

NONPARAMETRIC SURVIVAL ANALYSIS ON TIME-DEPENDENT COVARIATE EFFECTS IN CASE-COHORT SAMPLING DESIGN

Haimeng Zhang and Chunfeng Huang

Concordia College and North Dakota State University

Abstract: A nonparametric analysis of time-dependent covariate effects on failures determined by a regression function $\beta_0(t)$ in Cox's regression model based on case-cohort sampling design is developed. The analysis is carried out through maximizing appropriate penalized pseudolikelihoods. Weak uniform consistency and pointwise asymptotic normality of the resulting estimators are investigated under regularity conditions. Further generalization of the results is also discussed in the paper.

Key words and phrases: Asymptotic normality, case-cohort sampling, Cox model, consistency, maximum penalized pseudolikelihood estimator, survival analysis.

1. Introduction

The Cox proportional hazards model has played a prominent role in both the statistical literature and the analysis of right-censored survival data since it was first introduced in Cox (1972). It has been widely used for analysis of biomedical data from both longitudinal studies and clinical trials, mainly due to its appealing mathematical simplicity, as well as its general availability through most statistical packages. In this model, the data consist of *i.i.d.* copies (T_i, D_i, Z_i) , $i = 1, \dots, n$, of (T, D, Z) , where $T_i = \min\{T_i^0, V_i^0\}$, $D_i = I(T_i = T_i^0)$ the censoring indicator function, and T_i^0 , V_i^0 , and Z_i represent individual i 's failure time, censoring time, and random covariate, respectively. In addition, it is further assumed that the conditional hazard function given z at time $t \in [0, 1]$ is of the form $\lambda(t|z) = \lambda_0(t) \exp[\beta_0 z]$ with the unknown parameter β_0 and an unknown baseline hazard function $\lambda_0(t)$.

While the Cox proportional hazards model is relatively simple to present, the assumption of the constant effect of a covariate on survival over the entire follow-up period may not be met, and a generalization of the original model becomes necessary. There are various nonparametric approaches proposed in the literature. These include, for example, the sieve estimation procedure of Murphy and Sen (1991); the goodness-of-fit testing in Marzec and Marzec (1997)

based on the sieve estimator; a local partial likelihood estimation technique in Fan, Gijbels and King (1997) and in Cai and Sun (2003); and the smoothing splines approach in Zucker and Karr (1990). Other complementary works include O'Sullivan (1993) and Gray (1992, 1994) for the generalized Cox regression model with time-varying regression coefficients.

In this paper, we focus on the nonparametric smoothing approach proposed by Zucker and Karr (1990) (ZK90 for short). We think the penalized smoothing splines approach could enhance the robustness of the flexible estimates of time-dependent effects, especially in regions where there is little data. In ZK90, it is assumed that the hazard function at time $t \in [0, 1]$ for an individual with covariate z is given by

$$\lambda(t|z) = \lambda_0(t) \exp[\beta_0(t)z], \quad (1)$$

where $\beta_0(t)$ is an unknown function taking values in $\mathbf{R}$. In addition, one takes the maximum penalized partial likelihood estimator of $\beta_0(t)$ as the maximizer of $l_F(\beta|\text{data}) - \alpha J(\beta)$, where l_F is the negative log partial likelihood in the Cox model based on the data from entire cohort; J is a quadratic functional of the form $J \equiv [\beta, \beta] = \int [\beta^{(m)}]^2 (m \geq 1)$ which measures the roughness of β ; and $\alpha \equiv \alpha_n$ is a sequence of positive numbers depending on n , controlling the tradeoff between the smoothness and the goodness-of-fit of the estimate. Asymptotic properties of the maximum penalized partial likelihood estimator are presented in ZK90.

If $\beta_0(t) \equiv \beta_0$ is a constant parameter in (1), the estimation of β_0 based on counting processes and martingale theory has been widely discussed in the literature; see for example, Andersen and Gill (1982) and Fleming and Harrington (1991). Estimation might require the collection of information from all cohort members, generally impractical for many epidemiological studies and prevention trials, where questions relevant to public health require the follow-up of thousands of subjects over many years. There has been much interest in developing statistical methods to tackle this problem. Prentice (1986) proposed a case-cohort design, where a simple random sample is selected at time $t = 0$ and the estimation of β_0 is based only on the data from this random sample and failures outside this subcohort. The asymptotic properties of the estimator based on this case-cohort design are discussed in Self and Prentice (1988) (SP88 for short). Chen and Lo (1999) (CL99 for short) improved Prentice's estimator by incorporating information from all cases rather than only those cases included in the random sample. It has been known that the CL99 estimator performs better than the SP88 estimator. When the failure rate is small, both estimators generally perform well if the sampled risk set is of an appropriate size (Zhang and Goldstein (2003)).

In this paper, we present a maximum penalized pseudolikelihood approach under a case-cohort design as a generalization of ZK90's. More specifically, if l_{cc} denotes the negative log pseudolikelihood based on SP88's case-cohort sample data, then the maximum penalized pseudolikelihood estimator (MPPLE) of $\beta_0(t)$ is the maximizer of $l_{cc}(\beta|\text{data}) - \alpha J(\beta)$. One of the objectives of this paper is to investigate the asymptotic properties of MPPLE. In Section 2, after the introduction of notation and assumptions, we show the existence and weak uniform consistency of the estimator under regularity conditions, as well as asymptotic normality. It should be noted that under a "very smooth" assumption on $\beta_0(t)$ (see the remark in Section 2), both consistency and asymptotic normality hold for the most popular cubic smoothing spline estimate.

In Section 3, the extension of the approach in Section 2 to CL99 is considered, and the corresponding asymptotic properties of the proposed estimator are presented. Some discussion is given in Section 4 and detailed proofs are deferred to Section 5.

2. Penalized Estimator under the SP88 Approach

In this section, we investigate the estimation of the regression function $\beta_0(t)$ based on a case-cohort sampling design. To start with, we present the counting processes framework associated with the Cox regression model (Andersen and Gill (1982)). Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and $\{\mathcal{F}_t\}_{t \in [0,1]}$ a right continuous, nondecreasing family of sub- σ -algebras of $\mathcal{F}$, with $\mathcal{F}_0$ containing all P -null subsets of $\mathcal{F}$. We suppose that $\{\mathcal{F}_t\}$ includes failure time and covariate histories up to time t , and censoring histories to t^+ for all subjects in a cohort $\mathcal{R} = \{1, 2, \dots, n\}$. For the i^{th} individual, $i \in \mathcal{R}$, we associate the triplet $(N_i(t), Y_i(t), Z_i)$, these are independent replicates of $(N(t), Y(t), Z)$, where $N_i(t) = \sum_{j \geq 1} 1(t_j \leq t, j = i)$ is the counting process of failure times in $(0, t]$, Z_i is the time independent covariate, and $Y_i(t)$ is the censoring process: 1 if the i th individual is "at risk" for observable failure time and 0 otherwise. We also assume $N_i(1) < \infty$ *a.s.*, for every i . Note that N_i can only jump when $Y_i(t) = 1$.

Corresponding to each counting process $N_i(t)$, the intensity process

$$\lambda_i(t) = Y_i(t)\lambda_0(t)\exp(\beta_0(t)Z_i), \quad (2)$$

determines the rate at which individual i is observed to fail at time t , given the cohort history $\mathcal{F}_{t-}$ up to time t . Then $M_i(t) = N_i(t) - \int_0^t \lambda_i(u)du$ is a martingale (Andersen and Gill (1982)).

Let $\tilde{\mathcal{R}}$ of size $\tilde{n}$ be the simple random sample selected at $t = 0$ from the cohort $\mathcal{R}$. The Cox log pseudolikelihood under a case-cohort sampling design,

up to a constant $(1/n)$, is given by

$$l_{cc}(\beta) = \frac{1}{n} \sum_{i=1}^n D_i[\beta(T_i)Z_i - \log(\sum_{j \in \tilde{\mathcal{R}}} Y_j(T_i)e^{\beta(T_j)Z_j})].$$

The MPPLE, denoted as $\hat{\beta}_{cc}(t)$, is the minimizer of,

$$H_{cc}(\beta) = \frac{1}{2}\alpha[\beta, \beta] + \frac{1}{n} \sum_{i=1}^n \int_0^1 Y_i(u) (\log \frac{\tilde{S}^{(0)}(\beta, u)}{\tilde{S}^{(0)}(\beta_0, u)} - Z_i[\beta(u) - \beta_0(u)]) dN_i(u). \quad (3)$$

Here $[f, g] = \int_0^1 f^{(m)}(t)g^{(m)}(t)dt$ for f and g belonging to the Sobolev space $H^m = H^m[0, 1]$ of piecewise m -times ($m \geq 1$) differentiable functions f with $[f, f] < \infty$. The remainder of this section is organized as follows. We first introduce basic notation, then assert the existence and consistency of $\hat{\beta}_{cc}(t)$, and its asymptotic normality.

Notation. For $x \in \mathbf{R}$,

$$\begin{aligned} S^{(j)}(x; s) &= \frac{1}{n} \sum_{i=1}^n Y_i(s) Z_i^j e^{xZ_i}, j = 0, \dots, 3, & A(x; s) &= \frac{S^{(1)}(x; s)}{S^{(0)}(x; s)}, \\ V(x; s) &= \frac{S^{(2)}(x; s)}{S^{(0)}(x; s)} - A(x; s)^2, \\ C(x; s) &= \frac{\frac{1}{n} \sum_{i=1}^n Y_i(s) [Z_i - A(x; s)]^3 e^{xZ_i}}{S^{(0)}(x; s)}, \\ s^{(j)}(x; s) &= E[Y(s) Z^j e^{xZ}], j = 0, \dots, 3, & a(x; s) &= \frac{s^{(1)}(x; s)}{s^{(0)}(x; s)}, \\ v(x; s) &= \frac{s^{(2)}(x; s)}{s^{(0)}(x; s)} - a(x; s)^2, \\ w(s) &= \lambda_0(s) s^{(0)}(\beta_0(s); s) v(\beta_0(s); s). \end{aligned}$$

For a case-cohort design, we set

$$\begin{aligned} \tilde{S}^{(j)}(x; s) &= \frac{1}{\tilde{n}} \sum_{i \in \tilde{\mathcal{R}}} Y_i(s) Z_i^j e^{xZ_i}, j = 0, \dots, 3, \\ \tilde{A}(x; s) &= \frac{\tilde{S}^{(1)}(x; s)}{\tilde{S}^{(0)}(x; s)}, \quad \tilde{V}(x; s) = \frac{\tilde{S}^{(2)}(x; s)}{\tilde{S}^{(0)}(x; s)} - \tilde{A}(x; s)^2. \end{aligned}$$

For notational simplicity we write, for example, $S^{(j)}(\beta_0, s) \equiv S^{(j)}(\beta_0(s), s)$. Similar abbreviations for the other quantities are applied throughout the paper.

Assumptions. *The following assumptions are in force for the remainder of the section.*

Assumption 2.1. The parameters α_n are deterministic.

Assumption 2.2. $w(s) \geq w_0$ for some positive constant w_0 .

Assumption 2.3. $\beta_0 \in H^m$.

Assumption 2.4. $w(s)$ is $(2m - 1)$ times continuously differentiable on $[0, 1]$.

Assumption 2.5. The distributions of T^0 and V^0 are absolutely continuous. Furthermore for simplicity, we assume Z is a time-independent random variable, $0 \leq Z \leq 1$. Consequently, $0 \leq A(x, s) \leq 1$, $0 \leq V(x, s) \leq 1$, and $|C(x, s)| \leq 1$.

In addition to the requirements listed above from ZK90, we need the following assumptions for a case-cohort design.

Assumption 2.6. $\tilde{n}/n \rightarrow_P \gamma > 0$.

Let $D[0, 1]$ denote the space of right-continuous functions on $[0, 1]$ with left-hand limits functions. The following weak convergence of stochastic processes taking values in $D[0, 1]$ is defined in terms of the Skorohod topology (Billingsley (1968)).

Assumption 2.7. (Tightness condition) $n^{1/2}(A(\beta_0, s) - \tilde{A}(\beta_0, s))$ is tight on $D[0, 1]$.

Assumption 2.7 is required in SP88. It guarantees weak convergence in the later development of asymptotic properties.

Assumption 2.8. For any fixed $s, t \in [0, 1]$, $B(s, t) \equiv E[(Z - a(\beta_0, s))(Z - a(\beta_0, t))Y(s)Y(t)e^{\beta_0(s)Z + \beta_0(t)Z}] \lambda_0(s)\lambda_0(t) \geq B_0$, for some positive constant B_0 .

Assumption 2.8 is required for the development of asymptotic normality. In the case when Z is a binary covariate with $P(Z = 1) = \pi_0 > 0$ and $P(Z = 0) = 1 - \pi_0 > 0$, Assumption 2.8 holds automatically, see Section 4 for details.

Finally, we define

$$\langle f, g \rangle_w = \int_0^1 f(s)g(s)w(s)ds, \quad \langle f, g \rangle_{H^m} = \langle f, g \rangle_w + [f, g].$$

for $f, g \in H^m$. From the assumption that $0 < w_0 \leq w(s) \leq \|w\|_\infty$ and Sobolev space theory, as given in Silverman (1982), there exist functions $(\phi_\nu)_{\nu=0}^\infty$ in H^m and numbers $1 = \mu_0 \geq \mu_1 \geq \mu_2 \geq \dots \geq 0$ such that the sequence $(\phi_\nu)_{\nu=0}^\infty$ is an orthonormal basis for $L^2[0, 1]$ under the inner product $\langle f, g \rangle_w$ and $(\mu_\nu^{1/2}\phi_\nu)$ is an orthonormal basis for H^m under the inner product $\langle \cdot, \cdot \rangle_{H^m}$. In particular, with $\rho_\nu = \mu_\nu^{-1} - 1$,

$$\langle \phi_\nu, \phi_\eta \rangle_w = \delta_{\nu\eta}, \quad \langle \phi_\nu, \phi_\eta \rangle_{H^m} = \mu_\nu^{-1}\delta_{\nu\eta}, \quad [\phi_\nu, \phi_\eta] = \rho_\nu\delta_{\nu\eta}. \quad (4)$$

Existence. The existence and characterization of the MPPL $\hat{\beta}_{cc}(t)$ follows the same lines as those given in Theorem 1 of ZK90, except that the inequality (4.1) in ZK90 is replaced by $\int_0^1 \tilde{V}(\beta, s) h^2(s) d\bar{N}(s) > 0$ with $\bar{N}(s) = (1/n) \sum_{i=1}^n N_i(s)$. It follows that the computation of $\hat{\beta}_{cc}(t)$ reduces to a finite dimensional maximization as discussed in ZK90.

Consistency.

Theorem 1. *Suppose that $m \geq 3$ and that $\alpha \rightarrow 0$, with $n^{-1}\alpha^{-(4+\varepsilon)/2m} \rightarrow 0$ as $n \rightarrow \infty$ for sufficiently small $\varepsilon > 0$. Then $\|\hat{\beta}_{cc} - \beta_0\|_\infty \rightarrow_P 0$.*

The proof of the consistency of $\hat{\beta}_{cc}(t)$ is straightforward but tedious, paralleling the approach in ZK90. A sketch of the proof is given in Section 5.

Remark 1. It can be shown (Zhang and Huang (2004)) that we actually have

$$\begin{aligned} \|\hat{\beta}_{cc} - \beta_0\|_\infty &\leq O_P(n^{-1}\alpha^{-\frac{4+\varepsilon}{2m}}) + O_P(n^{-1}\alpha^{-\frac{7}{4m}}) + O_P(n^{-\frac{1}{2}}\alpha^{-\frac{2+\varepsilon}{4m}}) \\ &\quad + O_P(\alpha^{1-\frac{4+\varepsilon}{2m}}) + O_P(n^{-\frac{1}{2}}\alpha^{\frac{1}{2}-\frac{7}{4m}}) + O_P(\alpha^{\frac{1}{2}-\frac{2+\varepsilon}{4m}}) \\ &\quad + O_P(n^{-\frac{1}{2}}\alpha^{\frac{1}{2}-\frac{4+\varepsilon}{2m}}) + O_P(n^{-\frac{3}{4}}\alpha^{\frac{1}{4}-\frac{7}{4m}}) \\ &\quad + O_P(n^{-\frac{1}{4}}\alpha^{\frac{1}{4}-\frac{2+\varepsilon}{4m}}). \end{aligned} \quad (5)$$

Note that the first six terms are the same as those in ZK90 for a full cohort design (ZK90 missed the second term), the last three are due to the case-cohort sampling. Furthermore, the requirement of $m \geq 3$ is from the fourth term in order to obtain $\alpha^{1-(4+\varepsilon)/2m} \rightarrow 0$ as $\alpha \rightarrow 0$.

Remark 2. The requirement of $m \geq 3$ in Theorem 1 is only a sufficient condition. However, if we assume $\beta_0(t)$ is “very smooth,” i.e., $\beta_0(t) \in H^{2m}[0, 1]$, or equivalently $\sum_{\nu=0}^\infty \mu_\nu^{-2} b_{0\nu}^2 < \infty$ with $\beta_0(t) = \sum_{\nu=0}^\infty b_{0\nu} \phi_\nu(t)$, then the above conditions can be relaxed to $m \geq 1$ and $\alpha \rightarrow 0$ with $n^{-1}\alpha^{-\max\{(4+\varepsilon)/2m, (m+8)/6m\}} \rightarrow 0$. Note that the same assumption is also used by both Wahba (1977) and Silverman (1982).

Remark 3. The requirement of $m \geq 3$ for consistency in ZK90 can also be relaxed to $m \geq 1$ under the same “very smooth” assumption in a full cohort design.

Asymptotic Nomality. Let $\mathbf{X}(t) = (N(t), Y(t), Z)$ and, for fixed $s, t \in [0, 1]$,

$$\begin{aligned} R_\alpha(s, t) &= \sum_{\nu=0}^\infty \frac{1}{1 + \alpha \rho_\nu} \phi_\nu(s) \phi_\nu(t), & r_\alpha(s, t) &= \sum_{\nu=0}^\infty \frac{1}{(1 + \alpha \rho_\nu)^2} \phi_\nu(s) \phi_\nu(t), \\ f(\mathbf{X})(t) &= \int_0^1 [Z - a(\beta_0, s)] R_\alpha(s, t) Y(s) e^{\beta_0(s)Z} \lambda_0(s) ds. \end{aligned} \quad (6)$$

Theorem 2. Suppose that $\alpha \rightarrow 0$ and that

$$n^{-1}\alpha^{-\frac{7+2\varepsilon}{2m}} \rightarrow 0, \quad n\alpha^{1-\frac{1+\varepsilon}{2m}} \rightarrow 0. \quad (7)$$

as $n \rightarrow \infty$ for sufficient small $\varepsilon > 0$. Then, as $n \rightarrow \infty$, we have for each fixed $t \in [0, 1]$,

$$\frac{\hat{\beta}_{cc}(t) - \beta_0(t)}{\sqrt{\frac{\sigma_n^2(t)}{n} + \frac{1-\gamma}{\gamma} \frac{\text{Var}[f(\mathbf{X})](t)}{n}}} \rightarrow_d N(0, 1),$$

where $\sigma_n^2(t) = r_\alpha(t, t)$, and $\text{Var}[f(\mathbf{X})](t) = \int_0^1 \int_0^1 B(s, \eta) R_\alpha(s, t) R_\alpha(\eta, t) ds d\eta$.

The proof of Theorem 2 is given in Section 5.

Remark 1. The result reveals that the asymptotic variance of $\hat{\beta}_{cc}(t)$ can be decomposed into two terms, with $\sigma_n^2(t)/n$ being the asymptotic variance of the maximum penalized partial likelihood estimator based on the full cohort and $((1-\gamma)/\gamma)(\text{Var}[f(\mathbf{X})](t)/n)$ due to the case-cohort sampling design. In addition, it can be shown that $\text{Var}[f(\mathbf{X})](t)$ has the same convergence rate as $\sigma_n^2(t)$ under Assumption 2.8. It is then easy to provide an estimate of the asymptotic variance of $\hat{\beta}_{cc}(t)$ and thus to make inference for $\beta_0(t)$, following the discussion in ZK90, Section 9.

Remark 2. Theorem 2 actually requires $m \geq 5$. Note that in (7), $n\alpha^{1-(1+\varepsilon)/2m} = n\alpha^{(7+2\varepsilon)/2m} \cdot \alpha^{1-4/m-3\varepsilon/2m}$. The condition (7) holds only if $\alpha^{1-4/m-3\varepsilon/2m} \rightarrow 0$, or equivalently, $1 - 4/m - 3\varepsilon/2m > 0$, which necessitates $m \geq 5$.

Remark 3. Our requirements on m for consistency and asymptotic normality agree with those in ZK90. (There is an apparent mistake in Theorem 3 of ZK90, where they actually require $m \geq 5$ for their inequality (7.10) to hold.)

Remark 4. It should be noted that the requirement of $m \geq 5$ for asymptotic normality is only a sufficient condition. Under the assumption of a “very smooth” $\beta_0(t)$, the condition for asymptotic normality of $\hat{\beta}_{cc}(t)$ can be relaxed to $\alpha \rightarrow 0$ with $n^{-1}\alpha^{-(7+2\varepsilon)/2m} \rightarrow 0$ and $n\alpha^{4-(5+\varepsilon)/2m} \rightarrow 0$, so $m \geq 2$ suffices.

Remark 5. Note that if $\gamma = 1$, Theorem 2 recovers the asymptotic normality of the maximum penalized partial likelihood estimator of ZK90. Therefore, the requirement of $m \geq 5$ for asymptotic normality in ZK90 can also be relaxed to $m \geq 2$ using the “very smooth” assumption under a full cohort design.

3. Penalized Estimator under the CL99 Approach

In this section, we apply the same techniques as those used in Section 2 to derive the asymptotic properties of the penalized estimator mimicking the CL99

approach. Recall that $\mathcal{R}$ and $\tilde{\mathcal{R}}$ denote the sets of all individuals in the cohort of size of n and subcohort of size of $\tilde{n}$, respectively. Following the notation from CL99, we let $n_1(\tilde{n}_1)$ and $n_0(\tilde{n}_0)$ be the numbers of the cases and controls in the cohort (subcohort), respectively. We further let $\mathcal{R}^1, \tilde{\mathcal{R}}^1$ and $\mathcal{R}^0, \tilde{\mathcal{R}}^0$ denote, respectively, the index sets of all cases and all controls in $\mathcal{R}, \tilde{\mathcal{R}}$.

First we introduce

$$\begin{aligned}\check{S}^{(p)}(x; s) &= \frac{1}{n} \sum_{j \in \mathcal{R}^1} Z_j^p Y_j(w) e^{\beta Z_j} + \frac{n_0}{n \tilde{n}_0} \sum_{j \in \tilde{\mathcal{R}}^0} Z_j^p Y_j(w) e^{\beta Z_j}, p = 0, \dots, 3, \\ \check{A}(x; s) &= \frac{\check{S}^{(1)}(x; s)}{\check{S}^{(0)}(x; s)}, \quad \check{V}(x; s) = \frac{\check{S}^{(2)}(x; s)}{\check{S}^{(0)}(x; s)} - \check{A}(x; s)^2.\end{aligned}$$

Recall from Section 1 that $D \equiv I(T = T^0)$ is the indicator of an observed failure. We assume that $\pi = P(D = 1)$ with $0 < \pi < 1$. In addition to Assumptions 2.1–2.6, we need the following.

Assumption 3.7. (Tightness condition) $n^{1/2}(A(\beta_0, s) - \check{A}(\beta_0, s))$ is tight on $D[0, 1]$.

Assumption 3.8. For any fixed $s, t \in [0, 1]$, $\tilde{B}(s, t) \equiv E[(Z - a(\beta_0, s))(Z - a(\beta_0, t))Y(s)Y(t)e^{\beta_0(s)Z + \beta_0(t)Z} | D = 0]\lambda_0(s)\lambda_0(t) \geq \tilde{B}_0$, for some constant $\tilde{B}_0 > 0$.

Assumption 3.9. $n_1/n \rightarrow_P \pi > 0$.

Assumption 3.7 parallels Assumption 2.7 to ensure the weak convergence of random sequence $n^{1/2}(A(\beta_0, s) - \check{A}(\beta_0, s))$, and Assumption 3.8 parallels Assumption 2.8 for asymptotic normality of $\hat{\beta}_{\text{CL}}(t)$.

The existence of $\hat{\beta}_{\text{CL}}(t)$ follows as and the existence of $\hat{\beta}_{\text{CC}}(t)$ in Section 2, with $\tilde{V}(\beta, s)$ replaced by $\check{V}(\beta, s)$. For consistency and asymptotic normality, we have the following.

Theorem 3.

1. Suppose that $m \geq 3$, and that $\alpha \rightarrow 0$ with $n^{-1}\alpha^{-(4+\varepsilon)/2m} \rightarrow 0$ as $n \rightarrow \infty$ for sufficiently small $\varepsilon > 0$. Then $\|\hat{\beta}_{\text{CL}}(t) - \beta_0\|_\infty \rightarrow_P 0$.
2. Suppose that $\alpha \rightarrow 0$ with $n^{-1}\alpha^{-(7+2\varepsilon)/2m} \rightarrow 0$ and $n\alpha^{1-(1+\varepsilon)/2m} \rightarrow 0$ as $n \rightarrow \infty$ for sufficiently small $\varepsilon > 0$. Then, as $n \rightarrow \infty$, for each fixed $t \in [0, 1]$, we have $(\hat{\beta}_{\text{CL}}(t) - \beta_0(t))/\sqrt{\sigma_n^2(t)/n + [(1-\gamma)/\gamma](1-\pi)[E_0(t)/n]} \rightarrow_d N(0, 1)$, where, for each $t \in [0, 1]$, $E_0(t) = \int_0^1 \int_0^1 E[(Z - a(\beta_0, s))(Z - a(\beta_0, u))Y(s)Y(u)e^{\beta_0(s)Z + \beta_0(u)Z} | D = 0]R_\alpha(s, t)R_\alpha(u, t)\lambda_0(s)\lambda_0(u)dsdu$.

The proof of Theorem 3 is given in Section 5. Note that the remarks after Theorems 1 and 2 in Section 2 hold here as well.

4. Discussion

Several extensions of model (1) in ZK90 can be carried over from the full cohort case to a case-cohort design. These extensions include time-dependent covariates and the general relative risk function $r(\beta_0 z)$ as specified by SP88. However additional assumptions, such as regularity conditions mentioned in SP88 when the exponential relative risk function $e^{\beta_0 Z}$ is replaced by $r(\beta_0 Z)$, are necessary. Note that to extend our results to the multivariate case, we face the same difficulty as was pointed out by ZK90.

The MPPL presented here offers flexibility in exploring time-dependent covariate effects. To apply this method in practice, however, one needs to compute the resultant estimator efficiently. Hastie and Tibshirani (1993) develop an iterative algorithm to obtain the maximum penalized partial likelihood estimator of ZK90 through the computation of weighted smoothing splines. The same algorithm can clearly be applied to the computation of MPPL. In addition, various data driven smoothing parameter selection procedures, such as cross validation and generalized cross validation in O'Sullivan (1988) and Gu (2002), can be incorporated with the algorithm.

Now we visit Assumption 2.8 by considering a special case where Z is a binary covariate with $P(Z = 0) = 1 - \pi_0$ and $P(Z = 1) = \pi_0$, $0 < \pi_0 < 1$. Let $\bar{G}(t|z)$ be the survival function of continuous censored time V^0 when $Z = z$. Under the conditional independence of V^0 and T^0 given Z we have that, without loss of generality, for $s \leq t$,

$$\begin{aligned} w(t) &= \pi_0(1 - \pi_0) \frac{\bar{G}(t|1)e^{-\Lambda(t)}e^{\beta_0(t) - \tilde{\Lambda}(t)}}{\pi_0\bar{G}(t|1)e^{\beta_0(t) - \tilde{\Lambda}(t)} + (1 - \pi_0)\bar{G}(t|0)e^{-\Lambda(t)}} \lambda_0(t), \\ B(s, t) &= \pi_0(1 - \pi_0)\lambda_0(t) \frac{\bar{G}(t|1)e^{-\Lambda(t)}e^{\beta_0(t) - \tilde{\Lambda}(t)}}{\pi_0\bar{G}(t|1)e^{\beta_0(t) - \tilde{\Lambda}(t)} + (1 - \pi_0)\bar{G}(t|0)e^{-\Lambda(t)}} \\ &\quad \times \frac{(1 - \pi_0)\bar{G}(s|0)e^{\beta_0(s) - \Lambda(s)} + \pi_0\bar{G}(s|1)e^{\beta_0(s) - \tilde{\Lambda}(s)}}{\pi_0\bar{G}(s|1)e^{\beta_0(s) - \tilde{\Lambda}(s)} + (1 - \pi_0)\bar{G}(s|0)e^{-\Lambda(s)}} \lambda_0(s) \\ &\equiv w(t) \times \tilde{B}(s), \end{aligned}$$

say. Here $\tilde{\Lambda}(t) = \int_0^t \lambda_0(s)e^{\beta_0(s)}ds$ and $\Lambda(t) = \int_0^t \lambda_0(s)ds$. Let $d_0 = \min_{0 \leq s \leq 1} \beta_0(s) > -\infty$. Noticing that the assumption of $w(t) \geq w_0$ implies that $\lambda_0(t) \geq c_0$ for some positive constant $c_0 > 0$, one can easily see that $\tilde{B}(s) \geq C^* > 0$, with $C^* = c_0$ when $d_0 \geq 0$, $C^* = c_0e^{d_0}$ otherwise. In summary, if Z is a binary covariate, we have for some positive $C^* > 0$, $B(s, t) \geq C^*w(\max\{s, t\}) \geq C^*w_0 > 0$, which satisfies Assumption 2.8. The more general case might be more complicated and requires further investigation.

Appendix

A sketch of the proof of Theorem 1. We let

$$\begin{aligned}\tilde{H}_1(\beta) &= \frac{1}{2}\alpha[\beta, \beta] + \frac{1}{2} \int_0^1 w(s)[\beta(s) - \beta_0(s)]^2 ds \\ &\quad - \frac{1}{n} \sum_{i=1}^n \int_0^1 [Z_i - \tilde{A}(\beta_0, s)][\beta(s) - \beta_0(s)] dN_i(s)\end{aligned}$$

be an approximation of $H_{cc}(t)$ based on a two-term Taylor series expansion, patterned after ZK90. Let $\tilde{\beta}_1$ be the minimizer of $\tilde{H}_1(\beta)$. With (4), if we take (b_ν) and $(b_{0\nu})$ as the coefficients in the expansions of $\beta(t)$ and $\beta_0(t)$ in H^m , respectively, i.e., $\beta(t) = \sum_{\nu=0}^\infty b_\nu \phi_\nu(t)$, $\beta_0(t) = \sum_{\nu=0}^\infty b_{0\nu} \phi_\nu(t)$, the minimizer $\tilde{\beta}_1(t)$ is given by $\tilde{\beta}_1(t) = \sum_{\nu=0}^\infty \tilde{b}_{1\nu} \phi_\nu(t) \equiv \sum_{\nu=0}^\infty [(\tilde{X}_\nu + b_{0\nu})/(1 + \alpha\rho_\nu)] \phi_\nu(t)$, where $\tilde{X}_\nu = (1/n) \sum_{i=1}^n \int_0^1 [Z_i - \tilde{A}(\beta_0, s)] \phi_\nu(s) dN_i(s)$. Our proof then consists of the following steps.

Step 1. Show that $\tilde{\beta}_1$ converges to β_0 as $n \rightarrow \infty$. Specifically, that for a suitable choice of α , $E\|\tilde{\beta}_1 - \beta_0\|_\infty^2 \rightarrow 0$.

Step 2. Show that for n sufficiently large, $\tilde{\beta}_1$ is close to $\hat{\beta}_{cc}(t)$, i.e., for the suitable choice of α ,

$$\|\hat{\beta}_{cc} - \tilde{\beta}_1\|_{H^1} \rightarrow_P 0. \quad (8)$$

For more details, see Zhang and Huang (2004).

To prove Theorem 2, we first write

$$\begin{aligned}\hat{\beta}_{cc}(t) - \beta_0(t) &= (\hat{\beta}_{cc}(t) - \tilde{\beta}_1(t)) + (\tilde{\beta}_1(t) - \hat{\beta}_1(t)) + (\hat{\beta}_1(t) - \beta^*(t) + (\beta_\alpha(t) - \beta_0(t))) \\ &\quad + U(t) \\ &\equiv I + II + III + IV + U(t), \quad \text{where,} \\ X_\nu &= \frac{1}{n} \sum_{i=1}^n \int_0^1 [Z_i - A(\beta_0, s)] \phi_\nu(s) dN_i(s), \quad \hat{\beta}_1(t) = \sum_{\nu=0}^\infty \frac{X_\nu + b_{0\nu}}{1 + \alpha\rho_\nu} \phi_\nu(t), \\ X_\nu^* &= \frac{1}{n} \sum_{i=1}^n \int_0^1 [Z_i - a(\beta_0, s)] \phi_\nu(s) dM_i(s), \quad \beta^*(t) = \sum_{\nu=0}^\infty \frac{X_\nu^* + b_{0\nu}}{1 + \alpha\rho_\nu} \phi_\nu(t), \\ \beta_\alpha(t) &= \sum_{\nu=0}^\infty \frac{b_{0\nu}}{1 + \alpha\rho_\nu} \phi_\nu(t), \quad U(t) = \sum_{\nu=0}^\infty \frac{X_\nu^*}{1 + \alpha\rho_\nu} \phi_\nu(t).\end{aligned}$$

This decomposition is analogous to (7.2) of ZK90, with the extra breakdown of the first two terms due to the finite sampling of the case-cohort. It follows from (8) that term I converges to zero in probability if $m \geq 1$, for a suitable choice

of α . For terms III and IV, Lemmas 5 and 6 in ZK90 guarantee convergence to zero in probability. In addition one has, from ZK90 that if $n^{1/2}\alpha^{1/4m} \rightarrow \infty$, $U(t)/\sqrt{\text{Var}(U(t))} \rightarrow_d N(0, 1)$ as $n \rightarrow \infty$. Here $\text{Var}(U(t)) = \sigma_n^2(t)/n$ and $C_1\alpha^{-1/2m} \leq \sigma_n^2(t) \leq C_2\alpha^{-1/2m}$ for some $C_1, C_2 > 0$ (depending on t).

For term II, we first observe that $\tilde{X}_\nu(t) - X_\nu(t) = (1/n) \sum_{i=1}^n \int_0^1 [A(\beta_0, s) - \tilde{A}(\beta_0, s)] \lambda_i(s) \phi_\nu(s) dN_i(s)$. Therefore,

$$\begin{aligned} \tilde{\beta}_1(t) - \hat{\beta}_1(t) &= \sum_{\nu=0}^{\infty} \frac{\tilde{X}_\nu - X_\nu}{1 + \alpha \rho_\nu} \phi_\nu(t) \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^1 [A(\beta_0, s) - \tilde{A}(\beta_0, s)] R_\alpha(s, t) dM_i(s) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \int_0^1 [A(\beta_0, s) - \tilde{A}(\beta_0, s)] R_\alpha(s, t) \lambda_i(s) ds \\ &\equiv II_1 + II_2, \end{aligned}$$

say. For each fixed t , let $\sqrt{n/\sigma_n^2} \zeta(u, t) = [1/(\sqrt{n}\sigma_n)] \sum_{i=1}^n \int_0^u [A(\beta_0, s) - \tilde{A}(\beta_0, s)] R_\alpha(s, t) dM_i(s)$, which is a local square integrable martingale with $\xi(\cdot, t) = \langle \zeta(\cdot, t), \zeta(\cdot, t) \rangle = [1/(n\sigma_n^2)] \sum_{i=1}^n \int_0^1 [A(\beta_0, s) - \tilde{A}(\beta_0, s)]^2 R_\alpha^2(s, t) \lambda_i(s) ds$. Note that $\sup_s |A(\beta_0, s) - \tilde{A}(\beta_0, s)| = o_P(1)$, and that $(1/n) \sum_{i=1}^n \lambda_i(s) \rightarrow_{a.s.} s^{(0)}(\beta_0, s) \lambda_0(s)$. Hence, for some constant $C^* > 0$ and each $t \in [0, 1]$,

$$\begin{aligned} \xi(1, t) &\cong \frac{o_P(1)}{\sigma_n^2} \int_0^1 R_\alpha^2(s, t) s^{(0)}(s) \lambda_0(s) ds \\ &\leq o_P(1) \int_0^1 \frac{R_\alpha^2(s, t) w(s) ds}{\sigma_n^2 \inf_s v(\beta_0, s)} \leq C^* o_P(1) \rightarrow_P 0 \end{aligned}$$

if $n \rightarrow \infty$, $\sqrt{n/\sigma_n^2} II_1 \rightarrow_P 0$. Here the asymptotic equivalence $\cong$ is in the sense of almost sure convergence. Now let $U_{cc}(t) = II_2 + U(t)$. We need only find the asymptotic normality of $U_{cc}(t)$. II_2 involves the difference between subcohort and full-cohort means. Note that Proposition 1 in SP88 is not applicable in our situation, we need the following proposition to show that II_2 and $U(t)$ converge jointly to independent Gaussian random variables.

Proposition 1. *Let $\mathbf{X}_n = (X_{1n}, \dots, X_{nn})$ and $\delta_n = (\delta_{1n}, \dots, \delta_{nn})$ be independent random sequences satisfies the following.*

1. δ_n is a vector of $\tilde{n}$ ones and $n - \tilde{n}$ zeros, each possible configuration of zeros and ones is equally likely and $\tilde{n}n^{-1} \rightarrow \gamma \in (0, 1)$.
2. For some scalar functions $f_{in}(\mathbf{X}_n)$ of $\mathbf{X}_n$, $S_{f_n}^2(\mathbf{X}_n) - c_n^2 \rightarrow_P 0$ for some positive sequence $c_n \rightarrow \sigma_p^2$, with $0 < \sigma_p^2 \leq \infty$ as $n \rightarrow \infty$. Here $S_{f_n}^2 = n^{-1} \sum_{i=1}^n [f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]^2$.

3. For any $\varepsilon > 0$,

$$n^{-1} \sum_{i=1}^n \frac{[f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]^2 I_{\{|f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)| > n^{\frac{1}{2}} \varepsilon c_n\}}}{c_n^2} \rightarrow_P 0,$$

where $f_{\cdot n}(\mathbf{X}_n) = n^{-1} \sum_{i=1}^n f_{in}(\mathbf{X}_n)$.

4. The scalar functions $g_n(\mathbf{X}_n)$ of $\mathbf{X}_n$ converge in distribution to a Gaussian random variable with mean zero and variance 1.

Then for $h_n(\mathbf{X}_n, \delta_n) = \sqrt{n/c_n^2} [\tilde{n}^{-1} \sum_{i=1}^n \delta_{in} f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]$, we have $(g_n(\mathbf{X}_n), h_n(\mathbf{X}_n, \delta_n))$ converges in distribution to a bivariate normal random variable with mean zero and covariance matrix given by $\begin{pmatrix} 1 & 0 \\ 0 & (1 - \gamma)/\gamma \end{pmatrix}$.

Proof. The result when $0 < \sigma_p^2 < \infty$ holds, as given by Proposition 1 in SP88. We focus on the case when $\sigma_p^2 = \infty$. From Conditions 2 and 3, it is sufficient to prove that the conclusion holds for $(g_n(\mathbf{X}_n), \tilde{h}_n(\mathbf{X}_n, \delta_n))$, where $\tilde{h}_n(\mathbf{X}_n, \delta_n) = \sqrt{n/S_{f_n}^2(\mathbf{X}_n)} [\tilde{n}^{-1} \sum_{i=1}^n \delta_{in} f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]$, if for any $\varepsilon > 0$, $n^{-1} \sum_{i=1}^n [f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]^2 I_{\{|f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)| > n^{1/2} \varepsilon S_{f_n}\}} / S_{f_n}^2 \rightarrow_P 0$. Now the result above follows along the lines of the proof of Proposition 3 of Zhang and Goldstein (2003), with the modification of using Hájek's theorem on sampling without replacement from a finite population as stated in Cochran (1977, pp.39-40.)

Note that simple calculation gives

$$\begin{aligned} & n^{\frac{1}{2}} [A(\beta_0, s) - \tilde{A}(\beta_0, s)] \\ &= n^{\frac{1}{2}} [(\tilde{S}^{(0)}(\beta_0, s) - S^{(0)}(\beta_0, s))A(\beta_0, s) - (\tilde{S}^{(1)}(\beta_0, s) - S^{(1)}(\beta_0, s))\tilde{S}^{(0)}(\beta_0, s)^{-1}] \\ &= n^{\frac{1}{2}} [(\tilde{S}^{(0)}(\beta_0, s) - S^{(0)}(\beta_0, s))a(\beta_0, s) - (\tilde{S}^{(1)}(\beta_0, s) - S^{(1)}(\beta_0, s))s^{(0)}(\beta_0, s)^{-1}] \\ &\quad + O_P(n^{-1}). \end{aligned}$$

Hence, we can apply Proposition 1 with X_{in} representing $\{Y_i(u), N_i(u), Z_i; 0 \leq u \leq 1\}$. For fixed $t \in [0, 1]$,

$$\begin{aligned} U_{cc}(t) &= U(t) + \frac{1}{n} \sum_{i=1}^n \int_0^1 [A(\beta_0, s) - \tilde{A}(\beta_0, s)] R_\alpha(s, t) \lambda_i(s) ds \\ &\cong U(t) + [\tilde{n}^{-1} \sum_{i=1}^n \delta_{in} f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)] \\ &\quad + O_P(n^{-1}) \int_0^1 R_\alpha(s, t) s^{(0)}(\beta_0, s) \lambda_0(s) ds \\ &= U(t) + [\tilde{n}^{-1} \sum_{i=1}^n \delta_{in} f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)] + II_{2b}, \end{aligned} \tag{9}$$

where $\cong$ holds due to $(1/n) \sum_{i=1}^n \lambda_i(s) \rightarrow_{a.s.} s^{(0)}(\beta_0, s) \lambda_0(s)$ in the integration for II_2 . Here, for each fixed $t \in [0, 1]$,

$$\begin{aligned} f_{in}(\mathbf{X}_n)(t) &= \int_0^1 [Y_i(s) e^{\beta_0(s) Z_i} a(\beta_0, s) (s^{(0)}(\beta_0, s))^{-1} - Y_i(s) Z_i e^{\beta_0(s) Z_i} (s^{(0)}(\beta_0, s))^{-1}] \\ &\quad \times R_\alpha(s, t) \lambda_0(s) s^{(0)}(\beta_0, s) ds \\ &= \int_0^1 [Y_i(s) e^{\beta_0(s) Z_i} a(\beta_0, s) - Y_i(s) Z_i e^{\beta_0(s) Z_i}] R_\alpha(s, t) \lambda_0(s) ds. \end{aligned} \quad (10)$$

Now for II_{2b} , we have, for some constant $C^* > 0$,

$$\begin{aligned} \sqrt{\frac{n}{\sigma_n^2}} II_{2b} &= \sqrt{\frac{n}{\sigma_n^2}} O_P(n^{-1}) \int_0^1 R_\alpha(s, t) s^{(0)}(\beta_0, s) \lambda_0(s) ds \\ &\leq \frac{O_P(n^{-\frac{1}{2}})}{\sigma_n} \sup_{s, t} |R_\alpha(s, t)| \int_0^1 s^{(0)}(\beta_0, s) \lambda_0(s) ds \\ &\leq C^* O_P(n^{-\frac{1}{2}} \alpha^{-\frac{1}{4m}}) \rightarrow_P 0 \end{aligned}$$

as $n\alpha^{1/2m} \rightarrow \infty$. Therefore, it is sufficient to consider the convergence of the first two terms in the expression for $U_{cc}(t)$ in (9).

Proposition 2. (i) If $n\alpha^{3/2m} \rightarrow \infty$ as $n \rightarrow \infty$, $S_{f_n}^2(\mathbf{X}_n) - c_n^2 \rightarrow_P 0$ with $c_n^2 = \text{Var}[f(\mathbf{X})](t)$ and $C_1 \sigma_n^2(t) \leq \text{Var}[f(\mathbf{X})](t) \leq C_2 \sigma_n^2(t)$ for some constants $C_1, C_2 > 0$ (depending on t). (ii) If $n\alpha^{1/2m} \rightarrow \infty$ as $n \rightarrow \infty$, $n^{-1} \sum_{i=1}^n \{[f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]^2 / c_n^2\} I_{\{|f_{in}(\mathbf{X}) - f_{\cdot n}(\mathbf{X})| > n^{1/2} \varepsilon_{cn}\}} \rightarrow_P 0$.

Proof. (i) Noting that $n^{-1}(\sum_{i=1}^n f_{in}(\mathbf{X}_n)) \rightarrow_{a.s.} E[f(\mathbf{X})] = 0$, it is then sufficient to prove that

$$n^{-1} \left(\sum_{i=1}^n f_{in}^2(\mathbf{X}_n) \right) - E[f(\mathbf{X})^2] \rightarrow_P 0, \quad (11)$$

For (11), note that for constants $C_1^*, C_2^*, C^* > 0$,

$$\begin{aligned} E[f(\mathbf{X})^2 I_{\{|f(\mathbf{X})|^2 > n\}}] &\leq \frac{E[|f(\mathbf{X})|^4]}{n} \\ &\leq \frac{1}{n} C_1^* \sup_s |\lambda_0^3(s)| \sup_{s, t} R_\alpha^2(s, t) \int_0^1 R_\alpha^2(s, t) s^{(0)}(s) v(\beta_0, s) \lambda_0(s) ds \\ &\leq \frac{C_2^*}{n} \alpha^{-\frac{2}{2m}} \int_0^1 R_\alpha^2(s, t) w(s) ds \leq C^* n^{-1} \alpha^{-\frac{3}{2m}} \rightarrow 0 \end{aligned}$$

if $n\alpha^{3/2m} \rightarrow \infty$. It remains to prove that $n^{-1}(\sum_{i=1}^n f_{in}^2(\mathbf{X}_n)) - E[f(\mathbf{X})^2] I_{\{|f(\mathbf{X})|^2 \leq n\}} \rightarrow_P 0$. In view of the Weak Law of Large Numbers, we need to show

that $\sum_{i=1}^n P(|f_{in}(\mathbf{X}_n)|^2 > n) \rightarrow 0$, and $n^{-2} \sum_{i=1}^n E[|f_{in}(\mathbf{X}_n)|^4 I_{\{|f_{in}(\mathbf{X}_n)|^2 \leq n\}}] \rightarrow 0$. However, one can easily see that both terms above are bounded by $E|f(\mathbf{X})|^4/n \rightarrow 0$ if $n\alpha^{3/2m} \rightarrow \infty$. The first part follows.

For the second part, we show that there exist positive constants $C_1, C_2 > 0$ (depending on t), such that

$$C_1 \sigma_n^2(t) \leq \text{Var}[f(\mathbf{X})](t) \leq C_2 \sigma_n^2(t). \quad (12)$$

Following the notation in Section 2, we let $K(s, \eta)$, $0 \leq s, \eta \leq 1$, be given as $K(s, t) = \sum_{\nu=0}^{\infty} \rho_{\nu} \phi_{\nu}(s) \phi_{\nu}(t)$. Obviously $K(s, \eta)$ is a reproducing kernel with eigenfunctions $\phi_{\nu}(s)$ and corresponding eigenvalues ρ_{ν} . Therefore, $K(s, \eta)$ is a continuous function on $[0, 1] \times [0, 1]$ and hence bounded by a positive constant $M > 0$. Then from Assumption 2.8 and the boundedness of $w(\cdot)$, we have, for some positive constant $C_1^* > 0$, that $B(s, \eta) \geq C_1^* K(s, \eta) w(s) w(\eta)$. Then, for some constant $C_1 > 0$,

$$\begin{aligned} \text{Var}[f(\mathbf{X})](t) &\geq C_1^* \int_0^1 \int_0^1 K(s, \eta) R_{\alpha}(s, t) R_{\alpha}(\eta, t) w(s) w(\eta) ds d\eta \\ &= C_1^* \sum_{n_1} \sum_{n_2} \frac{\phi_{n_1}(t) \phi_{n_2}(t)}{(1 + \alpha \rho_{n_1})(1 + \alpha \rho_{n_2})} \int_0^1 \int_0^1 K(s, \eta) w(s) w(\eta) \phi_{n_1}(s) \phi_{n_2}(\eta) ds d\eta \\ &= C_1^* \sum_{n_1} \sum_{n_2} \frac{\phi_{n_1}(t) \phi_{n_2}(t)}{(1 + \alpha \rho_{n_1})(1 + \alpha \rho_{n_2})} \sum_{\nu} \rho_{\nu} \left(\int_0^1 \phi_{n_1}(s) \phi_{\nu}(s) w(s) ds \right) \\ &\quad \times \left(\int_0^1 \phi_{n_2}(\eta) \phi_{\nu}(\eta) w(\eta) d\eta \right) \\ &= C_1^* \sum_{\nu} \frac{\rho_{\nu}}{(1 + \alpha \rho_{\nu})^2} \phi_{\nu}^2(t) \geq C_1 \sigma_n^2(t). \end{aligned}$$

Here $\sum_{\nu}$ (and the same for others) is the abbreviation for $\sum_{\nu=0}^{\infty}$. The last inequality holds in view of $\rho_{\nu} = C_{\nu} \nu^{-2m}$ for $c_1^* \leq C_{\nu} \leq c_2^*$, from Naimark (1967) or ZK90, with constants $c_1^* > 0$ and $c_2^* > 0$. The left inequality of (12) follows.

Now we consider the right inequality of (12). Note that $B(s, \eta)$ is the covariance function of random process $[Y(s)(Z - a(\beta_0, s))e^{\beta_0(s)Z} \lambda_0(s)]$. Therefore, it can be expanded in a uniformly convergent series of its eigenfunctions $\varphi_k(s)$ with corresponding eigenvalues ξ_k (for example, Yeh (1973, p.288)), $B(s, \eta) = \sum_k \xi_k \varphi_k(s) \varphi_k(\eta)$. Here $\{\varphi_k(s)\}$ is a complete orthonormal basis for $L^2[0, 1]$ under the inner product $\langle \cdot, \cdot \rangle_w$. Therefore, there exists $\zeta_i^{(k)}$, such that for each k , $\varphi_k(s) = \sum_i \zeta_i^{(k)} \phi_i(s)$ and $\sum_i (\zeta_i^{(k)})^2 = 1$. Hence, $B(s, \eta) = \sum_k \xi_k (\sum_i \zeta_i^{(k)} \phi_i(s))$

$(\sum_j \zeta_j^{(k)} \phi_j(\eta))$. Then,

$$\begin{aligned}
\text{Var}[f(\mathbf{X})](t) &= \int_0^1 \int_0^1 R(s, t) R(\eta, t) \sum_k \xi_k \sum_i \zeta_i^{(k)} \phi_i(s) \sum_j \zeta_j^{(k)} \phi_j(\eta) ds d\eta \\
&\leq \frac{1}{w_0^2} \sum_k \xi_k \sum_{n_1} \sum_i \frac{\zeta_i^{(k)}}{1 + \alpha \rho_{n_1}} \phi_{n_1}(t) \int_0^1 \phi_{n_1}(s) \phi_i(s) w(s) ds \\
&\quad \times \sum_{n_2} \sum_j \frac{\zeta_j^{(k)}}{1 + \alpha \rho_{n_2}} \phi_{n_2}(t) \int_0^1 \phi_{n_2}(\eta) \phi_j(\eta) w(\eta) d\eta \\
&= \frac{1}{w_0^2} \sum_k \xi_k \left(\sum_i \frac{\zeta_i^{(k)}}{1 + \alpha \rho_i} \phi_i(t) \right)^2 \\
&= \frac{1}{w_0^2} \sum_k \xi_k \left(\sum_j \frac{1}{(1 + \alpha \rho_j)^2} \phi_j^2(t) \right) \\
&= \sigma_n^2(t) \left(\sum_k \frac{\xi_k}{w_0^2} \right).
\end{aligned}$$

Note from $B(s, s) = \sum \xi_k \varphi_k^2(s)$, we have $\int_0^1 B(s, s) w(s) ds = \sum_k \xi_k \int_0^1 \varphi_k^2(s) w(s) ds = \sum_k \xi_k$. Therefore, $\text{Var}[f(\mathbf{X})](t) \leq \sigma_n^2(t) (\int_0^1 B(s, s) w(s) ds / w_0^2)$. Take $C_2 = \int_0^1 B(s, s) w(s) ds / w_0^2 > 0$ and complete the proof.

(ii). The proof requires extensive usage of the Weak Law of Large Numbers as in Durrett (1996, p.41). We take $b_n = n$ in the theorem throughout the rest of the proof.

From (i) we may assume, for simplicity, that $c_n = \sigma_n^2(t)$ in the proof. From the inequality (for example, see SP88, p.72), $|a - b|^2 I_{\{|a-b|>\varepsilon\}} \leq 4|a|^2 I_{\{|a|>\varepsilon/2\}} + 4|b|^2 I_{\{|b|>\varepsilon/2\}}$, it is sufficient that, for $\varepsilon > 0$,

$$n^{-1} \sum_{i=1}^n \frac{|f_{in}(\mathbf{X}_n)|^2}{\sigma_n^2} I_{\{|f_{in}(\mathbf{X}_n)| > n^{\frac{1}{2}} \varepsilon \sigma_n\}} \rightarrow_P 0, \quad (13)$$

$$n^{-1} \sum_{i=1}^n \frac{|f_{\cdot n}(\mathbf{X}_n)|^2}{\sigma_n^2} I_{\{|f_{\cdot n}(\mathbf{X}_n)| > n^{\frac{1}{2}} \varepsilon \sigma_n\}} = \frac{|f_{\cdot n}(\mathbf{X}_n)|^2}{\sigma_n^2} I_{\{|f_{\cdot n}(\mathbf{X}_n)| > n^{\frac{1}{2}} \varepsilon \sigma_n\}} \rightarrow_P 0. \quad (14)$$

To prove (13), note the *i.i.d.* nature of $f_{in}(\mathbf{X}_n)$ and the Weak Law of Large Numbers of Durrett (1996), with

$$a_n = \sum_{i=1}^n E \left[\frac{|f_{in}(\mathbf{X}_n)|^2}{\sigma_n^2} I_{\{|f_{in}(\mathbf{X}_n)| > n^{\frac{1}{2}} \varepsilon \sigma_n\}} I_{\left\{ \frac{|f_{in}(\mathbf{X}_n)|^2}{\sigma_n^2} I_{\{|f_{in}(\mathbf{X}_n)| > n^{\frac{1}{2}} \varepsilon \sigma_n\}} \leq n \right\}} \right].$$

It is sufficient to prove

$$\sum_{i=1}^n P\left(\frac{|f_{in}(\mathbf{X}_n)|^2}{\sigma_n^2} I_{\{|f_{in}(\mathbf{X}_n)| > n^{\frac{1}{2}} \varepsilon \sigma_n\}} > n\right) \rightarrow 0, \quad (15)$$

$$n^{-2} \sum_{i=1}^n E\left[\frac{|f_{in}(\mathbf{X}_n)|^4}{\sigma_n^4} I_{\{|f_{in}(\mathbf{X}_n)| > n^{\frac{1}{2}} \varepsilon \sigma_n\}} I_{\{|f_{in}(\mathbf{X}_n)| \leq n \sigma_n\}}\right] \rightarrow 0, \quad (16)$$

$$n^{-1} a_n = n^{-1} \sum_{i=1}^n E\left[\frac{|f_{in}(\mathbf{X}_n)|^2}{\sigma_n^2} I_{\{|f_{in}(\mathbf{X}_n)| > n^{\frac{1}{2}} \varepsilon \sigma_n\}} I_{\left\{\frac{|f_{in}(\mathbf{X}_n)|^2}{\sigma_n^2} I_{\{|f_{in}(\mathbf{X}_n)| > n^{\frac{1}{2}} \varepsilon \sigma_n\}} \leq n\right\}}\right] \rightarrow 0. \quad (17)$$

For (15), we have for $\delta, C_1^*, C_1 > 0$,

$$\begin{aligned} & \sum_{i=1}^n P\left(\frac{|f_{in}(\mathbf{X}_n)|^2}{\sigma_n^2} I_{\{|f_{in}(\mathbf{X}_n)| > n^{\frac{1}{2}} \varepsilon \sigma_n\}} > n\right) \\ &= P\left(\frac{|f_{in}(\mathbf{X}_n)|^2}{\sigma_n^2} > n\right) \leq n \frac{E[|f(\mathbf{X})|^{2+\delta}]}{n^{1+\frac{\delta}{2}} \sigma_n^{2+\delta}} \leq \frac{\alpha^{-\frac{\delta}{2m}} E[|f(\mathbf{X})|^2]}{n^{\frac{\delta}{2}} \sigma_n^{2+\delta}} \\ &\leq C_1^* \frac{\alpha^{-\frac{1+\delta}{2m}}}{n^{\frac{\delta}{2}} \sigma_n^{2+\delta}} \leq \frac{C_1}{n^{\frac{\delta}{2}} \alpha^{\frac{\delta}{4m}}} \rightarrow 0 \end{aligned}$$

if $n\alpha^{1/2m} \rightarrow \infty$. Similarly,

$$\begin{aligned} & n^{-2} \sum_{i=1}^n E\left[\frac{|f_{in}(\mathbf{X}_n)|^4}{\sigma_n^4} I_{\{|f_{in}(\mathbf{X}_n)| > n^{\frac{1}{2}} \varepsilon \sigma_n\}} I_{\{|f_{in}(\mathbf{X}_n)| \leq n \sigma_n\}}\right] \\ &\leq n^{-1} E\left[\frac{|f(\mathbf{X})|^4}{\sigma_n^4} I_{\{|f(\mathbf{X})| > n^{\frac{1}{2}} \varepsilon \sigma_n\}}\right] \\ &\leq \frac{E|f(\mathbf{X})|^4}{n \sigma_n^4} \leq C^* (n\alpha^{\frac{1}{2m}})^{-1} \rightarrow 0, \end{aligned}$$

if $n\alpha^{1/2m} \rightarrow \infty$. Lastly, for $\delta, C_1^* > 0$,

$$\begin{aligned} & n^{-1} \sum_{i=1}^n E\left[\frac{|f_{in}(\mathbf{X}_n)|^2}{\sigma_n^2} I_{\{|f_{in}(\mathbf{X}_n)| > n^{\frac{1}{2}} \varepsilon \sigma_n\}} I_{\left\{\frac{|f_{in}(\mathbf{X}_n)|^2}{\sigma_n^2} I_{\{|f_{in}(\mathbf{X}_n)| > n^{\frac{1}{2}} \varepsilon \sigma_n\}} \leq n\right\}}\right] \\ &\leq E\left[\frac{|f_{in}(\mathbf{X}_n)|^2}{\sigma_n^2} I_{\{|f(\mathbf{X})| > n^{\frac{1}{2}} \varepsilon \sigma_n\}}\right] \leq C_1^* \frac{\alpha^{-\frac{1+\delta}{2m}}}{n^{\frac{\delta}{2}} \sigma_n^{2+\delta}} \rightarrow 0, \end{aligned}$$

if $n\alpha^{1/2m} \rightarrow \infty$. Now for (14), we have that

$$n^{-1} \sum_{i=1}^n \frac{|f_{in}(\mathbf{X}_n)|^2}{\sigma_n^2} I_{\{|f_{in}(\mathbf{X}_n)| > n^{\frac{1}{2}} \varepsilon \sigma_n\}} = \frac{|f_{\cdot n}(\mathbf{X}_n)|^2}{\sigma_n^2} I_{\{|f_{\cdot n}(\mathbf{X}_n)| > n^{\frac{1}{2}} \varepsilon \sigma_n\}} \xrightarrow{P} 0$$

is equivalent to $[|f_n(\mathbf{X}_n)|/\sigma_n]I_{\{|f_n(\mathbf{X}_n)| > n^{1/2}\varepsilon\sigma_n\}} \rightarrow_P 0$. Then for every $\delta > 0$ and some constant $C_1^* > 0$,

$$\begin{aligned} & P\left(\frac{|f_n(\mathbf{X}_n)|}{\sigma_n} I_{\{|f_n(\mathbf{X}_n)| > n^{\frac{1}{2}}\varepsilon\sigma_n\}} > \delta\right) \\ &= P\left(\frac{|f_n(\mathbf{X}_n)|}{\sigma_n} > n^{\frac{1}{2}}\varepsilon\right) \leq \frac{E[f_n(\mathbf{X}_n)^2]}{n\varepsilon\sigma_n^2} \cong \frac{E[f^2(\mathbf{X})]}{n\varepsilon\sigma_n^2} \leq C_1^* \frac{\alpha^{-\frac{1}{2m}}}{n\varepsilon\alpha^{-\frac{1}{2m}}} = \frac{C_1^*}{n\varepsilon} \rightarrow 0. \end{aligned}$$

This completes the proof for (ii) and the proof for Proposition 2.

In view of the first part of Proposition 2 and the direct application of Proposition 1 with $f_{in}(\mathbf{X}_n)(t)$ given by (10) and $g(\mathbf{X}) = U(t)$, we have the following result.

Proposition 3. *Provided $n\alpha^{3/2m} \rightarrow \infty$ as $n \rightarrow \infty$ for each fixed $t \in [0, 1]$,*

$$\frac{U_{cc}(t)}{\sqrt{\frac{\sigma_n^2(t)}{n} + \frac{1-\gamma}{\gamma} \frac{\text{Var}[f(\mathbf{X})](t)}{n}}} \rightarrow_d N(0, 1). \quad (18)$$

Proof of Theorem 2. By Slutsky's Theorem, Proposition 2, and Proposition 3, it suffices to show that $\sqrt{n/\sigma_n^2}[\hat{\beta}_{cc}(t) - \hat{\beta}_1(t)]$, $\sqrt{n/\sigma_n^2}[\hat{\beta}_1(t) - \beta^*(t)]$, and $\sqrt{n/\sigma_n^2}[\beta_\alpha(t) - \beta_0(t)]$ converge in probability to zero as $n \rightarrow \infty$. Under (7), these follow from (8), Lemmas 1 and 6 in ZK90, and (18), respectively.

Proof of Theorem 3. We first introduce a convenient representation of $\check{A}(\beta_0, s)$. For each $j = 0, 1$, we define

$$\check{S}^{(j)}(\beta_0, s) \equiv \frac{n_1}{n} \check{S}_1^{(j)}(\beta_0, s) + \frac{n_0}{n} \check{S}_0^{(j)}(\beta_0, s) \quad (19)$$

with $\check{S}_0^{(j)}(\beta_0, s) \equiv (1/\tilde{n}_0) \sum_{l \in \tilde{\mathcal{R}}^j} Y_l(s) Z_l^j e^{\beta_0 Z_l}$, $\check{S}_1^{(j)}(\beta_0, s) \equiv (1/n_1) \sum_{l \in \mathcal{R}^j} Y_l(s) Z_l^j e^{\beta_0 Z_l}$. Then $\check{A}(\beta_0, s) = \check{S}^{(1)}(\beta_0, s)/\check{S}^{(0)}(\beta_0, s)$. In addition, we write $S^{(j)}(\beta_0, s)$, $j = 0, 1$ in Section 2 as

$$S^{(j)}(\beta_0, s) = \frac{n_1}{n} S_1^{(j)}(\beta_0, s) + \frac{n_0}{n} S_0^{(j)}(\beta_0, s), \quad (20)$$

where $S_0^{(j)}(\beta_0, s) \equiv (1/n_0) \sum_{l \in \mathcal{R}^j} Y_l(s) Z_l^j e^{\beta_0 Z_l}$, $S_1^{(j)}(\beta_0, s) \equiv (1/n_1) \sum_{l \in \mathcal{R}^j} Y_l(s) Z_l^j e^{\beta_0 Z_l}$.

From (19) and (20), and applying the same calculation as in Section 2, we

have, from Zhang and Goldstein (2003), that

$$\begin{aligned}
 & n^{\frac{1}{2}}(\check{A}(\beta_0, s) - A(\beta_0, s)) \\
 &= n^{\frac{1}{2}}\left(\frac{\check{S}^{(1)}(\beta_0, s)}{\check{S}^{(0)}(\beta_0, s)} - \frac{S^{(1)}(\beta_0, s)}{S^{(0)}(\beta_0, s)}\right) \\
 &= -\frac{n_0^{\frac{1}{2}}}{n^{\frac{1}{2}}}n_0^{\frac{1}{2}}\{(\check{S}_0^{(0)}(\beta_0, s) - S_0^{(0)}(\beta_0, t))a(\beta_0, s) - (\check{S}_0^{(1)}(\beta_0, s) - S_0^{(1)}(\beta_0, s))\} \\
 &\quad \times s^{(0)}(\beta_0, s)^{-1} + O_P(n^{-1}).
 \end{aligned}$$

Therefore, as with $U_{cc}(t)$, we have

$$U_{cl}(t) = U(t) + \frac{n_0^{\frac{1}{2}}}{n^{\frac{1}{2}}}[\tilde{n}_0^{-1} \sum_{i \in \mathcal{R}^0} \delta_{in} f_{in}(\mathbf{X}_n) - f_n(\mathbf{X}_n)] + II_{2b}$$

with $f_{in}(\mathbf{X}_n)$ and $f(\mathbf{X})$ given by (10) and (6). Here II_{2b} is given in Section 2 with $\sqrt{n/\sigma_n^2} II_{2b} \rightarrow_P 0$ as $n^{1/2} \alpha^{1/4m} \rightarrow \infty$. Noting that $(n_0/n) \rightarrow_P 1 - \pi$ as $n \rightarrow \infty$, and with the application of Proposition 1, the consistency and asymptotic normality of $\hat{\beta}_{cl}(t)$ can be proved following the lines of those for $\hat{\beta}_{cc}(t)$.

Acknowledgements

This paper is based upon the work supported by Centennial Research Grant from Concordia College for H. Zhang, and by National Science Foundation under Grant No. #0132289 for C. Huang. The authors sincerely appreciate the reviews provided by an anonymous referee and editor. Their suggestions and comments have strengthened the paper significantly.

References

- Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100-1120.
- Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley, New York.
- Cai, Z. and Sun, Y. (2003). Local linear estimation for time-dependent coefficients in Cox's regression models. *Scand. J. Statist.* **30**, 93-111.
- Chen, K. and Lo, S. (1999). Case-cohort and case-control analysis with Cox's model. *Biometrika* **25**, 755-764.
- Cochran, W. (1977) *Sampling Techniques*. Wiley, New York.
- Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 182-220.
- Durrett, R. (1996). *Probability: Theory and Examples*. 2nd edition. Duxbury Press, Belmont, California.
- Fan, J., Gijbels, I. and King, M. (1997). Local likelihood and local partial likelihood in hazard regression. *Ann. Statist.* **25**, 1661-1690.

- Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.
- Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to Breast Cancer Prognosis. *J. Amer. Statist. Assoc.* **87**, 942-951.
- Gray, R. J. (1994). Spline-based tests in survival analysis. *Biometrics* **50**, 640-652.
- Gu, C. (2002). *Smoothing Spline ANOVA Models*. Springer-Verlag, New York.
- Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *J. Roy. Statist. Soc. Ser. B* **55**, 757-796.
- Marzec, L. and Marzec, P. (1997). On fitting Cox's regression model with time-dependent coefficients. *Biometrika* **84**, 901-908.
- Murphy, S. A. and Sen, P. K. (1991). Time-dependent coefficients in a Cox-type regression model. *Stochastic Process. Appl.* **39**, 153-180.
- Naimark, M. A. (1967). *Differential Operators, Part I*. Ungar, New York.
- O'Sullivan, F. (1988). Nonparametric estimation of relative risk using splines and cross-validation. *SIAM J. Sci. Statist. Comput.* **9**, 531-542.
- O'Sullivan, F. (1993). Nonparametric estimation in the Cox model. *Ann. Statist.* **21**, 124-145.
- Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73**, 1-11.
- Self, S. G. and Prentice, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann. Statist.* **16**, 64-81.
- Silverman, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Ann. Statist.* **10**, 795-810.
- Yeh, J. (1973). *Stochastic Processes and the Wiener Integral*. Marcel Dekker, New York.
- Wahba, G. (1977). Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.* **14**, 651-667.
- Zhang, H. and Goldstein, L. (2003). Information and asymptotic efficiency of case-cohort sampling design in Cox regression model. *J. Multivariate Anal.* **85**, 292-317.
- Zhang, H. and Huang, C. (2004). Nonparametric survival analysis on time-dependent covariate effects in case-cohort sampling design. Technical Report, Department of Statistics, North Dakota State University. <http://www.cord.edu/faculty/zhang/cc-technicalReport.pdf>.
- Zucker, D. and Karr, A. (1990). Nonparametric survival analysis with time-dependent covariate effects: A penalized partial likelihood approach. *Ann. Statist.* **18**, 329-353.

Department of Mathematics and Computer Science, Concordia College, Moorhead, MN 56562, U.S.A.

E-mail: zhang@cord.edu

Department of Statistics, North Dakota State University, Fargo, ND 58105, U.S.A.

E-mail: Chunfeng.Huang@ndsu.nodak.edu

(Received April 2004; accepted November 2004)