Bernoulli **23**(4B), 2017, 3385–3411
DOI: 10.3150/16-BEJ850

Sieve maximum likelihood estimation for a general class of accelerated hazards models with bundled parameters

XINGQIU ZHAO[1], YUANSHAN WU[2] and GUOSHENG YIN[3]

[1]*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong.*
E-mail: xingqiu.zhao@polyu.edu.hk
[2]*School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei, 430072, China.*
E-mail: wuyuanshan@gmail.com
[3]*Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong.*
E-mail: gyin@hku.hk

In semiparametric hazard regression, nonparametric components may involve unknown regression parameters. Such intertwining effects make model estimation and inference much more difficult than the case in which the parametric and nonparametric components can be separated out. We study the sieve maximum likelihood estimation for a general class of hazard regression models, which include the proportional hazards model, the accelerated failure time model, and the accelerated hazards model. Coupled with the cubic B-spline, we propose semiparametric efficient estimators for the parameters that are bundled inside the nonparametric component. We overcome the challenges due to intertwining effects of the bundled parameters, and establish the consistency and asymptotic normality properties of the estimators. We carry out simulation studies to examine the finite-sample properties of the proposed method, and demonstrate its efficiency gain over the conventional estimating equation approach. For illustration, we apply our proposed method to a study of bone marrow transplantation for patients with acute leukemia.

Keywords: accelerated failure time model; B-spline; proportional hazards model; semiparametric efficiency bound; sieve maximum likelihood estimator; survival data

1. Introduction

The Cox [6] proportional hazards model has been routinely used in survival analysis. Under the proportional hazards assumption, the Cox model takes the form of

$$\lambda(t|\mathbf{Z}) = \lambda_0(t)e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}, \tag{1}$$

where $\lambda_0(\cdot)$ is the unknown baseline hazard function, $\mathbf{Z}$ is the covariate vector, and $\boldsymbol{\beta}_0$ is the regression parameter of interest. Nevertheless, such constant proportionality between hazard functions may not hold in practice. As a result, alternative modeling structures, such as the accelerated failure time (AFT) model, have been proposed, which directly model the logarithm of the failure time in a linear regression form,

$$\log(T) = -\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z} + \epsilon, \tag{2}$$

where T is the failure time, and the distribution of the error ϵ is unspecified. In fact, model (2) can be rewritten as

$$S(t|\mathbf{Z}) = S_0\big(te^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big), \tag{3}$$

where $S(t|\mathbf{Z})$ is the conditional survival function given covariate $\mathbf{Z}$, and $S_0(t)$ is the baseline survival function. The inference procedure for model (2) or (3) can typically be carried out using the least squares or rank methods (Prentice [14], Buckley and James [1], Ritov [15], Tsiatis [20], Wei, Ying and Lin [22], Lai and Ying [13], and Jin *et al.* [10]), and the corresponding variance are often estimated by resampling algorithm, such as bootstrap. Clearly, the nonparametric function $S_0(\cdot)$ involves the parametric component $e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}$, which makes it difficult to derive the nonparametric maximum likelihood estimator (NPMLE). Whereas, Zeng and Lin [24] developed an efficient estimator for the AFT model (3) by maximizing a kernel-smoothed profile likelihood function for the regression parameter. However, their approach is restricted to the log-transformed linear model (2). Recently, when the failure time T is subject to any completely known and strictly increasing transformation, Ding and Nan [7] proposed a sieve maximum likelihood estimator (MLE) for the censored linear regression model where the bundled parameter problem is involved. Owning to its flexibility, the sieve MLE method has been widely adopted in various semiparametric models, such as the partly linear Cox model (Huang [8]), transformed hazard models (Zeng, Yin and Ibrahim [25]), and the proportional odds model for survival data under various interval censoring mechanisms (Rossini and Tsiatis [16], Huang and Rossini [9], Shen [18]). Chen [2] provided a comprehensive review on the sieve method in the semiparametric models.

Despite the popularity of the Cox model, it assumes the treatment effect to take place immediately after patients are randomized to different treatment groups; that is, the hazards for different groups are different from time $t = 0$. However, in a randomized clinical trial, the treatment groups are essentially identical at $t = 0$ due to randomization. Randomization makes different groups alike except for treatments. Particularly in oncology, it often takes some time to observe efficacy effects of the treatment, for example, tumor shrinkage. In other words, it may take a certain period of lag time for the treatment to fully exert the therapeutic effect instead of being immediately effective. Along this direction, Chen and Wang [4] proposed the accelerated hazards model by replacing the survival functions in (3) with the corresponding hazard functions, and thus the conditional hazard function of failure time T given covariate $\mathbf{Z}$ takes the form of

$$\lambda(t|\mathbf{Z}) = \lambda_0\big(te^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big). \tag{4}$$

This model is intuitive in the sense that the hazard functions for different values of $\mathbf{Z}$ in (4) are the same at time $t = 0$. As time goes by, the hazards in different groups would gradually change due to different treatment effects. In a more general framework, Chen and Jewell [3] proposed a class of hazards regression model,

$$\lambda(t|\mathbf{Z}) = \lambda_0\big(te^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big)e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{Z}}, \tag{5}$$

where $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ are vectors of regression parameters. Based on different parametrization, model (5) includes the proportional hazards model ($\boldsymbol{\beta}_0 = 0$), the AFT model ($\boldsymbol{\beta}_0 = \boldsymbol{\gamma}_0$), and the acceler-

ated hazards model ($\boldsymbol{\gamma}_0 = 0$) as special cases. Chen and Jewell [3] developed martingale estimating equations for parameter estimation and inference, which may not be semiparametric efficient. Due to the discontinuity of estimating equations with respect to the regression parameters, the estimation procedure may suffer from potential multiple roots. Furthermore, the variance estimation depends on the derivation of the baseline function, which makes it difficult to calculate in practice.

To enhance the estimation efficiency and modeling flexibility, we study the sieve maximum likelihood estimation for a general class of accelerated hazards regression models in the form of

$$\Lambda(t|\mathbf{Z}, \mathbf{X}) = \Lambda_0\big(te^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big)e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}, \tag{6}$$

where $\Lambda_0(\cdot)$ is an unknown baseline cumulative hazards function, and $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ are unknown vectors of regression parameters. Covariates $\mathbf{Z}$ and $\mathbf{X}$ are allowed to share some common components. It is easy to see that model (6) reduces to the proportional hazards model when $\boldsymbol{\beta}_0 = 0$ and to the AFT model when $\boldsymbol{\gamma}_0 = 0$. In the case where $\mathbf{Z}$ is the same as $\mathbf{X}$, model (6) reduces to the accelerated hazards model when $\boldsymbol{\beta}_0 + \boldsymbol{\gamma}_0 = 0$ and to model (5) by reparameterizing $\boldsymbol{\beta}_0 + \boldsymbol{\gamma}_0$ as a new parameter. Hence, model (6) has great flexibility and, more importantly, is able to simultaneously investigate the time-accelerated effect of covariate $\mathbf{Z}$ and the proportional hazards effect of covaraite $\mathbf{X}$. Noting that the parametric and nonparametric components are bundled together in $\Lambda_0(te^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}})$, the theoretical development and numerical implementation of model (6) are very challenging. In contrast to the conventional martingale-based estimating equations proposed by Chen and Wang [4] and Chen and Jewell [3], we propose an intuitive spline-based sieve maximum likelihood estimation procedure for model (6) to improve the estimation efficiency. The numerical implementation of the proposed method can be achieved through the conventional gradient-based search algorithm, such as the Newton–Raphson algorithm. The variance estimates can be obtained from the inverse of Fisher's information matrix, and thus achieves semiparametric efficiency.

The main contributions of this paper are threefold. First, we proposed a sieve MLE method for the general accelerated hazards model in which the nonparametric function and the regression parameter are entangled with each other. The asymptotic properties of the resulting estimators are established and the estimator for the regression parameter achieves the semiparametric efficiency bound. Second, compared with the weighted estimating equation approach where the optimal weight depends on the form of the baseline function and thus it is challenging to find such an optimal weight and difficult to implement in practice, the proposed sieve MLE method is easier to be carried out. Third, the standard error estimates are obtained directly by either inverting the observed information matrix of all the parameters or inverting the efficient information matrix of the regression parameters, and both methods are more computationally tractable compared with the resampling techniques.

The rest of this article is organized as follows. We propose the sieve maximum likelihood estimating procedure in Section 2 and establish the asymptotic properties of the resultant estimators in Section 3, while proofs are presented in Section 7. We conduct simulation studies to assess the proposed method with finite samples in Section 4. As an illustration, a real data set is analyzed in Section 5. Some concluding remarks are provided in Section 6.

2. Sieve maximum likelihood estimation

Let T be the failure time, let C be the censoring time, denote $a \wedge b$ as the minimum of a and b, and let $I(\cdot)$ be the indicator function. We observe the data $\{Y_i \equiv T_i \wedge C_i, \Delta_i \equiv I(T_i \le C_i), \mathbf{Z}_i, \mathbf{X}_i\}$, $i = 1, \ldots, n$, which are independent and identically distributed (i.i.d.) copies of $\{Y \equiv T \wedge C, \Delta \equiv I(T \le C), \mathbf{Z}, \mathbf{X}\}$. Covariates $\mathbf{Z}$ and $\mathbf{X}$ may share the same components. Assume that T and C are conditionally independent given covariates $\mathbf{Z}$ and $\mathbf{X}$. Under model (6), the conditional survival and density functions of T given both $\mathbf{Z}$ and $\mathbf{X}$ are $S(t|\mathbf{Z}, \mathbf{X}) = \exp\{-\Lambda_0(te^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}})e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\}$ and $f(t|\mathbf{Z}, \mathbf{X}) = S(t|\mathbf{Z}, \mathbf{X})\lambda_0(te^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}})e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}+\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}$, respectively, where $\lambda_0(t) = \mathrm{d}\Lambda_0(t)/\mathrm{d}t$ is the baseline hazard function. The likelihood function of parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda)$ based on the observed data can be derived as

$$\prod_{i=1}^{n}\Big[\big\{\lambda\big(Y_i e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i}\big)e^{(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i+\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}_i)}\big\}^{\Delta_i}\exp\big\{-\Lambda\big(Y_i e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i}\big)e^{\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}_i}\big\}\Big],$$

where $\lambda(t) = \mathrm{d}\Lambda(t)/\mathrm{d}t$. The log-likelihood function is given by

$$\begin{aligned} l_n(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda) = n^{-1}\sum_{i=1}^{n}\Bigg[&\Delta_i\big\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}_i + \log\lambda\big(Y_i e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i}\big)\big\} \\ &- \int_0^{Y_i e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i}} \lambda(s)\,\mathrm{d}s e^{\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}_i}\Bigg]. \end{aligned} \tag{7}$$

To overcome the nonnegative constraint on $\lambda(\cdot)$, let $g(t) = \log\lambda(t)$ and then (7) is recast as

$$\begin{aligned} l_n(\boldsymbol{\beta}, \boldsymbol{\gamma}, g) = n^{-1}\sum_{i=1}^{n}\Bigg[&\Delta_i\big\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}_i + g\big(Y_i e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i}\big)\big\} \\ &- \int_0^{Y_i e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i}} \exp\big\{g(s)\big\}\,\mathrm{d}s e^{\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}_i}\Bigg]. \end{aligned} \tag{8}$$

In what follows, we propose a spline-based method to estimate the function g. Denote $b = \sup_{y,\mathbf{z},\boldsymbol{\beta}} y\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{z})$, then $0 < b < \infty$ under conditions C1 and C2 listed in Section 3. Let $0 \equiv t_0 < t_1 < \cdots < t_{K_n} < t_{K_n+1} \equiv b$ be a partition of $[0, b]$ with $K_n = O(n^v)$ and $\max_{0\le j\le K_n}|t_{j+1} - t_j| = O(n^{-v})$ for $v \in (0, 0.5)$. Denote the set of partition points by $T_{K_n} = \{t_1, \ldots, t_{K_n}\}$, and let $\mathcal{S}_n(T_{K_n}, K_n, p)$ be the space of polynomial splines of order p defined in Schumaker ([17], page 108, Definition 4.1). According to Schumaker ([17], page 117, Corollary 4.10), there exists a local basis $\{B_j : 1 \le j \le q_n\}$ with $q_n = K_n + p$ such that for any $s \in \mathcal{S}_n(T_{K_n}, K_n, p)$, we can write

$$s(t) = \mathbf{a}^{\mathrm{T}}\mathbf{B}(t) = \sum_{j=1}^{q_n} a_j B_j(t),$$

where $\mathbf{a} = (a_1, \ldots, a_{q_n})^{\mathrm{T}}$ and $\mathbf{B} = (B_1, \ldots, B_{q_n})^{\mathrm{T}}$. Under some suitable smoothness assumptions, g_0, the true function of g, can be well approximated by some function in $\mathcal{S}_n(T_{K_n}, K_n, p)$.

Let $\mathcal{B} \subseteq \mathbb{R}^{d_1}$ and $\mathcal{T} \subseteq \mathbb{R}^{d_2}$ denote the parameter spaces of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively, where d_1 and d_2 are their corresponding dimensions. As a result, we seek a member of $\mathcal{S}_n(T_{K_n}, K_n, p)$ along with values of $(\boldsymbol{\beta}, \boldsymbol{\gamma}) \in \mathcal{B} \times \mathcal{T}$ that maximizes the log-likelihood function. Specifically, we define $(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n, \widehat{\mathbf{a}}_n)$ to be the parameter values that maximize

$$l_n(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{a}) = n^{-1} \sum_{i=1}^{n} \Bigg[\Delta_i \big\{ \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_i + \boldsymbol{\gamma}^{\mathrm{T}} \mathbf{X}_i + \mathbf{a}^{\mathrm{T}} \mathbf{B}\big(Y_i e^{\boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_i}\big) \big\} - \int_0^{Y_i e^{\boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_i}} \exp\big\{ \mathbf{a}^{\mathrm{T}} \mathbf{B}(s) \big\} \, \mathrm{d}s e^{\boldsymbol{\gamma}^{\mathrm{T}} \mathbf{X}_i} \Bigg].$$

3. Asymptotic properties

Denote the true parameter $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}_0, g_0)$ with $\boldsymbol{\alpha}_0 = (\boldsymbol{\beta}_0^{\mathrm{T}}, \boldsymbol{\gamma}_0^{\mathrm{T}})^{\mathrm{T}}$. To establish the asymptotic properties of the proposed estimator $(\widehat{\boldsymbol{\alpha}}_n, \widehat{g}_n)$ with $\widehat{\boldsymbol{\alpha}}_n = (\widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}, \widehat{\boldsymbol{\gamma}}_n^{\mathrm{T}})^{\mathrm{T}}$ and $\widehat{g}_n(t) = \widehat{\mathbf{a}}_n^{\mathrm{T}} \mathbf{B}(t)$, we need the following regularity conditions.

C1. The parameter spaces $\mathcal{B}$ and $\mathcal{T}$ are both compact and contain the true parameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ as their interior points, respectively.

C2. The domain of the covariate $\mathbf{V} \equiv (\mathbf{Z}^{\mathrm{T}}, \mathbf{X}^{\mathrm{T}})^{\mathrm{T}}$, denoted by $\mathcal{V}$, is a bounded subset of $\mathbb{R}^d$, where $d = d_1 + d_2$, and both $E(\mathbf{Z}\mathbf{Z}^{\mathrm{T}})$ and $E(\mathbf{X}\mathbf{X}^{\mathrm{T}})$ are nonsingular.

C3. For $i = 1, 2$, assume that $\boldsymbol{\beta}_i \in \mathcal{B}$, $\boldsymbol{\gamma}_i \in \mathcal{T}$, and $\log \lambda_i(\cdot) \in \mathcal{G}^p$, and denote $\Lambda_i(t) = \int_0^t \lambda_i(s) \, \mathrm{d}s$. If $\Lambda_1(te^{\boldsymbol{\beta}_1^{\mathrm{T}} \mathbf{z}}) e^{\boldsymbol{\gamma}_1^{\mathrm{T}} \mathbf{x}} = \Lambda_2(te^{\boldsymbol{\beta}_2^{\mathrm{T}} \mathbf{z}}) e^{\boldsymbol{\gamma}_2^{\mathrm{T}} \mathbf{x}}$ for any $t \in [0, b]$ and $\mathbf{v} = (\mathbf{z}^{\mathrm{T}}, \mathbf{x}^{\mathrm{T}})^{\mathrm{T}} \in \mathcal{V}$, then $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2$, and $\lambda_1 = \lambda_2$.

C4. Let $\epsilon_0 = Y e^{\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{Z}}$. There exists a truncation time $\tau < \infty$ such that, for some positive constant δ_0, $P(\epsilon_0 > \tau | \mathbf{V}) \geq \delta_0$ almost surely with respect to the probability measure of $\mathbf{V}$.

C5. The conditional density of C given $\mathbf{V}$ and its derivative are uniformly bounded for all possible values of $\mathbf{V}$.

C6. Let $\mathcal{G}^p$ denote the collection of bounded functions g on $[0, b]$ with bounded derivatives $g^{(j)}$, $j = 1, \ldots, k$, such that the kth derivative $g^{(k)}$ satisfies the following Lipschitz continuity condition,

$$\big| g^{(k)}(s) - g^{(k)}(t) \big| \leq L |s - t|^m \qquad \text{for } s, t \in [0, b],$$

where k is a positive integer and $m \in (0, 1]$ such that $p = m + k \geq 3$, and $L < \infty$ is a constant. The true log baseline hazard function $g_0(\cdot) = \log \lambda_0(\cdot)$ belongs to $\mathcal{G}^p$. For notational simplicity, we may also use g' and g'' to denote the first and second derivatives of g, respectively.

C7. For some $\eta \in (0, 1)$, $u^{\mathrm{T}} \mathrm{Var}(\mathbf{V} | \epsilon_0, \Delta = 1) u \geq \eta u^{\mathrm{T}} E(\mathbf{V}\mathbf{V}^{\mathrm{T}} | \epsilon_0, \Delta = 1) u$ almost surely for all $u \in \mathbb{R}^d$. $E(\Delta \mathbf{W}\mathbf{W}^{\mathrm{T}})$ is nonsingular, where $\mathbf{W} = (\{1 + \epsilon_0 g_0'(\epsilon_0)\} \mathbf{Z}^{\mathrm{T}}, \mathbf{X}^{\mathrm{T}})^{\mathrm{T}}$.

C8. Let $M(t) = \Delta I(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \le t) - \int_0^t I(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \ge s)e^{g_0(s)}e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\,\mathrm{d}s$,

$$l^*_{\boldsymbol{\beta}_0}(\mathcal{O}) = \int_0^b \left[\mathbf{Z} - \frac{E\{\mathbf{Z}I(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \ge t)e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\}}{E\{I(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \ge t)e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\}}\right]\{1 + tg_0'(t)\}\,\mathrm{d}M(t),$$

$$l^*_{\boldsymbol{\gamma}_0}(\mathcal{O}) = \int_0^b \left[\mathbf{X} - \frac{E\{\mathbf{X}I(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \ge t)e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\}}{E\{I(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \ge t)e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\}}\right]\mathrm{d}M(t),$$

$$l^*_{\boldsymbol{\alpha}_0}(\mathcal{O}) = \big(l^*_{\boldsymbol{\beta}_0}(\mathcal{O})^{\mathrm{T}}, l^*_{\boldsymbol{\gamma}_0}(\mathcal{O})^{\mathrm{T}}\big)^{\mathrm{T}}, \qquad \mathcal{I}(\boldsymbol{\alpha}_0) = E\big\{l^*_{\boldsymbol{\alpha}_0}(\mathcal{O})^{\otimes 2}\big\},$$

where $\mathcal{O} = (Y, \Delta, \mathbf{Z}, \mathbf{X})$ and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^{\mathrm{T}}$ for a column vector $\mathbf{a}$. Assume that $\mathcal{I}(\boldsymbol{\alpha}_0)$ is nonsingular.

Conditions C1–C2 and C4–C5 are common assumptions in the context of survival analysis. Condition C3 is required to guarantee the identifiability of model (6). Obviously, the model is unidentifiable if and only if $\Lambda_0(t) = c_1 t^{c_2}$ for some positive constants c_1 and c_2 (Chen and Jewell [3]). Condition C6 requires $p \ge 3$ to guarantee the desirable control of the spline approximation error rates of the first and second derivatives of g_0. Condition C7 is a technical assumption and can be justified in many applications. This assumption is also imposed by Wellner and Zhang [23] for the panel count data model and Ding and Nan [7] for the censored linear regression model. Condition C8 is a natural assumption that essentially requires the semiparametric efficiency information matrix to be invertible.

Following Ding and Nan [7], we define

$$\mathcal{H}^p = \big\{\xi(\cdot, \boldsymbol{\beta}) : \xi(t, \mathbf{z}, \boldsymbol{\beta}) = g\big(\psi(t, \mathbf{z}, \boldsymbol{\beta})\big), g \in \mathcal{G}^p, t \in [0, b], \mathbf{z} \in \mathcal{Z}, \boldsymbol{\beta} \in \mathcal{B}\big\},$$

where

$$\psi(t, \mathbf{z}, \boldsymbol{\beta}) = te^{(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{z}}.$$

Here ξ is a composite function of g composed with ψ, and $\xi(t, \mathbf{z}, \boldsymbol{\beta}_0) = g(t)$. We equip the functional space $\mathcal{H}^p$ with the norm $\|\cdot\|_2$ defined as

$$\big\|\xi(\cdot, \boldsymbol{\beta})\big\|_2 = \left[\int_{\mathcal{Z}}\int_0^b \big\{g\big(te^{(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{z}}\big)\big\}^2\,\mathrm{d}\Lambda_0(t)\,\mathrm{d}F_{\mathbf{Z}}(\mathbf{z})\right]^{1/2}$$

for any $\xi(\cdot, \boldsymbol{\beta}) \in \mathcal{H}^p$, where $F_{\mathbf{Z}}(\mathbf{z})$ is the cumulative distribution function of $\mathbf{Z}$. For any $\boldsymbol{\theta}_i = (\boldsymbol{\beta}_i, \boldsymbol{\gamma}_i, \xi_i(\cdot, \boldsymbol{\beta}_i))$, $i = 1, 2$, in the space $\boldsymbol{\Theta}^p \equiv \mathcal{B} \times \mathcal{T} \times \mathcal{H}^p$, define the distance,

$$d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \big(|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2|^2 + |\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2|^2 + \big\|\xi_1(\cdot, \boldsymbol{\beta}_1) - \xi_2(\cdot, \boldsymbol{\beta}_2)\big\|_2^2\big)^{1/2},$$

where $|\cdot|$ is the Euclidean norm. Let $\mathcal{G}_n^p = \mathcal{S}_n(T_{K_n}, K_n, p)$,

$$\mathcal{H}_n^p = \big\{\xi(\cdot, \boldsymbol{\beta}) : \xi(t, \mathbf{z}, \boldsymbol{\beta}) = g\big(\psi(t, \mathbf{z}, \boldsymbol{\beta})\big), g \in \mathcal{G}_n^p, t \in [0, b], \mathbf{z} \in \mathcal{Z}, \boldsymbol{\beta} \in \mathcal{B}\big\},$$

and $\boldsymbol{\Theta}_n^p = \mathcal{B} \times \mathcal{T} \times \mathcal{H}_n^p$. It is easy to see that $\boldsymbol{\Theta}_n^p \subseteq \boldsymbol{\Theta}_{n+1}^p \cdots \subseteq \boldsymbol{\Theta}^p$ for $n \ge 1$. Note that the sieve estimator $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n, \widehat{\xi}_n(\cdot, \widehat{\boldsymbol{\beta}}_n))$ is the maximizer of the empirical log-likelihood over

the sieve space $\boldsymbol{\Theta}_n^p$, where $\widehat{\xi}_n(t, \mathbf{z}, \widehat{\boldsymbol{\beta}}_n) = \widehat{g}_n(te^{(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{z}})$. The following theorem provides the convergence rate of the proposed estimator $\widehat{\boldsymbol{\theta}}_n$ to the true parameter $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \xi_0(\cdot, \boldsymbol{\beta}_0)) = (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, g_0)$.

Theorem 1. *Suppose that conditions* C1–C7 *hold and* $(2p+2)^{-1} < v < (2p)^{-1}$, *then*

$$d(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_p\big(n^{-\min\{pv, (1-v)/2\}}\big).$$

The proof of Theorem 1 is provided in Section 7 by verifying the conditions of Theorem 1 in Shen and Wong [19]. Theorem 1 implies that, if $v = (2p+1)^{-1}$, $d(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_p(n^{-p/(1+2p)})$, which is the optimal convergence rate in the nonparametric setting. Although the overall convergence rate is slower than $n^{-1/2}$, the proposed estimator for the regression parameter $\boldsymbol{\alpha}_0$ is still asymptotically normal at the rate of $n^{-1/2}$ and attains the semiparametric efficiency bound. We summarize these asymptotic results in the following theorem.

Theorem 2. *Suppose that conditions* C1–C8 *hold and* $(2p+2)^{-1} < v < (2p)^{-1}$, *then* $n^{1/2}(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0)$ *converges in distribution to a mean zero normal random vector with covariance matrix* $\mathcal{I}^{-1}(\boldsymbol{\alpha}_0)$ *equal to the semiparametric efficiency bound of* $\boldsymbol{\alpha}_0$.

The proof of Theorem 2 is also presented in Section 7 by checking the conditions in Theorem 6.1 of Wellner and Zhang [23], which relies heavily on the empirical process theory. A consistent estimator for the limiting covariance matrix is summarized by the following theorem.

Theorem 3. *Let* $\widehat{l}^*_{\boldsymbol{\alpha}_0}(\mathcal{O}) = (\widehat{l}^*_{\boldsymbol{\beta}_0}(\mathcal{O})^{\mathrm{T}}, \widehat{l}^*_{\boldsymbol{\gamma}_0}(\mathcal{O})^{\mathrm{T}})^{\mathrm{T}}$, *where*

$$\widehat{l}^*_{\boldsymbol{\beta}_0}(\mathcal{O}) = \int_0^b \big\{\mathbf{Z} - \bar{\mathbf{Z}}(t, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n)\big\}\big\{1 + t\widehat{g}'_n(t)\big\}\,\mathrm{d}\widehat{M}(t),$$

$$\widehat{l}^*_{\boldsymbol{\gamma}_0}(\mathcal{O}) = \int_0^b \big\{\mathbf{X} - \bar{\mathbf{X}}(t, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n)\big\}\,\mathrm{d}\widehat{M}(t),$$

$$\bar{\mathbf{Z}}(t, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{P_n\{\mathbf{Z} I(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}} \geq t) e^{\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}}\}}{P_n\{I(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}} \geq t) e^{\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}}\}},$$

$$\bar{\mathbf{X}}(t, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{P_n\{\mathbf{X} I(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}} \geq t) e^{\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}}\}}{P_n\{I(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}} \geq t) e^{\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}}\}},$$

$$\widehat{M}(t) = \Delta I\big(Ye^{\widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{Z}} \leq t\big) - \int_0^t I\big(Ye^{\widehat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{Z}} \geq s\big) \exp\big\{\widehat{g}_n(s)\big\} e^{\widehat{\boldsymbol{\gamma}}_n^{\mathrm{T}}\mathbf{X}}\,\mathrm{d}s,$$

and P_n *is the empirical measure with respect to* $\mathcal{O}$. *Suppose that conditions in Theorem 2 hold, then* $P_n\{\widehat{l}^*_{\boldsymbol{\alpha}_0}(\mathcal{O})^{\otimes 2}\}$ *converges to* $\mathcal{I}(\boldsymbol{\alpha}_0)$ *in probability.*

4. Simulation studies

We conducted simulation studies to assess the proposed sieve MLE for finite samples. We simulated covariates $\mathbf{Z}$ and $\mathbf{X}$ independently from the Bernoulli distribution with success probability 0.5 and then generated the survival times T from model (6). We set the true parameter values $\boldsymbol{\beta}_0 = 1.5$ and $\boldsymbol{\gamma}_0 = 0.5$ and considered four different baseline hazard functions for $\lambda_0(t)$: (i) $\lambda_0(t) = 1/(1+t)$; (ii) $\lambda_0(t) = (t-0.5)^2$; (iii) $\lambda_0(t) = \log(1+t)$; and (iv) $\lambda_0(t) = 1 + \cos(5t+10)$. For each case, we generated censoring times C from Unif(c_1, c_2) with truncation at $\tau = c_2 - 1$ to achieve censoring rates of 20% and 40%, respectively. We considered the sample size $n = 200$ and 400.

In the implementation of the sieve MLE, we chose the cubic B-spline and took the data-adaptive interior knots as the median of $\{Y_i e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i} : i = 1, \ldots, n\}$ with a given $\boldsymbol{\beta}$ in cases (i)–(iii) and the 20th, 40th, 60th, and 80th quantiles in case (iv). In particular, we adopted the following procedure to obtain the sieve MLE.

(1) Choose initial values $(\widetilde{\boldsymbol{\beta}}^{(0)}, \widetilde{\boldsymbol{\gamma}}^{(0)}, \widetilde{\mathbf{a}}^{(0)})$ and set $k = 0$.
(2) At step $k+1$, obtain $\widetilde{\mathbf{a}}^{(k+1)}$ by solving $\partial l_n(\widetilde{\boldsymbol{\beta}}^{(k)}, \widetilde{\boldsymbol{\gamma}}^{(k)}, \mathbf{a})/\partial \mathbf{a} = \mathbf{0}$ using the Newton–Raphson algorithm with the initial value $\widetilde{\mathbf{a}}^{(k)}$ until the maximum componentwise difference between the two consecutive values is less than 10^{-3}.
(3) Obtain $(\widetilde{\boldsymbol{\beta}}^{(k+1)}, \widetilde{\boldsymbol{\gamma}}^{(k+1)})$ by solving $\partial l_n(\boldsymbol{\beta}, \boldsymbol{\gamma}, \widetilde{\mathbf{a}}^{(k+1)})/\partial(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \mathbf{0}$ using the Newton–Raphson algorithm with initial value $(\widetilde{\boldsymbol{\beta}}^{(k)}, \widetilde{\boldsymbol{\gamma}}^{(k)})$ until the maximum componentwise difference between the two consecutive values is less than 10^{-3}.
(4) Repeat steps (2) and (3) until the maximum componentwise differences of two consecutive values are less than 10^{-3}. The resultant estimators, denoted by $(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n, \widehat{\mathbf{a}}_n)$, are taken as the sieve MLE.

Table 1 summarizes the estimates from 1000 replications for the censoring rate of 20% with $n = 200$ and 400, respectively. The column labeled "EST" is the average value of the estimates, "SE" is the sample standard error of the estimates, "ESE_1" is the average of standard error estimates by inverting the observed information matrix of all parameters including the basis spline coefficients, and "CP_1" is the corresponding coverage proportion of 95% confidence intervals. We also present the column "ESE_2", which is the average of standard error estimates by inverting the estimated information matrix of the regression parameter $\boldsymbol{\alpha}_0$ based on Theorem 3 and list the column "CP_2", which is the corresponding coverage proportion of 95% confidence intervals. The column "MSE" refers to the average value of the mean squared errors.

Clearly, the proposed sieve MLE method performs well under all of the four different baseline hazard functions. The parameter estimates are virtually unbiased for both $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and the bias decreases as the sample size increases. The estimated standard errors by inverting the observed information matrix of all parameters or those by inverting the information matrix based on the efficient score function agree well with the sample standard errors. The coverage probabilities are around the nominal level 95% for all cases. The estimated baseline hazard function using the B-spline approximation under $n = 200$ and $n = 400$ are presented in Figure 1. It can be seen that the estimated baseline hazard functions are reasonably close to the corresponding true curves. We also explored the situation with a censoring rate of 40%. The corresponding results of

Table 1. Simulation results under the proposed accelerated hazards model with a censoring rate of 20%

			Sieve MLE						
$\lambda_0(\cdot)$	n	True value	EST	SE	ESE_1	CP_1	ESE_2	CP_2	$MSE_{\times 10^2}$
(i)	200	$\boldsymbol{\beta}(1.5)$	1.496	0.238	0.232	0.939	0.225	0.936	5.657
		$\boldsymbol{\gamma}(0.5)$	0.511	0.162	0.160	0.947	0.160	0.949	2.628
	400	$\boldsymbol{\beta}(1.5)$	1.503	0.162	0.163	0.948	0.161	0.944	2.606
		$\boldsymbol{\gamma}(0.5)$	0.504	0.114	0.112	0.950	0.112	0.950	1.292
(ii)	200	$\boldsymbol{\beta}(1.5)$	1.505	0.046	0.043	0.935	0.040	0.922	0.210
		$\boldsymbol{\gamma}(0.5)$	0.514	0.168	0.165	0.950	0.167	0.952	2.817
	400	$\boldsymbol{\beta}(1.5)$	1.506	0.031	0.030	0.932	0.028	0.915	0.099
		$\boldsymbol{\gamma}(0.5)$	0.503	0.114	0.116	0.954	0.117	0.954	1.293
(iii)	200	$\boldsymbol{\beta}(1.5)$	1.501	0.099	0.098	0.955	0.096	0.947	0.984
		$\boldsymbol{\gamma}(0.5)$	0.514	0.166	0.163	0.945	0.166	0.950	2.774
	400	$\boldsymbol{\beta}(1.5)$	1.503	0.068	0.069	0.952	0.069	0.950	0.466
		$\boldsymbol{\gamma}(0.5)$	0.498	0.111	0.114	0.961	0.115	0.962	1.236
(iv)	200	$\beta(1.5)$	1.493	0.115	0.115	0.911	0.099	0.902	1.311
		$\boldsymbol{\gamma}(0.5)$	0.515	0.169	0.162	0.944	0.163	0.942	2.865
	400	$\boldsymbol{\beta}(1.5)$	1.498	0.079	0.077	0.924	0.072	0.913	0.623
		$\boldsymbol{\gamma}(0.5)$	0.504	0.117	0.114	0.954	0.114	0.951	1.358

*EST, the average value of the parameter estimates; SE, the sample standard error of the estimates; ESE_1, the estimate of the standard error by inverting the information matrix of all parameters; CP_1, the corresponding coverage probability of 95% confidence intervals; ESE_2, the estimate of the standard error by inverting the information matrix based on the efficient score function; CP_2, the corresponding coverage probability of 95% confidence intervals; MSE, the mean squared errors of the parameter estimates.

the estimates for the regression parameters based on 1000 replications are presented in Table 2, from which similar conclusions can be drawn as before. Moreover, the estimated baseline hazard functions are plotted in Figure 2, which deteriorate slightly compared with those in Figure 1.

5. Application

As an illustration, we applied the proposed general class of accelerated hazards models to a study of bone marrow transplantation with 137 patients of acute leukemia (Copelan *et al.* [5] and Klein and Moeschberger [11]). The disease-free survival time, including the time to relapse, death, or the end of study, is of primary interest. Patients were followed for approximate 7.2 years, of whom around 39.4% were censored. Several potential risk factors were measured at the time of transplantation. Patients were classified into three risk categories based on their disease status: 38 patients with acute lymphoblastic leukemia (ALL), 54 patients with acute myeloctic leukemia (AML) low risk, and 45 patients with AML high risk. Both patients and donors' ages and the waiting times from diagnosis to transplantation were recorded. The AML patients with their French–American–British (FAB) classification of grade 4 or 5 based on standard morphological criteria were also considered as a covariate in our regression model. Patients were either given

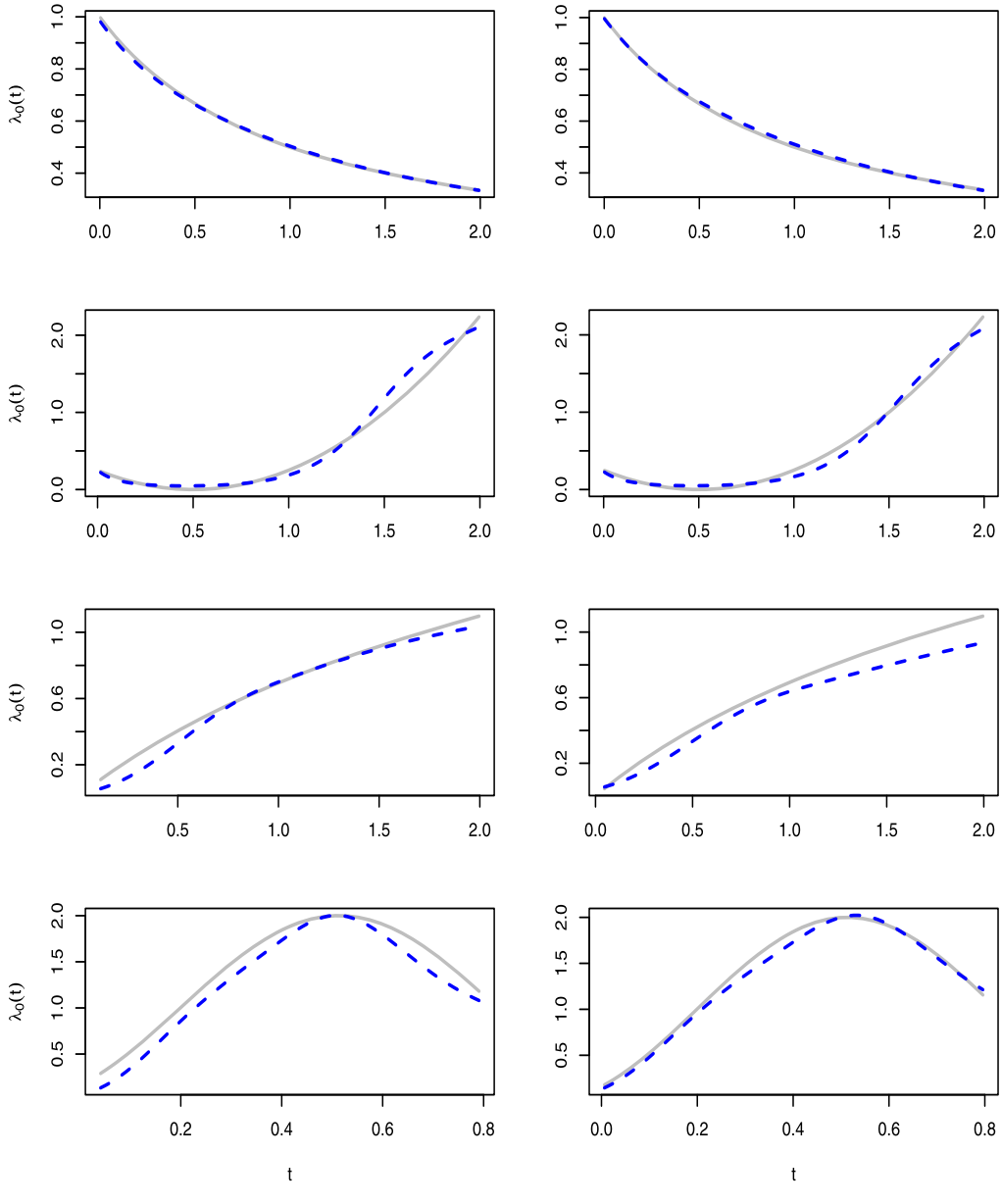

Figure 1. True baseline hazard function (solid line) and its estimate (dashed line) using the B-spline approximation under $n = 200$ (left panel) and $n = 400$ (right panel) with a censoring rate of 20%. From top to bottom, the plots correspond to cases (i) to (iv) for the baseline hazard functions.

Table 2. Simulation results under the proposed accelerated hazards model with a censoring rate of 40%

			Sieve MLE						
$\lambda_0(\cdot)$	n	True value	EST	SE	ESE_1	CP_1	ESE_2	CP_2	$MSE_{\times 10^2}$
(i)	200	$\boldsymbol{\beta}(1.5)$	1.501	0.258	0.262	0.952	0.250	0.949	6.625
		$\boldsymbol{\gamma}(0.5)$	0.518	0.188	0.185	0.950	0.185	0.949	3.539
	400	$\boldsymbol{\beta}(1.5)$	1.504	0.180	0.180	0.942	0.178	0.936	3.241
		$\boldsymbol{\gamma}(0.5)$	0.504	0.130	0.130	0.941	0.130	0.941	1.679
(ii)	200	$\boldsymbol{\beta}(1.5)$	1.501	0.069	0.063	0.921	0.061	0.916	0.472
		$\boldsymbol{\gamma}(0.5)$	0.523	0.196	0.189	0.947	0.194	0.949	3.864
	400	$\boldsymbol{\beta}(1.5)$	1.498	0.048	0.044	0.925	0.042	0.913	0.235
		$\boldsymbol{\gamma}(0.5)$	0.509	0.130	0.132	0.957	0.135	0.962	1.696
(iii)	200	$\boldsymbol{\beta}(1.5)$	1.511	0.142	0.138	0.943	0.139	0.943	2.025
		$\boldsymbol{\gamma}(0.5)$	0.520	0.196	0.189	0.952	0.191	0.953	3.843
	400	$\boldsymbol{\beta}(1.5)$	1.506	0.099	0.097	0.946	0.099	0.948	0.977
		$\boldsymbol{\gamma}(0.5)$	0.501	0.129	0.132	0.956	0.133	0.958	1.666
(iv)	200	$\boldsymbol{\beta}(1.5)$	1.504	0.127	0.146	0.936	0.117	0.935	1.605
		$\boldsymbol{\gamma}(0.5)$	0.516	0.196	0.187	0.947	0.187	0.943	3.843
	400	$\boldsymbol{\beta}(1.5)$	1.508	0.090	0.097	0.948	0.086	0.933	0.811
		$\boldsymbol{\gamma}(0.5)$	0.504	0.126	0.131	0.963	0.131	0.961	1.577

*EST, the average value of the parameter estimates; SE, the sample standard error of the estimates; ESE_1, the estimate of the standard error by inverting the information matrix of all parameters; CP_1, the corresponding coverage probability of 95% confidence intervals; ESE_2, the estimate of the standard error by inverting the information matrix based on the efficient score function; CP_2, the corresponding coverage probability of 95% confidence intervals; MSE, the mean squared errors of the parameter estimates.

a graft-versus-host prophylactic combining methotrexate (MTX) with cyclosporine and possibly methlprednisolone or given only a combination of cyclosporine and methlprednisolone. In our analysis, we used $X_1 = 1$ to indicate the patient with AML low risk and $X_1 = 0$ otherwise, $X_2 = 1$ to indicate the patient with AML high risk and $X_2 = 0$ otherwise, X_3 to denote the patient's age centered by 28 years, X_4 to denote the donor's age centered by 28 years, $X_5 = 1$ to indicate the AML patient with FAB grade 4 or 5 and $X_5 = 0$ otherwise, and X_6 to denote the patient's waiting time from diagnosis to transplantation centered by 9 months, and $X_7 = 1$ to indicate the patient treated with MTX and $X_7 = 0$ otherwise.

To make a preliminary investigation of whether the hazards of the patients within three different risk categories were identical at the beginning of study, we plotted the kernel-smoothed hazard rate functions with bandwidth 100 days in Figure 3. It can be observed that the smoothed hazards of patients with ALL and AML low risk are almost the same at the initiation of the study. However, the smoothed hazard of patients with AML high risk appears to be slightly higher than those of the other two at time $t = 0$, and increases to a higher level and then lies between those of patients with ALL and patients with AML low risk during the later follow-up of the study. Figure 3 reveals that the hazards may not be proportional from the very beginning of the study, which results in the usual proportional hazards assumption questionable. Intuitively, it is more appealing

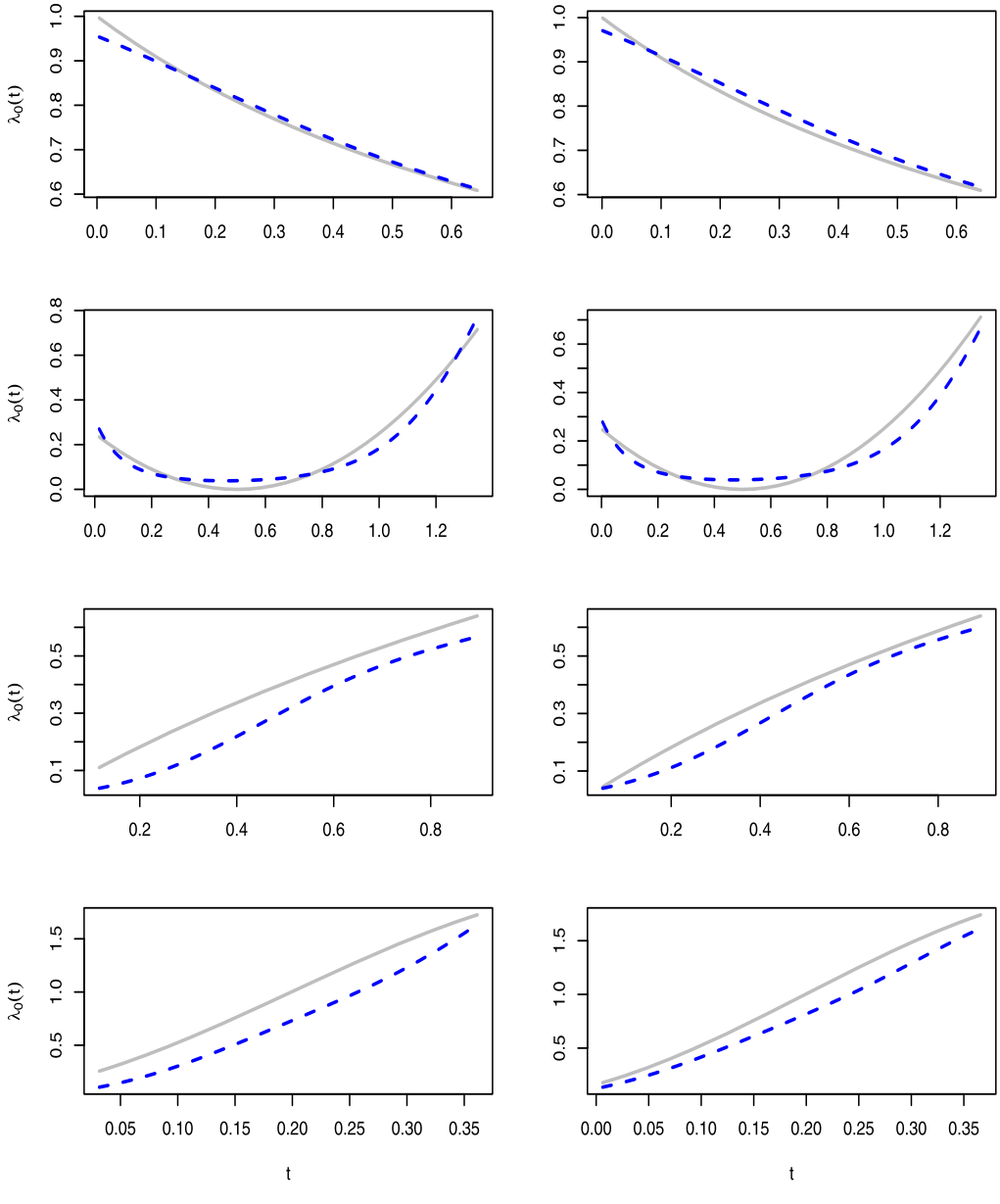

Figure 2. True baseline hazard function (solid line) and its estimate (dashed line) using the B-spline approximation under $n = 200$ (left panel) and $n = 400$ (right panel) with a censoring rate of 40%. From top to bottom, the plots correspond to cases (i) to (iv) for the baseline hazard functions.

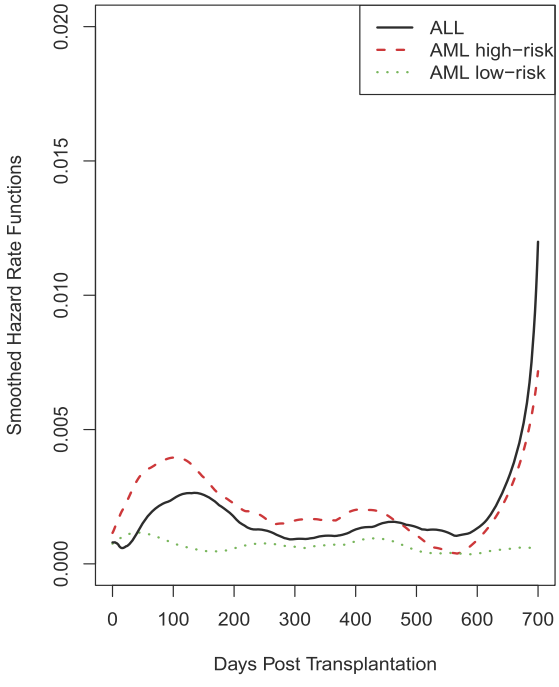

Figure 3. Smoothed hazard rate functions for patients with ALL, AML high-risk, and AML low-risk, respectively.

to consider the time scaled effects of risk status categories as well as the proportional effects of all the risk factors by employing our proposed general accelerated hazards model through setting $\mathbf{Z} = (X_1, X_2)^{\mathrm{T}}$ and $\mathbf{X} = (X_1, \ldots, X_7)^{\mathrm{T}}$. We applied the proposed sieve MLE with smoothing splines to fit the data. For comparison, we fitted the Cox proportional hazards model to the data without taking into consideration of the time scaled effects of risk status categories. The estimation results are summarized in Table 3. For the regression parameters in the proportional hazards component, all of the three considered methods agree in general: Patients without FAB grade 4 or 5 and those with AML low risk were associated with lower hazard risks and thus led to longer progression-free survival times, while the effects of other covariates were not significant. For the time scaled effects of risk status categories, the sieved MLE method shows that patients with AML low risk had significant decelerated hazard risks while the scaled time effect of patients with AML high risk was not significant.

Figure 4 exhibits the estimate of the baseline hazard function $\lambda_0(t)$ using the proposed sieve MLE method. Clearly, patients at the beginning of post transplantation would suffer from the drastically increasing risk due to the incompatibility between the donor and patient, and then the hazard gradually decreased with time.

Table 3. Analysis results of the bone marrow transplantation data

Estimation	Z_1 (AML L)	Z_2 (AML H)	X_1 (AML L)	X_2 (AML H)	X_3 (P Age)	X_4 (D Age)	X_5 (FAB)	X_6 (Wait T)	X_7 (MTX)
Sieve MLE									
EST	−0.651	−0.128	−0.716	−0.033	0.009	0.000	0.804	−0.011	0.348
ESE_1	0.045	0.214	0.370	0.380	0.020	0.018	0.276	0.012	0.252
p-value$_1$	<0.001	0.548	0.053	0.930	0.650	0.994	0.004	0.329	0.166
ESE_2	0.119	0.105	0.365	0.371	0.020	0.016	0.269	0.010	0.240
p-value$_2$	<0.001	0.220	0.050	0.929	0.652	0.993	0.003	0.262	0.147
Cox model									
EST			−1.051	−0.188	0.012	−0.001	0.812	−0.011	0.294
ESE			0.368	0.359	0.020	0.018	0.275	0.011	0.250
p-value			0.004	0.600	0.530	0.940	0.003	0.310	0.240

*EST, the parameter estimates; ESE_1, the estimate of the standard error by inverting the information matrix of all parameters; ESE_2, the estimate of the standard error by inverting the information matrix based on the efficient score function; ESE, the estimate of the standard error by inverting the information matrix based on the Cox proportional hazards regression model.

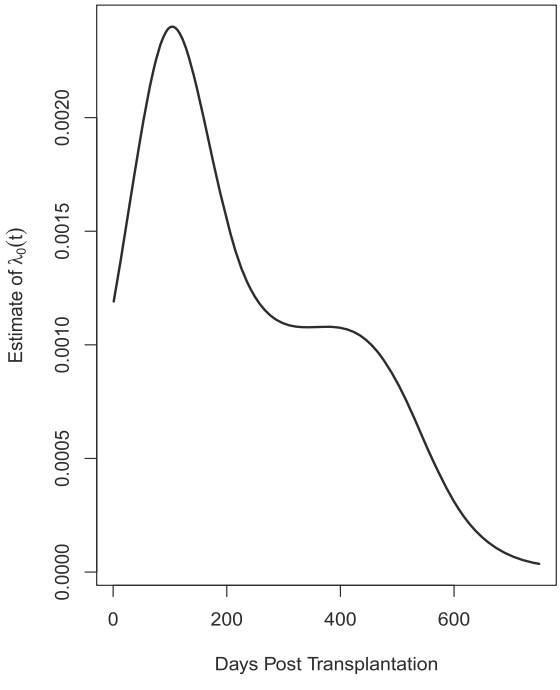

Figure 4. Estimated baseline hazard $\lambda_0(t)$ using the proposed sieve MLE method for the bone marrow transplantation data.

6. Remark

The general accelerated hazards model enables us to evaluate the time scaled effects and the proportional hazards effects of covariates simultaneously. However, it is difficult in practice to classify the risk factors rigorously into either the time scaled or the proportional hazards components of the model. It often depends on the objectives of the study, the interest of the investigator, and the underlying biological process. If there is no such biological information as guidance, some data-driven methods could be used for the classification of covariates. For example, when the number of covariates is small, all the possible models from different combinations of covariates in the time scaled and the proportional hazards parts can be considered. To facilitate the selection of the models, some criteria for evaluating the goodness of model fitting should be considered. When the number of covariates is moderately large, this exhaustive method could be time-consuming, while similar automatic structure discovery procedures as presented in Zhang, Cheng and Liu [26] may warrant further research.

7. Proofs of theorems

Before proving the theorems presented in Section 3, we introduce some useful lemmas. Define

$$\mathcal{H} = \left\{ h : h(\cdot, \boldsymbol{\beta}) = \left. \frac{\partial \xi_\eta(\cdot, \boldsymbol{\beta})}{\partial \eta} \right|_{\eta=0} = w\big(\psi(\cdot, \boldsymbol{\beta})\big), \xi_\eta \in \mathcal{H}^p \right\}.$$

Lemma 1. *Denote*

$$l\big(\boldsymbol{\beta}, \boldsymbol{\gamma}, \xi(\cdot, \boldsymbol{\beta}); \mathcal{O}\big) = \Delta\big\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X} + g\big(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\big)\big\} - \int_0^{Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}} \exp\big\{g(s) + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\big\}\,\mathrm{d}s.$$

Under conditions C1, C2, C4 *and* C6, l *has bounded and continuous first and second derivatives with respect to* $\boldsymbol{\beta} \in \mathcal{B}$, $\boldsymbol{\gamma} \in \mathcal{T}$, *and* $\xi(\cdot, \boldsymbol{\beta}) \in \mathcal{H}^p$.

Proof. After some algebraic calculations, we have

$$\begin{aligned}
l'_{\boldsymbol{\beta}}\big(\boldsymbol{\beta}, \boldsymbol{\gamma}, \xi(\cdot, \boldsymbol{\beta}); \mathcal{O}\big) &= \Delta\mathbf{Z}\big\{1 + Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}} g'\big(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\big)\big\} \\
&\quad - \mathbf{Z}Y \exp\big\{g\big(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\big) + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\big\}, \\
l'_{\boldsymbol{\gamma}}\big(\boldsymbol{\beta}, \boldsymbol{\gamma}, \xi(\cdot, \boldsymbol{\beta}); \mathcal{O}\big) &= \Delta\mathbf{X} - \mathbf{X}\int_0^{Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}} \exp\big\{g(s) + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\big\}\,\mathrm{d}s, \\
l'_{\xi}\big(\boldsymbol{\beta}, \boldsymbol{\gamma}, \xi(\cdot, \boldsymbol{\beta}); \mathcal{O}\big)\big[h(\cdot, \boldsymbol{\beta})\big] &= \Delta w\big(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\big) - \int_0^{Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}} \exp\big\{g(s) + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\big\} w(s)\,\mathrm{d}s, \\
l''_{\boldsymbol{\beta}\boldsymbol{\beta}}\big(\boldsymbol{\beta}, \boldsymbol{\gamma}, \xi(\cdot, \boldsymbol{\beta}); \mathcal{O}\big) &= \Delta\mathbf{Z}\mathbf{Z}^{\mathrm{T}} Y^2 e^{2\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}} g''\big(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\big) + \Delta\mathbf{Z}\mathbf{Z}^{\mathrm{T}} Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}} g'\big(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\big)
\end{aligned}$$

$$
\begin{aligned}
&-\mathbf{Z}\mathbf{Z}^{\mathrm{T}}Y^2 g'\big(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\big)\exp\big\{g\big(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\big)+2\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}+\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\big\}\\
&-\mathbf{Z}\mathbf{Z}^{\mathrm{T}}Y\exp\big\{g\big(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\big)+\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}+\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\big\},\\
l''_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\big(\boldsymbol{\beta},\boldsymbol{\gamma},\xi(\cdot,\boldsymbol{\beta});\mathcal{O}\big)&=-\mathbf{X}\mathbf{X}^{\mathrm{T}}\int_0^{Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}}\exp\big\{g(s)+\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\big\}\,\mathrm{d}s,\\
l''_{\boldsymbol{\beta}\boldsymbol{\gamma}}\big(\boldsymbol{\beta},\boldsymbol{\gamma},\xi(\cdot,\boldsymbol{\beta});\mathcal{O}\big)&=-Y\mathbf{Z}\mathbf{X}^{\mathrm{T}}\exp\big\{g\big(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\big)+\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}+\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\big\},\\
l''_{\boldsymbol{\beta}\xi}\big(\boldsymbol{\beta},\boldsymbol{\gamma},\xi(\cdot,\boldsymbol{\beta});\mathcal{O}\big)\big[h(\cdot,\boldsymbol{\beta})\big]&=\Delta Y\mathbf{Z}e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}w'\big(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\big)\\
&\quad-Y\mathbf{Z}w\big(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\big)\exp\big\{g\big(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\big)+\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}+\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\big\},\\
l''_{\boldsymbol{\gamma}\xi}\big(\boldsymbol{\beta},\boldsymbol{\gamma},\xi(\cdot,\boldsymbol{\beta});\mathcal{O}\big)[h]&=-\mathbf{X}\int_0^{Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}}\exp\big\{g(s)+\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\big\}w(s)\,\mathrm{d}s,\\
l''_{\xi\xi}\big(\boldsymbol{\beta},\boldsymbol{\gamma},\xi(\cdot,\boldsymbol{\beta});\mathcal{O}\big)[h_1,h_2]&=-\int_0^{Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}}\exp\big\{g(s)+\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\big\}w_1(s)w_2(s)\,\mathrm{d}s,
\end{aligned}
$$

where $h(\cdot,\boldsymbol{\beta})=w(\psi(\cdot,\boldsymbol{\beta}))$, $h_1=w_1(\psi(\cdot,\boldsymbol{\beta}))$, $h_2(\cdot,\boldsymbol{\beta})=w_2(\psi(\cdot,\boldsymbol{\beta}))\in\mathcal{H}$. Under conditions C1, C2, C4 and C6, all the above derivatives are continuous and bounded.

Employing Corollary 6.21 in Schumaker [17], we directly have the following lemma with its proof omitted. □

Lemma 2. *For $g_0\in\mathcal{G}^p$, there exists a function $g_{0n}\in\mathcal{G}_n^p$ such that*

$$\|g_{0n}-g_0\|_\infty=O\big(n^{-pv}\big),$$

where $\|\cdot\|_\infty$ is the sup-norm.

Lemma 3. *Let $\boldsymbol{\theta}_{0n}=(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\xi_{0n}(\cdot,\boldsymbol{\beta}_0))$ with $\xi_{0n}(\cdot,\boldsymbol{\beta}_0)=g_{0n}$ defined in Lemma 2, and $\mathcal{F}_n=\{l(\boldsymbol{\theta};\mathcal{O})-l(\boldsymbol{\theta}_{0n};\mathcal{O}):\boldsymbol{\theta}\in\boldsymbol{\Theta}_n^p\}$. If conditions* C1–C4 *and* C6 *hold, then the ε-bracketing number associated with $\|\cdot\|_\infty$ for $\mathcal{F}_n$, denoted by $N_{[]}(\varepsilon,\mathcal{F}_n,\|\cdot\|_\infty)$, is bounded by $(1/\varepsilon)^{cq_n+d}$, i.e.,*

$$N_{[]}\big(\varepsilon,\mathcal{F}_n,\|\cdot\|_\infty\big)\lesssim(1/\varepsilon)^{cq_n+d}$$

for a constant c. Hereafter, we use the symbol $\lesssim$ to denote that the left-hand side is bounded above by a constant times the right-hand side.

Proof. Denote the ceiling of x by $\lceil x\rceil$. By the calculation in Shen and Wong ([19], page 597), for any $\varepsilon>0$, there exists a set of brackets

$$\big\{\big[g_i^L,g_i^U\big]:i=1,\ldots,\lceil(1/\varepsilon)^{cq_n}\rceil\big\}$$

such that for any $g\in\mathcal{G}_n^p$, $g_i^L(t)\le g(t)\le g_i^U(t)$ over $t\in[0,b]$ for some $1\le i\le\lceil(1/\varepsilon)^{cq_n}\rceil$, where $\|g_i^U-g_i^L\|_\infty\le\varepsilon$ and c is a constant. Since $\mathcal{B}$ and $\mathcal{T}$ are both compact under condition

C1, $\mathcal{B}$ and $\mathcal{T}$ can be covered by $\lceil c_2(1/\varepsilon)^{d_1}\rceil$ and $\lceil c_3(1/\varepsilon)^{d_2}\rceil$ balls with radius ε, respectively. Thus, for any $\boldsymbol{\beta}\in\mathcal{B}$ and $\boldsymbol{\gamma}\in\mathcal{T}$, there exist $\boldsymbol{\beta}_\ell$ for some $1\le\ell\le\lceil c_2(1/\varepsilon)^{d_1}\rceil$ and $\boldsymbol{\gamma}_k$ for some $1\le k\le\lceil c_3(1/\varepsilon)^{d_2}\rceil$ such that $|\boldsymbol{\beta}-\boldsymbol{\beta}_\ell|\le\varepsilon$ and $|\boldsymbol{\gamma}-\boldsymbol{\gamma}_k|\le\varepsilon$. Hence, $|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}-\boldsymbol{\beta}_\ell^{\mathrm{T}}\mathbf{Z}|\le c_4\varepsilon$ and $|\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}-\boldsymbol{\gamma}_k^{\mathrm{T}}\mathbf{X}|\le c_4\varepsilon$ for some constant c_4 under condition C2. Define

$$m_{i,\ell,k}^{L}(\mathcal{O})=\Delta\bigl\{\boldsymbol{\beta}_\ell^{\mathrm{T}}\mathbf{Z}+\boldsymbol{\gamma}_k^{\mathrm{T}}\mathbf{X}-2c_4\varepsilon+g_i^{L}\bigl(Ye^{\boldsymbol{\beta}_\ell^{\mathrm{T}}\mathbf{Z}+\xi_{il}}\bigr)\bigr\}-\int_0^{Ye^{\boldsymbol{\beta}_\ell^{\mathrm{T}}\mathbf{Z}+c_4\varepsilon}}\exp\bigl\{g_i^{U}(s)+\boldsymbol{\gamma}_k^{\mathrm{T}}\mathbf{X}+c_4\varepsilon\bigr\}\,\mathrm{d}s-l(\boldsymbol{\theta}_{0n};U)$$

and

$$m_{i,\ell,k}^{U}(\mathcal{O})=\Delta\bigl\{\boldsymbol{\beta}_\ell^{\mathrm{T}}\mathbf{Z}+\boldsymbol{\gamma}_k^{\mathrm{T}}\mathbf{X}+2c_4\varepsilon+g_i^{U}\bigl(Ye^{\boldsymbol{\beta}_\ell^{\mathrm{T}}\mathbf{Z}+\xi_{iu}}\bigr)\bigr\}-\int_0^{Ye^{\boldsymbol{\beta}_\ell^{\mathrm{T}}\mathbf{Z}-c_4\varepsilon}}\exp\bigl\{g_i^{L}(s)+\boldsymbol{\gamma}_k^{\mathrm{T}}\mathbf{X}-c_4\varepsilon\bigr\}\,\mathrm{d}s-l(\boldsymbol{\theta}_{0n};\mathcal{O}),$$

where

$$g_i^{L}\bigl(e^{\boldsymbol{\beta}_\ell^{\mathrm{T}}\mathbf{Z}+\xi_{i\ell}}\bigr)=\min_{|s|\le c_4\varepsilon}g_i^{L}\bigl(e^{\boldsymbol{\beta}_\ell^{\mathrm{T}}\mathbf{Z}+s}\bigr)\quad\text{and}\quad g_i^{U}\bigl(e^{\boldsymbol{\beta}_\ell^{\mathrm{T}}\mathbf{Z}+\xi_{iu}}\bigr)=\max_{|s|\le c_4\varepsilon}g_i^{U}\bigl(e^{\boldsymbol{\beta}_\ell^{\mathrm{T}}\mathbf{Z}+s}\bigr).$$

After some calculations, we have $|m_{i,\ell,k}^{U}(\mathcal{O})-m_{i,\ell,k}^{L}(\mathcal{O})|\lesssim\varepsilon$ and for any $m(\boldsymbol{\theta};\mathcal{O})\in\mathcal{F}_n$, there exist some i, ℓ, and k such that $m(\boldsymbol{\theta};\mathcal{O})\in[m_{i,\ell,k}^{L}(\mathcal{O}),m_{i,\ell,k}^{U}(\mathcal{O})]$. Therefore, we have

$$N_{[]}\bigl(\varepsilon,\mathcal{F}_n,\|\cdot\|_\infty\bigr)\lesssim(1/\varepsilon)^{cq_n}(1/\varepsilon)^{d_1}(1/\varepsilon)^{d_2}=(1/\varepsilon)^{cq_n+d}$$

for a constant c, which completes the proof. □

Lemma 4. *Let*

$$\bigl(w_1^*(t),\ldots,w_{d_1}^*(t)\bigr)^{\mathrm{T}}=\bigl\{1+tg_0'(t)\bigr\}\frac{E\{\mathbf{Z}I(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\ge t)e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\}}{E\{I(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\ge t)e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\}}$$

and

$$\bigl(w_{d_1+1}^*(t),\ldots,w_d^*(t)\bigr)^{\mathrm{T}}=\frac{E\{\mathbf{X}I(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\ge t)e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\}}{E\{I(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\ge t)e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\}}.$$

If conditions C1–C6 *hold, then there exist* $w_{jn}^*\in\mathcal{G}_n^2$ $(j=1,\ldots,d)$ *such that* $\|w_{jn}^*-w_j^*\|_\infty=O(n^{-2v})$, $j=1,\ldots,d$.

Proof. Note that $E\{I(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\ge t)|\mathbf{Z},\mathbf{X}\}=S_{C|\mathbf{Z},\mathbf{X}}(te^{-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}})\exp\{-\Lambda_0(t)\exp(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X})\}$, where $S_{C|\mathbf{Z},\mathbf{X}}(\cdot)$ is the conditional survival function of C given $\mathbf{Z}$ and $\mathbf{X}$. It can be shown that the first and second derivatives of w_j^* are bounded under conditions C1–C6. Thus, according to Corollary 6.21 of Schumaker [17], the conclusion of this lemma follows. □

Lemma 5. *Let* $h_j^*(\cdot,\boldsymbol{\beta}) = w_j^*(\psi(\cdot,\boldsymbol{\beta}))$ *with* $h_j^*(t,\mathbf{z},\boldsymbol{\beta}_0) = w_j^*(t)$ *where* w_j^* *is defined in Lemma 4,* $j = 1,\dots,d$. *For* $\eta > 0$, *denote*

$$\mathcal{F}_{jn}(\eta) = \big\{l'_\xi(\boldsymbol{\theta};\mathcal{O})\big[h_j^* - h_j\big] : \boldsymbol{\theta}\in\boldsymbol{\Theta}_n^p, h_j\in\mathcal{H}_n^p, d(\boldsymbol{\theta},\boldsymbol{\theta}_0)\le\eta, \big\|h_j^* - h_j\big\|_\infty \le \eta\big\}.$$

If conditions C1–C6 *hold, then* $N_{[]}(\varepsilon,\mathcal{F}_{jn}(\eta),\|\cdot\|_\infty) \lesssim (\eta/\varepsilon)^{cq_n+d}$ *for a constant* c.

Lemma 6. *Define*

$$\mathcal{F}_{jn}^{\beta}(\eta) = \big\{l'_{\beta_j}(\boldsymbol{\theta};\mathcal{O}) - l'_{\beta_j}(\boldsymbol{\theta}_0;\mathcal{O}) : \boldsymbol{\theta}\in\boldsymbol{\Theta}_n^p, d(\boldsymbol{\theta},\boldsymbol{\theta}_0)\le\eta, \big\|g'\big(\psi(\cdot,\boldsymbol{\beta})\big) - g_0'\big(\psi(\cdot,\boldsymbol{\beta}_0)\big)\big\|_2\le\eta\big\},$$
$$j = 1,\dots,d_1,$$
$$\mathcal{F}_{jn}^{\gamma}(\eta) = \big\{l'_{\gamma_j}(\boldsymbol{\theta};\mathcal{O}) - l'_{\gamma_j}(\boldsymbol{\theta}_0;\mathcal{O}) : \boldsymbol{\theta}\in\boldsymbol{\Theta}_n^p, d(\boldsymbol{\theta},\boldsymbol{\theta}_0)\le\eta\big\}, \qquad j = 1,\dots,d_2,$$

and

$$\mathcal{F}_{jn}^{\xi}(\eta) = \big\{l'_\xi(\boldsymbol{\theta};\mathcal{O})\big[h_j^*(\cdot,\boldsymbol{\beta})\big] - l'_\xi(\boldsymbol{\theta}_0;\mathcal{O})\big[h_j^*(\cdot,\boldsymbol{\beta}_0)\big] : \boldsymbol{\theta}\in\boldsymbol{\Theta}_n^p, d(\boldsymbol{\theta},\boldsymbol{\theta}_0)\le\eta\big\},$$
$$j = 1,\dots,d,$$

where $l'_{\beta_j}(\boldsymbol{\theta};\mathcal{O})$ *and* $l'_{\gamma_j}(\boldsymbol{\theta};\mathcal{O})$ *are the* j*th element of* $l'_{\boldsymbol{\beta}}(\boldsymbol{\theta};\mathcal{O})$ *and* $l'_{\boldsymbol{\gamma}}(\boldsymbol{\theta};\mathcal{O})$, *respectively, and* h_j^* *is defined in Lemma 5. Suppose that conditions* C1–C6 *hold, then*

$$N_{[]}\big(\varepsilon,\mathcal{F}_{jn}^{\beta}(\eta),\|\cdot\|_\infty\big) \lesssim (\eta/\varepsilon)^{c_1q_n+d},$$
$$N_{[]}\big(\varepsilon,\mathcal{F}_{jn}^{\gamma}(\eta),\|\cdot\|_\infty\big) \lesssim (\eta/\varepsilon)^{c_2q_n+d},$$
$$N_{[]}\big(\varepsilon,\mathcal{F}_{jn}^{\xi}(\eta),\|\cdot\|_\infty\big) \lesssim (\eta/\varepsilon)^{c_3q_n+d},$$

for some constants c_1, c_2, *and* c_3.

The proofs of Lemmas 5 and 6 are similar to that of Lemma 3 and thus omitted here for the sake of space. The detailed proofs are available as supplementary materials from the authors.

Proof of Theorem 1. To obtain the convergence rate of the proposed estimator, we need to verify conditions C1–C3 of Theorem 1 in Shen and Wong [19]. Some algebraic calculations yield that

$$E\big\{l\big(\boldsymbol{\beta},\boldsymbol{\gamma},\xi(\cdot,\boldsymbol{\beta});\mathcal{O}\big)\big\} = E\big[\Delta\big\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X} + g\big(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\big)\big\}\big] - E\big[\Delta\exp\big\{\big(g\big(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\big) + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\big) - \big\{g_0\big(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big) + \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}\big\}\big\}\big]$$

and

$$E\big\{l\big(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\xi_0(\cdot,\boldsymbol{\beta}_0);\mathcal{O}\big)\big\} - E\big\{l\big(\boldsymbol{\beta},\boldsymbol{\gamma},\xi(\cdot,\boldsymbol{\beta});\mathcal{O}\big)\big\}$$
$$= E\big[\Delta\big\{\exp\big(\big\{g\big(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\big) + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\big\} - \big\{g_0\big(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big) + \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}\big\}\big)$$

$$
\begin{aligned}
&- 1 - \bigl(\bigl\{g\bigl(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\bigr) + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\bigr\} - \bigl\{g_0\bigl(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} + \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}\bigr)\bigr\}\bigr)\bigr\}\bigr] \qquad (9)\\
&\geq \frac{1}{2}E\bigl\{\Delta\bigl(\bigl\{g\bigl(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\bigr) + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\bigr\} - \bigl\{g_0\bigl(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\bigr) + \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}\bigr\}\bigr)^2\bigr\}\\
&\geq \frac{1}{2}E\bigl\{\Delta\bigl(\bigl\{g\bigl(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\bigr) - g_0\bigl(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\bigr)\bigr\} + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{Z} + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^{\mathrm{T}}\mathbf{X}\bigr)^2\bigr\}.
\end{aligned}
$$

Using the Taylor expansion, we have

$$
\begin{aligned}
&E\bigl\{\Delta\bigl(\bigl\{g\bigl(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\bigr) - g_0\bigl(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\bigr)\bigr\} + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{Z} + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^{\mathrm{T}}\mathbf{X}\bigr)^2\bigr\}\\
&\quad = E\bigl\{\Delta\bigl(\bigl\{g\bigl(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}\bigr) - g\bigl(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\bigr)\bigr\} + \bigl\{g\bigl(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\bigr) - g_0\bigl(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\bigr)\bigr\}\\
&\qquad + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{Z} + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^{\mathrm{T}}\mathbf{X}\bigr)^2\bigr\}\\
&\quad = E\bigl(\Delta\bigl[g'(\varepsilon_0)\varepsilon_0(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{Z} + O\bigl(|\boldsymbol{\beta} - \boldsymbol{\beta}_0|^2\bigr) + \bigl\{g(\varepsilon_0) - g_0(\varepsilon_0)\bigr\}\\
&\qquad + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{Z} + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^{\mathrm{T}}\mathbf{X}\bigr]^2\bigr)\\
&\quad = E\bigl(\Delta\bigl[\bigl\{g(\varepsilon_0) - g_0(\varepsilon_0)\bigr\} + (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^{\mathrm{T}}\mathbf{W}\bigr]^2\bigr) + o\bigl(d^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0)\bigr),
\end{aligned} \qquad (10)
$$

where $\mathbf{W} = (\{g_0'(\varepsilon_0)\varepsilon_0 + 1\}\mathbf{Z}^{\mathrm{T}}, \mathbf{X}^{\mathrm{T}})^{\mathrm{T}}$. Obviously,

$$
\begin{aligned}
&E\bigl(\Delta\bigl[\bigl\{g(\varepsilon_0) - g_0(\varepsilon_0)\bigr\} + (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^{\mathrm{T}}\mathbf{W}\bigr]^2\bigr)\\
&\quad \geq E\bigl[\Delta\bigl\{g(\varepsilon_0) - g_0(\varepsilon_0)\bigr\}^2\bigr] + E\bigl[\Delta\bigl\{(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^{\mathrm{T}}\mathbf{W}\bigr\}^2\bigr]\\
&\qquad - 2\bigl|E\bigl[\Delta\bigl\{g(\varepsilon_0) - g_0(\varepsilon_0)\bigr\}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^{\mathrm{T}}\mathbf{W}\bigr]\bigr|.
\end{aligned} \qquad (11)
$$

On the other hand, it follows from the Cauchy–Schwarz inequality and condition C7 that

$$
\begin{aligned}
&\bigl|E\bigl[\Delta\bigl\{g(\varepsilon_0) - g_0(\varepsilon_0)\bigr\}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^{\mathrm{T}}\mathbf{W}\bigr]\bigr|^2\\
&\quad = \bigl|E\bigl[\Delta\bigl\{g(\varepsilon_0) - g_0(\varepsilon_0)\bigr\}E\bigl\{(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^{\mathrm{T}}\mathbf{W}|\varepsilon_0, \Delta = 1\bigr\}\bigr]\bigr|^2\\
&\quad \leq E\bigl[\Delta\bigl\{g(\varepsilon_0) - g_0(\varepsilon_0)\bigr\}^2\bigr]E\bigl[E\bigl\{(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^{\mathrm{T}}\mathbf{W}|\varepsilon_0, \Delta = 1\bigr\}\bigr]^2\\
&\quad \leq (1 - \eta)E\bigl[\Delta\bigl\{g(\varepsilon_0) - g_0(\varepsilon_0)\bigr\}^2\bigr]E\bigl[\Delta\bigl\{(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^{\mathrm{T}}\mathbf{W}\bigr\}^2\bigr].
\end{aligned} \qquad (12)
$$

Note that

$$
d^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \lesssim E\bigl[\Delta\bigl\{g\bigl(\varepsilon_0 e^{(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{Z}}\bigr) - g_0(\varepsilon_0)\bigr\}^2\bigr] + |\boldsymbol{\beta} - \boldsymbol{\beta}_0|^2 + |\boldsymbol{\gamma} - \boldsymbol{\gamma}_0|^2 \lesssim d^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \qquad (13)
$$

and

$$
\begin{aligned}
&E\bigl[\Delta\bigl\{g\bigl(\varepsilon_0 e^{(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{Z}}\bigr) - g_0(\varepsilon_0)\bigr\}^2\bigr] + |\boldsymbol{\beta} - \boldsymbol{\beta}_0|^2 + |\boldsymbol{\gamma} - \boldsymbol{\gamma}_0|^2\\
&\quad \lesssim E\bigl[\Delta\bigl\{g(\varepsilon_0) - g_0(\varepsilon_0)\bigr\}^2\bigr] + |\boldsymbol{\beta} - \boldsymbol{\beta}_0|^2 + |\boldsymbol{\gamma} - \boldsymbol{\gamma}_0|^2\\
&\quad \lesssim E\bigl[\Delta\bigl\{g\bigl(\varepsilon_0 e^{(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{Z}}\bigr) - g_0(\varepsilon_0)\bigr\}^2\bigr] + |\boldsymbol{\beta} - \boldsymbol{\beta}_0|^2 + |\boldsymbol{\gamma} - \boldsymbol{\gamma}_0|^2
\end{aligned} \qquad (14)
$$

under conditions C1–C4. Therefore, it follows from (10)–(14) that

$$\begin{aligned} &E\{\Delta(\{g(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}) - g_0(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}})\} + (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{Z} + (\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)^{\mathrm{T}}\mathbf{X})^2\} \\ &\quad \geq \{1-(1-\eta)^{1/2}\}(E[\Delta\{g(\varepsilon_0)-g_0(\varepsilon_0)\}^2] + E[\Delta\{(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)^{\mathrm{T}}\mathbf{W}\}^2]) \\ &\quad \gtrsim E[\Delta\{g(\varepsilon_0)-g_0(\varepsilon_0)\}^2] + |\boldsymbol{\beta}-\boldsymbol{\beta}_0|^2 + |\boldsymbol{\gamma}-\boldsymbol{\gamma}_0|^2 \\ &\quad \gtrsim d^2(\boldsymbol{\theta},\boldsymbol{\theta}_0). \end{aligned} \tag{15}$$

Hence, condition C1 in Theorem 1 of Shen and Wong [19] holds from (9) and (15).

Next, we verify their condition C2. Note that under our conditions C2 and C6,

$$\begin{aligned} &\{l(\boldsymbol{\beta},\boldsymbol{\gamma},\xi(\cdot,\boldsymbol{\beta});\mathcal{O}) - l(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\xi_0(\cdot,\boldsymbol{\beta}_0);\mathcal{O})\}^2 \\ &\quad \lesssim \{(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{Z}\}^2 + \{(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)^{\mathrm{T}}\mathbf{X}\}^2 \\ &\qquad + \Delta\{g(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}) - g(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}})\}^2 + \Delta\{g(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}) - g_0(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}})\}^2 \\ &\qquad + \int_0^Y [\exp\{g(se^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}) + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\} - \exp\{g_0(se^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}) + \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}\}]^2\,\mathrm{d}s. \end{aligned} \tag{16}$$

Using the Taylor expansion, we have for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}^p$,

$$E[\Delta\{g(Ye^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}) - g(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}})\}^2] \lesssim |\boldsymbol{\beta}-\boldsymbol{\beta}_0|^2, \tag{17}$$

and

$$\begin{aligned} &E\int_0^Y [\exp\{g(se^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}}) + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}\} - \exp\{g_0(se^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}) + \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}\}]^2\,\mathrm{d}s \\ &\quad \lesssim E\int_0^Y \exp\{g_0(se^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}) + \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}\}\{g(se^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}) - g_0(se^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}})\}^2\,\mathrm{d}s \\ &\qquad + |\boldsymbol{\beta}-\boldsymbol{\beta}_0|^2 + |\boldsymbol{\gamma}-\boldsymbol{\gamma}_0|^2 \\ &\quad = E[\Delta\{g(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}) - g_0(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}})\}^2] + |\boldsymbol{\beta}-\boldsymbol{\beta}_0|^2 + |\boldsymbol{\gamma}-\boldsymbol{\gamma}_0|^2 \lesssim d^2(\boldsymbol{\theta},\boldsymbol{\theta}_0). \end{aligned} \tag{18}$$

Thus, combining (16)–(18), we obtain that

$$E\{l(\boldsymbol{\beta},\boldsymbol{\gamma},\xi(\cdot,\boldsymbol{\beta});\mathcal{O}) - l(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\xi_0(\cdot,\boldsymbol{\beta}_0);\mathcal{O})\}^2 \lesssim d^2(\boldsymbol{\theta},\boldsymbol{\theta}_0),$$

which implies condition C2 in Theorem 1 of Shen and Wong [19].

As $\widehat{\boldsymbol{\theta}}_n$ maximizes the log-likelihood $P_n l(\boldsymbol{\theta};\mathcal{O})$ over the sieve space $\boldsymbol{\Theta}_n^p$, $\widehat{\boldsymbol{\theta}}_n$ satisfies inequality (1.1) in Shen and Wong [19] with $\eta_n = 0$. It follows from Lemma 2 that there exists a $\xi_{0n}(\cdot,\boldsymbol{\beta}_0) \in \mathcal{H}_n^p$ such that $\|\xi_{0n} - \xi_0\|_\infty = O(n^{-pv})$. The Kullback–Leibler distance between

$\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \xi_0(\cdot, \boldsymbol{\beta}_0))$ and $\boldsymbol{\theta}_{0n} = (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \xi_{0n}(\cdot, \boldsymbol{\beta}_0))$ is given by

$$\begin{aligned} K(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{0n}) &= P\big\{l(\boldsymbol{\theta}_0; \mathcal{O}) - l(\boldsymbol{\theta}_{0n}; \mathcal{O})\big\} \\ &\lesssim \big\|\xi_0(\cdot, \boldsymbol{\beta}_0) - \xi_{0n}(\cdot, \boldsymbol{\beta}_0)\big\|_2^2 \\ &\lesssim \big\|\xi_{0n}(\cdot, \boldsymbol{\beta}_0) - \xi_0(\cdot, \boldsymbol{\beta}_0)\big\|_\infty^2 \\ &= O\big(n^{-2pv}\big). \end{aligned}$$

Thus, it follows from Theorem 1 of Shen and Wong [19] that

$$d(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_p\big(n^{-\min(pv,(1-v)/2)}\big),$$

which completes the proof of Theorem 1. □

Proof of Theorem 2. Employing Theorem 2.1 of Ding and Nan [7], it suffices to verify the following conditions to prove Theorem 2.

A1. $d(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_p(n^{-\rho})$ for some $\rho > 0$.

A2. $Pl_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}) = \mathbf{0}$ and $Pl_{\xi}(\boldsymbol{\alpha}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O})[h] = \mathbf{0}$ for all $h \in \mathcal{H}$.

A3. There exists an $\mathbf{h}^* = (h_1^*, \ldots, h_d^*)^{\mathrm{T}}$, where $h_j^* \in \mathcal{H}$ for $j = 1, \ldots, d$, such that

$$Pl''_{\boldsymbol{\alpha}\xi}\big(\boldsymbol{\alpha}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}\big)[h] - Pl''_{\xi\xi}\big(\boldsymbol{\alpha}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}\big)\big[\mathbf{h}^*, h\big] = \mathbf{0},$$

for all $h \in \mathcal{H}$. Furthermore, the matrix $P\{l''_{\boldsymbol{\alpha}\boldsymbol{\alpha}}(\boldsymbol{\alpha}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}) - l''_{\xi\boldsymbol{\alpha}}(\boldsymbol{\alpha}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O})[\mathbf{h}^*]\}$ is non-singular.

A4. $P_n l'_{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\alpha}}_n, \widehat{\xi}_n(\cdot, \widehat{\boldsymbol{\beta}}_n); \mathcal{O}) = o_p(n^{-1/2})$ and $P_n l'_{\xi}(\widehat{\boldsymbol{\alpha}}_n, \widehat{\xi}_n(\cdot, \widehat{\boldsymbol{\beta}}_n); \mathcal{O})[\mathbf{h}^*] = o_p(n^{-1/2})$.

A5. Let $G_n = n^{1/2}(P_n - P)$. For any $c > 0$,

$$\sup_{d(\boldsymbol{\theta},\boldsymbol{\theta}_0) \le cn^{-\rho}, \boldsymbol{\theta} \in \boldsymbol{\Theta}_n^p} \big|G_n l'_{\boldsymbol{\alpha}}\big(\boldsymbol{\alpha}, \xi(\cdot, \boldsymbol{\beta}); \mathcal{O}\big) - G_n l'_{\boldsymbol{\alpha}}\big(\boldsymbol{\alpha}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}\big)\big| = o_p(1)$$

and

$$\begin{aligned} &\sup_{d(\boldsymbol{\theta},\boldsymbol{\theta}_0) \le cn^{-\rho}, \boldsymbol{\theta} \in \boldsymbol{\Theta}_n^p} \big|G_n l'_{\xi}\big(\boldsymbol{\alpha}, \xi(\cdot, \boldsymbol{\beta}); \mathcal{O}\big)\big[\mathbf{h}^*(\cdot, \boldsymbol{\beta})\big] - G_n l'_{\xi}\big(\boldsymbol{\alpha}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}\big)\big[\mathbf{h}^*(\cdot, \boldsymbol{\beta})\big]\big| \\ &\quad = o_p(1). \end{aligned}$$

A6. For some $\zeta > 1$ satisfying that $\rho\zeta > 1/2$ and for any $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$, $\{\boldsymbol{\theta} : d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \le cn^{-\rho}, \boldsymbol{\theta} \in \boldsymbol{\Theta}_n^p\}$ say,

$$\begin{aligned} &\big|Pl'_{\boldsymbol{\alpha}}\big(\boldsymbol{\alpha}, \xi(\cdot, \boldsymbol{\beta}); \mathcal{O}\big) - Pl'_{\boldsymbol{\alpha}}\big(\boldsymbol{\alpha}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}\big) - Pl''_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\big(\boldsymbol{\alpha}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}\big)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) \\ &\qquad - Pl''_{\boldsymbol{\alpha}\xi}\big(\boldsymbol{\alpha}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}\big)\big[\xi(\cdot, \boldsymbol{\beta}) - \xi_0(\cdot, \boldsymbol{\beta}_0)\big]\big| \\ &\quad = O\big(d^{\zeta}(\boldsymbol{\theta}, \boldsymbol{\theta}_0)\big) \end{aligned}$$

and

$$\begin{aligned}&\big|Pl'_{\xi}\big(\boldsymbol{\alpha},\xi(\cdot,\boldsymbol{\beta});\mathcal{O}\big)\big[\mathbf{h}^*(\cdot,\boldsymbol{\beta})\big]-Pl'_{\xi}\big(\boldsymbol{\alpha}_0,\xi_0(\cdot,\boldsymbol{\beta}_0);\mathcal{O}\big)\big[\mathbf{h}^*(\cdot,\boldsymbol{\beta}_0)\big]\\&\quad-Pl''_{\xi\boldsymbol{\alpha}}\big(\boldsymbol{\alpha}_0,\xi_0(\cdot,\boldsymbol{\beta}_0);\mathcal{O}\big)\big[\mathbf{h}^*(\cdot,\boldsymbol{\beta}_0)\big](\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)\\&\quad-Pl'_{\xi\xi}\big(\boldsymbol{\alpha}_0,\xi_0(\cdot,\boldsymbol{\beta}_0);\mathcal{O}\big)\big[\mathbf{h}^*(\cdot,\boldsymbol{\beta}),\xi(\cdot,\boldsymbol{\beta})-\xi_0(\cdot,\boldsymbol{\beta}_0)\big]\big|\\&=O\big(d^{\zeta}(\boldsymbol{\theta},\boldsymbol{\theta}_0)\big).\end{aligned}$$

We briefly discuss these conditions prior to verification of each of them. The convergence rate in A1, established in Theorem 1, is a prerequisite while proving the asymptotic normality. Condition A2 evaluates the score function (at the population level) at the true value, while A3 states the least favorable direction and the nonsingular information matrix along such a direction; both of them are standard in the maximum likelihood theory. Condition A4 assesses the score function (at the sample level) at the estimator and the stochastic equicontinuity in A5 can typically be verified by either the Donsker property or the maximal inequality (van der Vaart and Wellner [21]). Finally, the Taylor expansion results in A6.

First, A1 holds by choosing $\rho=\min(pv,(1-v)/2)$ from Theorem 1. Using the fact of zero-mean score functions, it is easy to show A2 holds.

Next, we find $\mathbf{h}^*=(h_1^*,\dots,h_d^*)^{\mathrm{T}}$ with $\mathbf{h}^*(t,z,\boldsymbol{\beta}_0)=\mathbf{w}^*(t)$ such that A3 holds. Denote $\mathbf{h}^*_{\boldsymbol{\beta}}=(h_1^*,\dots,h_{d_1}^*)^{\mathrm{T}}$, $\mathbf{h}^*_{\boldsymbol{\gamma}}=(h_{d_1+1}^*,\dots,h_d^*)^{\mathrm{T}}$, $\mathbf{w}^*_{\boldsymbol{\beta}}=(w_1^*,\dots,w_{d_1}^*)^{\mathrm{T}}$, and $\mathbf{w}^*_{\boldsymbol{\gamma}}=(w_{d_1+1}^*,\dots,w_d^*)^{\mathrm{T}}$. For any $h\in\mathcal{H}$,

$$\begin{aligned}&Pl''_{\boldsymbol{\alpha}\xi}\big(\boldsymbol{\alpha}_0,\xi_0(\cdot,\boldsymbol{\beta}_0);\mathcal{O}\big)[h]-Pl''_{\xi\xi}\big(\boldsymbol{\alpha}_0,\xi_0(\cdot,\boldsymbol{\beta}_0);\mathcal{O}\big)\big[\mathbf{h}^*,h\big]\\&\quad=P\begin{pmatrix}l''_{\boldsymbol{\beta}\xi}\big(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\xi_0(\cdot,\boldsymbol{\beta}_0);\mathcal{O}\big)[h]-l''_{\xi\xi}\big(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\xi_0(\cdot,\boldsymbol{\beta}_0);\mathcal{O}\big)\big[\mathbf{h}^*_{\boldsymbol{\beta}},h\big]\\l''_{\boldsymbol{\gamma}\xi}\big(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\xi_0(\cdot,\boldsymbol{\beta}_0);\mathcal{O}\big)[h]-l''_{\xi\xi}\big(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\xi_0(\cdot,\boldsymbol{\beta}_0);\mathcal{O}\big)\big[\mathbf{h}^*_{\boldsymbol{\gamma}},h\big]\end{pmatrix}.\end{aligned}$$

Some calculation entails that

$$\begin{aligned}&Pl''_{\boldsymbol{\beta}\xi}\big(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\xi_0(\cdot,\boldsymbol{\beta}_0);\mathcal{O}\big)[h]-Pl''_{\xi\xi}\big(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\xi_0(\cdot,\boldsymbol{\beta}_0);\mathcal{O}\big)\big[\mathbf{h}^*_{\boldsymbol{\beta}},h\big]\\&\quad=E\big\{\Delta Y\mathbf{Z}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}w'\big(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big)\big\}-E\big[Y\mathbf{Z}w\big(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big)\exp\big\{g_0\big(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big)+\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}+\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}\big\}\big]\\&\qquad+E\bigg[\int_0^{Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}}\exp\big\{g_0(s)+\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}\big\}\mathbf{w}^*_{\boldsymbol{\beta}}(s)w(s)\,\mathrm{d}s\bigg].\end{aligned}$$

In what follows, we calculate the above expectations using the ordinary properties of conditional expectation. We denote the conditional survival function of T given $\mathbf{Z}$ and $\mathbf{X}$ by $S_{T|\mathbf{Z},\mathbf{X}}(\cdot|\mathbf{Z},\mathbf{X})$ and the corresponding conditional density function by $f_{T|\mathbf{Z},\mathbf{X}}(\cdot|\mathbf{Z},\mathbf{X})$. After some tedious but straightforward calculations, we have

$$\begin{aligned}&E\big\{\Delta Y\mathbf{Z}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}w'\big(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big)|C,\mathbf{Z},\mathbf{X}\big\}\\&\quad=\mathbf{Z}Cw\big(Ce^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big)f_{T|\mathbf{Z},\mathbf{X}}(C|\mathbf{Z},\mathbf{X})-\mathbf{Z}\int_0^C w\big(te^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big)f_{T|\mathbf{Z},\mathbf{X}}(t|\mathbf{Z},\mathbf{X})\,\mathrm{d}t\end{aligned}$$

$$- \mathbf{Z} \int_0^C t w\big(te^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big)\big[g_0'\big(te^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big)e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} - \exp\big\{g_0\big(te^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big) + \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}\big\}\big] f_{T|\mathbf{Z},\mathbf{X}}(t|\mathbf{Z},\mathbf{X})\,\mathrm{d}t,$$

$$E\big[Y\mathbf{Z}w\big(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big)\exp\big\{g_0\big(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big) + \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}\big\}|C,\mathbf{Z},\mathbf{X}\big]$$
$$= \mathbf{Z}\int_0^C t w\big(te^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big)\exp\big\{g_0\big(te^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big) + \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}\big\} f_{T|\mathbf{Z},\mathbf{X}}(t|\mathbf{Z},\mathbf{X})\,\mathrm{d}t$$
$$+ \mathbf{Z}Cw\big(Ce^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big) f_{T|\mathbf{Z},\mathbf{X}}(C|\mathbf{Z},\mathbf{X}),$$

and

$$E\bigg[\int_0^{Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}} \exp\big\{g_0(s) + \boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}\big\}\mathbf{w}_{\boldsymbol{\beta}}^*(s)h(s)\,\mathrm{d}s\,\Big|\,C,\mathbf{Z},\mathbf{X}\bigg]$$
$$= \int_0^C \mathbf{w}_{\boldsymbol{\beta}}^*\big(se^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big)w\big(se^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big) f_{T|\mathbf{Z},\mathbf{X}}(s|\mathbf{Z},\mathbf{X})\,\mathrm{d}s.$$

Thus, we obtain that

$$Pl''_{\boldsymbol{\beta}\xi}\big(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\xi_0(\cdot,\boldsymbol{\beta}_0);\mathcal{O}\big)[h] - Pl''_{\xi\xi}\big(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\xi_0(\cdot,\boldsymbol{\beta}_0);\mathcal{O}\big)\big[\mathbf{h}_{\boldsymbol{\beta}}^*,h\big]$$
$$= E\bigg[\int_0^C \big\{\mathbf{w}_{\boldsymbol{\beta}}^*\big(se^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big) - \mathbf{Z}\big(1 + se^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}g_0'\big(se^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big)\big)\big\}w\big(se^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big) f_{T|\mathbf{Z},\mathbf{X}}(s|\mathbf{Z},\mathbf{X})\,\mathrm{d}s\bigg]$$
$$= E\bigg[\int_0^{Ce^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}} \big\{\mathbf{w}_{\boldsymbol{\beta}}^*(t) - \mathbf{Z}\big(1 + tg_0'(t)\big)\big\}w(t) f_{T|\mathbf{Z},\mathbf{X}}\big(te^{-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}|\mathbf{Z},\mathbf{X}\big)e^{-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\,\mathrm{d}t\bigg]$$
$$= E\bigg[\int_0^{+\infty} I\big(Ce^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \ge t\big)\big\{\mathbf{w}_{\boldsymbol{\beta}}^*(t) - \mathbf{Z}\big(1 + tg_0'(t)\big)\big\}w(t) f_{T|\mathbf{Z},\mathbf{X}}\big(te^{-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}|\mathbf{Z},X\big)e^{-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\,\mathrm{d}t\bigg]$$
$$= \int_0^{+\infty} \big[E\big\{I\big(Ce^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \ge t\big)\exp\big\{-\Lambda_0(t)e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\big\}e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\big\}\mathbf{w}_{\boldsymbol{\beta}}^*(t)$$
$$- \big(1 + tg_0'(t)\big)E\big\{\mathbf{Z}I\big(Ce^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \ge t\big)\exp\big(-\Lambda_0(t)e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\big)e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\big\}\big]w(t)\lambda_0(t)\,\mathrm{d}t.$$

Therefore, we take $\mathbf{h}_{\boldsymbol{\beta}}^*$ with

$$\mathbf{h}_{\boldsymbol{\beta}}^*(t,z,\boldsymbol{\beta}_0) = \mathbf{w}_{\boldsymbol{\beta}}^*(t) = \frac{\{1 + tg_0'(t)\}E[\mathbf{Z}I(Ce^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \ge t)\exp\{-\Lambda_0(t)e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\}e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}]}{E[I(Ce^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \ge t)\exp\{-\Lambda_0(t)e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\}e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}]},$$

which makes $Pl''_{\boldsymbol{\beta}g}(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,g_0;\mathcal{O})[h] - Pl''_{gg}(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,g_0;\mathcal{O})[\mathbf{h}_{\boldsymbol{\beta}}^*,h] = \mathbf{0}$ for any $h \in \mathcal{H}$.

Based on similar but simpler calculations, we also have that

$$Pl''_{\boldsymbol{\gamma}g}(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,g_0;\mathcal{O})[h] - Pl''_{gg}(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,g_0;\mathcal{O})\big[\mathbf{h}_{\boldsymbol{\gamma}}^*,h\big]$$
$$= \int_0^{+\infty} \big[E\big\{I\big(Ce^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \ge t\big)\exp\big\{-\Lambda_0(t)e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\big\}e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\big\}\mathbf{w}_{\boldsymbol{\gamma}}^*(t)$$

$$- E\{\mathbf{X} I(Ce^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \geq t) \exp\{-\Lambda_0(t) e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\} e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\}] w(t)\lambda_0(t)\,\mathrm{d}t.$$

Thus, we take $\mathbf{h}^*_{\boldsymbol{\gamma}}$ with

$$\mathbf{h}^*_{\boldsymbol{\gamma}}(t, z, \boldsymbol{\beta}_0) = \mathbf{w}^*_{\boldsymbol{\gamma}}(t) = \frac{E[\mathbf{X} I(Ce^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \geq t) \exp\{-\Lambda_0(t) e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\} e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}]}{E[I(Ce^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \geq t) \exp\{-\Lambda_0(t) e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\} e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}]}$$

such that $Pl''_{\boldsymbol{\gamma} g}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, g_0; \mathcal{O})[h] - Pl''_{gg}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, g_0; \mathcal{O})[\mathbf{h}^*_{\boldsymbol{\gamma}}, h] = \mathbf{0}$ for any $h \in \mathcal{H}$.

Note that

$$\begin{aligned}
&P\big(Ye^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \geq t | C, \mathbf{Z}, \mathbf{X}\big) \\
&\quad = P\big(Te^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \geq t, T \leq C | C, \mathbf{Z}, \mathbf{X}\big) + P\big(Ce^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \geq t, T > C | C, \mathbf{Z}, \mathbf{X}\big) \\
&\quad = P\big(te^{-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}} \leq T \leq C | C, \mathbf{Z}, \mathbf{X}\big) I\big(C \geq te^{-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big) + P(T > C | C, \mathbf{Z}, \mathbf{X}) I\big(C \geq te^{-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big) \\
&\quad = S_{T|\mathbf{Z},\mathbf{X}}\big(te^{-\boldsymbol{\beta}_0^{\mathrm{T}} Z} | \mathbf{Z}, \mathbf{X}\big) I\big(C \geq te^{-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big) \\
&\quad = I\big(C \geq te^{-\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}}\big) \exp\big\{-\Lambda_0(t) e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\big\}.
\end{aligned}$$

Then, $\mathbf{w}^*_{\boldsymbol{\beta}}$ and $\mathbf{w}^*_{\boldsymbol{\gamma}}$ can be simplified as

$$\mathbf{w}^*_{\boldsymbol{\beta}}(t) = \frac{\{1 + tg'_0(t)\} E\{\mathbf{Z} I(\varepsilon_0 \geq t) e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\}}{E\{I(\varepsilon_0 \geq t) e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\}}$$

and

$$\mathbf{w}^*_{\boldsymbol{\gamma}}(t) = \frac{E\{\mathbf{X} I(\varepsilon_0 \geq t) e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\}}{E\{I(\varepsilon_0 \geq t) e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{X}}\}}.$$

Hence, we have found $\mathbf{h}^* = (h^*_1, \ldots, h^*_d)^{\mathrm{T}}$ such that for any $h \in \mathcal{H}$,

$$Pl''_{\boldsymbol{\alpha}\xi}\big(\boldsymbol{\alpha}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}\big)[h] - Pl''_{\xi\xi}\big(\boldsymbol{\alpha}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}\big)\big[\mathbf{h}^*, h\big] = \mathbf{0}.$$

Furthermore, we obtain

$$l'_{\boldsymbol{\beta}}\big(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}\big) - l'_{\xi}\big(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}\big)\big[\mathbf{h}^*_{\boldsymbol{\beta}}\big] = l^*_{\boldsymbol{\beta}_0}(\mathcal{O})$$

and

$$l'_{\boldsymbol{\gamma}}\big(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}\big) - l'_{\xi}\big(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}\big)\big[\mathbf{h}^*_{\boldsymbol{\gamma}}\big] = l^*_{\boldsymbol{\gamma}_0}(\mathcal{O}),$$

which are the efficient score functions for $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$, respectively. We can also show that

$$P\big\{l''_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\big(\boldsymbol{\alpha}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}\big) - l''_{\xi\boldsymbol{\alpha}}\big(\boldsymbol{\alpha}_0, \xi_0(\cdot, \boldsymbol{\beta}_0); \mathcal{O}\big)\big[\mathbf{h}^*\big]\big\} = -Pl^*_{\boldsymbol{\alpha}_0}(\mathcal{O})^{\otimes 2},$$

which is the negative information matrix for $\boldsymbol{\alpha}_0$. Thus, it is invertible under condition C8. Hence, A3 holds.

Using Lemmas 4–6, the Taylor expansion, the maximal inequality in Lemma 3.4.2 of van der Vaart and Wellner [21] or Theorem 11.3 of Kosorok [12], and the Markov inequality, we can show that assumptions A4–A6 hold for $\rho = \min\{pv, (1-v)/2\}$ and

$$\zeta = \min\bigl\{2(p-1)v, 1/2 + (p-5/2)v, 1-v\bigr\}/\min\bigl\{pv, (1-v)/2\bigr\} > 1.$$

Therefore, by Theorem 6.1 of Wellner and Zhang [23], we have

$$n^{1/2}(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0) = \bigl\{\mathcal{I}(\boldsymbol{\alpha}_0)\bigr\}^{-1} n^{1/2} P_n l^*_{\boldsymbol{\alpha}_0}(\boldsymbol{\theta}_0; \mathcal{O}) + o_p(1) \to_d N\bigl(0, \bigl\{\mathcal{I}(\boldsymbol{\alpha}_0)\bigr\}^{-1}\bigr),$$

where $l^*_{\boldsymbol{\alpha}_0}(\boldsymbol{\theta}_0; \mathcal{O}) = l'_{\boldsymbol{\alpha}}(\boldsymbol{\theta}_0; \mathcal{O}) - l'_{\xi}(\boldsymbol{\theta}_0; \mathcal{O})[\mathbf{h}^*]$ is the efficient score function for $\boldsymbol{\alpha}_0$. This completes the proof of Theorem 2. □

Proof of Theorem 3. Define

$$\mathbf{w}^*_n(t) = \bigl(\bigl\{1 + t\widehat{g}'_n(t)\bigr\}\bar{\mathbf{Z}}^{\mathrm{T}}(t; \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n), \bar{\mathbf{X}}^{\mathrm{T}}(t; \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n)\bigr)^{\mathrm{T}}.$$

Then we have

$$\widehat{l}^*_{\boldsymbol{\alpha}_0}(\widehat{\boldsymbol{\theta}}_n; \mathcal{O}) = l'_{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}_n; \mathcal{O}) - l'_{\xi}(\widehat{\boldsymbol{\theta}}_n; \mathcal{O})\bigl[\mathbf{h}^*_n\bigr].$$

Let

$$\begin{aligned}\mathcal{I}^{jk}(\boldsymbol{\theta}_0) &= P\bigl[\bigl\{l'_{\alpha_j}(\boldsymbol{\theta}_0; \mathcal{O}) - l'_{\xi}(\boldsymbol{\theta}_0; \mathcal{O})\bigl[h^*_j\bigr]\bigr\}\bigl\{l'_{\alpha_k}(\boldsymbol{\theta}_0; \mathcal{O}) - l'_{\xi}(\boldsymbol{\theta}_0; \mathcal{O})\bigl[h^*_k\bigr]\bigr\}\bigr] \\ &\equiv P A^{jk}(\boldsymbol{\theta}_0; \mathcal{O})\end{aligned}$$

and

$$\begin{aligned}\widehat{\mathcal{I}}^{jk}_n(\boldsymbol{\theta}_0) &= P_n\bigl[\bigl\{l'_{\alpha_j}(\widehat{\boldsymbol{\theta}}_n; \mathcal{O}) - l'_{\xi}(\widehat{\boldsymbol{\theta}}_n; \mathcal{O})\bigl[h^*_{jn}\bigr]\bigr\}\bigl\{l'_{\alpha_k}(\widehat{\boldsymbol{\theta}}_n; \mathcal{O}) - l'_{\xi}(\widehat{\boldsymbol{\theta}}_n; \mathcal{O})\bigl[h^*_{kn}\bigr]\bigr\}\bigr] \\ &\equiv P_n A^{jk}_n(\widehat{\boldsymbol{\theta}}_n; \mathcal{O})\end{aligned}$$

for $j, k = 1, \ldots, d$. It suffices to show that $P_n A^{jk}_n(\widehat{\boldsymbol{\theta}}_n; \mathcal{O})$ converges to $P A^{jk}(\boldsymbol{\theta}_0; \mathcal{O})$ in probability. Note that

$$\begin{aligned}&P_n A^{jk}_n(\widehat{\boldsymbol{\theta}}_n; \mathcal{O}) - P A^{jk}(\boldsymbol{\theta}_0; \mathcal{O}) \\ &\quad= P_n\bigl\{A^{jk}_n(\widehat{\boldsymbol{\theta}}_n; \mathcal{O}) - A^{jk}(\boldsymbol{\theta}_0; \mathcal{O})\bigr\} + (P_n - P)A^{jk}(\boldsymbol{\theta}_0; \mathcal{O}).\end{aligned}$$

Clearly, $(P_n - P)A^{jk}(\boldsymbol{\theta}_0; \mathcal{O}) = o_p(1)$. On the other hand, under conditions C2 and C6, we have

$$\begin{aligned}&P\bigl\{A^{jk}_n(\widehat{\boldsymbol{\theta}}_n; \mathcal{O}) - A^{jk}(\boldsymbol{\theta}_0; \mathcal{O})\bigr\}^2 \\ &\lesssim P\bigl\{l'_{\alpha_j}(\widehat{\boldsymbol{\theta}}_n; \mathcal{O}) - l'_{\alpha_j}(\boldsymbol{\theta}_0; \mathcal{O})\bigr\}^2 + P\bigl\{l'_{\alpha_k}(\widehat{\boldsymbol{\theta}}_n; \mathcal{O}) - l'_{\alpha_k}(\boldsymbol{\theta}_0; \mathcal{O})\bigr\}^2 \\ &\quad + P\bigl\{l'_{\xi}(\widehat{\boldsymbol{\theta}}_n; \mathcal{O})\bigl[h^*_j - h^*_{jn}\bigr]\bigr\}^2 + P\bigl\{l'_{\xi}(\widehat{\boldsymbol{\theta}}_n; \mathcal{O})\bigl[h^*_k - h^*_{kn}\bigr]\bigr\}^2 \\ &\quad + P\bigl\{l'_{\xi}(\widehat{\boldsymbol{\theta}}_n; \mathcal{O})\bigl[h^*_j\bigr] - l'_{\xi}(\boldsymbol{\theta}_0; \mathcal{O})\bigl[h^*_j\bigr]\bigr\}^2 + P\bigl\{l'_{\xi}(\widehat{\boldsymbol{\theta}}_n; \mathcal{O})\bigl[h^*_k\bigr] - l'_{\xi}(\boldsymbol{\theta}_0; \mathcal{O})\bigl[h^*_k\bigr]\bigr\}^2.\end{aligned}$$

It is easy to show that

$$P\{l'_{\alpha_j}(\widehat{\boldsymbol{\theta}}_n;\mathcal{O}) - l'_{\alpha_j}(\boldsymbol{\theta}_0;\mathcal{O})\}^2 \lesssim \|\widehat{g}'_n - g'_0\|_2^2 + d^2(\widehat{\boldsymbol{\theta}}_n,\boldsymbol{\theta}_0), \qquad j=1,\ldots,d_1,$$

$$P\{l'_{\alpha_j}(\widehat{\boldsymbol{\theta}}_n;\mathcal{O}) - l'_{\alpha_j}(\boldsymbol{\theta}_0;\mathcal{O})\}^2 \lesssim d^2(\widehat{\boldsymbol{\theta}}_n,\boldsymbol{\theta}_0), \qquad j=d_1+1,\ldots,d,$$

$$P\{l'_{\xi}(\widehat{\boldsymbol{\theta}}_n;\mathcal{O})[h_j^* - h_{jn}^*]\}^2 \lesssim d^2(\widehat{\boldsymbol{\theta}}_n,\boldsymbol{\theta}_0) + |\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|, \qquad j=1,\ldots,d,$$

$$P\{l'_{\xi}(\widehat{\boldsymbol{\theta}}_n;\mathcal{O})[h_j^*] - l'_{\xi}(\boldsymbol{\theta}_0;\mathcal{O})[h_j^*]\}^2 \lesssim d^2(\widehat{\boldsymbol{\theta}}_n,\boldsymbol{\theta}_0), \qquad j=1,\ldots,d.$$

Thus, it follows from Theorem 1 that $P_n\{A_n^{jk}(\widehat{\boldsymbol{\theta}}_n;\mathcal{O}) - A^{jk}(\boldsymbol{\theta}_0;\mathcal{O})\} = o_p(1)$, which completes the proof of the theorem. □

Acknowledgments

The authors would like to thank the Editor, Professor Eric Moulines, the Associate Editor and the two referees for their insightful comments and constructive suggestions that greatly improved the paper. Zhao's research was partly supported by the Research Grant Council of Hong Kong (504011, 503513), The Hong Kong Polytechnic University, and the National Natural Science Foundation of China (11371299); Wu's research was partly supported by the National Natural Science Foundation of China (11201350); and Yin's research was partly supported by the Research Grant Council of Hong Kong (17125814).

References

[1] Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66** 429–436.
[2] Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In *Handbook of Econometrics*, *Part B* (J.J. Heckman and E.E. Leamer, eds.) **6** 5549–5632. North-Holland: Elsevier.
[3] Chen, Y.Q. and Jewell, N.P. (2001). On a general class of semiparametric hazards regression models. *Biometrika* **88** 687–702. MR1859402
[4] Chen, Y.Q. and Wang, M.-C. (2000). Analysis of accelerated hazards models. *J. Amer. Statist. Assoc.* **95** 608–618. MR1803176
[5] Copelan, E.A., Biggs, J.C., Thompson, J.M., Crilley, P., Szer, J., Klein, J.P., Kapoor, N., Avalos, B.R., Cunningham, I., Atkinson, K., Downs, K., Harmon, G.S., Daly, M.B., Brodsky, I., Bulova, S.I. and Tutschka, P.J. (1991). Treatment for acute myelocytic leukemia with allogeneic bone marrow transplantation following preparation with Bu/Cy. *Blood* **78** 838–843.
[6] Cox, D.R. (1972). Regression models and life-tables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **34** 187–220. With discussion by F. Downton, Richard Peto, D. J. Bartholomew, D. V. Lindley, P. W. Glassborow, D. E. Barton, Susannah Howard, B. Benjamin, John J. Gart, L. D. Meshalkin, A. R. Kagan, M. Zelen, R. E. Barlow, Jack Kalbfleisch, R. L. Prentice and Norman Breslow, and a reply by D. R. Cox. MR0341758
[7] Ding, Y. and Nan, B. (2011). A sieve M-theorem for bundled parameters in semiparametric models, with application to the efficient estimation in a linear model for censored data. *Ann. Statist.* **39** 3032–3061. MR3012400

[8] Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *Ann. Statist.* **27** 1536–1563. MR1742499
[9] Huang, J. and Rossini, A.J. (1997). Sieve estimation for the proportional-odds failure-time regression model with interval censoring. *J. Amer. Statist. Assoc.* **92** 960–967. MR1482126
[10] Jin, Z., Lin, D.Y., Wei, L.J. and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90** 341–353. MR1986651
[11] Klein, J.P. and Moeschberger, M.L. (2003). *Survival Analysis*: *Techniques for Censored and Truncated Data*, 2nd ed. New York: Springer.
[12] Kosorok, M.R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. *Springer Series in Statistics*. New York: Springer. MR2724368
[13] Lai, T.L. and Ying, Z. (1991). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19** 531–556. MR1105835
[14] Prentice, R.L. (1978). Linear rank tests with right censored data. *Biometrika* **65** 167–179. MR0497517
[15] Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18** 303–328. MR1041395
[16] Rossini, A.J. and Tsiatis, A.A. (1996). A semiparametric proportional odds regression model for the analysis of current status data. *J. Amer. Statist. Assoc.* **91** 713–721. MR1395738
[17] Schumaker, L.L. (1981). *Spline Functions*: *Basic Theory*. *Pure and Applied Mathematics*. New York: Wiley. A Wiley-Interscience Publication. MR0606200
[18] Shen, X. (1998). Proportional odds regression and sieve maximum likelihood estimation. *Biometrika* **85** 165–177. MR1627289
[19] Shen, X. and Wong, W.H. (1994). Convergence rate of sieve estimates. *Ann. Statist.* **22** 580–615. MR1292531
[20] Tsiatis, A.A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18** 354–372. MR1041397
[21] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. *Springer Series in Statistics*. New York: Springer. MR1385671
[22] Wei, L.J., Ying, Z. and Lin, D.Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77** 845–851. MR1086694
[23] Wellner, J.A. and Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Ann. Statist.* **35** 2106–2142. MR2363965
[24] Zeng, D. and Lin, D.Y. (2007). Efficient estimation for the accelerated failure time model. *J. Amer. Statist. Assoc.* **102** 1387–1396. MR2412556
[25] Zeng, D., Yin, G. and Ibrahim, J.G. (2005). Inference for a class of transformed hazards models. *J. Amer. Statist. Assoc.* **100** 1000–1008. MR2201026
[26] Zhang, H.H., Cheng, G. and Liu, Y. (2011). Linear or nonlinear? Automatic structure discovery for partially linear models. *J. Amer. Statist. Assoc.* **106** 1099–1112. MR2894767

Received April 2014 and revised March 2016