Spring 2018

Cox-type model validation with recurrent event data

Muna Mohamed Hammuda

COX-TYPE MODEL VALIDATION WITH RECURRENT EVENT DATA

by

MUNA MOHAMED HAMMUDA

A DISSERTATION

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

MATHEMATICS WITH STATISTICS EMPHASIS

2018

Approved by

Dr. Akim Adekpedjou , Advisor
Dr. V. A. Samaranayake
Dr. Xuerong (Meggie) Wen
Dr. John Singler
Dr. Ivan Guardiola

ABSTRACT

Recurrent event data occurs in many disciplines such as actuarial science, biomedical studies, sociology, and environment to name a few. It is therefore important to develop models that describe the dynamic evolution of the event occurrences. One major problem of interest to researchers with these types of data is models for the distribution function of the time between events occurrences, especially in the presence of covariates that play a major role in having a better understanding of time to events.

This work pertains to statistical inference of the regression parameter and the baseline hazard function in a Cox-type model for recurrent events that accounts for the effective age and time varying covariates. Estimators of the regression parameters as well as baseline hazard function are obtained using the counting processes and martingales machinery techniques. Asymptotic properties of the proposed estimators and how they can be used to construct confidence intervals are investigated. The results of the simulation studies assessing the performance of the estimators and an application to a biomedical dataset illustrating the models are presented. The impact of unit effective age is also assessed.

To check the validity of the models used, many decision rules are developed for checking the validity of the various components of Cox-type model. Specifically, using martingales residuals, we proposed test statistics for checking the link function and the covariates functional form. Asymptotic properties of test statistics and simulation studies are presented as well.

Key words: Recurrent events, Effective age, Martingale residuals, Goodness of fit, Cox-model.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Dr. Akim Adekpedjou, for his guidance, encouragement, and endless patience throughout the development of this dissertation. It was his inspiration and enthusiasm that guaranteed the accomplishment of this work. I would also like to extend my gratitude to the members of my advisory committee members: Dr. V.A. Samaranayake, Dr. Xuerong (Meggie) Wen, Dr. John Singler, and Dr. Ivan Guardiola.

I would like to thank the Mathematics and Statistics Department faculty members, staff, and my fellow graduate students for making my graduate life at Missouri S&T memorable.

Finally, my deepest gratitude goes to my family, especially my parents, for their endless love, support and encouragement, confidence in my decisions, and patience as I strove to pave my own way in a foreign country.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure Page

LIST OF TABLES

1. INTRODUCTION

The processes which generate events repeatedly over time are referred to as recurrent events and the data they provide are called recurrent event data. In some settings, data may lie in a large number of processes generating a relatively small number of recurrent events. These types of processes are often seen in health and biomedical studies, where information is available on many individuals, each of whom may experience clinical events repeatedly over time. Examples in medical field includes occurrences of heart attacks, sickness leave from work, recurrence of cancer tumors, epileptic seizures in neurology studies, and deteriorating episodes of visual acuity. In actuarial science, examples include the filing of vehicle warranty claims, and property insurance claims for policy holders. In other settings the recurrent event may be available for a relatively small number of processes occurring with a large number of recurrent events. Examples include analyzing processes for software fault detection and removal, cracks in highways, and investigating the injuries incidence in manufacturing plants.

1.1. THE DATA STRUCTURE

Consider a recurrent event data with n independent subjects monitored for the occurrence of a recurrent event over a random time interval $[0,\tau_i]$, where $\tau_i, i = 1, 2, ..., n$ are independent and identically distributed (i.i.d) having distribution function $G(t) = P\{\tau_i \leq t\}$. For each i^{th} subject, let $\{S_{ij}, j = 1, 2, ...\}$ be the calender times of event occurrence, where $0 \equiv S_{i0} < S_{i1} < S_{i2} < ...$, and $\{T_{ij} = S_{ij} - S_{ij-1}, j = 1, 2, ...\}$ is the successive interoccurrence times, or gap times with a common absolutely continuous distribution function $F(t) = P\{T_{ij} \leq t\}$. The renewal function associated with $F(\cdot)$ is $\rho(t) = \sum_{k=1}^{\infty} F^{*k}(t)$, where $F^{*k}(\cdot)$ is the k^{th} convolution of $F(\cdot)$ with itself. A renewal process is the situation in which a subject is completely restored to a similar state after each event and is defined as a process in which the gap times T_{ij} between consecutive events are (i.i.d). The $T_{ij}s$ are to be assumed independent of τ_i. The censoring random variable τ_i is noninformative about the $S_{ij}s$. If $K_i = \max\{k \in \{0, 1, ...\} : S_{ik} \leq \tau_i\}$ is the total number of occurrences for unit i, then the observable data for n units is n independent copies of $\mathbf{D}(s) = \{\mathbf{D}_1(s), \mathbf{D}_2(s), ..., \mathbf{D}_n(s)\}$ where for $i = 1, 2, ..., n$

$$\mathbf{D}_i(\tau_i) = \{(\mathbf{X}_i(s) : 0 \leq s \leq \tau_i), K_i, \tau_i, T_{i1}, T_{i2}, ..., T_{iK_i}, C_{i,K_i+1}\}, \tag{1.1}$$

where $S_{iK_i} = \sum_{j=1}^{k} T_{ij_i}$ and $C_{i,K_i+1} = \tau_i - S_{iK_i}$ is the right-censoring time variable for $T_{i\{K_i+1\}}$. Note that since the right-censoring time variable is $C_{i,K_i+1} = \tau_i - \sum_{j=1}^{k} T_{ij}$ then this censoring variable is dependently functionally of the $T_{ij}s$, even though the τ_i and the $T_{ij}s$ are independent. Furthermore, note that K_i is informative about the distribution of the interoccurrence times. Thus, the data accrual scheme leads to dependent and informative censoring. The process $\mathbf{X}_i(s)$ is a q-dimensional time dependent covariates vector recorded every time an event occurs. For example, in studies on the frequency of visits to hospital emergency clinics because of breathing problems, air pollution measures, temperature, and humidity may be important covariates. In actuarial science, covariates can be age, driving history, Zip code, etc.

An intervention is often performed after each event occurrence, such as replacing or repairing failed components in a reliability system, or reducing or increasing physical activity after a heart attack in a medical settings. These interventions will typically impact the next occurrence of the event. Also, the inter-ocurrence times may be affected by an unobserved random variable, Z, called a frailty. Frailty are unobservable random variables that could inject an association between the inter-event times, Immune system and driver aggressiveness are typical frailty. These frailties could make some subjects have more recurrences than others as a result. A pictorial representation of recurrent events is given in Figure 1.1.

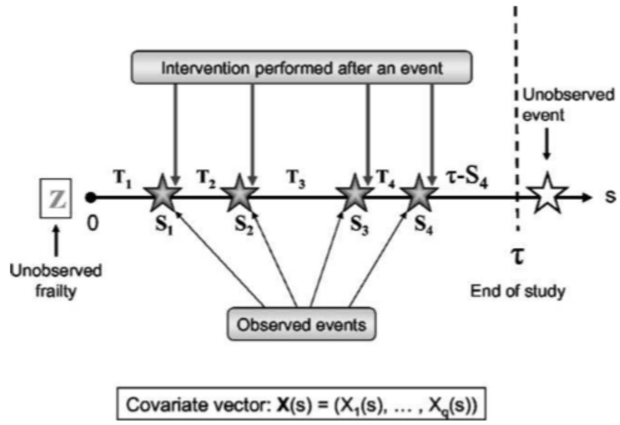


Figure 1.1. Pictorial representation of recurrent events

With respect to the interventions, in reliability systems for instance, interventions can be perfect repair (i.e. the system returns to the as-good-as-new state), minimal repair (i.e. the system in the same condition as it was just before the failure, so called as bad as old), or imperfect repair (i.e. the system returns to an intermediate state between as bad as old and as good as new). In the case of a renewal process, no intervention is performed.

Let the process $N(s)$ denote the number of events that occurred on or before calender time s and the baseline hazard function has the process $\mathcal{E}(\cdot)$ as its argument, called the effective age process. This process models the impact of performed interventions after each event occurrence. The perfect repair is modelled by $\mathcal{E}(s) = s - S_{N(s-)}$ which means that the age of the component at time s equals $s - S_{N(s-)}$, the time elapsed since the last event. Minimal repair is modelled by $\mathcal{E}(s) = s$, which means that the age at any time s equals the calendar time s. Imperfect repair can be modelled by $\mathcal{E}(s) = \Gamma_{N(s-)} + s - S_{N(s-)}$ where $0 \leq \Gamma_i \leq S_i$ is some measure of the effective age of the component immediately after the i^{th} event. In the BP model, Γ_i is defined indirectly by letting a failed component be given perfect repair with probability p, and minimal repair with probability 1-p.

1.2. MODELS FOR RECURRENT EVENT DATA

There are various ways for modeling recurrent events. Some of them are outlined here.

1.2.1. General Intensity-Based Models. One way to model recurrent events data is via the intensity function. Intensity-based models, including modulated Poisson and renewal processes are discussed by various authors in the field. The Poisson process is the simplest example of a model with global time, where time is measured from an initializing event. For the Poisson process, future occurrences of the event are not influenced by past occurrences under the renewal assumption. For the renewal process, the probability of an event only depends on the time elapsed since the last event. In the point processes literature, extensive probabilistic developments and many examples of applications have been discussed by Cox and Lewis [1], Cox and Isham [2], Daley and Vere-Jones [3, 4], Lewis [5], and Snyder and Miller [6]. Fleming and Harrington [7], and Andersen et al. [8] emphasized the dynamic route using counting processes and martingale framework. Aalen et al. [9] provide interesting discussion on intensity-based versus random effects modeling. Fosen et al. [10, 11] gives an interesting perspective on internal covariates and the use of path analysis methods to assess internal and external covariate effects. Cox [12] introduced semiparametric modulated renewal models, and Aalen and Husebye [13], Follmann and Goldberg [14], and Dabrowska et al. [15] emphasize renewal models. Anderson and Gill [16] in an intensity based approach proposed asymptotic properties of the parameters in Cox model with time varying covariates for single events whereas Peña et al. [17] and Peña et al. [18] consider the same model and provide small sample properties in the recurrent event context. Lawless [19], and Thall [20] consider semiparametric and parametric methods for regression Poisson models. More general multiplicative intensity-based models are considered by Gail et al. [21] and Prentice et al. [22].

1.2.2. Marginal Models. In some situations, it makes sense to regard the variation between individuals as a nuisance and to use a marginal approach, where one focuses on the effect of the fixed covariates averaged over the variation between individuals. In fact, a marginal model ignores the dependence on the past in a process. But ignoring the past entails some technical complications, and one will miss the opportunity to understand the details of the underlying process. Marginal models has been discussed by Lawless and Nadeau [23], Cook and Lawless [24], Lin et al. [25], Scheike [26], Chiang et al. [27], and Martinussen and Scheike [28].

1.2.3. Dynamic Recurrent Event Models. Dynamic modeling of recurrent events is an approach that outlines the time evolution of the recurrent events. The class of models in this case incorporates an effective age function encoding the impact of interventions after each event occurrence, the impact of accumulating event occurrences, the induced informative and dependent right censoring mechanism due to the data-accrual scheme, and the effect of covariate processes. The class was proposed by Peña and Hollander [29] and subsumes as special cases many of the recurrent event models that have been considered in biostatistics, reliability, and in the social sciences. Inferential based on this class were rigorously developed by Peña, Strawderman, and Hollander with recurrent events, Paul and Kvam in reliability, and Peña, Slate, and González on intensity based. Stocker and Peña [30] discussed the class of models under a fully parametric specification. Adekpedjou and Stocker [31] proposed Cox-model with more general effective age process that subsumes the one proposed here.

1.3. AIMS OF DISSERTATION

In this dissertation, we follow the intensity based modeling approach. Namely, let $\mathbf{X}$ be q-dimensional vector of covariates, $\boldsymbol{\beta}$ is regression parameter vector, $\lambda_0(s)$ is unknown baseline hazard function, and $R_i(s)$ is backward recurrence time. Consider

$$\lambda_i(s) = \lambda_0(R_i(s))\exp(\boldsymbol{\beta}'\mathbf{X}_i(s)), \quad s \geq 0.$$

This is a spacial case of the general class of models proposed by Peña and Hollander [29]

$$\lambda(s|\mathbf{X}) = \lambda_0[\mathcal{E}(s)]\rho[N(s-)]\psi[\boldsymbol{\beta}^t\mathbf{X}(s)],$$

with $\mathcal{E}(s) = s - S_{N_(s-)}$, $\rho(s) = 1$, and $\psi(\cdot) = \exp(\cdot)$.
Our main aim is to develop methods using the counting process and martingale machines for:

1. Deriving estimators of the hazard $\Lambda_0(s)$ and $\boldsymbol{\beta}$.

2. Obtain their asymptotic properties.

3. Develop decision rules for checking the goodness of the underlying model.

4. Perform a simulation studies to assess our methods.

5. Apply the method to a reliability dataset.

2. MATHEMATICAL PRELIMINARIES

Since the pioneering work of Aalen [32] counting processes and martingales have become the critical tools for analyzing failure time data. In this chapter, we discuss the mathematical prerequisites needed for analyzing these types of data. An excellent overview is given in Anderson et al. [33] and Fleming and Harrington [7]. Assume that $\mathscr{F}$ is a $\sigma - algebra$, and $\mathscr{P}$ is a probability measure on Ω. Let $\mathscr{T} \subset \mathbb{R}^+$. $\mathscr{T}$ is usually taken to be the interval $[0,\tau)$ or $[0,\tau\ ]$, where $\tau = \infty$ is allowed.

Definition 2.1. A stochastic process is a time-indexed family of random variables $X = \{X(t) : t \in \mathscr{T}\}$ on (Ω, $\mathscr{F}$, P).

Definition 2.2. A stochastic process X is

- Integrable if $\sup_{t\in\mathscr{T}} E|X(t)| < \infty$,
- Square integrable if $\sup_{t\in\mathscr{T}} EX(t)^2 < \infty$,
- Bounded if there exists a finite constant Γ such that

$$\mathrm{P}\ \{\sup_{t\in\mathscr{T}} |X(t)| < \Gamma\} = 1.$$

Definition 2.3. A filtration $\mathscr{F} = \{\mathscr{F}_t : t \in \mathscr{T}\}$ on (Ω, $\mathscr{F}$, P), is an increasing right-continuous family of $sub - \sigma - algebra$ of $\mathscr{F}$. That is, $\mathscr{F}_s \subseteq \mathscr{F}_t \subseteq \mathscr{F}$, for $s \leq t$.

Definition 2.4. A stochastic process $X = \{X(t) : t \geq 0\}$ is adapted to a filtration $\mathscr{F}_t$ if, $\forall t \geq 0$, X(t) is $\mathscr{F}_t - measurable.$

A stochastic process X is always adapted to its natural filtration $\mathscr{F}_t = \sigma\{X(s) : s \leq t\}$, the smallest $\sigma - algebra$ with respect to which all the variables $\{X(s) : s \leq t\}$ are measurable.

Definition 2.5. A stochastic process X is called cadlag if its sample paths $\{X(t,w) : t \in \mathscr{T}\}$, for almost all w, are right continuous with left hand limits.

Definition 2.6. A collection $M = \{M(t) : t \geq 0\}$ is an $\mathscr{F}$-martingale if M is an $\mathscr{F}$-cadlag adapted process and satisfies

$E|M(t)| < \infty$ for all $t \in \mathscr{T}$ (integrablility);

$E(M(t)|\mathscr{F}_s) = M(s)$ a.s. for all s < t (martingale property).

The process is submartingale if (2) is replaced by

$$E(M(t)|\mathscr{F}_s) \geq M(s) \text{ a.s. for all } s < t.$$

The process is supermartingale if (2) is replaced by

$$E(M(t)|\mathscr{F}_s) \leq M(s) \text{ a.s. for all } s < t.$$

Definition 2.7. Let $\mathscr{F}$ be a filtration on (Ω, $\mathscr{F}$, P). The $\sigma-$ algebra on $[0,\infty) \times \Omega$ generated by all sets of the form:

1. $[0] \times A$, $A \in \mathscr{F}_0$,
2. $(a,b] \times A$, $0 \leq a < b < \infty$, $A \in \mathscr{F}_a$

is called the predictable $\sigma-$ algebra for $\mathscr{F}$, where $\mathscr{F}_0$ is the information at time 0.

Definition 2.8. A stochastic process X is called predictable if, as a function of $(t,w) \in \mathscr{T} \times \Omega$, it is measurable with respect to the $\sigma-$ algebra on $\mathscr{T} \times \Omega$ generated by the left-continuous adapted processes.

Lemma 2.1. *Let $\mathscr{F}$ be a filtration, and X a left-continuous real-valued process adapted to $\mathscr{F}$. Then X is predictable.*

Proposition 2.2. *Let X be a $\mathscr{F}_t$-predictable process. Then X(t) is $\mathscr{F}_t$- measurable, for any $t > 0$.*

Theorem 2.3 (Doob-Meyer Decomposition)**.** *Let $\mathscr{M}$ be a right-continuous nonnegative submartingale with respect to $\mathscr{F}$, then there exists a right-continuous martingale M and an increasing right-continuous predictable process A such that $E\{A(t)\} < \infty$ and*

$$\mathscr{M}(t) = M(t) - A(t) \ a.s.$$

is a right-continuous $\mathscr{F}$-martingale.

Corollary 2.4. *Let $\{N(t) : t \geq 0\}$ be a counting process F-adapted and right-continuous with $E\{N(t) < \infty\}$ for any t. Then, there exists a unique increasing right-continuous $\mathscr{F}$-predictable process A such that A(0) = 0 a.s., $E\{N(t)\} < \infty$ for any t, and*

$$M(t) = N(t) - A(t), \, t \geq 0$$

is a right-continuous $\mathscr{F}$-martingale.

If M is a martingale with $E\{M^2(t)\} < \infty$ for any $t \geq 0$, Jensen's inequality indicates that $M^2(t)$ is a submartingale. Therefore, the square of a local square integrable martingale is a local submartingale and furthermore has a nondecreasing compensator.

Corollary 2.5. *Let M be a $\mathscr{F}$- cadlag martingale, and $E\{M^2(t)\} < \infty$. Then, there exists a unique increasing right-continuous predictable process $\langle M, M \rangle$, called the predictable variation process of M, such that $\langle M, M \rangle(0) = 0$, a.s., $E\langle M, M \rangle(t) < \infty$ for each t, and $\{M^2(t) - \langle M, M \rangle(t) : t \geq 0\}$ is a right-continuous martingale.*

Stochastic integrals with respect to the observations path are needed to solve the problems of inference for continuous time stochastic processes. The forming of the integral of one stochastic process with respect to another is considered. This will be a pathwise operation. For given $w \in \Omega$, one forms an ordinary Lebesgue-Stieltjes integral over the interval [0,t], see Chung [34]. The next theorem establishes that $\mathscr{L} = \int_0^t H(u)dM(u)$ is martingale for all t where H is a bounded predicatable process. $\mathscr{L}$ exists as a Lebesgue-Stieltjes integral for all paths of H and M.

Theorem 2.6. *Let N be a counting process, and* $\mathscr{F} = \{\mathscr{F}_t : t \geq 0\}$ *be a right-continuous filtration such that*

- M = *N-A is an* $\mathscr{F}$*-martingale, where* $A = \{A(t) : t \geq 0\}$ *is an increasing* $\mathscr{F}$*-predictable process with A(0) = 0;*
- *H is a bounded* $\mathscr{F}$*-predictable process.*

Then the process $\mathscr{L}$ *given by*

$$\mathscr{L} = \int_0^t H(u)dM(u)$$

is an $\mathscr{F}$*-martingale.*

Theorem 2.7. *Suppose M is a finite variation local square integrable martingale, H is a predictable process, [M] is optional variation process, and* $\int H^2 \mathrm{d}[M]$ *is locally integrable or* $\int H^2 \mathrm{d}\langle M \rangle$ *is locally finite. Then* $\int HdM$ *is a local square integrable martingale, and it's predictable and variation processes of stochastic integrals are given by*

$$[\textstyle\int HdM](t) = \int H^2 d[M],$$

$$\langle \textstyle\int HdM \rangle(t) = \int H^2 d\langle M \rangle.$$

Likelihood representations for general counting process models are first introduced by Jacod [35] [36]. The likelihood function can be written by using a product-integral notation $\prod$, which is a continuous version of the simple product Π. The martingale central limit theorem has been used for proving asymptotic properties for counting estimators, that arise in the models for failure time data.

Theorem 2.8 (Rebolledo's Martingale Central Limit Theorem). *Let $\mathscr{T}_0 \subseteq \mathscr{T}$ and consider the conditions*

$$\langle \boldsymbol{M}^{(n)} \rangle(t) \xrightarrow{P} \boldsymbol{V}(t) \qquad \text{for all } t \in \mathscr{T} \text{ as } n \to \infty,$$

$$[\boldsymbol{M}^{(n)}](t) \xrightarrow{P} \boldsymbol{V}(t) \qquad \text{for all } t \in \mathscr{T} \text{ as } n \to \infty,$$

$$\langle \boldsymbol{M}^{(n)}_{\varepsilon,h} \rangle(t) \xrightarrow{P} 0 \qquad \text{for all } t \in \mathscr{T}, h \text{ and } \varepsilon > 0 \text{ as } n \to \infty.$$

Then, as $n \to \infty$

$$(\boldsymbol{M}^{(n)}(t_1), ..., \boldsymbol{M}^{(n)}(t_k)) \xrightarrow{\mathscr{D}} (\boldsymbol{M}^{(\infty)}(t_1), ..., \boldsymbol{M}^{(\infty)}(t_k))(t_1, ..., t_k) \in \mathscr{T}_0$$

Furthermore, if $\mathscr{T}_0$ is dense in $\mathscr{T}$ and contains τ; $\tau \in \mathscr{T}$, then the same conditions imply

$$\boldsymbol{M}^{(n)} \xrightarrow{\mathscr{D}} \boldsymbol{M}^{(\infty)} \qquad \text{in } (D(\mathscr{T}))^k \text{ as } n \to \infty,$$

and $\langle \boldsymbol{M}^{(n)} \rangle$ converges uniformly on compact subsets of $\mathscr{T}$, in probability, to $\boldsymbol{V}$.

The next theorem pertains to the asymptotic properties of martingale transform that usually arise in the modeling and analysis of failure time data. Let $\mathscr{T}$ be a compact subset of $\mathbb{R}$. For $i = 1, ..., n$ and $(s,t) \in \mathscr{T}^2$, let $\mathbf{H}_i(s,t)$ be p-dimensional vector-valued processes on $(\Omega, \mathscr{F}, \mathscr{P})$ with $H_{ij}(\cdot,\cdot)$ bounded, and for each $t \in \mathscr{T}$, the process $\{\mathbf{H}_i(s,t) : s \in \mathscr{T}\}$ is $\mathscr{F}$-predictable. Let

$$\mathbf{W}(s,t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t \mathbf{H}_i(s,w) M_i(s,dw), \qquad (s,t) \in \mathscr{T}^2$$

be the integral-transformed processes which arise in recurrent event and renewal process models and introduce for later use

$$\mathbf{W}^{(n)}(s,t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t \mathbf{H}_i^{(n)}(s,w)M_i^{(n)}(s,dw);$$
$$\mathbf{V}^{(n)}(s,t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t [\mathbf{H}_i^{(n)}(s,w)]^{\otimes 2}Y_i^{(n)}(s,w)\lambda(w)dw.$$

Theorem 2.9 (Peña et al. [37])**.** *Fix an $s \in \mathscr{T}$. Suppose the following conditions are satisfied for $t, t_1, t_2 \in [0, t^*]$ where $t^* \in \mathscr{T}$:*

1. *The processes $\{\boldsymbol{H}_i(v,w) : 0 \le v \le s; 0 \le w \le t^*\}$ are left-continuous in (v, w), and there exists a deterministic function $\boldsymbol{h}(v,w)$ on $[0,s]\times[0,t^*]$ which is continuous in (v, w) and bounded such that*

$$\max_{1\le i\le n}\sup_{0\le w\le t^*}|\boldsymbol{H}_i(s,w) - \boldsymbol{h}(s,w)| \xrightarrow{p} 0;$$

2. *For each $s \in \mathscr{T}$, $\inf_{w\in[0,t^*]}y(s,w) > 0$ where $y(s,w) = E[Y_1(s,w)]$;*

3. *The matrix function*

$$\boldsymbol{\Sigma}(s,t) = \int_0^t \boldsymbol{h}(s,w)^{\otimes 2}y(s,w)\lambda(w)dw$$

is such that for each $t_1, t_2 \in [0,t^*]$ with $t_1 < t_2, 0 < det\{\boldsymbol{\Sigma}(s,t_2) - \boldsymbol{\Sigma}(s,t_1)\} < \infty$; and for each t, as $n \to \infty$, $\|\boldsymbol{V}^{(n)}(s,t) - \boldsymbol{\Sigma}(s,t)\| \xrightarrow{p} 0$.

Then, as $n \to \infty$, the sequence of processes $\{\boldsymbol{W}^{(n)}(s,t) : t \in [0,t^*], n = 1,2,...\}$ converges weakly on the Skorohod space $\mathscr{D}[0,t^*]$ to the Gaussian process $\{\boldsymbol{W}^{(\infty)}(s,t) : t \in [0,t^*]\}$ with zero mean function and covariance matrix function given by

$$\boldsymbol{Cov}\{\boldsymbol{W}^{(\infty)}(s,t_1), \boldsymbol{W}^{(\infty)}(s,t_2)\} = \begin{bmatrix}\boldsymbol{\Sigma}(s,t_1) & \boldsymbol{\Sigma}(s,t_1)\\ \boldsymbol{\Sigma}(s,t_1) & \boldsymbol{\Sigma}(s,t_2)\end{bmatrix}, \quad for \quad t_1 \le t_2.$$

The continuous mapping theorem and Slutsky theorem will be used in the sequel.

Theorem 2.10 (The Continuous Mapping Theorem). *Let $h : \mathbb{D} \mapsto \mathbb{E}$ be continuous at all points in $\mathbb{D}_0 \subset \mathbb{D}$, where $\mathbb{D}$ and $\mathbb{E}$ are metric spaces. Suppose that $P_n \Rightarrow P$ in $\mathbb{D}$ and $P(P \in \mathbb{D}_0) = 1$. Then $h(P_n) \Rightarrow h(P)$ in $\mathbb{E}$.*

Theorem 2.11 (Slutsky Theorem). *Let $X_n \Rightarrow X$ and $Y_n \xrightarrow{p} c$ constant as $n \to \infty$. Then*

1. *$X_n Y_n \Rightarrow cX$, and*
2. *$X_n Y_n \Rightarrow X + c$.*

Anderson and Gill [16] state that pointwise convergence of random concave functions implies uniform convergence on compact subspaces.

Theorem 2.12. *Let E be an open convex subset of $\mathbb{R}^p$ and let $F_1, F_2, ...,$ be a sequence of random concave functions on E such that $\forall x \in E$, $F_n(x) \xrightarrow{p} f(x)$ as $n \to \infty$ where f is some real function on E. Then f is also concave and all compact $A \subset E$,*

$$\sup_t |F_n(x) - f(x)| \xrightarrow{p} 0 \quad as\, n \to \infty.$$

Corollary 2.13 (Anderson and Gill [16]). *Suppose f(x) has a unique maximum at $\hat{x} \in E$. Let $\hat{X}_n$ maximizes $F_n(x)$. Then $\hat{X}_n \xrightarrow{p} \hat{x} \quad as\, n \to \infty.$*

3. MODELING AND ANALYSIS OF RECURRENT EVENTS

Denote by $(\Omega, \mathscr{F}, \mathscr{P})$ the common probability space on which all random entities are defined. We consider a study with n independent subjects are each under observation over a random time interval [0,τ_i], where τ_i are i.i.d. right-censoring random variables with distribution function G(t) = P$\{\tau_i \leq$ t$\}$. For the i^{th} subject, let $\{S_{ij}, j = 1, 2, ...\}$ be the calender times of event occurrence, where $0 \equiv S_{i0} < S_{i1} < S_{i2} < ...$, and $T_{ij} = S_{ij} - S_{ij-1}$ be i.i.d nonnegative random variables representing the successive interoccurrence times, or gap times, of the recurrent event of interest, with a common absolutely continuous distribution function F(t) = P$\{T_{ij} \leq t\}$. The renewal function associated with F is $\rho(t) = \sum_{k=1}^{\infty} F^{*k}(t)$, where F^{*k} is the k^{th} convolution of F with itself. We assume that T_{ij} and τ_i are independent. For the i^{th} subject, the random variable $K_i = \max\{k \in \{0, 1, ...\} : S_{ik} \leq \tau_i\}$ is the number of event occurrences observed over [0,τ_i] and $\{\mathbf{X}_i(s) : 0 \leq s \leq \tau_i\}$ is q-dimensional covariate process. Let $C_{i,K_i+1} = \tau_i - S_{iK_i}$ is the right-censoring time variable for T_{i,K_i+1}. The observable entities for the i^{th} subject are

$$\mathbf{D}_i(\tau_i) = \{(\mathbf{X}_i(s) : 0 \leq s \leq \tau_i), K_i, \tau_i, T_{i1}, T_{i2}, ..., T_{iK_i}, C_{i,K_i+1}\}, \tag{3.1}$$

Based on the data in (3.1), we define the calender time processes. For the i^{th} subject, the process $N_i(s)$ represents the number of events that occurs on or before calender time s; while $Y_i(s)$ represents the at-risk process, with

$$N_i(s) = \sum_{j=1}^{\infty} I\{S_{ij} \leq (s \wedge \tau_i) : s \geq 0\},$$

$$Y_i(s) = I\{\tau_i \geq s : s \geq 0\}.$$

Suppose $\mathscr{F} = \{\mathscr{F}_s : s \geq 0\}$ is the natural filtration generated by $\{(N_i(s), Y_i(s), \mathbf{X}_i(s)) : s \geq 0, i = 0, 1, .., n\}$. Hence $\mathscr{F}$ is the smallest $\sigma-$field for which $\mathbf{X}_i(s)$ and $N_i(s)$ are adapted cadlag processes. Moreover, because $\mathbf{X}_i(s)$ and $Y_i(s)$ are predictable, the filtration $\mathscr{F}_s$ is defined by

$$\mathscr{F}_s = \mathscr{F}_0 \vee \sigma\{(N_i(v), Y_i(v+), \mathbf{X}_i(v)) : s \geq v, i = 0, 1, .., n\},$$

where $\mathscr{F}_0$ represents the σ-field containing all information and events supposed to be fixed at time 0 and $\mathscr{F}_s$ represents the σ-field containing all information and events which have occurred up to and including time s, is right-continuous.

Let the observables for n subjects be $\mathbf{D}(s^*) = (\mathbf{D}_1(s^*), ..., \mathbf{D}_n(s^*))$, where $\mathbf{D}_i, i = 1, ..., n$ are independent copies of $\mathbf{D}_i$ and $s^* \geq \max_{1 \leq i \leq n} \tau_i$. Define the backward recurrence time process for each i, which is the elapsed time since the last event occurrence, by $R_i(s) = s - S_{iN_i(s-)}$. The process $R_i = \{R_i(s) : s \geq 0\}$ is $\mathcal{F}$-adapted and left-continuous, and hence $\mathcal{F}$-predictable. We shall consider Cox-type model. Using a counting process formulation, our model is that $\{N_i(s), i = 1, 2, ..., n\}$ are univariate counting process, having intensity process with respect to $\mathscr{F}_s$

$$\lambda_i(s) = \lambda_0(R_i(s))\exp(\boldsymbol{\beta}^{'}\mathbf{X}_i(s)), \quad s \geq 0, i = 1, ..., n$$

where $\lambda_0(\cdot)$ is the baseline hazard function whose argument is backward recurrence time process and $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_q)^{'}$ is a q-dimensional vector of unknown regression coefficients. Note that since the backward recurrence time process is the perfect repair effective age process, then the proposed model becomes a special case of that in Adekpedjou and Stocker [31]. Stocker and Peña [30] developed a general class of parametric models for recurrent event data.

From stochastic integration theory, the compensator process of $N_i(s)$ is $A_i = \{A_i(s;\boldsymbol{\beta}) : s \geq 0\}$, with

$$A_i(s;\boldsymbol{\beta}) = \int_0^s Y_i(v)\lambda_0(R_i(v))\exp(\boldsymbol{\beta}'\mathbf{X}_i(v))dv.$$

The martingale process with respect to the natural filtration $\mathcal{F}$ is $M_i = \{M_i(s;\boldsymbol{\beta}) : s \geq 0\}$ with $M_i(s;\boldsymbol{\beta}) = N_i(s) - A_i(s;\boldsymbol{\beta})$ being a square integrable martingale with respect to the filtration $\mathcal{F}_s$, with the predictable quadratic covariation process given by

$$\langle M_i, M_i\rangle(s) = A_i(s).$$

where the backward recurrence time process $R_i(s) = s - S_{N_i(s-)}$. So that the compensator of the counting process $N_i(s)$ is

$$A_i(s;\boldsymbol{\beta}) = \int_0^s Y_i(v)\lambda_0(R_i(v))\exp(\boldsymbol{\beta}'\mathbf{X}_i(v))dv.$$

Observe that $R_i(s)$ is random in $A_i(s;\boldsymbol{\beta})$, then direct applications of calender time counting process can not be applied. Instead, the techniques accounting simultaneously for calender time and gap time will be employed to set of the random argument. Extending an idea of Sellke [38] and Gill [39], Peña et al.[40] introduced a doubly-indexed process

$$Z_i(s,t) = I\{R_i(s) \leq t\}, \quad i = 1, 2, ..., n. \tag{3.2}$$

This process indicates whether, at calender time s, at most t time units have elapsed since the last event occurrence. To extend the development of the doubly-indexed process, we form the partition $0 = s_0 < s_1 < ... < s_{k_i} < s_{k_{i+1}} = \tau_i$, and $\left(S_{ij-1}, S_{ij}\right] \subseteq$

$\Big(s_{j-1}, s_j\Big]$, for each $i, j = 1, 2, .., k_{i+1}$. Let

$$Q_i(s,t) = \sum_{j=1}^{k_{i+1}} I\{s - S_{ij-1} \le t\} = I\{R_{ij}(s) \le t\}, \quad i = 1, 2, ..., n, \tag{3.3}$$

The process Q_i is used to obtain the doubly-indexed process via integral transformation. For each i,

$$\begin{aligned} N_i(s,\mathbf{t}) &= \int_0^s Q_i(v,t)dN_i(v), \\ A_i(s,t;\boldsymbol{\beta}) &= \int_0^s Z_i(v,t)A_i(dv;\boldsymbol{\beta}), \\ M_i(s,t;\boldsymbol{\beta}) &= \int_0^s Z_i(v,t)M_i(dv;\boldsymbol{\beta}) = N_i(s,t) - A_i(s,t;\boldsymbol{\beta}). \end{aligned} \tag{3.4}$$

For a fixed t, the process $M_i(.,t;\boldsymbol{\beta})$ is a zero-mean square integrable martingale. However, for a fixed s, the process $M_i(s,.;\boldsymbol{\beta})$ is not a square integrable martingale, but in spite of that, it also has a zero-mean. The process $N_i(s,t)$ represents the number of inter events for the i^{th} subject that occurred over [0, s] with interoccurence times at most t.

Proposition 3.1. *For each* $i = 1, ..., n$, $A_i(s,t;\boldsymbol{\beta}) = \int_0^t Y_i(s,w;\boldsymbol{\beta})\lambda_0(w)dw$, *where*

$$\begin{aligned} Y_i(s,t;\boldsymbol{\beta}) = &\sum_{j=1}^{N_{i((s\wedge\tau_i)-)}} \exp(\boldsymbol{\beta}' X_i(t + S_{ij-1}))I(T_{ij} \ge t) \\ &+ \exp(\boldsymbol{\beta}' X_i(t + S_{iN_{i((s\wedge\tau_i)-)}}))I((s\wedge\tau_i) - S_{iN_{i((s\wedge\tau_i)-)}} \ge t) \end{aligned}$$

is for each $t \in \mathscr{T}$ *an* $\mathscr{F}-$*predicatble process. Furthermore, for each*$(s,t) \in \mathscr{T}^2$, $Y_i(s,w;\boldsymbol{\beta}) \le N_i(s-) + 1$, *thus* $E\{Y_i(s,w;\boldsymbol{\beta})\} < \infty$.

Proof.

$$A_i(s,t;\boldsymbol{\beta}) = \int_0^{(s\wedge\tau_i)} Y_i(v)Z_i(v,t)\lambda_0(R_i(v))\exp(\boldsymbol{\beta}' X_i(v))dv$$

$$= \int_0^{(s\wedge\tau_i)} I\{\tau_i \geq v\} Z_i(v,t)\lambda_0(R_i(v))\exp(\boldsymbol{\beta}'\mathbf{X}_i(v))dv$$

$$= \sum_{j=1}^{N_{i((s\wedge\tau_i)-)}} \int_{S_{ij-1}}^{S_{ij}} I\{R_i(v) \leq t\}\lambda_0(R_i(v))\exp(\boldsymbol{\beta}' X_i(v))dv$$

$$+ \int_{S_{iN_{i((s\wedge\tau_i)-)}}}^{(s\wedge\tau_i)} I\{R_i(v) \leq t\}\lambda_0(R_i(v))\exp(\boldsymbol{\beta}' X_i(v))dv$$

Let $w = R_i(v) = v - S_{iN_i(v-)}$ so that dw = dv

with this substitution,

$$\begin{aligned} if \quad v &= S_{ij-1}, \quad w = 0, \\ if \quad v &= S_{ij}, \quad w = T_{ij} \end{aligned}$$

and

$$\begin{aligned} if \quad v &= S_{iN_{i((s\wedge\tau_i)-)}}, \quad w = 0, \\ if \quad v &= (s\wedge\tau_i), \quad w = R_i(s\wedge\tau_i) \end{aligned}$$

Thus,

$$\begin{aligned} A_i(s,t;\boldsymbol{\beta}) &= \sum_{j=1}^{N_{i((s\wedge\tau_i)-)}} \int_0^{T_{ij}} I\{w \leq t\}\lambda_0(w)\exp(\boldsymbol{\beta}'\mathbf{X}_i(w + S_{ij-1}))dw \\ &+ \int_{S_{iN_{i((s\wedge\tau_i)-)}}}^{(s\wedge\tau_i)} I\{w \leq t\}\lambda_0(w)\exp(\boldsymbol{\beta}'\mathbf{X}_i(w + S_{iN_{i((s\wedge\tau_i)-)}}))dw \\ &= \int_0^t \Bigg\{ \sum_{j=1}^{N_{i((s\wedge\tau_i)-)}} I\{T_{ij} \geq w\}\exp(\boldsymbol{\beta}'\mathbf{X}_i(w + S_{ij-1})) \\ &+ I\{R_i(s\wedge\tau_i) \geq w\}\exp(\boldsymbol{\beta}'\mathbf{X}_i(w + S_{iN_{i((s\wedge\tau_i)-)}}))\Bigg\}\lambda_0(w)dw \\ &= \int_0^t Y_i(s,w;\boldsymbol{\beta})\Lambda_0(dw). \end{aligned}$$

□

The process $Y_i(s,t;\boldsymbol{\beta})$ is the generalized at-risk process and it keeps track of the number of gap times that exceed t by calender time s. Furthermore, for each fixed t, the s-indexed process $Y_i(.,t;\boldsymbol{\beta})$ is left- continuous, and hence $\mathscr{F}-$predictable. As a consequence, for each fixed t,

$$M_i(\cdot,t;\boldsymbol{\beta}) = N_i(\cdot,t) - \int_0^t Y_i(\cdot,w;\boldsymbol{\beta})\Lambda_0(dw) \tag{3.5}$$

is a square integrable martingale with respect to $\mathscr{F}_s$.

3.1. METHOD OF MOMENT ESTIMATORS

To derive the estimator of Λ_0, we use the alternative martingale form in (3.5). Let $S^{(0)}(s,t;\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n Y_i(s,t;\boldsymbol{\beta})$ and $J(s,t;\boldsymbol{\beta}) = I\{S^{(0)}(s,t;\boldsymbol{\beta}) > 0\}$, then

$$\begin{aligned}\int_0^t \frac{J(s,w;\boldsymbol{\beta})}{S^{(0)}(s,t;\boldsymbol{\beta})} M(s,dw) &= \int_0^t \frac{J(s,w;\boldsymbol{\beta})}{S^{(0)}(s,w;\boldsymbol{\beta})} N(s,dw) \\ &\quad - \int_0^t J(s,w;\boldsymbol{\beta})\Lambda_0(dw),\end{aligned} \tag{3.6}$$

and

$$\begin{aligned}\int_0^t \frac{J(s,w;\boldsymbol{\beta})}{S^{(0)}(s,w;\boldsymbol{\beta})} M(s,dw) &= \sum_{i=1}^n \int_0^t \frac{J(s,w;\boldsymbol{\beta})}{S^{(0)}(s,w;\boldsymbol{\beta})} M_i(s,dw) \\ &= \sum_{i=1}^n \int_0^s \frac{J(s,R_i(v);\boldsymbol{\beta})}{S^{(0)}(s,R_i(v);\boldsymbol{\beta})} M_i(dv,t).\end{aligned}$$

Furthermore, since $M_i(.,t;\boldsymbol{\beta})$ is a square integrable martingale with respect to $\mathscr{F}_s$, it follows from stochastic integration theory that $\int_0^t \frac{J(s,w;\beta)}{S^{(0)}(s,w;\beta)} M_i(s,dw)$ is a square integrable martingale and

$$E\left\{\int_0^t \frac{J(s,w;\boldsymbol{\beta})}{S^{(0)}(s,w;\boldsymbol{\beta})} M_i(s,dw)\right\} = 0. \tag{3.7}$$

Consequently, the moment identity can be shown as follows:

$$E\left\{\int_0^t \frac{J(s,w;\boldsymbol{\beta})}{S^{(0)}(s,w;\boldsymbol{\beta})} M(s,dw)\right\} = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{J(s,w;\boldsymbol{\beta})}{S^{(0)}(s,w;\boldsymbol{\beta})} M_i(s,dw).$$

Thus, it follows from (3.6), (3.7) that for a given value of $\boldsymbol{\beta}$ over [0, s], the Nelson-Aalen estimator of Λ_0 is given by

$$\tilde{\Lambda}_0(s,t;\boldsymbol{\beta}) = \frac{1}{n}\int_0^t \frac{J(s,w;\boldsymbol{\beta})}{S^{(0)}(s,w;\boldsymbol{\beta})} N(s,dw). \tag{3.8}$$

It follows that the survivor function associated with $\Lambda_0(.)$ is defined by $\bar{F}(t) = \exp\{-\Lambda_0(t)\}$. The estimator of the survivor function sometimes known as the product-limit estimator of $\bar{F}(t)$, is obtained as

$$\tilde{\bar{F}}_0(s,t;\boldsymbol{\beta}) = \prod_{w=0}^{t}(1-\tilde{\Lambda}_0(s,dw;\boldsymbol{\beta})). \tag{3.9}$$

Note that $\tilde{\Lambda}_0$ is not set an estimator, since $\boldsymbol{\beta}$ is unknown. For the case when $\boldsymbol{\beta}$ is unknown, the estimator of the baseline survivor function can be evaluated from the estimated regression coefficients $\boldsymbol{\beta}$. The next section will discuss developing the profile likelihood for $\boldsymbol{\beta}$ from which the estimator of $\boldsymbol{\beta}$ is obtained.

3.2. ESTIMATION OF $\boldsymbol{\beta}$

Our aim in this section is to derive likelihood function for estimating $\boldsymbol{\beta}$. This can be done in two steps. One first obtain the full likelihood as we have two unknowns $\tilde{\Lambda}_0$ and $\boldsymbol{\beta}$. Since we have an expression for Λ_0 as a function of $\boldsymbol{\beta}$, one can just maximize $L(\cdot;\Lambda_0(\cdot),\boldsymbol{\beta})$ with respect to $\tilde{\Lambda}_0$ to obtain the partial likelihood which can later be used to estimate $\boldsymbol{\beta}$. To that end, the profile likelihood function for $\boldsymbol{\beta}$ is obtained by inserting $\tilde{\Lambda}_0$ into the partial likelihood function which in turn can be maximized to give $\hat{\boldsymbol{\beta}}$.

Following Jacod [36], the full likelihood process is:

$$\begin{aligned} L(s^*, t; \boldsymbol{\beta}) = &\left\{ \prod_{i=1}^{n} \prod_{v=0}^{s^*} \Big[Y_i(v)\lambda_0(R_i(v))\exp(\boldsymbol{\beta}'\mathbf{X}_i(v)) \Big]^{dN_i(v)Z_i(v,t)} \right\} \\ &\times \exp\Big[-\sum_{i=1}^{n} \int_0^{s^*} Y_i(v)Z_i(v,t)\lambda_0(R_i(v))\exp(\boldsymbol{\beta}'\mathbf{X}_i(v))dv \Big], \end{aligned} \tag{3.10}$$

where $s^* = \max\ \tau_i$. Let $s^* = t^* \geq \max\ \tau_i$, we get

$$\begin{aligned} L(s^*, t^*; \boldsymbol{\beta}) = &\left\{ \prod_{i=1}^{n} \prod_{v=0}^{t^*} \Big[Y_i(v)\lambda_0(R_i(v))\exp(\boldsymbol{\beta}'\mathbf{X}_i(v)) \Big]^{dN_i(v)Z_i(v,t^*)} \right\} \\ &\times \left\{ \exp\Big[-\sum_{i=1}^{n} \int_0^{t^*} Y_i(s^*, w; \boldsymbol{\beta})\Lambda_0(dw) \Big] \right\} \\ = &\left\{ \prod_{i=1}^{n} \prod_{v=0}^{t^*} \Big[Y_i(v)\lambda_0(R_i(v))\exp(\boldsymbol{\beta}'\mathbf{X}_i(v)) \Big]^{dN_i(v)Z_i(v,t^*)} \right\} \\ &\times \exp\Big[-\int_0^{t^*} nS^{(0)}(s^*, w; \boldsymbol{\beta})\Lambda_0(dw) \Big]. \end{aligned} \tag{3.11}$$

In (3.11), we may then replace the differentials of $dN_i(w)$ and $d\Lambda_0(w)$ by the increments $\Delta N_i(w)$ and $\Delta\Lambda_0(w)$; the integral $\int Y_i(w)d\Lambda_0(w)$ becomes the sum $\sum \Delta\Lambda_0(w)$. For fixed value of $\boldsymbol{\beta}$, maximization of (3.11) with respect to $\Delta\Lambda_0(t)$ leads to

$$\Delta\tilde{\Lambda}_0(s^*, t^*; \boldsymbol{\beta}) = \frac{\Delta N(s^*, t^*))}{nS^{(0)}(s^*, t^*; \boldsymbol{\beta})}. \tag{3.12}$$

Thus, for fixed value of $\boldsymbol{\beta}$, $\tilde{\Lambda}_0(s^*, t^*; \boldsymbol{\beta})$ can be estimated by the Nelson-Aalen estimator given by (3.8). Inserting (3.8) into (3.11), we obtain the following profile likelihood

$L_p(s^*, t^*; \boldsymbol{\beta})$:

$$L_p(s^*, t^*; \boldsymbol{\beta}) = \prod_{i=1}^{n} \prod_{v=0}^{s^*} \left[Y_i(v) \hat{\Lambda}_0(s^*, \Delta R_i(v); \boldsymbol{\beta}) \exp(\boldsymbol{\beta}' \mathbf{X}_i(v)) \right]^{\Delta N_i(v) Z_i(v, t^*)} \times \exp\{-N(s^*, t^*)\}$$

$$\propto \prod_{i=1}^{n} \prod_{v=0}^{s^*} \left[Y_i(v) \hat{\Lambda}_0(s^*, \Delta R_i(v); \boldsymbol{\beta}) \exp(\boldsymbol{\beta}' \mathbf{X}_i(v)) \right]^{\Delta N_i(v) Z_i(v, t^*)}$$
$$\propto \prod_{i=1}^{n} \prod_{v=0}^{s^*} \left[Y_i(v) \left\{ \frac{N(s^*, \Delta R_i(v)))}{S_0(s^*, R_i(v); \boldsymbol{\beta})} \right\} \exp(\boldsymbol{\beta}' \mathbf{X}_i(v)) \right]^{\Delta N_i(v) Z_i(v, t^*)},$$

Since $Y_i(v)$ and $N(s, \Delta R_i(v)))$ are independent of $\boldsymbol{\beta}$, then the partial likelihood profile can be written in the form

$$L_p(s^*, t^*; \boldsymbol{\beta}) = \prod_{i=1}^{n} \prod_{j=1}^{N_{i((s^* \wedge \tau_i)-)}} \left[\frac{\exp(\boldsymbol{\beta}' \mathbf{X}_i(S_{ij}))}{S_0(s^*, R_i(S_{ij}); \boldsymbol{\beta})} \right]^{\Delta N_i(S_{ij}) Z_i(S_{ij}, t^*)}. \tag{3.13}$$

Therefore, we can estimate $\boldsymbol{\beta}$ from (3.13), or equivalently from the log-partial likelihood profile takes the form

$$\begin{aligned} l_p(s^*, t^*; \boldsymbol{\beta}) &= \sum_{i=1}^{n} \int_0^{s^*} \left[\boldsymbol{\beta}' \mathbf{X}_i(v) - \log S_0(s, R_i(v); \boldsymbol{\beta}) \right] Z_i(v, t^*) dN_i(v) \\ &= \sum_{i=1}^{n} \int_0^{s^*} \left[\boldsymbol{\beta}' \mathbf{X}_i(v) - \log S_0(s^*, R_i(v); \boldsymbol{\beta}) \right] N_i(dv, t^*). \end{aligned} \tag{3.14}$$

The derivative of $l_p(s^*, t^*; \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ yields the vector of score statistics $\mathbf{U}(\boldsymbol{\beta}) = \nabla_{\boldsymbol{\beta}} l_p(s^*, t^*; \boldsymbol{\beta}) = \frac{\partial}{\partial \beta_j} l_p(s^*, t^*; \boldsymbol{\beta}) = U^j(\boldsymbol{\beta}),\ j = i, ..., q$, where

$$U^j(s^*, t^*; \boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{s^*} \frac{\partial}{\partial \beta_j} \left[\beta' \mathbf{X}_i(v) - \log S_0(s, R_i(v); \beta) \right] N_i(dv, t). \tag{3.15}$$

The score statistic can thus be written as

$$\mathbf{U}(s^*, t^*; \boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{j=1}^{N_{i(s^*-)}} \int_{S_{ij-1}}^{S_{ij}} \nabla_\beta \Big[\boldsymbol{\beta}' \mathbf{X}_i(v) - \log S_0(s, R_i(v); \boldsymbol{\beta}) \Big] N_i(dv, t)$$
$$+ \sum_{i=1}^{n} \sum_{j=1}^{N_{i(s^*-)}} \int_{S_{iN_{i(s^*-)}}}^{s^*} \nabla_\beta \Big[\boldsymbol{\beta}' \mathbf{X}_i(v) - \log S_0(s, R_i(v); \boldsymbol{\beta}) \Big] N_i(dv, t). \tag{3.16}$$

The maximum partial likelihood estimator $\hat{\boldsymbol{\beta}}$ is a solution to the equation $\mathbf{U}(s^*, t^*; \boldsymbol{\beta}) = 0$. It is obvious that numerical techniques, such as the Newton-Raphson Method can be applied to obtain the estimate $\hat{\boldsymbol{\beta}}$. The estimator of $\Lambda_0(t)$ based on the observable realization over $[0, s^*]$ is obtained by substituting $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ in the expression $\boldsymbol{\beta}$. Accordingly,

$$\hat{\Lambda}_0(s^*, t^*; \hat{\boldsymbol{\beta}}) = \frac{1}{n} \left\{ \int_0^{t^*} \frac{J(s^*, w; \hat{\boldsymbol{\beta}})}{S^{(0)}(s^*, w; \hat{\boldsymbol{\beta}})} N(s^*, dw) \right\}, \tag{3.17}$$

which is often called the Breslow estimator.

3.3. ASYMPTOTIC PROPERTIES

The asymptotic properties of the estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}$ are discussed in this section. To establish the asymptotic properties, we require some notation and regularity conditions. Some important definitions are:

$$\begin{aligned}
\mathbf{S}^{(1)}(s, t; \boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i(t) Y_i(s, t; \boldsymbol{\beta}) \\
\mathbf{S}^{(2)}(s, t; \boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i^{\otimes 2}(t) Y_i(s, t; \boldsymbol{\beta}) \\
\mathbf{E}(s, t; \boldsymbol{\beta}) &= \frac{\mathbf{S}^{(1)}(s, t; \boldsymbol{\beta})}{S^{(0)}(s, t; \boldsymbol{\beta})} \\
\mathbf{V}(s, t; \boldsymbol{\beta}) &= \frac{\mathbf{S}^{(2)}(s, t; \boldsymbol{\beta})}{S^{(0)}(s, t; \boldsymbol{\beta})} - \mathbf{E}(s, t; \boldsymbol{\beta})^{\otimes 2},
\end{aligned} \tag{3.18}$$

where for a $q-$ vector $a, a^{\otimes 2}$ is the $q \times q$ matrix aa'. The expressions $\mathbf{S}^{(1)}(s,t;\boldsymbol{\beta})$ and $\mathbf{S}^{(2)}(s,t;\boldsymbol{\beta})$ are the first and second-order partial derivatives, respectively, of $S^{(0)}(s,t;\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. Furthermore, the vector $\mathbf{E}(s,t;\boldsymbol{\beta})$ and the matrix $\mathbf{V}(s,t;\boldsymbol{\beta})$ are the expectation and the covariance, respectively, of the covariate vector $\mathbf{X}_i(t)$. In the proofs, we need the following regularity condition:

Condition 1. There exist a neighborhood $\mathscr{B}$ of $\boldsymbol{\beta}_0$ such that for all $s \in [0, s^*]$, $t \in [0, t^*]$, $(s,t) \in \mathscr{T}^2, \boldsymbol{\beta} \in \mathscr{B}$, and $m = 1, 2$:

(a) There exists a deterministic function $s^{(0)} : \mathscr{T} \times \mathscr{B} \to \mathbb{R}_+$ such that

$$\sup_{t\in\mathscr{T};\beta\in\mathscr{B}} |S^{(0)}(s,t;\boldsymbol{\beta}) - s^{(0)}(s,t;\boldsymbol{\beta})| \xrightarrow{p} 0,$$

and with $\inf_{t\in\mathscr{T}} s^{(0)}(s,t;\boldsymbol{\beta}) > 0$ and with $\Lambda_0(t) = \int_0^t \lambda_0(w)dw < \infty$;

(b) There exists deterministic functions $\mathbf{s}^{(1)} : \mathscr{T} \times \mathscr{B} \to \mathbb{R}^q$ and $\mathbf{s}^{(2)} : \mathscr{T} \times \mathscr{B} \to \mathbb{R}^{q\times q}$ such that

$$\sup_{t\in\mathscr{T}} |\mathbf{S}^{(m)}(s,t;\boldsymbol{\beta}) - \mathbf{s}^{(m)}(s,t;\boldsymbol{\beta})| \xrightarrow{p} 0 \quad n \to \infty;$$

(c) $\mathbf{s}^{(m)}(s,.;\boldsymbol{\beta}_0)$ is bounded on $\mathscr{T} \times \mathscr{B}$ and is a continuous function of $\beta \in \mathscr{B}$ uniformly in $t \in \mathscr{T}$;

(d) For $t \in \mathscr{T}$ and $\boldsymbol{\beta} \in \mathscr{B}$, $\mathbf{s}^{(0)}(s,.;\boldsymbol{\beta}_0)$ is bounded;

(e) $\mathbf{s}^{(1)}(s,t;\boldsymbol{\beta}) = \nabla_{\boldsymbol{\beta}} s^{(0)}(s,t;\boldsymbol{\beta}), \quad \mathbf{s}^{(2)}(s,t;\boldsymbol{\beta}) = \nabla_{\boldsymbol{\beta}} \nabla_{\boldsymbol{\beta}'} s^{(0)}(s,t;\boldsymbol{\beta})$;

(f) $\boldsymbol{\Sigma}(s,t) = \int_0^t \mathbf{v}(s,w;\boldsymbol{\beta}_0)\mathbf{s}^{(0)}(s,w;\boldsymbol{\beta}_0)\lambda(w)dw$ is positive definite, where $\mathbf{v} = \frac{\mathbf{s}^{(2)}(s,t;\boldsymbol{\beta}_0)}{s^{(0)}(s,t;\boldsymbol{\beta}_0)} - \mathbf{e}^{\otimes 2}(s,t;\boldsymbol{\beta}_0)$ and $\mathbf{e} = \frac{\mathbf{s}^{(1)}(s,t;\boldsymbol{\beta}_0)}{s^{(0)}(s,t;\boldsymbol{\beta}_0)}$.

Before we proceed further, we need to extend Theorem 1 of Peña et al.[37]. For $i = 1, ..., n, s \in [0, s^*], t \in [0, t^*]$, and $(s,t) \in \mathscr{T}^2$, let $\mathbf{H}_i(s,t;\boldsymbol{\beta})$ be q-dimensional vector-valued processes on $(\Omega,\ \mathscr{F},\ \mathscr{P})$ with $\left\{H_{i,k}(.,.;\beta_{i,k}) : k = 1, .., q\right\}$ bounded and $\mathscr{F}$-predictable. Let

$$\mathbf{W}(s,t;\boldsymbol{\beta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t \mathbf{H}_i(s,w;\boldsymbol{\beta})M_i(s,dw;\boldsymbol{\beta}),$$
$$\mathbf{V}(s,t;\boldsymbol{\beta}) = \sum_{i=1}^{n}\int_0^t [\mathbf{H}_i(s,w;\boldsymbol{\beta})]^{\otimes 2}S_i^{(0)}(s,w;\boldsymbol{\beta})\lambda(w)dw.$$

be the integral-transformed processes.

Lemma 3.2. *Suppose the following conditions are satisfied for $t \in [0, t^*]$ where $t^* \in \mathscr{T}$:*

1. *There exists a deterministic function $\boldsymbol{h}(v,w;\boldsymbol{\beta})$ on $[0,s^*]\times[0,t^*]$ which is continuous in (v, w) and bounded such that*

$$\max_{1\le i\le n}\sup_{0\le w\le t^*}|\boldsymbol{H}_i(s,w;\boldsymbol{\beta}) - \boldsymbol{h}(s,w;\boldsymbol{\beta})| \xrightarrow{p} 0;$$

2. *The function*

$$\breve{\boldsymbol{\Sigma}}(s,t;\boldsymbol{\beta}) = \int_0^t \boldsymbol{h}(s,w;\boldsymbol{\beta})^{\otimes 2}s^{(0)}(s,w;\boldsymbol{\beta})\lambda(w)dw$$

   is such that for each t, as $n\to\infty$, $\|\,\boldsymbol{V}(s,t;\boldsymbol{\beta}) - \breve{\boldsymbol{\Sigma}}(s,t;\boldsymbol{\beta})\| \xrightarrow{p} 0$.

Then, as $n\to\infty$, the processes $\{\,\boldsymbol{W}(s,t;\boldsymbol{\beta}) : t\in[0,t^*]\}$ converges weakly on the Skorohod space $\mathscr{D}[0,t^*]$ to the Gaussian process with zero mean function and covariance function given by

$$\breve{\boldsymbol{\Sigma}}(s,t;\boldsymbol{\beta}) = \int_0^t \boldsymbol{h}(s,w;\boldsymbol{\beta})^{\otimes 2}s^{(0)}(s,w;\boldsymbol{\beta})\lambda(w)dw$$

Proof. The proof is the similar to the proof of Theorem 1 of Peña et al. [37]. □

We are now able to prove the following consistency theorems.

3.3.1. Consistency of Estimators.

Theorem 3.3. *Under Condition (1), there exists a unique solution* $\hat{\boldsymbol{\beta}}$ *to the equation* $\boldsymbol{U}(\boldsymbol{\beta}) = 0$ *and* $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$ *as* $n \to \infty$.

Proof. For $i = 1, ..., n$ and $(s^*, t^*) \in \mathscr{T}^2$, consider the function $\mathbf{D}(s^*, t^*; \boldsymbol{\beta})$ given by

$$\begin{aligned}\mathbf{D}(s^*, t^*; \boldsymbol{\beta}) &= \frac{1}{n}\Big[l_p(s^*, t^*; \boldsymbol{\beta}) - l_p(s^*, t^*; \boldsymbol{\beta}_0)\Big] \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{t^*}\left\{(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{X}_i(s^*) - \log\frac{S^{(0)}(s^*, w; \boldsymbol{\beta})}{S^{(0)}(s^*, w; \boldsymbol{\beta}_0)}\right\}N_i(s^*, dw).\end{aligned} \tag{3.19}$$

where $l_p(s^*, t^*; \boldsymbol{\beta})$ is the log-profile likelihood given in (3.14). Now, we need to show that $\mathbf{D}(s^*, t^*; \boldsymbol{\beta})$ is a concave function which converges pointwise in probability to a concave function of $\boldsymbol{\beta}$ with a unique maximum at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Using the fact that $M_i(s^*, dw; \boldsymbol{\beta}) = N_i(s^*, dw) - Y_i(s^*, w; \boldsymbol{\beta})\Lambda_0(dw)$, we have

$$\begin{aligned}\mathbf{D}(s^*, t^*; \boldsymbol{\beta}) = &\frac{1}{n}\sum_{i=1}^{n}\int_0^{t^*}\left\{(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{X}_i(s^*) - \log\frac{S^{(0)}(s^*, w; \boldsymbol{\beta})}{S^{(0)}(s^*, w; \boldsymbol{\beta}_0)}\right\}M_i(s^*, dw; \boldsymbol{\beta}) \\ &+ \frac{1}{n}\sum_{i=1}^{n}\int_0^{t^*}\left\{(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{X}_i(s^*) - \log\frac{S^{(0)}(s^*, w; \boldsymbol{\beta})}{S^{(0)}(s^*, w; \boldsymbol{\beta}_0)}\right\}Y_i(s^*, w; \boldsymbol{\beta})\Lambda_0(dw),\end{aligned} \tag{3.20}$$

where the first term on the right-hand side of (3.20) is a local square integrable martingale with predictable variation process

$$\begin{aligned}\langle M, M\rangle(t^*) =&\frac{1}{n^2}\sum_{i=1}^{n}\int_0^{t^*}\bigg\{(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{X}_i(s^*) \\ &- \log\frac{S^{(0)}(s^*, w; \boldsymbol{\beta})}{S^{(0)}(s^*, w; \boldsymbol{\beta}_0)}\bigg\}^2 Y_i(s^*, w; \boldsymbol{\beta})\Lambda_0(dw).\end{aligned} \tag{3.21}$$

It follows from Lemma 3.2 and Condition 1 (a), (b), (c), and (d) that the first term in (3.20) is $o_p(1)$. Furthermore, the second term in(3.20) for $\boldsymbol{\beta} \in \mathscr{B}$ converges in probability to some function

$$\Upsilon(s^*, t^*; \boldsymbol{\beta}) = \int_0^{t^*} \Bigg[(\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \mathbf{s}^{(1)}(s^*, w; \boldsymbol{\beta}_0) \\ - \log \left(\frac{s^{(0)}(s^*, w; \boldsymbol{\beta})}{s^{(0)}(s^*, w; \boldsymbol{\beta}_0)} \right) s^{(0)}(s^*, w; \boldsymbol{\beta}_0) \Bigg] \Lambda_0(dw). \quad (3.22)$$

Thus, we have the result

$$\mathbf{D}(s^*, t^*; \boldsymbol{\beta}) \xrightarrow{p} \quad \Upsilon(s^*, t^*; \boldsymbol{\beta}).$$

By Condition 1 (c)-(f), and assuming that we may interchange the order of integration and differentiation in (3.20), we have for $\boldsymbol{\beta} \in \mathscr{B}$

$$\nabla_{\boldsymbol{\beta}} \Upsilon(s^*, t^*; \boldsymbol{\beta}) = \int_0^{t^*} \left(\mathbf{e}(s^*, w; \boldsymbol{\beta}_0) - \mathbf{e}(s^*, w; \boldsymbol{\beta}) \right) s^{(0)}(s^*, w; \boldsymbol{\beta}_0) \Lambda_0(dw), \quad (3.23)$$

which is zero for $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Moreover,

$$-\nabla_{\boldsymbol{\beta}} \nabla_{\boldsymbol{\beta}'} \Upsilon(s^*, t^*; \boldsymbol{\beta}) = \int_0^{t^*} \mathbf{v}(s^*, w; \boldsymbol{\beta}) s^{(0)}(s^*, w; \boldsymbol{\beta}_0) \Lambda_0(dw), \quad (3.24)$$

which is negative semidefinite and concave. It follows from Condition 1 (c)-(f) that the right side of (3.24) is positive definite for $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Thus, $\mathbf{D}(s^*, t^*; \boldsymbol{\beta})$ converges pointwise in probability to a concave function $\Upsilon(s^*, t^*; \boldsymbol{\beta})$ on $\mathscr{B}$ with a unique maximum at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Therefore, $\mathbf{D}(s^*, t^*; \boldsymbol{\beta})$ is also concave and has a maximum at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ when $\hat{\boldsymbol{\beta}}$ exists. Theorem 2.12 and Corollary 2.13 of Andersen and Gill [41] imply that the maximizing values $\hat{\boldsymbol{\beta}}$ of $\mathbf{D}(s^*, t^*; \boldsymbol{\beta})$ converges in probability to the maximizing value $\boldsymbol{\beta}_0$ of $\Upsilon(s^*, t^*; \boldsymbol{\beta})$. $\square$

Theorem 3.4. *Under Condition 1, $\hat{\Lambda}_0(s^*, t^*; \boldsymbol{\beta})$ converges uniformly in probability to $\Lambda_0(.)$ on $[0, t^*]$, that is*

$$\sup_{w \in [0,t]} |\hat{\Lambda}_0(s^*, w; \boldsymbol{\beta}) - \Lambda_0(w)| \xrightarrow{p} 0.$$

Proof. Let $\mathscr{D}[0, s^*]$ be the space of the cadlag function on $[0, s^*]$ and endow this space with the Skorohod metric. For $D \in \mathscr{D}[0, s^*]$, denote $\|D\|_\infty = \sup_{v \in [0,s^*]} |D(v)|$, let $\mathscr{G} = \mathscr{D}[0, s^*]^2 \times \mathbb{R}_+^q$, and define $d : \mathscr{G} \times \mathscr{G} \mapsto \mathbb{R}_+^q$ with

$$d\left([U_1, U_2, x], [V_1, V_2, y]\right) = \sqrt{\|U_1 - V_1\|_\infty^2 + \|U_2 - V_2\|_\infty^2 + |x - y|^2}.$$

Define

$$Q \equiv Q_n = \left\{ \left(\frac{N(s^*, w)}{n}, S^{(0)}(s^*, w; \hat{\boldsymbol{\beta}}) \right); \hat{\boldsymbol{\beta}} \right\},$$

where $Q \in \mathscr{G}$, and

$$Q_0 = \left\{ \left(\int_0^t s^{(0)}(s^*, w; \boldsymbol{\beta}_0) \lambda_0(w) dw, s^{(0)}(s^*, w; \boldsymbol{\beta}_0) \right); \boldsymbol{\beta}_0) \right\}.$$

We can show that for any sequence of elements of $\mathscr{G} \times \mathscr{G}$, Q_n converges to Q_0, $d(Q_n, Q_0) \xrightarrow{p} 0$. Therefore, it follows from Lemma 3.2 that $\left\{ M(s^*, t; \boldsymbol{\beta}_0)/\sqrt{n} : t \in [0, t^*] \right\}$ converges weakly to a Gaussian process, thus $\left| \frac{N(s^*,w)}{n} - \int_0^t s^{(0)}(s^*, w; \boldsymbol{\beta}_0) \lambda_0(w) dw \right|$ $= o_p(1)$. By Condition 1 (a) and the consistency of $\hat{\boldsymbol{\beta}}$, we have

$$d \left[\left\{ \left(\frac{N(s^*, w)}{n}, S^{(0)}(s^*, w; \hat{\boldsymbol{\beta}}) \right); \hat{\boldsymbol{\beta}} \right\}, \right.$$
$$\left. \left\{ \left(\int_0^t s^{(0)}(s^*, w; \boldsymbol{\beta}_0) \Lambda_0(dw), s^{(0)}(s^*, w; \boldsymbol{\beta}_0) \right); \boldsymbol{\beta}_0) \right\} \right] \xrightarrow{p} 0.$$

Now, we recall first that

$$\hat{\Lambda}_0(s^*, t; \hat{\boldsymbol{\beta}}) = \frac{1}{n}\left\{\int_0^t \frac{J(s^*, w; \hat{\boldsymbol{\beta}})}{S^{(0)}(s^*, w; \hat{\boldsymbol{\beta}})} N(s^*, dw)\right\}. \tag{3.25}$$

Let $H : \mathscr{G} \mapsto \mathscr{D}[0, t^*]$ which maps Q into $\hat{\Lambda}_0(s^*, t; \hat{\boldsymbol{\beta}})$. Then, we have $H(Q_n) = \hat{\Lambda}_0(s^*, t; \hat{\boldsymbol{\beta}})$ and $H(Q_0) = \Lambda_0(t)$. If we could show that the mapping H is continuous, the consistency of $\hat{\Lambda}$ will follow by the continuous mapping theorem and the fact that $d(Q_n, Q_0) \xrightarrow{p} 0$. To show continuity of this mapping, we have

$$\begin{aligned}
|H(Q_n) - H(Q_0)| &= \int_0^t \left| \frac{\frac{N(s^*,dw)}{n}}{S^{(0)}(s^*, w; \hat{\boldsymbol{\beta}})} - \Lambda_0(dw) \right| \\
&\leq \int_0^t \left| \frac{\frac{N(s^*,dw)}{n}}{S^{(0)}(s^*, w; \hat{\boldsymbol{\beta}})} - \frac{N(s^*, dw)}{ns^{(0)}(s^*, w; \boldsymbol{\beta}_0)} \right| \\
&+ \int_0^t \left| \frac{N(s^*, dw)}{ns^{(0)}(s^*, w; \boldsymbol{\beta}_0)} - \Lambda_0(dw) \right| \\
&= \int_0^t \left| \frac{1}{S^{(0)}(s^*, w; \hat{\boldsymbol{\beta}})} - \frac{1}{s^{(0)}(s^*, w; \boldsymbol{\beta}_0)} \right| \frac{N(s^*, dw)}{n} \\
&+ \int_0^t \left| \frac{N(s^*, dw)}{n} - s^{(0)}(s^*, w; \boldsymbol{\beta}_0)\Lambda_0(dw) \right| \frac{1}{s^{(0)}(s^*, w; \boldsymbol{\beta}_0)}.
\end{aligned} \tag{3.26}$$

Since $N(s^*, w)$ is a non-decreasing non-negative process in w and $s^{(0)}(s^*, w; \boldsymbol{\beta}_0)$ is a non-increasing process in w, it follows that the first term in (3.26) is bounded above by

$$\left\{ \sup_{w \in [0,t^*]} \left| \frac{1}{S^{(0)}(s^*, w; \hat{\boldsymbol{\beta}})} - \frac{1}{s^{(0)}(s^*, w; \boldsymbol{\beta}_0)} \right| \right\} \frac{N(s^*, t^*)}{n}. \tag{3.27}$$

It follows from Condition 1 that

$$\sup_{w \in [0,t^*]} \left| \frac{1}{S^{(0)}(s^*, w; \hat{\boldsymbol{\beta}})} - \frac{1}{s^{(0)}(s^*, w; \boldsymbol{\beta}_0)} \right| = o_p(1) \tag{3.28}$$

and $\frac{N(s^*,t^*)}{n} = O_p(1)$, it follows that the first term is asymptotically negligible. The second term is bounded above by

$$\left\{ \sup_{w \in [0,t^*]} \left| \frac{N(s^*, dw)}{n} - s^{(0)}(s^*, w; \boldsymbol{\beta}_0)\Lambda_0(dw) \right| \right\} \frac{1}{s^{(0)}(s^*, t^*; \boldsymbol{\beta}_0)}, \tag{3.29}$$

which is $o_p(1)$. This is because by the results in , $\left\{ M(s^*, t)/\sqrt{n} : t \in [0, \infty] \right\}$ converges weakly to a Gaussian process, implying

$$\left\{ \sup_{w \in [0,t^*]} \left| \frac{N(s^*, dw)}{n} - s^{(0)}(s^*, w; \boldsymbol{\beta}_0)\Lambda_0(dw) \right| \right\} = o_p(1);$$

whereas by Condition 1 (a) and (c) we have $\frac{1}{s^{(0)}(s^*,t^*;\boldsymbol{\beta}_0)} = O_p(1)$. □

3.3.2. Large Sample Distributional Properties. In this section, we derive the asymptotic Gaussianity for the partial likelihood score vector. The results are then used to establish the asymptotic normality of $\hat{\boldsymbol{\beta}}$ and the weak convergence of $\{W_n(s^*, t) : t \in \mathscr{T}; n = 1, 2, ...\}$, given by

$$W_n(s^*, t) = \sqrt{n} \left[\hat{\Lambda}_0(s^*, t; \hat{\boldsymbol{\beta}}) - \Lambda_0(t) \right].$$

To prove the limiting distributional properties of $\hat{\boldsymbol{\beta}}$ and $W_n(s^*, t)$, we will define the process $\{\Xi(s^*, t; \boldsymbol{\beta}_0) : t \in \mathscr{T}\}$ where

$$\Xi(s^*, t; \boldsymbol{\beta}_0) = \int_0^t J(s^*, w; \boldsymbol{\beta}_0) \frac{S^{(1)}(s^*, w; \boldsymbol{\beta}_0)}{S^{(0)}(s^*, w; \boldsymbol{\beta}_0)^2} N(s^*, dw),$$

and the process

$$\boldsymbol{\vartheta}(s^*, t) = \sqrt{n} \left[\hat{\boldsymbol{\Lambda}}(s^*, t; \hat{\boldsymbol{\beta}}) - \Lambda_0(t) \right] + \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \Xi(s^*, t; \boldsymbol{\beta}_0),$$

Also, let the gradient of $l_p(\cdot, \cdot; \boldsymbol{\beta})$ be given by

$$
\begin{aligned}
\dot{U}(s^*, t; \boldsymbol{\beta}) &= \nabla_\beta \{ \frac{1}{n} l_p(s^*, t; \boldsymbol{\beta}) \} \\
&= \nabla_\beta \left\{ \frac{1}{n} \sum_{i=1}^{n} \int_0^t \left[\boldsymbol{\beta}' \mathbf{X}_i(s^*) - \log S_0(s^*, w; \boldsymbol{\beta}) \right] N_i(s^*, dw) \right\} \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_0^{t^*} \left\{ \mathbf{X}_i(s^*) - \frac{S^{(1)}(s^*, w; \boldsymbol{\beta})}{S^{(0)}(s^*, w; \boldsymbol{\beta})} \right\} N_i(s^*, dw),
\end{aligned}
$$

and, the second partial derivative of $l_p(\cdot, \cdot; \boldsymbol{\beta})$ given by

$$
\begin{aligned}
\ddot{U}(s^*, t; \boldsymbol{\beta}) &= \nabla_\beta \nabla_{\beta'} \{ \frac{1}{n} l_p(s^*, t; \boldsymbol{\beta}) \} = \nabla_\beta \{ \dot{U}(s^*, t; \boldsymbol{\beta}) \} \\
&= \nabla_\beta \left\{ \frac{1}{n} \sum_{i=1}^{n} \int_0^{t^*} \left\{ \mathbf{X}_i(s^*) - \frac{S^{(1)}(s^*, w; \boldsymbol{\beta})}{S^{(0)}(s^*, w; \boldsymbol{\beta})} \right\} N_i(s^*, dw) \right\} \\
&= \frac{-1}{n} \sum_{i=1}^{n} \int_0^{t^*} \left[\frac{S^{(2)}(s^*, w; \boldsymbol{\beta})}{S^{(0)}(s^*, w; \boldsymbol{\beta})} - \left(\frac{S^{(1)}(s^*, w; \boldsymbol{\beta})}{S^{(0)}(s^*, w; \boldsymbol{\beta})} \right)^{\otimes 2} \right] N_i(s^*, dw) \\
&= \frac{-1}{n} \sum_{i=1}^{n} \int_0^{t^*} \mathbf{V}(s^*, w; \boldsymbol{\beta}) N_i(s^*, dw).
\end{aligned}
$$

The next lemma gives the in-probability limit of $\ddot{U}(s^*, t^*; \boldsymbol{\beta})$

Lemma 3.5. *Under Condition 1,*

$$
\begin{aligned}
\ddot{U}(s^*, t^*; \boldsymbol{\beta}) = \\
& - \sum_{i=1}^{n} \int_0^{t^*} \boldsymbol{V}(s^*, w; \boldsymbol{\beta}) S^{(0)}(s^*, w; \boldsymbol{\beta}) \Lambda_0(dw) + o_p(1) \\
& \xrightarrow{p} -\boldsymbol{\Sigma}(s^*, t^*).
\end{aligned}
$$

Proof. Straightforward calculations show that

$$\begin{aligned}
\ddot{U}(s^*, t^*; \boldsymbol{\beta}) &= \frac{-1}{n} \sum_{i=1}^{n} \int_0^{t^*} \mathbf{V}(s^*, w; \boldsymbol{\beta}) N_i(s^*, dw) \\
&= \frac{-1}{n} \sum_{i=1}^{n} \int_0^{t^*} \mathbf{V}(s^*, w; \boldsymbol{\beta}) [M_i(s^*, dw) + Y_i(s^*, w; \boldsymbol{\beta}) \Lambda_0(dw)] \\
&= \frac{-1}{n} \sum_{i=1}^{n} \int_0^{t^*} \mathbf{V}(s^*, w; \boldsymbol{\beta}) M_i(s^*, dw) \\
&+ \int_0^{t^*} \mathbf{V}(s^*, w; \boldsymbol{\beta}) S^{(0)}(s^*, w; \hat{\boldsymbol{\beta}}) \Lambda_0(dw).
\end{aligned} \tag{3.30}$$

By Condition 1, the second term in (3.30) converges in probability to $\boldsymbol{\Sigma}(s^*, t^*)$. Furthermore, by Lemma 3.2, we have that the process

$$\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{t} \mathbf{V}(s^*, w; \boldsymbol{\beta}) M_i(s^*, dw) : t \in [0, t^*] \right\}$$

converges weakly to a Gaussian process $W(s^*, t)$ with zero mean function and covariance function

$$\mathrm{Cov}(W(s^*, t_1), W(s^*, t_2)) = \int_0^{t_1 \wedge t_2} \mathbf{v}(s^*, w; \boldsymbol{\beta}_0) \mathbf{s}^{(0)}(s^*, w; \boldsymbol{\beta}_0) \Lambda_0(dw)$$

for $t_1, t_2 \in [0, t^*]$. As a result,

$$\sup_{t \in [0, t^*]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{t} \mathbf{V}(s^*, w; \boldsymbol{\beta}) M_i(s^*, dw) \right|$$

converges weakly to $\sup_{t\in[0,t^*]} |W(s^*, t)|$, which is $O_p(1)$. It follows that

$$\begin{aligned}
&\sup_{t\in[0,t^*]} \left| \frac{1}{n} \sum_{i=1}^{n} \int_0^t \mathbf{V}(s^*, w; \boldsymbol{\beta}) M_i(s^*, dw) \right| \\
&= \frac{1}{\sqrt{n}} \sup_{t\in[0,t^*]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t \mathbf{V}(s^*, w; \boldsymbol{\beta}) M_i(s^*, dw) \right| \\
&= o_p(1).
\end{aligned}$$

□

Theorem 3.6. *Under Condition 1, we have the representations*

$$\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \boldsymbol{\Sigma}(s^*, t^*)^{-1} \times& \\
&\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{t^*} [\boldsymbol{X}_i(s^*) - \boldsymbol{E}(s^*, w; \boldsymbol{\beta}_0)] M_i(s^*, dw) \right\} + o_p(1). \quad (3.31)
\end{aligned}$$

Proof. By first-order Taylor expansion of $\dot{U}(s^*, t^*; \hat{\boldsymbol{\beta}})$ around $\boldsymbol{\beta}_0$, we obtain

$$\dot{U}(s^*, t^*; \hat{\boldsymbol{\beta}}) = \dot{U}(s^*, t^*; \hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \ddot{U}(s^*, t^*; \tilde{\boldsymbol{\beta}}),$$

where $\tilde{\boldsymbol{\beta}}$ is on the line segment between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. Since $\dot{U}(s^*, t^*; \hat{\boldsymbol{\beta}}) = 0$, we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = [\sqrt{n} \dot{U}(s^*, t^*; \boldsymbol{\beta}_0)][-\ddot{U}(s^*, t^*; \tilde{\boldsymbol{\beta}})]^{-1},$$

Using the consistency of $\hat{\boldsymbol{\beta}}$ and by Lemma 3.5, we may write

$$[-\ddot{U}(s^*, t^*; \tilde{\boldsymbol{\beta}})]^{-1} = [\boldsymbol{\Sigma}(s^*, t^*)]^{-1} + o_p(1). \quad (3.32)$$

Furthermore, it is seen that

$$\sqrt{n}\dot{U}(s^*,t^*;\boldsymbol{\beta}_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{t^*}[\mathbf{X}_i(s^*)-\mathbf{E}(s^*,w;\boldsymbol{\beta}_0)]M_i(s^*,dw),$$

which, combined with (3.32), gives the representation for $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$. □

Theorem 3.7. *Assume Condition 1 holds. Then* $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ *and the processes*

$$\boldsymbol{\vartheta}(s^*,t^*)=\frac{1}{\sqrt{n}}\int_0^{t^*}\frac{J(s^*,w;\hat{\boldsymbol{\beta}})}{S^{(0)}(s^*,w;\boldsymbol{\beta}_0)}M(s^*,dw)+o_p(1),$$

are asymptotically independent.

Proof. By introducing

$$\begin{aligned}\sqrt{n}\left[\hat{\Lambda}_0(s^*,t^*;\hat{\boldsymbol{\beta}})-\Lambda_0(t^*)\right]=&\sqrt{n}\left[\hat{\Lambda}_0(s^*,t^*;\hat{\boldsymbol{\beta}})-\Lambda^*(s^*,t^*;\hat{\boldsymbol{\beta}})\right]\\&+\sqrt{n}\left[\Lambda^*(s^*,t^*;\hat{\boldsymbol{\beta}})-\Lambda_0(t^*)\right],\end{aligned}\tag{3.33}$$

where $\Lambda^*(s^*,t^*;\boldsymbol{\beta}_0)=\int_0^{t^*}J(s^*,w;\boldsymbol{\beta}_0)\Lambda_0(dw)$. the second term in(3.33) can be written as

$$\sqrt{n}\left[\Lambda^*(s^*,t^*;\hat{\boldsymbol{\beta}})-\Lambda_0(t^*)\right]=\sqrt{n}\int_0^{t^*}I\{S_0(s^*,w;\boldsymbol{\beta}_0)=0\}\Lambda_0(dw)$$

Since $\Lambda_0=\int_0^{t^*}\lambda_0(w)dw<\infty$ and $\inf_{t\in\mathscr{T}}s^{(0)}(s^*,t^*;\boldsymbol{\beta})>0$ by Condition 1 (a) and (c), it follows that

$$\sup_{w\in[0,t^*]}|\Lambda^*(s^*,w;\hat{\boldsymbol{\beta}})-\Lambda_0(dw)|=o_p(1).\tag{3.34}$$

The first-order Taylor expansion of the first term yields

$$
\begin{aligned}
&\sqrt{n}\Big[\hat{\Lambda}_0(s^*,t^*;\hat{\boldsymbol{\beta}}) - \Lambda^*(s^*,t^*;\hat{\boldsymbol{\beta}})\Big] = \\
&\quad \sqrt{n}\int_0^{t^*}\left[\frac{J(s^*,w;\hat{\boldsymbol{\beta}})}{nS^{(0)}(s^*,w;\hat{\boldsymbol{\beta}})}N(s^*,dw) - J(s^*,w;\hat{\boldsymbol{\beta}})\Lambda_0(dw)\right] \\
&\quad = \frac{1}{\sqrt{n}}\int_0^{t^*} J(s^*,w;\hat{\boldsymbol{\beta}})\left\{\frac{1}{S^{(0)}(s^*,w;\hat{\boldsymbol{\beta}})} - (\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)'\frac{S^{(1)}(s^*,w;\hat{\boldsymbol{\beta}})}{S^{(0)}(s^*,w;\hat{\boldsymbol{\beta}})^2}\right\}N(s^*,dw) \\
&\quad - \sqrt{n}\int_0^{t^*} J(s^*,w;\hat{\boldsymbol{\beta}})\Lambda_0(dw) \\
&\quad = \frac{1}{\sqrt{n}}\int_0^{t^*}\frac{J(s^*,w;\hat{\boldsymbol{\beta}})}{S^{(0)}(s^*,w;\hat{\boldsymbol{\beta}})}\Big[N(s^*,dw) - nS^{(0)}(s^*,w;\hat{\boldsymbol{\beta}})\Lambda_0(dw)\Big] \\
&\quad - \frac{1}{\sqrt{n}}\int_0^{t^*}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)'\frac{S^{(1)}(s^*,w;\hat{\boldsymbol{\beta}})}{S^{(0)}(s^*,w;\hat{\boldsymbol{\beta}})^2}N(s^*,dw) \\
&\quad = \frac{1}{\sqrt{n}}\int_0^{t^*}\frac{J(s^*,w;\hat{\boldsymbol{\beta}})}{S^{(0)}(s^*,w;\hat{\boldsymbol{\beta}})}M(s^*,dw) \\
&\quad - \frac{1}{\sqrt{n}}\int_0^{t^*}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)'\frac{S^{(1)}(s^*,w;\hat{\boldsymbol{\beta}})}{S^{(0)}(s^*,w;\hat{\boldsymbol{\beta}})^2}N(s^*,dw).
\end{aligned}
$$

It follows by the consistency of $\hat{\boldsymbol{\beta}}$ and Condition 1 (a),(b),and (c) that

$$
\begin{aligned}
\sqrt{n}\left[\hat{\Lambda}_0(s^*,t^*;\hat{\boldsymbol{\beta}}) - \Lambda^*(s^*,t^*;\hat{\boldsymbol{\beta}})\right] = &\frac{1}{\sqrt{n}}\int_0^{t^*}\frac{J(s^*,w;\hat{\boldsymbol{\beta}})}{S^{(0)}(s^*,w;\hat{\boldsymbol{\beta}})}M(s^*,dw) \\
&+o_p(1).
\end{aligned}
\tag{3.35}
$$

Thus, the representation for $\boldsymbol{\vartheta}(s^*,t)$ follows from (3.33) and (3.34). We may rewrite the established representations as follows:

$$
\begin{aligned}
\begin{bmatrix}\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \\ \boldsymbol{\vartheta}(s^*,t^*)\end{bmatrix} = &\begin{bmatrix}[\boldsymbol{\Sigma}(s^*,t^*)]^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\end{bmatrix} \\
&\times\frac{1}{\sqrt{n}}\int_0^{t^*}\begin{bmatrix}\sum_{i=1}^n[\mathbf{X}_i(s^*) - \mathbf{E}(s^*,w;\boldsymbol{\beta})]M_i(s^*,dw) \\ \frac{J(s^*,w;\hat{\beta})}{S^{(0)}(s^*,w;\beta_0)}M(s^*,dw)\end{bmatrix}.
\end{aligned}
$$

By Theorem 3.6 and Slutaky's theorem, it follow that

$$\begin{bmatrix} \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ \boldsymbol{\vartheta}(s^*, t^*) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{bmatrix} = \begin{bmatrix} [\boldsymbol{\Sigma}(s^*, t^*)]^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{Z}_1(s^*, t^*) \\ \mathbf{Z}_2(s^*, t^*) \end{bmatrix}, \tag{3.36}$$

where $\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1(s^*, t^*) \\ \mathbf{Z}_2(s^*, t^*) \end{bmatrix}$ is a $(q+p)$-dimensional zero mean multivariate normal random vector with covariance matrix

$$Cov(\mathbf{Z}, \mathbf{Z}) = \int_0^{min(t_1,t_2)} \begin{bmatrix} \mathbf{v}(s^*, w; \boldsymbol{\beta}) & \mathbf{0} \\ \mathbf{0} & \frac{1}{s^{(0)}(s^*,w;\boldsymbol{\beta}_0)^2} \end{bmatrix} s^{(0)}(s^*, w; \boldsymbol{\beta}_0)\Lambda_0(dw) \tag{3.37}$$

for $t_1, t_2 \in [0, t^*]$. Consequently, we establish that $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ and $\boldsymbol{\vartheta}(s^*, t^*)$ are asymptotically independent. □

The following two corollaries of Theorem 3.6 are immediate consequences of the preceding discussion.

Corollary 3.8. *Under the conditions of Theorem 3.6, as $n \to \infty$,*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathscr{N}(\mathbf{0}, [\boldsymbol{\Sigma}(s^*, t^*)]^{-1})$$

Proof. Immediate from Theorem 3.6. It follows from (3.36) and (3.37) that

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathscr{N}(\mathbf{0}, [\boldsymbol{\Sigma}(s^*, t^*)^{-1}]' \boldsymbol{\Sigma}(s^*, t^*) \boldsymbol{\Sigma}(s^*, t^*)^{-1}).$$

□

Corollary 3.9. *Under the same conditions as in Theorem 3.6, as $n \to \infty$, the process $W_n(s^*, t^*) = \sqrt{n}\left[\hat{\Lambda}_0(s^*, \cdot; \boldsymbol{\beta}) - \Lambda_0(\cdot)\right]$ converges weakly on the Skorohod space $\mathscr{D}[0, t^*]$*

to a p-variate Gaussian process with mean zero and covariance function

$$\varphi(s^*, t_1, t_2) = \int_0^{min(t_1,t_2)} \frac{\Lambda_0(dw)}{s^{(0)}(s^*, w; \boldsymbol{\beta}_0)} + \xi(s^*, t_1)^{'}[\boldsymbol{\Sigma}(s^*, t^*)]^{-1}\xi(s^*, t_2),$$

for $t_1, t_2 \in [0, t^*]$ *and with* $\xi(s^*, t_1) = \int_0^{t_1} \boldsymbol{e}(s^*, w; \boldsymbol{\beta}_0)\Lambda_0(dw)$.

Proof. From the results of Theorem 3.7, it follows that $\{W_n(s^*, t^*) : t^* \in \mathscr{T}\}$, converges weakly on $\mathscr{D}[0, t^*]$ to the zero-mean Gaussian process $W_\infty(s^*, t^*)$ with

$$W_\infty(s^*, t^*) = \mathbf{Z}_2(s^*, t^*) + \xi(s^*, t_1)^{'}\mathbf{W}_1(s^*, t^*),$$

and its covariance function is

$$\varphi(s^*, t_1, t_2) = \int_0^{min(t_1,t_2)} \frac{\Lambda_0(dw)}{s^{(0)}(s^*, w; \boldsymbol{\beta}_0)} + \xi(s^*, t_1)^{'}[\boldsymbol{\Sigma}(s^*, t^*)]^{-1}\xi(s^*, t_2),$$

for $t_1, t_2 \in [0, t^*]$.

This completes the proof of the corollary. □

3.4. SIMULATION STUDY

Simulation study is carried out to assess the large sample performance of the proposed approach in Section 3.2 and 3.3. The goals of these studies are: (i) to examine the effect of sample size on the properties of the estimators; (ii) to examine the bias, standard deviation, and root-mean-square-error of the estimators; (iii) to compare asymptotic to large sample results with the simulated results. It is interesting to compare the performance of the proposed approach with one designed for the analysis of recurrent event data. In order to accomplish this, another simulation study is conducted under the setup described in Cook and Lawless [24]. The Cook and Lawless (CL) model, an extension of the Cox proportional hazards model with time varying covariates, is a semiparametric analysis of the intensity of the recurrent event. For comparison, we also fit Cook and Lawless model using the function coxph in R.

3.4.1. Simulation Design. In the study, the interoccurrence times T_{ij} are generated from the Weibull distribution with the hazard function $\alpha\eta t^{\alpha-1}e^{\beta' \mathbf{X}_i(t)}$ and also the censoring variables τ_i from uniform distribution over the interval $(0, 360)$. For covariates, we consider a two-dimensional covariate vector (X_1, X_1) with X_1 having a standard normal distribution, and X_2 having Bernoulli distribution with success probability of 0.5. The true values of regression coefficient vector (β_1, β_2) is set to be (1, -1). The results given below are based on sample size n that varies in the set {30, 50, 80, 200} with 1000 replications. The Weibull shape parameter α is set to be 0.5 and 0.8, and parameter η with $\eta \in \{0.1, 0.5, .8, 1\}$. The results include the averages of the point estimates $\hat{\boldsymbol{\beta}}$ (Mean), and the sample standard deviations of the point estimate (SSD).

3.4.2. Discussion of Simulation Results. To find the maximum likelihood estimator of $\boldsymbol{\beta}$, we need to maximize the log-liklihood profile in (3.14) with respect to $\boldsymbol{\beta}$. This is done using mle2 function in R and which is based on the Nelder-Mead method. We show in tables below estimates of $\boldsymbol{\beta}$ that are based on the proposed model and Cook and Lawless model, which are given in the output from R function coxph. Tables 3.1 - 3.3 summarize the mean values, bias, and standard errors of the estimators of β_1 and β_2 for different values of α, η, and sample size n. The numerical results indicate that the proposed methods work well. The tables show that as the sample size increases, the performance of the estimators of $\boldsymbol{\beta}$ improved, with the biases decreasing and the standard errors also decreasing. The impact of changing Weibull parameters in the context of the accuracy and precision of the estimators can be observed because of the interplay between these parameters leads to differing observed number of event occurrences. Upon examining the mean number of events observed per unit μ_{Ev}, we notice that when $\alpha < 0.5$ and $\eta < 1$, the latter leading to a DFR Weibull baseline distribution, there tends to be a smaller number of observed events; whereas when $\alpha > 0.5$, the latter leading to a IFR Weibull baseline distribution, then there tends to be a larger number of observed events than the average of approximately 10 events per unit. Figures 3.1-3.6 show an interplay between the nature of the baseline hazard rate function (IFR/DFR) and the behavior of $\hat{\boldsymbol{\beta}}$. We observed that as the sample size increases, the $\hat{\boldsymbol{\beta}}$s exhibit negative bias. Moreover, both the standard errors and RMSEs of $\hat{\boldsymbol{\beta}}$ decrease steady.

Figures 3.7-3.9 present the simulation results on estimation of the cumulative hazard function Λ based on the simulated data generated under the Weibull distribution with $\alpha = 0.1, 0.5$ and $\eta = 0.1, 0.9$. The results given below are based on sample size n varies in the set {30, 50, 80, 200} and the vector (β_1, β_2) is set to be (1, -1) with 1000 replications. The results include the estimated bias (Bias) and the root mean square error (RMSE) for equally spaced values of t in the set [0,200] by increments of 20. The graphs shown below demonstrate the shapes of various values of α and η. It can be seen that as the time increases, the $\hat{\Lambda}$ exhibits positive bias and a steady increase in RMSE and SE. The shape of graphs in each case appears to be the same.

Table 3.1. Maximum likelihood estimates for $\boldsymbol{\beta} = (1, -1)$

				Proposed model		Cook-Lawless model	
α	η	n	μ_{Ev}	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.5	0.5	30	5.16	1.0316	-1.0247	0.5024	-0.5064
		50	5.21	1.0212	-1.0201	0.5057	-0.4780
		80	5.15	0.9856	-0.9871	0.4869	-0.4805
		200	5.16	0.9951	-0.9896	0.4710	-0.4821
0.5	0.8	30	8.88	1.0276	-1.0214	0.5948	-0.5735
		50	8.80	0.9821	-0.9863	0.5544	-0.5086
		80	8.61	0.9899	-0.9885	0.5269	-0.5436
		200	8.74	0.9960	-0.9924	0.5194	-0.5391
0.5	1	30	11.4	1.0149	-1.0157	0.5182	-0.5846
		50	11.4	0.9866	-0.9857	0.5835	-0.5792
		80	11.4	0.9930	-0.9953	0.5860	-0.5731
		200	11.4	0.9971	-0.9935	0.5548	-0.5763
0.8	0.5	30	17.6	0.9870	-0.9854	0.6682	-0.6496
		50	17.2	0.9892	-0.9871	0.6825	-0.6542
		80	17.7	0.9968	-0.9985	0.6550	-0.6247
		200	11.4	0.9982	-0.9987	0.6148	-0.6463

Table 3.2. Bias, RMSE and standard errors for $\hat{\beta}_1$

			Proposed model			Cook-Lawless model		
α	η	n	$Bias$	$RMSE$	$\hat{\sigma}_{\hat{\beta}_1}$	$Bias$	$RMSE$	$\hat{\sigma}_{\hat{\beta}_1}$
0.5	0.5	30	0.0109	0.0668	0.0804	-0.496	0.5423	0.0932
		50	-0.0029	0.0655	0.0406	-0.495	0.5247	0.0694
		80	-0.0044	0.0628	0.0211	-0.514	0.5316	0.0536
		200	-0.0077	0.0622	0.0062	-0.529	0.5368	0.0332
0.5	0.8	30	0.0103	0.0598	0.0393	-0.406	0.4698	0.0699
		50	0.0046	0.0577	0.0147	-0.446	0.4799	0.0514
		80	-0.0066	0.0526	0.0084	-0.474	0.4904	0.0401
		200	-0.0187	0.0525	0.0031	-0.481	0.4876	0.0244
0.5	1	30	0.0140	0.0547	0.0244	-0.411	0.4649	0.0595
		50	0.0029	0.0511	0.0108	-0.417	0.4568	0.0445
		80	-0.0058	0.0464	0.0054	-0.414	0.4375	0.0347
		200	-0.0221	0.0478	0.0015	-0.446	0.4536	0.0209
0.8	0.5	30	0.0019	0.0429	0.0117	-0.232	0.2775	0.0489
		50	-0.0091	0.0400	0.0043	-0.218	0.2713	0.0369
		80	-0.0165	0.0372	0.0034	-0.245	0.2744	0.0277
		200	-0.0229	0.0370	0.0008	-0.285	0.2976	0.0166

Table 3.3. Bias, RMSE and standard errors for $\hat{\beta}_2$

			Proposed model			Cook-Lawless model		
α	η	n	$Bias$	$RMSE$	$\hat{\sigma}_{\hat{\beta}_2}$	$Bias$	$RMSE$	$\hat{\sigma}_{\hat{\beta}_2}$
0.5	0.5	30	0.0090	0.0705	0.0804	-0.494	0.6065	0.0932
		50	-0.0038	0.0697	0.0406	-0.522	0.5782	0.0694
		80	-0.0041	0.0673	0.0211	-0.520	0.5517	0.0536
		200	-0.0047	0.0685	0.0062	-0.518	0.5318	0.0332
0.5	0.8	30	0.0067	0.0620	0.0393	-0.427	0.5458	0.0699
		50	0.0007	0.0603	0.0147	-0.431	0.4943	0.0514
		80	-0.0043	0.0565	0.0084	-0.457	0.5090	0.0401
		200	-0.0152	0.0556	0.0031	-0.461	0.4777	0.0244
0.5	1	30	0.0107	0.0561	0.0244	-0.418	0.5338	0.0595
		50	0.0038	0.0545	0.0108	-0.391	0.4709	0.0445
		80	-0.0047	0.0508	0.0054	-0.427	0.4746	0.0347
		200	-0.0205	0.0503	0.0015	-0.424	0.4464	0.0209
0.8	0.5	30	-0.0013	0.0450	0.0117	-0.251	0.3651	0.0489
		50	-0.0082	0.0438	0.0043	-0.246	0.3236	0.0369
		80	-0.0165	0.0384	0.0034	-0.276	0.3252	0.0277
		200	-0.0223	0.0398	0.0008	-0.254	0.2785	0.0166

3.5. APPLICATION TO REAL DATA

3.5.1. Data Description and Model Assumptions. To illustrate the model, we examine the data set of failure times for the hydraulic subsystems of load-haul-dump (LHD) machines that is given in Kumar [42]. Kumar [42] state that these machines are used “to pick up ore or waste rock from the mining points and for dumping it into trucks or ore passes” and that preliminary studies indicate that the engine and hydraulics are “the two most critical subsystems.” Additional investigation of the hydraulic subsystems was carried out by Kumar [42] due to the fact that they were in a developmental phase. They analyzed two years of maintenance data for these subsystems using a power law process model.

The data set consists of the failure times (in hours), excluding repair or down times, of six different machines. The machines are categorized based on their age with the first two being old, the next two being medium old, and the last two being new machines. To account for the differences in ages of the machines, we define an age covariate vector $\mathbf{X}$ as follows: $\mathbf{X} = (0, 0)$ represents old age; $\mathbf{X} = (1, 0)$ represents medium old age; and $\mathbf{X} = (0, 1)$ represents new machines. Since censoring information was not provided by Kumar [42], we set $\tau_i = S_{i,K_i}$ for all i. Information regarding the types of repairs performed was not given by Kumar [42]. In practice, this often leads to the researcher having to choose between “always perfect repair” ($\mathcal{E}_i(s) = s - S_{i,N_i^\dagger(s-)}$) or “always minimal repair” ($\mathcal{E}_i(s) = s$) effective age processes. For illustration purposes, we present the analyses for both choices and later discuss the appropriateness of these selections.

3.5.2. Analysis Results. Parameter estimates along with their associated standard errors and 95% confidence intervals are presented in Table 3.4. The main difference in the parameter estimates for the two different models is for β_1. In both models, the 95% confidence intervals indicate that the parameters associated with

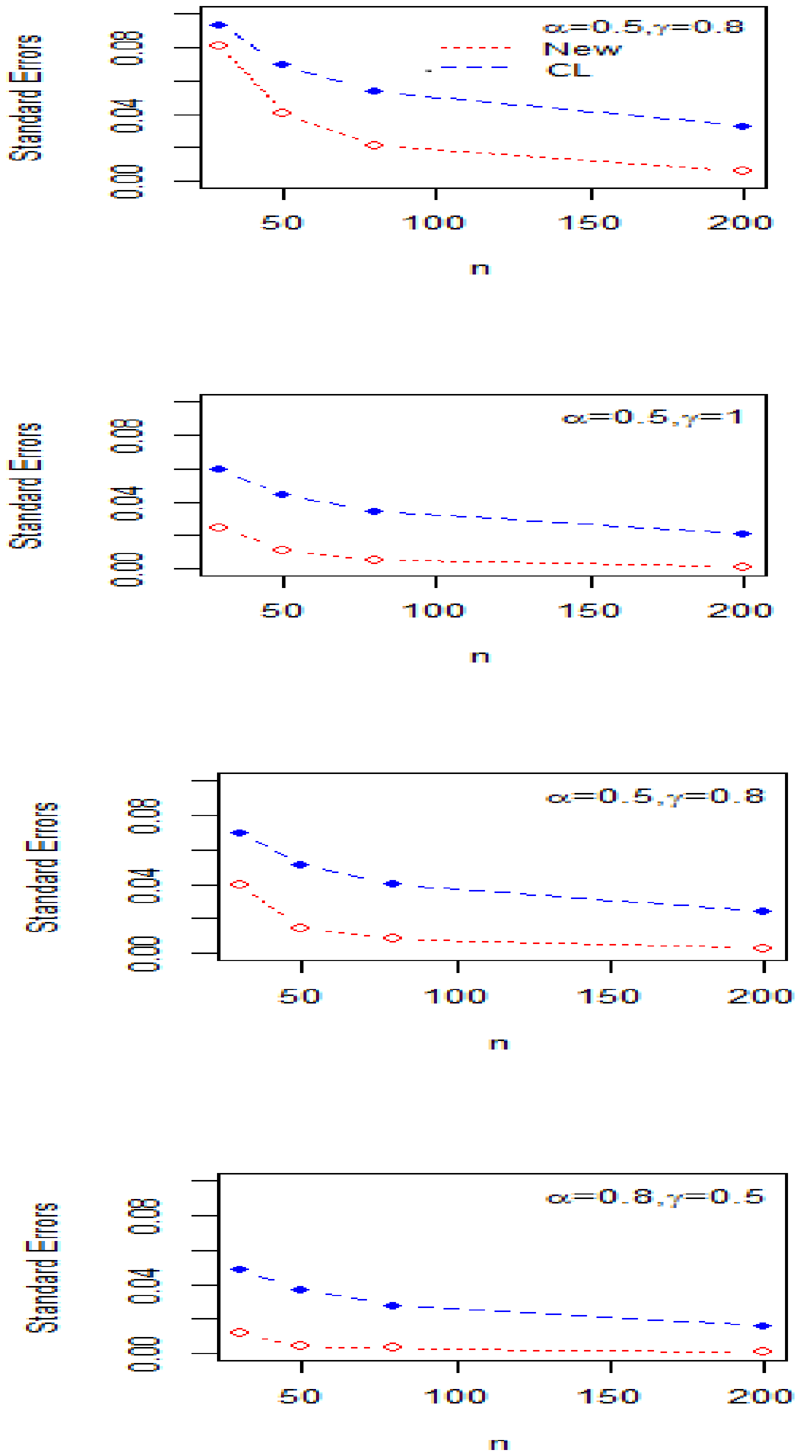
Table 3.4. Estimates of β_1 and β_2; the standard errors of the parameter estimates denoted as $\hat{\sigma}(\hat{\beta}_i)$; and 95% confidence intervals (CI) for β_1 and β_2. We assume that the interventions result in either an "always perfect repair" or "always minimal repair."

	Always Perfect Repair			Always Minimal Repair		
i	$\hat{\beta}_i$	$\hat{\sigma}(\hat{\beta}_i)$	95% CI	$\hat{\beta}_i$	$\hat{\sigma}(\hat{\beta}_i)$	95% CI
1	-0.0384	0.1996	(-.4296, .3528)	-0.1572	0.2094	(-.5676, .2532)
2	-0.0471	0.2054	(-.4497, .3555)	-0.0543	0.2051	(-.4563, .3477)

age are not statistically significant. Additionally, likelihood ratio tests of the null hypothesis, $H_0 : \beta_1 = \beta_2 = 0$, were performed to assess if there were statistically significant differences in the survival of the subsystems by age of the machines. These tests result in p-values of .9711 and .7523 for the "always perfect repair" and "always minimal repair" models respectively; indicating a lack of evidence to conclude that survival differs by age. 3.4 presents estimates of the survivor functions by age for each model using the expression given by

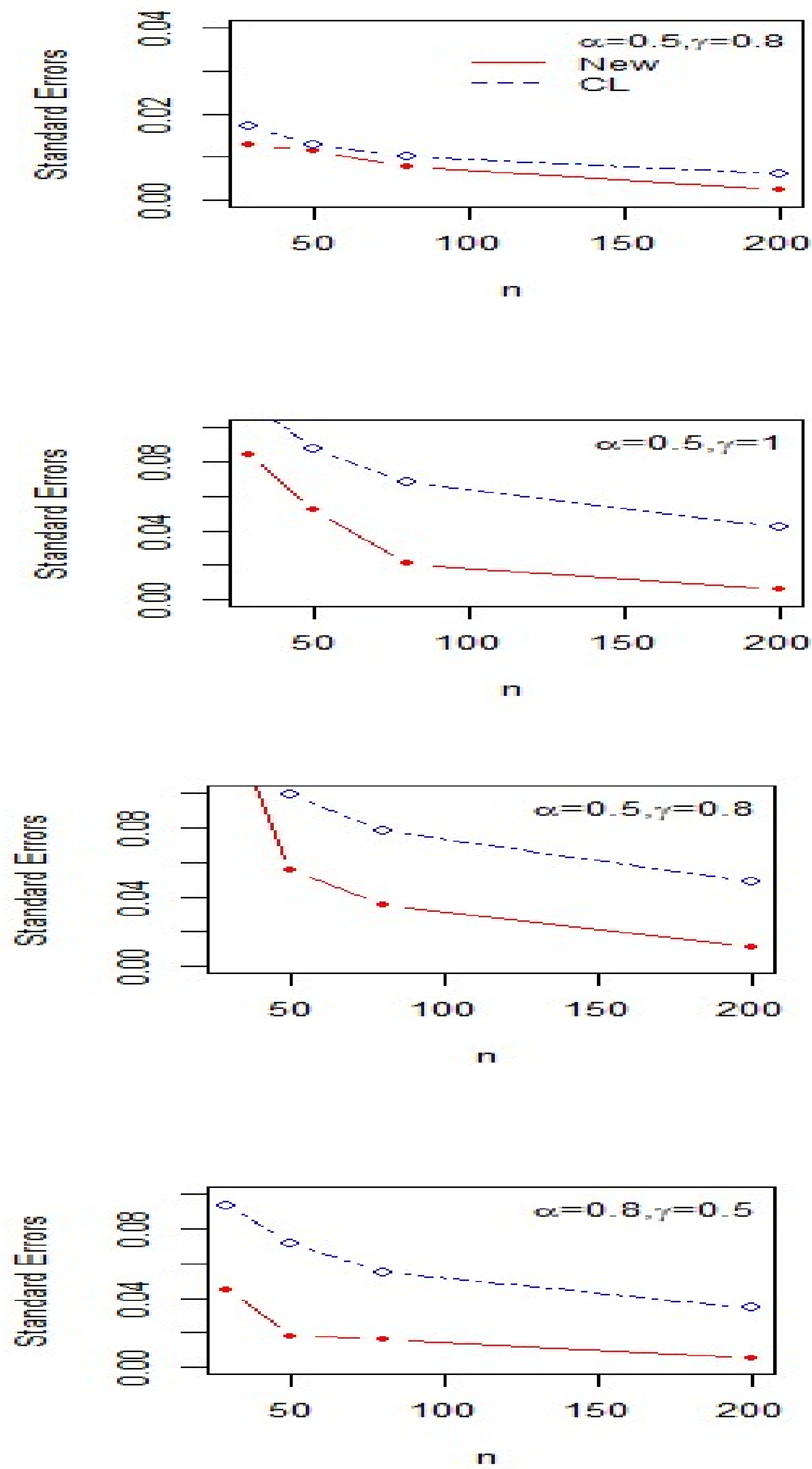
$$\hat{\bar{F}}_0(t|\hat{\beta}_1, \hat{\beta}_2)^{\exp(\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2)}.$$

These also indicate that survival does not differ by age in either model.



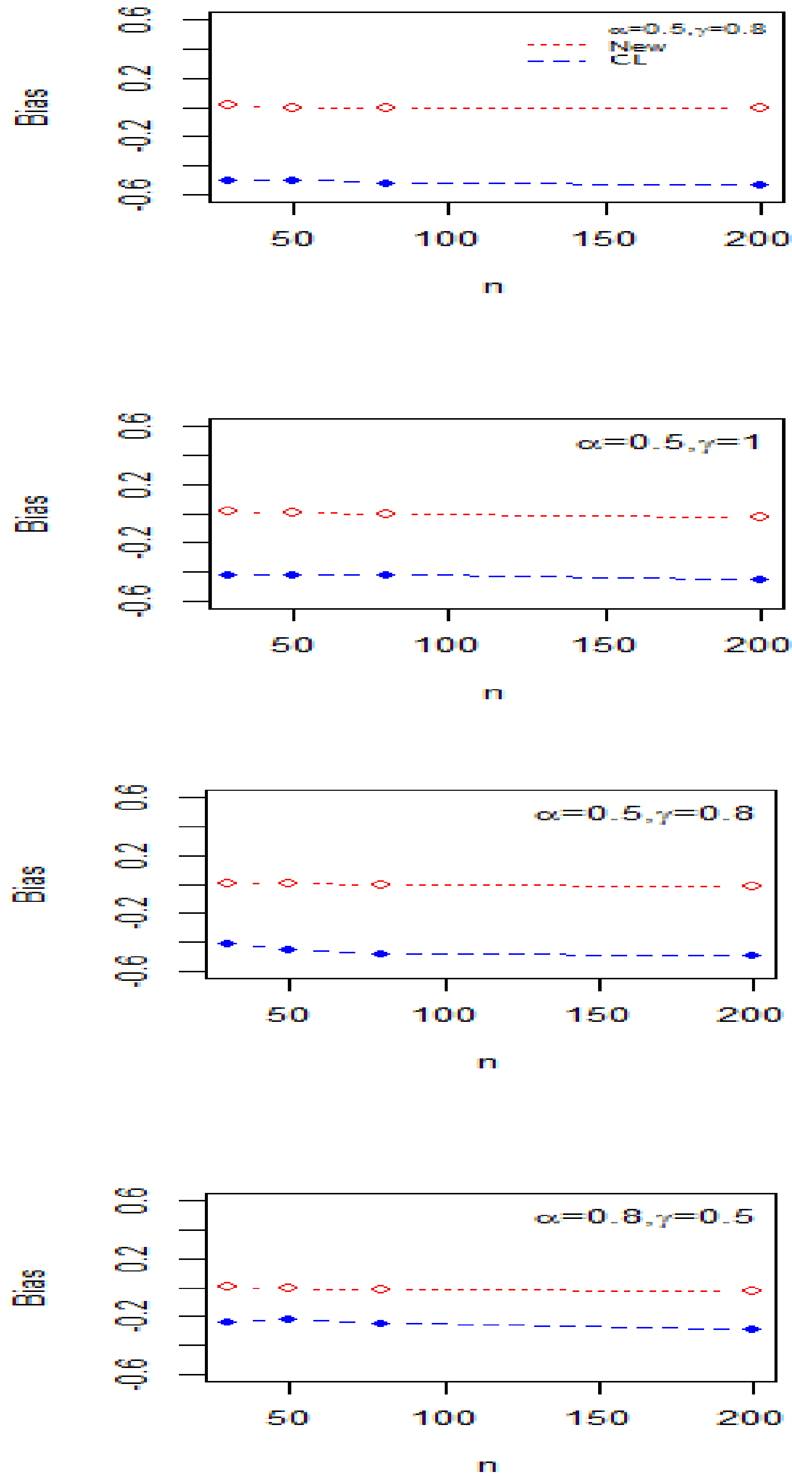
.

Figure 3.1. Graph of standard errors for the estimator of β_1 for different values of n, α and η



.

Figure 3.2. Graph of standard errors for the estimator of β_2 for different values of n, α and η



.

Figure 3.3. Biases in the estimator of β_1 for different values of n, α and η

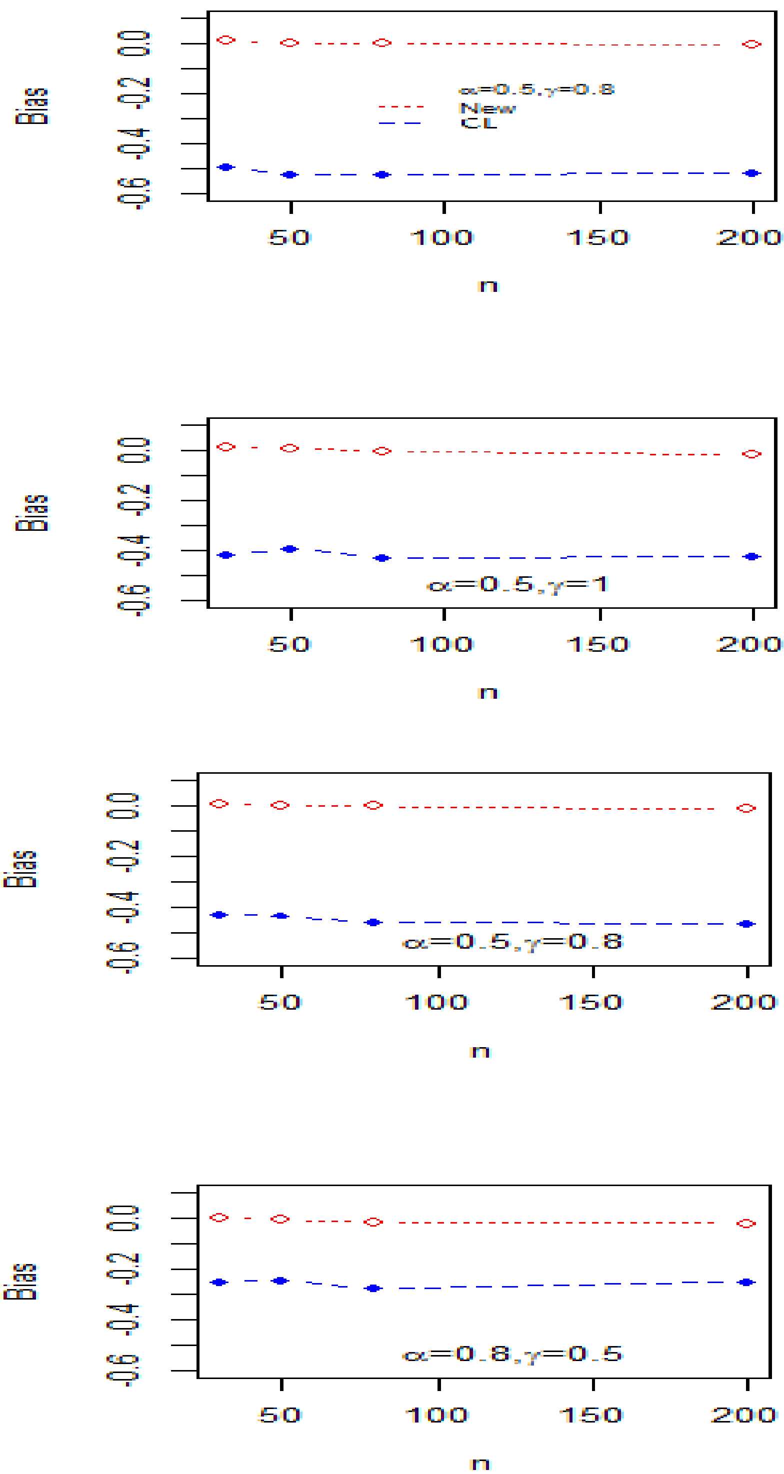
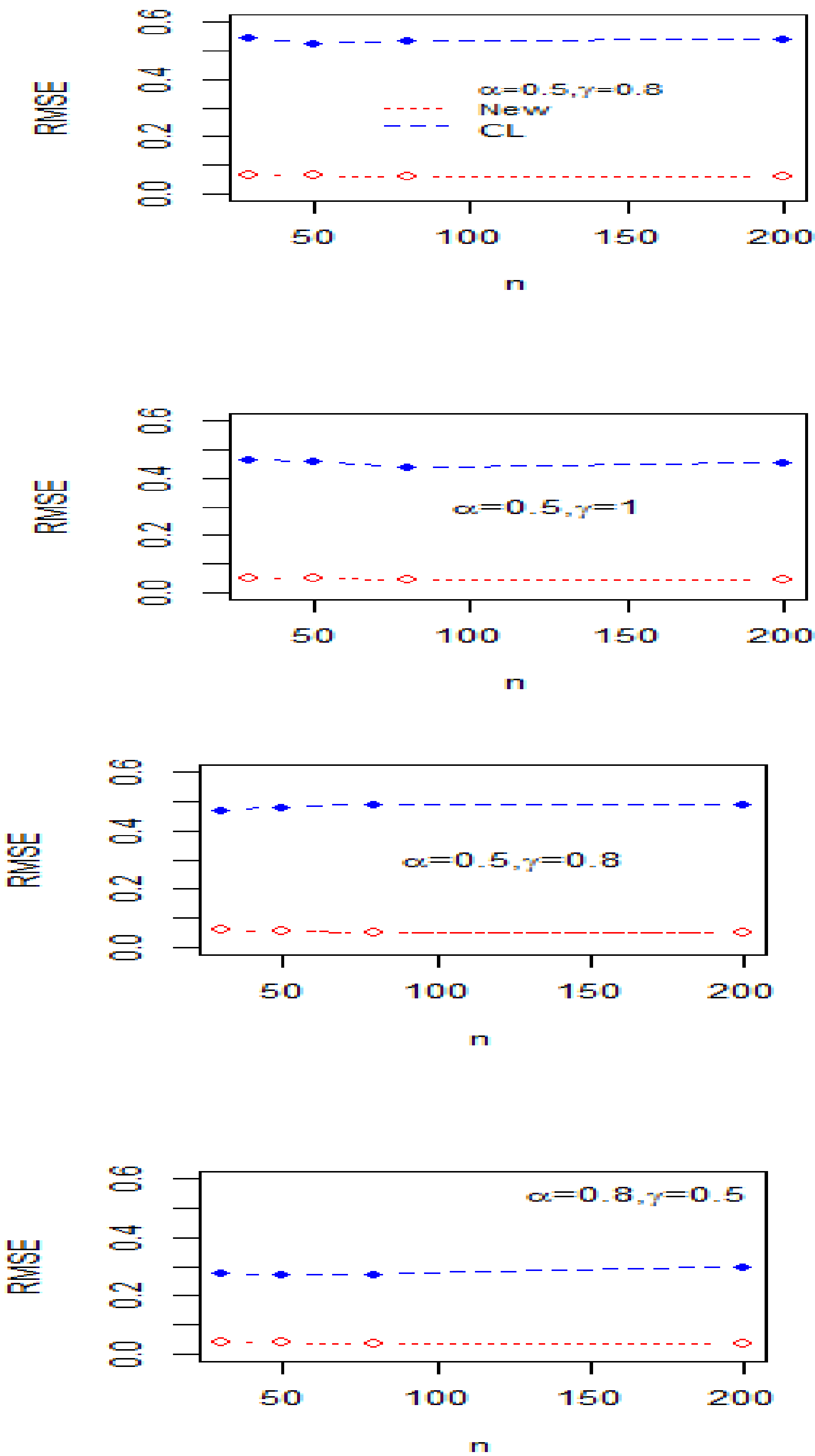
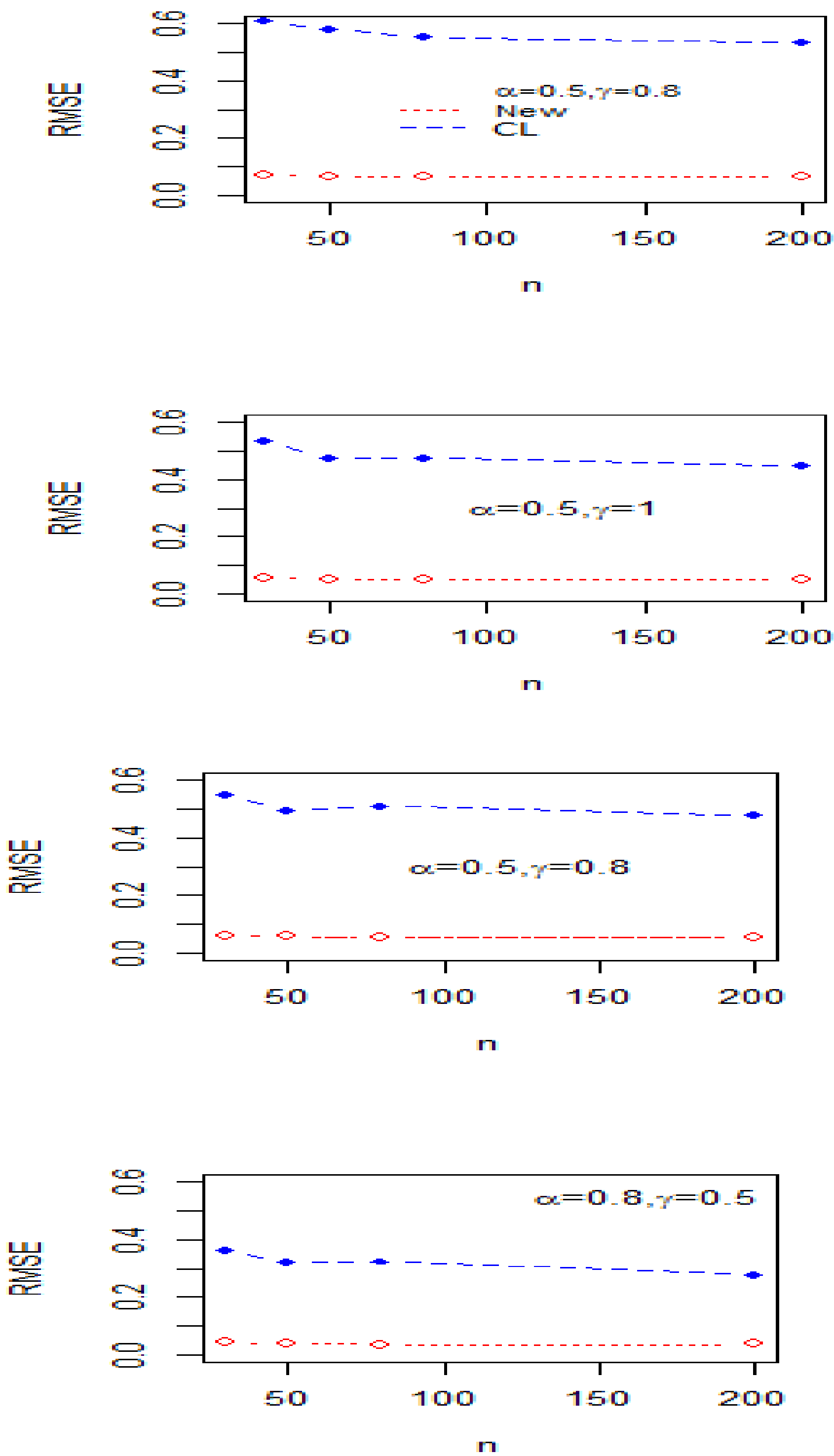


Figure 3.4. Biases in the estimator of β_2 for different values of n, α and η



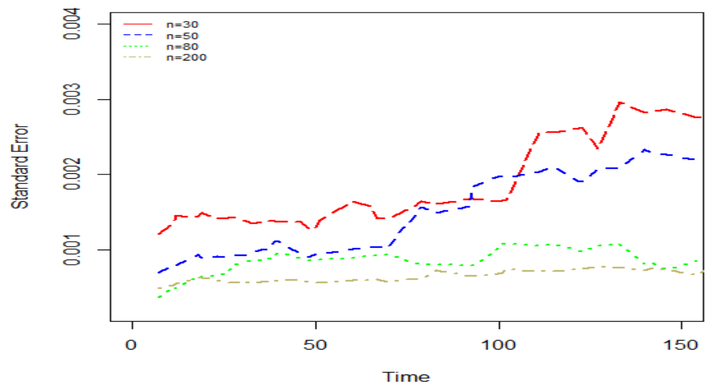
.

Figure 3.5. RMSE for the estimator of β_1 for different values of n, α and η

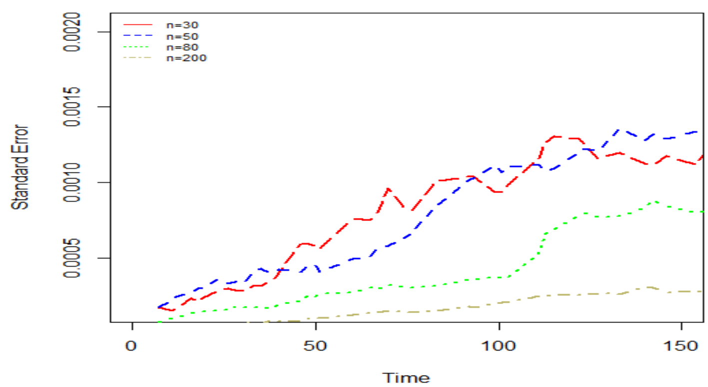


.

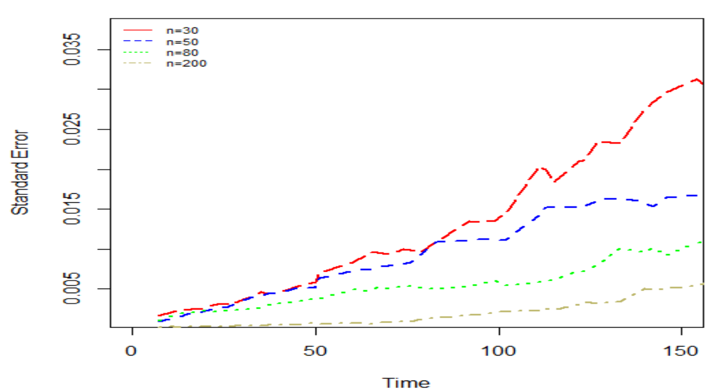
Figure 3.6. RMSE for the estimator of β_2 for different values of n, α and η



(a) $\alpha = 0.1$ and $\eta = 0.9$

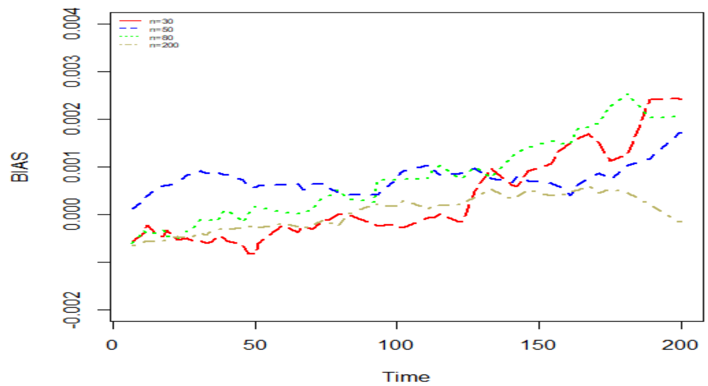


(b) $\alpha = 0.5$ and $\eta = 0.1$

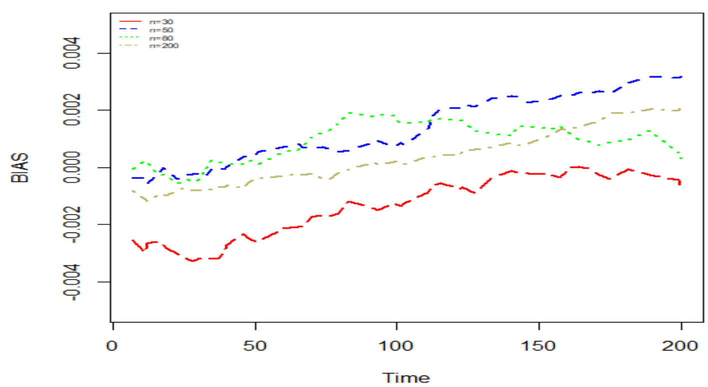


(c) $\alpha = 0.5$ and $\eta = 0.5$

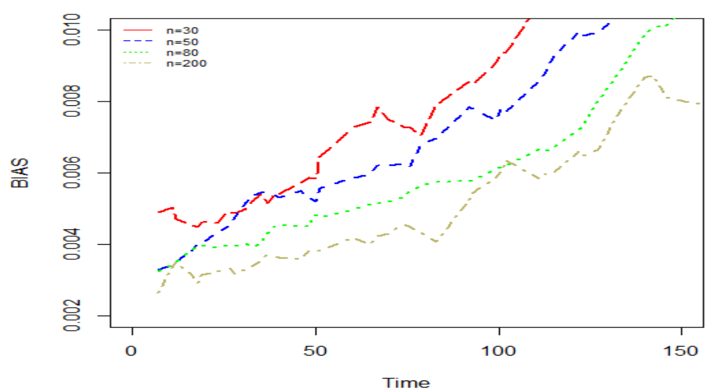
Figure 3.7. Graph of standard errors for the estimator of Λ_0 for different values of n, α and η



(a) $\alpha = 0.1$ and $\eta = 0.9$

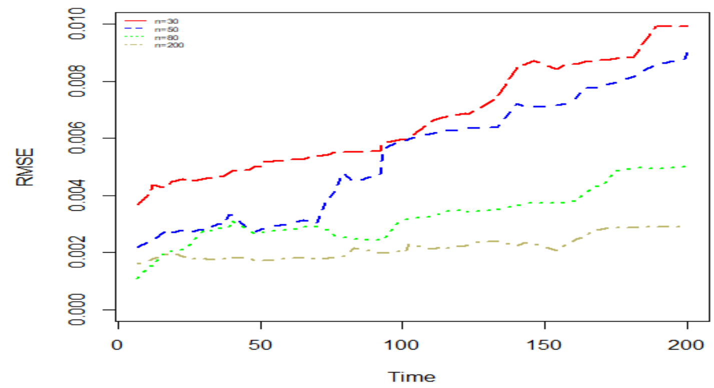


(b) $\alpha = 0.5$ and $\eta = 0.1$

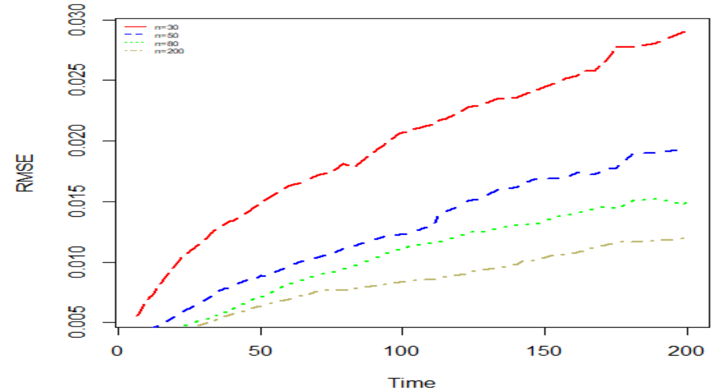


(c) $\alpha = 0.5$ and $\eta = 0.5$

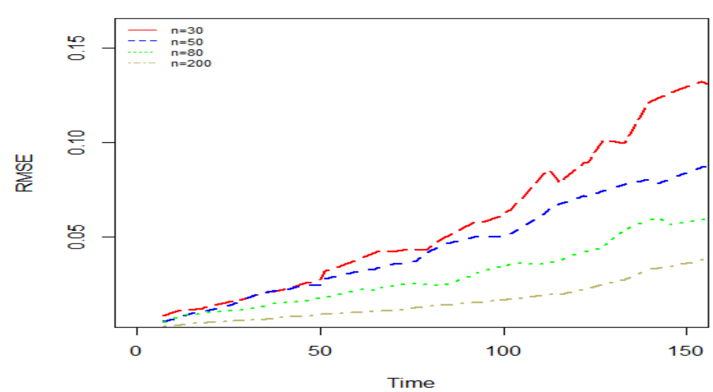
Figure 3.8. Graph of bias for $\hat{\Lambda}$ for different values of n, α and η



(a) $\alpha = 0.1$ and $\eta = 0.9$



(b) $\alpha = 0.5$ and $\eta = 0.1$



(c) $\alpha = 0.5$ and $\eta = 0.5$

Figure 3.9. Graph of RMSE for $\hat{\Lambda}$ for different values of n, α and η

4. MODEL CHECKING

In the first part of this work, we have assumed that the time to failure as a function of the covariates follows the Cox's model $\lambda(s|\mathbf{x}) = \lambda_0(\mathcal{E}_i(s)) \exp(\boldsymbol{\beta}^t \mathbf{x}_i(s))$ for recurrent event. This model assume the link function between the hazard function at time s is link to the baseline hazard function via an exponential link function $\exp(\boldsymbol{\beta}' \mathbf{x}_i(s))$. More, it assumes that the form of the covariates in the exponential is linear, and that the model is proportional hazards. At least one of these assumptions can fail. And, if any of the assumption does not hold, that can lead to inaccurate estimators thereby inaccurate inferential properties. Worst, wrong conclusions will be drawn from that data leading to devastating consequences, especially in biomedical studies where the model is often applied. To avoid such detrimental consequences, appropriate model checking procedures need to be developed to check models accuracy before being applied to any given dataset.

To that end, many techniques have been proposed to deal with the issue. Those techniques can be graphical, and a good reference summarizing various graphical techniques is Liu [43], and Wei [44]. Others include Margaret Sullivan Pepe and Jian Wen Cai [45], and Gill and Schumacher [46]. The graphical technique is used to check the global validation. Specific decision rules for checking any of the assumption are also provided in the literature. For instance, to check the functional forms of the covariates, Parzen and Lipsitz [47] proposed a global goodness of fit test that follows a chi-square distribution. Lin, Wei, and Ying [43, 48, 49] proposed various procedures for checking all the assumptions of the model based on martingale residuals. Martingales are similar to models errors in regression. Specifically, for the single event for instance, and if the Cox model is assumed, let $N(s)$ be the number of events occurrences by time s. The compensator of $N(s)$ is $A(s) = \int_0^s Y(u)\lambda_0(u) \exp(\boldsymbol{\beta}' \mathbf{x}) du$

making $M(s) = N(s) - A(s)$ a zero-mean martingale with respect to the filtration $\mathfrak{F}_s = \sigma\{N(u), Y(u), u \leq s\}$. If $\hat{\boldsymbol{\beta}}$ is the regression parameter estimator obtained via partial likelihood, the martingale residuals, n of them, are defined by, for unit i

$$\hat{M}_i(s) = N_i(s) - \int_0^s Y(u)\lambda_0(u|\hat{\boldsymbol{\beta}}) \exp(\hat{\boldsymbol{\beta}}'\mathbf{x})du,$$

$i = 1, ..., n$, where the estimator of the baseline hazard function is given by

$$\lambda_0(u|\hat{\boldsymbol{\beta}}) = \int_0^s \frac{d\bar{N}(u)}{\sum_{i=1}^n Y_i(u)} \exp(\hat{\boldsymbol{\beta}}'\mathbf{x}).$$

The martingale residuals are interpreted as the difference at time s between the observed number of events and the expected number of events which is estimated by the estimated cumulative hazard given by

$$\hat{A}(s) = \int_0^s Y(u)\lambda_0(u|\hat{\boldsymbol{\beta}}) \exp(\hat{\boldsymbol{\beta}}'\mathbf{x})du.$$

Moreover, their properties are similar to those of regression models. That is, if the Cox's model is valid, then $\sum_i \hat{M}_i(s) \approx 0$ and are uncorrelated. The martingales residuals can be transformed to develop test statistics for checking the validity of all the assumptions of the Cox models thereby assessing models departure, cf. Schoenfeld [50], Barlow and Prentice [51].

4.1. GRAPHICAL METHODS

Informal graphical procedures such as martingale residual plots are useful tools for checking the fit of the regression model for recurrent event data. To illustrate how the plots of the residual processes can be used to check the proposed model, we consider the residual processes for individual counting processes,that is

$$M_{ij}(s,t) = N_{ij}(s,t) - \int_0^t Y_{ij}(s,w)\psi(\boldsymbol{\beta}'\mathbf{X}_{ij})\Lambda_0(dw),\ j = 1,2,...,i = 1,...,n. \quad (4.1)$$

In general, we shall consider the cumulative intensity process of the form

$$\Lambda_{ij}(s,t;\boldsymbol{\beta}) = \int_0^t Y_{ij}(s,w)\psi(\boldsymbol{\beta}'\mathbf{X}_{ij})\Lambda_0(dw), \quad (4.2)$$

and it may be estimated by

$$\Lambda_{ij}(s,t;\hat{\boldsymbol{\beta}}) = \int_0^t Y_{ij}(s,w)\psi(\hat{\boldsymbol{\beta}}'\mathbf{X}_{ij})\Lambda_0(dw), \quad (4.3)$$

For the link function $\psi(\hat{\boldsymbol{\beta}}'\mathbf{X}_{ij})$, we have $\hat{\eta}_{ij} = \hat{\boldsymbol{\beta}}'\mathbf{X}_{ij}(t) = g(\Lambda_{ij}(s,t;\hat{\boldsymbol{\beta}})) = g(\Lambda_{ij})$. Now, let Z_{ij} be a linearized form of the link function applied to the data

$$\begin{aligned} Z_{ij}(t) &= g(\Lambda_{ij}) + (N_{ij}(s,t) - \Lambda_{ij}(s,t;\hat{\boldsymbol{\beta}}))g'(\Lambda_{ij}) \\ &= \hat{\eta}_{ij}(t) + M_{ij}(s,t;\hat{\boldsymbol{\beta}})\frac{d\hat{\eta}_{ij}}{d\Lambda_{ij}}. \end{aligned} \quad (4.4)$$

It follows by (4.4) that the aggregated process $Z_{ij}(t)$ is given by

$$Z(t) = \sum\nolimits_{j,i} Z_{ij}(t) = \sum\nolimits_{j,i}\left(\hat{\eta}_{ij}(t) + M_{ij}(s,t;\hat{\boldsymbol{\beta}})\frac{d\hat{\eta}_{ij}(t)}{d\Lambda_{ij}}\right) \quad (4.5)$$

4.1.1. Checking the Link Function. A plot of the points $(t, \hat{\eta}(t), Z(t))$ provides an informal check for the appropriateness of link function. If the link function is appropriate the plot should be approximately plane.

4.1.2. Checking the Covariates Functional Form. the partial residual plot is a useful and important a graphical technique for checking the covariates functional form. The partial residual is defined by

$$R(t) = Z(t) - \hat{\eta}_{ij}(t) + \hat{\gamma}x(t), \tag{4.6}$$

where $\hat{\gamma}$ the parameter estimate for $x(t)$. Thus a plot of R versus t and $x(t)$ should be roughly plane if the the covariates functional form is appropriate.

4.2. TESTS BASED ON MARTINGALES RESIDUALS

In classical linear models residuals are often useful when assessing goodness of fit of a given model. Assume that Y is integrable, so that

$$m(x) = E(Y|\mathbf{X} = \mathbf{x}) \qquad \mathbf{x} \in \mathbb{R}^d$$

is the regression function of Y on $\mathbf{X}$. In the problem of interest assumes m is a member of a parametric family of link functions defined on the real line

$$\mathcal{M} = \{m(.,\theta) : \theta \in \Theta \subset \mathbb{R}^p\}$$

and given independent observations $(X_i, Y_i), 1 \leq i \leq n$ such that the errors

$$\varepsilon_i = Y_i - m(\boldsymbol{\beta}'\mathbf{X}, \theta), 1 \leq i \leq n, \tag{4.7}$$

are independent, identically distributed r.v.'s with

$$E(\varepsilon|X) = 0 = E(\varepsilon|\boldsymbol{\beta}'\mathbf{X}) \qquad \mathbf{x} \in \mathbb{R}^d.$$

It is then of interest to know whether m belongs to $\mathcal{M}$ or not; that is, by a test for

$$H_0 : m(x) = m(\boldsymbol{\beta_0'}\mathbf{x}, \theta_0), \qquad for\, some\, \boldsymbol{\beta}_0 \in \mathbb{R}^d\, and\, \theta_0 \in \Theta \subset \mathbb{R}^p.$$

To begin with, pose we want to check whether the proposed Cox-model is adequate. Introduce the

$$\lambda_i(s) = \lambda_0(R_i(s))\exp(\boldsymbol{\beta}'\mathbf{X}_i(s)), \quad s \geq 0, i = 1, ..., n$$

where $\lambda_0(.)$ is the baseline hazard function and $\boldsymbol{\beta}$ is a q-dimensional vector of unknown regression coefficients, which under the model assumption is equal to $\lambda_i(s, \theta)$ for a certain value of the parameter. Residuals can be defined via the basic counting process decomposition where

$$M_i(., t; \boldsymbol{\beta}) = N_i(., t) - \int_0^t Y_i(., w; \boldsymbol{\beta})\Lambda_0(dw) \qquad i = 1, ..., n \tag{4.8}$$

are local square integrable martingale with respect to $\mathscr{F}_s$. Inserting the estimated parameter values into the compensator, we get the martingale residual process

$$\hat{M}_i(s, t) = N_i(s, t) - \int_0^t Y_i(s, w; \hat{\boldsymbol{\beta}})\hat{\Lambda}_0(dw), \tag{4.9}$$

where the estimated baseline cumulative hazard function is given by the Breslow estimator

$$\hat{\Lambda}_0(s, t; \hat{\boldsymbol{\beta}}) = \frac{1}{n}\left\{\int_0^t \frac{J(s, w; \hat{\boldsymbol{\beta}})}{S^{(0)}(s, w; \hat{\boldsymbol{\beta}})} N(s, dw)\right\}. \tag{4.10}$$

The martingale residual process is constructed based on calender and gap times which differs from the martingale residual process considered by Lin et al. [25]. Numerous techniques have been proposed for checking the adequacy of Cox model. In his original paper, Cox [52] proposed dummy time-varying covariates for model checking.

Key [53], Cox [54], and Kalbfeisch and Prentice [55, 56] proposed various graphical methods for checking the Cox model, but they are all subjective. Lin and Wei [57] extended the work of White [58] on an information matrix type of test. Lin et al. [48] proposed model checking based on cumulative sums of martingale-based residuals. Lin et al. [25] proposed a test for model checking for recurrent events whose asymptotic distribution can be approximated by a weighted sum of zero-mean Gaussian processes. Furthermore, it behaves like martingale type residuals. Huang et al. [59] proposed a model checking method for recurrent events, but the argument in Cox-model does not account for time elapsed.

To construct proper tests for checking the functional form of the covariates and the link function. In particular, we shall consider statistics of Kolmogorov-Smirnov type. We consider the time-varying covariates, the maximum likihood estimators, and the hypothesis

$$H_0 : F(K; \boldsymbol{\beta}) = F_0(K; \boldsymbol{\beta}) \qquad for\, some\, \boldsymbol{\beta} \in \mathbb{R}^p,$$

where K is a martingale residual process . The martingale residual process is then

$$K = K_n(t, u) = K_n(s, t, u; \boldsymbol{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} I(\boldsymbol{\beta}' \mathbf{X}_i(s) \leq u) M_i(s, t). \tag{4.11}$$

The tests based on a certain marked empirical or partial sum processes have been discussed by An and Cheng [60], Stute [61], Stute, González Manteiga and Presedo Quindimil [62], Stute, Thies and Zhu [62], Stute and Zhu [63]. We shall show in Section 2 that the proposed martingale residual processes converges in distribution to a certain zero-mean Gaussian process.

4.3. DISTRIBUTION OF TESTS STATISTICS

This section provides limit distribution results for the two classes of stochastic processes described in section 1. It is assumed that the parameter is unspecified. We consider an approach based on estimates $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$. Put

$$
\begin{aligned}
\hat{K}_n(t,u) &= K_n(s,t;u,\hat{\boldsymbol{\beta}}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} I(\hat{\boldsymbol{\beta}}'\mathbf{X}_i(s) \leq u)\hat{M}_i(s,t) \\
&= n^{-\frac{1}{2}} \sum_{i=1}^{n} I(\hat{\boldsymbol{\beta}}'\mathbf{X}_i(s) \leq u)N_i(s,t) - \int_0^t \frac{I(\hat{\boldsymbol{\beta}}'\mathbf{X}_i(s) \leq u)Y_i(s,w;\hat{\boldsymbol{\beta}})}{nS^{(0)}(s,w;\hat{\boldsymbol{\beta}})} N(s,dw)
\end{aligned}
\tag{4.12}
$$

In order to study the asymptotic distribution of $\hat{K}_n(t,u)$, we shall assume that Condition 1 (a-f) are fulfilled. We shall also throw in the following
Assumption g: There exists a deterministic function $y : \mathscr{T} \times \mathscr{B} \rightarrow \mathbb{R}_+$ which is continuous in (s,t) and bounded and such that

$$
\max_{1 \leq i \leq n} \sup_{t \in \mathscr{T}; \beta \in \mathscr{B}} \left| \frac{Y_i(s,t;\boldsymbol{\beta})}{n} - y(s,t;\boldsymbol{\beta}) \right| \xrightarrow{p} 0,
$$

Assumption h: Under H_0, that is $\lambda(x) = \lambda(\boldsymbol{\beta}_{\mathbf{0}}'\mathbf{x})$, for some $\boldsymbol{\beta}_0 \in \mathbb{R}^d$, we have

$$
\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &= \boldsymbol{\Sigma}(s,t)^{-1} \times \\
&\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t [\mathbf{X}_i(s) - \mathbf{E}(s,w;\boldsymbol{\beta}_0)] M_i(s,dw) \right\} + o_p(1),
\end{aligned}
\tag{4.13}
$$

where $\boldsymbol{\Sigma}(s,t)$ is given by part f of Condition 1.

Theorem 4.1. *Under H_0, assume that the assumptions a-h are satisfied, we have*

$K_n \longrightarrow_d K$ in the Skorohod space $D[-\infty,\infty]$,

where K is a zero-mean Gaussian process with covariance function

$$n^{\frac{1}{2}} \sum_{i=1}^{n} \int_0^t I(\boldsymbol{\beta}_{\mathbf{0}}' \boldsymbol{X}_i(s) \leq u)\, y_i(s,w;\boldsymbol{\beta}_{\mathbf{0}}) \Lambda_0(dw)$$
$$- \left\{ \sum_{i=1}^{n} \int_0^t I(\boldsymbol{\beta}_{\mathbf{0}}' \boldsymbol{X}_i(s) \leq u) \boldsymbol{X}_i(s) y_i(s,w;\boldsymbol{\beta}_{\mathbf{0}}) \Lambda_0(dw) \right\}' \boldsymbol{\Sigma}(s,t)^{-1} \tag{4.14}$$
$$\times\, n^{\frac{1}{2}} \sum_{i=1}^{n} \int_0^t [\boldsymbol{X}_i(s) - \boldsymbol{e}(s,w;\boldsymbol{\beta}_0)]^{\otimes 2} y_i(s,w;\boldsymbol{\beta}_{\mathbf{0}}) \Lambda_0(dw).$$

Proof. Taylor-expanding $K_n(t,u;\boldsymbol{\beta})$ around $\boldsymbol{\beta}_0$ gives,

$$\begin{aligned} K_n(t,u) = K_n(s,t;u,\boldsymbol{\beta}) &= n^{-\frac{1}{2}} \sum_{i=1}^{n} I(\boldsymbol{\beta}_{\mathbf{0}}' \mathbf{X}_i(s) \leq u) M_i(s,t) \\ &- (\boldsymbol{\beta} - \boldsymbol{\beta}_{\mathbf{0}})' n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^t I(\boldsymbol{\beta}^{*\prime} \mathbf{X}_i(s) \leq u) \mathbf{X}_i(s) Y_i(s,w;\boldsymbol{\beta}^*) \Lambda_0(dw), \end{aligned} \tag{4.15}$$

where $\boldsymbol{\beta}^*$ is on the line segment between $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_{\mathbf{0}}$. Insert $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ to get

$$\begin{aligned} K_n(t,u) = n^{-\frac{1}{2}} &\sum_{i=1}^{n} I(\boldsymbol{\beta}_{\mathbf{0}}' \mathbf{X}_i(s) \leq u) M_i(s,t) \\ &- n^{-\frac{1}{2}} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\mathbf{0}})' \sum_{i=1}^{n} \int_0^t I(\boldsymbol{\beta}^{*\prime} \mathbf{X}_i(s) \leq u) \mathbf{X}_i Y_i(s,w;\boldsymbol{\beta}^*) \Lambda_0(dw). \end{aligned} \tag{4.16}$$

By Assumption h, it follow that

$$\begin{aligned} K_n(t,u) = n^{-\frac{1}{2}} &\sum_{i=1}^{n} I(\boldsymbol{\beta}_{\mathbf{0}}' \mathbf{X}_i(s) \leq u) M_i(s,t) \\ &- n^{-\frac{1}{2}} \left\{ \sum_{i=1}^{n} \int_0^t I(\boldsymbol{\beta}^{*\prime} \mathbf{X}_i(s) \leq u) \mathbf{X}_i(s) Y_i(s,w;\boldsymbol{\beta}^*) \Lambda_0(dw) \right\}' \boldsymbol{\Sigma}(s,t)^{-1} \\ &\times n^{-1} \sum_{i=1}^{n} \int_0^t [\mathbf{X}_i(s) - \mathbf{E}(s,w;\boldsymbol{\beta}_0)] M_i(s,dw), \end{aligned} \tag{4.17}$$

for any $\boldsymbol{\beta}^*$ such that $\boldsymbol{\beta}^* \xrightarrow{p} \boldsymbol{\beta}_{\mathbf{0}}$, we note that

$$\begin{aligned} K_n(t,u) &= n^{-\frac{1}{2}} \sum_{i=1}^{n} I(\boldsymbol{\beta}_{\mathbf{0}}' \mathbf{X}_i(s) \le u) M_i(s,t) \\ &\quad - n^{-\frac{1}{2}} \left\{ \sum_{i=1}^{n} \int_0^t I(\boldsymbol{\beta}_{\mathbf{0}}' \mathbf{X}_i(s) \le u) \mathbf{X}_i(s) Y_i(s,w;\boldsymbol{\beta}_{\mathbf{0}}) \Lambda_0(dw) \right\}' \boldsymbol{\Sigma}(s,t)^{-1} \\ &\quad \times n^{-1} \sum_{i=1}^{n} \int_0^t [\mathbf{X}_i(s) - \mathbf{E}(s,w;\boldsymbol{\beta}_0)] M_i(s,dw). \end{aligned} \tag{4.18}$$

The predictable variation process of this martingale is

$$\begin{aligned} \langle K_n(t,u) \rangle &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^t I(\boldsymbol{\beta}_{\mathbf{0}}' \mathbf{X}_i(s) \le u) Y_i(s,w;\boldsymbol{\beta}_{\mathbf{0}}) \Lambda_0(dw) \\ &\quad - n^{-\frac{1}{2}} \left\{ \sum_{i=1}^{n} \int_0^t I(\boldsymbol{\beta}_{\mathbf{0}}' \mathbf{X}_i(s) \le u) \mathbf{X}_i(s) Y_i(s,w;\boldsymbol{\beta}_{\mathbf{0}}) \Lambda_0(dw) \right\}' \boldsymbol{\Sigma}(s,t)^{-1} \\ &\quad \times n^{-1} \sum_{i=1}^{n} \int_0^t [\mathbf{X}_i(s) - \mathbf{E}(s,w;\boldsymbol{\beta}_0)]^{\otimes 2} Y_i(s,w;\boldsymbol{\beta}_{\mathbf{0}}) \Lambda_0(dw). \end{aligned} \tag{4.19}$$

It now follows from assumptions a-g and Lemma 3.2 that K_n converges weakly on Skorohod's space to a zero-mean Gaussian process with covariance function

$$\begin{aligned} & n^{\frac{1}{2}} \sum_{i=1}^{n} \int_0^t I(\boldsymbol{\beta}_{\mathbf{0}}' \mathbf{X}_i(s) \le u) y_i(s,w;\boldsymbol{\beta}_{\mathbf{0}}) \Lambda_0(dw) \\ & - \left\{ \sum_{i=1}^{n} \int_0^t I(\boldsymbol{\beta}_{\mathbf{0}}' \mathbf{X}_i(s) \le u) \mathbf{X}_i y_i(s,w;\boldsymbol{\beta}_{\mathbf{0}}) \Lambda_0(dw) \right\}' \boldsymbol{\Sigma}(s,t)^{-1} \\ & \times n^{\frac{1}{2}} \sum_{i=1}^{n} \int_0^t [\mathbf{X}_i(s) - \mathbf{e}(s,w;\boldsymbol{\beta}_0)]^{\otimes 2} y_i(s,w;\boldsymbol{\beta}_{\mathbf{0}}) \Lambda_0(dw). \end{aligned} \tag{4.20}$$

□

Proposition 4.2. *Under H_0, assume that the assumptions a-h are satisfied, we have*

$$\hat{K}_n(t,u) = K_n(t,u) - n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^t \frac{I(\boldsymbol{\beta}_{\mathbf{0}} \boldsymbol{X}_i(s) \le u) Y_i(s,w;\boldsymbol{\beta}_0)}{n S^{(0)}(s,w;\boldsymbol{\beta}_0)} M_i(s,dw). \tag{4.21}$$

Proof.

$$\begin{aligned}
\hat{K}_n(t,u) &= \hat{K}_n(s,t;u,\hat{\boldsymbol{\beta}}) = n^{-\frac{1}{2}}\sum_{i=1}^{n} I(\hat{\boldsymbol{\beta}}'\mathbf{X}_i(s)\le u)\hat{M}_i(s,t)\\
&= n^{-\frac{1}{2}}\sum_{i=1}^{n}\Big[I(\hat{\boldsymbol{\beta}}'\mathbf{X}_i(s)\le u)N_i(s,t) - \int_0^t I(\hat{\boldsymbol{\beta}}'\mathbf{X}_i(s)\le u)Y_i(s,w;\hat{\boldsymbol{\beta}})\hat{\Lambda}_0(dw)\Big]\\
&= n^{-\frac{1}{2}}\sum_{i=1}^{n}\Big[I(\hat{\boldsymbol{\beta}}'\mathbf{X}_i(s)\le u)N_i(s,t) - \int_0^t \frac{I(\hat{\boldsymbol{\beta}}'\mathbf{X}_i(s)\le u)Y_i(s,w;\hat{\boldsymbol{\beta}})}{nS^{(0)}(s,w;\hat{\boldsymbol{\beta}})}N(s,dw)\Big]\\
&= n^{-\frac{1}{2}}\sum_{i=1}^{n}\Big[I(\hat{\boldsymbol{\beta}}'\mathbf{X}_i(s)\le u)N_i(s,t) - \int_0^t \frac{I(\hat{\boldsymbol{\beta}}'\mathbf{X}_i(s)\le u)Y_i(s,w;\hat{\boldsymbol{\beta}})}{nS^{(0)}(s,w;\hat{\boldsymbol{\beta}})}N(s,dw)\Big]\\
&- n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t I(\hat{\boldsymbol{\beta}}'\mathbf{X}_i(s)\le u)Y_i(s,w;\hat{\boldsymbol{\beta}})\Lambda_0(dw)\\
&+ n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t I(\hat{\boldsymbol{\beta}}'\mathbf{X}_i(s)\le u)Y_i(s,w;\hat{\boldsymbol{\beta}})\Lambda_0(dw).\\
&= n^{-\frac{1}{2}}\sum_{i=1}^{n}\Big[I(\hat{\boldsymbol{\beta}}'\mathbf{X}_i(s)\le u)M_i(s,t;\hat{\boldsymbol{\beta}})\\
&- \int_0^t \frac{I(\hat{\boldsymbol{\beta}}'\mathbf{X}_i(s)\le u)Y_i(s,w;\hat{\boldsymbol{\beta}})}{nS^{(0)}(s,w;\hat{\boldsymbol{\beta}})}M_i(s,dw;\hat{\boldsymbol{\beta}})\Big].
\end{aligned}$$

It follows by the consistency of $\hat{\boldsymbol{\beta}}$ that

$$\begin{aligned}
&= n^{-\frac{1}{2}}\sum_{i=1}^{n}\Big[I(\boldsymbol{\beta}_{\mathbf{0}}'\mathbf{X}_i(s)\le u)M_i(s,t;\boldsymbol{\beta}_{\mathbf{0}})\\
&- \int_0^t \frac{I(\boldsymbol{\beta}_{\mathbf{0}}'\mathbf{X}_i(s)\le u)Y_i(s,w;\boldsymbol{\beta}_{\mathbf{0}})}{nS^{(0)}(s,w;\boldsymbol{\beta}_{\mathbf{0}})}M_i(s,dw;\boldsymbol{\beta}_{\mathbf{0}})\Big].
\end{aligned}$$

□

Theorem 4.3. *Under H_0, assume that the assumptions a-h are satisfied, we have*

$\hat{K}_n \longrightarrow_d K^\star$ *in the Skorohod space* $D[-\infty,\infty]$,

where $K^\star$ is a zero-mean Gaussian process with covariance function

$$
\begin{aligned}
& n^{\frac{1}{2}} \sum_{i=1}^{n} \int_0^t I(\boldsymbol{\beta}_0' \boldsymbol{X}_i(s) \leq u)\, y_i(s, w|\boldsymbol{\beta}_0)\Lambda_0(dw) \\
& - \left\{ \sum_{i=1}^{n} \int_0^t I(\boldsymbol{\beta}_0' \boldsymbol{X}_i(s) \leq u) \boldsymbol{X}_i y_i(s, w|\boldsymbol{\beta}_0)\Lambda_0(dw) \right\}' \boldsymbol{\Sigma}(s,t)^{-1} \\
& \times n^{\frac{1}{2}} \sum_{i=1}^{n} \int_0^t [\boldsymbol{X}_i(s) - \boldsymbol{e}(s, w|\boldsymbol{\beta}_0]^{\otimes 2} y_i(s, w|\boldsymbol{\beta}_0)\Lambda_0(dw) \\
& + \left[\frac{I(\boldsymbol{\beta}_0' \boldsymbol{X}_i(s) \leq u) y_i(s, w|\boldsymbol{\beta}_0)}{n s^{(0)}(s, w|\boldsymbol{\beta}_0)} \right]^{\otimes 2} s^{(0)}(s, w|\boldsymbol{\beta}_0)\Lambda_0(dw).
\end{aligned}
\tag{4.22}
$$

Proof. The proof of this Theorem follows directly from Theorem (4.3), Lemma (3.2), Condition (1), and Theorem (4.1). □

4.4. KOLMOGOROV-SMIRNOV TEST

In this section, we shall discuss the procedure which permit us to detect the cause of departure from the GLM, including wrong choice of link function or functional form. The test can be constructed as a Kolmogorov-Smirnov test statistic based on function of the process $\hat{K}_n(t, u; \hat{\boldsymbol{\beta}})$ given by:

$$
T^{(n)} = \sup_{\substack{u \in \mathbb{R}, \\ t \geq 0}} |\hat{K}_n(t, u; \hat{\boldsymbol{\beta}})| \tag{4.23}
$$

Observe that the KS test is based on the estimated parameters which makes the asymptotic distribution of this test process has complicated structure subsequently does not allow computation of the critical values. Therefore, p-value can be obtained by simulation, as follows:

- Generate recurrent event data, that is, with $\boldsymbol{\beta_0}$;
- Obtain the m.l.e's $\hat{\boldsymbol{\beta}}$;

- By repeating M times, generate $K^*(t,u;\hat{\boldsymbol{\beta}})$ and find $T^* = \sup_{\substack{u\in\mathbb{R},\\ t\geq 0}} |K^*(t,u;\hat{\boldsymbol{\beta}})|$;
- Estimate the probability that T^* exceeds the observed value $t^{(n)}$.

4.5. SIMULATION STUDY

4.5.1. Simulation Design. The simulation study was conducted to assess the performance of the proposed methodology. In the study, the model has the log link. For covariates, we consider a three-dimensional covariate vector $(X_1, X_2, , X_3)$ with $X_1 \sim N(0,1)$, $X_2 \sim Ber(0.5)$, and $X_3 \sim N(0,3)$. The true values of regression coefficient vector $(\beta_1, \beta_2, \beta_3)$ is set to be (1, -1,2). We consider the interoccurrence times T_{ij} is the Weibull distribution with shape parameter $\alpha = 0.5$ and parameter $\eta = 0.1$, and also the censoring variables τ_i from uniform distribution over the interval $(0, 360]$. The results given below are based on sample size n that varies in the set {80, 200} with 10K replications. The alternative links were:

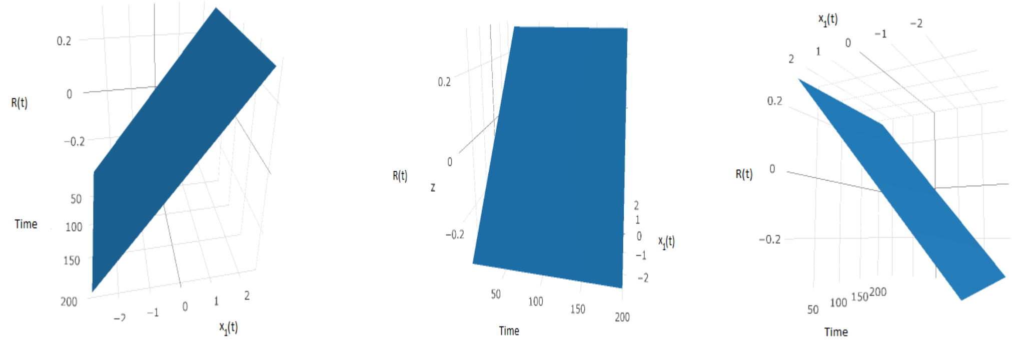
logit link:

$$\psi(\mathbf{x(t)}) = \frac{\exp(\beta_1 x_1(t) + \beta_2 x_2(t) + \beta_3 x_3(t))}{1 + \exp(\beta_1 x_1(t) + \beta_2 x_2(t) + \beta_3 x_3(t))}$$

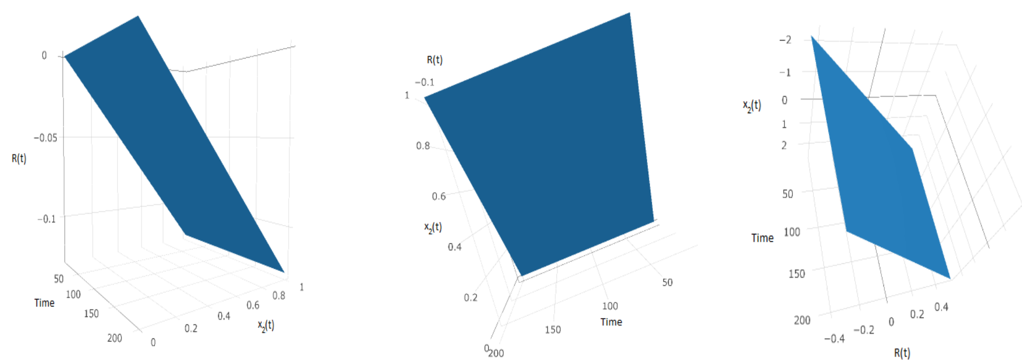
$\log_{NL}$ link:

$$\psi(\mathbf{x(t)}) = \exp(\beta_1 x_1(t) + \beta_2 x_2(t) + \beta_3 x_3^2(t)).$$

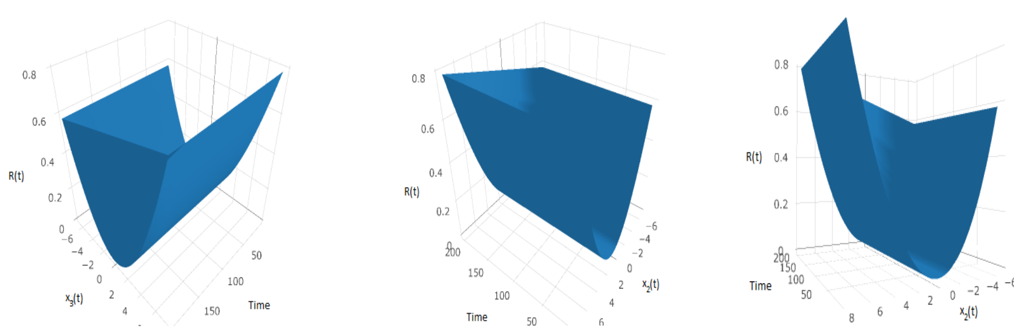
4.5.2. Discussion of Simulation Results. In Figures 4.1 - 4.3, we see that the proposed graphical method is consistent for all scenarios. Following the KS test statistic, Table 4.1 reports on the proportion of times the null hypothesis was rejected for $\alpha = 0.01$ and 0.05.



The residual from a fit of $x_1(t)$



The residual from a fit of $x_2(t)$



The residual from a fit of $x_2(t)$

Figure 4.1. Plot of partial residuals for the functional form $x_1(t) - x_2(t) + 2x_3^2(t)$

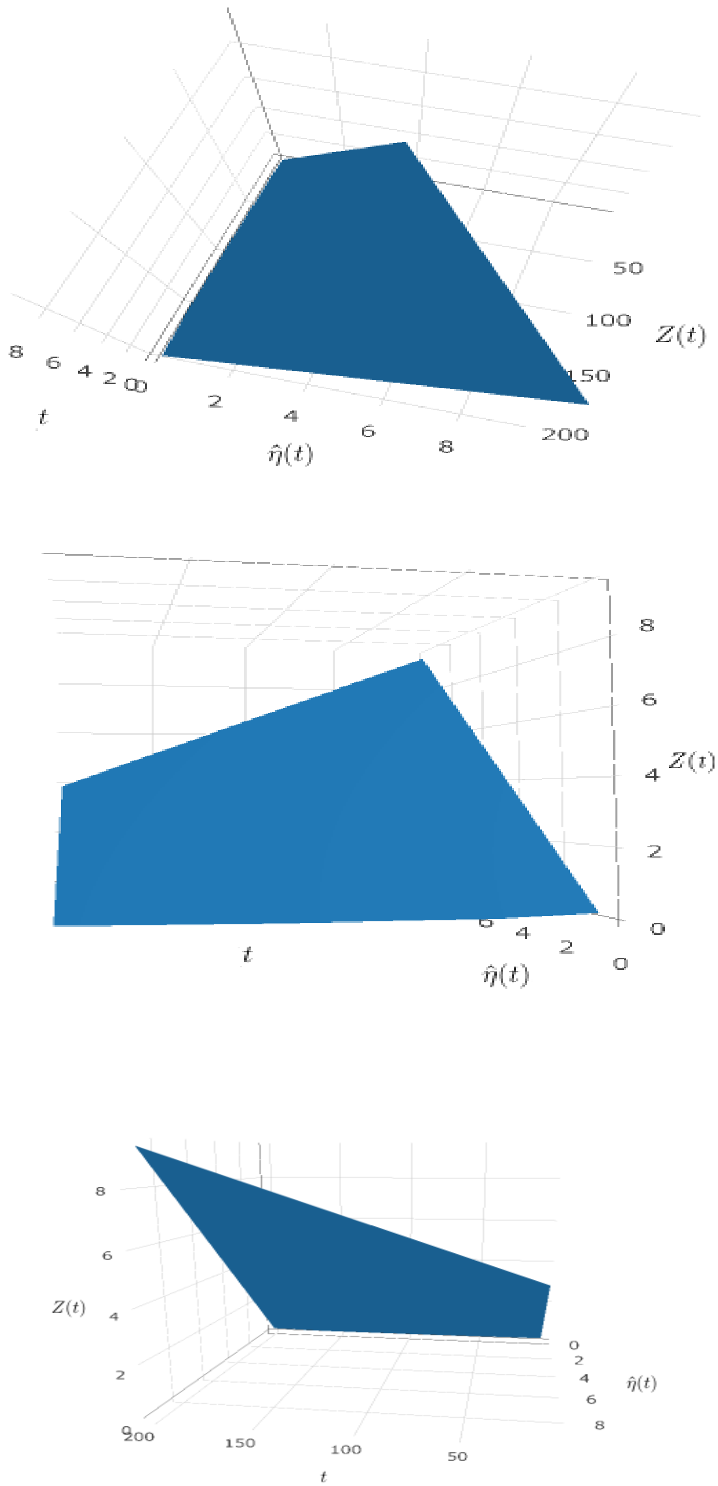


Figure 4.2. Plot of residuals for the log link function

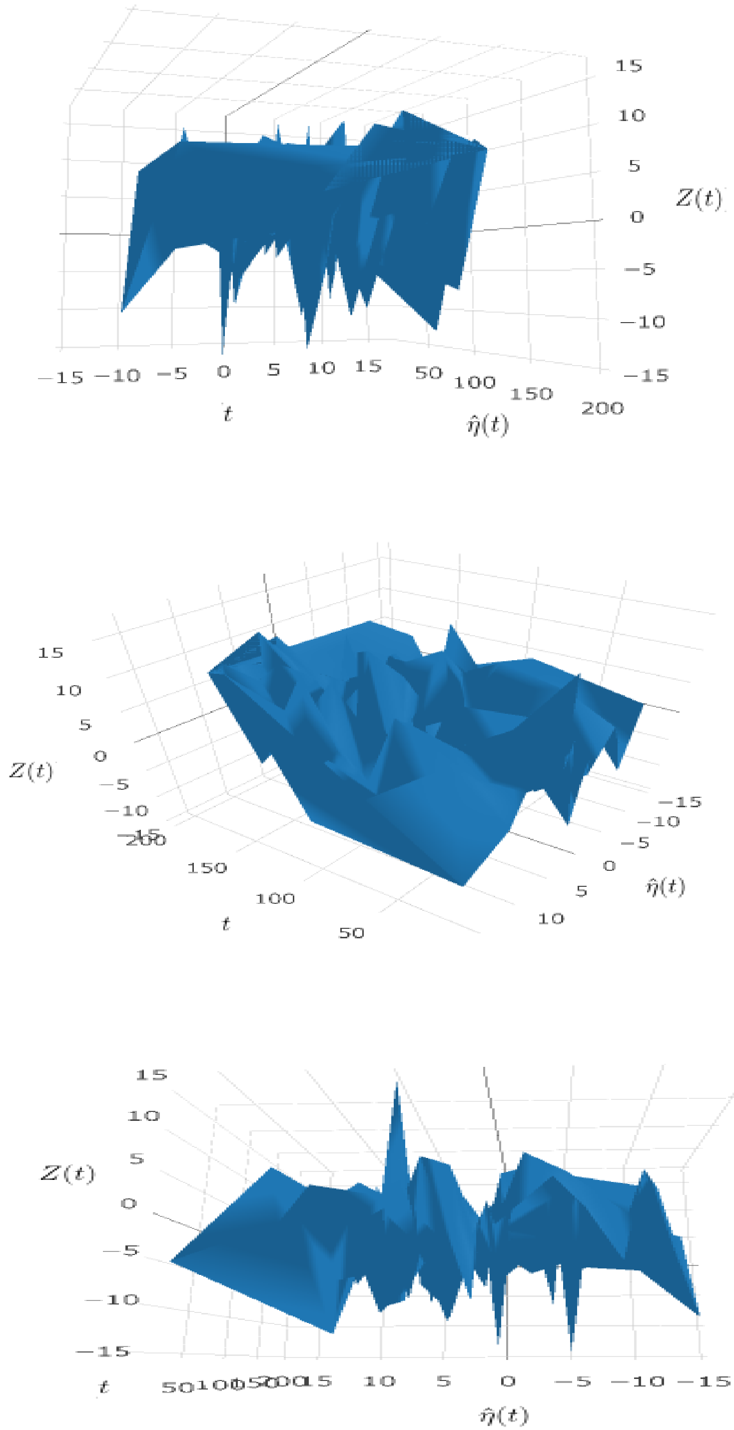


Figure 4.3. Plot of residuals for the logit link function

Table 4.1. Percentages of times H_0 was rejected

n	α	Log	Logit	$Log_N L$
80	0.05	3.1%	98.7%	96.6%
	0.01	1.5%	99.2%	97.8%
200	0.05	2.4%	98.7%	96.6%
	0.01	0.8%	100%	98.8%

BIBLIOGRAPHY

[1] Cox D.R. and Lewis P.A.W. The statistical analysis of. *Series of Events*, 1966.

[2] Cox D.R. and Isham V. *Point processes*, volume 12. CRC Press, 1980.

[3] Daley D.J. and Vere-Jones D. Basic properties of the poisson process. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*, pages 19–40, 2003.

[4] Daley D.J., Vere-Jones D., and Lewis P.A.W. A summary of the theory of point processes. *Stochastic point processes*, pages 299–383, 1972.

[5] Lewis P.A.W. Recent results in the statistical analysis of univariate point processes. *Stochastic point processes*, pages 1–54, 1972.

[6] Snyder D.L. and Miller M.I. *Random Point Processess in Time and Space*. Springer-Verlag & New York, 1991.

[7] Fleming T.R. and Harrington D.P. *Counting processes and survival analysis*, volume 169. John Wiley & Sons, 2011.

[8] Andersen P.K., Borgan O., Gill R.D., and Keiding N. *Statistical models based on counting processes*. Springer Science & Business Media, 2012.

[9] Aalen O.O., Fosen J., Weedon-Fekjaer H., Borgan O., and Husebye E. Dynamic analysis of multivariate failure time data. *Biometrics*, 60 3:764–73, 2004.

[10] Fosen J., Borgan Ø., Weedon-Fekjær H., and Aalen O.O. Dynamic analysis of recurrent event data using the additive hazard model. *Biometrical Journal*, 48(3):381–398, 2006.

[11] Fosen J., Ferkingstad E., Borgan Ø., and Aalen O.O. Dynamic path analysis?a new approach to analyzing time-dependent covariates. *Lifetime data analysis*, 12(2):143–167, 2006.

[12] Cox D.R. The statistical analysis of dependencies in point processes. *Stochastic point processes*, 55:66, 1972.

[13] Aalen O.O. and Husebye E. Statistical analysis of repeated events forming renewal processes. *Statistics in medicine*, 10(8):1227–1240, 1991.

[14] Follmann D.A. and Goldberg M.S. Distinguishing heterogeneity from decreasing hazard rates. *Technometrics*, 30(4):389–396, 1988.

[15] Dabrowska D.M., Sun G.W., and Horowitz M.M. Cox regression in a markov renewal model: an application to the analysis of bone marrow transplant data. *Journal of the American Statistical Association*, 89(427):867–877, 1994.

[16] Andersen P.K. and Gill R.D. Cox's regression model for counting processes: a large sample study. *The annals of statistics*, pages 1100–1120, 1982.

[17] Peña E.A. Dynamic modelling and statistical analysis of event times. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 21(4):1, 2006.

[18] Peña E.A., Slate E.H., and González J.R. Semiparametric inference for a general class of models for recurrent events. *Journal of Statistical Planning and Inference*, 137(6):1727–1747, 2007.

[19] Lawless J.F. *Statistical models and methods for lifetime data*, volume 362. John Wiley & Sons, 2011.

[20] Thall P.F. and Lachin J.M. Analysis of recurrent events: Nonparametric methods for random-interval count data. *Journal of the American Statistical Association*, 83(402):339–347, 1988.

[21] Gill R.D. Nonparametric estimation based on censored observations of a markov renewal process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 53(1):97–116, 1980.

[22] Prentice R.L., Williams B.J., and Peterson A.V. On the regression analysis of multivariate failure time data. *Biometrika*, 68(2):373–379, 1981.

[23] Lawless J.F. and Nadeau J.C. Nonparametric estimation of cumulative mean functions for recurrent events. *Technometrics*, 37:158–168, 1995.

[24] Cook R.J. and Lawless J. *The statistical analysis of recurrent events*. Springer Science & Business Media, 2007.

[25] Lin D.Y., Wei L.J., Yang I., and Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):711–730, 2000.

[26] Scheike T.H. The additive nonparametric and semiparametric aalen model as the rate function for a counting process. *Lifetime Data Analysis*, 8(3):247–262, 2002.

[27] Chiagn C.T., Wang M.C., and Huang C.Y. Kernel estimation of rate function for recurrent event data. *Scandinavian journal of statistics*, 32(1):77–91, 2005.

[28] Martinussen T. and Scheike T.H. *Dynamic regression models for survival data*. Springer Science & Business Media, 2007.

[29] Hollander M. and Peña E.A. Nonparametric methods in reliability. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 19(4):644, 2004.

[30] Stocker R. and Peña E.A. A general class of parametric models for recurrent event data. *Technometrics*, 49(2):210–221, 2007.

[31] Adekpedjou A. and Stocker R. A general class of semiparametric models for recurrent event data. *Journal of Statistical Planning and Inference*, 156:48–63, 2015.

[32] Aalen O.O. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, pages 701–726, 1978.

[33] Andersen P.K., Hansen L.S., and Keiding N. Non-and semi-parametric estimation of transition probabilities from censored observation of a non-homogeneous markov process. *Scandinavian Journal of Statistics*, pages 153–167, 1991.

[34] Kai Lai Chung. *A course in probability theory.* Academic press, New York, 1974.

[35] Jacod J. On the stochastic intensity of a random point process over the half-line. *Princeton Univ., Dept. Statist., Techn. Rep*, 51:33, 1973.

[36] Jacod J. Multivariate point processes: predictable projection, radon-nikodym derivatives, representation of martingales. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 31(3):235–253, 1975.

[37] Peña E.A., Strawderman R.L., and Hollander M. A weak convergence result relevant in recurrent and renewal models. In *Recent Advances in Reliability Theory*, pages 493–514. Springer, 2000.

[38] Sellke T. Weak convergence of the aalen estimator for a censored renewal process. *Statistical Decision Theory and Related Topics IV*, 2:183–194, 1988.

[39] Gill R.D. Testing with replacement and the product limit estimator. *The Annals of Statistics*, pages 853–860, 1981.

[40] Peña E.A., Strawderman R.L., and Hollander M. Nonparametric estimation with recurrent event data. *Journal of the American Statistical Association*, 96(456):1299–1315, 2001.

[41] Andersen P.K., Borgan Ø., Gill R., and Keiding N. Linear nonparametric tests for comparison of counting processes, with applications to censored survival data, correspondent paper. *International Statistical Review/Revue Internationale de Statistique*, pages 219–244, 1982.

[42] Kumar U. and Klefsjö B. Reliability analysis of hydraulic systems of lhd machines using the power law process model. *Reliability Engineering & System Safety*, 35(3):217 – 224, 1992.

[43] Lin D.Y. Cox regression analysis of multivariate failure time data: the marginal approach. *Statistics in medicine*, 13(21):2233–2247, 1994.

[44] Wei L.J. Testing goodness of fit for proportional hazards model with censored observations. 79:649–652, 1984.

[45] Pepe M.S. and Cai J. 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, 1993.

[46] Gill R. and Schumacher M. A simple test of the proportional hazards assumption. *Biometrika*, 74(2):289–300, 1987.

[47] Parzen M. and Lipsitz S.R. A global goodness-of-fit statistic for cox regression models. *Biometrics*, 55(2):580–584, 1999.

[48] Lin D.Y., Wei L.J., and Ying Z. Checking the cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80(3):557–572, 1993.

[49] Lin D.Y., Wei L.J., and Ying Z. Model-checking techniques based on cumulative residuals. *Biometrics*, 58(1):1–12, 2002.

[50] Schoenfeld D. Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika*, 67(1):145–153, 1980.

[51] Barlow W.E. and Prentice R.L. Residuals for relative risk regression. *Biometrika*, 75(1):65–74, 1988.

[52] Cox D.R. Regression models and life-tables (with discussions). *Jr Stat Soc B*, 34:187–220, 1972.

[53] Kay R. Proportional hazard regression models and the analysis of censored survival data. *Applied Statistics*, pages 227–237, 1977.

[54] Cox D.R. A note on the graphical analysis of survival data. *Biometrika*, 66(1):188–190, 1979.

[55] Kalbfleisch J.D. The propotional hazard model. *The statistical analysis of failure time data*, 1980.

[56] Kalbfleisch J.D. and Prentice R.L. *The statistical analysis of failure time data*, volume 360. John Wiley & Sons, 2011.

[57] Lin D.Y. and Wei L.J. Goodness-of-fit tests for the general cox regression model. *Statistica Sinica*, pages 1–17, 1991.

[58] White H. Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pages 1–25, 1982.

[59] Huang C.Y., Luo X., and Follmann D.A. A model checking method for the proportional hazards model with recurrent gap time data. *Biostatistics*, 12(3):535–547, 2010.

[60] Cheng K.F. and Wu J.W. Testing goodness of fit for a parametric family of link functions. *Journal of the American Statistical Association*, 89(426):657–664, 1994.

[61] Stute W. Nonparametric model checks for regression. *Ann. Statist.*, 25:613–641, 1997.

[62] Stute W., Thies S., and Zhu L.X. Model checks for regression: An innovation process approach. 26, 1998.

[63] Stute W. and ZHU L.X. Model checks for generalized linear models. *Scandinavian Journal of Statistics*, 29(3):535–545, 2002.

VITA

Muna Mohamed Hammuda attended Tripoli University, Tripoli, Libya. In July, 1993, she received her bachelor's degree in Statistics from University of Tripoli, Libya. From 1994 until 2006, Muna was employed as a teacher at Al-majed high school Tripoli, Libya. In February, 2006, Muna completed her MS in Statistics at Tripoli University, Tripoli, Libya. After receiving her Master's degree she was employed as a Lecturer at University of Tripoli from 2006 to 2010. In January, 2012, she joined Missouri University of Science and Technology to pursue her graduate studies and in May, 2017, she completed her second MS in Applied Mathematics. She received her PhD in Applied Mathematics with Statistics emphasis from Missouri University of Science and Technology in May, 2018.