Asymptotic Properties of Maximum Partial Likelihood Estimators When The Relative Risk Possesses First Order Derivative

Fei Tan[1] and Hanxiang Peng[1,*,†]

[1]Department of Mathematical Sciences
Indiana University – Purdue University At Indianapolis
Indianapolis, IN 46202-3216, USA
*Corresponding author: Hanxiang Peng
Email: hpeng@math.iupui.edu

†This research is supported by the US National Science Foundation under Grant No. DMS-0940365.

September 18, 2013

Abstract

Knots in a free-knot spline model are treated as parameters and can be threshold values such as changepoints. Motivated by quadratic splines which only possess continuous first order derivatives, this article investigates the asymptotic properties of a general semiparametric multiplicative hazard model when the relative risk is expressed as a first order continuously differentiable parametric function. It is shown that the logarithm of the partial likelihood function of the model is locally concave under suitable conditions. Using the convexity lemma [Andersen and Gill [1], Pollard [2]], it is proved that the maximum partial likelihood estimators of parameters uniquely exist in a neighborhood of the true values of parameter and are consistent and asymptotically normal. The developed theory is applied to derive the asymptotic normality of the maximum partial likelihood estimators of parameters in a free-knot quadratic spline model.

Key words and phrases: Asymptotic normality, convexity lemma, Cox model, free-knot spline, relative risk

1 Introduction and Assumptions

In a general semiparametric multiplicative hazard model, the hazard rate or intensity of failure $h(t)$ for the survival time of an individual with a covariate process $Z(t) \in \mathbb{R}^q$ of time t satisfies

$$h(t) = h_0(t) r_\theta\big(t, Z(t)\big), \quad t \geq 0, \tag{1}$$

where $r_\theta(t, z)$ is a parametric function of $\theta \in \subset \mathbb{R}^p$, and h_0 is a completely unspecified nonparametric baseline hazard function on $[0, \infty)$.

In this article, we are interested in the asymptotic behaviors of the maximum partial likelihood estimator (MPLE) of θ when $r_\theta(t, z)$ possesses continuous first order derivative with respect to θ for every t, z. As an application of the developed theory, we derive the consistency and asymptotic normality of MPLE's of parameters when the logarithm of the relative risk is expressed as a free-knot spline polynomial (truncated polynomial or B-spline) of order three or above. Knots in a free-knot spline polynomial are considered as parameters, and a quadratic spline polynomial possesses continuous first order derivative and a cubic spline polynomial has discontinuous second order derivative at coalescing knots. Quadratic and cubic spline polynomial functions are low-order polynomial splines; they allow the number of parameters to remain low while ensuring a sufficient degree of smoothness, the continuity of the functions and their gradients; they provide more flexibility to explain data than quadratic or cubic polynomial functions.

Intensely studied over the past several decades, hazard regression (or relative risk regression) has been fruitful in the analysis of survival data in parametric, semiparametric and nonparametric regression. Andersen and Gill (1982) [1] formulated the Cox linear regression model in a setup of multivariate counting processes and proved the consistency and asymptotic normality of MPLE's. Relaxing the exponential regression form in the popular Cox model to an arbitrary non-negative twice differentiable form, Prentice and Self (1983) [3] gave the asymptotic distribution theory for Cox-type linear regression models under conditions generalizing those in Andersen and Gill (1982) [1]. Borgan (1984) [4] studied MPLE's for a general parametric multiplicative intensity model and established the asymptotic results under a set of regularity conditions including the usual existence and boundedness of third order continuous differentiability of the relative risk. Here we assume first order continuous differentiability of the relative risk and prove the consistency and asymptotic normality of MPLE's. The proof of the latter is based on the convexity lemma [Andersen and Gill (1983) [1]; Pollard (1991) [2]; Fleming and Harrington (1991) [5]; Lemma 1 below]. Our assumptions are weaker than those in Borgan (1984) [4] and Prentice and Self (1983) [3] and, in fact, are minimal.

The key ingredient in our proofs is the local concavity of the log- partial likelihood function. We show in this article that the log- partial likelihood function is *locally concave* in probability (short for locally concave or simply concave, see Subsection 2.1) for an arbitrary non-negative first order continuously differentiable relative risk function $r_\theta(t, z)$. This is of course due to the unique structure of the log- partial likelihood functions. For the usual Cox model, where $\log r_\theta(t, z)$ is a linear function of parameter θ, it is well known in the literature that the log- partial likelihood function is *globally concave* (see page 1106, Andersen and Gill (1983) [1]). Interestingly, the global concavity also holds when $\log r_\theta(t, z)$ is a quadratic function of θ, see (4) and the discussions therein. The global concavity of the log- partial likelihood function is reduced to the local concavity when the usual exponential regression form $r_\theta(z) = \exp(\theta^\top z)$ in the Cox model is relaxed to a non-negative second order continuously differentiable regression form $r_\theta(z) = r(\theta^\top z)$ (see page 809, Prentice and Self (1983) [3]). For nonparametric relative risk regression in a generalized Cox model, using Frèchet derivatives, O'Sullivan showed that the logarithm of the partial likelihood and its limit are concave in some Sobolev space (see pages 130-131, O'Sullivan (1993) [6] or Remark 2 below). Thus the log- partial likelihood function in a semiparametric multiplicative hazard model is always concave for an arbitrary relative risk either nonparametric or parametric (continuously differentiable). Indeed, Hjort and Pollard [7] also established the concavity in their Lemma A2 in an alternative form.

Goetghebeur and Pocock (1995) [8] proposed a family of double quadratic models in which the relationship between a risk factor and disease outcome was expressed as two independent quadratic curves joined at a low point to be estimated. As Tan (2007) [9] pointed out, this is indeed a free-knot quadratic spline model with knots in covariates. For a log- relative risk expressed as a quadratic spline function, under the neighborhood restriction that there are no observations in a neighborhood of the true value of parameter, the second order derivative of the partial likelihood function is continuous. Under this restriction, Goetghebeur and Pocock (1995) [8] derived the asymptotic results of the MPLE's. When observations are allowed in the neighborhood of the true knot value, the log- partial likelihood function possesses continuous first order derivative, but does not have second order derivative. Our developed theory fills this gap and gives the consistency and asymptotic normality of the MPLE's.

Free-knot polynomial splines, taking the advantage of the freedom of knots, can approximate wider classes of smooth functions to higher order than fixed-knot polynomial splines. Free-knot spline functions can be used to model either the regression coefficient process or the covariate process, both of time. Knots can represent *threshold values* such as changepoints (e.g. the nadir of BMI). Spline partial likelihood functions can be computed using S-Plus package [10]. Recently, free-knot splines in regression have gained momentum in statistical inference, see e.g. [Mao and Zhao (2003) [11]; Molinari, *et al.* (2001) [12]; Giorgi, *et al.* (2003) [13]; Molinari, *et al.*

(2004) [14]].

In order to facilitate our theoretical analysis, we set $g_\theta(t, Z(t)) = \log r_\theta(t, Z(t))$ to relate the general multiplicative intensity model to the usual exponential form in the popular Cox model. We shall work with g_θ upon which the assumptions shall be imposed. This does not cause any loss of generality. Throughout we shall assume that *for every* (θ, t, z)*,* $g_\theta(t, z)$ *has continuous first order derivative* $\dot{g}_\theta(t, z)$ *w.r.t.* θ *which is locally bounded and predictable w.r.t. the filtrations, unless otherwise explicitly stated.* The most commonly used functional form of g_θ is perhaps the linear model $g_\theta(t, Z(t)) = \theta(t)^\top Z(t)$, where $\theta(t)$ is a coefficient process of time. This of course includes the popularly used Cox (1972) [15] proportional hazards model $g_\theta(t, Z(t)) = \theta^\top Z(t)$, where the coefficient θ is not time dependent. Our formulation may accommodate many other different parameterizations, see more discussions in Section 2.2.

We start now with a brief summary of the counting process framework for the case of independent and identically distributed observations. We have n independent individuals which are continuously monitored over time $t \geq 0$. For each subject, there is a process $(N_i(t), Y_i(t), Z_i(t))$ for $t \geq 0$. Here $N_i(t)$ is a counting process recording events (such as deaths, hospital visits and so on) to occur up to time t, $Y_i(t)$ takes binary values 1 and 0 depending on whether or not the subject is under observation immediately prior to time t, and $Z_i(t)$ is a q-dimensional covariate process. As in Gill (1984) and Fleming and Harrington (1991), we assume throughout the following conditions (i)-(iii) hold. (i) $(N_1, ..., N_n)$ is a multivariate counting process. (ii) For each i, $M_i = N_i - A_i$ is a local martingale w.r.t. a right-continuous filtration $\{\mathcal{F}_t : t \geq 0\}$ which represents the statistical information accruing over time, where A_i is the continuous compensator $A_i = \int Y_i(s) \exp(g_{\theta_0}(s, Z_i(s)))h_0(s)ds$ with g_θ being a predictable process and θ_0 denoting the true but unknown value of parameter. (iii)

Each of the at-risk process Y_i and covariate process Z_i are predictable w.r.t. the filtration $\{\mathcal{F}_t : t \geq 0\}$.

Let $\mathbb{S}_n$ be the operator defined by

$$\mathbb{S}_n(t, g)[h] = \frac{1}{n} \sum_{i=1}^{n} h\big(t, Z_i(t)\big) Y_i(t) \exp\big(g(t, Z_i(t))\big),$$

where g is a scalar function and h is a scalar, vector or matrix function. Set $S_n^{(0)}(\theta, t) = \mathbb{S}_n(t, g_\theta)[\imath]$, where $\imath$ is the identity map; $S_n^{(1)}(\theta, t) = \mathbb{S}_n(t, g_\theta)[\dot{g}_\theta]$; $S_n^{(2)}(\theta, t) = \mathbb{S}_n(t, g_\theta)[\dot{g}_\theta^{\otimes 2}]$, where $B^{\otimes 2} = BB^\top$ for a matrix B. In what follows we denote A, M, N, etc. i.i.d. copies of A_i, M_i, N_i, etc. respectively. Following Andersen and Gill (1982), we introduce the following regularity assumption.

Assumption

(I) There exists finite time τ such that $\int_0^\tau h_0(t)dt < \infty$.

(II) For a compact neighborhood Θ_0 of θ_0, there exists a scalar $s^{(0)}$, a vector $s^{(1)}$ and a matrix $s^{(2)}$ on $\Theta_0 \times [0, \tau]$ such that for $j = 0, 1, 2$,

$$\sup_{t \in [0,\tau],\, \theta \in \Theta_0} \|S_n^{(j)}(\theta, t) - s^{(j)}(\theta, t)\| \xrightarrow{P} 0, \tag{2}$$

where $\|B\| \equiv \max\{|B_{ij}| : \forall i, j\}$ is a matrix norm.

(III) Define $e \equiv s^{(1)}/s^{(0)}$ and $v \equiv s^{(2)}/s^{(0)} - e^{\otimes 2}$. Then for $\theta \in \Theta_0$ and $t \in [0, \tau]$, $\frac{\partial}{\partial \theta} s^{(0)}(\theta, t) = s^{(1)}(\theta, t)$.

(IV) For $j = 0, 1, 2$, the functions $s^{(j)}(\theta, t)$ are bounded; the function families $s^{(j)}(\cdot, t)$, $t \in [0, \tau]$ are equicontinuous at $\theta = \theta_0$; and $s^{(0)}(\theta, t)$ is bounded away from zero on $\Theta_0 \times [0, \tau]$.

(V) The matrix $\Sigma(\theta_0, \tau) = \int_0^\tau v(\theta_0, t) s^{(0)}(\theta_0, t) h_0(t)\, dt$ is positive definite.

(VI) There exists $\delta > 0$ such that as n tends to infinity,

$$n^{-1/2} \sup_{1\le i\le n,\, 0\le t\le \tau} \|\dot{g}_{\theta_0}(t, Z_i(t))\| Y_i(t) \mathbf{1}_{\left\{g_{\theta_0}(t,Z_i(t))>-\delta\|\dot{g}_{\theta_0}(t,Z_i(t))\|\right\}} \xrightarrow{P} 0.$$

Note that (VI) is a Lindeberg negligibility condition and trivially holds if covariates are bounded while (I)-(V) are regularity assumptions similar to those found in standard asymptotic likelihood theory.

Let T and U be the failure and censoring time of a person and Z be a covariate associated with the person. Suppose that the data can be summarized as n realizations of i.i.d. random vectors (X_i, δ_i, Z_i) for $i = 1, ..., n$, where $X_i \equiv \min(T_i, U_i)$, representing the observed time of person i; $\delta_i \equiv 1_{\{T_i \le U_i\}}$, indicating that the observed time is a death time not a censoring time. Let the counting process be $N_i(t) \equiv 1_{\{X_i \le t, \delta_i = 1\}}$, and the at-risk process be $Y_i(t) \equiv 1_{\{X_i \ge t\}}$. The following result can be considered as a sufficient condition for Assumption and the proof can be obtained analogous to the proof of Theorem 8.4.1 [Fleming and Harrington (1991) [5]] or Proposition 1 [Tan (2007) [9]].

Proposition 1. *Suppose that for $i = 1, ..., n$, the covariate Z_i is constant in time and takes value in a compact set $\mathbb{Z}$ of $\mathbb{R}^q$; the failure time T_i and the censoring U_i are conditionally independent given the covariate Z_i; and $\mathbb{P}\{Y_i(\tau) > 0\} > 0$ for some $\tau > 0$. Suppose that $g_\theta(Z)$ has continuous first order derivative for $Z \in \mathbb{Z}$ and θ in a compact neighborhood Θ of θ_0. Then* Assumption *holds with the exception of (V).*

The rest of the article is organized as follows. The main results are given in Section 2; Section 2.1 presents concavity, consistency and asymptotic normality; Section 2.2 discusses the applications of the results to free-knot spline polynomial models. In Section 3, we show concavity. In Section 4, we prove asymptotic normality of MPLE's.

2 The Main Results

In the first part of this section, we give the local concavity of the log- partial likelihood function, followed by consistency and asymptotic normality of MPLEs. In the second part, we apply the developed theory to derive the asymptotics of MPLE's of parameters in free-knot spline polynomial models.

2.1 Concavity, Consistency and Asymptotic Normality

As usual, the partial likelihood function can be expressed as

$$PL_n(\theta) = \prod_{i=1}^{n} \left\{ \frac{\exp\left[g_\theta(T_i, Z_i(T_i))\right]}{\Sigma_{j\in R_i} \exp\left[g_\theta(T_i, Z_j(T_i))\right]} \right\}^{\delta_i},$$

where δ_i is the indicator that the failure of individual i is observed and R_i is the set of those at risk at the time of the ith failure. The logarithm of the partial likelihood function can be written in an integral of a counting process,

$$l_n(\theta) = \log PL_n(\theta) = \sum_{i=1}^{n} \int_0^\tau \Big[g_\theta(t, Z_i(t)) - \log S_n^{(0)}(\theta, t) \Big] dN_i(t).$$

Our first result is that $l_n(\theta)$ is *concave in probability* in the sense that the probability of the event that $l_n(\theta)$ is concave converges to one as the sample size n tends to infinity. This is stated below with the proof delayed to Section 3.

Theorem 1. (Concavity) *Suppose that* Assumption *(I)–(V) hold. Assume that there exist a neighborhood* Θ_0 *of* θ_0 *and a bounded matrix function* $s^{(3)}$ *on* $\Theta_0^3 \times [0, \tau]$ *such that*

$$\sup_{t\in[0,\tau],\, \forall \theta_i \in \Theta_0} \|\mathbb{S}_n(t, g_{\theta_1})[\dot{g}_{\theta_2} \dot{g}_{\theta_3}^\top] - s^{(3)}(\theta_1, \theta_2, \theta_3, t)\| \xrightarrow{P} 0, \tag{3}$$

and that the family of matrix functions $s^{(3)}(\cdot, \cdot, \cdot, t)$, $t \in [0, \tau]$ *is equicontinuous at* $(\theta_0, \theta_0, \theta_0)$. *Then there exists a neighborhood* Θ *of the true value* θ_0 *of parameter such that* $l_n(\theta)$ *is concave in* Θ *in probability.*

Remark 1. *The uniform convergence in (3) implies the uniform convergence in* Assumption *(II) for $s^{(2)}(\theta, t) = s^{(3)}(\theta, \theta, \theta, t)$. Examining the proof of Proposition 1, one can see that (3) holds if the assumptions in Proposition 1 are met.*

When the log- relative risk is linear, i.e., $g_\theta(t, Z(t)) = \theta^\top Z(t)$, the concavity of the log-partial likelihood function is a well known fact, see e.g. page 1106, Andersen and Gill (1983) [1]. From Prentice and Self (1983) [3]) (page 809), we observe that the global concavity of the log- partial likelihood function is reduced to the local concavity when the usual exponential regression form is relaxed to an arbitrary non-negative twice continuously differentiable function. Theorem 1 extends the concavity to a much larger class of smooth functions, functions of continuous first order derivatives. The concavity of the log- partial likelihood function, in fact, also holds for nonparametric relative risks.

Remark 2. *O'Sullivan (1993) [6] investigated nonparametric estimation in the Cox model, where the logarithm of the log- relative risk is expressed as a nonparametric function of the covariate process,*

$$h(t) = h_0(t) \exp[\theta(Z(t))], \quad t \geq 0,$$

where θ is a nonparametric function from some bounded open simply connected set in $\mathbb{R}^q$ to reals $\mathbb{R}$. Using Fréchet derivatives, O'Sullivan showed that the logarithm of the partial likelihood $l_n(\theta)$ and its limit $l(\theta)$ are concave in θ in some Sobolev space. For details, see pages 130-131, O'Sullivan (1993).

A special but common case is that $\theta \mapsto g_\theta(t, z)$ has continuous second order derivative for every (t, z). In this case, the log- partial likelihood function has continuous second order derivative w.r.t. θ, which is a matrix given by

$$\frac{\partial^2 \log \mathrm{PL}_n(\theta)}{\partial\theta\partial\theta^\top} = \sum_{i=1}^n \int_0^\tau \Big[\ddot{g}_\theta(t, Z_i(t)) + \Big(\frac{S_n^{(1)}(\theta, t)}{S_n^{(0)}(\theta, t)}\Big)^{\otimes 2} - \frac{S_n^{(2)}(\theta, t)}{S_n^{(0)}(\theta, t)} - \frac{\sum_{j=1}^n \ddot{g}_\theta(t, Z_j(t)) Y_j(t) \exp(g_\theta(t, Z_j(t)))}{S_n^{(0)}(\theta, t)}\Big] dN_i(t).$$

As usual, in terms of the urn model, the above stochastic integral can be written as

$$
\begin{aligned}
-\frac{1}{n}\frac{\partial^2 \log \mathrm{PL}_n(\theta)}{\partial\theta\partial\theta^\top} &= \frac{1}{n}\sum_{j=1}^{n}\int_0^\tau V_n(\theta,t)dN_i(t) \\
&\quad -\frac{1}{n}\sum_{j=1}^{n}\int_0^\tau \left\{\ddot{g}_\theta(t,Z_i(t)) - \mathbb{E}^I_{\theta,t}[\ddot{g}_\theta(t,Z_I(t))]\right\} dN_i(t), \qquad (4)
\end{aligned}
$$

where $\mathbb{E}^I_{\theta_1,x}[\ddot{g_{\theta_2}}(t,Z_I(t))] = \sum_{i=1}^n \ddot{g}_{\theta_2}(t,Z_i(t))p_i(\theta_1,x)$, $\theta_1,\theta_2 \in \Theta$ denotes the expectation calculated under the discrete distribution

$$
p_i(\theta_1,x) \equiv \frac{Y_i(x)\exp(g_{\theta_1}(x,Z_i(x)))}{\sum_{j=1}^n Y_j(x)\exp(g_{\theta_1}(x,Z_j(x)))}, \quad i=1,...,n, \qquad (5)
$$

which can be viewed as the probability that, at time point x, index i is selected from an urn containing all the n indices. Defined this way, the selected index I is a random variable and $\dot{g}_\theta(t,Z_I(t))$, as a vector function of I, is a random vector. Consequently,

$$
\begin{aligned}
V_n(\theta,x) &= \frac{S_n^{(2)}(\theta,x)}{S_n^{(0)}(\theta,x)} - \left[\frac{S_n^{(1)}(\theta,x)}{S_n^{(0)}(\theta,x)}\right]^{\otimes 2} \\
&= \mathbb{E}^I_{\theta,x}[\dot{g}_\theta(x,Z_I(x))^{\otimes 2}] - \left(\mathbb{E}^I_{\theta,x}[\dot{g}_\theta(x,Z_I(x))]\right)^{\otimes 2} \\
&= \mathbb{V}\mathrm{ar}^I_{\theta,x}\left[\dot{g}_\theta(x,Z_I(x))\right]
\end{aligned}
$$

is the urn model variance matrix of $\dot{g}_\theta(x,Z_I(x))$ at time point x.

Observe that the second term on the right hand side in (4) is identically zero if g_θ is a linear or quadratic function of θ. It is this fact that results in the global concavity of the log- partial likelihood function. The linear case corresponds to the usual Cox model in which the logarithm of the relative risk is a linear function of the parameter. Interestingly, we have found that the global concavity also holds for a quadratic relative risk function of the parameter. In Section 3, we show that this term converges to zero in probability as Θ shrinks to θ_0 and n tends to infinity under suitable conditions. Since the first term is positive definite, it follows that the log-partial likelihood is concave. We present this result below as a corollary with an independent proof given in Section 3. It may serve as a quick intuition for the concavity of the log- partial likelihood.

Corollary 1. *Suppose* Assumption *(I)-(V) hold. Suppose that $\theta \mapsto g_\theta(t,z)$ possesses continuous second order derivative $\ddot{g}_\theta(t,z)$ for every (t,z). Assume there exist a neighborhood Θ_0 of θ_0 and a bounded matrix function s_2 on $\Theta_0^2 \times [0,\tau]$ such that*

$$\sup_{\theta_1,\theta_2\in\Theta_0,\, t\in[0,\tau]} \|\mathbb{S}_n(t,g_{\theta_1})[\ddot{g}_{\theta_2}] - s_2(\theta_1,\theta_2,t)\| \xrightarrow{P} 0, \tag{6}$$

for which the family of matrix functions $s_2(\cdot,\cdot,x)$, $x \in [0,\tau]$ is equicontinuous at (θ_0,θ_0). Furthermore,

$$\sup_{\theta\in\Theta_0} \mathbb{E}\left\{\int_0^\tau \|\ddot{g}_\theta(x,Z(x))\|^2 dA(x)\right\} < \infty. \tag{7}$$

Then there exists a neighborhood Θ of the true value θ_0 of parameter such that the log- partial likelihood function $l_n(\theta)$ is concave in $\theta \in \Theta$ in probability.

Examining the proof of Proposition 1, it can be seen that (6) holds under the assumptions in Proposition 1. As a consequence of the concavity of the log- partial likelihood function and with the aid of Lemma 1 below, we can prove the consistency of the MPLE analogous to the proof of Theorem 8.3.1 [Fleming and Harrington (1991) [5]] or Theorem 1 [Tan (2007) [9]].

Theorem 2. (Consistency) *Assume that the assumptions in Theorem 1 hold. Suppose there exist a neighborhood Θ of θ_0 and scalar functions m_j on $\Theta \times [0,\tau]$ such that*

$$\sup_{x\in[0,\tau],\,\theta\in\Theta_0} \|\mathbb{S}_n(x,g_\theta)[g_\theta^j] - m_j(\theta,x)\| \xrightarrow{P} 0, \quad j=1,2. \tag{8}$$

Suppose

$$\frac{\partial}{\partial\theta}\int_0^\tau \boldsymbol{s}(x,g_{\theta_0})[g_\theta]h_0(x)\,dx\Big|_{\theta=\theta_0} = \frac{\partial}{\partial\theta}\int_0^\tau \log(s^{(0)}(\theta,x))s^{(0)}(\theta_0,x)h_0(x)\,dx\Big|_{\theta=\theta_0}, \tag{9}$$

where $\boldsymbol{s}$ is the operator defined by $\boldsymbol{s}(t,g)[h] = \mathbb{E}\big(h(t,Z(t))Y(t)\exp(g(t,Z(t)))\big)$ for a scalar function h. Then the MPLE $\hat{\theta}_n$ is consistent for θ_0, i.e., $\hat{\theta}_n \xrightarrow{p} \theta_0$.

Remark 3. *One can verify that (8) is satisfied under the assumptions of Proposition 1, while (9) is met if differentiation and integration on both sides can be swapped.*

As usual, the score function $U_n(\theta)$ can be expressed as a stochastic integral,

$$U_n(\theta) \equiv \frac{\partial l_n(\theta)}{\partial \theta} = \sum_{i=1}^{n} \int_0^\infty \Big[\dot{g}_\theta(t, Z_i(t)) - \frac{S_n^{(1)}(\theta, t)}{S_n^{(0)}(\theta, t)}\Big] dN_i(t).$$

Since $dA_i(t) = Y_i(t)\exp(g_{\theta_0}(t, Z_i(t)))h_0(t)dt$, it follows

$$\sum_{i=1}^{n} \int_0^\infty \Big[\dot{g}_{\theta_0}(t, Z_i(t)) - \frac{S_n^{(1)}(\theta_0, t)}{S_n^{(0)}(\theta_0, t)}\Big] dA_i(t)$$
$$= \int_0^\infty nS_n^{(1)}(\theta_0, t)h_0(t)\,dt - \int_0^\infty \frac{S_n^{(1)}(\theta_0, t)}{S_n^{(0)}(\theta_0, t)} nS_n^{(0)}(\theta_0, t)h_0(t)\,dt = 0, \tag{10}$$

so that

$$U_n(\theta_0) = \sum_{i=1}^{n} \int_0^\infty \Big[\dot{g}_{\theta_0}(t, Z_i(t)) - \frac{S_n^{(1)}(\theta_0, t)}{S_n^{(0)}(\theta_0, t)}\Big]\, dM_i(t)$$

is a martingale. Let the score process be defined by

$$U_n(\theta_0, t) = \sum_{i=1}^{n} \int_0^t \Big[\dot{g}_{\theta_0}(x, Z_i(x)) - \frac{S_n^{(1)}(\theta_0, x)}{S_n^{(0)}(\theta_0, x)}\Big]\, dM_i(x).$$

For an arbitrary parametric log- relative risk g_θ which possesses continuous first order derivative, we can prove the asymptotic normality of the score process. The proof uses the martingale central limit theorem and can be completed analogous to the proof of Theorem 8.2.1 [Fleming and Harrington (1991) [5]] or Theorem 2 [Tan (2007) [9]].

Theorem 3. (Asymptotic Normality of the Score Process) *Suppose that* ASSUMPTION *is satisfied. Then the following hold.*

(a) $n^{-1/2}U_n(\theta_0, t)$ *converges in distribution to a Gaussian process, where each component of the Gaussian process has independent increments, the mean of the limiting process is zero and the covariance matrix of the limiting process at time* t *is given by*

$$\Sigma(\theta_0, t) = \int_0^t v(\theta_0, x)s^{(0)}(\theta_0, x)h_0(x)\,dx.$$

(b) If $\hat{\theta}_n$ *is a consistent estimator of* θ_0*, then the substitution estimator* $\Sigma(\hat{\theta}_n, t) = \frac{1}{n}\sum_{i=1}^n \int_0^t V_n(\hat{\theta}_n, x)dN_i(x)$ *of* $\Sigma(\theta_0, t)$ *satisfies*

$$\sup_{t\in[0,\tau]} \Big\|\frac{1}{n}\sum_{i=1}^{n} \int_0^t V_n(\hat{\theta}_n, x)\,dN_i(x) - \Sigma(\theta_0, x)\Big\| \xrightarrow{P} 0, \quad n \to \infty.$$

We now present the asymptotic normality of the MPLE under the assumption of first order continuous differentiability of g_θ. Our proof uses Lemma 1, which is a generalized version of the convexity lemma (Pollard (1991) [2]). As Pollard pointed out, the convexity lemma allows to derive the limit distribution directly, without preliminary consistency arguments, so that it simplifies the asymptotic theory for estimators defined by minimization of a convex criterion function. To apply the convexity lemma to our case, we have slightly relaxed the convexity assumption, see Lemma 1. Here is the main result of this article with the proof delayed to Section 4.

Theorem 4. (Asymptotic Normality of MPLE) *Suppose* Assumption *holds. Assume as θ tends to θ_0,*

$$\mathbb{E}\left\{\int_0^\tau \|\dot{g}_\theta - \dot{g}_{\theta_0}\|^2(x, Z(x))\, dA(x)\right\} = o(1). \tag{11}$$

Assume there allows a two-term Taylor expansion:

$$\mathbb{E}[n^{-1}l_n(\theta)] = \mathbb{E}[n^{-1}l_n(\theta_0)] - (1/2)(\theta-\theta_0)^\top\Sigma(\theta_0,\tau)(\theta-\theta_0) + o_p(\|\theta-\theta_0\|^2). \tag{12}$$

If (3) holds, then $\hat{\theta}_n$ satisfies the stochastic expansion,

$$n^{1/2}(\hat{\theta}_n - \theta_0) = \Sigma^{-1}(\theta_0,\tau)n^{-1/2}U_n(\theta_0,\tau) + o_p(1). \tag{13}$$

Accordingly $\hat{\theta}_n$ is asymptotically normal with mean zero and variance-covariance matrix $\Sigma^{-1}(\theta_0,\tau)$, i.e., $n^{1/2}(\hat{\theta}_n - \theta_0) \Longrightarrow \mathcal{N}(0, \Sigma^{-1}(\theta_0,\tau))$.

2.2 Free-Knot Polynomial Spline Models

Nonlinear models are widely used in order for better understanding of various phenomena in medical and industrial sciences and other areas. Our formulation of the general parametric relative risk $r_\theta(t, Z(t))$ allows many different parameterizations. One example of this is $r_\theta(z) = (1+[\theta_2^\top z]\theta_1)^{1/\theta_3}$, $\theta = (\theta_1,\ \theta_2,\ \theta_3) \in (0,\infty)\times(-\infty,\infty)^p\times(0,\infty)$. This is a family of models given in Cox and Oakes (1984) [15] and can be used to discriminate between the linear model $r_\theta(z) = 1 + \theta_2^\top z$ ($\theta_1 = \theta_3 = 1$) and log- linear (Cox) model $r_\theta(z) = \exp(\theta_2^\top z)$ ($\theta_1 = \theta_3 \to 0$). Another example is the logistic model $r_\theta(z) = \log(1 + \exp(\theta^\top z))$. Prentice and Self (1983) [3] investigated the model of the form $r_\theta(z) = r(\beta^\top z)$, where r is a non-negative twice differentiable function. With our developed theory, r is allowed to be a function of continuous first order derivative. Many other examples can be found in Andersen, Borgan, Gill and Keiding (1993) [16].

Here we shall derive the asymptotic normality of the MPLE when the log- relative risk is expressed as a free-knot spline polynomial. Since free-knot cubic or higher order spline polynomials possess continuous second order derivatives, it is not difficult to derive the asymptotic normality of the MPLE's, e.g., by an application of Theorem 4. However, the asymptotic normality of the MPLE in a free-knot cubic spline polynomial model does not follow from Borgan (1984) [4] since the latter assumed continuous third order differentiability, while a cubic spline polynomial only possesses continuous second order differentiability. In what follows we shall consider a free-knot quadratic spline polynomial model with knots in q risk factors:

$$g_\theta(Z) = \sum_{i=1}^{q} \Big(\beta_{1i} z_i + \beta_{2i} z_i^2 + \beta_{3i}(z_i - \kappa_{1i})_+^2 + \ldots + \beta_{(m_i+2)i}(z_i - \kappa_{m_i i})_+^2\Big),$$

where both regression coefficients and knots, $\theta^\top = \big(\beta_{1i},\, \ldots,\, \beta_{(m_i+2)i},\ \kappa_{1i},\, \ldots,\, \kappa_{m_i i}\big)_{i=1}^{q}$, are parameters of interest, $Z^\top = (z_1, \ldots, z_q)$ are risk factors, and $(z-\kappa)_+$ is the positive part of $z - \kappa$ (i.e., $(z-\kappa)_+ = z - \kappa$ if $z - \kappa > 0$ otherwise it equals zero). For ease of exposition, it is without loss of generality to assume $q = 1$ (a single risk factor). Let us denote the parameter by $\theta^\top = (\beta_1, \beta_2, \beta_3, \kappa)$, and the unknown true value of parameter by $\theta_0^\top = (\beta_{10}, \beta_{20}, \beta_{30}, \kappa_0)$. Then a free-knot truncated quadratic spline polynomial with one free-knot is

$$g_\theta(z) = \beta_1 z + \beta_2 z^2 + \beta_3 (z - \kappa)_+^2, \quad z \in \mathbb{R}. \tag{14}$$

Here the intercept is excluded since it can be absorbed in the baseline hazard function. This has first order derivative give by

$$\dot{g}_\theta(z) = \left(z,\, z^2,\, (z-\kappa)_+^2,\, -2\beta_3(z-\kappa)_+\right)^\top. \tag{15}$$

Clearly $\dot{g}_\theta(z)$ is continuous in θ, but the second order partial derivative of $g_\theta(Z)$ w.r.t. κ does not exist at knot $\kappa = z$. Under the assumption that there is a neighborhood of the true knot value in which there are no observations lie, Goetghebeur and Pocock (1995) [8] and Tan (2007) [9] proved the asymptotic normality of the MPLE of the parameter θ. The former conjectured that the neighborhood restriction can be relaxed. Here we confirm their conjecture and prove the asymptotic normality of the MPLE when g_θ is a free-knot quadratic spline with the proof postponed to Section 4.

Theorem 5. (Asymptotic Normality of MPLE for a Free-knot Quadratic Spline Model) *Consider the free-knot quadratic spline polynomial model specified in (14). Suppose the assumptions in Proposition 1 are satisfied. Assume that Q has a continuous density q with respect to the Lesbegue measure. Then the MPLE $\hat{\theta}_n$ satisfies* $n^{1/2}(\hat{\theta}_n - \theta_0) \Longrightarrow \mathcal{N}\left(0, \Sigma^{-1}(\theta_0, \tau)\right).$

Truncated polynomial splines are computationally expensive, for example, to evaluate a truncated polynomial spline at a point near the right end, it is necessary to evaluate all of the basis elements and compute the entire sum. B- splines overcome this difficulty and each B- spline basis function is nonzero only on a relatively small set. For knots $\kappa_1 < \kappa_2 < ... < \kappa_k$, a B-spline basis can be represented as

$$B_i(z) = \sum_{j=1}^{k}(\theta_1(\boldsymbol{\kappa}) + \theta_2(\boldsymbol{\kappa})(z-\kappa_j) + \theta_3(\boldsymbol{\kappa})(z-\kappa_j)_+^2),$$

where $\theta_i(\boldsymbol{\kappa}), i = 1, 2, 3$ are functions of the knot vector $\boldsymbol{\kappa} = (\kappa_1, ..., \kappa_k)$ that are at least twice continuously differentiable w.r.t. each knot. More details can be found in

Schumaker (1981) [17], in particular Theorem 4.9. Thus it is not difficult to obtain an analog of Theorem 5 for B-splines and we shall omit the details.

We can lean on splines to express the departure from proportional hazards to non-proportional hazards model. In the Cox model, one basic assumption is *proportional hazards*. To check this model assumption, consider a non-proportional hazards model with constant time in covariate z,

$$h(t) = h_0(t) \exp\left\{(\theta_1 + \theta_2 G_{\theta_3}(t))z\right\}, \tag{16}$$

where θ_1 can be viewed as the intercept and $G_{\theta_3}(t)$ reflects the the nature of time dependency. Then the assumption of proportional hazards can be tested via hypothesis $\theta_2 = 0$. This formulation also accommodates *linear or quadratic decline in the log- relative risk, changepoint, and crossing hazards situations*. Knots are the changepoints. O'Quigley and Natarajan (2004) assumed $G_\theta(t) = \mathbf{1}[t \leq \theta] - \mathbf{1}[t > \theta]$ to express a sudden change in effect at some unknown time point, where θ is an unknown changepoint in time. This is, in fact, a free-knot constant spline model and is discontinuous at the changepoint θ. When real situations require continuity in both the model and its derivative, one choice is a free-knot quadratic spline polynomial model, which is the simplest model in the sense that even though linear spline polynomials are simpler than quadratic splines, linear splines do not have derivatives at the knots. A free-knot quadratic spline with constant time in covariate z and knots in time is of the form

$$g_\theta(t, z) = (\beta_0 + \beta_1(t - \kappa) + \beta_2(t - \kappa)_+^2)z, \tag{17}$$

where $\theta = (\beta_0, \beta_1, \beta_2, \kappa)^\top$ is the parameter of interest. The first order derivative is

$$\dot{g}_\theta(t, z) = \frac{\partial g_\theta}{\partial \theta}(t, z) = \big(1,\ t - \kappa,\ (t - \kappa)_+^2,\ -2\beta_2(t - \kappa)_+\big)^\top z, \tag{18}$$

which is a continuous function. The second order derivative of $g_\theta(t, z)$ w.r.t. κ does not exist. Specifically, the fourth component, $\dot{g}_{\theta,4} = -2\beta_2(t - \kappa)_+$ of $g_\theta(t, z)$, is not differentiable w.r.t. to κ. Analogous to Theorem 5, we can derive the asymptotic normality of the MPLE with some additional work in establishing an analog of Proposition 1, and we shall omit the details.

3 Proofs of Concavity

Under the assumption of continuous first order differentiability of a semiparametric parametric model, we cannot use the usual second order derivative (matrix) to prove concavity. In this section, we employ Gâteaux-differential to prove the concavity of the log-partial likelihood function, that is, we show that the Gâteaux-differential of the log-partial likelihood function is monotonically decreasing. In the end, we prove the corollary.

Proof of Theorem 1. Fix $t \in [0, \tau]$, let

$$X_n(\theta, t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \left[(g_\theta - g_{\theta_0})(x, Z_i(x)) - \log\left(S_n^{(0)}(\theta, x)/S_n^{(0)}(\theta_0, x)\right)\right] dN_i(x).$$

Clearly $X_n(\theta, \tau) = n^{-1}(l_n(\theta) - l_n(\theta_0))$. The derivative is given by

$$\dot{X}_n(\theta, t) = \frac{\partial X_n(\theta, t)}{\partial \theta} = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \left[\dot{g}_\theta(x, Z_i(x)) - \frac{S_n^{(1)}(\theta, x)}{S_n^{(0)}(\theta, x)}\right] dN_i(x).$$

For $\theta_1, \theta_2 \in \Theta$, we write

$$\begin{aligned}\dot{X}_n(\theta_1, t) - \dot{X}_n(\theta_2, t) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^t [\dot{g}_{\theta_1}(x, Z_i(x)) - \dot{g}_{\theta_2}(x, Z_i(x))]\, dN_i(x) \\ &\quad - \int_0^t \left[\frac{S_n^{(1)}(\theta_1, x)}{S_n^{(0)}(\theta_1, x)} - \frac{S_n^{(1)}(\theta_2, x)}{S_n^{(0)}(\theta_2, x)}\right]\frac{d\bar{N}(x)}{n}.\end{aligned}$$

Break the last integrand,

$$\begin{aligned}\Delta(\theta_1, \theta_2, x) &\equiv \frac{S_n^{(1)}(\theta_1, x)}{S_n^{(0)}(\theta_1, x)} - \frac{S_n^{(1)}(\theta_2, x)}{S_n^{(0)}(\theta_2, x)} \\ &= \frac{S_n^{(1)}(\theta_1, x) - S_n^{(1)}(\theta_2, x)}{S_n^{(0)}(\theta_1, x)} + \frac{S_n^{(1)}(\theta_2, x)}{S_n^{(0)}(\theta_2, x)}\frac{S_n^{(0)}(\theta_2, x) - S_n^{(0)}(\theta_1, x)}{S_n^{(0)}(\theta_1, x)},\end{aligned}$$

and write

$$S_n^{(1)}(\theta_1,x)-S_n^{(1)}(\theta_2,x)=\frac{1}{n}\sum_{i=1}^n[\dot g_{\theta_1}-\dot g_{\theta_2}](x,Z_i(x))Y_i(x)\exp(g_{\theta_1}(x,Z_i(x)))$$
$$+\frac{1}{n}\sum_{j=1}^n\dot g_{\theta_2}(x,Z_i(x))Y_i(x)\left[\exp(g_{\theta_1}(x,Z_i(x)))-\exp(g_{\theta_2}(x,Z_i(x)))\right],$$

so that

$$\frac{S_n^{(1)}(\theta_1,x)-S_n^{(1)}(\theta_2,x)}{S_n^{(0)}(\theta_1,x)}=\mathbb{E}^I_{\theta_1,x}\left[(\dot g_{\theta_1}-\dot g_{\theta_2})(x,Z_I(x))\right]$$
$$+\frac{1}{nS_n^{(0)}(\theta_1,x)}\sum_{i=1}^n\dot g_{\theta_2}(x,Z_i(x))Y_i(x)\left[\exp(g_{\theta_1}(x,Z_i(x)))-\exp(g_{\theta_2}(x,Z_i(x)))\right].$$

Hence we have

$$\begin{aligned}\Delta(\theta_1,\theta_2,x) &= \mathbb{V}\mathrm{ar}^I_{\theta_1,x}\left[\dot g_{\theta_2}(x,Z_I(x))\right](\theta_1-\theta_2)+\mathbb{E}^I_{\theta_1,x}\left[(\dot g_{\theta_1}-\dot g_{\theta_2})(x,Z_I(x))\right]\\ &\quad +a_n(\theta_1,\theta_2,x)+b_n(\theta_1,\theta_2,x), \end{aligned}\tag{19}$$

where $a_n(\theta_1,\theta_2,x)=a_{1n}(\theta_1,\theta_2,x)/S_n^{(0)}(\theta_1,x)-\mathbb{E}^I_{\theta_1,x}\left[\dot g_{\theta_2}^{\otimes 2}(x,Z_I(x))\right](\theta_1-\theta_2)$,

$$a_{n1}(\theta_1,\theta_2,x)=\frac{1}{n}\sum_{i=1}^n\dot g_{\theta_2}(x,Z_i(x))Y_i(x)\left[\exp(g_{\theta_1}(x,Z_i(x)))-\exp(g_{\theta_2}(x,Z_i(x)))\right],$$

$$b_n(\theta_1,\theta_2,x)=\frac{S_n^{(1)}(\theta_2,x)}{S_n^{(0)}(\theta_2,x)}\frac{S_n^{(0)}(\theta_2,x)-S_n^{(0)}(\theta_1,x)}{S_n^{(0)}(\theta_1,x)}-\left(\mathbb{E}^I_{\theta_1,x}\left[\dot g_{\theta_2}(x,Z_I(x))\right]\right)^{\otimes 2}(\theta_2-\theta_1).$$

By the Taylor formula, there is θ_* lying between θ_1 and θ_2 and possibly depending on n such that

$$\begin{aligned}a_{n1}(\theta_1,\theta_2,x) &= \frac{1}{n}\sum_{i=1}^n[\dot g_{\theta_2}\dot g_{\theta_*}^\top](x,Z_i(x))Y_i(x)\exp(g_{\theta_*}(x,Z_i(x)))(\theta_1-\theta_2)\\ &= \mathbb{S}_n(t,g_{\theta_*})[\dot g_{\theta_2}\dot g_{\theta_*}^\top](\theta_1-\theta_2).\end{aligned}$$

Accordingly,

$$a_n(\theta_1,\theta_2,x)=\left(\mathbb{S}_n(x,g_{\theta_*})[\dot g_{\theta_2}\dot g_{\theta_*}^\top]-\mathbb{S}_n(x,g_{\theta_1})[\dot g_{\theta_2}\dot g_{\theta_2}^\top]\right)/S_n^{(0)}(\theta_1,x)(\theta_1-\theta_2).$$

Hence it follows from ASSUMPTION (II)-(IV) and (3) that $a_n(\theta_1,\theta_2,x) = (\theta_1-\theta_2)o_p(1)$. Similarly $b_n(\theta_1,\theta_2,x) = (\theta_1-\theta_2)o_p(1)$, both uniformly in $x \in [0,\tau]$. Consequently,

$$\begin{aligned}
\dot{X}_n(\theta_1,t) - \dot{X}_n(\theta_2,t) &= -\int_0^t \mathbb{V}\mathrm{ar}^I_{\theta_1,x}\left[\dot{g}_{\theta_2}(x,Z_I(x))\right] \frac{d\bar{N}(x)}{n}(\theta_1-\theta_2) \\
&\quad - \int_0^t (a_n+b_n)(\theta_1,\theta_2,x)\,\frac{d\bar{N}(x)}{n} + d_n(\theta_1,\theta_2,x) \\
= \ & -\int_0^t \mathbb{V}\mathrm{ar}^I_{\theta_1,x}\left[\dot{g}_{\theta_2}(x,Z_I(x))\right] \frac{d\bar{N}(x)}{n}(\theta_1-\theta_2) + d_n(\theta_1,\theta_2) + o_p(\theta_1-\theta_2),
\end{aligned}$$

where

$$d_n(\theta_1,\theta_2) = \frac{1}{n}\sum_{i=1}^n \int_0^t \left[(\dot{g}_{\theta_1}-\dot{g}_{\theta_2})(x,Z_i(x)) - \mathbb{E}^I_{\theta_1,x}[(\dot{g}_{\theta_1}-\dot{g}_{\theta_2})(x,Z_I(x))]\right] dN_i(x).$$

To $dN = dM - dA$ there corresponds the decomposition $d_n(\theta_1,\theta_2) = d_{n,1}(\theta_1,\theta_2) - d_{n,2}(\theta_1,\theta_2)$. By Lenglart's inequality (see e.g. Theorem 3.4.1, Fleming and Harrington (2005) [5]),

$$d_{n,1}(\theta_1,\theta_2) = \frac{1}{n}\sum_{i=1}^n \int_0^t [(\dot{g}_{\theta_1}-\dot{g}_{\theta_2})(x,Z_i(x)) - \mathbb{E}^I_{\theta_1,x}[(\dot{g}_{\theta_1}-\dot{g}_{\theta_2})(x,Z_I(x))]]dM_i(x) \xrightarrow{P} 0.$$

With $dA_i(x) = Y_i(x)\exp(g_{\theta_0}(t,Z_i(t)))h_0(x)\,dx$, we get

$$\begin{aligned}
d_{n,2}(\theta_1,\theta_2) &= \frac{1}{n}\sum_{i=1}^n \int_0^t (\dot{g}_{\theta_1}-\dot{g}_{\theta_2})(x,Z_i(x))Y_i(x)e^{g_{\theta_0}(x,Z_i(x))}h_0(x)\,dx \\
&\quad -\frac{1}{n}\sum_{i=1}^n \int_0^t (\dot{g}_{\theta_1}-\dot{g}_{\theta_2})(x,Z_i(x))Y_i(x)e^{g_{\theta_1}(x,Z_i(x))}\frac{S_n^{(0)}(\theta_0,x)}{S_n^{(0)}(\theta_1,x)}h_0(x)\,dx \\
&= \frac{1}{n}\sum_{i=1}^n \int_0^t (\dot{g}_{\theta_1}-\dot{g}_{\theta_2})(x,Z_i(x))Y_i(x)\left[e^{g_{\theta_0}(x,Z_i(x))} - e^{g_{\theta_1}(x,Z_i(x))}\right]h_0(x)dx \\
&\quad -\frac{1}{n}\sum_{i=1}^n \int_0^t (\dot{g}_{\theta_1}-\dot{g}_{\theta_2})(x,Z_i(x))Y_i(x)e^{g_{\theta_1}(x,Z_i(x))}\left[\frac{S_n^{(0)}(\theta_0,x)}{S_n^{(0)}(\theta_1,x)} - 1\right]h_0(x)\,dx \\
&= (\theta_1-\theta_2)o_p(1), \quad as \quad \theta_1,\theta_2 \to \theta_0.
\end{aligned} \tag{20}$$

To claim the last equality (20), we first note that (3) implies that for $\theta_1,\theta_2 \in \Theta$, uniformly in $t \in [0,\tau]$,

$$\mathbb{E}\left\{(\dot{g}_{\theta_1}-\dot{g}_{\theta_2})(t,Z(t))Y(t)[\exp(g_{\theta_1}(t,Z(t))) - \exp(g_{\theta_2}(t,Z(t)))]\right\} = o(\|\theta_1-\theta_2\|). \tag{21}$$

This, the law of large numbers, (I)-(IV) and (3) yield (20). Thus $d_n(\theta_1, \theta_2) = (\theta_1 - \theta_2)o_p(1)$. Combining the above, we conclude

$$(\theta_1 - \theta_2)^\top \left[\dot{X}_n(\theta_1, t) - \dot{X}_n(\theta_2, t)\right] \leq 0, \quad \theta_1, \theta_2 \in \Theta,$$

as n is sufficiently large and the neighborhood Θ of θ_0 is small. This implies, by Proposition 5.5 of Ekeland and Temam (1976) [18], that $X_n(\theta, \tau)$ hence $l_n(\theta)$ is concave in $\theta \in \Theta$ for large n and small neighborhood Θ of θ_0. $\square$

Proof of Corollary 1. We only need to prove that the second term in (4) converges to zero in probability as Θ shrinks to θ_0 and n tends to infinity. To this end, we break it as follows,

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{\ddot{g}_\theta(x, Z_i(x)) - \mathbb{E}^I_{\theta,x}[\ddot{g}_\theta(x, Z_I(x))]\right\} dN_i(x)\right\| \\
&\leq \left\|\frac{1}{n}\sum_{i=1}^n \int_0^\tau \Big\{\ddot{g}_\theta(x, Z_i(x)) - \ddot{g}_{\theta_0}(x, Z_i(x)) \\
&\quad -\mathbb{E}^I_{\theta,x}[\ddot{g}_\theta(x, Z_I(x))] + \mathbb{E}^I_{\theta_0,x}[\ddot{g}_{\theta_0}(x, Z_I(x))]\Big\} dN_i(x)\right\| \qquad (22) \\
&\quad +\left\|\frac{1}{n}\sum_{i=1}^n \int_0^\tau \Big\{\ddot{g}_{\theta_0}(t, Z_i(t)) - \mathbb{E}^I_{\theta_0,x}[\ddot{g}_{\theta_0}(x, Z_I(x))]\Big\} dN_i(x)\right\|. \qquad (23)
\end{aligned}$$

Thus, the desired zero limit is implied by both (22) and (23) converge to zero in probability. The proof of (23) is simple. Using $dN = dM + dA$, it is implied by the following two limits:

$$\frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{\ddot{g}_{\theta_0}(x, Z_i(x)) - \mathbb{E}^I_{\theta_0,x}[\ddot{g}_{\theta_0}(x, Z_I(x))]\right\} dM_i(x) \xrightarrow{P} 0, \qquad (24)$$

and

$$\frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{\ddot{g}_{\theta_0}(x, Z_i(x)) - \mathbb{E}^I_{\theta_0,x}[\ddot{g}_{\theta_0}(x, Z_I(x))]\right\} dA_i(x) \xrightarrow{P} 0.$$

Clearly the latter holds since it is equal to

$$\int_0^\tau \Big\{ \sum_{i=1}^n \ddot{g}_{\theta_0}(x, Z_i(x)) p_i(\theta_0, x) - \mathbb{E}^I_{\theta_0,x}[\ddot{g}_{\theta_0}(x, Z_I(x))]\Big\} S_n^{(0)}(\theta_0, x) h_0(x) dx = 0.$$

Applying Lenglart's inequality [Theorem 3.4.1, Fleming and Harrington (1991) [5]] to every entry of (24), we have for $\epsilon > 0$ and $\eta > 0$,

$$
\begin{aligned}
&P\Big\{|\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\Big[\ddot{g}_{\theta_0}(x,Z_i(x)) - \mathbb{E}^I_{\theta_0,x}[\ddot{g}_{\theta_0}(x,Z_I(x))]\Big]_{j,k} dM_i(x)|^2 \geq \epsilon\Big\} \\
&\leq \frac{\eta}{\epsilon} + P\Big\{n^{-2}\sum_{i=1}^{n}\int_0^{\tau}\Big[\ddot{g}_{\theta_0}(x,Z_i(x)) - \mathbb{E}^I_{\theta_0,x}[\ddot{g}_{\theta_0}(x,Z_I(x))]\Big]^2_{j,k} dA_i(x) \geq \eta\Big\} \\
&\leq \frac{\eta}{\epsilon} + P\Big\{n^{-2}\sum_{i=1}^{n}\int_0^{\tau}\Big[\ddot{g}_{\theta_0}(x,Z_i(x))\Big]^2_{j,k} dA_i(x) \geq \eta\Big\}.
\end{aligned}
$$

By the law of large numbers, it follows from (7) that the last probability is zero when n is sufficiently large for any $\eta > 0$. Therefore by taking $\eta = \epsilon^2$ we obtain (24) as ϵ tends to zero. Bound (22) by the sum of the following two expressions:

$$
\|\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{\ddot{g}_{\theta}(x,Z_i(x)) - \ddot{g}_{\theta_0}(x,Z_i(x))\}\, dN_i(x)\| \equiv D, \tag{25}
$$

and

$$
\|\int_0^{\tau}\{\mathbb{E}^I_{\theta,x}[\ddot{g}_{\theta}(x,Z_I(x))] - \mathbb{E}^I_{\theta_0,x}[\ddot{g}_{\theta_0}(x,Z_I(x))]\}\,\frac{d\bar{N}(x)}{n}\|. \tag{26}
$$

Using $dN = dM + dA$, we have $D \leq D_1 + D_2$, where D_1, D_2 are obtained from D with N replaced by M and A respectively. Analogously for $\epsilon > 0$ and $\eta > 0$,

$$
\begin{aligned}
\mathbb{P}([D_1]_{jk} \geq \epsilon) &= \mathbb{P}\Big\{|\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\Big[\ddot{g}_{\theta}(x,Z_i(x)) - \ddot{g}_{\theta_0}(x,Z_i(x))\Big]_{j,k} dM_i(x)|^2 \geq \epsilon\Big\} \\
&\leq \frac{\eta}{\epsilon} + \mathbb{P}\Big\{n^{-2}\sum_{i=1}^{n}\int_0^{\tau}\Big[\ddot{g}_{\theta}(x,Z_i(x)) - \ddot{g}_{\theta_0}(x,Z_i(x))\Big]^2_{j,k} dA_i(x) \geq \eta\Big\} \\
&\leq \frac{\eta}{\epsilon} + \mathbb{P}\Big\{n^{-2}\sum_{i=1}^{n}\int_0^{\tau}\Big[\ddot{g}_{\theta}(x,Z_i(x))\Big]^2_{j,k} dA_i(x) \geq \eta/4\Big\} \\
&\qquad + \mathbb{P}\Big\{n^{-2}\sum_{i=1}^{n}\int_0^{\tau}\Big[\ddot{g}_{\theta_0}(x,Z_i(x))\Big]^2_{j,k} dA_i(x) \geq \eta/4\Big\}.
\end{aligned}
$$

Again by the law of large number, it follows from (7) that the last two probabilities converge to zero as n tends to infinity. By taking $\eta = \epsilon^2$ we show D_1 converges to zero in probability as ϵ tends to zero. By Assumption (I) and the equicontinuity of s_2 in (6),

$$\begin{aligned} D_2 &= \|\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \{\ddot{g}_\theta(x, Z_i(x)) - \ddot{g}_{\theta_0}(x, Z_i(x))\}\, dA_i(x)\| \\ &\leq \int_0^\tau \|\mathbb{S}_n(x, g_{\theta_0})[\ddot{g}_\theta] - s_2(\theta_0, \theta, x)\| h_0(x)\, dx \\ &\quad + \int_0^\tau \|\mathbb{S}_n(x, g_{\theta_0})[\ddot{g}_{\theta_0}] - s_2(\theta_0, \theta_0, x)\| h_0(x)\, dx \\ &\quad + \int_0^\tau \|s_2(\theta_0, \theta, x) - s_2(\theta_0, \theta_0, x)\| h_0(x)\, dx \\ &\xrightarrow{P} 0. \end{aligned}$$

Therefore $D \xrightarrow{p} 0$ as Θ shrinks to θ_0 and n tends to infinity. Let $S_{n,2}(\theta, x) = \mathbb{S}_n(x, g_\theta)[\ddot{g}_\theta]$. Then the integrand in (26) can be broken as

$$\begin{aligned} &\mathbb{E}^I_{\theta,x}[\ddot{g}_\theta(x, Z_I(x))] - \mathbb{E}^I_{\theta_0,x}[\ddot{g}_{\theta_0}(x, Z_I(x))] = \frac{S_{n,2}(\theta, x)}{S_n^{(0)}(\theta, x)} - \frac{S_{n,2}(\theta_0, x)}{S_n^{(0)}(\theta_0, x)} \\ &= \frac{S_{n,2}(\theta, x) - S_{n,2}(\theta_0, x)}{S_n^{(0)}(\theta, x)} - \frac{S_{n,2}(\theta_0, x)[S_n^{(0)}(\theta, x) - S_n^{(0)}(\theta_0, x)]}{S_n^{(0)}(\theta_0, x) S_n^{(0)}(\theta, x)} \\ &\equiv B_n(\theta, x) - C_n(\theta, x), \quad \text{say.} \end{aligned}$$

Since $S_n^{(0)}(\theta, x)$ is bounded away from zero (by $1/\eta > 0$ say) for large n, it follows

$$\Big\| \int_0^\tau B_n(\theta, x) \frac{d\bar{N}(x)}{n} \Big\| \leq \frac{\eta}{n} \bar{N}(\tau) \sup_{x \in [0,\tau]} \|S_{n,2}(\theta, x) - S_{n,2}(\theta_0, x)\|.$$

The boundedness in probability of $\frac{1}{n}\bar{N}(\tau)$ and condition (6) imply that the above is negligible in probability as Θ shrinks to θ_0 and n tends to infinity. A similar argument verifies

$$\Big\| \int_0^\tau C_n(\theta, x) \frac{d\bar{N}(x)}{n} \Big\|$$

also becomes negligible in probability as Θ shrinks to θ_0 and n tends to infinity. Combining the above shows that (26) and hence (22) is negligible in probability again as Θ shrinks to θ_0 and n tends to infinity. Thus we have shown the negligibility of both (22) and (23).

It has been shown the last line in (4) converges to zero in probability for all $\theta \in \Theta$ when $n \to \infty$ and Θ shrinks to θ_0. Since $V_n(\theta, x)$ is the urn model variance-covariance matrix of $\dot{g}_\theta(x, Z_I(x))$ at time point x, it is positive definite and $-\frac{1}{n}\sum_{i=1}^n \int_0^\tau V_n(\theta, x) dN_i(x)$ is negative definite. This and Assumption (V) imply that $X_n(\theta, \tau)$ is a concave function of θ when n is large and $\theta \in \Theta$ with small neighborhood Θ of θ_0. $\square$

4 Proofs of Asymptotic Normality

In this Section, we first prove the convexity lemma which slightly relaxes the convexity assumption of the existing result (Andersen and Gill(1982)[1]; Pollard (1991)[2]; Lemma 8.3.1, Fleming and Harrington (1991) [5]). Using this lemma and the characterization of minimizers, we prove the asymptotic normality of MPLE's. As an application of the obtained results, we show the asymptotic normality of MPLE's in a free-knot quadratic spline polynomial model.

Lemma 1. *Let $\{\lambda_n(\theta) : \theta \in \Theta\}$ be a sequence of random functions defined on a convex, open subset Θ of $\mathbb{R}^d$. Suppose that there exists a sequence $\{C_n\}$ of measurable sets with $P(C_n) \to 1$ such that each $\lambda_n(\theta)$ is convex on C_n for every $\theta \in \Theta$. Suppose $\lambda(\cdot)$ is a real-valued function on Θ for which $\lambda_n(\theta) \to \lambda(\theta)$ in probability for each $\theta \in \Theta$. Then the following hold.*
(1) For each compact subset K of Θ, $\sup_{\theta\in K} |\lambda_n(\theta) - \lambda(\theta)| \xrightarrow{P} 0$. The function $\lambda(\cdot)$ is necessarily convex in Θ.
(2) If λ_n has a unique maximum at θ_n and λ has one at θ_0, then $\theta_n \xrightarrow{p} \theta_0$ as $n \to \infty$.

Proof. Fix $\epsilon > 0$. The proof proceeds the same as Pollard [2], but the last two limits in his proof are replaced with

$$\mathbb{P}\Big(\sup_{\theta\in K}(\lambda_n(\theta) - \lambda(\theta)) > 2\epsilon, C_n\Big) \to 0, \quad \mathbb{P}\Big(\inf_{\theta\in K}(\lambda_n(\theta) - \lambda(\theta)) < -3(d+1)\epsilon, C_n\Big) \to 0,$$

which clearly hold by assumption. Therefore,

$$\mathbb{P}\big(\sup_{\theta\in K}(\lambda_n(\theta)-\lambda(\theta))>2\epsilon\big)\to 0,\quad \mathbb{P}\big(\inf_{\theta\in K}(\lambda_n(\theta)-\lambda(\theta))<-3(d+1)\epsilon\big)\to 0.$$

These yield the desired uniform convergence. □

Proof of Theorem 4: Recall $X_n(\theta,t)$ in the proof of Theorem 1, and let $X_n(\theta)=X_n(\theta,\tau)$ and $B_n(\theta)=B_n(\theta,\tau)$, where

$$B_n(\theta,t)=\frac{1}{n}\sum_{i=1}^{n}\int_0^t\left[(g_\theta-g_{\theta_0})(x,Z_i(x))-\log\left(S_n^{(0)}(\theta,x)/S_n^{(0)}(\theta_0,x)\right)\right]\,dA_i(x).$$

Then $X_n(\theta)-A_n(\theta)$ is a martingale. The derivatives of $X_n(\theta)$ and $B_n(\theta)$ are

$$\begin{aligned}\dot{X}_n(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau\left(\dot{g}_\theta(x,Z_i(x))-\left(S_n^{(1)}(\theta,x)/S_n^{(0)}(\theta,x)\right)\right)\,dN_i(x),\\ \dot{B}_n(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau\left(\dot{g}_\theta(x,Z_i(x))-\left(S_n^{(1)}(\theta,x)/S_n^{(0)}(\theta,x)\right)\right)\,dA_i(x).\end{aligned}$$

For $\alpha\in\mathbb{R}^p$, let $\alpha_n=n^{-1/2}\alpha$ and $D_n(\alpha)=n\big(X_n(\theta_0+\alpha_n)-\alpha_n^\top\dot{X}_n(\theta_0)\big)$. The asymptotic behaviors of the MPLE can be derived from the investigation of $D_n(\alpha)$. Since $C_n(\alpha)\equiv X_n(\theta_0+\alpha_n)-B_n(\theta_0+\alpha_n)-\alpha_n^\top[\dot{X}_n(\theta_0)-\dot{B}_n(\theta_0)]$ is a martingale, it follows

$$\mathbb{V}\mathrm{ar}\,\{nC_n(\alpha)\}=n\mathbb{E}\Big(\int_0^\tau c_n^2(x,Z_1(x))\,dA_1(x)\Big),\tag{27}$$

where $c_n(x,z)=c_{n1}(x,z)-c_{n2}(x,z)$ with $c_{n1}(x,z)=(g_{\theta_0+\alpha_n}-g_{\theta_0}-\alpha_n^\top\dot{g}_{\theta_0})(x,z)$ and $c_{n2}(x,z)=\log[S_n^{(0)}(\theta_0+\alpha_n,x)/S_n^{(0)}(\theta_0,x)]-\alpha_n^\top[S_n^{(1)}(\theta_0,x)/S_n^{(0)}(\theta_0,x)]$. Using $c_n^2\le 2c_{n1}^2+2c_{n2}^2$, we bound $\mathbb{V}\mathrm{ar}\,\{nC_n(\alpha)\}\le 2V_{n1}(\alpha)+2V_{n2}(\alpha)$, where

$$V_{ni}(\alpha)=n\mathbb{E}\Big(\int_0^\tau c_{ni}^2(x,Z_1(x))\,dA_1(x)\Big),\quad i=1,2.$$

Then it follows from (11) and the Taylor expansion that $V_{n1}\to 0$ as $n\to\infty$. By Assumption (I), the uniform convergence in (II), the equicontinuity and the positivity of $s^{(0)}$ in (IV), and the Taylor expansion at $\alpha=0$, it can be seen that $V_{n2}(\alpha)\to 0$, so that $\mathbb{V}\mathrm{ar}\,\{nC_n(\alpha)\}\to 0$ as $n\to\infty$. Using again (11), it can be shown analogously that $\mathbb{V}\mathrm{ar}\{n[B_n(\theta_0+\alpha_n)-\alpha_n^\top \dot{B}_n(\theta_0)]\} \to 0$. Therefore, we conclude $\mathbb{V}\mathrm{ar}\{D_n(\alpha)\} \to 0$ as $n \to \infty$. This last limit immediately yields as n tends to infinity,

$$D_n(\alpha) - \mathbb{E}(D_n(\alpha)) \xrightarrow{P} 0, \quad \alpha \in \mathbb{R}^p. \tag{28}$$

Noticing that $\mathbb{E}(\dot{X}_n(\theta_0)) = 0$ by (10) and $X_n(\theta) = n^{-1}(l_n(\theta) - l_n(\theta_0))$, the two-term Taylor expansion (12) now yields

$$D_n(\alpha) = -\frac{1}{2}\alpha^\top \Sigma(\theta_0, \tau)\alpha + o_p(1), \quad \alpha \in \mathbb{R}^p \tag{29}$$

for large n. Since $-X_n(\theta)$ is convex in θ in a small neighborhood Θ of θ_0, it follows from Lemma 1 that for arbitrary $M > 0$ and large n,

$$\sup_{\|\alpha\| \le M} \left| n\left\{ X_n(\theta_0 + n^{-1/2}\alpha) - n^{-1/2}\alpha^\top \dot{X}_n(\theta_0)\right\} + \frac{1}{2}\alpha^\top \Sigma(\theta_0,\tau)\alpha \right| = o_p(1). \tag{30}$$

Let $\tilde{X}_n(\alpha) = X_n(\theta_0 + n^{-1/2}\alpha)$ and $\hat{\alpha}_n = \arg\max_{\alpha \in \mathbb{R}^p} \tilde{X}_n(\alpha)$, so that $\hat{\alpha}_n = n^{1/2}(\hat{\theta}_n - \theta_0)$. Then for any random variable γ bounded in probability, the above equality implies

$$\tilde{X}_n(\gamma) = \gamma^\top n^{-1/2} U_n(\theta_0) - \frac{1}{2}\gamma^\top \Sigma(\theta_0,\tau)\gamma + o_p(1). \tag{31}$$

This shows that $\tilde{X}_n(\gamma)$ can be approximated by a quadratic function in γ, which is uniquely maximized by $\hat{\gamma}_n = \Sigma^{-1}(\theta_0,\tau) n^{-1/2} U_n(\theta_0)$. The maximized value of $\tilde{X}_n(\gamma)$ is approximately $\tilde{X}_n(\hat{\gamma}_n) = \frac{1}{2}\hat{\gamma}_n^\top \Sigma(\theta_0,\tau)\hat{\gamma}_n$. This together with the replacement of $n^{-1/2}U_n(\theta_0)$ by $\Sigma(\theta_0,\tau)\hat{\gamma}_n$ in (31) gives

$$\tilde{X}_n(\gamma) = \tilde{X}_n(\hat{\gamma}_n) - \frac{1}{2}(\hat{\gamma}_n - \gamma)^\top \Sigma(\theta_0,\tau)(\hat{\gamma}_n - \gamma) + o_p(1) \tag{32}$$

for any random γ bounded in probability. Now using the concavity of $\tilde{X}_n(\cdot)$ in a small neighborhood of the origin and the characterization of the maximizer, one can show $\hat{\gamma}_n - \hat{\alpha}_n \xrightarrow{p} 0$ as $n \to \infty$ as in Pollard (1991) [2]. Specifically, fix $\epsilon > 0$. If $\|\hat{\alpha}_n - \hat{\gamma}_n\| > \epsilon$, then there exists $\hat{\gamma}_n^*$ on the line segment joining $\hat{\alpha}_n$ and $\hat{\gamma}_n$ such that $\hat{\gamma}_n^* - \hat{\gamma}_n = \epsilon \upsilon_n$, where υ_n is a unit vector. From this equality and the boundedness of

$\hat{\gamma}_n$ in probability it immediately follows that $\hat{\gamma}_n^*$ is bounded in probability, so that we can substitute $\gamma = \hat{\gamma}_n^*$ in (32) to obtain

$$\tilde{X}_n(\hat{\gamma}_n^*) = \tilde{X}_n(\hat{\gamma}_n) - \frac{1}{2}\epsilon^2 \upsilon_n^\top \Sigma(\theta_0, \tau)\upsilon_n + o_p(1). \tag{33}$$

Since $\tilde{X}_n(\cdot)$ is concave and $\tilde{X}_n(\hat{\alpha}_n) \geq \tilde{X}_n(\hat{\gamma}_n)$, it follows that $\tilde{X}_n(\hat{\gamma}_n^*) \geq \tilde{X}_n(\hat{\gamma}_n)$. Accordingly, the preceding display implies

$$\frac{1}{2}\epsilon^2 \upsilon_n^\top \Sigma(\theta_0, \tau)\upsilon_n + o_p(1) \leq 0.$$

This shows that $\mathbb{P}(\|\hat{\alpha}_n - \hat{\gamma}_n\| > \epsilon) \leq \mathbb{P}(\frac{1}{2}\epsilon^2 \upsilon_n^\top \Sigma(\theta_0, \tau)\upsilon_n + o_p(1) \leq 0) \to 0$ as n tends to infinity. Hence $\hat{\alpha}_n = \hat{\gamma}_n + o_p(1)$ and the desired (13) follows. □

Proof of Theorem 5: We shall prove the theorem by verifying the assumptions in Theorem 4. First, we verify Assumption by Proposition 1 and (3) by Remark 1. As discussed in Subsection 2.2, a free-knot truncated quadratic polynomial spline $g_\theta(z) = \beta_1 z + \beta_2 z^2 + \beta_3 (z - \kappa)_+$ has continuous first order derivative given by

$$\dot{g}_\theta(z) = \big(z, z^2, (z-\kappa)_+^2, -2\beta_3(z-\kappa)_+\big)^\top$$

at θ_0, where $\theta = (\beta_1, \beta_1, \beta_3, \kappa)^\top$. By assumption, Z is bounded, say $|Z| \leq B$ for some $B > 0$. This, the continuity of $\dot{g}_\theta$ at θ_0 and an application of the dominated convergence theorem yield (11). Therefore we are left to verify the two-term expansion (12). Recall in the proof of Theorem 4, for fixed $\alpha \in \mathbb{R}^4$ we denote $D_n(\alpha) = n\big(X_n(\theta_0 + n^{-1/2}\alpha) - \alpha_n^\top \dot{X}_n(\theta_0)\big)$ with $\alpha_n = \theta_0 + n^{-1/2}\alpha$ and $X_n(\theta) = n^{-1}(l_n(\theta) - l_n(\theta_0))$. Since $\mathbb{E}(\dot{X}_n(\theta_0)) = 0$ by (10), the desired (12) can be derived from

$$\mathbb{E}[D_n(\alpha)] = -\frac{1}{2}\alpha^\top \Sigma(\theta_0, \tau)\alpha + o_p(1), \quad n \to \infty. \tag{34}$$

To show this, we first prove

$$\frac{\partial^2}{\partial\alpha\partial\alpha^\top}\mathbb{E}[D_n(\alpha)]\big|_{\alpha=0} = -\Sigma(\theta_0, \tau) + o_p(1). \tag{35}$$

By the dominated convergence theorem, the following derivative can pass the expectation, so that

$$\frac{\partial}{\partial\theta}\mathbb{E}\left\{\int_0^\tau \left\{g_\theta(Z)\,dN(t) - \log\left[S_n^{(0)}(\theta,t)/S_n^{(0)}(\theta_0,t)\right]\,n^{-1}d\bar{N}(t)\right\}\right\}\Big|_{\theta=\theta_0}$$
$$= \mathbb{E}\left\{\int_0^\tau \left\{\dot{g}_\theta(Z)\,dA(t) - \left[S_n^{(1)}(\theta,t)/S_n^{(0)}(\theta,t)\right]\,n^{-1}d\bar{A}(t)\right\}\big|_{\theta=\theta_0}\right\}.$$

Next we consider further differentiation of the above expectation. Notice that the first three components of $\dot{g}_\theta(z)$ possess continuous partial derivatives while the fourth component is not differentiable w.r.t. κ at z. Let $\dot{g}_{\theta,4}(z)$ denote the fourth component of $\dot{g}_\theta(z)$, so that $\dot{g}_{\theta,4}(z) = -2\beta_3(z-\kappa)_+$. Then the fourth component of $\mathbb{E}\int_0^\tau \dot{g}_\theta(Z)\,dA(t)$ can be expressed as

$$\mathbb{E}\int_0^\tau \dot{g}_{\theta,4}(Z)\,dA(t) = \mathbb{E}\left\{\int_0^\tau Y(t)h_0(t)\,dt\int_\kappa^B -2\beta_3(z-\kappa)\exp(g_{\theta_0}(z))q(z)\,dz\right\},$$

where q is the density of Q. Since q is continuous, it follows that the right side of the preceding equality is differentiable w.r.t. κ at κ_0 with the partial derivative

$$\frac{\partial}{\partial\kappa}\mathbb{E}\left\{\int_0^\tau \dot{g}_{\theta,4}(Z)\,dA(t)\right\}\Big| = \mathbb{E}\left\{\int_0^\tau Y(t)h_0(t)\,dt\int_\kappa^B 2\beta_3\exp(g_{\theta_0}(z))q(z)\,dz\right\} \quad (36)$$

Thus $\mathbb{E}\left\{\int g_\theta(Z)\,dN(t)\right\}$ has second order derivative at θ_0. Analogously, one can show that $s^{(1)}(\theta,t)$ is differentiable w.r.t. θ at θ_0. In particular, the fourth component of $s^{(1)}(\theta,t)$ can be written as

$$\mathbb{E}\left\{\int_{-B}^B \dot{g}_{\theta,4}(z)Y(t)\exp(g_\theta(z))\,dQ(z)\right\} = \mathbb{E}\left\{\int_\kappa^B -2\beta_3(z-\kappa)Y(t)\exp(g_\theta(z))\,dQ(z)\right\}.$$

This is differentiable w.r.t. κ at κ_0 with the partial derivative equal to

$$\mathbb{E}\left\{\dot{g}_{\theta,4}^2(Z)Y(t)\exp(g_\theta(Z))\right\} + \mathbb{E}\left\{\int_\kappa^B 2\beta_3 Y(t)\exp(g_\theta(z))\,dQ(z)\right\}. \quad (37)$$

After some algebra, we get

$$\frac{\partial}{\partial\theta^\top}s^{(1)}(\theta_0,t) = \mathbb{E}\left\{\Delta(Z,\theta_0)Y(t)\exp(g_{\theta_0}(Z))\right\} + \mathbb{E}\left\{\dot{g}_{\theta_0}(Z)^{\otimes 2}Y(t)\exp(g_{\theta_0}(Z))\right\}, \quad (38)$$

where $\Delta(z,\theta)$ is a 4×4 matrix with entries $\Delta_{i,j}(z,\theta) = 0$ except $\Delta_{3,4}(z,\theta) = \Delta_{4,3}(z,\theta) = -2(z-\kappa)_+$ and $\Delta_{4,4}(z,\theta) = 2\beta_3 \mathbf{1}_{[\kappa,B]}(z)$. Similarly, we derive

$$\frac{\partial^2}{\partial\theta\partial\theta^\top}\mathbb{E}\left\{\int_0^\tau g_\theta(Z)\,dN(t)\right\}\Big|_{\theta=\theta_0} = \mathbb{E}\left\{\int_0^\tau \Delta(Z,\theta_0)Y(t)\exp(g_{\theta_0}(Z))h_0(t)\,dt\right\}. \quad (39)$$

In view of Proposition 1, (II), (38) and (39), we obtain

$$\begin{aligned}
&\frac{\partial^2}{\partial\theta\partial\theta^\top}\mathbb{E}\left\{\int_0^\tau g_\theta(Z)\,dN(t) - \log\left[S_n^{(0)}(\theta,t)/S_n^{(0)}(\theta_0,t)\right]\, n^{-1}d\bar{N}(t)\right\}\Big|_{\theta=\theta_0} \\
&= \frac{\partial}{\partial\theta^\top}\mathbb{E}\left\{\int_0^\tau \left(\dot{g}_\theta(Z) - \frac{s^{(1)}}{s^{(0)}}(\theta,t)\right) dA(t)\right\}\Big|_{\theta=\theta_0} + o_p(1) \\
&= -\int_0^\tau v(\theta_0,t)s^{(0)}(\theta_0,t)h_0(t)\,dt + o_p(1) = -\Sigma(\theta_0,\tau) + o_p(1) \qquad (40)
\end{aligned}$$

for large n. This shows (35). Expanding $\mathbb{E}[D_n(\alpha)]$ at $\alpha = 0$ in a second order Taylor formula and noticing $\mathbb{E}[D_n(0)] = 0$ and $(\partial/\partial\alpha)\mathbb{E}[D_n(\alpha)]|_{\alpha=0} = 0$, we arrive at

$$\mathbb{E}[D_n(\alpha)] = \frac{1}{2}\alpha^\top \frac{\partial^2}{\partial\alpha\partial\alpha^\top}\mathbb{E}[D_n(\alpha)]\Big|_{\alpha=0}\alpha + r_n,$$

where r_n is the remainder. Since the distribution function Q has a continuous density q, both (36) and (37) are continuous functions at $\theta = \theta_0$. Therefore, it follows from (40) that $r_n = o_p(1)$. This shows (34). We now apply Theorem 4 to complete the proof. □

References

[1] P.K. Andersen and R.D. Gill. Cox's Regression Model for Counting Processes: A Large Sample Study. *Ann. Statist.*, 10:1100–1120, 1982.

[2] D. Pollard. Asymptotics for Least Absolute Deviation Regression Estimators. *Econometric Theory*, 7:186–199, 1991.

[3] R.L. Prentice and S. G. Self. Asymptotic distribution theory for Cox-type Regression Models with general risk form. *Ann. Statist.*, 11:804–813, 1983.

[4] ∅. Borgan. Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scand. J. Statist.*, 11:1–16, 1982.

[5] T. Fleming and D. Harrington. *Counting Processes and Survival Analysis.* Wiley, Hoboken, New Jersey, 2005.

[6] F. O'Sullivan. Nonparametric Estimation in the Cox Model. *Ann. Statist.*, 21:124–145, 1993.

[7] N.L. Hjort and D. Pollard. Asymptotics for minimisers of convex processes. *Statistical Research Report*, 1993.

[8] E. Goetghebeur and S. Pocock. Detection and Estimation of J-shaped Risk-Response Relationships. *J. Roy. Statist. Soc. Ser. A*, 158, Part 1:107–121, 1995.

[9] F. Tan. A method for finding the nadir. *Ph.D Dissertation, Department of Statistics, Florida State University*, 2007.

[10] Mathsoft. *S-Plus version 3.4 for Unix Supplement.* Data Analysis Products Division, Mathsoft: Seattle, 1996.

[11] W. Mao and L.H. Zhao. Free-knot polynomial splines with confidence intervals. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 65:901–919, 2003.

[12] N. Molinari, J.P. Daures, and J.F. Durand, J.F. Regression splines for threshold selection in survival data analysis. *Stat. Med.*, 20:237–247, 2001.

[13] R. Giorgi, M. Abrahamowicz, C. Quantin, P. Bolard, J. Esteve, J. Gouvernet and J. Faivre. A relative survival regression model using B-spline functions to model non-proportional hazards. *Stat. Med.*, 22:2767–2784, 2003.

[14] N. Molinari, J.P. Durand, and J.F. Sabatier. Bounded optimal knots for regression splines. *Comput. Statist. Data Anal.*, 45:159–178, 2004.

[15] D.R. Cox. Regression models and life-tables (with discussion). *J.R. Statist. Soc. B*, 34:187–220, 1972.

[16] P.K. Andersen, O. Borgan, R.D. Gill, and N. Keiding. *Statistical models based on Counting processes*. Spring, New York, USA, 1993.

[17] L. Schumaker. *Spline Functions: Basic Theory*. Wiley, New York, USA, 1981.

[18] I. Ekeland and R. Teman. *Convex Analysis and Variational Problems*. American Elsevier Publishing Company, Inc., New York, 1976.