</th>
<th>Bias</th>
<th>SEE</th>
<th>ESE</th>
<th>CP</th>
<th>$\mu_{0j}(t)$</th>
<th>$\theta$</th>
<th>K</th>
<th>Bias</th>
<th>SEE</th>
<th>ESE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.125t$^2$</td>
<td>0.25</td>
<td>200</td>
<td>-0.0042</td>
<td>0.1168</td>
<td>0.1140</td>
<td>0.9370</td>
<td>0.375t$^2$</td>
<td>0.25</td>
<td>200</td>
<td>-0.0032</td>
<td>0.0580</td>
<td>0.0579</td>
<td>0.9450</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0013</td>
<td>0.0745</td>
<td>0.0731</td>
<td>0.9480</td>
<td></td>
<td>500</td>
<td>0.0000</td>
<td>0.0395</td>
<td>0.0375</td>
<td>0.9260</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.0000</td>
<td>0.0605</td>
<td>0.0581</td>
<td>0.9370</td>
<td></td>
<td>800</td>
<td>-0.0010</td>
<td>0.0299</td>
<td>0.0298</td>
<td>0.9570</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>200</td>
<td>-0.0055</td>
<td>0.1527</td>
<td>0.1445</td>
<td>0.9280</td>
<td></td>
<td>0.50</td>
<td>200</td>
<td>-0.0038</td>
<td>0.0921</td>
<td>0.0891</td>
<td>0.9260</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0012</td>
<td>0.0967</td>
<td>0.0940</td>
<td>0.9450</td>
<td></td>
<td>500</td>
<td>-0.0038</td>
<td>0.0591</td>
<td>0.0587</td>
<td>0.9310</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>-0.0029</td>
<td>0.0756</td>
<td>0.0745</td>
<td>0.9430</td>
<td></td>
<td>800</td>
<td>0.0003</td>
<td>0.0494</td>
<td>0.0473</td>
<td>0.9340</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>200</td>
<td>-0.0006</td>
<td>0.2424</td>
<td>0.2224</td>
<td>0.9050</td>
<td></td>
<td>1</td>
<td>200</td>
<td>-0.0182</td>
<td>0.1821</td>
<td>0.1673</td>
<td>0.8860</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0045</td>
<td>0.1482</td>
<td>0.1478</td>
<td>0.9350</td>
<td></td>
<td>500</td>
<td>-0.0051</td>
<td>0.1143</td>
<td>0.1123</td>
<td>0.9330</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.0033</td>
<td>0.1212</td>
<td>0.1185</td>
<td>0.9260</td>
<td></td>
<td>800</td>
<td>-0.0032</td>
<td>0.0922</td>
<td>0.0904</td>
<td>0.9310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25t$^2$</td>
<td>0.25</td>
<td>200</td>
<td>-0.0009</td>
<td>0.0713</td>
<td>0.0717</td>
<td>0.9370</td>
<td>0.5t$^2$</td>
<td>0.25</td>
<td>200</td>
<td>0.0004</td>
<td>0.0524</td>
<td>0.0515</td>
<td>0.9380</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0012</td>
<td>0.0454</td>
<td>0.0462</td>
<td>0.9530</td>
<td></td>
<td>500</td>
<td>-0.0011</td>
<td>0.0342</td>
<td>0.0331</td>
<td>0.9360</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>-0.0012</td>
<td>0.0370</td>
<td>0.0368</td>
<td>0.9420</td>
<td></td>
<td>800</td>
<td>0.0004</td>
<td>0.0266</td>
<td>0.0264</td>
<td>0.9410</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>200</td>
<td>-0.0042</td>
<td>0.1075</td>
<td>0.1040</td>
<td>0.9320</td>
<td></td>
<td>0.50</td>
<td>200</td>
<td>-0.0082</td>
<td>0.0890</td>
<td>0.0835</td>
<td>0.9120</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0005</td>
<td>0.0683</td>
<td>0.0678</td>
<td>0.9390</td>
<td></td>
<td>500</td>
<td>-0.0001</td>
<td>0.0556</td>
<td>0.0553</td>
<td>0.9380</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>-0.0027</td>
<td>0.0564</td>
<td>0.0538</td>
<td>0.9390</td>
<td></td>
<td>800</td>
<td>-0.0009</td>
<td>0.0437</td>
<td>0.0438</td>
<td>0.9500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>200</td>
<td>-0.0170</td>
<td>0.1893</td>
<td>0.1781</td>
<td>0.8960</td>
<td></td>
<td>1</td>
<td>200</td>
<td>-0.0058</td>
<td>0.1789</td>
<td>0.1624</td>
<td>0.8860</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0008</td>
<td>0.1201</td>
<td>0.1208</td>
<td>0.9350</td>
<td></td>
<td>500</td>
<td>-0.0037</td>
<td>0.1173</td>
<td>0.1091</td>
<td>0.9240</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>-0.0068</td>
<td>0.1040</td>
<td>0.0963</td>
<td>0.9210</td>
<td></td>
<td>800</td>
<td>-0.0043</td>
<td>0.0895</td>
<td>0.0863</td>
<td>0.9190</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 13: Scenario III - Summary of Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP. The Rate Ratio is covariate dependent, where true values $\rho(\theta, s, t; Z_k) = \theta_1 I(Z_k = 1) + \theta_2 I(Z_k = 0)$, with true value $\theta_1 = 1.25$ and $\theta_2 = 1.75$. Each entry is based on 1000 simulations with correctly specified multiplicative marginals and Rate Ratio form.

<table>
<thead>
<tr>
<th>$\mu_0(t)$</th>
<th>K</th>
<th>Bias</th>
<th>SEE</th>
<th>ESE</th>
<th>CP</th>
<th>Bias</th>
<th>SSE</th>
<th>ESE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.125$t^2$</td>
<td>200</td>
<td>-0.0045</td>
<td>0.0885</td>
<td>0.0843</td>
<td>0.9200</td>
<td>-0.0164</td>
<td>0.2187</td>
<td>0.1880</td>
<td>0.8950</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0020</td>
<td>0.0590</td>
<td>0.0557</td>
<td>0.9420</td>
<td>-0.0139</td>
<td>0.1352</td>
<td>0.1266</td>
<td>0.9120</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.0007</td>
<td>0.0447</td>
<td>0.0444</td>
<td>0.9530</td>
<td>-0.0033</td>
<td>0.1069</td>
<td>0.1039</td>
<td>0.9320</td>
</tr>
<tr>
<td></td>
<td>1100</td>
<td>0.0006</td>
<td>0.0378</td>
<td>0.0382</td>
<td>0.9560</td>
<td>0.0007</td>
<td>0.0950</td>
<td>0.0899</td>
<td>0.9400</td>
</tr>
<tr>
<td>0.25$t^2$</td>
<td>200</td>
<td>-0.0040</td>
<td>0.0639</td>
<td>0.0608</td>
<td>0.9170</td>
<td>-0.0174</td>
<td>0.1691</td>
<td>0.1527</td>
<td>0.8770</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0010</td>
<td>0.0440</td>
<td>0.0401</td>
<td>0.9270</td>
<td>-0.0007</td>
<td>0.1149</td>
<td>0.1061</td>
<td>0.9210</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>-0.0009</td>
<td>0.0326</td>
<td>0.0319</td>
<td>0.9410</td>
<td>-0.0007</td>
<td>0.0934</td>
<td>0.0860</td>
<td>0.9150</td>
</tr>
<tr>
<td></td>
<td>1100</td>
<td>-0.0009</td>
<td>0.0278</td>
<td>0.0278</td>
<td>0.9490</td>
<td>-0.0032</td>
<td>0.0826</td>
<td>0.0737</td>
<td>0.9240</td>
</tr>
<tr>
<td>0.375$t^2$</td>
<td>200</td>
<td>-0.0039</td>
<td>0.0563</td>
<td>0.0532</td>
<td>0.9150</td>
<td>-0.0129</td>
<td>0.1622</td>
<td>0.1434</td>
<td>0.8900</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0029</td>
<td>0.0364</td>
<td>0.0347</td>
<td>0.9250</td>
<td>-0.0050</td>
<td>0.1048</td>
<td>0.0983</td>
<td>0.9280</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.0002</td>
<td>0.0292</td>
<td>0.0283</td>
<td>0.9310</td>
<td>-0.0022</td>
<td>0.0846</td>
<td>0.0804</td>
<td>0.9310</td>
</tr>
<tr>
<td></td>
<td>1100</td>
<td>0.0011</td>
<td>0.0258</td>
<td>0.0245</td>
<td>0.9380</td>
<td>-0.0019</td>
<td>0.0713</td>
<td>0.0698</td>
<td>0.9390</td>
</tr>
<tr>
<td>0.5$t^2$</td>
<td>200</td>
<td>-0.0035</td>
<td>0.0530</td>
<td>0.0499</td>
<td>0.9150</td>
<td>-0.0142</td>
<td>0.1643</td>
<td>0.1399</td>
<td>0.8820</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.0025</td>
<td>0.0335</td>
<td>0.0326</td>
<td>0.9360</td>
<td>-0.0014</td>
<td>0.1041</td>
<td>0.0959</td>
<td>0.9310</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>-0.0007</td>
<td>0.0268</td>
<td>0.0263</td>
<td>0.9490</td>
<td>-0.0029</td>
<td>0.0853</td>
<td>0.0784</td>
<td>0.9190</td>
</tr>
<tr>
<td></td>
<td>1100</td>
<td>-0.0003</td>
<td>0.0229</td>
<td>0.0227</td>
<td>0.9470</td>
<td>-0.0012</td>
<td>0.0693</td>
<td>0.0668</td>
<td>0.9250</td>
</tr>
</tbody>
</table>
Table 14: Scenario IV - estimates $\theta_1$, $\theta_2$ in the underline models where $\rho(\theta, s, t|Z_k) = 1 + \theta_1 \frac{(0.25e^{0.1Z_k})}{(0.25+0.25e^{0.1Z_k})(0.25+0.25e^{0.2Z_k})}$ and $\rho(\theta, s, t|Z_k) = 1 + \theta_2 \frac{(0.5e^{0.1Z_k})}{(0.25+0.5e^{0.1Z_k})(0.25+0.5e^{0.2Z_k})}$. With the true values of $\theta_1, \theta_2$ equal to 0.25, 0.5, 0.75 and 1.00. Summary of Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP of $\theta$ where each entry is based on 1000 simulations. The averaged observed events for type 1(2) event is 2.44(2.56)

<table>
<thead>
<tr>
<th>$\theta_1$</th>
<th>N</th>
<th>Bias</th>
<th>SEE</th>
<th>ESE</th>
<th>CP</th>
<th>$\theta_2$</th>
<th>Bias</th>
<th>SEE</th>
<th>ESE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>200</td>
<td>-0.0020</td>
<td>0.0804</td>
<td>0.0798</td>
<td>0.9420</td>
<td>0.25</td>
<td>-0.0036</td>
<td>0.0523</td>
<td>0.0498</td>
<td>0.9280</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0012</td>
<td>0.0508</td>
<td>0.0518</td>
<td>0.9550</td>
<td>0.0000</td>
<td>0.0335</td>
<td>0.0328</td>
<td>0.9360</td>
<td></td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.0009</td>
<td>0.0420</td>
<td>0.0411</td>
<td>0.9490</td>
<td>-0.0002</td>
<td>0.0275</td>
<td>0.0260</td>
<td>0.9390</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1100</td>
<td>-0.0006</td>
<td>0.0345</td>
<td>0.0350</td>
<td>0.9500</td>
<td>0.0010</td>
<td>0.0232</td>
<td>0.0224</td>
<td>0.9490</td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>200</td>
<td>-0.0012</td>
<td>0.1237</td>
<td>0.1130</td>
<td>0.9170</td>
<td>0.50</td>
<td>-0.0015</td>
<td>0.0908</td>
<td>0.0835</td>
<td>0.9220</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0004</td>
<td>0.0763</td>
<td>0.0735</td>
<td>0.9260</td>
<td>-0.0059</td>
<td>0.0551</td>
<td>0.0531</td>
<td>0.9170</td>
<td></td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>-0.0007</td>
<td>0.0617</td>
<td>0.0590</td>
<td>0.9400</td>
<td>0.0008</td>
<td>0.0448</td>
<td>0.0436</td>
<td>0.9370</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1100</td>
<td>0.0000</td>
<td>0.0488</td>
<td>0.0503</td>
<td>0.9420</td>
<td>-0.0020</td>
<td>0.0377</td>
<td>0.0370</td>
<td>0.9320</td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>200</td>
<td>-0.0094</td>
<td>0.1584</td>
<td>0.1471</td>
<td>0.9090</td>
<td>0.75</td>
<td>-0.0097</td>
<td>0.1286</td>
<td>0.1149</td>
<td>0.8800</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0004</td>
<td>0.1067</td>
<td>0.0993</td>
<td>0.9240</td>
<td>-0.0033</td>
<td>0.0841</td>
<td>0.0793</td>
<td>0.9070</td>
<td></td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>-0.0002</td>
<td>0.0275</td>
<td>0.0260</td>
<td>0.9390</td>
<td>-0.0020</td>
<td>0.0612</td>
<td>0.0628</td>
<td>0.9400</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1100</td>
<td>0.0010</td>
<td>0.0232</td>
<td>0.0224</td>
<td>0.9490</td>
<td>-0.0030</td>
<td>0.0524</td>
<td>0.0539</td>
<td>0.9430</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>200</td>
<td>-0.0015</td>
<td>0.0908</td>
<td>0.0835</td>
<td>0.9220</td>
<td>1.00</td>
<td>-0.0035</td>
<td>0.1797</td>
<td>0.1587</td>
<td>0.8730</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0059</td>
<td>0.0551</td>
<td>0.0531</td>
<td>0.9170</td>
<td>0.0011</td>
<td>0.1079</td>
<td>0.1056</td>
<td>0.9280</td>
<td></td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.0008</td>
<td>0.0448</td>
<td>0.0436</td>
<td>0.9370</td>
<td>-0.0028</td>
<td>0.0833</td>
<td>0.0827</td>
<td>0.9330</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1100</td>
<td>-0.0020</td>
<td>0.0377</td>
<td>0.0370</td>
<td>0.9320</td>
<td>-0.0074</td>
<td>0.0771</td>
<td>0.0719</td>
<td>0.9190</td>
<td></td>
</tr>
</tbody>
</table>
4.2 Hypothesis testing of the rate ratio

4.2.1 Procedure description

For the case that the marginal mean rate model is additive, we developed a supreme test statistic to check the null hypothesis $\rho(s, t, \theta) = \theta$. We apply the same procedure and illustrate the test statistic below for hypothesis testing purposes. Define the residual process under the Multiplicative Marginal Mean Rate Model as

$$V'(s, t, \theta) = N^{-1/2} \sum_{k=1}^{N} \int_{0}^{t} \int_{0}^{s} w(u, v) \frac{\partial \rho(u, v, \theta)}{\partial \theta} \bigg|_{\theta=\theta} \left\{ dN_{k1}(u) dN_{k2}(v) - \rho(u, v, \theta) Y_{k1}(u) d\mu_{01}(u) e^{\beta_{1}^{T} Z_{k1}(u)} \cdot Y_{k2}(v) d\mu_{02}(v) e^{\beta_{2}^{T} Z_{k2}(v)} \right\}.$$

(4.21)

Denote $\tilde{V}(s, t, \tilde{\theta})$ the empirical value of $V'(s, t, \theta)$ as

$$\tilde{V}(s, t, \tilde{\theta}) =$$

$$N^{-1/2} \sum_{k=1}^{N} \int_{0}^{t} \int_{0}^{s} \frac{\partial \rho(u, v, \theta)}{\partial \theta} \bigg|_{\theta=\tilde{\theta}} \left\{ dN_{k1}(u) dN_{k2}(v) - \rho(u, v, \tilde{\theta}) Y_{k1}(u) d\tilde{\mu}_{01}(u) e^{\tilde{\beta}_{1}^{T} Z_{k1}(u)} \cdot Y_{k2}(v) d\tilde{\mu}_{02}(v) e^{\tilde{\beta}_{2}^{T} Z_{k2}(v)} \right\}.$$

and the Supreme Test Statistic as

$$\tilde{T} = \sup_{s, t \in [0, \tau]} \| \tilde{V}(s, t, \tilde{\theta}) \|.$$

(4.22)

Similarly, to access the empirical distribution of $T$, firstly we approximate it by the first-order Taylor expansion,

$$\tilde{V}(s, t, \tilde{\theta}) = \tilde{V}(s, t, \theta) + N^{-1/2} \frac{\partial \tilde{V}(s, t, \tilde{\theta})}{\partial \theta} N^{1/2}(\tilde{\theta} - \theta) + o_{p}(1)$$

(4.23)
where

$$\hat{V}(s, t, \theta) = V^c(s, t, \theta) + N^{-1/2} \sum_{k=1}^{N} \left\{ \hat{Y}_{k1}^c(s, t) + \hat{Y}_{k2}^c(s, t) \right\} + o_p(1),$$

$$N^{-1/2} \frac{\partial \hat{V}(s, t, \theta)}{\partial \theta} N^{1/2}(\hat{\theta} - \theta) = N^{-1/2} \left\{ \hat{\zeta}_{k1}^c(s, t, \theta) + \hat{\zeta}_{k2}^c(s, t, \theta) \right\} + o_p(1). \quad (4.24)$$

Next, we apply the Gaussian multiplier method by multiplying random numbers $G_k$ from normal distribution, so that

$$\hat{V}^*(s, t)$$

$$= \left\{ \hat{V}(s, t, \hat{\theta}) + N^{-1/2} \sum_{k=1}^{N} \hat{Y}_{k1}^c(s, t, \hat{\theta}) + \hat{Y}_{k2}^c(s, t, \hat{\theta}) + \hat{\zeta}_{k1}^c(s, t) + \hat{\zeta}_{k2}^c(s, t) \right\} \cdot G_k \quad (4.25)$$

By taking the supremum of $\hat{V}^*(s, t)$ among mesh grids of $(s, t)$, we obtain $\hat{T}^*$ from the empirical distribution of $\sup_{s,t \in [0, \tau]} \| \hat{V}^*(s, t, \theta) \|$. Repeating above the process 1000 times enables us to have enough observations and we would reject the $H_0: \rho(s, t, \theta) = \theta_0$ when $\hat{T}^*$ exceeds the 95th percentile of the observations.

4.2.2 Simulation studies

Here, we hope to answer two questions: (1) Are the two event processes independent? (2) If not, is the association constant? Firstly, to detect the dependency, we consider the independent bivariate counting processes as the null model and the constant rate ratio as alternative model. Secondly, we propose the constant rate ratio model as the null and Piecewise Constant (PWC), Time Dependent (TD), Time and Covariate Dependent (TCD) models as the corresponding alternatives.

To investigate the performance of the model checking procedure, finite sample
studies are conducted, with multiplicative mean rate marginal model. The size and power of the hypothesis test are also computed via Gaussian Multiplier Method.

4.2.2.1 Test for constant association with multiplicative marginal models

We consider the Shared Frailty Model below as the null model

\[
E[dN^*_{k1}(s)|R_k, Z_{k1}(s)] = R_k e^{\beta_1^T Z_{k1}(s)} d\mu_{01}(s),
\]

\[
E[dN^*_{k2}(t)|R_k, Z_{k2}(t)] = R_k e^{\beta_2^T Z_{k2}(t)} d\mu_{02}(t),
\]

where \( R_k \) is independent and comes from a Gamma Distribution. Following from Proposition 1, under the null model, we have \( \rho(\theta, s, t) = 1 + \sigma^2 / \mu^2 \) where \( \sigma^2 \) and \( \mu^2 \) represent \( E(R_k) \) and \( \text{var}(R_k) \). Let \( \beta_{01} = 0.2, \beta_{02} = 0.4, \tau = 4 \) and the censoring time follow uniform(3, 4). We take baseline rate \( \mu_{01}(t) = \mu_{02}(t) \) and set the values equal to 0.25, 0.375, 0.5, 0.75, 1 so that \( \rho = 1.25, 1.5, 1.75 \) and 2 respectively.

As we can see, the null model corresponds to \( H_0 : \rho(\theta, s, t) = \theta \). Implementing the Gaussian Multiplier method enables us to approach the empirical distribution of the supreme residuals under the \( H_0 \). Therefore the rejection rate under the \( H_0 \) can be used as an empirical size of the test and should be around its nominee value. The simulation result summarized in Table 16 shows the test has size below or around 0.05 consistently which agrees with the theoretical value.

Similar to the illustration in section 3.2.2.2, we propose the PWC, TD and TCD
model as alternative models to exam the power of the testing procedure. The adjustment is concerned with the marginal mean rate, which should be multiplicative in the following sections.

(I) The piecewise constant rate ratio model - PWC Model

Assume $\tau = 4$ and analogous to equation (3.29), the counting process $N_{kj}^*(t)$ is from

$$E[dN_{kj}^*(t)|R_k(t), Z_{kj}(t)] = R_k(t)\{d\mu_{0j}(t)e^{\beta_0 T_{kj}Z_{kj}(t)}\}.$$  \hspace{1cm} (4.26)

where $R_k(t) = I(t < 2)R_{k0} + I(t > 2)R_{k1}$ is time varying frailty. Let $\beta_{01} = 0.2$, $\beta_{02} = 0.4$, $C_{kj}$ be uniform on $(3, 4)$ and $Z_{kj}$ follows Uniform(0, 1). To modify the events observed before censoring, we take $\mu_{0j}(t)$ equal to $0.125t^2, 0.25t^2, 0.375t^2, 0.5t^2$. Consider $R_{k0}$ and $R_{k1}$ are independently generated from Gamma($a_0, b_0$) and Gamma($a_1, b_1$), where the choice of parameters represent the value of the piecewise rate ratio. The simulation settings are summarized in the Table 15 and Figure 3.

PWC models are alternatives to the null model and therefore the residuals calculated under $H_0$ should depart far away from zero. We would expect the supreme test statistic to go beyond threshold with high likelihood and a high rejection rate is an indicator of the power. 17 shows the proposed procedure can correctly detect non constant Rate Ratio at or above 95% of the cases when sample size is large ($N = 800$) and the accuracy is improved by increasing the sample size.

(II) Time dependent rate ratio - TD Model

Consider the Bivariate Counting Process described by equation (3.30). Assume the
Poisson process $N_{k0}(t)$ has conditional mean rate

$$E[dN_{k0}(t)|Z_{kj}(t), R_k] = \lambda_{k0}(t|Z_{kj}(t), R_k)\, dt$$

and

$$\lambda_{k0}(t|Z_{kj}(t), R_k)\, dt = R_k \cdot d\mu_{0j}(t)e^{\beta_{0j}Z_{kj}(t)}, \tag{4.27}$$

with $R_k$ is the cluster level random effect. Let the conditional mean rate of Poisson process be $\tilde{\lambda}_{kj}(t|Z_{kj}(t))$ and by assigning an appropriate value, we can generate the counting processes $N_{k1}(t)$ and $N_{k2}(s)$ with rate ratio

$$\rho(\theta, s, t) = 1 + \theta(-0.15s + 0.9)(-0.15t + 0.9).$$

where $\theta = \frac{\sigma^2}{\mu^2}$. To consider rare, moderate and high time dependent association, we generate $\theta = 0.5, 1, 1.5, 2$ by taking $R_k$ from Gamma distribution, where the shape and scale parameter pairs in the Gamma Distribution are $(2, 0.5), (1, 1), (0.67, 1.5)$ and $(0.5, 2)$. The color plots for the four settings are also illustrated by Figure 4.

The goodness of fit procedure is more likely to detect non-constant rate ratio for a more varying scenario or a larger sample case. It is observed in Table 18 that the time dependent rate ratio and piecewise constant rate ratio model have similar simulation performance.

(III) Time and covariate dependent model - TCD Model

The Time and Covariate Dependent Rate Ratio can be derived by comparing to section 3.2.2.2. Assume the Poisson process $N_{k0}(t)$ has marginal conditional rate $\lambda_0(t)$ where $\lambda_0(t|Z_{kj}(t), R_k)\, dt = R_k \cdot d\mu_{0j}(t)e^{\beta_{0j}Z_{kj}(t)}$ with $R_k$ the random frailty. By
Proposition 1, \( \rho(\theta, s, t|z_1, z_2) = 1 + \sigma^2/\mu^2 \), where \( \sigma^2 \) and \( \mu \) represent the variance and mean of \( R_k \). Let the Poisson process \( \tilde{N}_{kj}(t) \) has rate \( \tilde{\lambda}_{kj} = 1 \). Following Proposition 3, conditional on covariates

\[
\rho(\theta, s, t|z_1, z_2) = 1 + \theta \frac{\lambda_0(s|z_1)\lambda_0(t|z_2)}{(1 + \lambda_0(s|z_1))(1 + \lambda_0(t|z_2))},
\]

where \( \rho(\theta, s, t|z_1, z_2) \) represents the rate ratio of \{\( N_{k1}(s), N_{k2}(t) \)\} and \( \theta \) is \( \sigma^2/\mu^2 \).

To generate \( \theta = 0.25, 0.5, 1, 2 \), we consider \( R_k \) be from Gamma distribution with \( \mu = 1 \) and \( \sigma^2 = 0.25, 0.5, 1, 2 \). Let \( \tau = 4, \beta_{01} = 0.1, \beta_{02} = 0.2 \) and \( \mu_0j(t) = 0.125t^2, 0.25t^2, 0.375t^2, 0.5t^2 \) for \( j = 1 \) or 2. Take the censoring time and covariates from uniform distribution on \((3, 4)\) and \((1, 2)\) respectively. The rate ratio is in form of

\[
\rho(\theta, s, t|z_1, z_2) = 1 + \theta \frac{\lambda_0(s|z_1)\lambda_0(t|z_2)}{(1 + \lambda_0(s|z_1))(1 + \lambda_0(t|z_2))},
\]

Table 19 summarizes of the simulation result for the above settings, from which similar patterns of PWC and TD Models are shown. In general the test performs well and can distinguish the null model and alternative models with high precision, especially when the sample size is large or the variability of association is increasing.
Table 15: Summary of simulation settings under the PWC model with the corresponding $\rho$ values followed from Proposition 2. The Marginal model is multiplicative.

<table>
<thead>
<tr>
<th>Settings</th>
<th>PWC1</th>
<th>PWC2</th>
<th>PWC3</th>
<th>PWC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{k0}$</td>
<td>$(a_0, b_0)$</td>
<td>$(4, 0.25)$</td>
<td>$(4, 0.25)$</td>
<td>$(4, 0.25)$</td>
</tr>
<tr>
<td>$R_{k1}$</td>
<td>$(a_1, b_1)$</td>
<td>$(2, 0.5)$</td>
<td>$(1.33, 0.75)$</td>
<td>$(1, 1)$</td>
</tr>
<tr>
<td>$\rho(s &lt; 2, t &lt; 2)$</td>
<td>1.25</td>
<td>1.25</td>
<td>1.5</td>
<td>1.25</td>
</tr>
<tr>
<td>$\rho(s &gt; 2, t &lt; 2)$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\rho(s &gt; 2, t &gt; 2)$</td>
<td>1.5</td>
<td>1.75</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Figure 3: Visualization of Piecewise Constant $\rho(s, t, \theta)$ (PWC) under the Additive Marginal Models. The variation of $\rho(s, t)$ between different pieces is growing from PWC1 to PWC4.
Figure 4: The contour plot of the Rate Ratio $\rho(s, t)$ under the Multiplicative Marginal Models. The x-axis and y-axis represents the observation time for type1 and type2 events. From upper left to lower right, the heterogeneity of $\rho(s, t)$ is increased.

(a) TD1 Model $\rho(s, t)$
(b) TD2 Model $\rho(s, t)$
(c) TD3 Model $\rho(s, t)$
(d) TD4 Model $\rho(s, t)$
Table 16: Observed size of the test statistic $T$ for the proposed model-checking procedure under $H_0: \rho(\theta, s, t) = \theta$ is parametric vs $H_a: \rho(\theta, s, t)$ is not parametric, at significance level 0.05. The numbers in the parentheses represent the average observed count of type 1 and type 2 event after censoring. Each entry is calculated based on 1000 Gaussian multiplier samples and 1000 replicates.

<table>
<thead>
<tr>
<th>event counts</th>
<th>$\mu_{0j}(t)$</th>
<th>$K$</th>
<th>$\rho = 1.25$</th>
<th>$\rho = 1.5$</th>
<th>$\rho = 1.75$</th>
<th>$\rho = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4.18, 5.67)</td>
<td>0.25$t^2$</td>
<td>200</td>
<td>0.041</td>
<td>0.025</td>
<td>0.037</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.038</td>
<td>0.042</td>
<td>0.033</td>
<td>0.033</td>
</tr>
<tr>
<td>(6.25, 8.48)</td>
<td>0.375$t^2$</td>
<td>200</td>
<td>0.042</td>
<td>0.035</td>
<td>0.030</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.040</td>
<td>0.037</td>
<td>0.039</td>
<td>0.037</td>
</tr>
<tr>
<td>(8.34, 11.30)</td>
<td>0.5$t^2$</td>
<td>200</td>
<td>0.040</td>
<td>0.037</td>
<td>0.003</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.043</td>
<td>0.043</td>
<td>0.032</td>
<td>0.037</td>
</tr>
</tbody>
</table>
Table 17: Power of $H_0: \rho(\theta, s, t) = \theta_0$ vs $H_a: \rho(\theta, s, t)$ is not parametric. The $H_a$ model has Piecewise Constant Rate Ratio (PWC model). Each entry is calculated based on 1000 Gaussian multiplier samples with 1000 replicates.

<table>
<thead>
<tr>
<th>event counts</th>
<th>$\mu_{ij}(t)$</th>
<th>K</th>
<th>PWC1</th>
<th>PWC2</th>
<th>PWC3</th>
<th>PWC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2.09, 2.83)</td>
<td>0.125$t^2$</td>
<td>200</td>
<td>0.443</td>
<td>0.882</td>
<td>0.777</td>
<td>0.942</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.934</td>
<td>0.999</td>
<td>0.994</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>0.995</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>(4.16, 5.65)</td>
<td>0.25$t^2$</td>
<td>200</td>
<td>0.867</td>
<td>0.977</td>
<td>0.951</td>
<td>0.987</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.998</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>(6.25, 8.48)</td>
<td>0.375$t^2$</td>
<td>200</td>
<td>0.955</td>
<td>0.993</td>
<td>0.968</td>
<td>0.991</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>(8.34, 11.32)</td>
<td>0.5$t^2$</td>
<td>200</td>
<td>0.985</td>
<td>0.994</td>
<td>0.986</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Table 18: Power of $H_0 : \rho(\theta, s, t) = \theta_0$ vs $H_a : \rho(\theta, s, t)$ is not parametric. The $H_a$ model is Time and Dependent (TD). Each entry is calculated based on 1000 Gaussian multiplier samples with 1000 replicates.

<table>
<thead>
<tr>
<th>$\mu_{0j}(t)$</th>
<th>K</th>
<th>TD1</th>
<th>TD2</th>
<th>TDC3</th>
<th>TD4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.125$t^2$</td>
<td>200</td>
<td>0.129</td>
<td>0.252</td>
<td>0.308</td>
<td>0.355</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.295</td>
<td>0.560</td>
<td>0.706</td>
<td>0.782</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.463</td>
<td>0.817</td>
<td>0.906</td>
<td>0.932</td>
</tr>
<tr>
<td>0.25$t^2$</td>
<td>200</td>
<td>0.238</td>
<td>0.415</td>
<td>0.524</td>
<td>0.556</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.587</td>
<td>0.862</td>
<td>0.929</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.768</td>
<td>0.974</td>
<td>0.990</td>
<td>0.986</td>
</tr>
<tr>
<td>0.375$t^2$</td>
<td>200</td>
<td>0.337</td>
<td>0.518</td>
<td>0.598</td>
<td>0.675</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.748</td>
<td>0.933</td>
<td>0.961</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.931</td>
<td>0.991</td>
<td>0.994</td>
<td>0.995</td>
</tr>
<tr>
<td>0.5$t^2$</td>
<td>200</td>
<td>0.433</td>
<td>0.578</td>
<td>0.691</td>
<td>0.674</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.826</td>
<td>0.949</td>
<td>0.962</td>
<td>0.968</td>
</tr>
</tbody>
</table>
Table 19: Power of $H_0 : \rho(\theta, s, t) = \theta_0$ vs $H_a : \rho(\theta, s, t)$ is not parametric. The $H_a$ model is Time and Covariate Dependent (TCD). Each entry is calculated based on 1000 Gaussian multiplier samples and 1000 replicates.

<table>
<thead>
<tr>
<th>$\mu_{ij}(t)$</th>
<th>K</th>
<th>TCD1</th>
<th>TCD2</th>
<th>TCD3</th>
<th>TCD4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.125$t^2$</td>
<td>200</td>
<td>0.102</td>
<td>0.226</td>
<td>0.453</td>
<td>0.706</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.208</td>
<td>0.520</td>
<td>0.916</td>
<td>0.991</td>
</tr>
<tr>
<td>0.25$t^2$</td>
<td>200</td>
<td>0.175</td>
<td>0.417</td>
<td>0.704</td>
<td>0.798</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.508</td>
<td>0.923</td>
<td>0.988</td>
<td>0.977</td>
</tr>
<tr>
<td>0.375$t^2$</td>
<td>200</td>
<td>0.246</td>
<td>0.480</td>
<td>0.701</td>
<td>0.727</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.650</td>
<td>0.946</td>
<td>0.983</td>
<td>0.963</td>
</tr>
<tr>
<td>0.5$t^2$</td>
<td>200</td>
<td>0.210</td>
<td>0.437</td>
<td>0.566</td>
<td>0.631</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.658</td>
<td>0.952</td>
<td>0.972</td>
<td>0.949</td>
</tr>
</tbody>
</table>
REFERENCES


Proof of Proposition 1

By the conditional expectation property and the conditional independent increment of \(N_{k1}, N_{k2}\), we have:

\[
E\{dN_{k1}(s) dN_{k2}(t) | Z_{k1}(s), Z_{k2}(t) \} \\
= E\left\{ E\{dN_{k1}(s) dN_{k2}(t) | Z_{k1}(s), Z_{k2}(t), R_k \} \right\} \\
= E\left\{ E\{dN_{k1}(s) | Z_{k1}(s), R_k \} E\{dN_{k2}(t) | Z_{k2}(t), R_k \} \right\} \\
= E\left\{ R_k \{d\mu_0(s) + \beta_1^T Z_{k1}(s) \, ds \} R_k \{d\mu_0(t) + \beta_2^T Z_{k2}(t) \, dt \} \right\} \\
= E\{R_k^2 \} \{d\mu_0(s) + \beta_1^T Z_{k1}(s) \, ds \} \{d\mu_0(t) + \beta_2^T Z_{k2}(t) \, dt \} \\
\tag{A.1}
\]

and

\[
E\{dN_{k1}(s) | Z_{k1}(s) \} = E\left\{ E\{dN_{k1}(s) | Z_{k1}(s), R_k \} \right\} = E\{R_k \} \{d\mu_0(s) + \beta_1^T Z_{k1}(s) \, ds \}, \\
E\{dN_{k2}(t) | Z_{k2}(t) \} = E\left\{ E\{dN_{k2}(t) | Z_{k2}(t), R_k \} \right\} = E\{R_k \} \{d\mu_0(t) + \beta_2^T Z_{k2}(t) \, dt \}.
\]

Therefore, follows from the definition of the rate ratio in (2.1),

\[
\rho = \frac{E\{dN_{k1}(s) dN_{k2}(t) | Z_{k1}(s), Z_{k2}(t) \}}{E\{dN_{k1}(s) | Z_{k1}(s) \} E\{dN_{k2}(t) | Z_{k2}(t) \}} = \frac{E\{R_k^2 \}}{E\{R_k\} E\{R_k\}} = \frac{\mu^2 + \sigma^2}{\mu^2} = 1 + \frac{\sigma^2}{\mu^2} \\
\tag{A.2}
\]

□

Proof of Proposition 2
Similar to the proof of Proposition 1,

\[
E\{dN_{k1}(s)dN_{k2}(t)|Z_{k1}(s), Z_{k2}(t)\}
\]

\[
= E\{E\{dN_{k1}(s)|Z_{k1}(s), R_k\}E\{dN_{k2}(t)|Z_{k2}(t), R_k\}\}
\]

\[
= E\{R_k(s)R_k(t)\} \cdot \{d\mu_1(s) + \beta_1^T Z_{k1}(s) \, ds\} \{d\mu_2(t) + \beta_2^T Z_{k2}(t) \, dt\} \quad (A.3)
\]

and

\[
E\{dN_{k1}(s)|Z_{k1}(s)\} = E\{R_k(s)\}\{d\mu_1(s) + \beta_1^T Z_{k1}(s) \, ds\}
\]

\[
E\{dN_{k2}(t)|Z_{k2}(t)\} = E\{R_k(t)\}\{d\mu_2(t) + \beta_2^T Z_{k2}(t) \, dt\}. \quad (A.4)
\]

Since \(R_k(u)\) is piecewise constant, we have

\[
E\{R_k(s)R_k(t)\} = \begin{cases} 
E(R_{k0}R_{k0}) = (a_0 b_0 + \delta_0)^2 + a_0 b_0^2 & \text{if } s, t \in (0, c_0] \\
E(R_{k1}R_{k1}) = (a_1 b_1 + \delta_1)^2 + a_1 b_1^2 & \text{if } s, t \in (c_0, \tau] \\
E(R_{k0}R_{k1}) = (a_0 b_0 + \delta_0)(a_1 b_1 + \delta_1) & \text{otherwise}
\end{cases} \quad (A.5)
\]

\[
E\{R_k(s)\}E\{R_k(t)\} = \begin{cases} 
E(R_{k0})E(R_{k0}) = (a_0 b_0 + \delta_0)^2 & \text{if } s, t \in (0, c_0] \\
E(R_{k1})E(R_{k1}) = (a_1 b_1 + \delta_1)^2 & \text{if } s, t \in (c_0, \tau] \\
E(R_{k0})E(R_{k1}) = (a_0 b_0 + \delta_0)(a_1 b_1 + \delta_1) & \text{otherwise}
\end{cases} \quad (A.6)
\]
This yields the piecewise constant rate ratio below:

\[ \rho(\theta, s, t) = \frac{E\{R_k(s)R_k(t)\}}{E\{R_k(s)\}E\{R_k(t)\}} = \begin{cases} 
1 + \frac{a_0 b_0^2}{(a_0 b_0 + \delta_0)^2} & \text{if } s, t \in (0, c_0] \\
1 + \frac{a_1 b_1^2}{(a_1 b_1 + \delta_1)^2} & \text{if } s, t \in (c_0, \tau] \\
1 & \text{otherwise} 
\end{cases} \quad (A.7) \]

\[ \square \]

**Proof of Proposition 3**

By the definition of mean event rate

\[ E[dN_1(s) dN_2(t) | z_1, z_2] \]

\[ = P\{dN_1(s) = 1, dN_2(t) = 1 | Z_1(s) = z_1, Z_2(t) = z_2\} \]

\[ = P\{d\tilde{N}_1(s) + dN_0(s) = 1, d\tilde{N}_2(t) + dN_0(t) = 1 | z_1, z_2\} \]

since \( \tilde{N}_j(\cdot) \) and \( N_0(\cdot) \) are conditional independent to each other, we have

\[ P\{d\tilde{N}_1(s) + dN_0(s) = 1, d\tilde{N}_2(t) + dN_0(t) = 1 | z_1, z_2\} \]

\[ = P\{d\tilde{N}_1(s) = 1, dN_0(s) = 0, d\tilde{N}_2(t) + dN_0(t) = 1 | z_1, z_2\} \]

\[ + P\{d\tilde{N}_1(s) = 0, dN_0(s) = 1, d\tilde{N}_2(t) + dN_0(t) = 1 | z_1, z_2\} \quad (A.8) \]
On the right hand side of (A.8),

\[ P\{d\tilde{N}_1(s) = 1, dN_0(s) = 0, d\tilde{N}_2(t) + dN_0(t) = 0 \mid z_1, z_2 \} \]

\[ = P\{d\tilde{N}_1(s) = 1 \mid z_1 \} \cdot P\{dN_0(s) = 0, d\tilde{N}_2(t) = 0, dN_0(t) = 0 \mid z_1, z_2 \} \]

\[ + P\{d\tilde{N}_1(s) = 1 \mid z_1 \} \cdot P\{dN_0(s) = 0, d\tilde{N}_2(t) = 1, dN_0(t) = 0 \mid z_1, z_2 \} \]

\[ = \tilde{\lambda}_1(s \mid z_1) ds \cdot \lambda_0(t \mid z_2) dt + \tilde{\lambda}_1(s \mid z_1) ds \cdot \tilde{\lambda}_2(t \mid z_2) dt \]  \hspace{1cm} (A.9)

Similarly,

\[ P\{d\tilde{N}_1(s) = 0, dN_0(s) = 1, d\tilde{N}_2(t) + dN_0(t) = 1 \mid z_1, z_2 \} \]

\[ = P\{d\tilde{N}_1(s) = 0 \mid z_1 \} \cdot P\{dN_0(s) = 1, d\tilde{N}_2(t) = 0, dN_0(t) = 1 \mid z_1, z_2 \} \]

\[ + P\{d\tilde{N}_1(s) = 0 \mid z_1 \} \cdot P\{dN_0(s) = 1, d\tilde{N}_2(t) = 1, dN_0(t) = 0 \mid z_1, z_2 \} \]

\[ = 1 \cdot \rho_0(\theta, s, t \mid z_1, z_2) \lambda_0(s \mid z_1) ds \cdot \lambda_0(t \mid z_2) dt + 1 \cdot \tilde{\lambda}_2(t \mid z_2) dt \lambda_0(s \mid z_1) ds \]  \hspace{1cm} (A.10)

Combine equation (A.9) and (A.10) allows us to represent equation (A.8) as below

\[ E[\ N_1(s) \ N_2(t) \mid z_1, z_2] \]

\[ = P\{d\tilde{N}_1(s) + dN_0(s) = 1, d\tilde{N}_2(t) + dN_0(t) = 1 \mid z_1, z_2 \} \]

\[ = \tilde{\lambda}_1(s \mid z_1) ds \cdot \lambda_0(t \mid z_2) dt + \tilde{\lambda}_1(s \mid z_1) ds \cdot \tilde{\lambda}_2(t \mid z_2) dt \]

\[ + 1 \cdot \rho_0(\theta, s, t \mid z_1, z_2) \lambda_0(s \mid z_1) ds \cdot \lambda_0(t \mid z_2) dt + \tilde{\lambda}_2(t \mid z_2) dt \cdot \lambda_0(s \mid z_1) ds \]

\[ = \{\tilde{\lambda}_1(s \mid z_1) + \lambda_0(s \mid z_1)\}\{\tilde{\lambda}_0(t \mid z_2) + \lambda_0(t \mid z_2)\} ds dt \]

\[ + \{\rho_0(\theta, s, t \mid z_1, z_2) - 1\}\lambda_0(s \mid z_1)\lambda_0(t \mid z_2) ds dt \]

\[ = \lambda_1(s \mid z_1)\lambda_2(t \mid z_2) ds dt + \{\rho_0(\theta, s, t \mid z_1, z_2) - 1\}\lambda_0(s \mid z_1)\lambda_0(t \mid z_2) ds dt \]  \hspace{1cm} (A.11)
and

\[ E[dN_1(s) | z_1] E[dN_2(t) | z_2] = \lambda_1(s | z_1) \lambda_2(t | z_2) \, ds \, dt. \tag{A.12} \]

By definition the rate ratio of bivariate counting processes \( \{N_1(s), N_2(t)\} \) is

\[ \rho(\theta, s, t | z_1, z_2) = \frac{E[dN_1(s) dN_2(t) | z_1, z_2]}{E[dN_1(s) | z_1] E[dN_2(t) | z_2]} \tag{A.13} \]

and substituting equations (A.11) and (A.12) into the (A.13) gives us

\[ \rho(\theta, s, t | z_1, z_2) = 1 + \left( \rho_0(\theta, s, t | z_1, z_2) - 1 \right) \frac{\lambda_0(s | z_1) \lambda_0(t | z_2)}{\lambda_1(s | z_2) \lambda_2(t | z_2)} \]

The rate ratio of \( N_1(s) \) and \( N_2(t) \) depends on that of \( N_0(s) \) and \( N_0(t) \). If \( N_0(s) \) and \( N_0(t) \) are independent, \( \rho_0(\theta, s, t | z_1, z_2) \) would be 1, which leads to \( \rho(\theta, s, t | z_1, z_2) = 1 \) as well. If the occurrence of events at time \( s, t \) are positively correlated, \( \rho_0(\theta, s, t | z_1, z_2) \) will be greater than 1 and therefore \( \rho(\theta, s, t | z_1, z_2) > 1 \). For negatively associated event occurrence, both \( \rho_0(\theta, s, t | z_1, z_2) \) and \( \rho(\theta, s, t | z_1, z_2) \) will be both less than 1.

\[ \square \]

APPENDIX B: PROOFS OF THE THEOREMS IN CHAPTER 3

Condition I.

Adapting from H. Scheike (2002), we show the asymptotic properties of the first-stage estimators in our proposed method. The following regularity conditions are assumed for \( j = 1, 2 \):

C.1. \( \{N_{kj}^*(\cdot), C_{kj}, Z_{kj}(\cdot)\} \) are independent and identically distributed for \( k = 1, 2, \ldots, N \).
C.2. \( \Pr(C_{kj} > \tau) > 0 \), where \( \tau \) is predetermined constant; \( N_{kj}(\tau) < \eta < \infty \) are bounded by a constant almost surely.

C.3. \( N_{kj}(\tau) \) are bounded by a constant;

C.4. \(|Z_{kj}(0)| + \int_0^\tau |dZ_{kj}(s)| < c_Z < \infty \), almost surely, where \( c_Z > 0 \) is a constant.

C.5. Denote the positive-definiteness matrix \( A_j \) as

\[
A_j = E\{ \int_0^\tau \{ Z_{kj}(u) - \bar{Z}_j(\beta_j, u) \} \otimes^2 ds \},
\]

where \( \bar{Z}_j(t) = \lim_{N \to \infty} \bar{Z}_j(t) \) and \( \bar{Z}_j(t) = \frac{\sum_{k=1}^N Z_{kj}(t) Y_{kj}(t)}{\sum_{k=1}^N Y_{kj}(t)} \).

Proof of Theorem 3.1

Denote the likelihood function as

\[
L_j(\beta_j) = \sum_{k=1}^N \int_0^\tau \{ Z_{kj}(u) - \bar{Z}_j(u) \} dM_{kj}(u, \beta_j),
\]

and with the first order Taylor expansion with respect to \( \beta_j \) gives us

\[
(\hat{\beta}_j - \beta_j) = \hat{A}_j^{-1}(\beta^*) N^{-1} \int_0^\tau \{ Z_{kj}(u) - \bar{Z}_j(u) \} dM_{kj}(u, \beta_j),
\]

where

\[
dM_{kj}(t; \beta_j) = dN_{kj}(t) - Y_{kj}(t)\{d\mu_{kj}(t) + \beta_j^T Z_{kj}(t) dt\}
\]

\[
\hat{A}_j(\beta_j) = - N^{-1} \sum_{k=1}^N \int_0^\tau \{ Z_{kj}(u) - Z_j(\hat{\beta}_j, u) \} \otimes^2 du,
\]

with \( \beta^* \) a value falls between \( \hat{\beta}_j \) and \( \beta_j \).
By (C.4) and the strong law of large numbers (SLLN), $\hat{\beta}_j$ converges almost surely to $\beta_j$. From the Slutsky’s theorem and (A.15), $\sqrt{N}(\hat{\beta}_j - \beta_j)$ is asymptotically normal with mean zero and covariance matrix $A_j^{-1}\Sigma_j A_j^{-1}$, where
\[
\Sigma_j = E\left[\int_0^\tau \{Z_{1j}(u) - \bar{Z}_j(u)\} dM_{1j}(u, \beta_j) \int_0^\tau \{Z_{1j}(v) - \bar{Z}_j(v)\} dM_{1j}(v, \beta_j)\right].
\]

From (A.15) it is straightforward to show
\[
\sqrt{N}(\hat{\beta}_j - \beta_j) = A_j^{-1/2} \sum_{k=1}^N \xi_{kj} + o_p(1). \tag{A.16}
\]
where
\[
\xi_{kj} = \int_0^\tau \{Z_{kj}(u) - \bar{Z}_j(u)\} dM_{kj}(u, \beta_j). \tag{A.17}
\]

The asymptotic covariance matrix of $\sqrt{N}(\hat{\beta}_j - \beta_j)$ can be consistently estimated by $\hat{A}_j^{-1}\hat{\Sigma}_j \hat{A}_j^{-1}$, with the corresponding estimators
\[
d\hat{M}_{kj}(t; \hat{\beta}_j) = dN_{kj}(t) - Y_{kj}(t)\{d\hat{\mu}_0(t) + \hat{\beta}_j^T Z_{kj}(t) dt\},
\]
\[
\hat{\xi}_{kj} = \int_0^\tau \{Z_{kj}(u) - \bar{Z}_j(u)\} d\hat{M}_{kj}(u; \hat{\beta}_j),
\]
\[
\hat{\Sigma}_j = N^{-1} \sum_{k=1}^N \hat{\xi}_{kj}^2.
\]
\[\square\]

**Proof of Theorem 3.2**

Consider
\[
\hat{\mu}_{0j}(t) - \mu_{0j}(t) = \{\hat{\mu}_{0j}(t; \hat{\beta}_j) - \hat{\mu}_{0j}(t; \beta_j)\} + \{\hat{\mu}_{0j}(t; \beta_j) - \mu_{0j}(t)\} \tag{A.18}
\]
By the first order Taylor approximation, we have

\[
\hat{\mu}_{0j}(t; \hat{\beta}_j) - \mu_{0j}(t; \beta_j) = - (\hat{\beta}_j - \beta_j) \int_0^t \bar{Z}_j^T(u) \, du + o_p(N^{-1}), \tag{A.19}
\]

\[
\hat{\mu}_{0j}(t; \beta_j) - \mu_{0j}(t) = N^{-1} \sum_{k=1}^N \int_0^t \frac{dM_{kj}(u; \beta_j)}{\bar{\pi}_j(u)} + o_p(N^{-1}). \tag{A.20}
\]

Using the strong convergence of \(\beta_j\) in Theorem 3.1 and the Uniform SLLN (Pollard 1990), \(\{\hat{\mu}_{0j}(t; \hat{\beta}_j) - \hat{\mu}_{0j}(t; \beta_j)\}\) converges almost surely to 0 uniformly in \(t \in [0, \tau]\).

Similarly, \(\mu_{0j}(t; \beta_j)\) converges strongly to \(\mu_{0j}(t)\) uniformly.

By the Triangle Inequality,

\[
|\hat{\mu}_{0j}(t) - \mu_{0j}(t)| \leq |\hat{\mu}_{0j}(t; \hat{\beta}_j) - \hat{\mu}_{0j}(t; \beta_j)| + |\hat{\mu}_{0j}(t; \beta_j) - \mu_{0j}(t)|.
\]

Therefore, \(\hat{\mu}_{0j}(t)\) converges almost surely to \(\mu_{0j}(t)\) uniformly in \(t \in [0, \tau]\) as well.

Substituting (A.19), (A.20) into (A.18) and multiplying both sides by \(\sqrt{N}\) gives,

\[
\sqrt{N}\{\hat{\mu}_{0j}(t) - \mu_{0j}(t)\} = N^{-1/2} \sum_{k=1}^N \phi_{kj}(t) + o_p(1), \tag{A.21}
\]

where

\[
\phi_{kj}(t; \beta_j) = \int_0^t \frac{dM_{kj}(u; \beta_j)}{\bar{\pi}_j(u)} - H(t)A_j^{-1} \int_0^\tau \{Z_{kj}(u) - \bar{z}_j(u)\} \, dM_{kj}(u, \beta_j), \tag{A.22}
\]

with \(H(t) = \int_0^t \bar{z}_j(u) \, du\).

Thus \(\sqrt{N}\{\hat{\mu}_{0j}(t) - \mu_{0j}(t)\}\) converges weakly to a mean-zero Gaussian process with covariance function \(\Gamma_j(s, t) = E[\phi_{1j}(s; \beta_j) \phi_{1j}(t; \beta_j)]\), which can be consistently approximated by

\[
\hat{\Gamma}_j(s, t) = N^{-1} \sum_{k=1}^N \hat{\phi}_{kj}(s; \hat{\beta}_j) \hat{\phi}_{kj}(t; \hat{\beta}_j),
\]
where

\[ \hat{\phi}_{kj}(t; \hat{\beta}_j) = \int_0^t \frac{d\hat{M}_{kj}(u; \hat{\beta}_j)}{\hat{\pi}_j(u)} - \hat{H}^T(t) \hat{A}_j^{-1} \int_0^t \{ Z_{kj}(u) - \hat{Z}_j(u) \} d\hat{M}_{kj}(u; \hat{\beta}_j) , \]

with

\[ \hat{\pi}_j(t) = N^{-1} \sum_{k=1}^N Y_{kj}(t), \]

\[ \hat{H}^T(t) = \int_0^t Z^T_j(u) du. \]

\[ \square \]

**Proof of Theorem 3.3**

To prove the asymptotic of \( \{ U(\theta, \hat{\beta}_1, \hat{\mu}_1(\cdot), \hat{\beta}_2, \hat{\mu}_2(\cdot)) - U(\theta, \beta_1, \mu_01(\cdot), \beta_2, \mu_02(\cdot)) \} \)

where \( U(\theta, \hat{\beta}_1, \hat{\mu}_01(\cdot), \hat{\beta}_2, \hat{\mu}_02(\cdot)) = \sum_{k=1}^N U_k(\theta, \hat{\beta}_1, \hat{\mu}_01(\cdot), \hat{\beta}_2, \hat{\mu}_02(\cdot)) \), we consider the following decomposition:

\[ U_k(\theta, \hat{\beta}_1, \mu_01(\cdot), \hat{\beta}_2, \mu_02(\cdot)) = U_k(\theta, \beta_1, \mu_01(\cdot), \beta_2, \mu_02(\cdot)) \]

\[ + \{ U_k(\theta, \hat{\beta}_1, \hat{\mu}_01(\cdot), \hat{\beta}_2, \hat{\mu}_02(\cdot)) - U_k(\theta, \beta_1, \mu_01(\cdot), \hat{\beta}_2, \hat{\mu}_02(\cdot)) \} \]

\[ + \{ U_k(\theta, \beta_1, \mu_01(\cdot), \hat{\beta}_2, \hat{\mu}_02(\cdot)) - U_k(\theta, \beta_1, \mu_01(\cdot), \beta_2, \hat{\mu}_02(\cdot)) \} \]

(A.23)

The third term in (A.23) can be further expressed as

\[ \int_0^T \int_0^T - \frac{\partial \rho(s, t, \theta)}{\partial \theta} \rho(s, t, \theta) \]

\[ \{ Y_{k2}(t) \{ d\hat{\mu}_02(t) + \hat{\beta}_2^T Z_{k2}(t) dt - d\mu_02(t) - \beta_2^T Z_{k2}(t) dt \} Y_{k1}(s) \{ d\mu_01(s) + \beta_1^T Z_{k1}(s) ds \} \}, \]

(A.24)
by replacing $(\hat{\beta}_2 - \beta_2)$ and $\hat{\mu}_{0j}(t) - \mu_{0j}(t)$ with (A.16) and (A.21) respectively, we have

\[
U_k(\theta, \beta_1, \mu_{01}(s), \hat{\beta}_2, \hat{\mu}_{02}(t)) - U_k(\theta, \beta_1, d\mu_{01}(s), \beta_2, d\mu_{02}(t))
= \int_0^\tau \int_0^\tau -\frac{\partial \rho(s, t, \theta)}{\partial \theta} \rho(s, t, \theta) Y_{k1}(s) \{d\mu_{01}(s) + \beta_1^T Z_{k1}(s) ds\}
\]

\[
Y_{k2}(t) \left\{ Z_{k2}^T(t) dt A_2^{-1} N^{-1} \sum_{l=1}^N \xi_{l2} + N^{-1} \sum_{l=1}^N d\phi_{l2}(t; \beta_2) \right\} + o_p(N^{-1}) \tag{A.25}
\]

Similarly $\{U_k(\theta, \hat{\beta}_1, \hat{\mu}_{01}(s), \hat{\beta}_2, \hat{\mu}_{02}(t)) - U_k(\theta, \beta_1, \mu_{01}(s), \hat{\beta}_2, \hat{\mu}_{02}(t))\}$ in (A.23) is equivalent to

\[
\int_0^\tau \int_0^\tau -\frac{\partial \rho(s, t, \theta)}{\theta} \rho(s, t, \theta) Y_{k1}(s) Y_{k2}(t) [d\mu_{02}(t) + \beta_2^T Z_{k2}(t) dt]
\]

\[
\left\{ Z_{k1}^T(s) ds A_1^{-1} N^{-1} \sum_{l=1}^N \xi_{l1} + N^{-1} \sum_{l=1}^N d\phi_{l1}(s; \beta_1) \right\} + o_p(N^{-1}) \tag{A.26}
\]

It follows from (A.25), (A.26) and the definition in (3.9) that

\[
N^{-1/2} \left\{ U(\theta, \hat{\beta}_1, \hat{\mu}_{01}(\cdot), \hat{\beta}_2, \hat{\mu}_{02}(\cdot)) - U(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) \right\}
= N^{-1/2} \sum_{k=1}^N \left\{ h_{1,N} \xi_{k1} A_1^{-1} + g_{1,N,k} + h_{2,N} \xi_{k2} A_2^{-1} + g_{2,N,k} \right\} + o_p(N^{-1/2}) \tag{A.27}
\]

where the terms are denoted by

\[
q_l(s, t) = -\frac{\partial \rho(s, t, \theta)}{\partial \theta} \rho(s, t, \theta) Y_{l1}(s) Y_{l2}(t),
\]

\[
h_{1,N} = N^{-1} \sum_{l=1}^N \int_0^\tau \int_0^\tau q_l(s, t) \{d\mu_{02}(t) + \beta_2^T Z_{l2}(t) dt\} Z_{l1}^T(s) ds,
\]

\[
h_{2,N} = N^{-1} \sum_{l=1}^N \int_0^\tau \int_0^\tau q_l(s, t) \{d\mu_{01}(s) + \beta_1^T Z_{l1}(s) ds\} Z_{l2}^T(t) dt,
\]

\[
g_{1,N,k} = N^{-1} \sum_{l=1}^N \int_0^\tau \int_0^\tau q_l(s, t) \{d\mu_{02}(t) + \beta_2^T Z_{l2}(t) dt\} d\phi_{l1}(s; \beta_1),
\]

\[
g_{2,N,k} = N^{-1} \sum_{l=1}^N \int_0^\tau \int_0^\tau q_l(s, t) \{d\mu_{01}(s) + \beta_1^T Z_{l1}(s) ds\} d\phi_{l2}(t; \beta_2).
\]
Deriving from (A.27) the covariance matrix can be estimated by

\[
\hat{\Omega} = N^{-1} \sum_{k=1}^{N} \{ \hat{h}_{1,N} \xi_{k1} A_{1}^{-1} + \hat{g}_{1,N,k} + \hat{h}_{2,N} \xi_{k2} A_{2}^{-1} + \hat{g}_{2,N,k} \} \otimes 2, \tag{A.28}
\]

with

\[
\hat{q}_{l}(s, t) = -\frac{\partial \rho(s, t, \theta)}{\partial \theta} \rho(s, t) Y_{l1}(s) Y_{l2}(t)
\]

\[
\hat{h}_{1,N} = N^{-1} \sum_{l=1}^{N} \int_{0}^{\tau} \int_{0}^{\tau} q_{l}(s, t) \{ d\hat{\mu}_{02}(t) + \hat{\beta}_{2}^{T} Z_{l2}(t) dt \} Z_{l1}^{T}(s) ds,
\]

\[
\hat{h}_{2,N} = N^{-1} \sum_{l=1}^{N} \int_{0}^{\tau} \int_{0}^{\tau} q_{l}(s, t) \{ d\hat{\mu}_{01}(s) + \hat{\beta}_{1}^{T} Z_{l1}(s) ds \} Z_{l2}^{T}(t) dt,
\]

\[
\hat{g}_{1,N,k} = N^{-1} \sum_{l=1}^{N} \int_{0}^{\tau} \int_{0}^{\tau} q_{l}(s, t) \{ d\hat{\mu}_{02}(t) + \hat{\beta}_{2}^{T} Z_{l2}(t) dt \} d\hat{\phi}_{k1}(s; \hat{\beta}_{1}),
\]

\[
\hat{g}_{2,N,k} = N^{-1} \sum_{l=1}^{N} \int_{0}^{\tau} \int_{0}^{\tau} q_{l}(s, t) \{ d\hat{\mu}_{01}(s) + \hat{\beta}_{1}^{T} Z_{l1}(s) ds \} d\hat{\phi}_{k2}(t; \hat{\beta}_{2}).
\]

\[
(A.29)
\]

\[\square\]

**Proof of Theorem 3.4**

Denote

\[
W_{k}(\theta, \beta_{1}, \mu_{01}(\cdot), \beta_{2}, \mu_{02}(\cdot))
\]

\[
= U_{k}(\theta, \beta_{1}, \mu_{01}(\cdot), \beta_{2}, \mu_{02}(\cdot)) + \left\{ h_{1,N} \xi_{k1} A_{1}^{-1} + g_{1,N,k} + h_{2,N} \xi_{k2} A_{2}^{-1} + g_{2,N,k} \right\}, \tag{A.30}
\]

which follows from equation (A.27) and let

\[
\mathcal{I}(\theta, \beta_{1}, \mu_{01}(\cdot), \beta_{2}, \mu_{02}(\cdot)) = -N^{-1} \sum_{k=1}^{N} \left( \frac{\partial U_{k}(\theta, \beta_{1}, \mu_{01}(\cdot), \beta_{2}, \mu_{02}(\cdot))}{\partial \theta} \right)^{T}. \tag{A.31}
\]

The First-order Taylor expansion of the estimation equation around the true values
gives us,

\[
\sqrt{N}(\hat{\theta} - \theta) = N^{-1/2} \left\{ \mathcal{I}(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) \right\}^{-1} \sum_{k=1}^{N} W_k(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) + o_p(1).
\]

(A.32)

By the central limit theorem that \(\sqrt{N}(\hat{\theta} - \theta)\) is asymptotically normal with mean 0 and its variance that can be estimated by \(\hat{\Phi} = N^{-1} (\hat{\mathcal{I}})^{-1} \sum_{k=1}^{N} (\hat{W}_k)^{\otimes 2} \{ (\hat{\mathcal{I}}^T)^{-1} \}^{-1},\) with

\[
\hat{\mathcal{I}} = \mathcal{I}(\hat{\theta}, \hat{\beta}_1, \hat{\mu}_{01}(\cdot), \hat{\beta}_2, \hat{\mu}_{02}(\cdot)),
\]

\[
\hat{W}_k = W_k(\hat{\theta}, \hat{\beta}_1, \hat{\mu}_{01}(\cdot), \hat{\beta}_2, \hat{\mu}_{02}(\cdot)),
\]

obtained with the plugged in estimators \(\hat{\theta}, \hat{\beta}_1, \hat{\mu}_{01}(\cdot), \hat{\beta}_2, \hat{\mu}_{02}(\cdot), \hat{\xi}_{k1}\) and \(\hat{\xi}_{k2}.\)

\(\square\)

APPENDIX C: PROOFS OF THE MODEL CHECKING PROCEDURE IN

CHAPTER 3

Recall (3.26)

\[
V(s, t, \hat{\theta}, \hat{\mu}_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})) = V(s, t, \theta, \hat{\mu}_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2}))
\]

\[
+ N^{-1/2} \frac{\partial V(s, t, \theta, \hat{\mu}_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2}))}{\partial \theta} N^{1/2}(\hat{\theta} - \theta)
\]

\[
+ o_p(1),
\]
Note that \( V\left(s, t, \theta, \hat{\mu}_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\right) \) can be further decomposed by

\[
V\left(s, t, \theta, \hat{\mu}_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\right)
= V\left(s, t, \theta, \mu_1(\cdot; Z_{k1}), \mu_2(\cdot; Z_{k2})\right)
+ V\left(s, t, \theta, \hat{\mu}_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\right)
- V\left(s, t, \theta, \mu_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\right)
+ V\left(s, t, \theta, \mu_1(\cdot; Z_{k1}), \mu_2(\cdot; Z_{k2})\right),
\]

(A.33)

Applying the same techniques in the proof of Theorem 3.3, the third and fourth lines in equation (A.33) are

\[
\sqrt{N}\left\{ V\left(s, t, \theta, \hat{\mu}_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\right) - V\left(s, t, \theta, \mu_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\right) \right\}
= \sum_{k=1}^{N} \int_0^t \int_0^s \frac{\partial \rho(u, v, \theta; Z_{k1}, Z_{k2})}{\partial \theta} \rho(u, v, \theta; Z_{k1}, Z_{k2})
Y_{k1}(u)Y_{k2}(v) \left\{ d\mu_{02}(v) + \beta^T Z_{k2}(v) dv \right\}
\left\{ Z_{k1}^T(s) ds A_1^{-1} \sum_{l=1}^{N} \xi_{l1} + N^{-1} \sum_{l=1}^{N} d\phi_{l1} \right\}
+ o_p(N^{-1}),
\]

(A.34)

and

\[
\sqrt{N}\left\{ V\left(s, t, \theta, \mu_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\right) - V\left(s, t, \theta, \mu_1(\cdot; Z_{k1}), \mu_2(\cdot; Z_{k2})\right) \right\}
= \sum_{k=1}^{N} \int_0^t \int_0^s \frac{\partial \rho(u, v, \theta; Z_{k1}, Z_{k2})}{\partial \theta} \rho(u, v, \theta; Z_{k1}, Z_{k2})
Y_{k2}(v)Y_{k1}(u) \left\{ d\mu_{01}(u) + \beta^T Z_{k1}(u) du \right\}
\left\{ Z_{k2}^T(v) dv A_2^{-1} \sum_{l=1}^{N} \xi_{l2} + N^{-1} \sum_{l=1}^{N} d\phi_{l2} \right\}
+ o_p(N^{-1}).
\]

(A.35)
Combine (A.34) and (A.35), we have

\[
\sqrt{N} \left\{ V\left(s, t, \theta, \hat{\mu}_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\right) - V\left(s, t, \theta, \mu_{01}(\cdot; Z_{k1}), \mu_{02}(\cdot; Z_{k2})\right) \right\} \\
= \sum_{k=1}^{N} \left\{ h_{1,k}(s, t)\xi_{k1}A_1^{-1} + g_{1,k}(s, t) + h_{2,k}(s, t)\xi_{k2}A_2^{-1} + g_{2,k}(s, t) \right\} \\
+ o_p(N^{-1}) \tag{A.36}
\]

where

\[
q_l(u, v) = -\frac{\partial \rho(u, v, \theta)}{\partial \theta} \rho(u, v, \theta) Y_{l1}(u) Y_{l2}(v)
\]

\[
h_{2,N}(s, t) = N^{-1} \sum_{l=1}^{N} \int_{0}^{t} \int_{0}^{s} w(u, v)q_l(u, v)\{d\mu_{01}(u) + \beta_1^T Z_{l1}(u) du\} Z_{l2}^T(u) dv
\]

\[
g_{2,N,k}(s, t) = N^{-1} \sum_{l=1}^{N} \int_{0}^{t} \int_{0}^{s} w(u, v)q_l(u, v)\{d\mu_{02}(u) + \beta_1^T Z_{l2}(u) dv\} d\phi_{k2}(v)
\]

\[
h_{1,N}(s, t) = N^{-1} \sum_{l=1}^{N} \int_{0}^{t} \int_{0}^{s} w(u, v)q_l(u, v)\{d\mu_{01}(v) + \beta_2^T Z_{l1}(v) du\} Z_{l1}^T(u) du
\]

\[
g_{1,N,k}(s, t) = N^{-1} \sum_{l=1}^{N} \int_{0}^{t} \int_{0}^{s} w(u, v)q_l(u, v)\{d\mu_{02}(v) + \beta_2^T Z_{l2}(v) dv\} d\phi_{k1}(u) \tag{A.37}
\]

To simplify the notation, we define

\[
\Upsilon_{k1}(s, t, \theta) = N^{-1} \left\{ h_{1,k}(s, t)\xi_{k1}A_1^{-1} + g_{1,k}(s, t) \right\} + o_p(N^{-1})
\]

\[
\Upsilon_{k2}(s, t, \theta) = N^{-1} \left\{ h_{2,k}(s, t)\xi_{k2}A_2^{-1} + g_{2,k}(s, t) \right\} + o_p(N^{-1}) \tag{A.38}
\]

so that (A.36) can be rewritten as

\[
V\left(s, t, \theta, \hat{\mu}_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\right) - V\left(s, t, \theta, \mu_{01}(\cdot), \mu_{02}(\cdot)\right) \\
= N^{-1/2} \sum_{k=1}^{N} \left\{ \Upsilon_{k1}(s, t, \theta) + \Upsilon_{k2}(s, t, \theta) \right\} + o_p(1). \tag{A.39}
\]
Following the empirical approximation of $\sqrt{N}(\hat{\theta} - \theta)$ in equation (3.10),

\[ N^{-1/2} \frac{\partial V(s, t, \hat{\mu}_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2}))}{\partial \theta} N^{1/2}(\hat{\theta} - \theta) \]

\[ = \Psi_\theta(s, t) \left\{ -N^{-1/2} \{ I(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) \}^{-1} \sum_{k=1}^{N} W_k(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) \right\} \]

\[ + o_p(1), \tag{A.40} \]

where $\Psi_\theta(s, t) = \lim_{N \to \infty} N^{-1/2} \frac{\partial \hat{V}(s, t, \theta)}{\partial \theta}$. We reform (A.40) as

\[ N^{-1/2} \frac{\partial V(s, t, \hat{\beta}_1, \hat{\mu}_{01}(\cdot), \hat{\beta}_2, \hat{\mu}_{02}(\cdot))}{\partial \theta} N^{1/2}(\hat{\theta} - \theta) = N^{-1/2} \{ \zeta_{k1}(s, t, \theta) + \zeta_{k2}(s, t, \theta) \} \]

\[ \tag{A.41} \]

where

\[ \zeta_{k1}(s, t, \theta) = -\Psi_\theta(s, t) \{ I(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) \}^{-1} N^{-1} \left\{ h_{1,N} \xi_{k1} A_1^{-1} + g_{1,N,k} \right\} \]

\[ \zeta_{k2}(s, t, \theta) = -\Psi_\theta(s, t) \{ I(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) \}^{-1} N^{-1} \left\{ h_{2,N} \xi_{k2} A_2^{-1} + g_{2,N,k} \right\} \]

\[ \tag{A.42} \]

Plugging (A.39) and (A.41) back into equation (3.26) gives us (3.27)

\[ V(s, t, \hat{\theta}, \hat{\mu}_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})) \]

\[ = V(s, t, \theta, \mu_1(\cdot; Z_{k1}), \mu_2(\cdot; Z_{k2})) \]

\[ + N^{-1/2} \sum_{k=1}^{N} \left\{ r_{k1}(s, t, \theta) + r_{k2}(s, t, \theta) + \zeta_{k1}(s, t, \theta) + \zeta_{k2}(s, t, \theta) \right\} + o_p(1). \]

**APPENDIX D: THE PROOFS OF THEOREMS IN CHAPTER 4**

**Condition II.**

In this section, we investigate the asymptotic properties of $\hat{\theta}^c$ under the indepen-
dent censoring assumption and that the distribution functions of the censoring times are independent from covariates. Following regularity conditions in Lin et al. (2000):

(C*1) \(\{N_{kj}(.), Y_{kj}(.), Z_{kj}(.\}) (k = 1, 2, \ldots, N; j = 1, 2)\) are independent and identically distributed;

(C*2) \(Pr(C_{kj} > \tau) > 0\), where \(\tau\) is predetermined constant;

(C*3) \(N_{kj}(\tau)\) are bounded by a constant;

(C*4) \(Z_{kj}(\cdot)\) has bounded total variation, i.e. \(|Z_{kjl}(0)| + \int_0^\tau |dZ_{kj}(t)| \leq C_z\) for all \(j = 1, 2\) and \(k = 1, 2, \ldots, N\), where \(Z_{kjl}\) is the \(l\)th component of \(dZ_{kj}\) and \(C_z\) is a constant.

(C*5) \(A_j^c \equiv E\left[\int_0^\tau \{Z_{kj}(u) - \tilde{Z}_j(\beta_j, u)\} \otimes^2 Y_{kj}(u)e^{\beta_j^T Z_{kj}(u)} d\mu_{0j}(u)\right]\) is positive definite, where \(E\) is the expectation.

We summarize the asymptotic properties of \(\hat{\beta}_j^c\) in the following theorem, where the subscription \(c\) denote that the estimator is derived when the marginal model is multiplicative.

**Proof of Theorem 4.1**

Adapting A.2 in Lin et al. (2000), the partial likelihood score function for \(\beta_j\) is

\[L_j(\beta_j, \tau) = \sum_{k=1}^N \int_0^\tau \{Z_{kj}(u) - \tilde{Z}_j(\beta_j, u)\} dM_{kj}^c(u; \beta_j),\]

with \(M_{kj}^c(t; \beta_j) = N_{kj}(t) - \int_0^t Y_{kj}(u)e^{\beta_j^T Z_{kj}(u)}.\)

It is shown that \(N^{-1/2}L_j^c(\beta_j, t)(0 \leq t \leq \tau)\) converges weakly to a continuous zero-
mean Gaussian process with covariance function

\[ \Sigma_j(s, t) = E[\int_0^s \{ Z_{1j}(u) - \bar{z}_j(\beta_j, u)\} dM_{1j}^c(u) \int_0^t \{ Z_{1j}(v) - \bar{z}_j(\beta_j, v)\} dM_{1j}^c(v)], \]

\[ 0 \leq s, t \leq \tau, \]

between time points \( s \) and \( t \).

By Taylor series expansion,

\[ \sqrt{N}(\bar{\beta}_j - \beta_j) = \bar{A}_j^{-1}(\beta^*) N^{-1/2} \sum_{k=1}^N \{ Z_{kj}(u) - \tilde{Z}_j(\beta_j, u)\} dM_{kj}^c(u), \quad (A.43) \]

where \( \bar{A}_j(\beta_j) = -N^{-1}\partial L_j^c(\beta_j, \tau)/\partial \beta_j \), and \( \beta^* \) is on the line segment between \( \bar{\beta}_j \) and \( \beta_j \), with \( \bar{\beta}_j \) is the solution to \( L_j^c(\beta_j, \tau) = 0 \).

The almost sure convergence of \( \bar{\beta}_j \) and \( \bar{A}_j(\beta_j) \) for \( \beta_j \) and \( A_j^c \) imply that \( \sqrt{N}(\bar{\beta}_j - \beta_j) \) converges in distribution to a mean-zero normal random vector with covariance matrix \((A_j^c)^{-1}\Sigma_j(A_j^c)^{-1}\) and \( \Sigma_j = \Sigma_j(\tau, \tau) \). For future reference, we denote the asymptotic approximation as

\[ \sqrt{N}(\bar{\beta}_j - \beta_j) = (A_j^c)^{-1} N^{-1/2} \sum_{k=1}^N \xi_{kj}^c(u; \beta_j) + o_p(1). \quad (A.44) \]

where

\[ \xi_{kj}^c(u; \beta_j) = \int_0^\tau \{ Z_{kj}(u) - \bar{z}_j(u; \beta_j)\} dM_{kj}^c(u; \beta_j). \]

The consistency estimators of \( A_j \) and \( \Sigma_j \) are denoted by

\[ \bar{A}_j = N^{-1} \int_0^\tau \{ Z_{kj}(u) - \bar{Z}_j(\bar{\beta}_j, u)\} \otimes Y_{kj}(u) e^{\bar{\beta}_j Z_{kj}(u)} d\tilde{\mu}_0(u), \]

\[ \bar{\Sigma}_j = N^{-1} \sum_{k=1}^N \xi_{kj}^{\otimes 2}, \]
with

\[ \tilde{\xi}_{kj} = N^{-1} \sum_{k=1}^{N} \int_{0}^{\tau} \{ Z_{kj}(u) - \tilde{Z}_{kj}(u, \tilde{\beta}_j) \} \, d\tilde{M}_{kj}(u; \tilde{\beta}_j), \]

\[ \tilde{M}_{kj}(t; \tilde{\beta}_j) = N_{kj}(t) - \int_{0}^{t} Y_{kj}(u) e^{\tilde{\beta}_j t} Z_{kj}(u) \, d\tilde{\mu}_{0j}(u). \]  \quad (A.45)

\[ \square \]

**Proof of Theorem 4.2**

Let \( \tilde{\mu}_{0j}(t) \equiv \tilde{\mu}_{0j}(t, \tilde{\beta}_j) = \int_{0}^{t} \frac{d\tilde{N}_j(u)}{NS_{0j}(u, \beta_j)} \) and decompose \( \tilde{\mu}_{0j}(t) \) as

\[ \tilde{\mu}_{0j}(t) - \mu_{0j}(t) = \{ \tilde{\mu}_{0j}(t, \tilde{\beta}_j) - \tilde{\mu}_{0j}(t, \beta_j) \} + \{ \tilde{\mu}_{0j}(t, \beta_j) - \mu_{0j}(t) \}. \]  \quad (A.46)

The uniform strong law of large numbers Pollard (1990) implies \( S_{0j}(\beta_j, t) \to s_{0j}(\beta_j, t) \) and \( \tilde{N}_j(t)/N \to E[N_j(t)] \) uniformly in \( t \) and \( \beta_j \), and hence the uniform convergence of

\[ \tilde{\mu}_{0j}(t, \beta_j) = \int_{0}^{t} \frac{d\tilde{N}_j(u)}{NS_{0j}(u, \beta_j)} \] to \( \mu_{0j}(t) = \int_{0}^{t} \frac{s_{0j}(u, \beta_j)}{s_{0j}(u, \beta_j)} \, d\mu_{0j}(u) \). Furthermore, we can represent the second term in (A.46) as

\[ \tilde{\mu}_{0j}(t, \beta_j) - \mu_{0j}(t) = \int_{0}^{t} \frac{d\tilde{N}_j(u)}{NS_{0j}(u, \beta_j)} - d\mu_{0j}(u) \]

\[ = N^{-1} \int_{0}^{t} \frac{\sum_{k=1}^{N} dM_{kj}(u; \beta_j)}{S_{0j}(u, \beta_j)}, \]

\[ = N^{-1} \int_{0}^{t} \frac{\sum_{k=1}^{N} dM_{kj}(u; \beta_j)}{s_{0j}(u, \beta_j)} + o_p(N^{-1}). \]  \quad (A.47)
The first term in (A.46) can be rewritten as
\[
\tilde{\mu}_0(t, \tilde{\beta}_j) - \tilde{\mu}_0(t, \beta_j) = \int_0^t \frac{d\tilde{N}_j(u)}{NS\beta_j(u, \beta_j)} - \frac{d\tilde{N}_j(u)}{NS\beta_j(u, \beta_j)},
\]
\[
= \int_0^t -\tilde{Z}_j(u, \beta_j) \frac{d\tilde{N}_j(u)}{NS\beta_j(u, \beta_j)} (\tilde{\beta}_j - \beta_j) + o_p(N^{-1}),
\]
\[
= -\int_0^t \tilde{z}_j(u, \beta_j) d\mu_0(t, \beta_j)(\tilde{\beta}_j - \beta_j) + o_p(N^{-1}).
\]

The asymptotic approximation of \(\{\tilde{\beta}_j - \beta_j\}\) in (A.44) entails
\[
\tilde{\mu}_0(t, \tilde{\beta}_j) - \tilde{\mu}_0(t, \beta_j) = \left[H^c(t; \beta_j)\right]^T (A^c_j)^{-1} N^{-1} \sum_{k=1}^N \int_0^\tau \{Z_{kj}(u) - \tilde{z}_j(u, \beta_j)\} dM^c_{kj}(u; \beta_j) + o_p(N^{-1}),
\]

(A.48)

with \(H^c(t; \beta_j) = \int_0^t \tilde{z}_j(u, \beta_j) d\mu_0(u, \beta_j)\). Plugging (A.48), (A.47) into equation (A.46) and multiplying both sides by \(\sqrt{N}\) yield
\[
\sqrt{N} \left\{\tilde{\mu}_0(t) - \mu_0(t)\right\} = N^{-1/2} \sum_{k=1}^N \phi^c_{kj}(t; \beta_j) + o_p(1),
\]

(A.49)

where
\[
\phi^c_{kj}(t; \beta_j)
\]
\[
= \int_0^t \frac{dM^c_{kj}(u; \beta_j)}{s^\theta_j(u, \beta_j)} - \left[H^c(t; \beta_j)\right]^T (A^c_j)^{-1} \sum_{k=1}^N \int_0^\tau \{Z_{kj}(u) - \tilde{z}_j(u, \beta_j)\} dM^c_{kj}(u; \beta_j).
\]

(A.50)

Since \(\phi_{kj}(t)\) is independent mean-zero normal random variable, \(\sqrt{N} \left\{\tilde{\mu}_0(t) - \mu_0(t)\right\}\)
converges to a zero-mean Gaussian process with covariance function at \((s,t)\) as

\[
\Gamma_j(s,t) \equiv E[\phi^c_{kj}(s;\beta_j)\phi^c_{kj}(t;\beta_j)],
\] (A.51)

which can be approached by its consistent estimator

\[
\tilde{\Gamma}_j(s,t) = N^{-1} \sum_{k=1}^{N} \tilde{\phi}_{kj}(s;\beta_j)\tilde{\phi}_{kj}(t;\beta_j),
\]

where

\[
\tilde{\phi}_{kj}(t) = \int_0^t \frac{d\tilde{M}_{kj}(u)}{S_j^0(\beta_j,u)} - [\tilde{H}(t)]^T \tilde{\beta}^T \{Z_{kj}(u) - \tilde{Z}_{kj}(u,\beta_j)\} d\tilde{M}_{kj}(u;\tilde{\beta}_j),
\]

and

\[
\tilde{H}(t) = \int_0^t \tilde{Z}_j(u,\tilde{\beta}_j) d\tilde{\mu}_0(t,\tilde{\beta}_j).
\] (A.52)

\[\square\]

**Proof of Theorem 4.3**

Considering the decomposition:

\[
U^c_k(\theta,\tilde{\beta}_1,\tilde{\mu}_01(\cdot),\tilde{\beta}_2,\tilde{\mu}_02(\cdot)) - U_k(\theta,\beta_1,\mu_01(\cdot),\beta_2,\mu_02(\cdot))
\]

\[
= \left\{ U_k(\theta,\beta_1,\mu_01(s),\beta_2,\mu_02(\cdot)) - U_k(\theta,\beta_1,\mu_01(\cdot),\beta_2,\tilde{\mu}_02(t)) \right\}
\]

\[
+ \left\{ U_k(\theta,\beta_1,\mu_01(\cdot),\beta_2,\tilde{\mu}_02(\cdot)) - U_k(\theta,\beta_1,\mu_01(\cdot),\beta_2,\mu_02(\cdot)) \right\}
\] (A.53)

The first term on the right hand side of (A.53) is equivalent to

\[
\int_0^\tau \int_0^\tau -\rho(\theta,s,t) \frac{\partial \rho(\theta,s,t)}{\partial \theta} Y_{k2}(t)e^{\tilde{\beta}_2^T Z_{k2}(t)} d\tilde{\mu}_02(t)
\]

\[
\left\{ Y_{k1}(s)e^{\tilde{\beta}_1^T Z_{k1}(s)} d\tilde{\mu}_01(s) - Y_{k1}(s)e^{\beta_2^T Z_{k1}(s)} d\mu_01(s) \right\}
\] (A.54)
In (A.54), \( Y_{k1}(s)e^{\beta T Z_{k1}(s)} d\mu_0(s) - Y_{k1}(s)e^{\beta T Z_{k1}(s)} d\mu_0(s) \) can be further rewritten as

\[
Y_{k1}(s)\left\{ e^{\beta T Z_{k1}(s)} d\mu_0(s) - e^{\beta T Z_{k1}(s)} d\mu_0(s) + e^{\beta T Z_{k1}(s)} d\mu_0(s) - e^{\beta T Z_{k1}(s)} d\mu_0(s) \right\}
\]

\[
= Y_{k1}(s)\left\{ e^{\beta T Z_{k1}(s)} Z_{k1}(s) d\mu_0(s)(\hat{\beta}_1 - \beta_1) + e^{\beta T Z_{k1}(s)} (d\mu_0(s) - d\mu_0(s)) \right\}
\]

\[+ o_p(\beta_1 - \beta_1)^{0.2}\]

Applying the asymptotic properties of the first-stage estimators from (A.44) and (A.49) gives

\[
Y_{k1}(s)e^{\beta T Z_{k1}(s)} d\mu_0(s) - Y_{k1}(s)e^{\beta T Z_{k1}(s)} d\mu_0(s)
\]

\[
= Y_{k1}(s)\left\{ e^{\beta T Z_{k1}(s)} Z_{k1}(s) d\mu_0(s)(A_1) N^{-1} \sum_{l=1}^{N} \xi_{1l}^c + e^{\beta T Z_{k1}(s)} N^{-1} \sum_{l=1}^{N} d\phi_{1l}^c(s) \right\}
\]

\[+ o_p(N^{-1}). \quad (A.55)\]

By Combining (A.54) (A.55), and (A.57) we have

\[
\left\{ U_0^c(\theta, \tilde{\beta}_1, \bar{\mu}_0(s), \tilde{\beta}_2, \bar{\mu}_0(t)) - U_0^c(\theta, \beta_1, \mu_0(\cdot), \tilde{\beta}_2, d\bar{\mu}_0(t)) \right\}
\]

\[
= \int_{0}^{T} \int_{0}^{T} -p(\theta, s, t) \frac{\partial p(\theta, s, t)}{\partial \theta} Y_{k1}(s)e^{\beta T Z_{k1}(s)} \cdot Y_{k2}(t)e^{\beta T Z_{k2}(t)}
\]

\[
N^{-1} \sum_{l=1}^{N} \left\{ Z_{k1}(s) d\mu_0(s) d\mu_0(t)(A_1) N^{-1} \xi_{1l}^c + d\phi_{1l}^c(s) d\mu_0(t) \right\} + o_p(1) \quad (A.56)
\]

In a similar fashion,

\[
Y_{k2}(t)e^{\beta T Z_{k2}(t)} d\mu_0(t) - Y_{k2}(t)e^{\beta T Z_{k2}(t)} d\mu_0(t)
\]

\[
= Y_{k2}(t)\left\{ e^{\beta T Z_{k2}(t)} Z_{k2}(t) d\mu_0(t)(A_2) N^{-1} \sum_{l=1}^{N} \xi_{2l}^c + e^{\beta T Z_{k2}(t)} N^{-1} \sum_{l=1}^{N} d\phi_{2l}^c(t) \right\}
\]

\[+ o_p(N^{-1}). \quad (A.57)\]

Since the \( Y_{k2}(t)e^{\beta T Z_{k2}(t)} d\mu_0(t) \) and \( Y_{k1}(s)e^{\beta T Z_{k1}(s)} d\mu_0(s) \) only have \( o_p(N^{-1}) \) differ
ence compared to their true values, the product term has negligible difference of even higher orders.

The second part of (A.53) via a similar technique can be proved as

\[
\left\{ U^c_k(\theta, \beta_1, d\mu_0(s), \tilde{\beta}_2, d\tilde{\mu}_0(t)) - U^c_k(\theta, \beta_1, d\mu_0(s), \beta_2, d\mu_0(t)) \right\}
\]

\[
= \int_0^T \int_0^T -\rho(\theta, s, t) \frac{\partial \rho(\theta, s, t)}{\partial \theta} Y_{k1}(s)e^{\beta^T_1 Z_{k1}(s)} \cdot Y_{k2}(t)e^{\beta^T_2 Z_{k2}(t)}
\]

\[
\cdot N^{-1} \sum_{l=1}^N \left\{ Z_{k1}^T(t) d\mu_0(s) d\mu_0(t)(A^c_2)^{-1} \xi^c_{l2} + e^{\beta^T_2 Z_{l2}(t)} d\mu_0(s) d\phi^c_{l2}(t) \right\} + o_p(1)
\]

(A.58)

Since

\[
U^c(\theta, \tilde{\beta}_1, \tilde{\mu}_0(\cdot), \tilde{\beta}_2, \tilde{\mu}_0(\cdot)) - U^c(\theta, \beta_1, \mu_0(\cdot), \beta_2, \mu_0(\cdot))
\]

\[
= \sum_{k=1}^N \left\{ U^c_k(\theta, \tilde{\beta}_1, \tilde{\mu}_0(\cdot), \tilde{\beta}_2, \tilde{\mu}_0(\cdot)) - U^c_k(\theta, \beta_1, \mu_0(\cdot), \beta_2, \mu_0(\cdot)) \right\}
\]

by exchanging the order of the double summations, as well as switching the notations between l and k, it can be shown that

\[
N^{-1/2} \left\{ U^c(\theta, \tilde{\beta}_1, \tilde{\mu}_0(s), \tilde{\beta}_2, \tilde{\mu}_0(t)) - U^c(\theta, \beta_1, \mu_0(\cdot), \beta_2, \mu_0(\cdot)) \right\}
\]

\[
= N^{-1/2} \sum_{k=1}^N \left\{ h^c_{1,N}(A^c_1)^{-1} \xi^c_{k1} + g^c_{1,N} + h^c_{2,N}(A^c_2)^{-1} \xi^c_{k2} + g^c_{2,N} \right\} + o_p(1).
\]

(A.59)
where

\[ q_i^c(\theta, s, t) = -\rho(\theta, s, t) \frac{\partial \rho(\theta, s, t)}{\partial \theta} Y_{l1}(s)e^{\beta^T Z_{l1}(s)}Y_{l2}(t)e^{\beta^T Z_{l2}(t)}, \]

\[ h_{1,N}^c = N^{-1} \sum_{l=1}^{N} \int_{0}^{\tau} \int_{0}^{\tau} q_i^c(\theta, s, t) Z_{l1}^T(s) d\mu_{01}(s) d\mu_{02}(t), \]

\[ g_{1,N}^c = N^{-1} \sum_{l=1}^{N} \int_{0}^{\tau} \int_{0}^{\tau} q_i^c(\theta, s, t) d\mu_{02}(t) d\phi_{k1}(s), \]

\[ h_{2,N}^c = N^{-1} \sum_{l=1}^{N} \int_{0}^{\tau} \int_{0}^{\tau} q_i^c(\theta, s, t) Z_{l2}^T(t) d\mu_{02}(t) d\mu_{01}(s), \]

\[ g_{2,N}^c = N^{-1} \sum_{l=1}^{N} \int_{0}^{\tau} \int_{0}^{\tau} q_i^c(\theta, s, t) d\mu_{01}(s) d\phi_{k2}(t). \]  \hspace{1cm} (A.60)

□

**Proof of Theorem 4.4**

By the first order Taylor expansion of the estimation equation,

\[ \sqrt{N}(\hat{\theta} - \theta) \]

\[ = \left\{ -N^{-1} \frac{\partial U(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot))}{\partial \theta} \right\}^{-1} N^{-1/2} U(\theta, \hat{\beta}_1, \hat{\mu}_{01}(\cdot), \hat{\beta}_2, \hat{\mu}_{02}(\cdot)) + o_p(N^{-1/2}) \]  \hspace{1cm} (A.61)

Denote

\[ \mathcal{I}^c(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) = -N^{-1} \sum_{k=1}^{N} \left( \frac{\partial U_k(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot))}{\partial \theta} \right)^T \]  \hspace{1cm} (A.62)
and applying (A.59) and (A.62), (A.61) can be rewritten as

\[
\sqrt{N} (\hat{\theta} - \theta) = N^{-1/2} \{ \mathcal{I}^c(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) \}^{-1} \sum_{k=1}^{N} W_k^c(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) + o_p(1),
\]

where

\[
W_k^c(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) = \left\{ U_k^c(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) + h_{1,N}(A_1^{c})^{-1}\xi_{k1}^c + g_{1,N}^{c} + h_{2,N}(A_2^{c})^{-1}\xi_{k2}^c + g_{2,N}^{c} \right\}.
\]

(A.64)

By the central limit theorem that \( \sqrt{N}(\hat{\theta} - \theta) \) is asymptotically normal with mean 0 and a variance that can be estimated by \( \hat{\Phi} = N^{-1} \mathcal{I}^{-1}(\sum_{k=1}^{N} \hat{W}_k^{c \otimes 2})(\hat{\mathcal{I}}^T)^{-1} \), where \( \hat{\mathcal{I}} \) and \( \hat{W}_k \) are the empirical counterparts of

\[
\mathcal{I}(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot))
\]

and

\[
W_k(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)),
\]

respectively, obtained by plugging in the estimators of \( \tilde{\theta}, \tilde{\beta}_1, \tilde{\mu}_{01}(\cdot), \tilde{\beta}_2, \tilde{\mu}_{02}(\cdot) \).