

Efficient semiparametric estimation of time-censored intensity-reduction models for repairable systems

Wang, Jinyang ; Chen, Piao; Ye, Zhisheng

DOI

[10.1111/sjos.12564](https://doi.org/10.1111/sjos.12564)

Publication date

2022

Document Version

Final published version

Published in

Scandinavian Journal of Statistics

Citation (APA)

Wang, J., Chen, P., & Ye, Z. (2022). Efficient semiparametric estimation of time-censored intensity-reduction models for repairable systems. *Scandinavian Journal of Statistics*, 49(4), 1860-1888. <https://doi.org/10.1111/sjos.12564>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Efficient semiparametric estimation of time-censored intensity-reduction models for repairable systems

Jinyang Wang¹ | Piao Chen²  | Zhisheng Ye¹

¹Department of Industrial Systems Engineering & Management, National University of Singapore, Singapore

²Delft Institute of Applied Mathematics, Delft University of Technology, Delft, The Netherlands

Correspondence

Piao Chen, Delft Institute of Applied Mathematics, Delft University of Technology, Delft, The Netherlands.
Email: p.chen-6@tudelft.nl

Funding information

National Natural Science Foundation of China, Grant/Award Number: 72071138; Singapore MOE AcRF Tier 2, Grant/Award Number: R-266-000-143-112

Abstract

The rate reduction models have been widely used to model the recurrent failure data for their capabilities in quantifying the repair effects. Despite the widespread popularity, there have been limited studies on statistical inference of most failure rate reduction models. In view of this fact, this study proposes a semiparametric estimation framework for a general class of such models, called extended geometric failure rate reduction (EGFRR) models. Covariates are considered in our analysis and their effects are modeled as a log-linear factor on the baseline failure rate. Unlike the existing inference methods for the EGFRR models that assume the failure data are censored at a fixed number of failures, our study considers covariates and time-censoring, which are more common in practice. The semiparametric maximum likelihood (ML) estimators are obtained by carefully constructing the likelihood function. Asymptotic properties including consistency and weak convergence of the ML estimators are established by using the properties of the martingale process. In addition, we show that the semiparametric estimators are asymptotically efficient. A real example from the automobile industry illustrates the usefulness of the proposed framework and

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Scandinavian Journal of Statistics* published by John Wiley & Sons Ltd on behalf of The Board of the Foundation of the Scandinavian Journal of Statistics.

extensive simulations show its outstanding performance when comparing with the existing methods.

KEYWORDS

imperfect repair, martingale, recurrent events, repairable systems, semiparametric efficiency

1 | INTRODUCTION

Many physical systems such as vehicles, commercial products, and manufacturing equipment are repairable in the sense that they are repaired rather than replaced in the presence of failures. Reliability evaluation of these systems is mainly based on modeling of the recurrent failure processes. Due to the complicated effects of the repair actions on the system health, however, modeling of the repairable systems is not an easy task, and it is much more difficult than modeling of the nonrepairable systems.

In the literature, a multitude of models have been proposed to account for the repair effects in the recurrent failures. Most of these models can be defined based on the intensity process. Consider a repairable system observed over time $t \geq 0$, and let S_1, S_2, \dots be the failure time sequence of the system. Denote $N^*(t)$ as the number of failures/repairs in the time interval $[0, t]$. Define $\mathcal{F}(t) = \sigma\{N^*(s) : 0 \leq s \leq t\}$ to be the filtration, that is, the history of the process up to time t . The intensity process $\nu(t) \equiv \nu(t; \mathcal{F}_{t-})$ of the counting process $N^*(t)$ is defined as the conditional probability that an event occurs in $[t, t + dt)$ given the filtration \mathcal{F}_{t-} , that is,

$$\nu(t)dt = P\{dN^*(t) = 1 | \mathcal{F}_{t-}\}.$$

Let $\lambda_0(t)$ be the baseline intensity function, and $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ be the corresponding baseline cumulative intensity function. The intensity processes of many imperfect maintenance models are based on a modification of the baseline intensity λ_0 . When the repair effect is assumed to be as good as new (AGAN), $\{N^*(t), t \geq 0\}$ is a renewal process and the intensity process is given by $\nu(t) = \lambda_0\{t - S_{N^*(t-)}\}$. If the repair is as bad as old (ABAO), $\{N^*(t), t \geq 0\}$ is a nonhomogeneous Poisson process and the intensity process is the same as λ_0 , that is, $\nu(t) = \lambda_0(t)$. In practice, it is more common that the repair effect is between AGAN and ABAO, which is known as imperfect repair. In terms of modeling of the imperfect repair, Kijima's Type I and Type II models (Kijima, 1989) represent the first attempts by introducing the concept of virtual age. The Kijima's models have been further extended in many follow-up studies with applications to optimal decisions in preventive maintenance. For an overview of these extensions, see Finkelstein (2008), Zhang and Xie (2017) and Alaswad and Xiang (2017).

The Kijima's models treat the repair effect as a shift on the original survival function with a virtual age. In the literature, models with a change on the failure rate after repair have also been widely used for imperfect repairs. Among them, the Geometric Failure Rate Reduction (GFRR) model proposed by Finkelstein (2008) is a flexible model with a simple intensity process $\lambda(t) = \rho^{N^*(t-)} \lambda_0(t - S_{N^*(t-)})$, where the parameter $\rho > 0$ characterizes the repair effect. The empirical studies in Doyen et al. (2017) and Syamsundar and Kumar (2017) have revealed that the GFRR model outperforms most existing imperfect repair models when fitting to real

datasets. To mitigate the explosion problem, for which $\rho^{N^*(t^-)}$ becomes too large when ρ is far from 1 or when $N^*(t)$ is large, Dauouis et al. (2019) proposed an Extended GFRR (EGFRR) model with the intensity process $v(t) = \rho^{\alpha(N^*(t^-)+1)} \lambda_0(t - S_{N^*(t^-)})$, where $\alpha(k)$ is a known non-decreasing function such that $\alpha(1) = 0$. This model implies that the hazard function of the k th interoccurrence time equals $\rho^{\alpha(k)} \lambda_0(\cdot)$. The EGFRR model degenerates to the GFRR model when $\alpha(k) = k - 1$. For the EGFRR model, Dauouis et al. (2019) have developed a semiparametric inference procedure assuming that the failure data are censored at a fixed number of failures.

In most maintenance applications, however, the observed repair process is time-censored because of an end-of-study date (Ye & Ng, 2014), which is also called administrative censoring in biostatistics. This time-censoring assumption is widely used in recurrence data; see Pena et al. (2001), Pena et al. (2007) and Rahman et al. (2014). In addition, the covariates may also play a critical role in affecting the failure rates; see Cox (1972), Wang and Xu (2010) and Qu et al. (2016). In view of this fact, we propose a semiparametric inference procedure for the recurrent failure data by considering the covariates and the time-censoring scheme. That is, in each experiment, we can only observe the failures/repairs before a censoring or follow-up time, which might incorporate the covariates. In specific, the intensity process considered is

$$v(t; \mathbf{X}) = \rho^{\alpha(N^*(t^-)+1)} e^{\beta^T \mathbf{X}} \lambda_0(t - S_{N^*(t^-)}), \quad (1)$$

where β is a $p \times 1$ vector and $\mathbf{X} = (x_1, x_2, \dots, x_p)^T$ is a p -dimensional covariate, such as the use rate, brand, and other observable features that vary from unit to unit. In this study, the maximum likelihood (ML) estimators and their asymptotic properties are thoroughly investigated. In particular, semiparametric efficiency properties, which are important in knowing how much the estimators can be improved, are established for the proposed estimators. The semiparametric efficiency is rarely discussed in the literature of statistical inference for the repairable systems, and *our study bridges this gap by examining the efficiency of both the parametric and the nonparametric parts of the estimators*.

The rest of the paper is organized as follows. In Section 2 we derive the likelihood and the ML estimators based on the time-censored repair data with covariates. Section 3 establishes the asymptotic properties including the consistency, the weak convergence and the semiparametric efficiency for the ML estimators. Comprehensive simulations are conducted in Section 4, and a real example from the automobile industry is used to illustrate the proposed inference procedures in Section 5. At last, Section 6 concludes the paper.

2 | DATA AND LIKELIHOOD

Consider a system whose failure/repair process $\{N^*(t), t \geq 0\}$ is governed by the EGFRR model (1) with a covariate vector \mathbf{X} . By letting $\gamma = \log(\rho)$, the intensity can be written as

$$v(t; \mathbf{X}) = e^{\gamma \cdot \alpha(N^*(t^-)+1) + \beta^T \mathbf{X}} \lambda_0(t - S_{N^*(t^-)}).$$

Denote T_j as the interoccurrence time between the $(j - 1)$ th and the j th failures, and $S_j = \sum_{l=1}^j T_l$ as the occurrence time of the j th failure, $j = 1, 2, \dots$. For notational convenience, we stipulate $T_0 = 0$ and $S_0 = 0$. Let C denote the censoring or follow-up time of the system. The censoring

mechanism is supposed to be conditionally independent of the failure process, that is, C and $N^*(t)$ are independent conditional on \mathbf{X} for any $t \geq 0$. Define $\tilde{N}(t) = N^*(t \wedge C)$ and $\tilde{Y}(t) = I\{C \geq t\}$, where $a \wedge b = \min\{a, b\}$, and $I\{A\}$ is the indicator function that equals 1 if A is true, and 0 otherwise. The observed failure data from the system is $\{\mathbf{X}, \tilde{N}(\cdot), \tilde{Y}(\cdot)\}$. We are interested in estimating the true parameters, denoted as $\gamma_0, \boldsymbol{\beta}_0$ and Λ_0 , from the data.

2.1 | The likelihood

Based on the above problem setting, $\tilde{N}(s)$ is a submartingale with respect to the filtration \mathcal{F}_s with compensator $\tilde{A}(s; \gamma_0, \boldsymbol{\beta}_0, \Lambda_0) = \int_0^s \tilde{Y}(u) e^{\gamma_0 \cdot \alpha(\tilde{N}(u-)+1) + \boldsymbol{\beta}_0^T \mathbf{X}} \lambda_0(R(u)) du$, where $R(u) = u - S_{\tilde{N}(u-)}$ is the time from the last failure to time u . However, the compensator is a complicated function of the model parameters since the integrand contains a random argument $R(u)$. Therefore, decomposition of this submartingale may not be useful in parameter estimation. In this study, we follow Pena et al. (2001) and introduce a two-dimensional counting process

$$N(s, t) = \int_0^s Z(u, t) d\tilde{N}(u), \tag{2}$$

where $Z(u, t) = I\{R(u) \leq t\}$. Here, $N(s, t)$ is the failure count in the time window $[0, s]$ whose interoccurrence times are shorter than t . It is readily seen that $N(s, t) = N(s)$ when $s < t$. Let $\mathcal{F}_s = \sigma\{\mathbf{X}, \tilde{N}(u), \tilde{Y}(u+) : 0 \leq u \leq s\}$ be the event history up to time s . The following lemma summarizes the decomposition results for $N(s, t)$, which can be easily verified by proposition 31.1 in Pena et al. (2000).

Lemma 1. *For each fixed t , consider the counting process $\{N(s, t) : 0 \leq s \leq \tau\}$ indexed by s as defined above. Then there exists a unique increasing right-continuous \mathcal{F}_s -predictable process $\{A(s, t; \gamma_0, \boldsymbol{\beta}_0, \Lambda_0) : s \leq \tau\}$ given by*

$$A(s, t; \gamma, \boldsymbol{\beta}, \Lambda) = \int_0^t e^{\boldsymbol{\beta}^T \mathbf{X}} Y(s, x; \gamma) \Lambda(dx), \tag{3}$$

such that $N(s, t) - A(s, t; \gamma_0, \boldsymbol{\beta}_0, \Lambda_0)$, denoted by $M(s, t; \gamma_0, \boldsymbol{\beta}_0, \Lambda_0)$, is a right-continuous local martingale with respect to the filtration \mathcal{F}_s for fixed t . Here,

$$Y(s, t; \gamma) = \sum_{j=1}^{\tilde{N}(s-)} I\{T_j \geq t\} \cdot e^{\gamma \cdot \alpha(j)} + I\{R(s \wedge C) \geq t\} \cdot e^{\gamma \cdot \alpha(\tilde{N}(s-)+1)}.$$

In the lemma, $Y(s, t; \gamma)$ can be roughly understood as the weighted count of subintervals in $[0, s \wedge C]$ whose lengths are longer than t , where $[0, s \wedge C]$ is partitioned by the mesh defined by all failure times within this time interval. The lemma above enables us to write the likelihood function. According to corollary II.7.3 in Andersen et al. (2012), the likelihood based on the observation $\{\mathbf{X}, \tilde{N}(u), \tilde{Y}(u+), u \in [0, s]\}$ is

$$L(\tau, \tau; \gamma, \boldsymbol{\beta}, \Lambda) = \prod_{u \in [0, \tau]} A(du, \tau; \gamma, \boldsymbol{\beta}, \Lambda)^{N(du, \tau)} \cdot \exp\{-A(u, \tau; \gamma, \boldsymbol{\beta}, \Lambda)\},$$

where we stipulate $0^0 = 1$. Based on the definition of $N(s, t)$ and $A(s, t; \gamma, \beta, \Lambda)$, the log-likelihood function can be derived as

$$\begin{aligned} \ell(\gamma, \beta, \Lambda) &= \int_0^\tau \log(\Lambda\{R(u)\})Z(u, \tau)\tilde{N}(du) - \int_0^\tau e^{\beta^\top \mathbf{X}}Y(\tau, x; \gamma)\Lambda(dx) \\ &\quad + \int_0^\tau (\gamma \cdot \alpha(\tilde{N}(u-) + 1) + \beta^\top \mathbf{X})Z(u, \tau)\tilde{N}(du), \end{aligned} \tag{4}$$

where $\Lambda\{x\}$ denotes the pointmass of the baseline cumulative intensity function $\Lambda(x)$, which equals $\lambda(x)$ when $\Lambda(x)$ is differentiable (Van der Vaart, 2000, pp. 425). By proposition 1 in Pena et al. (2001), the first term of the right-hand side of the above display can be expressed as

$$\int_0^\tau \log(\Lambda\{R(u)\})Z(u, \tau)\tilde{N}(du) = \int_0^\tau \log(\Lambda\{x\})N(\tau, dx).$$

2.2 | ML estimator

For notational convenience, we let $\theta = (\gamma, \beta^\top)^\top$. Based on the log-likelihood in (4), the score function for the parametric part θ is

$$\dot{\ell}_1(\gamma, \beta, \Lambda) = - \int_0^\tau \begin{pmatrix} e^{\beta^\top \mathbf{X}}Y^{(1)}(\tau, x; \gamma) \\ e^{\beta^\top \mathbf{X}}Y(\tau, x; \gamma)\mathbf{X} \end{pmatrix} \Lambda(dx) + \int_0^\tau \begin{pmatrix} \alpha(\tilde{N}(u-) + 1) \\ \mathbf{X} \end{pmatrix} Z(u, \tau)\tilde{N}(du), \tag{5}$$

where $Y^{(k)}$ represents the k th derivative of Y with respect to γ , $k = 1, 2, \dots$. In order to derive the ML estimator for the nonparametric part, for a fixed Λ and each $h(\cdot) \in \ell^\infty([0, \tau])$, which is the space of uniform bounded functions, we consider the parametric submodel $\Lambda_{\epsilon, h}$ indexed by ϵ in a neighborhood of zero given by

$$\epsilon \mapsto \Lambda_{\epsilon, h}(\cdot) : \Lambda_{\epsilon, h}(u) = \int_0^u (1 + \epsilon h(x)) \Lambda(dx).$$

Notice that $\Lambda_{\epsilon, t}$ is nonnegative and monotone increasing when ϵ takes its value in a neighborhood of zero. Intuitively, an optimal nonparametric estimator should maximize the log-likelihood among submodels going through it. We will first show that the nonparametric part is unique under the submodel, and demonstrate the asymptotic properties of the estimator in the next section. For this submodel, we can calculate the score function by

$$\dot{\ell}_2(\gamma, \beta, \Lambda)(t) = \left. \frac{\partial}{\partial \epsilon} \ell(\gamma, \beta, \Lambda_{\epsilon, h}) \right|_{\epsilon=0} = \int_0^\tau h(t) \left(N(\tau, dx) - e^{\beta^\top \mathbf{X}}Y(\tau, x; \gamma)\Lambda(dx) \right).$$

Now, suppose we observe n i.i.d. copies of $\{\mathbf{X}, \tilde{N}(\cdot), \tilde{Y}(\cdot)\}$, denoted as $\{\mathbf{X}_i, \tilde{N}_i(\cdot), \tilde{Y}_i(\cdot)\}$, $i = 1, 2, \dots, n$. Let \mathbb{P}_n be the empirical measure based on the n i.i.d. realizations. By fixing γ , we can solve $\mathbb{P}_n \dot{\ell}_2(\gamma, \beta, \Lambda)(t) = 0$ for any $h(\cdot) \in \ell^\infty([0, \tau])$ to get

$$\hat{\Lambda}_n(t; \gamma, \beta) = \int_0^t \frac{I\left\{ \sum_{i=1}^n e^{\beta^\top \mathbf{X}_i} Y_i(\tau, x; \gamma) > 0 \right\}}{\sum_{i=1}^n e^{\beta^\top \mathbf{X}_i} Y_i(\tau, x; \gamma)} \bar{N}(\tau, dx), \tag{6}$$

where $\bar{N}(\tau, dx) = \sum_{i=1}^n N_i(\tau, dx)$, as with Fleming and Harrington (2005). By substituting the expression (6) into (4), the profile log-likelihood function for $\theta = (\gamma, \beta^\top)^\top$ can be obtained as

$$\begin{aligned} \ell_n(\gamma, \beta) &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau (\gamma \cdot \alpha(\tilde{N}_i(u-)) + 1) + \beta^\top \mathbf{X}_i Z_i(u, \tau) \tilde{N}_i(du) \\ &\quad - \frac{1}{n} \int_0^\tau \log \left(\frac{1}{n} \sum_{i=1}^n e^{\beta^\top \mathbf{X}_i} Y_i(\tau, x; \gamma) \right) \bar{N}(\tau, dx). \end{aligned} \tag{7}$$

Here we multiply the log-likelihood by n^{-1} for notational convenience when we consider the asymptotic properties. The following proposition shows that ℓ_n is concave.

Proposition 1. *The function $\ell_n : \mathbb{R} \times \mathbb{R}^p \mapsto \mathbb{R}$ defined in (7) is continuous and concave.*

Let $\hat{\theta}_n^\top = (\hat{\gamma}_n, \hat{\beta}_n^\top)$ be a solution to the estimating equation $\frac{\partial}{\partial \theta} \ell_n(\gamma, \beta) = 0$. Because of the concavity, existing convex optimization packages can be used to find $\hat{\theta}_n$ numerically and stably. Then the estimator of the nonparametric part is $\hat{\Lambda}_n(t; \hat{\gamma}_n, \hat{\beta}_n)$.

2.3 | Time-varying covariates

In this section, we consider the case of time-varying covariates. The intensity function is given by

$$v(t; \mathbf{X}) = e^{\gamma \cdot \alpha(N^*(t-)+1) + \beta^\top \mathbf{X}(t)} \lambda_0(t - S_{N^*(t-)}),$$

where $\mathbf{X}(t)$ is the time-dependent covariate. As we have shown in Section 2.2, according to Cox (1972), Lin (1991) and Huang et al. (2010), the log-likelihood based on the data $\{\mathbf{X}(\cdot), \tilde{N}(\cdot), \check{Y}(\cdot)\}$ is given by

$$\begin{aligned} \ell(\gamma, \beta, \Lambda) &= \int_0^\tau \log(\Lambda\{x\}) N(\tau, dx) - \int_0^\tau \check{Y}(\tau, x; \gamma, \beta, \mathbf{X}) \Lambda(dx) \\ &\quad + \int_0^\tau (\gamma \cdot \alpha(\tilde{N}(u-)) + 1) + \beta^\top \mathbf{X}(u) Z(u, \tau) \tilde{N}(du), \end{aligned}$$

where the weighted count $\check{Y}(s, t; \gamma, \beta, \mathbf{X})$ is

$$\check{Y}(s, t; \gamma, \beta, \mathbf{X}) = \sum_{j=1}^{\tilde{N}(s-)} I\{T_j \geq t\} \cdot e^{\gamma \cdot \alpha(j) + \beta^\top \mathbf{X}(t+S_{j-1})} + I\{R(s \wedge C) \geq t\} \cdot e^{\gamma \cdot \alpha(\tilde{N}(s-)+1) + \beta^\top \mathbf{X}(t+S_{\tilde{N}(s-)})}.$$

Therefore, we can again consider the family of submodels $\Lambda_{\epsilon, h}(u) = \int_0^u (1 + \epsilon h(x)) \Lambda(dx)$ for $h(\cdot) \in \ell^\infty([0, \tau])$ to see that the nonparametric estimator by fixing γ is

$$\hat{\Lambda}_n(t; \gamma, \beta) = \int_0^t \frac{I\{\sum_{i=1}^n \check{Y}_i(\tau, x; \gamma, \beta, \mathbf{X}_i) > 0\}}{\sum_{i=1}^n \check{Y}_i(\tau, x; \gamma, \beta, \mathbf{X}_i)} \bar{N}(\tau, dx),$$

which simply replaces $e^{\beta^\top \mathbf{X}} Y(\tau, x; \gamma)$ by $\check{Y}(\tau, x; \gamma, \beta, \mathbf{X})$ in (6). The parametric estimator can then be readily obtained by maximizing the profile likelihood function. For the purpose of concise

presentation, we use \mathbf{X} to denote the covariate vector in the remaining of the paper. In the presence of time-varying covariates, the derivatives of $\check{Y}(s, t; \gamma, \beta, \mathbf{X})$ with respect to γ and β can replace the derivatives of $e^{\beta^\top \mathbf{X}} Y(s, t; \gamma)$ in all the theoretical findings.

2.4 | Failure censoring

In this section, we verify that the derived ML estimators in Section 2.2 remain valid when the data are failure censored, which is the setting in Dauxois et al. (2019). The authors considered a single recurrent event with imperfect maintenances in their paper. Assume that the number of imperfect maintenances of a repairable system is censored at m , which is a positive integer-valued random variable upper bounded by a fixed given number N . Further assume that the distribution of m does not involve the parameters θ and Λ . Based on the observation (T_1, \dots, T_m) , the likelihood function can be written as

$$L(\gamma, \beta, \Lambda) = \prod_{j=1}^m \{ \lambda_j(T_j) \exp[-\Lambda_j(T_j)] \} = \prod_{j=1}^m \left\{ e^{\gamma \alpha(j) + \beta^\top \mathbf{X}} \lambda(T_j) \cdot \exp \left[-e^{\gamma \alpha(j) + \beta^\top \mathbf{X}} \Lambda(T_j) \right] \right\}.$$

This likelihood is equal to (4) by letting $C = S_m$, where S_m is the occurrence time of the m th failure. Therefore, the estimators $\hat{\theta}_n$ and $\hat{\Lambda}(\cdot)$ given in Section 2.2 are also ML estimators under the failure censoring mechanism. In Section 4, extensive simulations will be conducted to compare the proposed ML estimators with the estimators in Dauxois et al. (2019) based on the failure censored data.

3 | ASYMPTOTIC PROPERTIES

In this section, we study the asymptotic properties of the estimators derived in Section 2, including the consistency, weak convergence, and semiparametric efficiency.

3.1 | Consistency and asymptotic normality

Following the conventional notation in the large sample theory in survival analysis, we let

$$\begin{aligned} S_n^{(0)}(s, t; \gamma, \beta) &= n^{-1} \sum_{i=1}^n e^{\beta^\top \mathbf{X}_i} Y_i(s, t; \gamma), \\ S_n^{(1)}(s, t; \gamma, \beta) &= \begin{pmatrix} n^{-1} \sum_{i=1}^n e^{\beta^\top \mathbf{X}_i} Y_i^{(1)}(s, t; \gamma) \\ n^{-1} \sum_{i=1}^n e^{\beta^\top \mathbf{X}_i} Y_i(s, t; \gamma) \mathbf{X}_i \end{pmatrix}, \\ S_n^{(2)}(s, t; \gamma, \beta) &= \begin{pmatrix} n^{-1} \sum_{i=1}^n e^{\beta^\top \mathbf{X}_i} Y_i^{(2)}(s, t; \gamma) & n^{-1} \sum_{i=1}^n e^{\beta^\top \mathbf{X}_i} Y_i^{(1)}(s, t; \gamma) \mathbf{X}_i^\top \\ n^{-1} \sum_{i=1}^n e^{\beta^\top \mathbf{X}_i} Y_i^{(1)}(s, t; \gamma) \mathbf{X}_i & n^{-1} \sum_{i=1}^n e^{\beta^\top \mathbf{X}_i} Y_i(s, t; \gamma) \mathbf{X}_i^{\otimes 2} \end{pmatrix}, \end{aligned}$$

where $\mathbf{X}^{\otimes 2} = \mathbf{X}\mathbf{X}^\top$. Notice that $S_n^{(1)}$ is the gradient of $S_n^{(0)}$ and $S_n^{(2)}$ is the Hessian matrix of $S_n^{(0)}$ with respect to the parametric part $(\gamma, \beta^\top)^\top \in \mathbb{R}^{p+1}$. Further define

$$V_n(s, t; \gamma, \beta) = \frac{S_n^{(0)}(s, t; \gamma, \beta) S_n^{(2)}(s, t; \gamma, \beta) - S_n^{(1)}(s, t; \gamma, \beta)^{\otimes 2}}{S_n^{(0)}(s, t; \gamma, \beta)^2}, \tag{8}$$

and $s^{(k)}(\tau, t; \gamma, \beta) = E[S_1^{(k)}(\tau, t; \gamma, \beta)]$, for $k = 0, 1, 2$. It is obvious that $\sup_{t \in [0, \tau]} \|S_n^{(k)}(\tau, t; \gamma, \beta) - s^{(k)}(\tau, t; \gamma, \beta)\| \rightarrow_P 0$ for every fixed γ and β satisfying $s^{(k)}(\tau, 0; \gamma, \beta) < \infty$. Furthermore, we can obtain a result of uniform convergence with respect to θ according to the following lemma.

Lemma 2. Consider a convex neighborhood Θ of θ_0 . Let M_γ and M_β be the supremum of $|\gamma|$ and $\|\beta\|$ in Θ . Suppose $\sup_{t \in [0, \tau]} |\mathbf{X}(t)| < M_X$, and the random variables defined below has finite second moment:

$$Y^{(k)}(\tau, 0; M_\gamma) = \sum_{j=1}^{\tilde{N}(\tau-)} \alpha^k(j) e^{M_\gamma \alpha(j)} + \alpha^k (\tilde{N}(\tau-) + 1) e^{M_\gamma \alpha(\tilde{N}(\tau-)+1)},$$

where $k = 0, 1, 2$, and 3. Then

$$\sup_{t \in [0, \tau], \theta \in \Theta} \|S_n^{(k)}(\tau, t; \gamma, \beta) - s^{(k)}(\tau, t; \gamma, \beta)\| \rightarrow_P 0.$$

We should remark that the conclusion in Lemma 2, that is, the last display, is usually treated as a regularity condition in the literature of survival analysis. Using the advanced tools of empirical process, the above lemma directly establishes these uniform convergence results, which will be useful in establishing the asymptotic properties of our estimators. Actually, under the conditions given in Lemma 2, it is easy to see that

$$\sup_{t \in [0, \tau], \theta \in \Theta} \|s^{(k)}(\tau, t; \gamma, \beta)\|^2 \leq E\|Y^{(k)}(\tau, 0; M_\gamma)\|^2 e^{2\beta^\top \mathbf{X}} < \infty.$$

That is, $s^{(k)}(\tau, t; \gamma, \beta)$ is uniformly bounded on $[0, \tau]$ for $k = 0, 1$ and 2. In addition, notice that $ES_n^{(0)}(\tau, \tau; \gamma, \beta) \geq P\{C \geq \tau, T_1 \geq \tau\} \cdot Ee^{\gamma\alpha(1)+\beta^\top \mathbf{X}}$ for every fixed γ and β . Therefore, if $P\{C \geq \tau\} > 0$ and $\Lambda_0(\tau) < \infty$, then such $s^{(0)}(\tau, t; \gamma, \beta)$ is uniformly bounded away from 0.

Before presenting the asymptotic results, we list the following regularity conditions.

- (i) $P\{C \geq \tau\} > 0$ for the predetermined constant τ , and the baseline cumulative intensity function satisfies $\Lambda_0(\tau) < \infty$.
- (ii) The covariate X is uniformly bounded, and there exists an $M_\gamma > \gamma$ such that the second moment of following random variables

$$Y^{(k)}(\tau, 0; M_\gamma) = \sum_{j=1}^{\tilde{N}(\tau-)} \alpha^k(j) e^{M_\gamma \alpha(j)} + \alpha^k (\tilde{N}(\tau-) + 1) e^{M_\gamma \alpha(\tilde{N}(\tau-)+1)},$$

are finite, where $k = 0, 1, 2, 3$.

- (iii) Define $v = s^{(2)}/s^{(0)} - (s^{(1)}/s^{(0)})^{\otimes 2}$, where $s^{(k)}$ is the limit of $S_n^{(k)}$ in Lemma 2. The matrix

$$\Sigma(\gamma_0, \beta_0) = \int_0^\tau v(\tau, x; \gamma_0, \beta_0) s^{(0)}(\tau, x; \gamma_0, \beta_0) \lambda_0(x) dx,$$

is positive definite.

Condition (i) is a mild requirement on the underlying intensity process, which is commonly adopted in recurrent data analysis. See Lin et al. (2000), Gao et al. (2019) and Huang et al. (2010). Condition (ii) is the sufficient condition for Lemma 2. Conditions (iii) are widely used in survival

analysis, for example, see Andersen and Gill (1982), Lin et al. (2000) and Adekpedjou and Stocker (2015), among others.

The consistency of the estimators $\hat{\gamma}_n$, $\hat{\beta}_n$, and $\hat{\Lambda}_n(t; \hat{\gamma}_n, \hat{\beta}_n)$ is summarized as follows.

Theorem 1. *Under regularity conditions (i)–(iii), $\hat{\theta}_n = (\hat{\gamma}_n, \hat{\beta}_n)^\top$ converges to $\theta_0 = (\gamma_0, \beta_0^\top)^\top$ in probability and*

$$\sup_{0 \leq t \leq \tau} \left| \hat{\Lambda}_n(t; \hat{\gamma}_n, \hat{\beta}_n) - \Lambda_0(t) \right| \rightarrow 0$$

in probability as $n \rightarrow \infty$.

Next, we investigate the asymptotic normality of the estimators. For notational convenience, define

$$D(t; \gamma_0, \beta_0, \Lambda_0) = \int_0^t \frac{s^{(1)}(\tau, x; \gamma_0, \beta_0)}{s^{(0)}(\tau, x; \gamma_0, \beta_0)} \Lambda_0(dx),$$

and

$$D_n(t; \gamma, \beta) = \int_0^t \frac{S_n^{(1)}(\tau, x; \gamma, \beta)}{S_n^{(0)}(\tau, x; \gamma, \beta)} \hat{\Lambda}_n(dx; \gamma, \beta) = \frac{1}{n} \int_0^t \frac{S_n^{(1)}(\tau, x; \gamma, \beta)}{S_n^{(0)}(\tau, x; \gamma, \beta)^2} \bar{N}(\tau, dx).$$

By Theorem 1, Lemma 2, and a continuous mapping argument, it is readily seen that $D_n(\cdot; \hat{\gamma}_n, \hat{\beta}_n)$ is a consistent estimator of D . Note that $D_n(t; \gamma, \beta)$ defined above is the derivative of $\hat{\Lambda}_n(t; \gamma, \beta)$ in (6) with respect to θ , and it will naturally appear in the estimator of the asymptotic variance of $\hat{\Lambda}_n(t; \hat{\gamma}_n, \hat{\beta}_n)$ when the delta method is applied. With the regularity conditions (i)–(iii), the weak convergence results of $\hat{\theta}_n$ and $\hat{\Lambda}_n(t; \hat{\gamma}_n, \hat{\beta}_n)$ are stated below.

Theorem 2. *Under regularity conditions (i)–(iii), we have*

(i) *For the parametric part, $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotic normal with mean zero and variance*

$$\Sigma_\theta = \left(\int_0^\tau \frac{s^{(2)}(\tau, x; \gamma_0, \beta_0) s^{(0)}(\tau, x; \gamma_0, \beta_0) - s^{(1)}(\tau, x; \gamma_0, \beta_0)^{\otimes 2}}{s^{(0)}(\tau, x; \gamma_0, \beta_0)} \Lambda_0(dx) \right)^{-1},$$

which can be consistently estimated by

$$\hat{\Sigma}_\theta = \left(\frac{1}{n} \int_0^\tau V_n(\tau, x; \hat{\gamma}_n, \hat{\beta}_n) \bar{N}(\tau, dx) \right)^{-1},$$

where V_n is given in Equation 8.

(ii) *For the nonparametric part, $\{\sqrt{n}(\hat{\Lambda}_n(t; \hat{\gamma}_n, \hat{\beta}_n) - \Lambda_0(t))\}$ converges weakly to a mean-zero Gaussian process with covariance function*

$$\Sigma_\Lambda(t_1, t_2) = \int_0^{t_1 \wedge t_2} \frac{\Lambda_0(dx)}{s^{(0)}(\tau, x; \gamma_0, \beta_0)} + D(t_1; \gamma_0, \beta_0, \Lambda_0)^\top \Sigma_\theta D(t_2; \gamma_0, \beta_0, \Lambda_0), \tag{9}$$

and the covariance function can be consistently estimated by

$$\hat{\Sigma}_\Lambda(t_1, t_2) = \frac{1}{n} \int_0^{t_1 \wedge t_2} \frac{\bar{N}(\tau, dx)}{S_n^{(0)}(\tau, x; \hat{\gamma}_n, \hat{\beta}_n)^2} + D_n(t_1; \hat{\gamma}_n, \hat{\beta}_n)^\top \hat{\Sigma}_\theta D_n(t_2; \hat{\gamma}_n, \hat{\beta}_n). \tag{10}$$

3.2 | Asymptotic efficiency of the estimators

We show that the estimators derived above are asymptotically efficient by deriving the efficient score functions for both the parametric and the nonparametric parts. For this purpose, we consider a generic point (θ, Λ) in the parameter space, where $\theta = (\gamma, \beta^\top)^\top$. Consider the log-likelihood function (4) based on the observations from one system. The score function for the parameter θ can be derived by taking derivative on the log-likelihood with respect to θ , which can be neatly written as

$$\dot{\ell}_1(\theta, \Lambda) = \frac{\partial}{\partial \theta} \ell(\tau, \tau; \gamma, \beta, \Lambda) = \int_0^\tau \begin{pmatrix} \alpha(\tilde{N}(u^-) + 1) \\ \mathbf{X} \end{pmatrix} M(du, \tau; \gamma, \beta, \Lambda). \quad (11)$$

To derive the tangent space for Λ , consider a parametric path $\epsilon \mapsto \lambda_{\epsilon, h}$ through λ with

$$\partial \log \lambda_{\epsilon, h}(t) / \partial \epsilon|_{\epsilon=0} = h(t)\lambda(t),$$

where $h \in \ell^\infty[0, \tau]$, the space of bounded functions over $[0, \tau]$. For this parametric submodel, as we have already done in Section 2.2, the score function for the nonparametric part with respect to this submodel is

$$B_{\theta, \Lambda} h = \frac{\partial}{\partial \epsilon} \Big|_{\epsilon=0} \ell(\gamma, \beta, \Lambda_{\epsilon, h}) = \int_0^\tau h(x) M(\tau, dx; \gamma, \beta, \Lambda) = \int_0^\tau h(R(u)) M(du, \tau; \gamma, \beta, \Lambda), \quad (12)$$

where $B_{\theta, \Lambda} : \ell^\infty[0, \tau] \mapsto L_2^0(P_{\theta, \Lambda})$ is a linear operator, and $L_2^0(P_{\theta, \Lambda})$ is the space of mean-zero functions with finite second moments under $P_{\theta, \Lambda}$. A similar result could be found in proposition 1 of Pena et al. (2001).

We begin with the efficiency of the parametric estimator $\hat{\theta}_n$. Notice that the efficient score for θ is given by $\tilde{\ell}_1(\theta, \Lambda) = \dot{\ell}_1(\theta, \Lambda) - \Pi_{\theta, \Lambda} \dot{\ell}_1(\theta, \Lambda)$, where $\Pi_{\theta, \Lambda} \dot{\ell}_1(\theta, \Lambda)$ is the orthogonal projection of $\dot{\ell}_1(\theta, \Lambda)$ onto the closure of the linear space spanned by $B_{\theta, \Lambda} h$ for $h \in \ell^\infty[0, \tau]$. In the following, we show that $B_{\theta, \Lambda} \mathbf{h}^*$ is the projection of $\dot{\ell}_1(\theta, \Lambda)$ onto the nuisance space, in which

$$\mathbf{h}^*(x) = \frac{s^{(1)}(\tau, x; \gamma, \beta)}{s^{(0)}(\tau, x; \gamma, \beta)}. \quad (13)$$

According to Lemma 2, $\mathbf{h}^* \in \ell^\infty[0, \tau]$ because $s^{(1)}(\tau, x; \gamma, \beta)$ is uniformly bounded, and $s^{(0)}(\tau, x; \gamma, \beta)$ is bounded from zero. Next, we show that \mathbf{h}^* satisfies the following orthogonal condition

$$E_{\theta, \Lambda} [B_{\theta, \Lambda} h \cdot (\dot{\ell}_1(\theta, \Lambda) - B_{\theta, \Lambda} \mathbf{h}^*)] = \mathbf{0}_{p+1},$$

for all $h \in \ell^\infty[0, \tau]$. According to the score functions for the parameter part θ and nonparametric part Λ in Equations 11 and (12), the left-hand side of the above display can be expressed as

$$E_{\theta, \Lambda} \int_0^\tau h(R(u)) \left[\begin{pmatrix} \alpha(\tilde{N}(u^-) + 1) \\ \mathbf{X} \end{pmatrix} - \mathbf{h}^*(R(u)) \right] A(du, \tau; \gamma, \beta, \Lambda). \quad (14)$$

Use proposition 1 in Pena et al. (2001) to see that the first term in the expectation is

$$\int_0^\tau h(R(u)) \begin{pmatrix} \alpha(\tilde{N}(u^-) + 1) \\ \mathbf{X} \end{pmatrix} A(du, \tau; \gamma, \boldsymbol{\beta}, \Lambda) = \int_0^\tau h(x) \begin{pmatrix} e^{\boldsymbol{\beta}^\top \mathbf{X} Y^{(1)}(\tau, x; \gamma)} \\ e^{\boldsymbol{\beta}^\top \mathbf{X} Y(\tau, x; \gamma) \mathbf{X}} \end{pmatrix} \Lambda(dx),$$

and the second term is

$$\int_0^\tau h(R(u)) \mathbf{h}^*(R(u)) A(du, \tau; \gamma, \boldsymbol{\beta}, \Lambda) = \int_0^\tau h(x) \mathbf{h}^*(x) e^{\boldsymbol{\beta}^\top \mathbf{X} Y(\tau, x; \gamma)} \Lambda(dx).$$

Notice that

$$E \left[\begin{pmatrix} e^{\boldsymbol{\beta}^\top \mathbf{X} Y^{(1)}(\tau, x; \gamma)} \\ e^{\boldsymbol{\beta}^\top \mathbf{X} Y(\tau, x; \gamma) \mathbf{X}} \end{pmatrix} \right] = s^{(1)}(\tau, x; \gamma, \boldsymbol{\beta}),$$

and

$$E[e^{\boldsymbol{\beta}^\top \mathbf{X} Y(\tau, x; \gamma)}] = s^{(0)}(\tau, x; \gamma, \boldsymbol{\beta}).$$

Therefore, (14) is equal to

$$\int_0^\tau h(x) (s^{(1)}(\tau, x; \gamma, \boldsymbol{\beta}) - \mathbf{h}^*(x) s^{(0)}(\tau, x; \gamma, \boldsymbol{\beta})) \Lambda(dx).$$

With \mathbf{h}^* given in (13), it is readily seen that the above display equals $\mathbf{0}_{p+1}$ for any $h \in \ell^\infty[0, \tau]$. That is, the function \mathbf{h}^* satisfies the orthogonal condition (14). and thus, the efficient score for $\boldsymbol{\theta}$ is

$$\begin{aligned} \tilde{\ell}_1(\boldsymbol{\theta}, \Lambda) &= \dot{\ell}_1(\boldsymbol{\theta}, \Lambda) - \Pi_{\boldsymbol{\theta}, \Lambda} \dot{\ell}_1(\boldsymbol{\theta}, \Lambda) \\ &= \int_0^\tau \left[\begin{pmatrix} \alpha(\tilde{N}(u^-) + 1) \\ \mathbf{X} \end{pmatrix} - \frac{s^{(1)}(\tau, R(u); \gamma, \boldsymbol{\beta})}{s^{(0)}(\tau, R(u); \gamma, \boldsymbol{\beta})} \right] M(du, \tau; \gamma, \boldsymbol{\beta}, \Lambda), \end{aligned}$$

and the efficient information is given by

$$\tilde{I}_{\boldsymbol{\theta}, \Lambda} = E[\tilde{\ell}_1(\boldsymbol{\theta}, \Lambda) \tilde{\ell}_1(\boldsymbol{\theta}, \Lambda)^\top] = \int_0^\tau \frac{s^{(2)}(\tau, x; \gamma, \boldsymbol{\beta}) \cdot s^{(0)}(\tau, x; \gamma, \boldsymbol{\beta}) - s^{(1)}(\tau, x; \gamma, \boldsymbol{\beta})^2}{s^{(0)}(\tau, x; \gamma, \boldsymbol{\beta})} \Lambda(dx). \tag{15}$$

Compare (15) with the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_n$ in Theorem 2 to see that the estimator $\hat{\boldsymbol{\theta}}_n$ is asymptotically efficient.

Next, we verify the asymptotic efficiency of the nonparametric estimator. Recall that the statistical model is $\mathcal{P} = \{P_{\boldsymbol{\theta}, \Lambda} : \boldsymbol{\theta} \in \mathbb{R}^{p+1}, \Lambda \in \mathcal{M}\}$ where \mathcal{M} is the class of nondecreasing and right-continuous functions with $\Lambda(0) = 0$. Define a map $\psi : \mathcal{P} \mapsto \mathcal{M}$ as $\psi(P_{\boldsymbol{\theta}, \Lambda}) = \Lambda$. According to theorem 25.48 of Van der Vaart (2000), verification of the asymptotic efficiency of $\hat{\Lambda}(\cdot)$ at $P_{\boldsymbol{\theta}, \Lambda}$ is equivalent to verifying that

- (i) the map $\psi : \mathcal{P} \mapsto \ell^\infty[0, \tau]$ is differentiable at $P_{\boldsymbol{\theta}, \Lambda}$ in the sense of Van der Vaart (2000); and
- (ii) For each fixed t , the estimator $\hat{\Lambda}(t)$ is asymptotically efficient at $P_{\boldsymbol{\theta}, \Lambda}$ for estimating $\Lambda(t)$ for every $t \in [0, \tau]$.

First, we prove that ψ is differentiable. Consider the submodel $\{P_\epsilon, \epsilon \in \mathcal{E}\}$ defined by $P_\epsilon = P_{\theta+\epsilon a, \Lambda_{\epsilon, h}}$, where $a \in \mathbb{R}^{p+1}$, $\Lambda_{\epsilon, h}(t) = \int_0^t (1 + \epsilon h(u)) \Lambda(du)$, $h(t) \in \mathcal{L}^\infty[0, \tau]$ and $\mathcal{E} = \{\epsilon : \epsilon \geq 0, 1 + \epsilon h(t) > 0, \forall t \in [0, \tau]\}$. Hence the score function of this submodel is $g(a, h) = a^\top \dot{\ell}_1(\theta, \Lambda) + B_{\theta, \Lambda} h$, and the tangent space $\dot{P}_{\theta, \Lambda}$ is the collection of all score functions in this form. Using the definition in Van der Vaart (2000), ψ is differentiable at $P_{\theta, \Lambda}$ relative to a given tangent set $\dot{P}_{\theta, \Lambda}$ if there exists a continuous linear map $\dot{\psi}_{\theta, \Lambda} : L_2^0(P_{\theta, \Lambda}) \mapsto \mathcal{L}^\infty[0, \tau]$, such that for every $g \in \dot{P}_{\theta, \Lambda}$ and a submodel $\epsilon \mapsto P_\epsilon$ with score function g ,

$$\frac{\psi(P_\epsilon) - \psi(P)}{\epsilon} \rightarrow \dot{\psi}_{\theta, \Lambda} g. \tag{16}$$

We show that the linear operator $\dot{\psi}_{\theta, \Lambda} : L_2^0(P_{\theta, \Lambda}) \mapsto \mathcal{L}^\infty[0, \tau]$ pointwisely defined by $\dot{\psi}_{\theta, \Lambda}(g)(t) = \langle \tilde{\psi}_{\theta, \Lambda}(t), g \rangle_{P_{\theta, \Lambda}}$ satisfies the above display, where

$$\tilde{\psi}_{\theta, \Lambda}(t) = \int_0^t \frac{M(\tau, dx; \gamma, \beta, \Lambda)}{s^{(0)}(\tau, x; \gamma, \beta)} - \left[\int_0^t \frac{s^{(1)}(\tau, x; \gamma, \beta)}{s^{(0)}(\tau, x; \gamma, \beta)} \Lambda(dx) \right]^\top \tilde{I}_{\theta, \Lambda}^{-1} \tilde{\ell}_1(\theta, \Lambda).$$

It is obvious that $\dot{\psi}_{\theta, \Lambda}$ defined above is a linear operator on $L_2^0(P_{\theta, \Lambda})$. In the following, we will first show that it is a continuous operator, and $\dot{\psi}_{\theta, \Lambda}(g)$ equals $\int_0^t h(x) \Lambda(dx)$, the limit of the left side of (16). Use the Cauchy–Schwarz inequality to see that

$$\dot{\psi}_{\theta, \Lambda}(g)(t) = \langle \tilde{\psi}_{\theta, \Lambda}(t), g \rangle_{P_{\theta, \Lambda}} \leq \|\tilde{\psi}_{\theta, \Lambda}(t)\|_{L_2^0(P_{\theta, \Lambda})} \|g\|_{L_2^0(P_{\theta, \Lambda})}.$$

Using proposition 1 in Pena et al. (2001), the $L_2(P_{\theta, \Lambda})$ -norm of $\tilde{\psi}_{\theta, \Lambda}(t)$ can be expressed as

$$\left\{ \int_0^t \frac{\Lambda(dx)}{s^{(0)}(\tau, x; \gamma, \beta)} + \left[\int_0^t \frac{s^{(1)}(\tau, x; \gamma, \beta)}{s^{(0)}(\tau, x; \gamma, \beta)} \Lambda(dx) \right]^\top \tilde{I}_{\theta, \Lambda}^{-1} \left[\int_0^t \frac{s^{(1)}(\tau, x; \gamma, \beta)}{s^{(0)}(\tau, x; \gamma, \beta)} \Lambda(dx) \right] \right\}^{1/2},$$

which is bounded above by

$$\left\{ \int_0^\tau \frac{\Lambda(dx)}{|s^{(0)}(\tau, x; \gamma, \beta)|} + \int_0^\tau \left\| \frac{s^{(1)}(\tau, x; \gamma, \beta)}{s^{(0)}(\tau, x; \gamma, \beta)} \right\|^2 \Lambda(dx) \|\tilde{I}_{\theta, \Lambda}^{-1}\|_2 \right\}^{1/2} \equiv M,$$

where $\|\tilde{I}_{\theta, \Lambda}^{-1}\|_2$ is the largest eigenvalue of $\tilde{I}_{\theta, \Lambda}^{-1}$, which is bounded according to assumption (iii). Besides, according to assumptions (i), (ii), and Lemma 2, we also have $\Lambda(\tau) < \infty$ and $s^{(k)}(\tau, x; \gamma, \beta) < \infty$ for all $x \in [0, \tau]$, which implies that $M < \infty$. Notice that the bound M does not depend on t . Hence we have $\sup_{t \in [0, \tau]} |\dot{\psi}_{\theta, \Lambda}(g)(t)| \leq M \|g\|_{L_2(P_{\theta, \Lambda})}$, which implies that $\dot{\psi}_{\theta, \Lambda}(g)$ is a bounded operator, and thus it is a continuous operator by theorem 1.32 in Rudin (1991).

Next, we show that for every $g(a, h) = a^\top \dot{\ell}_1(\theta, \Lambda) + B_{\theta, \Lambda} h \in \dot{P}_{\theta, \Lambda}$, the linear operator $\dot{\psi}_{\theta, \Lambda}$ satisfies $\dot{\psi}_{\theta, \Lambda}(g) = \int_0^t h(x) \Lambda(dx)$, that is, $\langle \tilde{\psi}_{\theta, \Lambda}(t), g \rangle_{P_{\theta, \Lambda}} = \int_0^t h(x) \Lambda(dx)$ for every $t \in [0, \tau]$. This is equivalent to show that

$$\begin{aligned} \mathbb{E}_{\theta, \Lambda} [\tilde{\psi}_{\theta, \Lambda}(t) \dot{\ell}_1(\theta, \Lambda)] &= \mathbf{0}_{(p+1) \times 1}, \\ \mathbb{E}_{\theta, \Lambda} [\tilde{\psi}_{\theta, \Lambda}(t) B_{\theta, \Lambda} h] &= \int_0^t h(x) \Lambda(dx). \end{aligned} \tag{17}$$

To derive the first equation, we use Equation 12 to see that

$$\int_0^t \frac{M(\tau, dx; \gamma, \beta, \Lambda)}{s^{(0)}(\tau, x; \gamma, \beta)} = \int_0^\tau \frac{I\{x \leq t\} M(\tau, dx; \gamma, \beta, \Lambda)}{s^{(0)}(\tau, x; \gamma, \beta)} = B_{\theta, \Lambda} \frac{I\{\cdot \leq t\}}{s^{(0)}(\tau, \cdot; \gamma, \beta)}.$$

Therefore, the left side of the first equation of (17) is given by

$$\begin{aligned} & E_{\theta, \Lambda} [\tilde{\psi}_{\theta, \Lambda}(t) \dot{\ell}_1(\theta, \Lambda)] \\ &= E_{\theta, \Lambda} \left[\left(\int_0^t \frac{M(\tau, dx; \gamma, \beta, \Lambda)}{s^{(0)}(\tau, x; \gamma, \beta)} - \left[\int_0^t \frac{s^{(1)}(\tau, x; \gamma, \beta)}{s^{(0)}(\tau, x; \gamma, \beta)} \Lambda(dx) \right]^\top \cdot \tilde{I}_{\theta, \Lambda}^{-1} \tilde{\ell}_1(\theta, \Lambda) \right) \cdot \dot{\ell}_1(\theta, \Lambda) \right] \\ &= E_{\theta, \Lambda} \left[B_{\theta, \Lambda} \frac{I\{\cdot \leq t\}}{s^{(0)}(\tau, x; \gamma, \beta)} \cdot \dot{\ell}_1(\theta, \Lambda) \right] - E_{\theta, \Lambda} \left[\tilde{\ell}_1(\theta, \Lambda) \dot{\ell}_1(\theta, \Lambda)^\top \right] \cdot \tilde{I}_{\theta, \Lambda}^{-1} \cdot \int_0^t \frac{s^{(1)}(\tau, x; \gamma, \beta)}{s^{(0)}(\tau, x; \gamma, \beta)} \Lambda(dx). \end{aligned}$$

Note that $\dot{\ell}_1(\theta, \Lambda) = \tilde{\ell}_1(\theta, \Lambda) + \Pi_{\theta, \Lambda} \dot{\ell}_1(\theta, \Lambda)$, and $E_{\theta, \Lambda} \tilde{\ell}_1(\theta, \Lambda) B_{\theta, \Lambda} h = \mathbf{0}_{(p+1) \times 1}$. Hence, the above display is equal to

$$\begin{aligned} & E_{\theta, \Lambda} \left[B_{\theta, \Lambda} \frac{I\{\cdot \leq t\}}{s^{(0)}(\tau, \cdot; \gamma, \beta)} \cdot \Pi_{\theta, \Lambda} \dot{\ell}_1(\theta, \Lambda) \right] - \tilde{I}_{\theta, \Lambda} \tilde{I}_{\theta, \Lambda}^{-1} \cdot \int_0^t \frac{s^{(1)}(\tau, x; \gamma, \beta)}{s^{(0)}(\tau, x; \gamma, \beta)} \Lambda(dx) \\ &= E_{\theta, \Lambda} \left[\int_0^\tau \frac{I\{x \leq t\} M(\tau, dx; \gamma, \beta, \Lambda)}{s^{(0)}(\tau, x; \gamma, \beta)} \cdot \int_0^\tau \frac{s^{(1)}(\tau, x; \gamma, \beta) M(\tau, dx; \gamma, \beta, \Lambda)}{s^{(0)}(\tau, x; \gamma, \beta)} \right] \\ &\quad - \int_0^t \frac{s^{(1)}(\tau, x; \gamma, \beta)}{s^{(0)}(\tau, x; \gamma, \beta)} \Lambda(dx) \\ &= \mathbf{0}_{p+1}, \end{aligned}$$

which yields the first half of (17).

In order to verify the second half of (17), recall that $E_{\theta, \Lambda} \tilde{\ell}_1(\theta, \Lambda) B_{\theta, \Lambda} h = \mathbf{0}_{p+1}$. Thus, we have

$$\begin{aligned} & E_{\theta, \Lambda} \tilde{\psi}_{\theta, \Lambda}(t) B_{\theta, \Lambda} h \\ &= E_{\theta, \Lambda} \left[\left(\int_0^t \frac{M(\tau, dx; \gamma, \beta, \Lambda)}{s^{(0)}(\tau, x; \gamma, \beta)} - \left[\int_0^t \frac{s^{(1)}(\tau, x; \gamma, \beta)}{s^{(0)}(\tau, x; \gamma, \beta)} \Lambda(dx) \right]^\top \cdot \tilde{I}_{\theta, \Lambda}^{-1} \tilde{\ell}_1(\theta, \Lambda) \right) \cdot B_{\theta, \Lambda} h \right] \\ &= E_{\theta, \Lambda} \left[\int_0^t \frac{M(\tau, dx; \gamma, \beta, \Lambda)}{s^{(0)}(\tau, x; \gamma, \beta)} \cdot \int_0^\tau h(x) M(\tau, dx; \gamma, \beta, \Lambda) \right] = \int_0^t h(x) \Lambda(dx). \end{aligned}$$

Therefore, we have shown that the two equations in (17) hold. This implies that $\tilde{\psi}_{\theta, \Lambda}(g) = \int_0^t h(x) \Lambda(dx)$ for every $g \in \dot{P}_{\theta, \Lambda}$. It is readily seen that $\tilde{\psi}_{\theta, \Lambda}$ is a continuous linear operator. As a result, ψ is differentiable relative to $\dot{P}_{\theta, \Lambda}$ at $P_{\theta, \Lambda}$.

Next, we show that $\hat{\Lambda}(t)$ is asymptotically efficient at $P_{\theta, \Lambda}$ for estimating $\psi(P_{\theta, \Lambda})(t)$. Notice that $\Lambda(t)$ can be considered as a parameter of interest for a fixed t . We use a similar argument when proving the efficiency of parametric part based on the differentiability of ψ , which implies that $\psi(\cdot)(t)$ is differentiable at $P_{\theta, \Lambda}$. Following lemma 25.19 and relevant definitions in Van der Vaart (2000), it is obvious that the efficient influence function for $\Lambda(t)$ is $\tilde{\psi}_{\theta, \Lambda}(t)$. Thus, the efficient information for $\Lambda(t)$ is given by

$$E_{\theta, \Lambda} (\tilde{\psi}_{\theta, \Lambda}(t))^2 = \int_0^t \frac{\Lambda(dx)}{s^{(0)}(\tau, x; \gamma, \beta)} + \left[\int_0^t \frac{s^{(1)}(\tau, x; \gamma, \beta)}{s^{(0)}(\tau, x; \gamma, \beta)} \Lambda(dx) \right]^2 \tilde{I}_{\theta, \Lambda}^{-1}.$$

Comparing the above display with the covariance function $\Sigma_{\Lambda}(t, t)$ stated in Theorem 2, we can conclude that the estimator $\hat{\Lambda}_n(t)$ for a fixed t with $t \in [0, \tau]$ is asymptotically efficient. So far, we have verified all conditions in theorem 25.48 of Van der Vaart (2000). As a result, the nonparametric estimator $\hat{\Lambda}_n(\cdot)$ is asymptotically efficient at $P_{\theta, \Lambda}$.

4 | SIMULATION

Simulations are conducted to assess the performance of the proposed semiparametric framework. Because the traditional method in Dauxois et al. (2019) is only applicable to a fixed number of failures, Section 4.1 first considers the failure censoring case for the comparison purpose. In Section 4.2, the time censoring setting is then considered and some modifications are made to the traditional method in order to make it applicable. At last, Section 4.3 considers the scenario with covariates.

4.1 | Scenario without covariates (failure censoring)

In this subsection, we compare the proposed method with the traditional method in the case of failure censoring. The covariates are not considered here as the traditional method is not able to incorporate the covariate information. In specific, let T_{ij} be the interoccurrence time between the $(j - 1)$ th and j th failures in the i th system, where $i = 1, \dots, n$ and $j = 1, \dots, n_i$. Define

$$T_{\inf} = \max_{i=1, \dots, n} \min_{j=1, \dots, n_i} T_{ij} \quad \text{and} \quad T_{\sup} = \min_{i=1, \dots, n} \max_{j=1, \dots, n_i} T_{ij}. \quad (18)$$

To avoid $T_{\inf} > T_{\sup}$, let

$$T_{\sup} = \max \left\{ \min_{i=1, \dots, n} \max_{j=1, \dots, n_i} T_{ij}, (T_{\inf} + 0.1) \right\}. \quad (19)$$

In Dauxois et al. (2019), two estimators of γ have been proposed and the one with better performance is considered here. In specific, the estimator of γ is given by

$$\hat{\gamma}_1 = \frac{\sum_{j=2}^N \alpha(j) \int_{T_{\inf}}^{T_{\sup}} (\ln \hat{\Lambda}_{T_j}(t) - \ln \hat{\Lambda}_{T_1}(t)) dt}{(T_{\sup} - T_{\inf}) \sum_{j=2}^N \alpha^2(j)}, \quad (20)$$

where $N = \max\{n_1, \dots, n_n\}$, and $\hat{\Lambda}_{T_j}(t)$ is the empirical cumulative hazard function of the j th interoccurrence time T_j .

The simulation settings are as follows. We set the baseline cumulative intensity function as $\Lambda_0(t) = t^2$, which corresponds to a Weibull distribution with shape parameter 2 and scale parameter 1. In addition, let the fixed number of the failures $N = 5$ and $\alpha(j) = j - 1$. To investigate the performance under a variety of scenarios, let $\gamma \in \{0.1, 0.5, 1\}$ and $n \in \{20, 50, 100\}$. We first consider the accuracy in estimating the parameter γ . Based on 2000 replications, the estimated absolute biases and root mean square errors (rMSEs) under each setting are shown in Figure 1. As seen, the proposed ML estimator performs satisfactorily in all the scenarios and it outperforms the estimators by Dauxois et al. (2019). We then consider the nonparametric estimation part. The

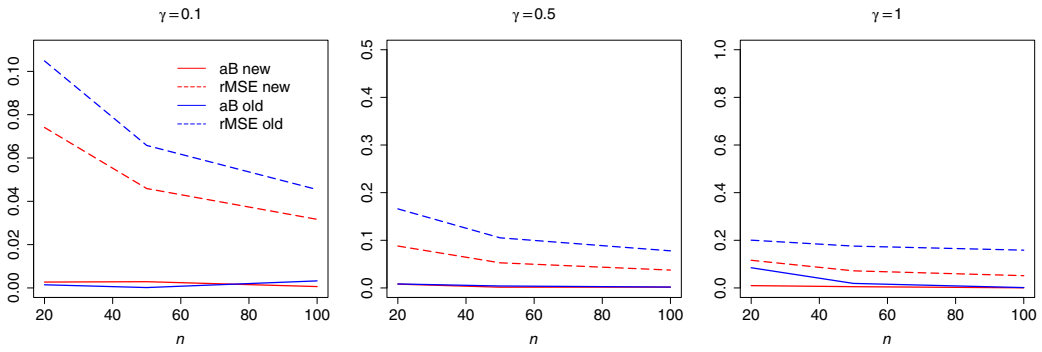


FIGURE 1 Absolute bias (aB) and root mean square error (rMSE) in estimating γ based on the proposed method (new) and the traditional method (old) by considering $n = 20, 50,$ and 100 under failure censoring [Colour figure can be viewed at wileyonlinelibrary.com]

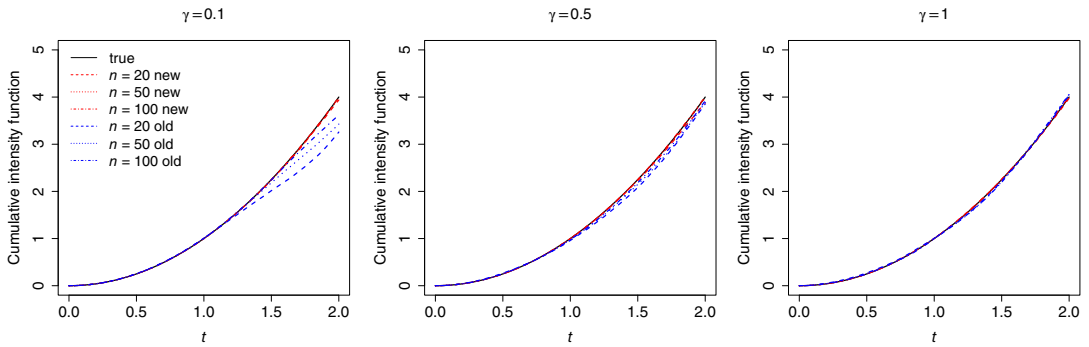


FIGURE 2 The true baseline cumulative intensity function (true) and the estimated ones by the proposed method (new) and the traditional method (old) by considering $n = 20, 50,$ and 100 under failure censoring [Colour figure can be viewed at wileyonlinelibrary.com]

mean values of 2000 estimated baseline cumulative intensity functions are shown in Figure 2, where the true curves are also plotted. Generally, the estimated cumulative intensity functions based on the proposed method match very well with the true curves in all the scenarios, while the estimated curves by the traditional method perform not so well when $\gamma = 0.1$.

4.2 | Scenario without covariates (time censoring)

In this subsection, we make a comparison between the proposed method and the traditional method under the time censoring mechanism, which is a more practical setting in real applications. In order to implement the traditional method, Equations 18–(20) are again used. Because the data are time censored, this treatment essentially means the time between the last repair and the censoring time is discarded, and only the interoccurrence times are utilized.

The censoring time C is assumed to follow the exponential distribution with mean 5 and the other settings are the same as those in Section 4.1. Figures 3 and 4, respectively, show the performance in estimating the parameter γ and the baseline cumulative intensity function. As seen,

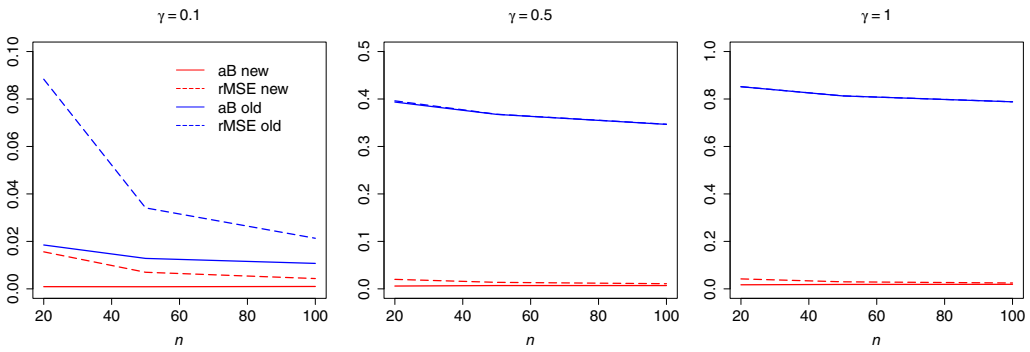


FIGURE 3 Absolute bias (aB) and root mean square error (rMSE) in estimating γ based on the proposed method (new) and the traditional method (old) by considering $n = 20, 50,$ and 100 and Weibull shape 2 and scale 1 under time censoring [Colour figure can be viewed at [wileyonlinelibrary.com](#)]

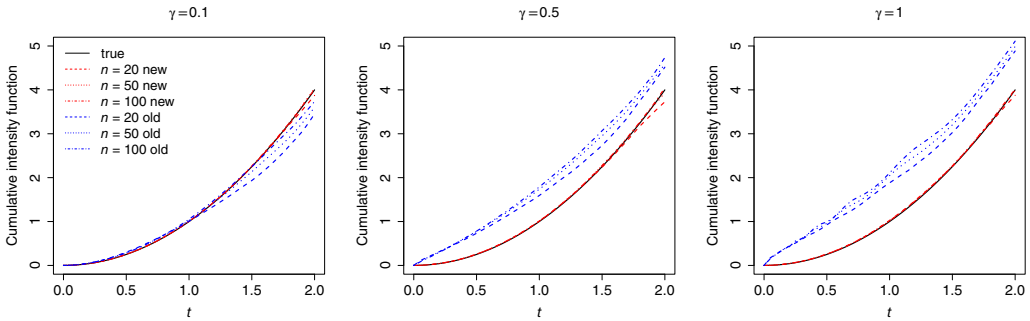


FIGURE 4 The true baseline cumulative intensity function (true) and the estimated ones by the proposed method (new) and the traditional method (old) by considering $n = 20, 50,$ and 100 and Weibull shape 2 and scale 1 under time censoring [Colour figure can be viewed at [wileyonlinelibrary.com](#)]

the proposed method achieves high estimation accuracies in all the combinations of n and γ and it outperforms the traditional method by a large margin. The poor performance of the traditional method is not surprising as its implementation requires a truncation of the original data and hence the information is not fully utilized.

To further validate our method, we then consider a different baseline cumulative intensity function $\Lambda_0(t) = \sqrt{t}/3$, which corresponds to a Weibull distribution with shape parameter 0.5 and scale parameter 3, and the simulations results are shown in Figures 5 and 6. As expected, the proposed method is more accurate in estimating γ . More importantly, the nonparametric estimators based on the proposed method tally well with the true curves in all the scenarios, while the estimated curves by the traditional method deviate significantly from the true values.

4.3 | Scenario with covariates

In this subsection, we investigate the performance of the proposed methods when the covariates are incorporated. We consider

$$\beta_0^\top \mathbf{X} = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3,$$

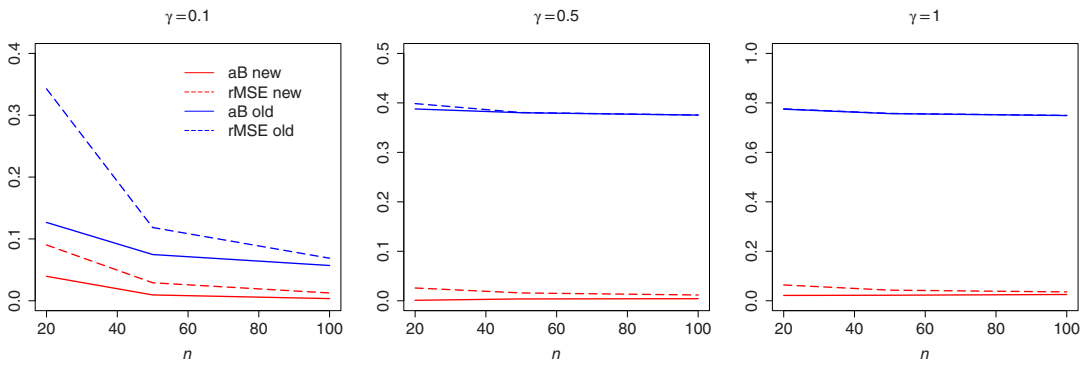


FIGURE 5 Absolute bias (aB) and root mean square error (rMSE) in estimating γ based on the proposed method (new) and the traditional method (old) by considering $n = 20, 50$, and 100 and Weibull shape 0.5 and scale 3 under time censoring [Colour figure can be viewed at wileyonlinelibrary.com]

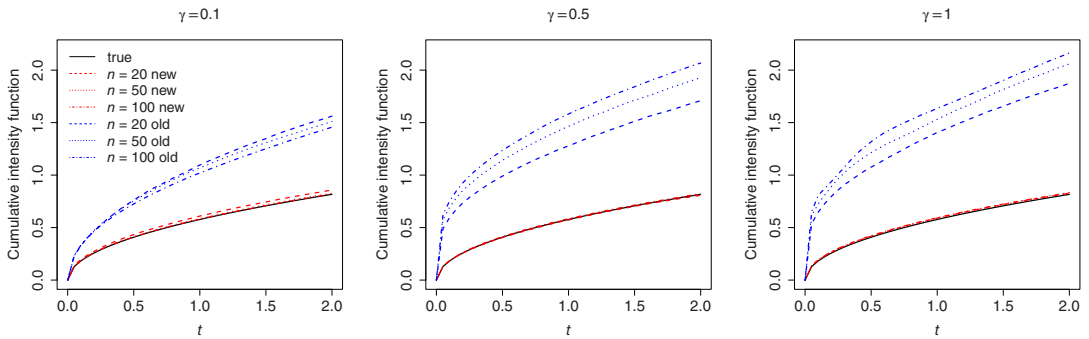


FIGURE 6 The true baseline cumulative intensity function (true) and the estimated ones by the proposed method (new) and the traditional method (old) by considering $n = 20, 50$, and 100 and Weibull shape 0.5 and scale 3 under time censoring [Colour figure can be viewed at wileyonlinelibrary.com]

where $(\beta_1, \beta_2, \beta_3)^T = (0.5, -1, 0.1)$, and X_1, X_2 , and X_3 , respectively, follow the standard normal distribution, the standard exponential distribution and the standard uniform distribution. All the other settings are the same as those in Section 4.1.

We first consider estimating the parameters $\theta_0 = (\gamma, \beta_1, \beta_2, \beta_3)^T$. Figures 7 and 8, respectively, show the estimated absolute biases and rMSEs based on 2000 replications by considering $n = 20, 50$ and 100 . As seen, all the parameters can be satisfactorily estimated and the estimation accuracy generally improves with the sample size n . On the other hand, the estimated baseline cumulative intensity function as well as the true values are plotted in Figure 9. It is observed that the nonparametric part can also be well estimated by the proposed method and when $n \geq 50$, the estimated baseline cumulative intensity functions are almost identical to the true ones.

5 | REAL DATA APPLICATION

In this section, we demonstrate the proposed models and methods by a real example from the automobile industry. The automobile manufacturer considered here is one of the largest

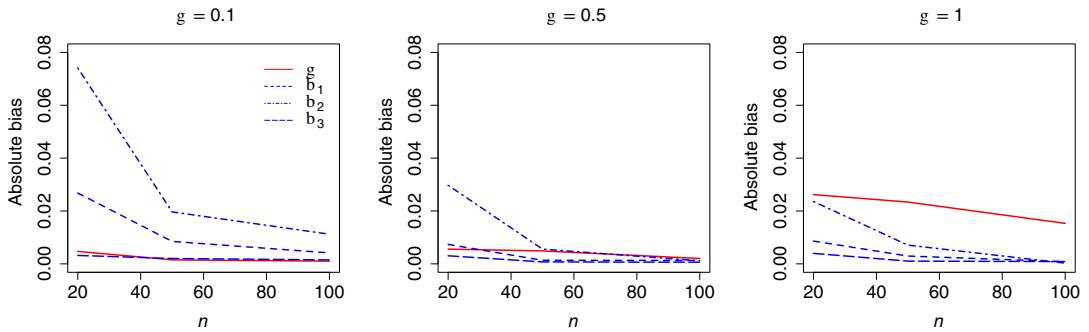


FIGURE 7 Absolute bias in estimating θ based on the proposed method by considering $n = 20, 50,$ and 100 [Colour figure can be viewed at wileyonlinelibrary.com]

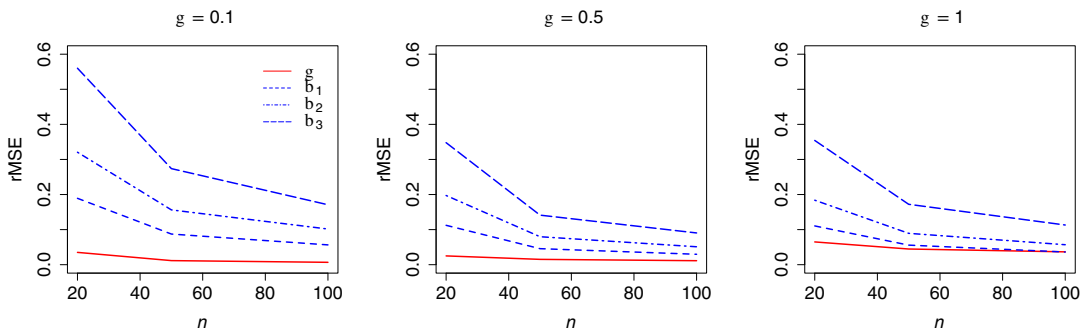


FIGURE 8 Root mean square error in estimating θ based on the proposed method by considering $n = 20, 50,$ and 100 [Colour figure can be viewed at wileyonlinelibrary.com]

automakers in the world and it has maintained a comprehensive warranty database which recorded detailed warranty claim data of different types of automobiles. The warranty claim data contain valuable information about product quality and reliability, and they can be used to, for example, give early warnings of faulty designs, predict future claims and warranty costs, and compare reliability with competing products. To achieve these, it is important to estimate the reliability of the components wisely.

As an illustrative example, the component window regulator is considered in this study. From the warranty claim database, we extract the repair records of the window regulator installed in 69 cars of the same model sold in a district, and the data are illustrated in Figure 10. Here, the red circle and the blue cross represent the date of sale and the date of each repair, respectively. To illustrate the methods, the end of study date is set as January 1, 2013, which is denoted by the red vertical line in Figure 10. Among the 69 cars, 54 cars have at least one failure record of the window regulator while the remaining 15 cars sold within the study period have no repair records. Based on the use rates and the sales dates, it is readily seen that all the cars are still within the two-dimensional warranty limit (i.e., 24 months or 60,000 miles warranty, whichever comes first) on the data-freeze date, and the censoring is caused by administrative censoring. In addition, it is observed from Figure 10 that the component failure time generally decreases with each subsequent repair, which indicates the effects brought by the imperfect repair. Thus, we use the EGFR model to fit this dataset.

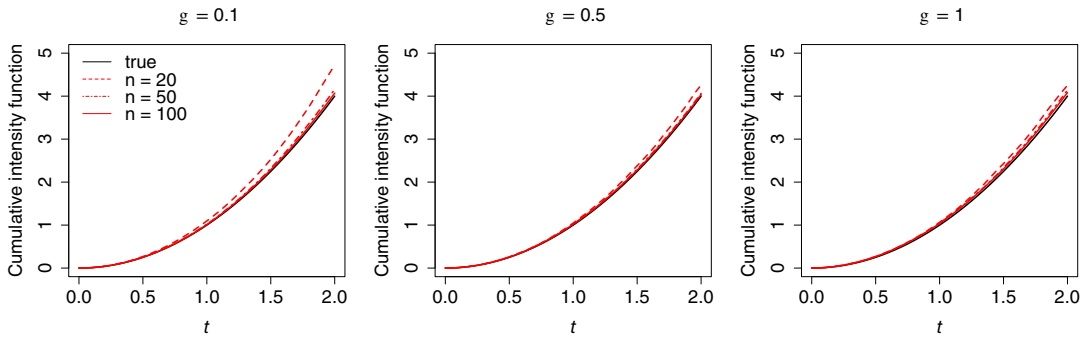


FIGURE 9 The estimated baseline cumulative intensity function based on the proposed method by considering $n = 20, 50,$ and 100 [Colour figure can be viewed at wileyonlinelibrary.com]

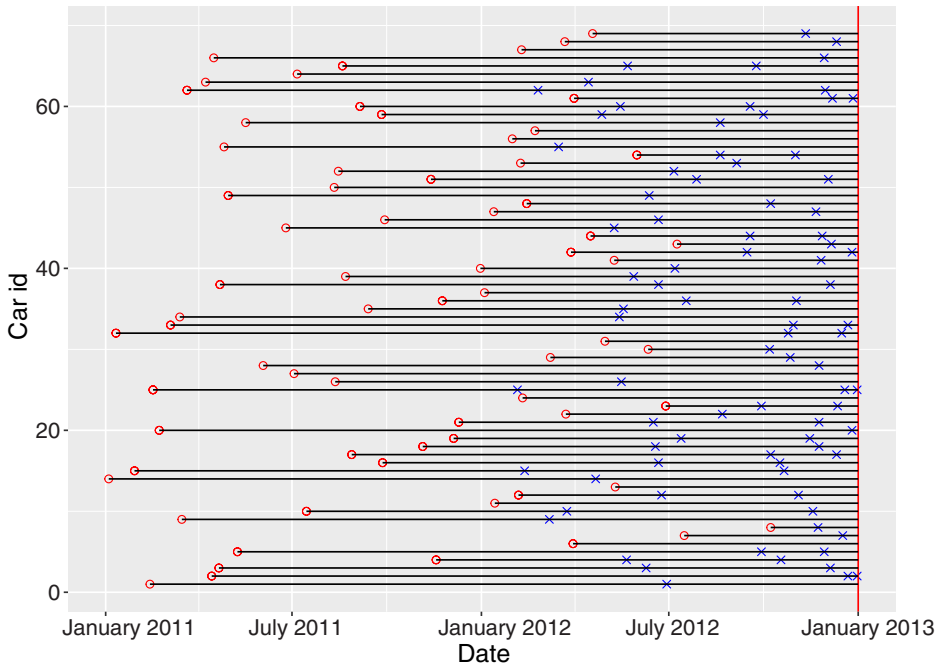


FIGURE 10 Repair records of the window regulator in 69 cars. The red circle denotes the sale date, the blue cross denotes the repair date and the red vertical line denotes the censoring time [Colour figure can be viewed at wileyonlinelibrary.com]

In the dataset, the mileage information is available and it reflects the actual use rate of the car. Hence, it is likely that the component failure rate depends on the rate at which mileage accumulates on a car. In this study, the value of usage rate, which is defined as mileage per day, was calculated for each car with an observed mileage value prior to the end-of-study date. Let $\alpha(j) = j - 1$. Based on the estimation procedures in Section 2, we obtain that $\hat{\gamma} = 0.867$ and $\hat{\beta} = 0.997$, and the respective 95% confidence intervals are $(0.391, 1.347)$ and $(0.344, 1.650)$, based on the large-sample normal approximation. These results indicate that the imperfect repair and the use rate have adverse effects on the component failure rate. In addition, Figure 11 shows the

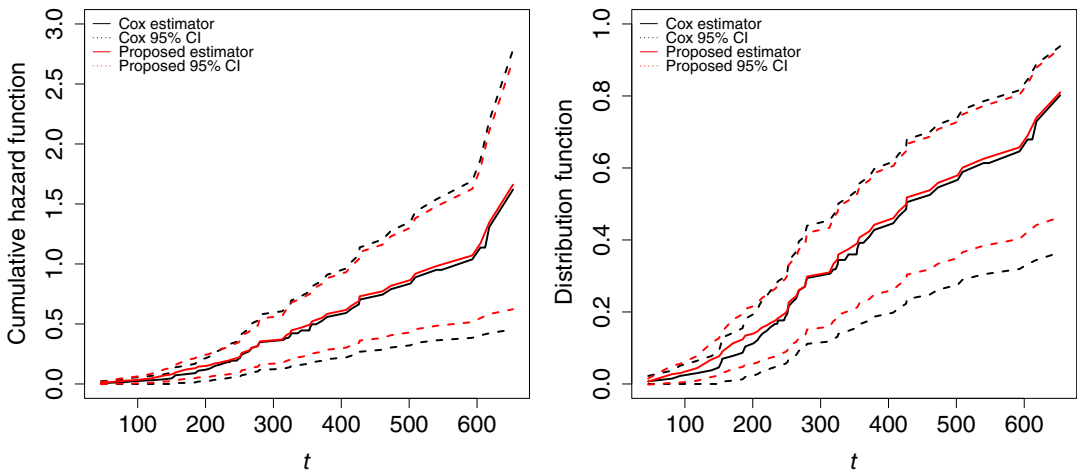


FIGURE 11 Estimated baseline cumulative hazard function and baseline distribution function of the first interoccurrence time, as well as the 95% confidence intervals based on the proposed model and the Cox proportional hazards model [Colour figure can be viewed at wileyonlinelibrary.com]

estimated baseline cumulative hazard function and the corresponding baseline distribution function of the first interoccurrence time, that is, the time between the sale date and the first repair, as well as the 95% confidence intervals based on the asymptotic results in Section 3. Note that the baseline cumulative hazard function of the first interoccurrence time is essentially equivalent to the baseline cumulative intensity function $\Lambda_0(\cdot)$. For comparison purpose, we have also applied the classical Cox proportional hazards model to fit the first interoccurrence times. This is because there is no repair effect on the first failure times, and a satisfactory baseline cumulative hazard/distribution function estimator of the first interoccurrence time should match well with that from the Cox model. In particular, the baseline cumulative hazard function based on the Cox model can be obtained by using the famous Breslow estimator and then the baseline distribution function can be readily obtained (Lin, 2007). As seen from Figure 11, the estimated baseline cumulative hazard function and distribution function of the first interoccurrence time by the proposed method tally reasonably well with the curves by the Cox model, indicating an adequate fit of the EGFPP model. In addition, it is observed that the 95% pointwise confidence intervals by the proposed method are narrower than those based on the Cox model. This is reasonable as more failure data are utilized in the proposed method and hence more accurate estimation of the component lifetime distribution is expected.

6 | CONCLUSIONS AND REMARKS

In this paper, we have proposed the semiparametric inference framework for the time-censored recurrent failure data with covariates under the EGFRR model. By careful projecting the score functions to the tangent spaces, we have shown that both the parametric and nonparametric estimators are asymptotically efficient. Extensive simulations and real examples have illustrated that the proposed framework outperforms the existing methods and it is useful to extract valuable reliability information from the recurrent failure data in presence of time censoring and covariates.

There are several research topics worth exploration. The first potential topic is about the exponent function $\alpha(j)$. In this paper as well as many other existing studies, $\alpha(j)$ is pre-given with the constraints that it is nondecreasing and $\alpha(1) = 0$. In practice, a more plausible way to determine $\alpha(j)$ is to base on the recurrent failure data. One feasible solution is to first estimate $\alpha(j)$ based on the empirical process and then verify the asymptotic properties of the estimator of $\alpha(j)$. In addition, the χ^2 test may be used to test the goodness-of-fit of $\alpha(j)$. Nevertheless, more thorough investigation is needed to look into the selection of $\alpha(j)$. Another possible research direction is to consider the statistical inference of other popular imperfect repair models with time-censoring, such as the models in Finkelstein (2008), Wu and Zuo (2010) and Alaswad and Xiang (2017). If a similar martingale process can be constructed, the proposed method in this paper can be extended to those models and the asymptotic properties may be similarly established.

ACKNOWLEDGMENTS

We are grateful to the editors and two referees for their insightful comments that have led to a substantial improvement to an earlier version of the paper. Wang and Ye were supported by the National Science Foundation of China (72071138) and Singapore MOE AcRF Tier 2 grant (R-266-000-143-112).

ORCID

Piao Chen  <https://orcid.org/0000-0001-9714-450X>

REFERENCES

- Adepedjou, A., & Stocker, R. (2015). A general class of semiparametric models for recurrent event data. *Journal of Statistical Planning and Inference*, 156, 48–63.
- Alaswad, S., & Xiang, Y. (2017). A review on condition-based maintenance optimization models for stochastically deteriorating system. *Reliability Engineering & System Safety*, 157, 54–63.
- Andersen, P. K., Borgan, O., Gill, R. D., & Keiding, N. (2012). *Statistical models based on counting processes*. Springer Science & Business Media.
- Andersen, P. K., & Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics*, 10(4), 1100–1120.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society Series B (Methodological)*, 34(2), 187–220.
- Dauxois, J.-Y., Gasmi, S., & Gaudoin, O. (2019). Semiparametric inference for an extended geometric failure rate reduction model. *Journal of Statistical Planning and Inference*, 199, 14–28.
- Doyen, L., Gaudoin, O., & Syamsundar, A. (2017). On geometric reduction of age or intensity models for imperfect maintenance. *Reliability Engineering & System Safety*, 168, 40–52.
- Finkelstein, M. (2008). *Failure rate modelling for reliability and risk*. Springer Science & Business Media.
- Fleming, T. R., & Harrington, D. P. (2005). *Counting processes and survival analysis*. Wiley.
- Gao, F., Zeng, D., Couper, D., & Lin, D. Y. (2019). Semiparametric regression analysis of multiple right and interval censored events. *Journal of the American Statistical Association*, 114(527), 1232–1240.
- Huang, C. Y., Qin, J., & Wang, M. C. (2010). Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. *Biometrics*, 66(1), 39–49.
- Kijima, M. (1989). Some results for repairable systems with general repair. *Journal of Applied Probability*, 26(1), 89–102.
- Lin, D. Y. (1991). Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators. *Journal of the American Statistical Association*, 86(415), 725–728.
- Lin, D. Y. (2007). On the Breslow Estimator. *Lifetime Data Analysis*, 13(4), 471–480.
- Lin, D. Y., Wei, L. J., Yang, I., & Ying, Z. (2000). 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.

- Pena, E. A., Slate, E. H., & Gonzalez, J. R. (2007). Semiparametric inference for a general class of models for recurrent events. *Journal of Statistical Planning and Inference*, 137(6), 1727–1747.
- Pena, E. A., Strawderman, R. L., & Hollander, M. (2000). A weak convergence result relevant in recurrent and renewal models. In N. Limnios, & M. Nikulin (Eds.), *Recent advances in reliability theory* (pp. 493–514). Springer.
- Pena, E. A., Strawderman, R. L., & Hollander, M. (2001). Nonparametric estimation with recurrent event data. *Journal of the American Statistical Association*, 96(456), 1299–1315.
- Qu, S., Wang, J.-L., & Wang, X. (2016). Optimal estimation for the functional Cox model. *The Annals of Statistics*, 44(4), 1708–1738.
- Rahman, A. F., Lynch, J. D., & Peña, E. A. (2014). Nonparametric bayes estimation of gap-time distribution with recurrent event data. *Journal of Nonparametric Statistics*, 26(3), 575–598.
- Rockafellar, R. T. (1970). *Convex analysis*. Princeton University Press.
- Rudin, W. (1991). *Functional analysis*. McGraw-Hill.
- Spiekerman, C. F., & Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association*, 93(443), 1164–1175.
- Syamsundar, A., & Kumar, D. E. V. (2017). *Geometric failure rate reduction model for the analysis of repairable systems*. Proceedings of the 2017 Annual Reliability and Maintainability Symposium (RAMS), Orlando, Florida, USA: IEEE.
- Van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge University Press.
- Wang, X., & Xu, D. (2010). An inverse Gaussian process model for degradation data. *Technometrics*, 52(2), 188–197.
- Wu, S., & Zuo, M. J. (2010). Linear and nonlinear preventive maintenance models. *IEEE Transactions on Reliability*, 59(1), 242–249.
- Ye, Z., & Ng, H. K. T. (2014). On analysis of incomplete field failure data. *The Annals of Applied Statistics*, 8(3), 1713–1727.
- Zhang, M., & Xie, M. (2017). An ameliorated improvement factor model for imperfect maintenance and its goodness of fit. *Technometrics*, 59(2), 237–246.

How to cite this article: Wang, J., Chen, P., & Ye, Z. (2022). Efficient semiparametric estimation of time-censored intensity-reduction models for repairable systems. *Scandinavian Journal of Statistics*, 1–29. <https://doi.org/10.1111/sjost.12564>

APPENDIX

Proof of Lemma 1

We use a similar proof as that in Pena et al. (2001). According to the definition of $N(\cdot, t)$, for each fixed t the compensator is given by $A(s, t; \gamma_0, \beta_0, \Lambda_0) = \int_0^s Z(u, t) \bar{A}(du; \gamma_0, \beta_0, \Lambda_0)$. Therefore,

$$\begin{aligned}
 A(s, t; \gamma, \beta, \Lambda) &= \int_0^s I\{R(u) \leq t\} I\{C \geq t\} e^{\gamma \cdot \alpha(N^*(u-) + 1) + \beta^T \mathbf{X}} \lambda(R(u)) du \\
 &= \int_0^{s \wedge C} \left[\sum_{j=1}^{N^*((s \wedge C)-)} I\{S_{j-1} < u \leq S_j\} + I\{u > S_{N^*((s \wedge C)-)}\} \right] I\{R(u) \leq t\} e^{\gamma \cdot \alpha(N^*(u-) + 1) + \beta^T \mathbf{X}} \lambda(R(u)) du \\
 &= \sum_{j=1}^{N^*((s \wedge C)-)} \int_0^{T_j} I\{x \leq t\} e^{\gamma \cdot \alpha(j) + \beta^T \mathbf{X}} \lambda(x) dx + \int_0^{R(s \wedge C)} I\{x \leq t\} e^{\gamma \cdot \alpha(j) + \beta^T \mathbf{X}} \Lambda(dx) \\
 &= \int_0^t \left[\sum_{j=1}^{N^*((s \wedge C)-)} I\{T_j \geq x\} e^{\gamma \cdot \alpha(j)} + I\{R(s \wedge C) \geq x\} e^{\gamma \cdot \alpha(N^*((s \wedge C)-) + 1)} \right] e^{\beta^T \mathbf{X}} \Lambda(dx).
 \end{aligned}$$

That is, $A(s, t; \gamma, \beta, \Lambda) = \int_0^t e^{\beta^T \mathbf{X}} Y(s, x; \gamma) \Lambda(dx)$.

Proof of Proposition 1

It is obvious that the profile likelihood function ℓ_n is continuously differentiable since Y is a continuous differentiable function with respect to $\theta = (\gamma, \beta^\top)^\top$. According to theorem 4.5 in Rockafellar (1970), verification of the concavity of ℓ_n is equivalent to showing that the Hessian matrix of ℓ_n is negative semidefinite. With the profile log-likelihood function defined in (7), the profile likelihood score can be written as

$$U_n(\gamma, \beta) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \begin{pmatrix} \alpha(\tilde{N}_i(u^-) + 1) \\ \mathbf{X}_i \end{pmatrix} Z_i(u, \tau) \tilde{N}_i(du) - \frac{1}{n} \int_0^\tau \begin{pmatrix} \sum_{i=1}^n e^{\beta^\top \mathbf{X}_i} Y_i^{(1)}(\tau, \mathbf{x}; \gamma) \\ \sum_{i=1}^n e^{\beta^\top \mathbf{X}_i} Y_i(\tau, \mathbf{x}; \gamma) \mathbf{X}_i \end{pmatrix} \frac{1}{\sum_{i=1}^n e^{\beta^\top \mathbf{X}_i} Y_i(\tau, \mathbf{x}; \gamma)} \bar{N}(\tau, dx). \tag{A1}$$

Notice that the Hessian matrix of ℓ_n is given by

$$U_n^{(1)}(\gamma, \beta) = \frac{1}{n} \int_0^\tau \frac{S_n^{(1)}(\tau, \mathbf{x}; \gamma, \beta)^{\otimes 2} - S_n^{(0)}(\tau, \mathbf{x}; \gamma, \beta) S_n^{(2)}(\tau, \mathbf{x}; \gamma, \beta)}{S_n^{(0)}(\tau, \mathbf{x}; \gamma, \beta)^2} \bar{N}(\tau, dx).$$

In order to show that $U_n^{(1)}(\gamma, \beta)$ is negative semidefinite, we verify that the integrand

$$S_n^{(1)}(\tau, \mathbf{x}; \gamma, \beta)^{\otimes 2} - S_n^{(0)}(\tau, \mathbf{x}; \gamma, \beta) S_n^{(2)}(\tau, \mathbf{x}; \gamma, \beta),$$

is negative semidefinite for any γ, β , and x . For every $y \in \mathbb{R}^{p+1}$, we have

$$\mathbf{y}^\top S_n^{(1)\otimes 2} \mathbf{y} = \left[\frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^{\tilde{N}_i(\tau^-)} \mathbf{y}^\top \begin{pmatrix} \alpha(j) \\ \mathbf{X}_i \end{pmatrix} e^{\gamma \cdot \alpha(j) + \beta^\top \mathbf{X}_i} I\{T_{ij} \geq t\} + \mathbf{y}^\top \begin{pmatrix} \alpha(\tilde{N}_i(\tau^-)) \\ \mathbf{X}_i \end{pmatrix} e^{\gamma \cdot \alpha(\tilde{N}_i(\tau^-)) + \beta^\top \mathbf{X}_i} I\{R_i((\tau \wedge C_i) -) \geq t\} \right) \right]^2,$$

and

$$\mathbf{y}^\top S_n^{(2)} \mathbf{y} = \frac{1}{n} \sum_{i=1}^n \left[\sum_{j=1}^{\tilde{N}_i(\tau^-)} \left(\mathbf{y}^\top \begin{pmatrix} \alpha(j) \\ \mathbf{X}_i \end{pmatrix} \right)^2 e^{\gamma \cdot \alpha(j) + \beta^\top \mathbf{X}_i} I\{T_{ij} \geq t\} + \left(\mathbf{y}^\top \begin{pmatrix} \alpha(\tilde{N}_i(\tau^-)) \\ \mathbf{X}_i \end{pmatrix} \right)^2 e^{\gamma \cdot \alpha(\tilde{N}_i(\tau^-)) + \beta^\top \mathbf{X}_i} I\{R_i((\tau \wedge C_i) -) \geq t\} \right],$$

since $\mathbf{y}^\top \mathbf{z}^{\otimes 2} \mathbf{y} = (\mathbf{y}^\top \mathbf{z})^2$. Combine the two equations to see that

$$\mathbf{y}^\top \left[S_n^{(1)}(\tau, \mathbf{x}; \gamma, \beta)^{\otimes 2} - S_n^{(0)}(\tau, \mathbf{x}; \gamma, \beta) S_n^{(2)}(\tau, \mathbf{x}; \gamma, \beta) \right] \mathbf{y} \leq 0,$$

for all γ, β and x according to Cauchy–Schwarz inequality. That is, the Hessian matrix $U_n^{(1)}$ is negative semidefinite and the profile likelihood function ℓ_n is thus concave.

Proof of Lemma 2

We only show the proof for $k = 0$ here, since the proof for $k = 1$ and 2 is almost the same as that for $k = 0$. Consider the class of functions $\mathcal{G} = \{g_{t,\theta}(X, \tilde{N}, \tilde{Y}) = e^{\beta^\top \mathbf{X}} Y(\tau, t; \gamma) : t \in [0, \tau], \theta \in \Theta\}$ indexed by t and θ . It is easy to see that

$$E g_{0,\theta}^2 = E \left[\left(\sum_{j=1}^{\tilde{N}(\tau-)} e^{M_\gamma \alpha(j)} + e^{M_\gamma \alpha(\tilde{N}(\tau-)+1)} \right)^2 e^{2\beta^\top \mathbf{X}} \right],$$

which is finite because X has bounded support, and $E [Y^{(0)}(\tau, 0; M_\gamma)^2] < \infty$.

Further, the map $t \mapsto g_{t,\theta}(X, \tilde{N}, \tilde{Y})$, when treated as a function of t , is right-continuous and monotone decreasing. Thus, we can use dominated convergence theorem to see that $E g_{t,\theta}^2$ is right continuous and monotone decreasing. Therefore, for a given $\epsilon > 0$, we first fix $\theta \in \Theta$ and consider the partition $0 = t_1 < \dots < t_m = \tau$ such that $E (g_{(t_k-),\theta} - g_{t_{k-1},\theta})^2 < \epsilon^2$. Then, $[g_{(t_k-),\theta}, g_{t_{k-1},\theta}]$ for $k = 2, \dots, m$ are brackets of $L_2(P)$ -size ϵ , and the total number of brackets can be chosen of order $(1/\epsilon)^2$.

Next, we allow θ to change. Notice that

$$\begin{aligned} \left| \frac{\partial}{\partial \theta} g_{t,\theta} \right| &\leq |M_{\mathbf{X}}| e^{M_{\beta} M_{\mathbf{X}}} Y(\tau, 0, M_\gamma) + e^{M_{\beta} M_{\mathbf{X}}} \sum_{j=1}^{\tilde{N}(\tau-)} \alpha(j) e^{M_\gamma \alpha(j)} \\ &+ e^{M_{\beta} M_{\mathbf{X}}} \alpha(\tilde{N}(\tau-) + 1) e^{M_\gamma \alpha(\tilde{N}(\tau-)+1)} \equiv \dot{g}(\mathbf{X}, \tilde{N}, \tilde{Y}). \end{aligned}$$

Let $\tilde{\epsilon} = \epsilon / \sqrt{E \dot{g}^2}$. For a fixed ϵ , we first find an $\tilde{\epsilon}$ -net $\theta_1, \dots, \theta_q$ of Θ with respect to the Euclidean metric in \mathbb{R}^{p+1} . For each $j \leq q$, we follow the above procedure at the beginning of the proof to find brackets of $L_2(P)$ -size ϵ for the functions $\{g_{t,\theta_j} : t \in [0, \tau]\}$. Enlarge each of the above brackets to $[g_{(t_k-),\theta_j} - \tilde{\epsilon} \dot{g}, g_{t_{k-1},\theta_j} + \tilde{\epsilon} \dot{g}]$. All of these brackets have $L_2(P)$ -size 3ϵ . Repeat the above procedure for all $j \leq q$. Here, q is not larger than the ϵ -covering number of Θ , and is of order $(1/\epsilon)^{p+1}$. For each fixed $j \leq q$, the ϵ -bracketing number for $\{g_{t,\theta_j} : t \in [0, \tau]\}$ is of order $(1/\epsilon)^2$. Therefore, the total number of ϵ -brackets under the $L_2(P)$ norm is of order $(1/\epsilon)^{p+3}$. This shows that the class of functions \mathcal{G} is P -Donsker.

Proof of Theorem 1

To establish the consistency of parametric estimator $\hat{\theta}_n$, we let $\Theta(\epsilon) = \{\theta : \|\theta - \theta_0\| \leq \epsilon, \theta \in \mathbb{R}^{p+1}\}$, which is a compact set in \mathbb{R}^{p+1} . Notice that to prove the weak consistency is equivalent to verifying that for each $\epsilon > 0$, $P(\hat{\theta}_n \in \Theta(\epsilon))$ converges to 1. Without loss of generality, we consider an arbitrary ϵ such that $\Theta(\epsilon) \subset \Theta$.

First we are going to find a continuous function $\ell(\theta)$ such that

$$\sup_{\theta \in \Theta(\epsilon)} |\ell_n(\theta) - \ell(\theta)| \rightarrow_P 0, \quad (\text{A3})$$

where we denote $\ell_n(\gamma, \beta)$ by $\ell_n(\theta)$ for notational convenience. Define

$$\tilde{\ell}(\theta) = \int_0^\tau (\gamma \cdot \alpha(\tilde{N}(u-) + 1) + \beta^\top \mathbf{X}) Z(u, \tau) \tilde{N}(du) - \int_0^\tau \log(s^{(0)}(\tau, x; \gamma, \beta)) N(\tau, dx),$$

and let $\ell(\theta)$ be the expectation of $\tilde{\ell}(\theta)$ under γ_0, β_0 and Λ_0 . Then $\frac{1}{n} \sum_{i=1}^n \tilde{\ell}_i(\theta) \rightarrow_P \ell(\theta)$ by the law of large numbers for every fixed θ . In addition, the difference between $\ell_n(\theta)$ and $\frac{1}{n} \sum_{i=1}^n \tilde{\ell}_i(\theta)$ equals

$$\frac{1}{n} \int_0^\tau \log \left(\frac{s^{(0)}(\tau, \mathbf{x}; \gamma, \beta)}{S_n^{(0)}(\tau, \mathbf{x}; \gamma, \beta)} \right) \bar{N}(\tau, d\mathbf{x}).$$

According to condition (ii) and Lemma 2, it converges to 0 in probability since

$$\sup_{x \in [0, \tau]} \left| \log \left(\frac{s^{(0)}(\tau, \mathbf{x}; \gamma, \beta)}{S_n^{(0)}(\tau, \mathbf{x}; \gamma, \beta)} \right) \right| \rightarrow_P 0,$$

and $\frac{1}{n} \bar{N}(\tau, \tau)$ is uniformly tight. Thus, the difference $\ell_n(\theta) - \frac{1}{n} \sum_{i=1}^n \tilde{\ell}_i(\theta) \rightarrow_P 0$, which yields $\ell_n(\theta) \rightarrow_P \ell(\theta)$ for every θ . Then following theorem II.1 in Andersen and Gill (1982), we use concavity and pointwise convergence of ℓ_n to see that $\ell(\theta)$ is a concave function, and ℓ_n converges to ℓ uniformly on every compact set $K \subset \Theta$. That is,

$$\sup_{\theta \in K} |\ell_n(\theta) - \ell(\theta)| \rightarrow_P 0.$$

Since $\Theta(\epsilon)$ is compact in \mathbb{R}^{p+1} , we let $K = \Theta(\epsilon)$ to derive (A3). Furthermore, the continuity of ℓ can be easily derived from the continuity of ℓ_n and the uniform convergence on every compact set in Θ .

Next, we demonstrate that $\ell(\theta)$ is maximized uniquely at θ_0 . Because ℓ is continuous and concave on Θ , we consider the maximum set $\Theta_0 = \{\theta \in \Theta : \ell(\theta) = \sup_{\theta \in \Theta} \ell(\theta)\}$. In order to show that $\Theta_0 = \{\theta_0\}$, we are going to verify that the derivative vector $\ell^{(1)}(\theta_0) = \mathbf{0}_{(p+1) \times 1}$ and the Hessian matrix $\ell^{(2)}(\theta_0)$ is negative definite. Because $\tilde{\ell}^{(1)}(\theta)$ is upper bounded by

$$\int_0^\tau \begin{pmatrix} \alpha(\tilde{N}(u-) + 1) \\ \mathbf{X} \end{pmatrix} Z(u, \tau) \tilde{N}(du) + \sup_{x \in [0, \tau], \theta \in \Theta} \left\| \frac{s^{(1)}(\tau, \mathbf{x}; \gamma, \beta)}{s^{(0)}(\tau, \mathbf{x}; \gamma, \beta)} \right\| N(\tau, \tau),$$

we use dominated convergence theorem to see that $\ell^{(1)}(\theta) = E[\tilde{\ell}^{(1)}(\theta)]$. Similarly, we can also present that $\ell^{(2)}(\theta) = E[\tilde{\ell}^{(2)}(\theta)]$. Then taking derivative on both side of Equation 3 with respect to θ , we have

$$\int_0^\tau \begin{pmatrix} \alpha(\tilde{N}(u-) + 1) \\ \mathbf{X} \end{pmatrix} Z(u, \tau) \tilde{A}(du; \gamma, \beta, \Lambda) = \int_0^\tau \begin{pmatrix} e^{\beta^\top \mathbf{X}} Y^{(1)}(\tau, \mathbf{x}; \gamma) \\ e^{\beta^\top \mathbf{X}} Y(\tau, \mathbf{x}; \gamma) \mathbf{X} \end{pmatrix} \Lambda(d\mathbf{x}). \tag{A4}$$

Thus, the derivative of $\tilde{\ell}$ evaluated at θ_0 is given by

$$\begin{aligned} \tilde{\ell}^{(1)}(\theta_0) &= \int_0^\tau \begin{pmatrix} \alpha(\tilde{N}(u-) + 1) \\ \mathbf{X} \end{pmatrix} Z(u, \tau) \tilde{M}(du; \gamma_0, \beta_0, \Lambda_0) \\ &\quad - \int_0^\tau \frac{s^{(1)}(\tau, \mathbf{x}; \gamma_0, \beta_0)}{s^{(0)}(\tau, \mathbf{x}; \gamma_0, \beta_0)} M(\tau, d\mathbf{x}; \gamma_0, \beta_0, \Lambda_0). \\ &\quad + \int_0^\tau \left[\begin{pmatrix} e^{\beta^\top \mathbf{X}} Y^{(1)}(\tau, \mathbf{x}; \gamma) \\ e^{\beta^\top \mathbf{X}} Y(\tau, \mathbf{x}; \gamma) \mathbf{X} \end{pmatrix} - \frac{s^{(1)}(\tau, \mathbf{x}; \gamma_0, \beta_0)}{s^{(0)}(\tau, \mathbf{x}; \gamma_0, \beta_0)} e^{\beta^\top \mathbf{X}} Y(\tau, \mathbf{x}; \gamma) \right] \Lambda_0(d\mathbf{x}), \end{aligned} \tag{A5}$$

Notice that according to the martingale structure in Lemma 1, the expectation of $\tilde{\ell}^{(1)}(\theta_0)$ equals $\mathbf{0}_{(p+1) \times 1}$. That is, $\ell^{(1)}(\theta_0) = E[\tilde{\ell}^{(1)}(\theta_0)] = \mathbf{0}_{(p+1) \times 1}$. Furthermore, since

$$\tilde{\ell}^{(2)}(\theta_0) = \int_0^\tau \frac{s^{(2)}(\tau, x; \gamma_0, \beta_0) s^{(0)}(\tau, x; \gamma_0, \beta_0) - s^{(1)}(\tau, x; \gamma_0, \beta_0)^{\otimes 2}}{s^{(0)}(\tau, x; \gamma_0, \beta_0)^2} N(\tau, dx),$$

we have $\ell^{(2)}(\theta_0) = -\Sigma(\gamma_0, \beta_0)$, which is negative definite by assumption (iii). Therefore, we have proved that $\Theta_0 = \{\theta_0\}$, which implies that ℓ is maximized uniquely at θ_0 .

After that, we consider the value of the parametric estimator $\hat{\theta}_n$. Notice that the continuous function ℓ has a maximizer θ_ϵ on $\partial\Theta(\epsilon) = \{\theta : \|\theta - \theta_0\| = \epsilon, \theta \in \mathbb{R}^{p+1}\}$, the boundary of $\Theta(\epsilon)$, which is a compact set in \mathbb{R}^{p+1} . Then we let $\delta = \ell(\theta_0) - \ell(\theta_\epsilon)$ and $A_n = \{\omega : \sup_{\theta \in \Theta(\epsilon)} |\ell_n(\theta; \omega) - \ell(\theta)| < \delta/2\}$. Obviously, $\delta > 0$ since ℓ has a uniquely maximizer θ_0 . Thus, for every $\omega \in A_n$ and every $\theta \in \Theta(\epsilon)^C$, we have the fact that

$$\ell_n(\tilde{\theta}, \omega) < \ell(\tilde{\theta}) + \delta/2 \leq \ell(\theta_\epsilon) + \delta/2 = \ell(\theta_0) - \delta/2 < \ell_n(\theta_0, \omega) \leq \ell_n(\hat{\theta}_n, \omega),$$

where $\tilde{\theta} = \theta_0 + \epsilon(\theta - \theta_0)/\|\theta - \theta_0\|$ is the intersection point between $\partial\Theta(\epsilon)$ and the line segment between θ and θ_0 . Notice that $\ell_n(\theta, \omega) < \ell_n(\tilde{\theta}, \omega) < \ell_n(\hat{\theta}_n, \omega)$ since ℓ_n is concave. Therefore, $\hat{\theta}_n \notin \Theta(\epsilon)^C$ for every $\omega \in A_n$, which implies that $\hat{\theta}_n \in \Theta(\epsilon)$. According to the uniform convergence of ℓ_n on $\Theta(\epsilon)$, we find that $P(A_n) \rightarrow 1$. As a result, $P(\hat{\theta}_n \in \Theta(\epsilon)) \rightarrow 1$ for arbitrary ϵ , which implies weak consistency for parametric estimator $\hat{\theta}_n$.

On the other hand, to show the consistency of the nonparametric part, we telescope using $\hat{\Lambda}_n(\cdot; \gamma_0, \beta_0)$ to get

$$|\hat{\Lambda}_n(t; \hat{\gamma}_n, \hat{\beta}_n) - \Lambda_0(t)| \leq |\hat{\Lambda}_n(t; \hat{\gamma}_n, \hat{\beta}_n) - \hat{\Lambda}_n(t; \gamma_0, \beta_0)| + |\hat{\Lambda}_n(t; \gamma_0, \beta_0) - \Lambda_0(t)|.$$

According to theorem 1 in Pena et al. (2001),

$$\sup_{1 \leq t \leq \tau} |\hat{\Lambda}_n(t; \gamma_0, \beta_0) - \Lambda_0(t)| \rightarrow_P 0.$$

Besides, use the definition of $\hat{\Lambda}_n(\cdot; \gamma, \beta)$ to see that

$$\begin{aligned} & \sup_{0 \leq t \leq \tau} \left| \hat{\Lambda}_n(t; \hat{\gamma}_n, \hat{\beta}_n) - \hat{\Lambda}_n(t; \gamma_0, \beta_0) \right| \\ & \leq \frac{1}{n} \int_0^\tau \left| \frac{I\{S_n^{(0)}(\tau, x; \hat{\gamma}_n, \hat{\beta}_n) > 0\}}{S_n^{(0)}(\tau, x; \hat{\gamma}_n, \hat{\beta}_n)} - \frac{I\{S_n^{(0)}(\tau, x; \gamma_0, \beta_0) > 0\}}{S_n^{(0)}(\tau, x; \gamma_0, \beta_0)} \right| \overline{N}(\tau, dx). \end{aligned}$$

Because $(\hat{\gamma}_n, \hat{\beta}_n)^\top \rightarrow_P (\gamma_0, \beta_0)^\top$ and $S_n^{(0)}(\tau, x; \gamma, \beta)$ is a uniformly bounded continuous function with respect to θ , the second line above converges to 0 in probability. Combine the results to see that the nonparametric estimator $\hat{\Lambda}_n(\cdot)$ is weakly consistent.

Proof of theorem 2

Define

$$\begin{aligned} U_n(\gamma_0, \beta_0)(s) &= \frac{1}{n} \sum_{i=1}^n \int_0^s \left[\begin{pmatrix} \alpha(\tilde{N}_i(u-) + 1) \\ \mathbf{X}_i \end{pmatrix} - \frac{S_n^{(1)}(\tau, R_i(u); \gamma_0, \beta_0)}{S_n^{(0)}(\tau, R_i(u); \gamma_0, \beta_0)} \right] \\ & \quad \times Z_i(u, \tau) \tilde{M}_i(du; \gamma_0, \beta_0, \Lambda_0), \end{aligned} \tag{A6}$$

which is a local martingale with respect to the filtration \mathcal{F}_s . Notice that when $s = \tau$, $U_n(\gamma_0, \beta_0)(\tau)$ equals the profile likelihood score. To verify the weak convergence of $\sqrt{n}(\hat{\theta}_n - \theta_0)$, we use Taylor expansion to see that

$$0 - \sqrt{n}U_n(\gamma_0, \beta_0)(\tau) = \sqrt{n}U_n^{(1)}(\gamma_0, \beta_0)(\tau) (\hat{\theta}_n - \theta_0) + o_p(\sqrt{n} \|\hat{\theta}_n - \theta_0\|).$$

Then we can derive the weak convergence of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ based on the weak convergence of $\sqrt{n}U_n(\gamma_0, \beta_0)$ using the delta method.

For the nonparametric estimator, notice that

$$\begin{aligned} \sqrt{n}(\hat{\Lambda}_n(t) - \Lambda(t)) &= \sqrt{n}(\hat{\Lambda}_n(\tau, t; \hat{\gamma}_n, \hat{\beta}_n) - \hat{\Lambda}_n(\tau, t; \gamma_0, \beta_0)) \\ &\quad + \sqrt{n}(\hat{\Lambda}_n(\tau, t; \gamma_0, \beta_0) - \tilde{\Lambda}_n(\tau, t; \gamma_0, \beta_0)) + \sqrt{n}(\tilde{\Lambda}_n(\tau, t; \gamma_0, \beta_0) - \Lambda(t)), \end{aligned} \tag{A7}$$

where $\tilde{\Lambda}_n(s, t; \gamma, \beta) = \int_0^t I\{S_n^{(0)}(s, x; \gamma, \beta) > 0\} \Lambda(dx)$. The first term converges to a tight limit in distribution according to the delta method and the weak convergence of $\sqrt{n}(\hat{\theta}_n - \theta_0)$, while the third term converges to 0 in probability by the proof of theorem 1 in Pena et al. (2001).

Therefore, we construct the joint random variable by

$$Z_{n;t_1, t_2, \dots, t_m}(s) = \left(\sqrt{n}U_n(\gamma_0, \beta_0)(s)^\top, Z_{n,(1+p+1)}(s), Z_{n,(1+p+2)}(s), \dots, Z_{n,(1+p+m)}(s) \right)^\top,$$

where the first term is a $(1 + p)$ -dimensional vector incorporating $\sqrt{n}(\hat{\theta}_n - \theta)$, and

$$Z_{n,(1+p+i)}(s) = \int_0^{t_i} \frac{\bar{M}(s, dx; \gamma_0, \beta_0, \Lambda_0)}{\sqrt{n}S_n^{(0)}(\tau, x; \gamma_0, \beta_0)} = \sum_{i=1}^n \int_0^s \frac{M_i(du, t_i; \gamma_0, \beta_0, \Lambda_0)}{\sqrt{n}S_n^{(0)}(\tau, R_i(u); \gamma_0, \beta_0)}. \tag{A8}$$

Furthermore, according to the definition in (A6) and (A8), the random variable $Z_{n;t_1, t_2, \dots, t_m}(s)$ can be written as

$$Z_{n;t_1, t_2, \dots, t_m}(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^s H_i^{(n)}(u) \tilde{M}_i(du; \gamma_0, \beta_0, \Lambda_0),$$

where

$$H_i^{(n)}(u) = \begin{pmatrix} \left[\begin{pmatrix} \alpha(\tilde{N}_i(u-) + 1) \\ \mathbf{X}_i \end{pmatrix} - \frac{S_n^{(1)}(\tau, R_i(u); \gamma_0, \beta_0)}{S_n^{(0)}(\tau, R_i(u); \gamma_0, \beta_0)} \right] \\ \frac{I\{R_i(u) \leq t_1\}}{S_n^{(0)}(\tau, R_i(u); \gamma_0, \beta_0)} \\ \vdots \\ \frac{I\{R_i(u) \leq t_m\}}{S_n^{(0)}(\tau, R_i(u); \gamma_0, \beta_0)} \end{pmatrix}.$$

For notational convenience, we denote this $(1 + p + m)$ -dimensional vector by

$$H_i^{(n)}(u) = \left(H_{i,1}^{(n)}(u), H_{i,2}^{(n)}(u), \dots, H_{i,1+p+m}^{(n)}(u) \right).$$

Next, we use the Martingale CLT given in theorem 5.3.5, Fleming and Harrington (2005) to see that the random process $Z_{n;t_1, t_2, \dots, t_m}(s), s \in [0, \tau]$ converges weakly to a Gaussian process. Actually, using a similar proof as we have done in the proof of Theorem 1, it is easy to see that for all $s \in [0, \tau]$, there exists a matrix $(\Gamma_{k,l}(s))_{(1+p+m) \times (1+p+m)}$ such that

$$\langle Z_{n;t_1, \dots, t_m; k}, Z_{n;t_1, \dots, t_m; l} \rangle (s) \rightarrow_P \Gamma_{k,l}(s), \forall k, l \in [1, 1 + p + m], \tag{A9}$$

where the covariance matrix evaluated at τ satisfies $(\Gamma_{k,l}(\tau))_{(1+p+m) \times (1+p+m)} = \Gamma$, where

$$\Gamma = \begin{pmatrix} \int_0^\tau \frac{s^{(2)}(\tau, x) s^{(0)}(\tau, x) - s^{(1)}(\tau, x)^2}{s^{(0)}(\tau, x)^2} \Lambda_0(dx) & 0 & 0 & \dots & 0 \\ 0 & \int_0^{t_1} \frac{\Lambda_0(dx)}{s^{(0)}(\tau, x)} & \int_0^{t_1 \wedge t_2} \frac{\Lambda_0(dx)}{s^{(0)}(\tau, x)} & \dots & \int_0^{t_1 \wedge t_m} \frac{\Lambda_0(dx)}{s^{(0)}(\tau, x)} \\ 0 & \int_0^{t_2 \wedge t_1} \frac{\Lambda_0(dx)}{s^{(0)}(\tau, x)} & \int_0^{t_2} \frac{\Lambda_0(dx)}{s^{(0)}(\tau, x)} & \dots & \int_0^{t_2 \wedge t_m} \frac{\Lambda_0(dx)}{s^{(0)}(\tau, x)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \int_0^{t_m \wedge t_1} \frac{\Lambda_0(dx)}{s^{(0)}(\tau, x)} & \int_0^{t_m \wedge t_2} \frac{\Lambda_0(dx)}{s^{(0)}(\tau, x)} & \dots & \int_0^{t_m} \frac{\Lambda_0(dx)}{s^{(0)}(\tau, x)} \end{pmatrix},$$

and $s^{(k)}(\tau, x)$ is the abbreviation for $s^{(k)}(\tau, x; \gamma_0, \beta_0)$. Besides, the Linderberg condition

$$\langle Z_{n;t_1, \dots, t_m; k, \epsilon}, Z_{n;t_1, \dots, t_m; l, \epsilon} \rangle (s) = \frac{1}{n} \sum_{i=1}^n \int_0^s \left(H_{i,k}^{(n)}(u) \right)^2 I\{H_{i,k}^{(n)}(u) \geq \sqrt{n}\epsilon\} \tilde{A}_i(du; \gamma_0, \beta_0, \Lambda_0), \tag{A10}$$

can be shown by a similar proof as in theorem 31.1, Pena et al. (2000). Therefore, we use theorem 5.3.5 in Fleming and Harrington (2005) to see that the random process $Z_{n;t_1, t_2, \dots, t_m}(\cdot)$ converges to a multivariate Gaussian process with mean $\mathbf{0}_{1+p+m}$ and independent increment, and the covariance matrix between time s_1 and s_2 is $\Gamma(s_1 \wedge s_2)$.

As a result, the weak convergency of the parametric estimators can be derived from the convergence of $Z_{n;t_1, t_2, \dots, t_m}(\tau)$. For the parametric part, notice that the first $(1 + p)$ elements of Z_n is $\sqrt{n}U_n(\gamma_0, \beta_0)(\tau)^\top$. Then we use the delta method to see that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotic normal with mean zero and covariance matrix

$$\Sigma_\theta = \left(\frac{1}{n} \int_0^\tau \frac{s^{(0)}(\tau, x; \gamma_0, \beta_0) s^{(2)}(\tau, x; \gamma_0, \beta_0) - s^{(1)}(\tau, x; \gamma_0, \beta_0)^2}{s^{(0)}(\tau, x; \gamma_0, \beta_0)} \Lambda_0(dx) \right)^{-1}.$$

Actually, the consistency of the estimator denoted by $\hat{\Sigma}_\theta$ can be shown using the same method in the proof of Theorem 1. For the of nonparametric estimator $\hat{\Lambda}_n$, we use the decomposition in (A7) to see that

$$\begin{aligned} \sqrt{n} \left(\hat{\Lambda}_n(t_i) - \tilde{\Lambda}(t_i) \right) &= D_n(t_i, \gamma_0, \beta_0)^\top \left(U_n^{(1)}(\gamma_0, \beta_0)(\tau) \right)^{-1} U_n(\gamma_0, \Lambda_0)(\tau) \\ &\quad + \int_0^{t_i} \frac{n^{-1} \bar{M}(\tau, dx; \gamma_0, \beta_0, \Lambda_0)}{S_n^{(0)}(\tau, x; \gamma_0, \beta_0)} + o_p \left(\sqrt{n} \left\| \hat{\theta}_n - \theta_0 \right\| \right) + o_p(1), \end{aligned}$$

where

$$D_n(t_i, \gamma_0, \boldsymbol{\beta}_0) = \int_0^{t_i} \frac{S_n^{(1)}(\tau, \mathbf{x}; \gamma_0, \boldsymbol{\beta}_0)}{S_n^{(0)}(\tau, \mathbf{x}; \gamma_0, \boldsymbol{\beta}_0)^2} \bar{N}(\tau, d\mathbf{x}),$$

is the derivative of $\hat{\Lambda}_n$ with respect to $\boldsymbol{\theta}$. According to the delta method, the random vector

$$\sqrt{n} \left(\left(\hat{\Lambda}_n(t_1) - \tilde{\Lambda}(t_1) \right), \left(\hat{\Lambda}_n(t_2) - \tilde{\Lambda}(t_2) \right), \dots, \left(\hat{\Lambda}_n(t_m) - \tilde{\Lambda}(t_m) \right) \right)^\top,$$

converges weakly to an multivariate normal distribution with mean $\mathbf{0}_m$ and covariance matrix $\Sigma_\Lambda = (\Sigma_{ij})_{m \times m}$, where

$$\Sigma_{ij} = D_n(t_i, \gamma_0, \boldsymbol{\beta}_0)^\top \Sigma_\theta D_n(t_j, \gamma_0, \boldsymbol{\beta}_0) + \int_0^{t_i \wedge t_j} \frac{\Lambda(d\mathbf{x})}{s^{(0)}(\tau, \mathbf{x}; \gamma_0, \boldsymbol{\beta}_0)}.$$

In addition, according to the proof of theorem 3 in Spiekerman and Lin (1998) and theorem 31.1 in Pena et al. (2000), the random process $\sqrt{n} \left(\hat{\Lambda}_n(\cdot) - \Lambda(\cdot) \right)$ is tight. As a result, following theorem B.1.3 in Fleming and Harrington (2005), the random process $\sqrt{n} \left(\hat{\Lambda}_n(t) - \tilde{\Lambda}(t) \right)$ converges weakly to a mean-zero Gaussian process with covariance matrix

$$\Sigma_\Lambda(t_1, t_2) = \int_0^{t_1 \wedge t_2} \frac{\Lambda_0(d\mathbf{x})}{s^{(0)}(\tau, \mathbf{x}; \gamma_0, \boldsymbol{\beta}_0)} + D(t_1; \gamma_0, \boldsymbol{\beta}_0, \Lambda_0)^\top \Sigma_\theta D(t_2; \gamma_0, \boldsymbol{\beta}_0, \Lambda_0),$$

since $D_n(t, \gamma_0, \boldsymbol{\beta}_0)$ converges to $D(t_1; \gamma_0, \boldsymbol{\beta}_0, \Lambda_0)$ in probability for every $t \in [0, \tau]$. Similarly, we can establish the consistency of $\hat{\Sigma}_\Lambda(t_1, t_2)$ using the same way as we have done in Theorem 1.