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

LIST OF TABLES

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.

Aalen (1980) proposed semiparametric additive regression models for the rate function. Later literatures worked by McKeague and Sasieni, Martinussen and Scheike provide more comprehensive discussion of semiparametric additivie models. Studies based on Poisson and related processes have been discussed in literatures such as Andersen (1982) , Cheuvarte (1988), Lawless (1987a, 1987b) Thall (1988) Lawless and Nadeau (1995). Pepe and Cai (1993) considered robust methods for parametric or semeparametric regression analysis for the rate and mean functions. Lin et al. (2000) developed the asymptotic properties for the semiparametric regression analysis of Cox proportional mean functions whereas H Scheike (2002) considered the additive model.

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

$$\lambda_Z(t) = e^{\beta_0^T Z(t)} \lambda_0(t) \tag{1.1}$$

is a special case under the assumptions that (a) $E[dN^*(t)|\mathcal{F}_t] = E[dN^*(t)|Z(t)]$ and (b) $E[dN^*(t)|Z(t)] = e^{\beta_0^T Z(t)} \lambda_0(t)dt$.

Lin (2000) proposed a mean rate model

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

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

$$d\mu_Z(t) = e^{\beta_0^T Z(t)} \, d\mu_0(t),$$

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

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

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

$j = 1, 2, ..., J$, $k = 1, 2, ..., K$ where r_{kj} denote the multivariate random effect. With multivariate random effects, it is often assumed that conditional on r_{kj} and $\mathcal{F}_{kt} = \{N_{kj}(s), Z_{kj} : 0 \leq s \leq t\}$, type i and type j event are independent if $i \neq j$, that is

$$\lambda_{kj}(t|\mathcal{F}_{kt}, r_{kj}) = r_{kj}\lambda_{kj}(t|\mathcal{F}_{kt}) \tag{1.4}$$

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

$$\prod_{j=1}^{J} \Big\{ \prod_{l=1}^{n_{kj}} r_{kj}\lambda_{kj}(t_{kjl}|\mathcal{F}_{kt}) exp\Big(- r_{kj} \int_0^{\tau_k} \lambda_{kj}(u|\mathcal{F}_{kt})\, du\Big)\Big)\Big\}, \tag{1.5}$$

and the marginal likelihood for individual k as

$$\int \prod_{j=1}^{J} \Big\{ \prod_{l=1}^{n_{kj}} r_{kj}\lambda_{kj}(t_{kjl}|\mathcal{F}_{kt}) exp\Big(- r_{kj} \int_0^{\tau_k} \lambda_{kj}(u|\mathcal{F}_{kt})\, du\Big)\Big)\Big\} dG(r_k; \phi) \tag{1.6}$$

Analogous to the derivation above, we obtain Mixed Poisson Models as well as their overall and marginal likelihood function by letting $\lambda_{kj}(t|\mathcal{F}_{kt}) = \lambda_{kj}(t)$. 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

$$\rho(s,t) = \frac{\lambda_{1|2}(s|t)}{\lambda_1(s)}, \quad s,t \geq 0, \tag{1.7}$$

where the conditional rate function is defined as

$$\lambda_{1|2}(s|t) = \lim_{\Delta \to 0^+} Pr\{N_1(s+\Delta) - N_1(s) > 0 | N_2(t+\Delta) - N_2(t) > 0\}/\Delta \tag{1.8}$$

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 construct. 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 $N_{kj}^*(t)$ be a counting process registering the number of event occurrences by time t for the jth subject in cluster k (or equivalently the type j event for subject k), for $j = 1, 2$ and $k = 1, ..., N$. Suppose $\big(N_{k1}^*(s), N_{k2}^*(t)\big)$ 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

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

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

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

which is a measure of how the occurence 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,

$$E\{dN_{k1}^*(s)dN_{k2}^*(t)\,|Z_{k1}=z_{k1},Z_{k2}=z_{k2}\} = \rho(s,t;z_{k1},z_{k2})\mu_1(s;z_{k1})\mu_2(t;z_{k2})\,dsdt, \tag{2.2}$$

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 $\rho(s,t|z_{i1},z_{i2})$ 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 *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

Θ is a dimensional compact set. The estimating equation for θ can be constructed as

$$U(\theta, \hat{\mu}_1(\cdot; z_{k1}), \hat{\mu}_2(\cdot; z_{k2})) = \sum_{k=1}^{N} \int_0^{\tau} \int_0^{\tau} \frac{\partial \rho(s,t,\theta; z_{k1}, z_{k2})}{\partial \theta} \Big\{ dN_{k1}(s)\, dN_{k2}(t) - \rho(s,t,\theta; z_{k1}, z_{k2}) Y_{k1}(s) \hat{\mu}_1(s; z_{k1})\, Y_{k2}(t) \hat{\mu}_2(t; z_{k2})\, ds\, dt \Big\}. \tag{2.3}$$

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

$$\begin{aligned} & V(s, t, \hat{\theta}, \hat{\mu}_1(\cdot; z_{k1}), \hat{\mu}_2(\cdot; z_{k2})) \\ &= N^{-1/2} \sum_{k=1}^{N} \int_0^{t} \int_0^{s} W_n(u,v) \frac{\partial \rho(u,v,\theta; z_{k1}, z_{k2})}{\partial \theta} \Big\{ dN_{k1}(u)\, dN_{k2}(v) \\ &\quad - \rho(u,v,\theta; z_{k1}, z_{k2}) Y_{k1}(u) \hat{\mu}_1(u; z_{k1})\, Y_{k2}(v) \hat{\mu}_2(v; z_{k2})\, du\, dv \Big\}, \end{aligned} \tag{2.4}$$

$W_n(u, v)$ 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

$$E[dN^*_{kj}(t)|Z_{kj}(t)] = d\mu_j(t|Z_{kj}(t)),$$

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

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

In the following sections, we first review the estimation procedure of β_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

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

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

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

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

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

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

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

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

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 β_j as

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

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* β_j*, and has the following asymptotic approximation*

$$\sqrt{N}\{\hat{\beta}_j - \beta_j\} = A_j^{-1} N^{-1/2} \sum_{k=1}^{N} \xi_{kj} + o_p(1)$$

where $\xi_{kj} = \int_0^\tau \{Z_{kj}(u) - \bar{z}_j(u)\}\, dM_{kj}(u, \beta_j)$ *and* $\bar{z}_j(t) = \lim_{N\to\infty} \bar{Z}_j(t)$.

$\sqrt{N}(\hat{\beta}_j - \beta_j)$ *is asymptotically normal with mean zero and covariance matrix* $A_j^{-1}\Sigma_j A_j^{-1}$*, where*

$$A_j = E\{\int_0^\tau \{Z_{kj}(u) - \bar{z}_j(\beta_j, u)\}^{\otimes 2}\, ds\},$$
$$\Sigma_j = E[\int_0^\tau \{Z_{1j}(u) - \bar{Z}_j(u)\} dM_{1j}(u, \beta_j) \int_0^\tau \{Z_{1j}(v) - \bar{Z}_j(v)\}\, dM_{1j}(v, \beta_j)].$$

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*

$$\Gamma_j(s, t) = E[\phi_{kj}(s)\phi_{kj}(t)] \tag{3.7}$$

where

$$\phi_{kj}(t) = \int_0^t \pi_j^{-1}(u)dM_{kj}(u;\beta_j) - H^T(t)A_j^{-1}\int_0^\tau \{Z_{kj}(u) - \bar{z}_j(u)\}\, dM_{kj}(u;\beta_j), \quad (3.8)$$

with $H(t) = \int_0^t \bar{z}_j(u)du$, $\bar{z}_j^T(t) = \lim_{N\to\infty} \bar{Z}_j^T(t)$ *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 $\hat{\phi}_{kj}(t) = \int_0^t \hat{\pi}_j^{-1}(u)\, d\hat{M}_{kj}(u;\hat{\beta}_j) - \hat{H}^T(t)\hat{A}_j^{-1}\int_0^\tau \{Z_{kj}(u) - \bar{Z}_j(u)\}\, d\hat{M}_{kj}(u;\hat{\beta}_j)$, ,$\hat{\pi}_j(t) = N^{-1}\sum_{k=1}^N Y_{kj}(t)$ and $\hat{H}(t) = \int_0^t \bar{Z}_j(u)\, du$.

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

$$U(\theta, \hat{\beta}_1, \hat{\beta}_2, \hat{\mu}_{01}(\cdot), \hat{\mu}_{02}(\cdot)) = \sum_{k=1}^N U_k(\theta, \hat{\beta}_1, \hat{\beta}_2, \hat{\mu}_{01}(\cdot), \hat{\mu}_{02}(\cdot)), \quad (3.9)$$

where

$$U_k(\theta, \hat{\beta}_1, \hat{\beta}_2, \hat{\mu}_{01}(\cdot), \hat{\mu}_{02}(\cdot)) = \int_0^\tau\int_0^\tau \frac{\partial\rho(s,t,\theta;Z_{k1},Z_{k2})}{\partial\theta}\Big\{ dN_{k1}(s)\, dN_{k2}(t)$$
$$- \rho(s,t,\theta;Z_{k1},Z_{k2})Y_{k1}(s)[\, d\hat{\mu}_{01}(s) + \hat{\beta}_1^T Z_{k1}(s)\, ds]Y_{k2}(t)[\, d\hat{\mu}_{02}(t) + \hat{\beta}_2^T Z_{k2}(t)\, dt\,]\Big\}.$$

Denote $\hat{\theta}$ the solution to $U(\theta, \hat{\beta}_1, \hat{\beta}_2, \hat{\mu}_{01}(\cdot), \hat{\mu}_{02}(\cdot)) = 0$. We investigate the asymptotic properties of $U(\hat{\theta}, \hat{\beta}_1, \hat{\beta}_2, \hat{\mu}_{01}(\cdot), \hat{\mu}_{02}(\cdot))$ and $\hat{\theta}$ in Theorem 3.3 and 3.4 below.

Theorem 3.3 $N^{-1/2}\{U(\theta, \hat{\beta}_1, \hat{\beta}_2, \hat{\mu}_{01}(\cdot), \hat{\mu}_{02}(\cdot)) - U(\theta, \beta_1, \beta_2, \mu_{01}(\cdot), \mu_{02}(\cdot))\}$ *converges to a mean-zero Gaussian process, with covariance*

$$\Omega = \lim_{N\to\infty} N^{-1}\sum_{k=1}^N \Big\{h_{1,N}\xi_{k1}A_1^{-1} + g_{1,N,k} + h_{2,N}\xi_{k2}A_2^{-1} + g_{2,N,k}\Big\}^{\otimes 2}.$$

The consistent estimates of Ω *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*

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

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*

$$d\mu_j(t) = R_k \cdot \{d\mu_{0j}(t) + \beta_j^T Z_{kj}(t)\, dt\}, \tag{3.11}$$

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

$$\rho(s, t, \theta) = \rho = 1 + \frac{\sigma^2}{\mu^2}. \tag{3.12}$$

Proposition 2 *Let τ be the maximum observation time and c_0 lies in the middle of 0 and τ. Suppose the shared frailty mean rate model for $N^*_{kj}(t)$ is*

$$d\mu_j(t\,|Z_{kj}(t), R_k(t)) = R_k(t)\{d\mu_{0j}(t) + \beta_j^T Z_{kj}(t)\,dt\} \tag{3.13}$$

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, δ) as

$$f(x|a, b, \delta) = \frac{1}{\Gamma(a)b^a}(x-\delta)^{a-1}e^{\frac{-(x-\delta)}{b}}, x \in [\delta, \infty), \quad \delta \geq 0 \tag{3.14}$$

for $x \in [\delta, \infty)$, $\delta \geq 0$ and here $\Gamma(\cdot)$ denotes the Gamma function. Let X come from shifted Gamma(a, b, δ) then we have E(X)=$a \cdot b + \delta$ and 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, δ_0) and (a_1, b_1, δ_1), then the rate ratio is piecewise constant:

$$\begin{aligned}
\rho(\theta, s \leq c_0, t \leq c_0) &= 1 + \frac{a_0 b_0^2}{(a_0 b_0 + \delta_0)^2}, \\
\rho(\theta, s > c_0, t > c_0) &= 1 + \frac{a_1 b_1^2}{(a_1 b_1 + \delta_1)^2}, \\
\rho(\theta, s \leq c_0, t > c_0) &= \rho(\theta, s > c_0, t \leq c_0) = 1.
\end{aligned} \tag{3.15}$$

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,τ], and covariates Z_{kj} are from a uniform$[1, 2]$ for $j = 1, 2$. The observed events for the jth 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) Constant $\rho(s, t, \theta) = \theta_0$

Recall the shared frailty model in equation (3.11)

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

Let R_k follows i.i.d Gamma(a, b) with E$(R_k) = ab$ and 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 θ_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 β_{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 β_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 ith 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. θ_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 kth cluster, let

$$N_{kj}(t) = \tilde{N}_{kj}(t) + N_{k0}(t), \quad \text{for} \quad j = 1, 2 \tag{3.17}$$

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 $\lambda_{k0}(t|\, Z_{kj})\, dt = d\mu_{0j}(t) + \beta_{0j} Z_{kj}(t)$, and R_k is the frailty and variable is from a positive i.i.d Distribution. Let E(R_k)=μ_0 and var(R_k)=σ_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

$$\rho(\theta, s, t) = 1 + \theta_0 \times \frac{1}{(1+m_1(t))(1+m_2(s))}, \tag{3.18}$$

where $\theta_0 = \frac{\sigma_0^2}{\mu_0^2}$. 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

$$\rho(\theta, s, t) = 1 + \theta_0 \times (-0.15t + 0.9)(-0.15s + 0.9). \tag{3.19}$$

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 θ_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 θ_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

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

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

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

We take $\beta_1 = 0.5$, $\beta_2 = 1$, $\mu_{01}(t) = \mu_{02}(t) = 0.25t, 0.5t, 0.75t$. Let Z_k come from Bernoulli($p = 0.5$), so that X_k has equal chance to be 0 or 1. We generate R_k from Gamma$(4, 0.25)$ and 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).

(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 $\widetilde{N}_{kj}(\cdot)$ 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 $\widetilde{\lambda}_{kj}(t) = 0.5t$, 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 β_{0j} and $\mu_{0j}(t)$ plugged into equation (3.23), the estimates of σ^2/μ^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 θ.

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 (β_{01}, β_{02}). Each entry is based on 1000 simulations.

$\mu_{0j}(t)$	θ_0	N	Bias(β_{01}, β_{02})	SEE (β_{01}, β_{02})	ESE (β_{01}, β_{02})	CP(β_{01}, β_{02})
$0.75t$	1	200	(0.0097, 0.0127)	(0.1871, 0.2223)	(0.1902, 0.2335)	(0.9560, 0.9590)
		500	(0.0076, -0.0051)	(0.1183, 0.1464)	(0.1202, 0.1466)	(0.9500, 0.9500)
		800	(-0.0001, -0.0028)	(0.0944, 0.1159)	(0.0947, 0.1163)	(0.9540, 0.9460)
	1.25	200	(-0.0126, 0.0145)	(0.2938, 0.3876)	(0.2830, 0.3936)	(0.9450, 0.9580)
		500	(-0.0005, 0.0057)	(0.1746 0.2579)	(0.1803, 0.2497)	(0.9600, 0.9470)
		800	(0.0040, 0.0025)	(0.1478, 0.1976)	(0.1428, 0.1984)	(0.9450, 0.9540)
	1.5	200	(0.0106, 0.0110)	(0.3490, 0.5128)	(0.3524, 0.5073)	(0.9560, 0.9440)
		500	(0.0109, 0.0108)	(0.2337, 0.3232)	(0.2244, 0.3217)	(0.9440, 0.9610)
		800	(0.0039, 0.0254)	(0.1826, 0.263)1	(0.1773, 0.2550)	(0.9470, 0.9460)
	1.75	200	(0.0027, -0.0388)	(0.4227, 0.6018)	(0.4096, 0.5902)	(0.9540, 0.9490)
		500	(0.0067, -0.0002)	(0.2623, 0.3876)	(0.2597, 0.3790)	(0.9570, 0.9480)
		800	(-0.0019, 0.0056)	(0.2042, 0.2906)	(0.2072, 0.3024	(0.9470, 0.9580)
t	1	200	(0.0091, -0.0117)	(0.2041, 0.2467)	(0.2048, 0.2433)	(0.9500, 0.9580)
		500	(-0.0087, -0.0027)	(0.1265, 0.1538)	(0.1293, 0.1545)	(0.9480, 0.9530)
		800	(0.0050, -0.0011)	(0.1009, 0.1246)	(0.1022, 0.1223)	(0.9500, 0.9600)
	1.25	200	(-0.0110, -0.0028)	(0.3289, 0.4274)	(0.3202, 0.4282)	(0.9410, 0.9560)
		500	(-0.0039, -0.0064)	(0.2022, 0.2755)	(0.2037, 0.2730)	(0.9460, 0.9550)
		800	(-0.0013, 0.0005)	(0.1656, 0.2209)	(0.1611, 0.2164)	(0.9460, 0.9530)
	1.5	200	(0.0295, -0.0237)	(0.4212, 0.5470)	(0.4010, 0.5476)	(0.9520, 0.9530)
		500	(0.0150, 0.0041)	(0.2649, 0.3599)	(0.2566, 0.3519)	(0.9510, 0.9450)
		800	(-0.0039, -0.0123)	(0.2067, 0.2745)	(0.2029, 0.2793)	(0.9540, 0.9580)
	1.75	200	(0.0181, 0.0012)	(0.4745, 0.6696)	(0.4692, 0.6541)	(0.9480, 0.9500)
		500	(0.0093, 0.0208)	(0.3089, 0.4311)	(0.3007, 0.4203)	(0.9500, 0.9450)
		800	(0.0005, -0.0082)	(0.2418, 0.3357)	(0.2381, 0.3333)	(0.9570, 0.9530)

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.

$\mu_{0j}(t)$	θ_0	N	Bias	SEE	ESE	CP	$\mu_{0j}(t)$	θ_0	N	Bias	SEE	ESE	CP
$0.25t$	1.25	200	-0.0031	0.0845	0.0818	0.9350	$0.75t$	1.25	200	-0.0011	0.0770	0.0726	0.9330
		500	0.0005	0.0567	0.0536	0.9420			500	-0.0011	0.0520	0.0476	0.9300
		800	0.0001	0.0432	0.0428	0.9520			800	-0.0001	0.0395	0.0385	0.9420
	1.50	200	-0.0131	0.1392	0.1248	0.8960		1.50	200	-0.0079	0.1280	0.1179	0.8970
		500	-0.0024	0.0896	0.0849	0.9200			500	-0.0025	0.0851	0.0791	0.9290
		800	-0.0045	0.0672	0.068	0.9500			800	-0.0041	0.0667	0.0636	0.9310
	1.75	200	-0.0160	0.1926	0.1715	0.8930		1.75	200	-0.0126	0.1978	0.1659	0.8850
		500	0.0006	0.1311	0.1212	0.9130			500	-0.0082	0.1212	0.1121	0.9160
		800	-0.0049	0.1031	0.0962	0.9290			800	.-0.0018	0.0954	0.0914	0.9260
$0.5t$	1.25	200	-0.0016	0.0858	0.0762	0.9090	t	1.25	200	-0.0057	0.0747	0.0699	0.9230
		500	-0.0019	0.0517	0.0500	0.9370			500	0.0003	0.0472	0.0465	0.9410
		800	-0.0009	0.0402	0.0402	0.9460			800	-0.0012	0.0379	0.0374	0.9460
	1.5	200	-0.0138	0.1361	0.1192	0.8830		1.50	200	-0.0125	0.1235	0.1127	0.9130
		500	0.0002	0.0862	0.0827	0.9310			500	-0.0026	0.0847	0.0777	0.9050
		800	-0.0014	0.0672	0.0652	0.9290			800	-0.0052	0.0658	0.0621	0.9210
	1.75	200	-0.0101	0.1916	0.1656	0.8840		1.75	200	-0.0154	0.1742	0.1577	0.8840
		500	0.0015	0.1224	0.1172	0.9160			500	0.0031	0.1207	0.1135	0.9250
		800	-0.0047	0.1019	0.0933	0.9270			800	0.0018	0.0980	0.0912	0.9300

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 θ_0 in $\rho(s,t,\theta)$. Each entry is based on 1000 simulations with correctly specified marginals and Rate Ratio form.

$\mu_{0j}(t)$	θ_0	N	Bias	SEE	ESE	CP	$\mu_{0j}(t)$	θ_0	N	Bias	SEE	ESE	CP
$0.5t$	0.25	200	0.0114	0.1793	0.1697	0.9410	t	0.25	200	-0.0073	0.1205	0.1207	0.9450
		500	0.0002	0.1134	0.1089	0.9440			500	0.0029	0.0760	0.0781	0.9520
		800	0.0001	0.0891	0.0860	0.9520			800	0.0007	0.0613	0.0620	0.9490
	0.50	200	-0.0053	0.2325	0.2162	0.9260		0.5	200	-0.0047	0.1832	0.1698	0.9160
		500	-0.0038	0.1427	0.1406	0.9310			500	-0.0036	0.1124	0.1099	0.9310
		800	-0.0026	0.1156	0.1118	0.9360			800	-0.0038	0.0904	0.0881	0.9290
	1	200	-0.0096	0.3406	0.3295	0.9170		1	200	-0.0193	0.2988	0.2811	0.8960
		500	-0.0066	0.2120	0.2180	0.9370			500	-0.0048	0.1779	0.1890	0.9490
		800	0.0007	0.1697	0.1765	0.9440			800	-0.0031	0.1472	0.1518	0.9420
	1.6	200	-0.0112	0.4884	0.4857	0.9150		1.6	200	-0.0101	0.4317	0.4388	0.9060
		500	-0.0120	0.3338	0.3325	0.9140			500	-0.0065	0.3049	0.3030	0.9130
		800	-0.0073	0.2697	0.2685	0.9240			800	-0.0085	0.2427	0.2452	0.9370
	2	200	-0.0324	0.6739	0.6011	0.8740		2	200	-0.0270	0.6112	0.5526	0.8780
		500	-0.0166	0.4226	0.4164	0.9040			500	-0.0249	0.3853	0.3821	0.9100
		800	-0.0059	0.3300	0.3431	0.9300			800	-0.0053	0.3039	0.3198	0.9390

Table 4: Scenario III - Estimation of θ'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.

		θ_1				θ_2			
$\mu_{0j}(t)$	N	Bias	SEE	ESE	CP	Bias	SSE	ESE	CP
$0.25t$	200	-0.0094	0.1284	0.1177	0.9250	-0.0012	0.4543	0.3610	0.8860
	500	-0.0046	0.0817	0.0776	0.9280	-0.0033	0.2675	0.2476	0.9190
	800	-0.0028	0.0653	0.0626	0.9430	-0.0043	0.2218	0.2016	0.9130
	1100	0.0010	0.0561	0.0545	0.9510	0.0151	0.1906	0.1798	0.9380
$0.50t$	200	-0.0055	0.1193	0.1075	0.9190	-0.0160	0.3267	0.2703	0.8720
	500	-0.0046	0.0765	0.0715	0.9320	-0.0226	0.2118	0.1873	0.8950
	800	-0.0013	0.0597	0.0583	0.9530	-0.0081	0.1674	0.1542	0.9180
	1100	-0.0003	0.0494	0.0502	0.9640	-0.0066	0.1408	0.1355	0.9240
$0.75t$	500	-0.0103	0.1131	0.1011	0.8960	-0.0182	0.2826	0.2469	0.8820
	500	-0.0020	0.0722	0.0689	0.9340	-0.0095	0.1933	0.1706	0.9110
	800	-0.0009	0.0564	0.0552	0.9370	-0.0037	0.1512	0.1401	0.9200
	1100	-0.0020	0.0486	0.0470	0.9420	0.0033	0.1251	0.1231	0.9330

Table 5: Scenario IV - estimates θ_1, θ_2 in the underline models where $\rho(\theta, s, t|Z_k) = 1 + \theta_1 \frac{(0.25+0.1Z_k)(0.25+0.2Z_k)}{(0.5t+0.25+0.1Z_k)(0.5s+0.25+0.2Z_k)}$ and $\rho(\theta, s, t|Z_k) = 1+\theta_2 \frac{(0.5+0.1Z_k)(0.5+0.2Z_k)}{(0.5t+0.5+0.1Z_k)(0.5s+0.5+0.2Z_k)}$. Summary of Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP of θ where each entry is based on 1000 simulations. The averaged observed events for type1 (2) event is 2.44(2.56)

θ_1	N	Bias	SEE	ESE	CP	θ_2	Bias	SEE	ESE	CP
0.25	200	-0.0017	0.4030	0.4401	0.9580	0.25	-0.0027	0.2143	0.2513	0.9700
	500	-0.0221	0.2529	0.2840	0.9650		0.0005	0.1320	0.1625	0.9810
	800	0.0014	0.1916	0.2255	0.9770		-0.0020	0.1078	0.1287	0.9790
	1100	0.0038	0.1728	0.1933	0.9720		-0.0001	0.0897	0.1104	0.9820
0.50	200	-0.0177	0.4404	0.4948	0.9730	0.50	-0.0202	0.2703	0.3050	0.9560
	500	-0.0178	0.2843	0.3201	0.9570		-0.0047	0.1703	0.2001	0.9670
	800	-0.0001	0.2235	0.2570	0.9750		-0.0001	0.1337	0.1605	0.9820
	1100	0.0026	0.2003	0.2197	0.9660		0.0014	0.1123	0.1378	0.9850
0.75	200	-0.0154	0.5316	0.5598	0.9510	0.75	-0.0438	0.3170	0.3607	0.9470
	500	-0.0031	0.3184	0.3641	0.9680		-0.0079	0.2029	0.2443	0.9750
	800	-0.0105	0.2513	0.2899	0.9720		0.0030	0.1554	0.1952	0.9830
	1100	-0.0050	0.2156	0.2496	0.9720		-0.0029	0.1385	0.1671	0.9720
1.00	200	0.0008	0.5731	0.6270	0.9630	1.00	-0.0282	0.3863	0.4341	0.9530
	500	0.0002	0.3891	0.4154	0.9590		-0.0047	0.2473	0.2920	0.9590
	800	-0.0007	0.2939	0.3308	0.9730		-0.0019	0.1955	0.2331	0.9740
	1100	0.0010	0.2584	0.2834	0.9620		-0.0107	0.1614	0.1985	0.9740

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. $H_0 : \rho(s, t, \theta; z_1, z_2) = \theta_0$, 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

$$
\begin{aligned}
&V(s, t, \hat{\theta}, \hat{\beta}_1, \hat{\mu}_{01}(\cdot), \hat{\beta}_2, \hat{\mu}_{02}(\cdot)) \\
&= N^{-1/2} \sum_{k=1}^{N} \int_0^t \int_0^s W_N(u, v) \frac{\partial \rho(u, v, \theta)}{\partial \theta}\bigg|_{\theta=\hat{\theta}} \Big\{ dN_{k1}(u) dN_{k2}(v) \\
&\quad - \rho(u, v, \hat{\theta}) Y_{k1}(u) \{ d\hat{\mu}_{01}(u) + \hat{\beta}_1^T Z_{k1}(u)\, du \} Y_{k2}(v) \{ d\hat{\mu}_{02}(v) + \hat{\beta}_2^T Z_{k2}(v)\, dv \} \Big\},
\end{aligned}
\tag{3.25}
$$

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 $\rho(s, t, \theta_0; z_{k1}, z_{k2})$, one would expect the value of equation (3.25) to fluctuate around zero at the any $(s, t) \in [0, \tau]^2$.

Let $T = \sup_{s,t \in [0,\tau]^2} \parallel V(s, t, \hat{\theta}, \hat{\beta}_1, \hat{\mu}_{01}(\cdot), \hat{\beta}_2, \hat{\mu}_{02}(\cdot)) \parallel$ 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 θ is

$$
\begin{aligned}
V(s,t,\hat{\theta},\hat{\beta}_1,\hat{\mu}_{01}(\cdot),\hat{\beta}_2,\hat{\mu}_{02}(\cdot)) = & V\Big(s,t,\theta,\hat{\beta}_1,\hat{\mu}_{01}(\cdot),\hat{\beta}_2,\hat{\mu}_{02}(\cdot)\Big) \\
& + N^{-1/2}\frac{\partial V\Big(s,t,\theta,\hat{\beta}_1,\hat{\mu}_{01}(\cdot),\hat{\beta}_2,\hat{\mu}_{02}(\cdot)\Big)}{\partial\theta}N^{1/2}(\hat{\theta}-\theta) \\
& + o_p(1), \qquad (3.26)
\end{aligned}
$$

which can be further decomposed as

$$
\begin{aligned}
& V\Big(s,t,\hat{\theta},\hat{\beta}_1,\hat{\mu}_{01}(\cdot),\hat{\beta}_2,\hat{\mu}_{02}(\cdot)\Big) \\
& = V(s,t,\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot)) \\
& \quad + N^{-1/2}\sum_{k=1}^{N}\Big\{\Upsilon_{k1}(s,t,\theta)+\Upsilon_{k2}(s,t,\theta)+\zeta_{k1}(s,t,\theta)+\zeta_{k2}(s,t,\theta)\Big\}+o_p(1), \qquad (3.27)
\end{aligned}
$$

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

$$
\begin{aligned}
& V^*(s,t,\hat{\theta}) \\
& = \Big\{V(s,t,\hat{\theta},\hat{\beta}_1,\hat{\mu}_{01}(\cdot),\hat{\beta}_2,\hat{\mu}_{02}(\cdot)) \\
& \quad + N^{-1/2}\textstyle\sum_{k=1}^{N}\hat{\Upsilon}_{k1}(s,t,\hat{\theta})+\hat{\zeta}_{k1}(s,t,\hat{\theta})+\hat{\Upsilon}_{k2}(s,t,\hat{\theta})+\hat{\zeta}_{k2}(s,t,\hat{\theta})\Big\}G_k, \qquad (3.28)
\end{aligned}
$$

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 $\hat{V}(s,t,\hat{\theta},\hat{\mu}_1(\cdot;Z_{k1}),\hat{\beta}_1,\hat{\mu}_{01}(\cdot),\hat{\beta}_2,\hat{\mu}_{02}(\cdot))$ 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 $H_0 : \rho = 1 \quad \text{vs} \quad H_a : \rho \neq 1$. 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 θ_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),

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

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

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

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

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, δ_0) and Gamma(a_1, b_1, δ_1) respectively. The shifted Gamma Distribution with (a, b, δ) 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 ρ values followed from Proposition 2.

Settings	PWC1	PWC2	PWC3	PWC4
$R_{k0}: \ (a_0, b_0, \delta_0)$	(0.25,1,0.75)	(0.5,1,0.5)	(0.25,2,0.5)	(0.25,2,0.5)
$R_{k1}: \ (a_1, b_1, \delta_1)$	(0.25,1,0.75)	(0.25,1,0.75)	(0.5,1,0.5)	(0.25,1,0.75)
$\rho(s < 2.5, t < 2.5)$	1.25	1.5	2	2
$\rho(s > 2.5, t < 2.5)$	1	1	1	1
$\rho(s > 2.5, t > 2.5)$	1.25	1.25	1.5	1.25

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

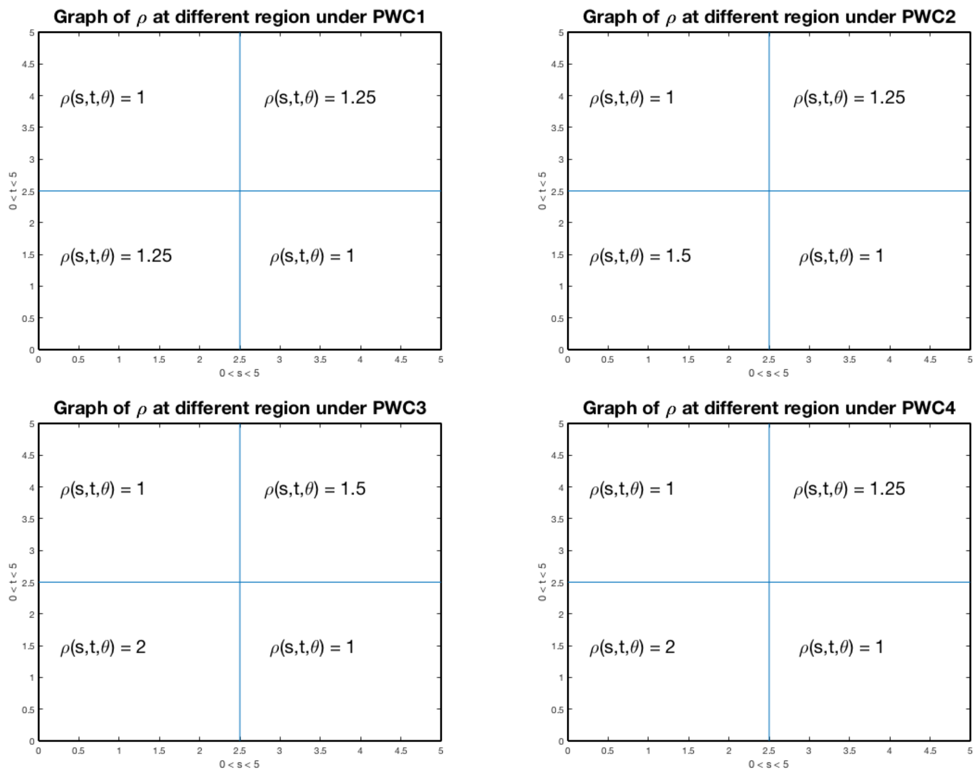
Assuming $\{N_{k1}(s), N_{k2}(t)\}$ follows the Bivariate Counting processes below

$$\begin{aligned} N_{k1}(s) &= \widetilde{N}_{k1}(s) + N_{k0}(s), \\ N_{k2}(t) &= \widetilde{N}_{k2}(t) + N_{k0}(t), \end{aligned} \tag{3.30}$$

where $\widetilde{N}_{k1}(\cdot)$, $\widetilde{N}_{k2}(\cdot)$ 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 μ and variance σ^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 $\widetilde{\lambda}_{kj}(t|Z_{kj}(t)) = m_j(t)\lambda_{k0}(t|Z_{kj}(t))$, with a nonnegative multiplier function $m_j(t)$. Following simulation settings in equation 3.19 to generate data that share the rate ratio as

$$\text{TD model: } \rho(\theta, s, t) = 1 + \theta_0 \times (-0.15s + 0.9)(-0.15t + 0.9). \tag{3.31}$$

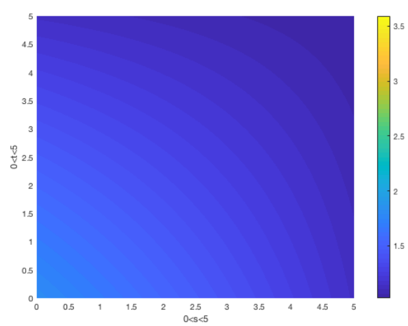
where $\theta_0 = \frac{\sigma^2}{\mu^2}$ reflects the time varying component in $\rho(\theta, s, t)$ proportionally. To capture different time varying levels, we take R_k from a shifted gamma distribution, with parameters $(a, b, \delta) = (0.25, 2, 0.5)$,$(0.2, 3, 0.4)$ $(0.25, 3, 0.25)$ and $(0.2, 4, 0.2)$ 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.

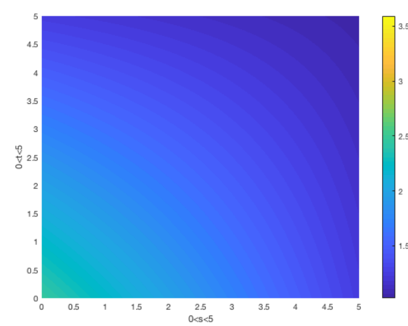
Table 7: Simulation settings of the Time Varying Rate Ratio (TD models). From TD1 to TD4, the value of σ^2/μ^2 is increasing and so is the association between the bivariate recurrent event processes.

Settings	TD1	TD2	TD3	TD4
(μ, σ^2)	(1, 1)	(1, 1.8)	(1, 2.25)	(1, 3.2)
$\frac{\sigma^2}{\mu^2}$	1	1.8	2.25	3.2

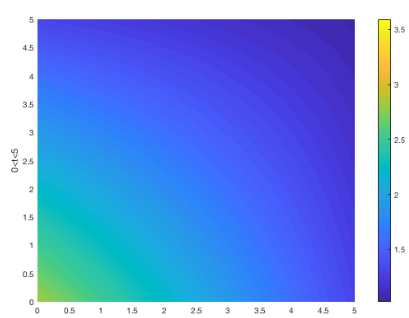
Figure 2: The contour plot of the Rate Ratio $\rho(s,t)$ under the additive marginal mean rate models. The x-axis and y-axis represents the observation time for type1 and type2 events. From upper left to lower right, the heterogeneity of $\rho(s,t)$ is increased.



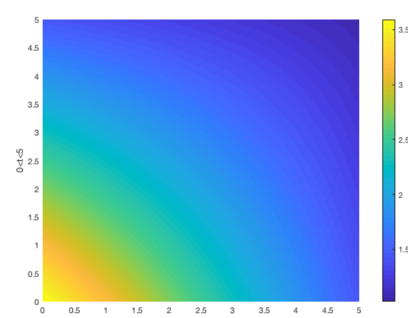
(a) TD1 contour plot



(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 $\widetilde{N}_{kj}(t)$ are Poisson processes with rate conditional on covariates and unobservable frailty R_k as $\lambda_{k0}(t|Z_{kj}, R_k) = R_k\{0.25 + \beta_{0j}Z_{kj}\}$ and $\widetilde{\lambda}_{kj}(t) = t$ 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 $\mathrm{E}(R_k) = 1$ and $\mathrm{var}(R_k) = 0.5, 0.8, 1, 2$. Denoted by $\rho(\theta, s, t|Z_{k1}, Z_{k2})$ the rate ratio of $\{N_{k1}(s), N_{k2}(t)\}$, where

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

is obtained by Proposition 3, with true θ 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 $\mathcal{G}$ composed by i.i.d Standard Gaussian random numbers, so that each row is an N dimensional vector:
$$\begin{bmatrix} G_{11} & G_{12} & G_{13} & \dots & G_{1n} \\ G_{21} & G_{22} & G_{23} & \dots & G_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ G_{B1} & G_{B2} & G_{B3} & \dots & G_{Bn} \end{bmatrix}$$
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 $Ha : \rho(s, t, \theta) = \theta$ 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.

		Size		
event count	ρ	N=200	N=500	N=800
(2.50, 4.37)	1	0.043	0.052	0.051
(3.13, 5.02)	1	0.051	0.057	0.051
(3.76, 5.64)	1	0.043	0.053	0.041
(4.37, 6.26)	1	0.045	0.049	0.054
event count		Power		
	ρ	N=200	N=500	N=800
(2.50, 4.37)	1.25	0.995	1.000	1.000
	1.5	1.000	1.000	1.000
	1.75	1.000	1.000	1.000
	2	1.000	1.000	1.000

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 $Ha : \rho$ is not constant, at 0.05 significance level. Each entry is calculated based on 1000 Gaussian multiplier samples with 1000 replicates.

		Size			
event count	N	$\rho = 1.25$	$\rho = 1.5$	$\rho = 2$	$\rho = 2.25$
	200	0.038	0.038	0.031	0.038
(3.50, 4.67)	500	0.057	0.037	0.032	0.040
	800	0.042	0.042	0.051	0.046
		Power			
event count	N	PWC1	PWC2	PWC3	PWC4
	200	0.173	0.579	0.638	0.755
(2.91, 4.80)	500	0.421	0.912	0.958	0.983
	800	0.622	0.979	0.990	0.999
		TD1	TD2	TD3	TD4
	200	0.197	0.231	0.311	0.307
(4.27, 4.27)	500	0.556	0.621	0.760	0.738
	800	0.773	0.821	0.887	0.894
		TCD1	TCD2	TCD3	TCD4
	200	0.250	0.336	0.405	0.455
(6.67, 8.54)	500	0.514	0.678	0.756	0.823
	800	0.735	0.900	0.917	0.933

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

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

where β_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 $N^*_{k2}(s)$. θ 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

$$\begin{aligned} d\bar{N}_{\cdot j}(t) &= \sum_{k=1}^{N} dN_{kj}(t), \\ M_{kj}^{c}(t) &= N_{kj}(t) - \int_0^t Y_{kj}(u) e^{\beta_j^T Z_{kj}(u)}\, d\mu_{01}(u), \\ S_j^d(t,\beta) &= N^{-1} \sum_{k=1}^{N} Y_{kj}(t) Z_{kj}^{\otimes d}(t) e^{\beta^T Z_{kj}(t)}, \qquad d = 0,1,2 \end{aligned} \tag{4.2}$$

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

Denote $\tilde{\beta}_j$ the solution to $L_j^c(\beta,\tau) = 0$, where $L_j^c(\beta,\tau) = \sum_{k=1}^{N} \int_0^{\tau} \{Z_{kj}(u) - \tilde{Z}_j(u,\beta)\} dN_{kj}(u)$ is the partial likelihood score function.

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

$$\tilde{\mu}_{0j}(t,\tilde{\beta}_j) = \int_0^t \frac{d\bar{N}_j(u)}{N S_j^0(u,\tilde{\beta}_j)}, \quad t \in [0,\tau]. \tag{4.3}$$

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 β_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*

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

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

The asymptotic approximation of $\tilde{\beta}_j$ *is*

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

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

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

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

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

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

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*

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

where

$$\phi_{kj}^c(t) = \int_0^t \frac{dM_{kj}^c(u;\beta_j)}{s_j^0(u,\beta_j)} - H^T(t;\beta_j)(A_j^c)^{-1} \int_0^\tau \{Z_{kj}(u) - \tilde{z}_{kj}(u,\beta_j)\}\, d\tilde{M}_{kj}(u), \quad k = 1, ..., N \quad (4.6)$$

and

$$H(t;\beta_j) = \int_0^t \tilde{z}_j(u,\beta_j)d\mu_{0j}(u) \tag{4.7}$$

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

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

where

$$\tilde{\phi}_{kj}(t) = \int_0^t \frac{d\tilde{M}_{kj}(u)}{S_j^0(u,\tilde{\beta}_j)} - \tilde{H}^T(t;\tilde{\beta}_j)\tilde{A}_j^{-1}\int_0^\tau \{Z_{kj}(u) - \tilde{Z}_{kj}(u,\tilde{\beta}_j)\}\, d\tilde{M}_{kj}(u)$$

and

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

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

$$U^c\Big(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot)\Big) = \sum_{k=1}^{N} U_k^c\Big(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot)\Big), \tag{4.9}$$

where

$$\begin{aligned} &U_k^c\Big(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot)\Big)\\ &= \int_0^\tau\int_0^\tau \frac{\partial\rho(\theta,s,t)}{\partial\theta}\cdot\\ &\quad \Big\{dN_{k1}(s)\,dN_{k2}(t) - \rho(\theta,s,t)Y_{k1}(s)e^{\beta_1^T Z_{k1}(s)}\,d\mu_{01}(s)Y_{k2}(t)e^{\beta_2^T Z_{k2}(t)}\,d\mu_{02}(t)\Big\}. \end{aligned} \tag{4.10}$$

with $\beta_1,\ \beta_2,\ \mu_{01}(\cdot),\ \mu_{02}(\cdot)$ replaced by estimator $\tilde{\beta}_1,\ \tilde{\beta}_2,\ \tilde{\mu}_{01}(\cdot),\ \tilde{\mu}_{02}(\cdot)$ from the first stage.

The resulting estimator $\tilde{\theta}$ does not have an explicit form. We adapt the asymptotic properties of $\tilde{\beta}_j$ and $\tilde{\mu}_{0j}(\cdot)$ from Theorem 4.1 and Theorem 4.2 to show the weak covergence of $\tilde{\theta}$.

Theorem 4.3 $N^{-1/2}\Big\{U^c\Big(\theta, \tilde{\beta}_1, \tilde{\mu}_{01}(\cdot), \tilde{\beta}_2, \tilde{\mu}_{02}(\cdot)\Big) - U^c\Big(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)\Big)\Big\}$ *follows a mean zero Gaussian process and has the following approximation*

$$\begin{aligned} & N^{-1/2}\{U^c(\theta, \tilde{\beta}_1, d\tilde{\mu}_{01}(\cdot), \tilde{\beta}_2, d\tilde{\mu}_{02}(\cdot)) - U^c(\theta, \beta_1, d\mu_{01}(\cdot), \beta_2, d\mu_{02}(\cdot))\} \\ &= N^{-1/2}\sum_{k=1}^{N}\Big\{h_{1,N}^c(A_1)^{-1}\xi_{k1}^c + g_{1,N,k}^c + h_{2,N}^c(A_2)^{-1}\xi_{k2} + g_{2,N,k}^c\Big\} + o_p(N^{-1/2}), \quad (4.11) \end{aligned}$$

where

$$\begin{aligned} q_l^c(\theta, s, t) &= -\rho(\theta, s, t)\frac{\partial \rho(\theta, s, t)}{\partial \theta} Y_{l1}(s) e^{\beta_1^T Z_{l1}(s)} Y_{l2}(t) e^{\beta_2^T Z_{l2}(t)}, \\ h_{1,N}^c &= N^{-1}\sum_{l=1}^{N}\int_0^\tau\int_0^\tau q_l^c(\theta, s, t) Z_{l1}^T(s)\, d\mu_{01}(s) d\mu_{02}(t), \\ h_{2,N}^c &= N^{-1}\sum_{l=1}^{N}\int_0^\tau\int_0^\tau q_l^c(\theta, s, t) Z_{l2}^T(s)\, d\mu_{02}(t) d\mu_{01}(s), \\ g_{1,N,k}^c &= N^{-1}\sum_{l=1}^{N}\int_0^\tau\int_0^\tau q_l^c(\theta, s, t)\, d\mu_{02}(t)\, d\phi_{k1}^c(s), \\ g_{2,N,k}^c &= N^{-1}\sum_{l=1}^{N}\int_0^\tau\int_0^\tau q_l^c(\theta, s, t)\, d\mu_{01}(s)\, d\phi_{k2}^c(t). \end{aligned} \quad (4.12)$$

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

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

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*

$\tilde{\phi}_{kj}(t)$ *into equation (4.12).*

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

$$\sqrt{N}(\tilde{\theta}-\theta)$$
$$= N^{-1/2}\{\mathcal{I}^c(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))\}^{-1}\sum_{k=1}^{N} W_k^c(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot)) + o_p(1) \quad (4.13)$$

where

$$\mathcal{I}^c(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot)) = -N^{-1}\sum_{k=1}^{N}\Big\{\frac{\partial U_k^c(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))}{\partial\theta}\Big\}^T, \quad (4.14)$$

and

$$W_k^c(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))$$
$$= U_k^c(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot)) + \Big\{h_{1,N}^c(A_1^c)^{-1}\xi_{k1}^c + g_{1,N,k}^c + h_{2,N}^c(A_2^c)^{-1}\xi_{k2}^c + g_{2,N,k}^c\Big\}. \quad (4.15)$$

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

$$\mathcal{I}^c(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))$$
$$W_k^c(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))$$

respectively, obtained by substituting $\tilde{\theta}$, $\tilde{\beta}_1$, $\tilde{\mu}_{01}(\cdot)$, $\tilde{\beta}_1$, $\tilde{\mu}_{02}(\cdot)$ *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, $E[dN^*_{kj}(t)|R_k, Z_{kj}(t)] = R_k\{e^{\beta_j^T Z_{kj}(t)}\, d\mu_{0j}(t)\}$, 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 Proposition1 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 β_1, β_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 ρ 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 ρ 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

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

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

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

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 type1 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

$$\rho(\theta, s, t) = 1 + \theta(-0.15s + 0.9)(-0.15t + 0.9), \tag{4.16}$$

with the parameter θ 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 θ_1 and θ_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 θ_2 are consistently larger than that of θ_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 $N_{kj}(t) = \widetilde{N}_{kj}(t) + N_{k0}(t)$. Let

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

$$E\{d\widetilde{N}_{kj}(t)|Z_{kj}, R_k\} = \widetilde{\lambda}_{kj}(t)\, dt$$

where $\lambda_{k0}(t|Z_{kj}, R_k) = R_k e^{\beta_{0j} Z_{kj}} 0.25t$ and $\widetilde{\lambda}_{kj}(t) = 0.25$.

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

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

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 $\lambda_{k0}(t|Z_{kj}, R_k) = R_k \cdot 0.5e^{\beta_{0j} Z_{kj}}$. 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

$$\rho(\theta, s, t|Z_k) = 1 + \theta \frac{(0.5t\, e^{0.1Z_{k1}})(0.5s\, e^{0.2Z_{k2}})}{(0.25 + 0.5t\, e^{0.1Z_{k2}})(0.25 + 0.5s\, e^{0.2Z_{k2}})}. \tag{4.20}$$

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 θ.

Table 10: Scenario I: Numerical results for (β_1, β_2) with true value equals $(0.2, 0.4)$. Bias, SEE (Standard Error of Estimates), ESE (Estimated Standard Error), CP summarized for (β_1, β_2). Each entry is based on 1000 simulated datasets under shared random effect model with Multiplicative marginals.

$\mu_j(t)$	ρ	K	bias of $(\hat{\beta}_1, \hat{\beta}_2)$	SEE $(\hat{\beta}_1, \hat{\beta}_2)$	ESE$(\hat{\beta}_1, \hat{\beta}_2)$	CP
$0.25t^2$	1.25	200	(-0.0063, 0.0019)	(0.1679, 0.1569)	(0.1710,0.1600)	(0.9620, 0.9550)
		500	(0.0042, 0.0021)	(0.1080, 0.1037)	(0.1088, 0.1021)	(0.9480, 0.9490)
		800	(0.0032, -0.0046)	(0.0871, 0.0822)	(0.0861, 0.0807)	(0.9480, 0.9420)
	1.5	200	(0.0061, 0.0092)	(0.2139, 0.2072)	(0.2107, 0.2028)	(0.9460, 0.9410)
		500	(-0.0075, 0.0092)	(0.1335, 0.1320)	(0.1341, 0.1286)	(0.9470, 0.9470)
		800	(0.0029, 0.0059)	(0.1100, 0.1050)	(0.1061, 0.1020)	(0.9410, 0.9430)
	1.75	200	(-0.0199, 0.0017)	(0.2457, 0.2394)	(0.2428, 0.2365)	(0.9430, 0.9500)
		500	(-0.0066, 0.0016)	(0.1589, 0.1529)	(0.1545, 0.1507)	(0.9390, 0.9480)
		800	(-0.0066, 0.0011)	(0.1236, 0.1192)	(0.1227, 0.1190)	(0.9500, 0.9560)
$0.375t^2$	1.25	200	(0.0068, -0.0028)	(0.1565, 0.1571)	(0.1571, 0.1506)	(0.9540, 0.9420)
		500	(-0.0039, 0.0042)	(0.0990, 0.0959)	(0.0996, 0.0950)	(0.9450, 0.9480)
		800	(-0.0004, 0.0013)	(0.0816, 0.0751)	(0.0790, 0.0751)	(0.9470, 0.9550)
	1.5	200	(0.0071, 0.0017)	(0.2106, 0.1958)	(0.1987, 0.1936)	(0.9350, 0.9500)
		500	(0.0012, 0.0011)	(0.1237, 0.1316)	(0.1267, 0.1228)	(0.9600, 0.9460)
		800	(0.0023, -0.0018)	(0.1041, 0.0966)	(0.1005, 0.0975)	(0.9450, 0.9600)
	1.75	200	(-0.0046, 0.0038)	(0.2314, 0.2376)	(0.2318, 0.2279)	(0.9510, 0.9320)
		500	(-0.0051, 0.0033)	(0.1499, 0.1514)	(0.1482, 0.1456)	(0.9460, 0.9370)
		800	(-0.0002, -0.0054)	(0.1213, 0.1199)	(0.1174, 0.1156)	(0.9460, 0.9480)
$0.5t^2$	1.25	200	(-0.0028, -0.0030)	(0.1531, 0.1462)	(0.1489, 0.1433)	(0.9460, 0.9480)
		500	(0.0005, 0.0007)	(0.0943, 0.0941)	(0.0945, 0.0912)	(0.9480, 0.9410)
		800	(0.0021, 0.0028)	(0.0748, 0.0710)	(0.0751, 0.0722)	(0.9490, 0.9510)
	1.5	200	(0.0095, -0.0104)	(0.1940, 0.1907)	(0.1929, 0.1884)	(0.9510, 0.9460)
		500	(0.0009, -0.0004)	(0.1239, 0.1244)	(0.1229, 0.1199)	(0.9470, 0.9390)
		800	(-0.0015, 0.0003)	(0.0940, 0.0941)	(0.0975, 0.0953)	(0.9600, 0.9540)
	1.75	200	(-0.0074, 0.0031)	(0.2372, 0.2272)	(0.2270, 0.2247)	(0.9340, 0.9460)
		500	(0.0023, 0.0046)	(0.1473, 0.1486)	(0.1456, 0.1438)	(0.9400, 0.9400)
		800	(0.0063, 0.0013)	(0.1119, 0.1109)	(0.1152, 0.1136)	(0.9550, 0.9560)

Table 11: Scenario I - Estimation of ρ 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

$\mu_{0j}(t)$	ρ	K	Bias	SEE	ESE	CP	$\mu_{0j}(t)$	ρ	K	Bias	SEE	ESE	CP
$0.125t^2$	1.25	200	-0.0041	0.0911	0.0878	0.9210	$0.375t^2$	1.25	200	-0.0010	0.0411	0.0386	0.9180
		500	0.0004	0.0588	0.0572	0.9420			500	-0.0003	0.0257	0.0252	0.9380
		800	-0.0002	0.0444	0.0450	0.9530			800	-0.0002	0.0206	0.0199	0.9400
	1.50	200	-0.0055	0.1273	0.1199	0.9190		1.50	200	-0.0045	0.0742	0.0702	0.9130
		500	-0.0046	0.0816	0.0789	0.9290			500	-0.0031	0.0493	0.0459	0.9170
		800	-0.0028	0.0646	0.0630	0.9380			800	-0.0010	0.0388	0.0370	0.9260
	1.75	200	-0.0082	0.1593	0.1556	0.9170		1.75	200	-0.0094	0.1081	0.1025	0.9050
		500	-0.0007	0.1109	0.1042	0.9260			500	-0.0051	0.0733	0.0697	0.9270
		800	-0.0056	0.0838	0.0825	0.9410			800	-0.0014	0.0572	0.0560	0.9350
$0.25t^2$	1.25	200	-0.0022	0.0443	0.0432	0.9220	$0.5t^2$	1.25	200	-0.0006	0.0370	0.0362	0.9300
		500	-0.0005	0.0285	0.0278	0.9370			500	-0.0004	0.0235	0.0234	0.9480
		800	-0.0009	0.0228	0.0223	0.9440			800	-0.0010	0.0192	0.0186	0.9450
	1.50	200	0.0002	0.0850	0.0759	0.9100		1.50	200	-0.0040	0.0725	0.0673	0.9230
		500	-0.0015	0.0512	0.0492	0.9440			500	-0.0015	0.0454	0.0446	0.9250
		800	-0.0008	0.0382	0.0391	0.9480			800	0.0013	0.0376	0.0362	0.9320
	1.75	200	-0.0082	0.1212	0.1078	0.8970		1.75	200	-0.0057	0.1088	0.1020	0.9170
		500	-0.0063	0.0743	0.0715	0.9230			500	0.0005	0.0715	0.0686	0.9320
		800	-0.0020	0.0613	0.0589	0.9340			800	0.0008	0.0554	0.0556	0.9370

Table 12: Scenario II - Estimation of θ 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.

$\mu_{0j}(t)$	θ	K	Bias	SEE	ESE	CP	$\mu_{0j}(t)$	θ	K	Bias	SEE	ESE	CP
$0.125t^2$	0.25	200	-0.0042	0.1168	0.1140	0.9370	$0.375t^2$	0.25	200	-0.0032	0.0580	0.0579	0.9450
		500	0.0013	0.0745	0.0731	0.9480			500	0.0000	0.0395	0.0375	0.9260
		800	0.0000	0.0605	0.0581	0.9370			800	-0.0010	0.0299	0.0298	0.9570
	0.50	200	-0.0055	0.1527	0.1445	0.9280		0.50	200	-0.0038	0.0921	0.0891	0.9260
		500	0.0012	0.0967	0.0940	0.9450			500	-0.0038	0.0591	0.0587	0.9310
		800	-0.0029	0.0756	0.0745	0.9430			800	-0.0003	0.0494	0.0473	0.9340
	1	200	-0.0006	0.2424	0.2224	0.9050		1	200	-0.0182	0.1821	0.1673	0.8860
		500	-0.0045	0.1482	0.1478	0.9350			500	-0.0051	0.1143	0.1123	0.9330
		800	0.0033	0.1212	0.1185	0.9260			800	-0.0032	0.0922	0.0904	0.9310
$0.25t^2$	0.25	200	-0.0009	0.0713	0.0717	0.9370	$0.5t^2$	0.25	200	0.0004	0.0524	0.0515	0.9380
		500	-0.0012	0.0454	0.0462	0.9530			500	-0.0011	0.0342	0.0331	0.9360
		800	-0.0012	0.0370	0.0368	0.9420			800	0.0004	0.0266	0.0264	0.9410
	0.50	200	-0.0042	0.1075	0.1040	0.9320		0.50	200	-0.0082	0.0890	0.0835	0.9120
		500	-0.0005	0.0683	0.0678	0.9390			500	-0.0001	0.0556	0.0553	0.9380
		800	-0.0027	0.0564	0.0538	0.9390			800	-0.0009	0.0437	0.0438	0.9500
	1	200	-0.0170	0.1893	0.1781	0.8960		1	200	-0.0058	0.1789	0.1624	0.8860
		500	0.0008	0.1201	0.1208	0.9350			500	-0.0037	0.1173	0.1091	0.9240
		800	-0.0068	0.1040	0.0963	0.9210			800	-0.0043	0.0895	0.0863	0.9190

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.

		θ_1				θ_2			
$\mu_{0j}(t)$	K	Bias	SEE	ESE	CP	Bias	SSE	ESE	CP
$0.125t^2$	200	-0.0045	0.0885	0.0843	0.9200	-0.0164	0.2187	0.1880	0.8950
	500	-0.0020	0.0590	0.0557	0.9420	-0.0139	0.1352	0.1266	0.9120
	800	0.0007	0.0447	0.0444	0.9530	-0.0033	0.1069	0.1039	0.9320
	1100	0.0006	0.0378	0.0382	0.9560	0.0007	0.0950	0.0899	0.9400
$0.25t^2$	200	-0.0040	0.0639	0.0608	0.9170	-0.0174	0.1691	0.1527	0.8770
	500	0.0010	0.0440	0.0401	0.9270	-0.0007	0.1149	0.1061	0.9210
	800	-0.0009	0.0326	0.0319	0.9410	-0.0007	0.0934	0.0860	0.9150
	1100	-0.0009	0.0278	0.0278	0.9490	-0.0032	0.0826	0.0737	0.9240
$0.375t^2$	200	-0.0039	0.0563	0.0532	0.9150	-0.0129	0.1622	0.1434	0.8900
	500	-0.0029	0.0364	0.0347	0.9250	-0.0050	0.1048	0.0983	0.9280
	800	0.0002	0.0292	0.0283	0.9310	-0.0022	0.0846	0.0804	0.9310
	1100	0.0011	0.0258	0.0245	0.9380	-0.0019	0.0713	0.0698	0.9390
$0.5t^2$	200	-0.0035	0.0530	0.0499	0.9150	-0.0142	0.1643	0.1399	0.8820
	500	-0.0025	0.0335	0.0326	0.9360	-0.0014	0.1041	0.0959	0.9310
	800	-0.0007	0.0268	0.0263	0.9490	-0.0029	0.0853	0.0784	0.9190
	1100	-0.0003	0.0229	0.0227	0.9470	-0.0012	0.0693	0.0668	0.9250

Table 14: Scenario IV - estimates θ_1, θ_2 in the underline models where $\rho(\theta, s, t|Z_k) = 1 + \theta_1 \frac{(0.25te^{0.1Z_{k1}})(0.25se^{0.2Z_{k2}})}{(0.25+0.25te^{0.1Z_{k1}})(0.25+0.25se^{0.2Z_{k2}})}$ and $\rho(\theta, s, t|Z_k) = 1 + \theta_2 \frac{(0.5te^{0.1Z_{k1}})(0.5se^{0.2Z_{k2}})}{(0.25+0.5te^{0.1Z_{k1}})(0.25+0.5se^{0.2Z_{k2}})}$. With the true values of θ_1, θ_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 θ where each entry is based on 1000 simulations. The averaged observed events for type 1(2) event is 2.44(2.56)

θ_1	N	Bias	SEE	ESE	CP	θ_2	Bias	SEE	ESE	CP
0.25	200	-0.0020	0.0804	0.0798	0.9420	0.25	-0.0036	0.0523	0.0498	0.9280
	500	0.0012	0.0508	0.0518	0.9550		0.0000	0.0335	0.0328	0.9360
	800	0.0009	0.0420	0.0411	0.9490		-0.0002	0.0275	0.0260	0.9390
	1100	-0.0006	0.0345	0.0350	0.9500		0.0010	0.0232	0.0224	0.9490
0.50	200	-0.0012	0.1237	0.1130	0.9170	0.50	-0.0015	0.0908	0.0835	0.9220
	500	0.0004	0.0763	0.0735	0.9260		-0.0059	0.0551	0.0531	0.9170
	800	-0.0007	0.0617	0.0590	0.9400		0.0008	0.0448	0.0436	0.9370
	1100	0.0000	0.0488	0.0503	0.9420		-0.0020	0.0377	0.0370	0.9320
0.75	200	-0.0094	0.1584	0.1471	0.9090	0.75	-0.0097	0.1286	0.1149	0.8800
	500	0.0004	0.1067	0.0993	0.9240		-0.0033	0.0841	0.0793	0.9070
	800	-0.0002	0.0275	0.0260	0.9390		-0.0020	0.0612	0.0628	0.9400
	1100	0.0010	0.0232	0.0224	0.9490		-0.0030	0.0524	0.0539	0.9430
1.00	200	-0.0015	0.0908	0.0835	0.9220	1.00	-0.0035	0.1797	0.1587	0.8730
	500	-0.0059	0.0551	0.0531	0.9170		0.0011	0.1079	0.1056	0.9280
	800	0.0008	0.0448	0.0436	0.9370		-0.0028	0.0833	0.0827	0.9330
	1100	-0.0020	0.0377	0.0370	0.9320		-0.0074	0.0771	0.0719	0.9190

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

$$\begin{aligned}&V^c(s,t,\theta)=\\&N^{-1/2}\sum_{k=1}^{N}\int_0^t\int_0^s w(u,v)\frac{\partial\rho(u,v,\theta)}{\partial\theta}\Big|_{\theta=\theta}\Big\{dN_{k1}(u)dN_{k2}(v)\\&-\rho(u,v,\theta)Y_{k1}(u)d\mu_{01}(u)e^{\beta_1^T Z_{k1}(u)}\cdot Y_{k2}(v)d\mu_{02}(v)e^{\beta_2^T Z_{k2}(v)}\Big\}.\end{aligned} \tag{4.21}$$

Denote $\tilde{V}(s,t,\tilde{\theta})$ the empirical value of $V^c(s,t,\theta)$ as

$$\begin{aligned}&\tilde{V}(s,t,\theta)=\\&N^{-1/2}\sum_{k=1}^{N}\int_0^t\int_0^s \frac{\partial\rho(u,v,\theta)}{\partial\theta}\Big|_{\theta=\tilde{\theta}}\Big\{dN_{k1}(u)dN_{k2}(v)\\&-\rho(u,v,\tilde{\theta})Y_{k1}(u)d\tilde{\mu}_{01}(u)e^{\tilde{\beta}_1^T Z_{k1}(u)}\cdot Y_{k2}(v)d\tilde{\mu}_{02}(v)e^{\tilde{\beta}_2^T Z_{k2}(v)}\Big\}.\end{aligned}$$

and the Supreme Test Statistic as

$$\tilde{T}=\sup_{s,t\in[0,\tau]}\parallel\tilde{V}(s,t,\tilde{\theta})\parallel. \tag{4.22}$$

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

$$\tilde{V}(s,t,\tilde{\theta})=\tilde{V}(s,t,\theta)+N^{-1/2}\frac{\partial\tilde{V}(s,t,\tilde{\theta})}{\partial\theta}N^{1/2}(\tilde{\theta}-\theta)+o_p(1) \tag{4.23}$$

where

$$\tilde{V}(s,t,\theta) = V^c(s,t,\theta) + N^{-1/2}\sum_{k=1}^{N}\Big\{\Upsilon^c_{k1}(s,t) + \Upsilon^c_{k2}(s,t)\Big\} + o_p(1),$$

$$N^{-1/2}\frac{\partial\tilde{V}(s,t,\theta)}{\partial\theta}N^{1/2}(\tilde{\theta}-\theta) = N^{-1/2}\{\zeta^c_{k1}(s,t,\theta) + \zeta^c_{k2}(s,t,\theta)\} + o_p(1). \qquad (4.24)$$

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

$$\tilde{V}^*(s,t)$$
$$= \Big\{\tilde{V}(s,t,\tilde{\theta}) + N^{-1/2}\sum_{k=1}^{N}\hat{\Upsilon}^c_{k1}(s,t,\hat{\theta}) + \hat{\Upsilon}^c_{k2}(s,t,\hat{\theta}) + \hat{\zeta}^c_{k1}(s,t) + \hat{\zeta}^c_{k2}(s,t)\Big\}\cdot G_k \qquad (4.25)$$

By taking the supremum of $\tilde{V}^*(s,t)$ among mesh grids of (s,t), we obtain $\tilde{T}^*$ from the empirical distribution of $\sup_{s,t\in[0,\tau]} \parallel \tilde{V}^*(s,t,\theta) \parallel$. 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 $\tilde{T}^*$ excesses 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 σ^2 and μ^2 represent E(R_k) and 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 $0.25t^2, 0.375t^2, 0.5t^2$ to represent moderately or more frequently observed events. The event count after censoring ranges from 4.18 to 11.30. To accommodate the association strength, we generate R_k from Gamma distribution with E$(R_k) = 1$ and var$(R_k) = 0.25, 0.5, 0.75, 1$ 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

$$E[dN^*_{kj}(t)|R_k(t), Z_{kj}(t)] = R_k(t)\{d\mu_{0j}(t)e^{\beta_j^T Z_{kj}(t)}\}. \tag{4.26}$$

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 $\widetilde{\lambda}_{kj}(t|Z_{kj}(t))$ 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 σ^2 and μ represent the variance and mean of R_k. Let the Poisson process $\widetilde{N}_{kj}(t)$ has rate $\widetilde{\lambda}_{kj}$ =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 θ is σ^2/μ^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 ρ values followed from Proposition 2. The Marginal model is multiplicative.

Settings	PWC1	PWC2	PWC3	PWC4
$R_{k0}: \quad (a_0, b_0)$	(4, 0.25)	(4, 0.25)	(2, 0.5)	(4, 0.25)
$R_{k1}: \ (a_1, b_1)$	(2, 0.5)	(1.33,0.75)	(1, 1)	(1, 1)
$\rho(s < 2, t < 2)$	1.25	1.25	1.5	1.25
$\rho(s > 2, t < 2)$	1	1	1	1
$\rho(s > 2, t > 2)$	1.5	1.75	2	2

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.

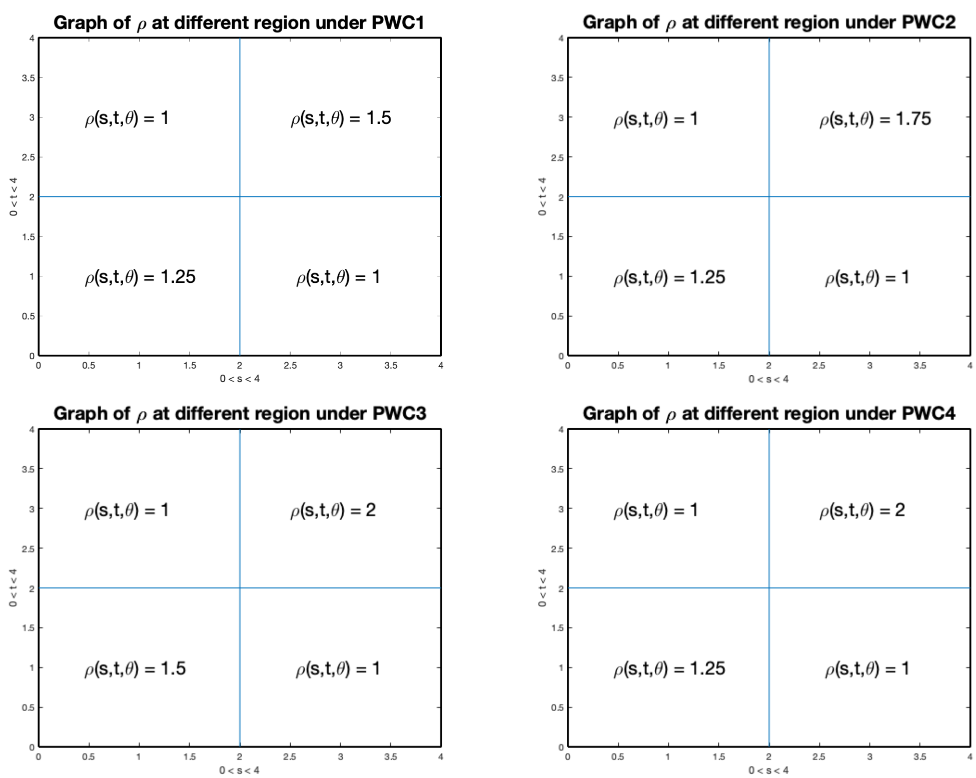
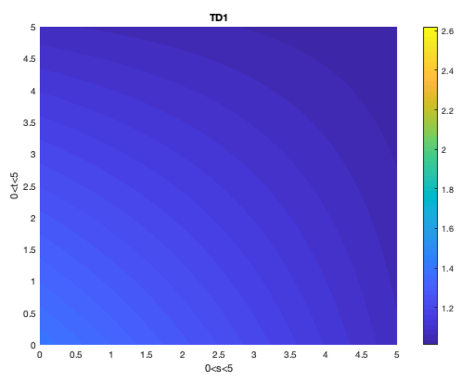
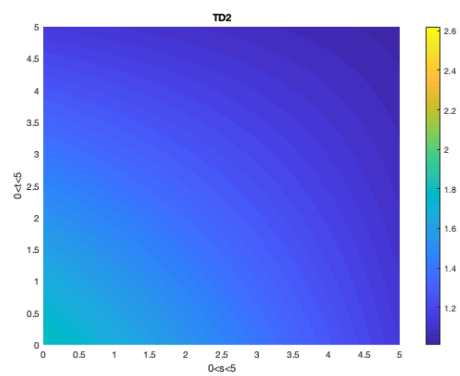


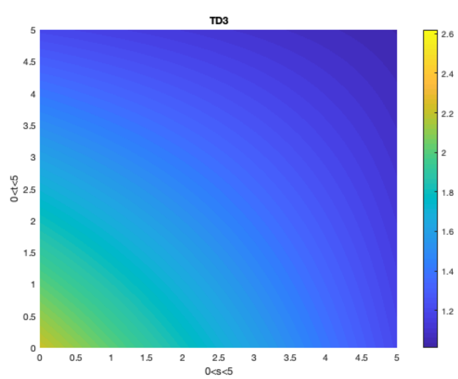
Figure 4: The contour plot of the Rate Ratio $\rho(s,t)$ under the Multiplicative Marginal Models. The x-axis and y-axis represents the observation time for type1 and type2 events. From upper left to lower right, the heterogeneity of $\rho(s,t)$ is increased.



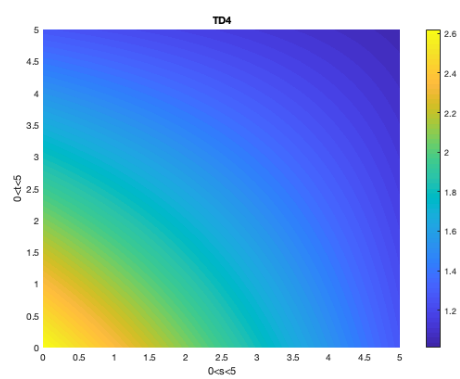
(a) TD1 Model $\rho(s,t)$



(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 $Ha : \rho(\theta, s, t)$ 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.

			Size			
event counts	$\mu_{0j}(t)$	K	$\rho = 1.25$	$\rho = 1.5$	$\rho = 1.75$	$\rho = 2$
(4.18,5.67)	$0.25t^2$	200	0.041	0.025	0.037	0.034
		500	0.038	0.042	0.033	0.033
(6.25, 8.48)	$0.375t^2$	200	0.042	0.035	0.030	0.021
		500	0.040	0.037	0.039	0.037
(8.34, 11.30)	$0.5t^2$	200	0.040	0.037	0.003	0.030
		500	0.043	0.043	0.032	0.037

Table 17: Power of $H_0 : \rho(\theta, s, t) = \theta_0$ vs $Ha : \rho(\theta, s, t)$ 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.

			Power			
event counts	$\mu_{0j}(t)$	K	PWC1	PWC2	PWC3	PWC4
(2.09, 2.83)	$0.125t^2$	200	0.443	0.882	0.777	0.942
		500	0.934	0.999	0.994	0.999
		800	0.995	1.000	1.000	1.000
(4.16, 5.65)	$0.25t^2$	200	0.867	0.977	0.951	0.987
		500	0.998	1.000	1.000	0.999
		800	1.000	1.000	1.000	1.000
(6.25, 8.48)	$0.375t^2$	200	0.955	0.993	0.968	0.991
		500	1.000	1.000	1.000	1.000
		800	1.000	1.000	1.000	1.000
(8.34, 11.32)	$0.5t^2$	200	0.985	0.994	0.986	0.995
		500	1.000	1.000	0.999	1.000
		800	1.000	1.000	1.000	1.000

Table 18: Power of $H_0 : \rho(\theta, s, t) = \theta_0$ vs $Ha : \rho(\theta, s, t)$ 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.

		Power			
$\mu_{0j}(t)$	K	TD1	TD2	TDC3	TD4
$0.125t^2$	200	0.129	0.252	0.308	0.355
	500	0.295	0.560	0.706	0.782
	800	0.463	0.817	0.906	0.932
$0.25t^2$	200	0.238	0.415	0.524	0.556
	500	0.587	0.862	0.929	0.940
	800	0.768	0.974	0.990	0.986
$0.375t^2$	200	0.337	0.518	0.598	0.675
	500	0.748	0.933	0.961	0.947
	800	0.931	0.991	0.994	0.995
$0.5t^2$	200	0.433	0.578	0.691	0.674
	500	0.826	0.949	0.962	0.968

Table 19: Power of $H_0 : \rho(\theta, s, t) = \theta_0$ vs $Ha : \rho(\theta, s, t)$ 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.

		Power			
$\mu_{0j}(t)$	K	TCD1	TCD2	TCD3	TCD4
$0.125t^2$	200	0.102	0.226	0.453	0.706
	500	0.208	0.520	0.916	0.991
$0.25t^2$	200	0.175	0.417	0.704	0.798
	500	0.508	0.923	0.988	0.977
$0.375t^2$	200	0.246	0.480	0.701	0.727
	500	0.650	0.946	0.983	0.963
$0.5t^2$	200	0.210	0.437	0.566	0.631
	500	0.658	0.952	0.972	0.949

REFERENCES

Abu-Libdeh, H., Turnbull, B. W., and Clark, L. C. (1990). Analysis of multi-type recurrent events in longitudinal studies; application to a skin cancer prevention trial. *Biometrics*, 46(4):1017–1034.

Anderson, J. E., Louis, T. A., Holm, N. V., and Harvald, B. (1992). Time-dependent association measures for bivariate survival distributions. *Journal of the American Statistical Association*, 87(419):641–650.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151.

Cook, R. J. and Lawless, J. (2007). *The Statistical Analysis of Recurrent Events*. Springer.

H Scheike, T. (2002). The additive nonparametric and semiparametric aalen model as the rate function for a counting process. *Lifetime data analysis*, 8:247–62.

He, S., Wang, F., and Liuquan, S. (2013). A semiparametric additive rates model for clustered recurrent event data. *Acta Mathematicae Applicatae Sinica*, 29(1):55–62.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics*, 37(2):158–168.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the rate and mean function of recurrent events. *J Royal Statistical Society B*, 62.

Ng, E. T. M. and Cook, R. J. (1999). Robust inference for bivariate point processes. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 27(3):509–524.

Ning, J., Chen, Y., Cai, C., Huang, X., and Wang, M. C. (2015). On the dependence structure of bivariate recurrent event processes: Inference and estimation. *Biometrika*, 102(2):345–358.

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84(406):487–493.

Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association*, 88(423):811–820.

Schaubel, D. E. and Cai, J. (2004). Regression methods for gap time hazard functions of sequentially ordered multivariate failure time data. *Biometrika*, 91(2):291–303.

Schaubel, D. E. and Cai, J. (2005). Analysis of clustered recurrent event data with application to hospitalization rates among renal failure patients. *Biostatistics*, 6(3):404–419.

Scheike, T. H. (2012). On cross-odds ratio for multivariate competing risks data. pages 680–694.

APPENDIX A: PROOFS OF THE PROPOSITIONS IN CHAPTER 3

Proof of Proposition 1

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

$$
\begin{aligned}
&E\{dN_{k1}(s)dN_{k2}(t)|Z_{k1}(s), Z_{k2}(t)\} \\
&= E\Big\{E\{dN_{k1}(s)dN_{k2}(t)|Z_{k1}(s), Z_{k2}(t), R_k\}\Big\} \\
&= E\Big\{E\{dN_{k1}(s)|Z_{k1}(s), R_k\}E\{dN_{k2}(t)|Z_{k2}(t), R_k\}\Big\} \\
&= E\Big\{R_k\{d\mu_{01}(s) + \beta_1^T Z_{k1}(s)\,ds\}R_k\{d\mu_{02}(t) + \beta_2^T Z_{k2}(t)\,dt\}\Big\} \\
&= E\{R_k^2\}\{d\mu_{01}(s) + \beta_1^T Z_{k1}(s)\,ds\}\{d\mu_{02}(t) + \beta_2^T Z_{k2}(t)\,dt\} \qquad \text{(A.1)}
\end{aligned}
$$

and

$$
E\{dN_{k1}(s)\,|\,Z_{k1}(s)\} = E\Big\{E\{dN_{k1}(s)\,|\,Z_{k1}(s), R_k\}\Big\} = E\{R_k\}\{d\mu_{01}(s) + \beta_1^T Z_{k1}(s)\,ds\},
$$

$$
E\{dN_{k2}(t)\,|\,Z_{k2}(t)\} = E\Big\{E\{dN_{k2}(t)\,|\,Z_{k2}(t), R_k\}\Big\} = E\{R_k\}\{d\mu_{02}(t) + \beta_2^T Z_{k2}(t)\,dt\}.
$$

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

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

□

Proof of Proposition 2

Similar to the proof of Proposition 1,

$$
\begin{aligned}
&E\{dN_{k1}(s)dN_{k2}(t)|Z_{k1}(s), Z_{k2}(t)\} \\
&= E\Big\{E\{dN_{k1}(s)|Z_{k1}(s), R_k\}E\{dN_{k2}(t)|Z_{k2}(t), R_k\}\Big\} \\
&= E\{R_k(s)R_k(t)\} \cdot \{d\mu_{01}(s) + \beta_1^T Z_{k1}(s)\, ds\}\{d\mu_{02}(t) + \beta_2^T Z_{k2}(t)\, dt\} \qquad \text{(A.3)}
\end{aligned}
$$

and

$$
\begin{aligned}
E\{dN_{k1}(s)\,|\,Z_{k1}(s)\} &= E\{R_k(s)\}\{d\mu_{01}(s) + \beta_1^T Z_{k1}(s)\, ds\} \\
E\{dN_{k2}(t)\,|\,Z_{k2}(t)\} &= E\{R_k(t)\}\{d\mu_{02}(t) + \beta_2^T Z_{k2}(t)\, dt\}. \qquad \text{(A.4)}
\end{aligned}
$$

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

$$
E\{R_k(s)R_k(t)\} = \begin{cases} E(R_{k0}R_{k0}) = (a_0b_0+\delta_0)^2 + a_0b_0^2 & \text{if } s,t \in (0, c_0] \\ E(R_{k1}R_{k1}) = (a_1b_1+\delta_1)^2 + a_1b_1^2 & \text{if } s,t \in (c_0, \tau] \\ E(R_{k0}R_{k1}) = (a_0b_0+\delta_0)(a_1b_1+\delta_1) & \text{otherwise} \end{cases} \qquad \text{(A.5)}
$$

$$
E\{R_k(s)\}E\{R_k(t)\} = \begin{cases} E(R_{k0})E(R_{k0}) = (a_0b_0+\delta_0)^2 & \text{if } s,t \in (0, c_0] \\ E(R_{k1})E(R_{k1}) = (a_1b_1+\delta_1)^2 & \text{if } s,t \in (c_0, \tau] \\ E(R_{k0})E(R_{k1}) = (a_0b_0+\delta_0)(a_1b_1+\delta_1) & \text{otherwise} \end{cases} \qquad \text{(A.6)}
$$

This yields the piecewise constant rate ratio below :

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

□

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),

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

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

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

$$= \tilde{\lambda}_1(s \,|\, z_1)\, ds \cdot \lambda_0(t \,|\, z_2)\, dt + \tilde{\lambda}_1(s \,|\, z_1)\, ds \cdot \tilde{\lambda}_2(t \,|\, z_2)\, dt \tag{A.9}$$

Similarly,

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

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

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

$$= 1 \cdot \rho_0(\theta, s, t|\, z_1, z_2)\lambda_0(s \,|\, z_1)\, ds \cdot\ \lambda_0(t \,|\, z_2)\, dt + 1 \cdot \tilde{\lambda}_2(t \,|\, z_2)\, dt\, \lambda_0(s \,|\, z_1)\, ds \tag{A.10}$$

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

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

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

$$= \tilde{\lambda}_1(s \,|\, z_1)\, ds \cdot \lambda_0(t \,|\, z_2)\, dt + \tilde{\lambda}_1(s \,|\, z_1)\, ds \cdot \tilde{\lambda}_2(t \,|\, z_2)\, dt$$

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

$$= \{\tilde{\lambda}_1(s \,|\, z_1) + \lambda_0(s \,|\, z_1)\}\{\tilde{\lambda}_0(t \,|\, z_2) + \lambda_0(t \,|\, z_2)\}\, ds\, dt$$

$$+ \{\rho_0(\theta, s, t|\, z_1, z_2) - 1\}\lambda_0(s \,|\, z_1)\lambda_0(t \,|\, z_2)\, ds\, dt$$

$$= \lambda_1(s \,|\, z_1)\lambda_2(t \,|\, z_2)\, ds\, dt + \{\rho_0(\theta, s, t|\, z_1, z_2) - 1\}\lambda_0(s \,|\, z_1)\lambda_0(t \,|\, z_2)\, ds\, dt \tag{A.11}$$

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

$$\rho(\theta, s, t|\,z_1, z_2) = 1 + \{\rho_0(\theta, s, t|\,z_1, z_2) - 1\}\frac{\lambda_0(s|z_1)\lambda_0(t|\,z2)}{\lambda_1(s|\,z_2)\lambda_2(t|\,z_2)}$$

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.

□

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 $k = 1, 2, ..., N$.

C.2. $Pr(C_{kj} > \tau) > 0$, where τ 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

$$L_j(\beta_j) = \sum_{k=1}^N \int_0^\tau \{Z_{kj}(u) - \bar{Z}_j(u)\}\, dM_{kj}(u, \beta_j), \tag{A.14}$$

and with the first order Taylor expansion with respect to β_j gives us

$$(\hat{\beta}_j - \beta_j) = \hat{A}_j^{-1}(\beta^*)N^{-1} \int_0^\tau \{Z_{kj}(u) - \bar{Z}_j(u)\}\, dM_{kj}(u, \beta_j), \tag{A.15}$$

where

$$dM_{kj}(t; \beta_j) = dN_{kj}(t) - Y_{kj}(t)\{d\mu_{0j}(t) + \beta_j^T Z_{kj}(t)\, dt\}$$

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

with β^* a value falls between $\hat{\beta}_j$ and β_j.

By (C.4) and the strong law of large numbers (SLLN), $\hat{\beta}_j$ converges almost surely to β_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

$$\Sigma_j = E[\int_0^\tau \{Z_{1j}(u) - \bar{Z}_j(u)\}dM_{1j}(u, \beta_j) \int_0^\tau \{Z_{1j}(v) - \bar{Z}_j(v)\}\, dM_{1j}(v, \beta_j)].$$

From (A.15) it is straight forward to show

$$\sqrt{N}\{\hat{\beta}_j - \beta_j\} = A_j^{-1} N^{-1/2} \sum_{k=1}^{N} \xi_{kj} + o_p(1). \tag{A.16}$$

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

$$d\hat{M}_{kj}(t; \hat{\beta}_j) = dN_{kj}(t) - Y_{kj}(t)\{d\hat{\mu}_{0j}(t) + \hat{\beta}_j^T Z_{kj}(t)\, dt\},$$
$$\hat{\xi}_{kj} = \int_0^\tau \{Z_{kj}(u) - \bar{Z}_j(u)\}\, d\hat{M}_{kj}(u; \hat{\beta}_j),$$
$$\hat{\Sigma}_j = N^{-1} \sum_{k=1}^{N} \hat{\xi}_{kj}^{\otimes 2}.$$

□

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

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

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

Using the strong convergence of β_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

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

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

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

with

$$\hat{\pi}_j(t) = N^{-1}\sum_{k=1}^N Y_{kj}(t),$$
$$\hat{H}^T(t) = \int_0^t \bar{Z}_j^T(u)\, du.$$

□

Proof of Theorem 3.3

To prove the asymptotic of $\Big\{U(\theta,\hat{\beta}_1,\hat{\mu}_1(\cdot),\hat{\beta}_2,\hat{\mu}_2(\cdot)) - U(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))\Big\}$ 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:

$$\begin{aligned}
&U_k(\theta,\hat{\beta}_1,\hat{\mu}_{01}(\cdot),\hat{\beta}_2,\hat{\mu}_{02}(\cdot))\\
&= U_k(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))\\
&+ \{U_k(\theta,\hat{\beta}_1,\hat{\mu}_{01}(\cdot),\hat{\beta}_2,\hat{\mu}_{02}(\cdot)) - U_k(\theta,\beta_1,\mu_{01}(\cdot),\hat{\beta}_2,\hat{\mu}_{02}(\cdot))\}\\
&+ \{U_k(\theta,\beta_1,\mu_{01}(\cdot),\hat{\beta}_2,\hat{\mu}_{02}(\cdot)) - U_k(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))\} \qquad \text{(A.23)}
\end{aligned}$$

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

$$\int_0^\tau\int_0^\tau -\frac{\partial\rho(s,t,\theta)}{\partial\theta}\rho(s,t,\theta)\cdot$$
$$\Big\{Y_{k2}(t)\{d\hat{\mu}_{02}(t) + \hat{\beta}_2^T Z_{k2}(t)\,dt - d\mu_{02}(t) - \beta_2^T Z_{k2}(t)\,dt\}Y_{k1}(s)\{d\mu_{01}(s) + \beta_1^T Z_{k1}(s)\,ds\}\Big\}, \qquad \text{(A.24)}$$

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

$$
\begin{aligned}
& U_k(\theta, \beta_1, \mu_{01}(s), \hat\beta_2, \hat\mu_{02}(t)) - U_k(\theta, \beta_1, d\mu_{01}(s), \beta_2, d\mu_{02}(t)) \\
&= \int_0^\tau \int_0^\tau -\frac{\partial \rho(s,t,\theta)}{\partial\theta} \rho(s,t,\theta) Y_{k1}(s)\{d\mu_{01}(s) + \beta_1^T Z_{k1}(s)\, ds\} \\
& Y_{k2}(t)\Big\{ Z_{k2}^T(t) dt\, A_2^{-1} N^{-1} \sum_{l=1}^N \xi_{l2} + N^{-1} \sum_{l=1}^N d\phi_{l2}(t;\beta_2)\Big\} + o_p(N^{-1}) \qquad \text{(A.25)}
\end{aligned}
$$

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

$$
\begin{aligned}
& \int_0^\tau \int_0^\tau -\frac{\partial \rho(s,t,\theta)}{\theta} \rho(s,t,\theta) Y_{k1}(s) Y_{k2}(t) [\, d\mu_{02}(t) + \beta_2^T Z_{k2}(t) dt\,] \\
& \Big\{ Z_{k1}^T(s)\, ds\, A_1^{-1} N^{-1} \sum_{l=1}^N \xi_{l1} + N^{-1} \sum_{l=1}^N d\phi_{l1}(s;\beta_1)\Big\} + o_p(N^{-1}) \qquad \text{(A.26)}
\end{aligned}
$$

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

$$
\begin{aligned}
& N^{-1/2}\Big\{U(\theta, \hat\beta_1, \hat\mu_{01}(\cdot), \hat\beta_2, \hat\mu_{02}(\cdot)) - U(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot))\Big\} \\
&= N^{-1/2} \sum_{k=1}^N \Big\{ h_{1,N} \xi_{k1} A_1^{-1} + g_{1,N,k} + h_{2,N} \xi_{k2} A_2^{-1} + g_{2,N,k} \Big\} + o_p(N^{-1/2}) \qquad \text{(A.27)}
\end{aligned}
$$

where the terms are denoted by

$$
\begin{aligned}
q_l(s,t) &= -\frac{\partial \rho(s,t,\theta)}{\partial\theta} \rho(s,t,\theta) Y_{l1}(s) Y_{l2}(t), \\
h_{1,N} &= N^{-1} \sum_{l=1}^N \int_0^\tau \int_0^\tau q_l(s,t)\{d\mu_{02}(t) + \beta_2^T Z_{l2}(t)\, dt\} Z_{l1}^T(s)\, ds, \\
h_{2,N} &= N^{-1} \sum_{l=1}^N \int_0^\tau \int_0^\tau q_l(s,t)\{d\mu_{01}(s) + \beta_1^T Z_{l1}(s)\, ds\} Z_{l2}^T(t)\, dt, \\
g_{1,N,k} &= N^{-1} \sum_{l=1}^N \int_0^\tau \int_0^\tau q_l(s,t)\{d\mu_{02}(t) + \beta_2^T Z_{l2}(t)\, dt\}\, d\phi_{k1}(s;\beta_1), \\
g_{2,N,k} &= N^{-1} \sum_{l=1}^N \int_0^\tau \int_0^\tau q_l(s,t)\{d\mu_{01}(s) + \beta_1^T Z_{l1}(s)\, ds\}\, d\phi_{k2}(t;\beta_2).
\end{aligned}
$$

Deriving from (A.27) the covariance matrix can be estimated by

$$\hat{\Omega} = N^{-1} \sum_{k=1}^{N} \Big\{ \hat{h}_{1,N} \hat{\xi}_{k1} \hat{A}_1^{-1} + \hat{g}_{1,N,k} + \hat{h}_{2,N} \hat{\xi}_{k2} \hat{A}_2^{-1} + \hat{g}_{2,N,k} \Big\}^{\otimes 2}, \tag{A.28}$$

with

$$\begin{aligned}
\hat{q}_l(s,t) &= -\frac{\partial \rho(s,t,\theta)}{\partial \theta} \rho(s,t,\theta) Y_{l1}(s) Y_{l2}(t) \\
\hat{h}_{1,N} &= N^{-1} \sum_{l=1}^{N} \int_0^\tau \int_0^\tau q_l(s,t) \{ d\hat{\mu}_{02}(t) + \hat{\beta}_2^T Z_{l2}(t)\, dt \} Z_{l1}^T(s)\, ds, \\
\hat{h}_{2,N} &= N^{-1} \sum_{l=1}^{N} \int_0^\tau \int_0^\tau q_l(s,t) \{ d\hat{\mu}_{01}(s) + \hat{\beta}_1^T Z_{l1}(s)\, ds \} Z_{l2}^T(t)\, dt, \\
\hat{g}_{1,N,k} &= N^{-1} \sum_{l=1}^{N} \int_0^\tau \int_0^\tau q_l(s,t) \{ d\hat{\mu}_{02}(t) + \hat{\beta}_2^T Z_{l2}(t)\, dt \}\, d\hat{\phi}_{k1}(s; \hat{\beta}_1), \\
\hat{g}_{2,N,k} &= N^{-1} \sum_{l=1}^{N} \int_0^\tau \int_0^\tau q_l(s,t) \{ d\hat{\mu}_{01}(s) + \hat{\beta}_1^T Z_{l1}(s)\, ds \}\, d\hat{\phi}_{k2}(t; \hat{\beta}_2).
\end{aligned} \tag{A.29}$$

□

Proof of Theorem 3.4

Denote

$$\begin{aligned}
& W_k(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) \\
&= U_k(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) + \Big\{ h_{1,N} \xi_{k1} A_1^{-1} + g_{1,N,k} + h_{2,N} \xi_{k2} A_2^{-1} + g_{2,N,k} \Big\},
\end{aligned} \tag{A.30}$$

which follows from equation (A.27) and let

$$\mathcal{I}(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot)) = -N^{-1} \sum_{k=1}^{N} \Big(\frac{\partial U_k(\theta, \beta_1, \mu_{01}(\cdot), \beta_2, \mu_{02}(\cdot))}{\partial \theta} \Big)^T. \tag{A.31}$$

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

$$\sqrt{N}(\hat{\theta}-\theta)$$
$$= N^{-1/2}\{\mathcal{I}(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))\}^{-1}\sum_{k=1}^{N} W_k(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot)) + o_p(1). \tag{A.32}$$

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 $\hat{\Phi} = N^{-1}(\hat{\mathcal{I}})^{-1}\sum_{k=1}^{N}(\hat{W}_k)^{\otimes 2}\{(\hat{\mathcal{I}}^T)\}^{-1}$, with

$$\hat{\mathcal{I}} = \mathcal{I}\big(\hat{\theta},\hat{\beta}_1,\hat{\mu}_{01}(\cdot),\hat{\beta}_2,\hat{\mu}_{02}(\cdot)\big),$$
$$\hat{W}_k = W_k(\hat{\theta},\hat{\beta}_1,\hat{\mu}_{01}(\cdot),\hat{\beta}_2,\hat{\mu}_{02}(\cdot)),$$

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

□

APPENDIX C: PROOFS OF THE MODEL CHECKING PROCEDURE IN CHAPTER 3

Recall (3.26)

$$\begin{aligned} V(s,t,\hat{\theta},\hat{\mu}_1(\cdot;Z_{k1}),\hat{\mu}_2(\cdot;Z_{k2})) =& V\Big(s,t,\theta,\hat{\mu}_1(\cdot;Z_{k1}),\hat{\mu}_2(\cdot;Z_{k2})\Big) \\ &+ N^{-1/2}\frac{\partial V\Big(s,t,\theta,\hat{\mu}_1(\cdot;Z_{k1}),\hat{\mu}_2(\cdot;Z_{k2})\Big)}{\partial\theta}N^{1/2}(\hat{\theta}-\theta) \\ &+ o_p(1), \end{aligned}$$

Note that $V\Big(s, t, \theta, \hat{\mu}_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\Big)$ can be further decomposed by

$$
\begin{aligned}
& V\Big(s, t, \theta, \hat{\mu}_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\Big) \\
& = V\Big(s, t, \theta, \mu_1(\cdot; Z_{k1}), \mu_2(\cdot; Z_{k2})\Big) \\
& \quad + V\Big(s, t, \theta, \hat{\mu}_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\Big) - V\Big(s, t, \theta, \mu_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\Big) \\
& \quad + V\Big(s, t, \theta, \mu_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\Big) - V\Big(s, t, \theta, \mu_1(\cdot; Z_{k1}), \mu_2(\cdot; Z_{k2})\Big). \qquad \text{(A.33)}
\end{aligned}
$$

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

$$
\begin{aligned}
& \sqrt{N}\Big\{V\Big(s, t, \theta, \hat{\mu}_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\Big) - V\Big(s, t, \theta, \mu_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\Big)\Big\} \\
& = \sum_{k=1}^{N} \int_0^t \int_0^s \frac{\partial \rho(u, v, \theta; Z_{k1}, Z_{k2})}{\theta} \rho(u, v, \theta; Z_{k1}, Z_{k2}) \\
& \quad Y_{k1}(u) Y_{k2}(v) [\, d\mu_{02}(v) + \beta_2^T Z_{k2}(v) dv\,] \Big\{ Z_{k1}^T(s)\, ds\, A_1^{-1} N^{-1} \sum_{l=1}^{N} \xi_{l1} + N^{-1} \sum_{l=1}^{N} d\phi_{l1} \Big\} \\
& \quad + o_p(N^{-1}), \qquad \text{(A.34)}
\end{aligned}
$$

and

$$
\begin{aligned}
& \sqrt{N}\Big\{V\Big(s, t, \theta, \mu_1(\cdot; Z_{k1}), \hat{\mu}_2(\cdot; Z_{k2})\Big) - V\Big(s, t, \theta, \mu_1(\cdot; Z_{k1}), \mu_2(\cdot; Z_{k2})\Big)\Big\} \\
& = \sum_{k=1}^{N} \int_0^t \int_0^s \frac{\partial \rho(u, v, \theta; Z_{k1}, Z_{k2})}{\partial \theta} \rho(u, v, \theta; Z_{k1}, Z_{k2}) \\
& \quad Y_{k2}(v) Y_{k1}(u) \{\mu_{01}(u) + \beta_1^T Z_{k1}(u)\, du\,\} \Big\{ Z_{k2}^T(v)\, dv\, A_2^{-1} N^{-1} \sum_{l=1}^{N} \xi_{l2} + N^{-1} \sum_{l=1}^{N} d\phi_{l2} \Big\} \\
& \quad + o_p(N^{-1}). \qquad \text{(A.35)}
\end{aligned}
$$

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

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

where

$$q_l(u,v)=-\frac{\partial\rho(u,v,\theta)}{\partial\theta}\rho(u,v,\theta)Y_{l1}(u)Y_{l2}(v)$$
$$h_{2,N}(s,t)=N^{-1}\sum_{l=1}^{N}\int_0^t\int_0^s w(u,v)q_l(u,v)\{d\mu_{01}(u)+\beta_1^TZ_{l1}(u)\,du\}Z_{l2}^T(u)\,dv$$
$$g_{2,N,k}(s,t)=N^{-1}\sum_{l=1}^{N}\int_0^t\int_0^s w(u,v)q_l(u,v)\{d\mu_{01}(u)+\beta_1^TZ_{l1}(u)\,du\}\,d\phi_{k2}(v)$$
$$h_{1,N}(s,t)=N^{-1}\sum_{l=1}^{N}\int_0^t\int_0^s w(u,v)q_l(u,v)\{d\mu_{02}(v)+\beta_2^TZ_{l2}(v)\,dv\}Z_{l1}^T(u)\,du$$
$$g_{1,N,k}(s,t)=N^{-1}\sum_{l=1}^{N}\int_0^t\int_0^s w(u,v)q_l(u,v)\{d\mu_{02}(v)+\beta_2^TZ_{l2}(v)\,dv\}\,d\phi_{k1}(u) \tag{A.37}$$

To simplify the notation, we define

$$\varUpsilon_{k1}(s,t,\theta)=N^{-1}\Big\{h_{1,N}(s,t)\xi_{k1}A_1^{-1}+g_{1,N,k}(s,t)\Big\}+o_p(N^{-1})$$
$$\varUpsilon_{k2}(s,t,\theta)=N^{-1}\Big\{h_{2,N}(s,t)\xi_{k2}A_2^{-1}+g_{2,N,k}(s,t)\Big\}+o_p(N^{-1}) \tag{A.38}$$

so that (A.36) can be rewritten as

$$V\Big(s,t,\theta,\hat{\mu}_1(\cdot;Z_{k1}),\hat{\mu}_2(\cdot;Z_{k2})\Big)-V\Big(s,t,\theta,\mu_{01}(\cdot),\mu_{02}(\cdot)\Big)$$
$$=N^{-1/2}\sum_{k=1}^{N}\{\varUpsilon_{k1}(s,t,\theta)+\varUpsilon_{k2}(s,t,\theta)\}+o_p(1). \tag{A.39}$$

Following the empirical approximation of $\sqrt{N}(\hat{\theta}-\theta)$ in equation (3.10),

$$N^{-1/2}\frac{\partial V(s,t,\theta,\hat{\mu}_1(\cdot;Z_{k1}),\hat{\mu}_2(\cdot;Z_{k2}))}{\partial\theta}N^{1/2}(\hat{\theta}-\theta)$$
$$=\Psi_\theta(s,t)\Big\{-N^{-1/2}\{\mathcal{I}(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))\}^{-1}\sum_{k=1}^{N}W_k(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))\Big\}$$
$$+o_p(1), \tag{A.40}$$

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

$$N^{-1/2}\frac{\partial V\Big(s,t,\theta,\hat{\beta}_1,\hat{\mu}_{01}(\cdot),\hat{\beta}_2,\hat{\mu}_{02}(\cdot)\Big)}{\partial\theta}N^{1/2}(\hat{\theta}-\theta)=N^{-1/2}\{\zeta_{k1}(s,t,\theta)+\zeta_{k2}(s,t,\theta)\} \tag{A.41}$$

where

$$\zeta_{k1}(s,t,\theta)=-\Psi_\theta(s,t)\{\mathcal{I}(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))\}^{-1}N^{-1}\Big\{h_{1,N}\xi_{k1}A_1^{-1}+g_{1,N,k}\Big\}$$
$$\zeta_{k2}(s,t,\theta)=-\Psi_\theta(s,t)\{\mathcal{I}(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))\}^{-1}N^{-1}\Big\{h_{2,N}\xi_{k2}A_2^{-1}+g_{2,N,k}\Big\} \tag{A.42}$$

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

$$V\Big(s,t,\hat{\theta},\hat{\mu}_1(\cdot;Z_{k1}),\hat{\mu}_2(\cdot;Z_{k2}))$$
$$=V(s,t,\theta,\mu_1(\cdot;Z_{k1}),\mu_2(\cdot;Z_{k2}))$$
$$+N^{-1/2}\sum_{k=1}^{N}\Big\{\Upsilon_{k1}(s,t,\theta)+\Upsilon_{k2}(s,t,\theta)+\zeta_{k1}(s,t,\theta)+\zeta_{k2}(s,t,\theta)\Big\}+o_p(1).$$

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

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

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

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

(C*.5) $A_j^c \equiv E\Big[\int_0^\tau \{Z_{kj}(u) - \bar{z}_j(\beta_j, u)\}^{\otimes 2} Y_{kj}(u) e^{\beta_j^T Z_{kj}(u)}\, d\mu_{0j}(u)\Big]$ 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 β_j is $L_j(\beta_j, \tau)$, where

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

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

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

$$0 \le s, t \le \tau,$$

between time points s and t.

By Taylor series expansion,

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

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

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

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

where

$$\xi_{kj}^c(u; \beta_j) = \int_0^\tau \{Z_{kj}(u) - \tilde{z}_j(u; \beta_j)\}\, dM_{kj}^c(u; \beta_j).$$

The consistency estimators of A_j and Σ_j are denoted by

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

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

with

$$\tilde{\xi}_{kj} = N^{-1} \sum_{k=1}^{N} \int_0^{\tau} \{Z_{kj}(u) - \tilde{Z}_{kj}(u, \tilde{\beta}_j)\} \, d\tilde{M}_{kj}(u; \tilde{\beta}_j),$$

$$\tilde{M}_{kj}(t; \tilde{\beta}_j) = N_{kj}(t) - \int_0^t Y_{kj}(u) e^{\tilde{\beta}_j^T Z_{kj}(u)} \, d\tilde{\mu}_{0j}(u). \tag{A.45}$$

□

Proof of Theorem 4.2

Let $\tilde{\mu}_{0j}(t) \equiv \tilde{\mu}_{0j}(t, \tilde{\beta}_j) = \int_0^t \frac{d\bar{N}_j(u)}{NS_j^0(u, \tilde{\beta}_j)}$ and decompose $\tilde{\mu}_{0j}(t)$ as

$$\tilde{\mu}_{0j}(t) - \mu_{0j}(t) = \big\{\tilde{\mu}_{0j}(t, \tilde{\beta}_j) - \tilde{\mu}_{0j}(t, \beta_j)\big\} + \big\{\tilde{\mu}_{0j}(t, \beta_j) - \mu_{0j}(t)\big\}. \tag{A.46}$$

The uniform strong law of large numbers Pollard (1990) implies $S_j^0(\beta_j, t) \to s_j^0(\beta_j, t)$ and $\bar{N}_j(t)/N \to E[N_j(t)]$ uniformly in t and β_j, and hence the uniform convergence of $\tilde{\mu}_{0j}(t, \beta_j) = \int_0^t \frac{d\bar{N}_j(u)}{NS_j^0(u, \beta_j)}$ to $\mu_{0j}(t) = \int_0^t \frac{s_j^0(u, \beta_j)}{s_j^0(u, \beta_j)} d\mu_{0j}(u)$. Furthermore, we can represent the second term in (A.46) as

$$\begin{aligned} \tilde{\mu}_{0j}(t, \beta_j) - \mu_{0j}(t) &= \int_0^t \frac{d\bar{N}_j(u)}{NS_j^0(u, \beta_j)} - d\mu_{0j}(u) \\ &= N^{-1} \int_0^t \frac{\sum_{k=1}^N dM_{kj}^c(u; \beta_j)}{S_j^0(u, \beta_j)}, \\ &= N^{-1} \int_0^t \frac{\sum_{k=1}^N dM_{kj}^c(u; \beta_j)}{s_j^0(u, \beta_j)} + o_p(N^{-1}). \end{aligned} \tag{A.47}$$

The first term in (A.46) can be rewritten as

$$\begin{aligned}\tilde{\mu}_{0j}(t,\tilde{\beta}_j)-\tilde{\mu}_{0j}(t,\beta_j) &= \int_0^t \frac{d\bar{N}_j(u)}{NS_j^0(u,\tilde{\beta}_j)}-\frac{d\bar{N}_j(u)}{NS_j^0(u,\beta_j)},\\ &= \int_0^t -\tilde{Z}_j^T(u,\beta_j)\frac{d\bar{N}_j(u)}{NS_j^0(u,\beta_j)}(\tilde{\beta}_j-\beta_j)+o_p(N^{-1}),\\ &= -\int_0^t \tilde{z}_j(u,\beta_j)\,d\mu_{0j}(t,\beta_j)(\tilde{\beta}_j-\beta_j)+o_p(N^{-1}).\end{aligned}$$

The asymptotic approximation of $\{\tilde{\beta}_j-\beta_j\}$ in (A.44) entails

$$\begin{aligned}&\tilde{\mu}_{0j}(t,\tilde{\beta}_j)-\tilde{\mu}_{0j}(t,\beta_j)\\ &=[H^c(t;\beta_j)]^T(A_j^c)^{-1}N^{-1}\sum_{k=1}^N\int_0^\tau\{Z_{kj}(u)-\tilde{z}_j(u,\beta_j)\}\,dM_{kj}^c(u;\beta_j)+o_p(N^{-1}),\end{aligned} \tag{A.48}$$

with $H^c(t;\beta_j)=\int_0^t \tilde{z}_j(u,\beta_j)\,d\mu_{0j}(u,\beta_j)$. Plugging (A.48), (A.47) into equation (A.46) and multiplying both sides by $\sqrt{N}$ yield

$$\sqrt{N}\Big\{\tilde{\mu}_{0j}(t)-\mu_{0j}(t)\Big\}=N^{-1/2}\sum_{k=1}^N\phi_{kj}^c(t;\beta_j)+o_p(1), \tag{A.49}$$

where

$$\begin{aligned}&\phi_{kj}^c(t;\beta_j)\\ &=\int_0^t\frac{dM_{kj}^c(u;\beta_j)}{s_j^0(u,\beta_j)}-[H^c(t;\beta_j)]^T(A_j^c)^{-1}\sum_{k=1}^N\int_0^\tau\{Z_{kj}(u)-\tilde{z}_j(u,\beta_j)\}\,dM_{kj}^c(u;\beta_j).\end{aligned} \tag{A.50}$$

Since $\phi_{kj}(t)$ is independent mean-zero normal random variable, $\sqrt{N}\Big\{\hat{\mu}_{0j}(t)-\mu_{0j}(t)\Big\}$

converges to a zero-mean Gaussian process with covariance function at (s,t) as

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

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

$$\tilde{\phi}_{kj}(t) = \int_0^t \frac{d\tilde{M}_{kj}(u)}{S_j^0(\tilde{\beta}_j,u)} - [\tilde{H}(t)]^T\tilde{A}_j^{-1}\sum_{k=1}^{N}\int_0^\tau \{Z_{kj}(u) - \tilde{Z}_{kj}(u,\tilde{\beta}_j)\}\, d\tilde{M}_{kj}(u;\tilde{\beta}_j),$$

and

$$\tilde{H}(t) = \int_0^t \tilde{Z}_j(u,\tilde{\beta}_j)\, d\tilde{\mu}_{0j}(t,\tilde{\beta}_j). \tag{A.52}$$

□

Proof of Theorem 4.3

Considering the decomposition:

$$\begin{aligned}
&U_k^c(\theta,\tilde{\beta}_1,\tilde{\mu}_{01}(\cdot),\tilde{\beta}_2,\tilde{\mu}_{02}(\cdot)) - U_k(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot)) \\
&= \Big\{U_k(\theta,\tilde{\beta}_1,\tilde{\mu}_{01}(s),\tilde{\beta}_2,\tilde{\mu}_{02}(\cdot)) - U_k(\theta,\beta_1,\mu_{01}(\cdot),\tilde{\beta}_2,d\tilde{\mu}_{02}(t))\Big\} \\
&\quad + \Big\{U_k(\theta,\beta_1,\mu_{01}(\cdot),\tilde{\beta}_2,\tilde{\mu}_{02}(\cdot)) - U_k(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))\Big\}
\end{aligned} \tag{A.53}$$

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

$$\begin{aligned}
&\int_0^\tau\int_0^\tau -\rho(\theta,s,t)\frac{\partial\rho(\theta,s,t)}{\partial\theta}Y_{k2}(t)e^{\tilde{\beta}_2^T Z_{k2}(t)}\, d\tilde{\mu}_{02}(t) \\
&\Big\{Y_{k1}(s)e^{\tilde{\beta}_1^T Z_{k1}(s)}\, d\tilde{\mu}_{01}(s) - Y_{k1}(s)e^{\beta_1^T Z_{k1}(s)}\, d\mu_{01}(s)\Big\}
\end{aligned} \tag{A.54}$$

In (A.54), $Y_{k1}(s)e^{\tilde{\beta}_1^T Z_{k1}(s)}\, d\tilde{\mu}_{01}(s) - Y_{k1}(s)e^{\beta_1^T Z_{k1}(s)}\, d\mu_{01}(s)$ can be further rewritten as

$$\begin{aligned}
&Y_{k1}(s)\Big\{e^{\tilde{\beta}_1^T Z_{k1}(s)}\, d\tilde{\mu}_{01}(s) - e^{\beta_1^T Z_{k1}(s)}\, d\tilde{\mu}_{01}(s) + e^{\beta_1^T Z_{k1}(s)}\, d\tilde{\mu}_{01}(s) - e^{\beta_1^T Z_{k1}(s)}\, d\mu_{01}(s)\Big\} \\
&= Y_{k1}(s)\Big\{e^{\beta_1^T Z_{k1}(s)} Z_{k1}^T(s)\, d\tilde{\mu}_{01}(s)(\hat{\beta}_1 - \beta_1) + e^{\beta_1^T Z_{k1}}(d\tilde{\mu}_{01}(s) - d\mu_{01}(s))\Big\} \\
&\quad + o_p(\tilde{\beta}_1 - \beta_1)^{\otimes 2}
\end{aligned}$$

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

$$\begin{aligned}
&Y_{k1}(s)e^{\tilde{\beta}_1^T Z_{k1}(s)}\, d\tilde{\mu}_{01}(s) - Y_{k1}(s)e^{\beta_1^T Z_{k1}(s)}\, d\mu_{01}(s) \\
&= Y_{k1}(s)\Big\{e^{\beta_1^T Z_{k1}(s)} Z_{k1}^T(s)\, d\mu_{01}(s)(A_1^c)^{-1} N^{-1} \sum_{l=1}^{N} \xi_{l1}^c + e^{\beta_1^T Z_{k1}(s)} N^{-1} \sum_{l=1}^{N} d\phi_{l1}^c(s)\Big\} \\
&\quad + o_p(N^{-1}).
\end{aligned} \tag{A.55}$$

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

$$\begin{aligned}
&\Big\{U_k^c(\theta, \tilde{\beta}_1, \tilde{\mu}_{01}(s), \tilde{\beta}_2, \tilde{\mu}_{02}(\cdot)) - U_k^c(\theta, \beta_1, \mu_{01}(\cdot), \tilde{\beta}_2, d\tilde{\mu}_{02}(t))\Big\} \\
&= \int_0^\tau \int_0^\tau -\rho(\theta, s, t)\frac{\partial \rho(\theta, s, t)}{\partial \theta} Y_{k1}(s)e^{\beta_1^T Z_{k1}(s)} \cdot Y_{k2}(t)e^{\beta_2^T Z_{k2}(t)} \\
&\quad \cdot N^{-1} \sum_{l=1}^{N} \Big\{Z_{k1}^T(s)\, d\mu_{01}(s)\, d\mu_{02}(t) A_1^{-1} \xi_{l1}^c + d\phi_{l1}^c(s)\, d\mu_{02}(t)\Big\} + o_p(1)
\end{aligned} \tag{A.56}$$

In a similar fashion,

$$\begin{aligned}
&Y_{k2}(t)e^{\tilde{\beta}_2^T Z_{k2}(t)}\, d\tilde{\mu}_{02}(t) - Y_{k2}(t)e^{\beta_2^T Z_{k2}(t)}\, d\mu_{02}(t) \\
&= Y_{k2}(t)\{e^{\beta_2^T Z_{k2}(t)} Z_{k2}^T(t)\, d\mu_{02}(t)(A_2^c)^{-1} N^{-1} \sum_{l=1}^{N} \xi_{l2}^c + e^{\beta_2^T Z_{k2}(t)} N^{-1} \sum_{l=1}^{N} d\phi_{l2}^c(t)\} \\
&\quad + o_p(N^{-1}).
\end{aligned} \tag{A.57}$$

Since the $Y_{k2}(t)e^{\tilde{\beta}_2^T Z_{k2}(t)}\, d\tilde{\mu}_{02}(t)$ and $Y_{k1}(s)e^{\tilde{\beta}_1^T Z_{k1}(s)}\, d\tilde{\mu}_{01}(s)$ only have $o_p(N^{-1})$ difference 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

$$\Big\{U_k^c(\theta,\beta_1,d\mu_{01}(s),\tilde{\beta}_2,d\tilde{\mu}_{02}(t)) - U_k^c(\theta,\beta_1,d\mu_{01}(s),\beta_2,d\mu_{02}(t))\Big\}$$
$$= \int_0^\tau\int_0^\tau -\rho(\theta,s,t)\frac{\partial\rho(\theta,s,t)}{\partial\theta}Y_{k1}(s)e^{\beta_1^T Z_{k1}(s)}\cdot Y_{k2}(t)e^{\beta_2^T Z_{k2}(t)}$$
$$\cdot N^{-1}\sum_{l=1}^{N}\Big\{Z_{k2}^T(t)\,d\mu_{01}(s)\,d\mu_{02}(t)(A_2^c)^{-1}\xi_{l2}^c + e^{\beta_2^T Z_{l2}(t)}\,d\mu_{01}(s)\,d\phi_{l2}^c(t)\Big\} + o_p(1) \tag{A.58}$$

Since

$$U^c(\theta,\tilde{\beta}_1,\tilde{\mu}_{01}(\cdot),\tilde{\beta}_2,\tilde{\mu}_{02}(\cdot)) - U^c(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))$$
$$= \sum_{k=1}^{N}\Big\{U_k^c(\theta,\tilde{\beta}_1,\tilde{\mu}_{01}(\cdot),\tilde{\beta}_2,\tilde{\mu}_{02}(\cdot)) - U_k^c(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))\Big\}$$

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

$$N^{-1/2}\{U^c(\theta,\tilde{\beta}_1,\tilde{\mu}_{01}(s),\tilde{\beta}_2,\tilde{\mu}_{02}(t)) - U^c(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))\}$$
$$= N^{-1/2}\sum_{k=1}^{N}\Big\{h_{1,N}^c(A_1^c)^{-1}\xi_{k1}^c + g_{1,N}^c + h_{2,N}(A_2^c)^{-1}\xi_{k2}^c + g_{2,N}^c\Big\} + o_p(1). \tag{A.59}$$

where

$$
\begin{aligned}
q_l^c(\theta,s,t) &= -\rho(\theta,s,t)\frac{\partial\rho(\theta,s,t)}{\partial\theta}Y_{l1}(s)e^{\beta_1^T Z_{l1}(s)}Y_{l2}(t)e^{\beta_2^T Z_{l2}(t)},\\
h_{1,N}^c &= N^{-1}\sum_{l=1}^N \int_0^\tau\int_0^\tau q_l^c(\theta,s,t)Z_{l1}^T(s)\,d\mu_{01}(s)d\mu_{02}(t),\\
g_{1,N}^c &= N^{-1}\sum_{l=1}^N \int_0^\tau\int_0^\tau q_l^c(\theta,s,t)\,d\mu_{02}(t)\,d\phi_{k1}^c(s),\\
h_{2,N}^c &= N^{-1}\sum_{l=1}^N \int_0^\tau\int_0^\tau q_l^c(\theta,s,t)Z_{l2}^T(t)\,d\mu_{02}(t)d\mu_{01}(s),\\
g_{2,N}^c &= N^{-1}\sum_{l=1}^N \int_0^\tau\int_0^\tau q_l^c(\theta,s,t)d\mu_{01}(s)\,d\phi_{k2}^c(t).
\end{aligned}
\tag{A.60}
$$

□

Proof of Theorem 4.4

By the first order Taylor expansion of the estimation equation,

$$
\begin{aligned}
&\sqrt{N}(\tilde\theta-\theta)\\
&= \{-N^{-1}\frac{\partial U(\theta,\beta_1,d\mu_{01}(\cdot),\beta_2,d\mu_{02}(\cdot))}{\partial\theta}\}^{-1}N^{-1/2}U(\theta,\hat\beta_1,\hat\mu_{01}(\cdot),\hat\beta_2,\hat\mu_{02}(\cdot))+o_p(N^{-1/2})
\end{aligned}
\tag{A.61}
$$

Denote

$$
\mathcal{I}^c(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot)) = -N^{-1}\sum_{k=1}^N\Big(\frac{\partial U_k(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))}{\partial\theta}\Big)^T \tag{A.62}
$$

and applying (A.59) and (A.62), (A.61) can be rewritten as

$$\sqrt{N}(\tilde{\theta}-\theta)$$
$$= N^{-1/2}\{\mathcal{I}^c(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))\}^{-1}\sum_{k=1}^{N} W_k^c(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot)) + o_p(1), \tag{A.63}$$

where

$$W_k^c(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot))$$
$$= \Big\{U_k^c(\theta,\beta_1,\mu_{01}(\cdot),\beta_2,\mu_{02}(\cdot)) + h_{1,N}^c(A_1^c)^{-1}\xi_{k1}^c + g_{1,N}^c + h_{2,N}^c(A_2^c)^{-1}\xi_{k2}^c + g_{2,N}^c\Big\}. \tag{A.64}$$

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 $\tilde{\Phi} = N^{-1}\tilde{\mathcal{I}}^{-1}(\sum_{k=1}^{N}\tilde{W}_k^{\otimes 2})(\tilde{\mathcal{I}}^T)^{-1}$, where $\tilde{\mathcal{I}}$ and $\tilde{W}_k$ 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 $\tilde{\theta},\tilde{\beta}_1,\tilde{\mu}_{01}(\cdot),\tilde{\beta}_1,\tilde{\mu}_{02}(\cdot)$.