Statistica Sinica **32** (2022), 983-1006
doi:https://doi.org/10.5705/ss.202020.0141

MAXIMUM LIKELIHOOD ESTIMATION FOR COX PROPORTIONAL HAZARDS MODEL WITH A CHANGE HYPERPLANE

Yu Deng[1], Jianwen Cai[2] and Donglin Zeng[2]

[1] *Genentech, Inc.* and [2] *University of North Carolina at Chapel Hill*

Abstract: We propose a Cox proportional hazards model with a change hyperplane to allow the effect of risk factors to differ depending on whether a linear combination of baseline covariates exceeds a threshold. The proposed model is a natural extension of the change-point hazards model. We maximize the partial likelihood function for estimation and suggest an m-out-of-n bootstrapping procedure for inference. We establish the asymptotic distribution of the estimators and show that the estimators for the change hyperplane converge in distribution to an integrated composite Poisson process defined on a multidimensional space. Finally, the numerical performance of the proposed approach is demonstrated using simulation studies and an analysis of the Cardiovascular Health Study.

Key words and phrases: Change hyperplane, m-out-of-n bootstrap, proportional hazards model.

1. Introduction

The Cox proportional hazards model with a change point is often used to identify subjects whose risk profiles are substantially different from others. These subjects are characterized by a biomarker exceeding a threshold (Tapp et al. (2006); Marquis et al. (2002); Zhao et al. (2014)). More recently, such models have been increasingly used in subgroup analyses of clinical trials in order to determine treatment-respondents based on a threshold of some potentially predictive biomarker. Inferences for the change-point model have been studied extensively (Liang, Self and Liu (1990); Luo (1996); Pons (2002); Luo (1996); Gandy, Jensen and Lütkebohmert (2005); Gandy and Jensen (2005); Jensen and Lütkebohmert (2008); Luo and Boyett (1997); Pons (2003); Kosorok and Song (2007)). In particular, Pons (2003) shows that the asymptotic distribution of the maximum likelihood estimator for the change point is given by a composite Poisson process.

Corresponding author: Donglin Zeng, Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, USA. E-mail: dzeng@email.unc.edu.

In practice, it is rather restrictive to assume a change point is determined by a single biomarker. For example, Zhao et al. (2014) investigated the change point of leukocyte telomere length (LTL) for diabetes incidence in the Strong Heart Family Study. In the same study, the change point based on LTL has been observed to depend on triglycerides and high-density lipoproteins (HDL), indicating that the incidence of diabetes can change dramatically depending on a combination of all these biomarkers. To better model this general change-point pattern, a natural extension of the change-point model, considered here, is a Cox proportional hazards model with a change hyperplane. More specifically, we assume that the log-hazard ratios of some covariates differs, depending on whether a linear combination of baseline biomarkers is larger than an unknown threshold. In other words, the risk profiles for subjects whose baseline biomarkers are above the hyperplane can be very different from those who are below.

Estimation and inference for the Cox proportional hazards model with a change hyperplane are much more challenging. We propose maximum likelihood approach for estimation in which all parameters, including the coefficients of the change hyperplane, are estimated by maximizing the Cox partial likelihood function. Because the likelihood function is not continuous in the latter parameters, we adopt a genetic optimization algorithm (Sekhon and Mebane (1998)) for optimization. For inference purposes, we suggest using an m-out-of-n bootstrap procedure to construct the confidence intervals. Because the hyperplane is defined by more than one biomarker, existing theory for the change-point model is no longer applicable. To establish the asymptotic distribution of the estimators for the change hyperplane, we need to carefully partition the support of the hyperplane, and then show that its asymptotic distribution is determined by an integrated composite Poisson process defined on a multidimensional space of the covariates. To the best of our knowledge, this is a novel finding. Furthermore, when there are no covariates except a constant term in the change plane, the derived asymptotic distribution reduces to the change-point distribution given in Pons (2003).

Note that although the proposed model can be viewed as one single-index hazard model, which is studied in Wang (2004) and Huang and Liu (2006), the link function for our model is discontinuous. In contrast, the usual single-index model assumes a smooth link function. This leads to substantially different properties for the maximum likelihood estimators. For example, we show that the estimators for the single index, that is, the coefficient in the hyperplane, has a convergence rate of $1/n$, in contrast to the standard $1/\sqrt{n}$ rate in Wang (2004) and Huang and Liu (2006).

2. Methods

2.1. Model and parameter estimation

For subject i, let $\widetilde{T}_i$ denote the failure time, $\boldsymbol{X}_i$ consist of the baseline biomarkers of the p_1-dimension and constant one and $\boldsymbol{Z}_i(t)$ be the potential time-dependent covariates with dimension p_2. A Cox proportional hazards model with a change hyperplane assumes that the hazard rate function for $\widetilde{T}_i$ given $\boldsymbol{W}_i(t) \equiv \left\{\boldsymbol{X}_i^{\mathrm{T}}, \boldsymbol{Z}_i^{\mathrm{T}}(t)\right\}^{\mathrm{T}}$ takes the form

$$\lambda(t|\boldsymbol{W}_i) = \lambda_0(t) \exp\left\{\boldsymbol{\beta}_1^{\mathrm{T}} \boldsymbol{Z}_i(t) + \beta_2 I(\boldsymbol{\eta}^{\mathrm{T}} \boldsymbol{X}_i > 0) + \boldsymbol{\beta}_3^{\mathrm{T}} \boldsymbol{Z}_i(t) I(\boldsymbol{\eta}^{\mathrm{T}} \boldsymbol{X}_i > 0)\right\},$$

where $\lambda_0(t)$ is an unknown baseline function, $\boldsymbol{\beta} \equiv \left(\boldsymbol{\beta}_1^{\mathrm{T}}, \beta_2, \boldsymbol{\beta}_3^{\mathrm{T}}\right)^{\mathrm{T}}$ is a vector of $2p_2 + 1$ unknown parameters, and $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_{p_1}, \eta_0)^{\mathrm{T}}$ is a vector of $p_1 + 1$ unknown change-hyperplane parameters. Because the model remains the same if we replace $\boldsymbol{\eta}$ with any rescaled $\boldsymbol{\eta}$, for model identifiability, we further assume that $\eta_1^2 + \eta_2^2 + \cdots + \eta_{p_1}^2 = 1$ and η_1 is positive. Additionally, we assume $(\beta_2, \boldsymbol{\beta}_3^{\mathrm{T}}) \neq 0$; otherwise, any $\boldsymbol{\eta}$ gives the same model. In the model, the change hyperplane is given by $\eta_1 X_{i1} + \eta_2 X_{i2} + \cdots + \eta_{p_1} X_{ip_1} + \eta_0$. The effect of $\boldsymbol{Z}_i(t)$ is $\boldsymbol{\beta}_1$ when $\boldsymbol{\eta}^{\mathrm{T}} \boldsymbol{X}_i \leq 0$, and becomes $(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_3)$ when $\boldsymbol{\eta}^{\mathrm{T}} \boldsymbol{X}_i > 0$. Furthermore, the hazard ratio between two groups $\boldsymbol{\eta}^{\mathrm{T}} \boldsymbol{X}_i > 0$ and $\boldsymbol{\eta}^{\mathrm{T}} \boldsymbol{X}_i \leq 0$ is $\exp\left\{\beta_2 + \boldsymbol{\beta}_3^{\mathrm{T}} \boldsymbol{Z}_i(t)\right\}$. When $p_1 = 1$, it reduces to the change-point model in Pons (2003).

Suppose that right-censored data are obtained from n independent and identically distributed (i.i.d.) subjects and we denote them as $(T_i = \widetilde{T}_i \wedge C_i, \Delta_i = I(\widetilde{T}_i \leq C_i), \boldsymbol{W}_i)$, for $i = 1, \ldots, n$, where C_i is the censoring time and is assumed to be noninformative. We propose estimating all the parameters by maximizing the observed likelihood function. After profiling the nuisance parameter for $\lambda_0(t)$, we obtain the following partial likelihood to be maximized for the estimation:

$$L_n(\boldsymbol{\eta}, \boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{\exp\left[r_{\boldsymbol{\eta},\boldsymbol{\beta}}\left\{\boldsymbol{W}_i(T_i)\right\}\right]}{\sum_{l=1}^{n} I(T_l \geq T_i) \exp\left[r_{\boldsymbol{\eta},\boldsymbol{\beta}}\left\{\boldsymbol{W}_l(T_i)\right\}\right]} \right)^{\Delta_i},$$

where $r_{\boldsymbol{\eta},\boldsymbol{\beta}}\left\{\boldsymbol{W}_i(t)\right\} \equiv \boldsymbol{\beta}_1^{\mathrm{T}} \boldsymbol{Z}_i(t) + \beta_2 I(\boldsymbol{\eta}^{\mathrm{T}} \boldsymbol{X}_i > 0) + \boldsymbol{\beta}_3^{\mathrm{T}} \boldsymbol{Z}_i(t) I(\boldsymbol{\eta}^{\mathrm{T}} \boldsymbol{X}_i > 0)$. We adopt a similar two-step procedure (Luo and Boyett (1997)) to compute the maximum likelihood estimators. In the first step, for any fixed value of $\boldsymbol{\eta}$, we obtain the estimates of $\boldsymbol{\beta}$ by applying the Newton–Raphson method to maximize the logarithm of the partial likelihood function. The algorithm for this step guarantees convergence to the global maximum, owing to the strict concavity of the log-partial likelihood function in terms of $\boldsymbol{\beta}$. In the second step, we apply an evolutionary algorithm with a quasi-Newton method to maximize the profile function for $\boldsymbol{\eta}$, subject to the constraints for $\boldsymbol{\eta}$ (Sekhon and Mebane (1998)). This evolutionary algorithm has been widely applied to optimize the function when the objective function is not a continuous function of the parameter of interest. We iterate between these two steps till convergence. Finally, we denote $(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\beta}}) = \text{argmax}_{\eta_1^2+\eta_2^2+\cdots+\eta_{p_1}^2=1,\eta_1>0,\boldsymbol{\beta}}\, l_n(\boldsymbol{\eta}, \boldsymbol{\beta})$, where $l_n(\boldsymbol{\eta}, \boldsymbol{\beta}) = \log L_n(\boldsymbol{\eta}, \boldsymbol{\beta})$.

2.2. Inference

We prove that $\widehat{\boldsymbol{\eta}}$ and $\widehat{\boldsymbol{\beta}}$ are asymptotically independent, and that their convergence rates are $1/n$ and $1/\sqrt{n}$, respectively. In addition, the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ remains normal, regardless of whether or not $\boldsymbol{\eta}$ is known. Consequently, the inference of $\widehat{\boldsymbol{\beta}}$ can be carried out in the same way as for the usual Cox proportional hazards model, as if $\widehat{\boldsymbol{\eta}}$ were a fixed constant. As a result, the corresponding confidence intervals are generated by a normal approximation.

The inference for $\boldsymbol{\eta}$ is more challenging because the asymptotic distribution of $\widehat{\boldsymbol{\eta}}$ is no longer normal and, in fact, has no explicit expression. For parameters like $\widehat{\boldsymbol{\eta}}$ that are estimated at the nonstandard n-rate, Shao (1994), Bickel, Götze and van Zwet (2012), Politis and Romano (1999), and Xu, Sen and Ying (2014) proposed using the m-out-of-n bootstrap to generate the 95% confidence intervals, where m is determined by a data-driven approach (Hall, Horowitz and Jing (1995); Lee (1999); Cheung, Lee and Young (2005); Bickel and Sakov (2005); Bickel and Sakov (2008)). Xu, Sen and Ying (2014) showed the theoretical consistency of the m-out-of-n bootstrap for the Cox proportional hazards model with a change point.

Therefore, for the inference in our approach, we suggest adopting a similar m-out-of-n bootstrap algorithm. Specifically, we choose to adapt the algorithm proposed by Bickel and Sakov (2008) to select m. In this algorithm, for each $j = 0, 1, \ldots, p_1$, we first determine m_j as the maximum sample size that achieves the stable empirical distribution of the bootstrap estimators for η_j. Then, the final m is defined as the minimum of m_j. Both the standard error estimator for $\widehat{\boldsymbol{\eta}}$ and the confidence interval for $\boldsymbol{\eta}$ are adjusted by n/m, based on the convergence rate $1/n$ of $\widehat{\boldsymbol{\eta}}$ (Theorem 3). In particular, the equal-tailed 95% confidence intervals are generated as $\left(\widehat{\boldsymbol{\eta}} - (Q_{\widehat{\boldsymbol{\eta}},0.95}/(n/m)), \widehat{\boldsymbol{\eta}} + (Q_{\widehat{\boldsymbol{\eta}},0.95}/(n/m))\right)$, where $Q_{\widehat{\boldsymbol{\eta}},0.95}$ is the 95th quantile of the absolute value $|\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{\eta}}_m^{(b)}|$, for $b = 1, 2 \ldots, B$.

2.3. Hypothesis testing for the change hyperplane

In practice, an important question is whether the change hyperplane exists. Equivalently, we wish to test the null hypothesis $H_0 : \beta_2 = 0, \boldsymbol{\beta}_3^{\mathrm{T}} = \mathbf{0}$ in our proposed model. Because the estimation of the change hyperplane relies on either β_2 or $\boldsymbol{\beta}_3$ being not equal to zero, the model is not identifiable given that both β_2 and $\boldsymbol{\beta}_3$ are zero under the null hypothesis. The supremum (SUP) test has been proposed to verify the existence of the change point based on a single covariate (Davies (1977), Davies (1987), Kosorok and Song (2007)). Here, we extend this SUP test with score statistics to the case of multidimensional covariates. Specifically, our test statistic is

$$\mathrm{SUP}_{k^{p_1}} = \sup_{\eta_j \in \{\eta_{j1},\ldots,\eta_{jk}\}, j=0,2,\ldots,p_1} \boldsymbol{U}(\boldsymbol{\eta})^{\mathrm{T}} \boldsymbol{\Sigma}(\boldsymbol{\eta})^{-1} \boldsymbol{U}(\boldsymbol{\eta}),$$

where $\boldsymbol{U}(\boldsymbol{\eta}) = \partial l_n(\boldsymbol{\eta}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}$, $\boldsymbol{\Sigma}(\boldsymbol{\eta}) = -\partial^2 l_n(\boldsymbol{\eta}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}^2$, and $\{\eta_{j1}, \ldots, \eta_{jk}\}$ is the set of k predetermined values for each η_j, for $j = 0, 2, \ldots, p_1$. We use a permutation to generate the null distribution of the proposed test statistic. Under the null hypothesis, there is no change-hyperplane effect on the response. Thus, we randomly shuffle the covariate $\boldsymbol{X}_i$ to obtain the permutation distribution of the proposed test statistics. We reject the null hypothesis at a significance level of α if $\mathrm{SUP}_{k^{p_1}}$ is larger than the upper α-quantile of the permutation distribution.

3. Asymptotic Properties

The consistency and asymptotic distributions of the estimators for both the change hyperplane and the regression parameters are established in this section. Let τ be the study duration, which is assumed to be finite. First, we define $Y_i(t) = I(T_i \geq t)$ as the at-risk process for subject i, and let $\boldsymbol{s}^{(r)}(t; \boldsymbol{\eta}, \boldsymbol{\beta}) = \mathrm{E}\{Y_i(t)\widetilde{\boldsymbol{Z}}_i^{\otimes r}(t; \boldsymbol{\eta}) \exp[r_{\boldsymbol{\eta},\boldsymbol{\beta}}\{\boldsymbol{W}_i(t)\}]\}$, for $r = 0, 1, 2$, and $\widetilde{\boldsymbol{Z}}_i(t; \boldsymbol{\eta}) = \{\boldsymbol{Z}_i^{\mathrm{T}}(t), I(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}_i > 0), \boldsymbol{Z}_i^{\mathrm{T}}(t) I(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}_i > 0)\}^{\mathrm{T}}$. In addition to assuming $\eta_1^2 + \eta_2^2 + \cdots + \eta_{p_1}^2 = 1$ with $\eta_1 > 0$, we assume the following conditions.

(C.1) The joint density of $(X_{i1}, X_{i2}, \ldots, X_{ip_1})$ with respect to a dominating measure has a support containing zero and is assumed to be strictly positive, bounded, and continuous in a neighborhood $V_0 = \{\boldsymbol{x} : |\boldsymbol{\eta}_0^{\mathrm{T}}\boldsymbol{x}| < \epsilon\}$, where $\boldsymbol{\eta}_0$ is the true value of $\boldsymbol{\eta}$. In addition, each $Z_{ij}(t)$ has a finite total variation with probability one, and the joint density of $\{Z_{i1}(t), \ldots, Z_{ip_2}(t)\}$ given $\boldsymbol{X}_i$ is assumed to be strictly positive and bounded for any t in $[0, \tau]$.

(C.2) The matrix $E\{(1, \boldsymbol{X}_i)^{\mathrm{T}}(1, \boldsymbol{X}_i)\}$ has a full rank. In addition, conditional on $\boldsymbol{X}_i$, if with probability one, $a(t) + b^{\mathrm{T}}\boldsymbol{Z}_i(t) = 0$ holds for any $t \in [0, \tau]$ for some deterministic function $a(t)$ and constant b, then $a(t) = 0$ and $b = 0$.

(C.3) For any $V_\delta(\boldsymbol{\eta}_0) = \{\boldsymbol{\eta} : \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| < \delta\}$, the covariance matrix $I(\boldsymbol{\eta}, \boldsymbol{\beta}) =$

$\int_0^\tau v(t;\boldsymbol{\eta},\boldsymbol{\beta})s^{(0)}(t;\boldsymbol{\eta},\boldsymbol{\beta})\lambda_0(t)dt$ is positive definite, where

$$\boldsymbol{v}(t;\boldsymbol{\eta},\boldsymbol{\beta}) = \frac{\boldsymbol{s}^{(2)}(t;\boldsymbol{\eta},\boldsymbol{\beta})}{s^{(0)}(t;\boldsymbol{\eta},\boldsymbol{\beta})} - \left\{\frac{\boldsymbol{s}^{(1)}(t;\boldsymbol{\eta},\boldsymbol{\beta})}{s^{(0)}(t;\boldsymbol{\eta},\boldsymbol{\beta})}\right\}^{\otimes 2}.$$

In addition, the smallest eigenvalue of $\int_0^\tau \mathrm{E}\left[Y_i(t)\{1,\boldsymbol{Z}_i(t)\}^{\otimes 2}\,|\boldsymbol{\eta}_0^{\mathrm{T}}\boldsymbol{X}_i = 0\right] d\Lambda_0(t)$ is positive.

(C.4) We assume $\boldsymbol{\beta}$ is bounded by a known constant B, and $\lambda_0(t)$ is continuously differentiable in $[0,\tau]$. Additionally, $P\{Y(\tau)=1\} > 0$.

(C.1) and (C.2) are needed for the identifiability of the change hyperplane and the regression coefficients. (C.2) holds if $\boldsymbol{Z}_i$ is time-independent and $E\{(1,\boldsymbol{Z}_i)(1,\boldsymbol{Z}_i)^T|\boldsymbol{X}_i\}$ is full rank. (C.3) requires that $\lambda_0(t)$ is bounded and that the at-risk probability is nonzero for $t \in [0,\tau]$. Condition (C.4) holds if the study ends at a fixed time τ so subjects who are alive at τ are censored at τ. Our first theorem establishes the identifiability of the change-hyperplane parameters and the regression coefficient parameters.

Theorem 1. *Under the condition that at least one of the elements in β_2 or $\boldsymbol{\beta}_3$ is nonzero, the change-hyperplane parameters $\boldsymbol{\eta}$ and the regression parameters $\boldsymbol{\beta}$ are identifiable.*

Theorem 2 and Theorem 3 show the consistency and convergence rates of the change-hyperplane estimators and the regression coefficients estimators. Theorem 3 implies that the convergence rates for $\widehat{\boldsymbol{\eta}}$ and $\widehat{\boldsymbol{\beta}}$ are $1/n$ and $1/\sqrt{n}$, respectively. These rates are applied in Theorem 4 to establish the asymptotic distributions of the estimators.

Theorem 2. *Under conditions* (C.1)–(C.4)*, $\widehat{\boldsymbol{\eta}}$ and $\widehat{\boldsymbol{\beta}}$ converge in probability to $\boldsymbol{\eta}_0$ and $\boldsymbol{\beta}_0$, respectively as $n \to \infty$.*

Theorem 3. *Under conditions* (C.1)–(C.4)*, the following hold:*

$$\lim_{A\to\infty}\lim_{n\to\infty} P_0\left(n\left\|\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\right\| > A\right) = 0,$$
$$\lim_{A\to\infty}\lim_{n\to\infty} P_0\left(n^{1/2}\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\| > A\right) = 0.$$

In other words, $\|\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\| = O_p(1/n)$ and $\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\| = O_p(1/\sqrt{n})$.

To give the asymptotic distributions for $\widehat{\boldsymbol{\eta}}$ and $\widehat{\boldsymbol{\beta}}$, we use W for $\boldsymbol{\eta}_0^{\mathrm{T}}\boldsymbol{X}$, and let

$$\eta = \Delta\left[\left\{\beta_{20} + \boldsymbol{\beta}_{30}^{\mathrm{T}}\boldsymbol{Z}(T)\right\}\right] + \int_0^\tau \phi(t)d\Lambda_0(t),$$

where $\phi(t) = Y(t)\exp\left\{\boldsymbol{\beta}_{10}^{\mathrm{T}}\boldsymbol{Z}(t)\right\}\left[1 - \exp\left\{\beta_{20} + \boldsymbol{\beta}_{30}^{\mathrm{T}}\boldsymbol{Z}(t)\right\}\right]$. Additionally, we define $\Gamma(\boldsymbol{x}, t)$ as a random process that is independent for any $\boldsymbol{x}$ and t, each with the conditional distribution of η given $W = 0$ and $\boldsymbol{X} = \boldsymbol{x}$. Furthermore, $v(\omega, t)$ is a multivariate Poisson process defined on $\Omega \times (0, \infty)$, where Ω is the probability measure space generating data, with Poisson intensity $E\{v(\omega \in \mathcal{A}, u \in [t, t+dt))\} \equiv P(\mathcal{A})dt$ for any measurable set $\mathcal{A}$ in the σ-field of the probability measure space and for any $t > 0$. Finally, we define the following integrated compound Poisson process:

$$Q^-(\boldsymbol{u}_1) \equiv \int_\Omega \int_0^{\max(0, \boldsymbol{X}(\omega)^{\mathrm{T}}\boldsymbol{u}_1)} \Gamma(\boldsymbol{X}(\omega), t)v(d\omega, dt)$$

and

$$Q^+(\boldsymbol{u}_1) \equiv \int_\Omega \int_0^{\max(0, -\boldsymbol{X}(\omega)^{\mathrm{T}}\boldsymbol{u}_1)} \Gamma(\boldsymbol{X}(\omega), t)v(d\omega, dt).$$

That is, the integrals inside Q^+ and Q^- are both some compound Poisson process. With these definitions, we have the following theorem.

Theorem 4. *Under conditions* (C.1)–(C.4), $n(\boldsymbol{\eta}_0 - \widehat{\boldsymbol{\eta}})$ *and* $n^{1/2}(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}})$ *are asymptotically independent. Furthermore,* $n(\boldsymbol{\eta}_0 - \widehat{\boldsymbol{\eta}})$ *converges weakly to* $\inf\{\boldsymbol{u}_1 : \arg\max Q(\boldsymbol{u}_1)\}$*, where* $Q = Q^+ - Q^-$*, and* $n^{1/2}(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}})$ *converges weakly to* $N(\mathbf{0}, \boldsymbol{I}(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)^{-1})$*, where* $\boldsymbol{I}(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)^{-1}$ *is the efficient information bound for* $\boldsymbol{\beta}_0$*, assuming* $\boldsymbol{\eta}_0$ *is known.*

Because the change hyperplane can be determined precisely by a finite number of observations near the true location, the estimator for the parameter in the change hyperplane $\widehat{\boldsymbol{\eta}}$ has a convergence rate in the order of n^{-1}. Thus, the randomness in $\widehat{\boldsymbol{\eta}}$ has no effect on the random behavior of $\widehat{\boldsymbol{\beta}}$, the variability of which is in the order of $1/\sqrt{n}$. This explains why the two distributions are asymptotically independent. The proof of this theorem relies on the derivation of the asymptotic process for $Q(\boldsymbol{u}_1)$. Because the change hyperplane depends on the random variable $\boldsymbol{X}$, this derivation is more challenging than the case with a change point. The proof is given in the appendix.

4. Simulation Studies

We conducted simulation studies to evaluate the performance of our proposed method. Our first set of studies was designed to assess the performance of the estimators and the coverage rate of the confidence interval. We considered one covariate $Z \sim$ Uniform(-1,1) and the change hyperplane with two covariates $X_1 \sim N(2, 1.5^2)$ and $X_2 \sim N(0,1)$. We generated the survival times $\widetilde{T}_i$ under the proportional hazards model $\Lambda(t|X_1, X_2, Z) = t\exp\{\beta_1 Z + \beta_2 I(\eta_1 X_1 + \eta_2 X_2 - \eta_0 > 0) + \beta_3 Z I(\eta_1 X_1 + \eta_2 X_2 - \eta_0 > 0)\}$, where $(\beta_1, \beta_2, \beta_3) = (-1, 1.8, 0.5)$, $(\eta_1, \eta_2, \eta_0) = (0.8, -0.6, 1.7)$, and $\eta_1^2 + \eta_2^2 = 1$. In order to obtain censoring rates of 10%, 30%, and 50%, we generated the censoring time from Uniform (0,680), Uniform(0,220), and Uniform(0,118), respectively. The number of subjects is 200 or 300. To use the m-out-of-n bootstrap, we consider a sequence of candidates, $[nk/10]$, where $k = 1, \ldots, 10$ and $[x]$ denotes the integer part of x. Following Bickel and Sakov (2008) and the description in Section 2.2, we first determine m_j as the maximal sample size in this sequence for each η_j that gives the stable bootstrap distribution. Then, the final m is chosen as the minimal size of these m_j. All results are based on 500 replications, and each m-out-of-n bootstrap consists of 100 replicates.

In Table 1, the proposed method provides approximately unbiased estimates for the change-hyperplane parameters η_2 and η_0. Here, we present only the results for η_2 and η_0, because η_1 and η_2 satisfy $\eta_1^2 + \eta_2^2 = 1$. In addition, the m-out-of-n bootstrap confidence interval generates proper coverage rates. When the number of subjects increases or the censoring rate decreases, the bias of the change point estimate and the variance estimates decrease. In Table 2, the results show that the estimates for the regression coefficients $\boldsymbol{\beta}$ are also approximately unbiased, and that the confidence intervals using a normal approximation have proper coverage rates.

Our second set of simulation studies compare the type-I errors and power of the SUP_{5^2}, SUP_{10^2}, and SUP_{20^2} tests under various scenarios. Because our test is based on two change-hyperplane parameters, the SUP test is evaluated on the set with k^2 points, where k is the number of grids in the prespecified range $[-1, 1]$ for η_2 and $[-10, 10]$ for η_0. The range for η_2 is determined by the conditions in Theorem 1. The range of η_0 is determined by the range of each covariate and the value of η_2. For example, the test SUP_{5^2} is evaluated on the grids $(-1, -0.5, 0, 0.5, 1) \times (-10, -5, 0, 5, 10)$. We examine the performance of these tests with sample sizes 200, 300, and 400. The results for the type-I errors and power are based on 10,000 and 1,000 replicates, respectively. All other

Table 1. Simulation Results for the Change-Hyperplane Parameters.

Censoring Rate	Sample Size	Parameters	Bias ($\times 10^{-2}$)	SSD ($\times 10^{-2}$)	95% CI ($\times 10^{-2}$)	Length ($\times 10^{-2}$)
50%	200	$\widehat{\eta}_2$	0.17	8.3	96.0	42.7
		$\widehat{\eta}_0$	1.11	15.7	95.2	73.9
	300	$\widehat{\eta}_2$	0.06	5.2	95.6	28.3
		$\widehat{\eta}_0$	0.83	9.5	94.6	50.6
30%	200	$\widehat{\eta}_2$	0.40	6.4	96.4	32.5
		$\widehat{\eta}_0$	0.76	11.7	96.6	56.9
	300	$\widehat{\eta}_2$	-0.13	4.0	96.2	21.0
		$\widehat{\eta}_0$	1.23	7.7	95.0	37.3
10%	200	$\widehat{\eta}_2$	-0.23	5.0	97.2	26.9
		$\widehat{\eta}_0$	1.81	9.8	95.8	46.8
	300	$\widehat{\eta}_2$	0.20	4.0	95.4	17.9
		$\widehat{\eta}_0$	0.86	7.1	95.6	31.7

NOTE: SSD stands for sample standard deviation. 95% CI is the coverage rate for the 95% confidence interval coverage. Length is the length of the 95% CI.

specifications are the same as the first set of simulations.

Table 3 shows that the type-I errors of all three tests are, in general, close to 0.05. As the sample sizes increase and the censoring rates decrease, the type-I errors get closer to 0.05. For the power, the performance of the supremum tests is determined by the numbers of grids, sample sizes, and censoring rates. Given the same sample size and censoring rate, the power stabilizes after the number of grids reaches 10 for each parameter. Given the tests with the same number of grids, the power increases as the sample size increases and the censoring rate decreases.

5. Application to the Cardiovascular Health Study

Here, we apply the proposed method to the Cardiovascular Health Study (CHS). The CHS recruited 5,888 participants aged 65 years and older from four U.S. communities to study the development and progression of CHD and stroke. We apply our approach to the cohort of male participants, who were free of CHD at baseline. The data contains 995 subjects, after excluding those with missing responses and covariates. Among them, 851 subjects developed CHD before the end of the study. We include a linear combination of HDL, systolic blood pressure, and cholesterol level to form the risk categories (high vs. low). We investigate the association between these risk categories and the risk of CHD using a Cox

Table 2. Simulation Results for the Regression Parameters.

Censoring Rate	Sample Size	Parameters	Bias ($\times 10^{-2}$)	SSD ($\times 10^{-2}$)	SSE ($\times 10^{-2}$)	95% CI ($\times 10^{-2}$)
50%	200	$\widehat{\beta}_1$	-4.69	33.2	34.4	94.4
		$\widehat{\beta}_2$	11.63	25.5	24.4	94.4
		$\widehat{\beta}_3$	3.92	39.9	40.7	95.0
	300	$\widehat{\beta}_1$	-2.54	26.7	27.0	95.4
		$\widehat{\beta}_2$	6.57	20.4	20.5	95.0
		$\widehat{\beta}_3$	0.51	32.1	32.4	95.2
30%	200	$\widehat{\beta}_1$	-3.46	25.0	24.5	96.2
		$\widehat{\beta}_2$	8.54	21.8	20.9	95.2
		$\widehat{\beta}_3$	1.89	32.0	31.4	95.6
	300	$\widehat{\beta}_1$	-2.40	20.2	20.6	94.8
		$\widehat{\beta}_2$	5.76	17.5	17.1	95.0
		$\widehat{\beta}_3$	0.35	25.9	26.4	95.6
10%	200	$\widehat{\beta}_1$	-2.28	21.0	20.7	95.0
		$\widehat{\beta}_2$	6.92	19.7	19.1	95.2
		$\widehat{\beta}_3$	1.12	28.1	27.3	96.2
	300	$\widehat{\beta}_1$	-1.53	17.0	18.1	94.2
		$\widehat{\beta}_2$	4.57	16.0	16.9	92.6
		$\widehat{\beta}_3$	0.31	22.8	22.9	95.2

NOTE: See Table 1. SSE stands for average standard error estimate.

proportional hazards model, adjusting for the baseline confounding covariates of age, hypertension, diabetes, and smoking status.

The analysis is conducted in two steps. First, we apply the SUP_{10^3} test to verify the existence of these risk categories. The test is significant, with a p-value of less than 0.01. Second, we obtain the parameter estimates to form the risk categories by applying the two-step estimation procedures. The corresponding 95% confidence intervals are generated using the m-out-of-n bootstrap. The results are summarized in Table 4. All estimates are significant and included in the final model. The change point in Table 4 refers to the estimated cut-off, which is used to form the risk categories (high vs. low) based on this linear combination for each individual subject. Based on these risk categories, the regression coefficient estimates are summarized in Table 5. Except for hypertension, all the other covariates have statistically significant effects. The hazard ratio of CHD for the low risk group $I(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X} > 0)$ versus the high-risk group $I(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X} < 0)$ is 0.652. To show the survival functions of these two risk groups, we show the Kaplan-Meier

Table 3. Type-I Errors and Power for SUP Tests for the Existence of the Change Hyperplane ($\times 10^{-2}$)

			Sample Size		
(β_{20}, β_{30})	Censoring Rate	Test	200	300	400
$\beta_{20} = \beta_{30} = 0$	10%	$SUP5^2$	5.6	5.0	5.1
		$SUP10^2$	5.1	5.3	5.2
		$SUP20^2$	4.9	5.3	5.1
	30%	$SUP5^2$	5.4	4.8	5.3
		$SUP10^2$	5.2	5.1	5.4
		$SUP20^2$	4.9	5.8	5.4
	50%	$SUP5^2$	5.4	4.9	5.1
		$SUP10^2$	5.5	5.0	5.2
		$SUP20^2$	5.1	5.5	5.2
$\beta_{20} = 0.8, \beta_{30} = -0.4$	10%	$SUP5^2$	14.4	26.0	29.4
		$SUP10^2$	71.8	84.6	97.2
		$SUP20^2$	74.8	94.0	99.6
	30%	$SUP5^2$	11.0	28.0	29.4
		$SUP10^2$	70.0	85.2	95.8
		$SUP20^2$	74.4	94.8	98.8
	50%	$SUP5^2$	9.4	19.6	23.2
		$SUP10^2$	60.0	77.2	90.4
		$SUP20^2$	60.2	87.6	96.0

Table 4. Change-Hyperplane Covariates Coefficient Estimates in the CHS

Change Hyperplane Covariate	Estimate ($\times 10^{-2}$)	95% CI ($\times 10^{-2}$)
HDL	67.1	[33.8, 100.3]
SBP	−60.4	[−79.6, −41.2]
CHOL	−43.1	[−81.1, −5.1]
Intercept	−20.9	–

curves in Figure 1.

6. Discussion

Although a number of approaches have been developed to estimate change points based on a single covariate, no rigorous theory has been developed for a change hyperplane based on multiple covariates. In this study, we developed a novel two-step approach to estimate the change-hyperplane parameters and a testing procedure to verify the existence of a change hyperplane for univariate survival data. We have developed an adaptive m-out-of-n bootstrap to construct

Table 5. Regression Coefficient Estimates in CHS

	Estimate ($\times 10^{-2}$)	exp(Est) ($\times 10^{-2}$)	p$-$value ($\times 10^{-2}$)
Age	7.1	107.3	< 1
Change Hyperplane	-42.8	65.2	< 1
Diabetes	38.5	146.9	< 1
Smoke	31.5	137.0	< 1
Hypertension	2.7	102.8	70.7

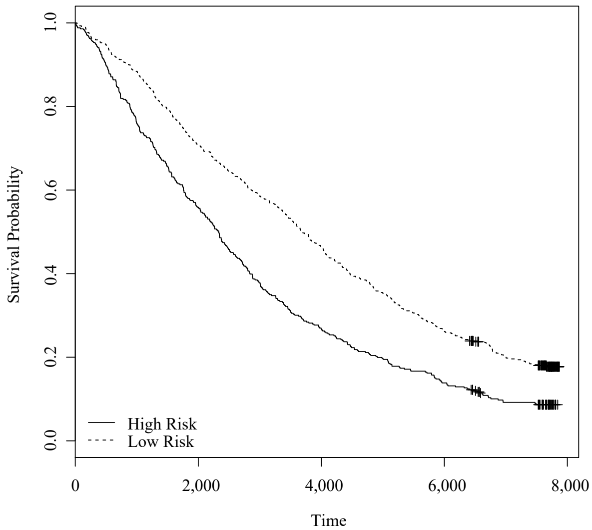

Figure 1. The Kaplan–Meier plot of the risk groups based on the change hyperplane (logrank test: p< 0.001).

the confidence interval, and provide an easy way to determine the appropriate m. We proved the asymptotic properties of the proposed change-hyperplane estimators. To the best of our knowledge, no previous works have derived an asymptotic distribution for a change-plane estimator. As shown in our simulation studies, the estimator is approximately unbiased and its confidence interval has a good coverage rate.

For the proposed test procedure, there is no general rule for choosing the number of grids k. The SUP test based on a larger k is likely to detect a change hyperplane under the alternative, and so may lead to greater power. However, for a fixed sample size, a larger k introduces greater variability into the test, which may reduce the power. Our numerical experience suggests $k = 10$ is a reasonable choice in terms of both the type-I errors and the power, but a more thorough investigation into the choice k is warranted.

We have considered the situation in which the linear combination of the multiple risk factors has only one change point. In reality, the change hyperplane may have multiple change points. Instead of categorizing the participants into low and high risk groups, we may further define a moderate risk group. In this situation, the inference procedures and the asymptotic properties cannot be extended directly to the change hyperplane with multiple thresholds. Thus, it is essential to devise valid and efficient inference procedures for general change-hyperplane models. Moreover, when the proportional hazards assumption is violated, we could extend the change-hyperplane model to other survival models, such as the additive hazard models and accelerated failure-time model. Such an extension will have wide application in univariate survival analysis.

Appendix

A. Proof of Theorems

An equivalent constraint for $\eta_1^2 + \eta_2^2 + \cdots + \eta_{p_1}^2 = 1$ with $\eta_1 > 0$ is to only restrict $\eta_1 = 1$. The maximum likelihood estimator for η_j under this new constraint is 1 for $j = 1$ and is $\widehat{\eta}_j/\widehat{\eta}_1$ for $j > 1$. The following proofs assumes this new equivalent constraint.

For convenience, we define $V_\delta(\boldsymbol{\eta}_0) = \{\boldsymbol{\eta} : \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| < \delta\}$, $V_\epsilon(\boldsymbol{\beta}_0) = \{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < \epsilon\}$,

$$\boldsymbol{s}^{(r)+}(t;\boldsymbol{\eta},\boldsymbol{\beta}) = \mathrm{E}\left\{Y(t)I(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X} > 0)\boldsymbol{Z}^{\otimes r}(t)e^{\boldsymbol{\beta}_1^{\mathrm{T}}\boldsymbol{Z}(t)+\beta_2+\boldsymbol{\beta}_3^{\mathrm{T}}\boldsymbol{Z}(t)}\Big|\boldsymbol{X}\right\},$$
$$\boldsymbol{s}^{(r)-}(t;\boldsymbol{\eta},\boldsymbol{\beta}) = \mathrm{E}\left\{Y(t)I(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X} \le 0)\boldsymbol{Z}^{\otimes r}(t)e^{\boldsymbol{\beta}_1^{\mathrm{T}}\boldsymbol{Z}(t)}\Big|\boldsymbol{X}\right\},$$

where $r = 0, 1, 2$.

Proof of Theorem 1. Suppose that two set of parameters, $(\boldsymbol{\eta}, \boldsymbol{\beta}, \lambda_0)$ and $(\widetilde{\boldsymbol{\eta}}, \widetilde{\boldsymbol{\beta}}, \widetilde{\lambda}_0)$, give the same likelihood functions. We set $\Delta = 1$ then after integrating the likelihood function from 0 to t, we obtain

$$\begin{aligned}&\int_0^{\mathrm{T}} \lambda_0(s)\exp\{\boldsymbol{\beta}_1^{\mathrm{T}}\boldsymbol{Z}(s) + \beta_2 I(X_1 + \eta_2 X_2 + \cdots + \eta_{p_1}X_{p_1} > \eta_0)\\ &\quad +\boldsymbol{\beta}_3^{\mathrm{T}}\boldsymbol{Z}(s)I(X_1 + \eta_2 X_2 + \cdots + \eta_{p_1}X_{p_1} > \eta_0)\}\,ds\\ &= \int_0^{\mathrm{T}} \widetilde{\lambda}_0(s)\exp\Big\{\widetilde{\boldsymbol{\beta}}_1^{\mathrm{T}}\boldsymbol{Z}(s) + \widetilde{\beta}_2 I(X_1 + \widetilde{\eta}_2 X_2 + \cdots + \widetilde{\eta}_{p_1}X_{p_1} > \widetilde{\eta}_0)\\ &\quad +\widetilde{\boldsymbol{\beta}}_3^{\mathrm{T}}\boldsymbol{Z}(s)I(X_1 + \widetilde{\eta}_2 X_2 + \cdots + \widetilde{\eta}_{p_1}X_{p_1} > \widetilde{\eta}_0)\Big\}\,ds.\end{aligned}$$

Thus, letting $X_2 = \cdots = X_{p_1} = 0$, we have

$$\log\lambda_0(s)+\boldsymbol{\beta}_1^{\mathrm T}\boldsymbol{Z}(s)+\beta_2 I(X_1>\eta_0)+\boldsymbol{\beta}_3^{\mathrm T}\boldsymbol{Z}(s)I(X_1>\eta_0)$$
$$=\log\widetilde{\lambda}_0(s)+\widetilde{\boldsymbol{\beta}}_1^{\mathrm T}\boldsymbol{Z}(s)+\widetilde{\beta}_2 I(X_1>\widetilde{\eta}_0)+\widetilde{\boldsymbol{\beta}}_3^{\mathrm T}\boldsymbol{Z}(s)I(X_1>\widetilde{\eta}_0)$$

for $s\in[0,\tau]$.

If $\eta_0\neq\widetilde{\eta}_0$, without loss of generality, we assume $\eta_0>\widetilde{\eta}_0$ then choose X_1 to be a value larger than η_0 and another value between $\widetilde{\eta}_0$ and η_0. We obtain

$$\log\lambda_0(s)+\boldsymbol{\beta}_1^{\mathrm T}\boldsymbol{Z}(s)+\beta_2+\boldsymbol{\beta}_3^{\mathrm T}\boldsymbol{Z}(s)=\log\widetilde{\lambda}_0(s)+\widetilde{\boldsymbol{\beta}}_1^{\mathrm T}\boldsymbol{Z}(s)+\widetilde{\beta}_2+\widetilde{\boldsymbol{\beta}}_3^{\mathrm T}\boldsymbol{Z}(s)$$

and

$$\log\lambda_0(s)+\boldsymbol{\beta}_1^{\mathrm T}\boldsymbol{Z}(s)=\log\widetilde{\lambda}_0(s)+\widetilde{\boldsymbol{\beta}}_1^{\mathrm T}\boldsymbol{Z}(s)+\widetilde{\beta}_2+\widetilde{\boldsymbol{\beta}}_3^{\mathrm T}\boldsymbol{Z}(s).$$

This gives $\beta_2+\boldsymbol{\beta}_3^{\mathrm T}\boldsymbol{Z}(s)=0$ for all $s\in[0,\tau]$ so $\beta_2=0$ and $\boldsymbol{\beta}_3=0$ by condition (C.2). This gives a contradiction to the condition in Theorem 3.1. We conclude $\eta_0=\widetilde{\eta}_0$. This further gives

$$\log\lambda_0(s)+\beta_2 I(X_1>\eta_0)=\log\widetilde{\lambda}_0(s)+\widetilde{\beta}_2 I(X_1>\eta_0)$$

and

$$\boldsymbol{\beta}_1+\boldsymbol{\beta}_3 I(X_1>\eta_0)=\widetilde{\boldsymbol{\beta}}_1+\widetilde{\boldsymbol{\beta}}_3 I(X_1>\eta_0).$$

We immediately conclude $\lambda_0(s)=\widetilde{\lambda}_0(s),\beta_2=\widetilde{\beta}_2$, $\boldsymbol{\beta}_1=\widetilde{\boldsymbol{\beta}}_1$ and $\boldsymbol{\beta}_3=\widetilde{\boldsymbol{\beta}}_3$.

This further gives

$$I(X_1+\eta_2X_2+\cdots+\eta_{p_1}X_{p_1}>\eta_0)=I(X_1+\widetilde{\eta}_2X_2+\cdots+\widetilde{\eta}_{p_1}X_{p_1}>\eta_0).$$

For fixed $X_2,\cdots,X_{p_1}$, the same arguments as before yield

$$\eta_2X_2+\cdots+\eta_{p_1}X_{p_1}=\widetilde{\eta}_2X_2+\cdots+\widetilde{\eta}_{p_1}X_{p_1}$$

so it holds $\eta_j=\widetilde{\eta}_j$ for $j=2,\ldots,p_1$. Theorem 1 is proved.

Proof of Theorem 2. To prove the consistency, since the class

$$[r_{\boldsymbol{\eta},\boldsymbol{\beta}}\{\boldsymbol{W}(t)\}:\eta_1=1,\|\boldsymbol{\beta}\|\le B]$$

is a P-Donsker so P-Glivenko-Cantelli class (van der Vaart and Wellner (1996)), it holds

$$\sup_{\boldsymbol{\eta},\|\boldsymbol{\beta}\|\le B}\left|n^{-1}l_n(\boldsymbol{\eta},\boldsymbol{\beta})+\log n-l(\boldsymbol{\eta},\boldsymbol{\beta})\right|\to 0$$

almost surely, where

$$l(\boldsymbol{\eta},\boldsymbol{\beta})=E\left[\Delta\log r_{\boldsymbol{\eta},\boldsymbol{\beta}}\{\boldsymbol{W}(T)\}-\widetilde{E}(I(\widetilde{T}\ge T)\exp[r_{\boldsymbol{\eta},\boldsymbol{\beta}}\{\widetilde{\boldsymbol{W}}(T)\}])\right],$$

where $\widetilde{E}$ is the expectation with respect to $(\widetilde{T}, \widetilde{\boldsymbol{W}})$, which is an independent copy of $(T, \boldsymbol{W})$.

Note that $l(\boldsymbol{\eta}, \boldsymbol{\beta}) \leq l(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)$ based on the standard result for the Cox partial likelihood theory. Furthermore, the equality holds if and only if there exits some $\lambda(t)$ such that the two sets of parameters, $(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0, \lambda_0)$ and $(\boldsymbol{\eta}, \boldsymbol{\beta}, \lambda)$, give the same likelihood functions. However, Theorem 1 implies that the equality holds if and only if $\boldsymbol{\eta}_0 = \boldsymbol{\eta}$ and $\boldsymbol{\beta}_0 = \boldsymbol{\beta}$. In other words, $l(\boldsymbol{\eta}, \boldsymbol{\beta})$ has the unique maximum at $(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)$. By Theorem 5.9 (van der Vaart (1998)), we conclude that $(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\beta}})$ converges to $(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)$ almost surely. Thus, Theorem 2 holds.

Proof of Theorem 3. First, we define $U_\epsilon(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0) = \{(\boldsymbol{\eta}, \boldsymbol{\beta}) : A < n^{1/2}(\|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2)^{1/2} \leq n^{1/2}\epsilon\}$ and $V_\epsilon(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0) = \{(\boldsymbol{\eta}, \boldsymbol{\beta}) : (\|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2)^{1/2} < \epsilon\}$, for a given ϵ. From Theorem 2, $P_0\{(\boldsymbol{\eta}, \boldsymbol{\beta}) \in V_\epsilon(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)\} > 1-\zeta$ for any $\zeta > 0$, when n is large enough. Hence,

$$\begin{aligned}
&P_0\left\{n^{1/2}\left(\|\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\| + \left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2\right)^{1/2} > A\right\} \\
&= P_0\left\{(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\beta}}) \in U_\epsilon(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)\right\} + P_0\left\{(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\beta}}) \in V_\epsilon^C(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)\right\} \\
&\leq P_0\left\{\sup_{\boldsymbol{\eta}, \boldsymbol{\beta} \in U_\epsilon(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)} L_n(\boldsymbol{\eta}, \boldsymbol{\beta}) \geq L_n(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)\right\} + \zeta \\
&= P_0\left\{\sup_{\boldsymbol{\eta}, \boldsymbol{\beta} \in U_\epsilon(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)} G_n(\boldsymbol{\eta}, \boldsymbol{\beta}) \geq 0\right\} + \zeta,
\end{aligned}$$

where $G_n(\boldsymbol{\eta}, \boldsymbol{\beta}) = \log L_n(\boldsymbol{\eta}, \boldsymbol{\beta}) - \log L_n(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)$. Let $G(\boldsymbol{\eta}, \boldsymbol{\beta})$ be the expectation of $G_n(\boldsymbol{\eta}, \boldsymbol{\beta})$. The Taylor expression gives

$$G(\boldsymbol{\eta}, \boldsymbol{\beta}) = \dot{G}_{\boldsymbol{\eta}}(\boldsymbol{\eta}, \boldsymbol{\beta})(\boldsymbol{\eta} - \boldsymbol{\eta}_0)^{\mathrm{T}} - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^{\mathrm{T}}\boldsymbol{I}(\boldsymbol{\eta}^*, \boldsymbol{\beta}^*)(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o(1),$$

where $\boldsymbol{\beta}^*$ is between $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$. The second order term in the expansion is due to the fact that the second order derivatives of the observed log likelihood function at the true value converges to the true negative information matrix by the strong law of large numbers. By linearization, we can show that $\dot{G}_{\boldsymbol{\eta}}(\boldsymbol{\eta}, \boldsymbol{\beta})(\boldsymbol{\eta} - \boldsymbol{\eta}_0)^{\mathrm{T}}$ is negative. In addition, the matrix $I(\boldsymbol{\eta}^*, \boldsymbol{\beta}^*)$ is positive definite by (C.3). Therefore, there exists a positive constant k_0 which ensures $G(\boldsymbol{\eta}, \boldsymbol{\beta}) \leq -k_0(\|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2)$. Additionally, we split $U_\epsilon(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)$ into subsets

$$H_{n,j} = \left\{ (\boldsymbol{\eta}, \boldsymbol{\beta}) : g(j) \le n^{1/2} \left(\|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 \right)^{1/2} < g(j+1) \right\},$$

where $g(j) = 2^j$, and $j = 1, 2, \ldots$. Similar to Lemma 3 in Pons (2003), there exists a constant $k > 0$ such that for any $\widetilde{\epsilon}$, $\mathrm{E} \sup_{(\boldsymbol{\eta},\boldsymbol{\beta}) \in V_{\widetilde{\epsilon}}(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)} \left| n^{1/2} \{G_n(\boldsymbol{\eta}, \boldsymbol{\beta}) - G(\boldsymbol{\eta}, \boldsymbol{\beta})\} \right| \le k\widetilde{\epsilon}$ as $n \to \infty$. Thus, we obtain

$$\begin{aligned}
&\limsup_n \sum_{j:g(j)>A} P_0 \left[\sup_{H_{n,j}} n^{1/2} \{G_n(\boldsymbol{\eta}, \boldsymbol{\beta}) - G(\boldsymbol{\eta}, \boldsymbol{\beta})\} \ge n^{-1/2} g^2(j) k_0 \right] \\
&\le \limsup_n \sum_{j:g(j)>A} \frac{\mathrm{E}\left[\sup_{H_{n,j}} n \{G_n(\boldsymbol{\eta}, \boldsymbol{\beta}) - G(\boldsymbol{\eta}, \boldsymbol{\beta})\} \right]^2}{g^4(j) k_0^2} \\
&\le \sum_{j:g(j)>A} \frac{k^2 g^2(j+1)}{k_0^2 g^4(j)} \to 0,
\end{aligned}$$

as A goes to infinity. Hence, it gives $\lim_A \limsup_n P_0\{n^{1/2}(\|\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\| + \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2)^{1/2} > A\} = 0$. Theorem 3 has been proved.

Proof of Theorem 4. Let $\boldsymbol{\eta}_0 = \boldsymbol{\eta}_{n,u_1} + n^{-1}\boldsymbol{u}_1$, $\boldsymbol{\beta}_0 = \boldsymbol{\beta}_{n,u_2} + n^{-1/2}\boldsymbol{u}_2$, and $W_{i0} = \boldsymbol{\eta}_0^{\mathrm{T}} \boldsymbol{X}_i$, where $\boldsymbol{u}_1 = (a_1, a_2, \ldots, a_{p_1})^{\mathrm{T}}$ and $\boldsymbol{u}_2 = (b_1, b_2, \ldots, b_{2p_2+1})^{\mathrm{T}}$ assumed to have norm bounded by a large constant A. Note that from Theorem 3, the probability $n(\boldsymbol{\eta}_0 - \widehat{\boldsymbol{\eta}})$ and $\sqrt{n}(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}})$ bounded by A tends to 1 when A diverges.

First, after some algebra, we can rewrite $l_n(\boldsymbol{\eta}_{n,u_1}, \boldsymbol{\beta}_{n,u_2}) - l_n(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)$ as

$$\begin{aligned}
&l_n(\boldsymbol{\eta}_{n,u_1}, \boldsymbol{\beta}_{n,u_2}) - l_n(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0) \\
&= (\boldsymbol{\beta}_{n,u_2} - \boldsymbol{\beta}_0)^{\mathrm{T}} \left\{ \sum_{i=1}^n \int_0^\tau \widetilde{\boldsymbol{Z}}_i(t; \boldsymbol{\eta}_{n,u_1}) dN_i(t) \right\} \\
&\quad - \int_0^\tau \log \left\{ \frac{S_n^{(0)}(t; \boldsymbol{\eta}_{n,u_1}, \boldsymbol{\beta}_{n,u_2})}{S_n^{(0)}(t; \boldsymbol{\eta}_0, \boldsymbol{\beta}_0)} \right\} d\bar{N}_n(t) \\
&\quad + \sum_{i=1}^n \Delta_i \{\beta_{20} + \boldsymbol{\beta}_{30}^{\mathrm{T}} \boldsymbol{Z}_i(T_i)\} \left\{ I\left(0 \ge W_{i0} > n^{-1} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u}_1\right) I(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u}_1 < 0) \right. \\
&\quad \left. - I\left(0 < W_{i0} \le n^{-1} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u}_1\right) I(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u}_1 \ge 0) \right\},
\end{aligned}$$

where $\bar{N}_n(t) = \sum_{i=1}^n \Delta_i I(T_i \le t)$,

$$S_n^{(k)}(t; \boldsymbol{\eta}, \boldsymbol{\beta}) \equiv n^{-1} \sum_{i=1}^n Y_i(t) \boldsymbol{Z}_i^{\otimes k}(t) e^{r_{\boldsymbol{\eta},\boldsymbol{\beta}}(\boldsymbol{W}_i(t))}$$

for $k=0,1$, and $W_{i0}=\boldsymbol{\eta}_0^{\mathrm{T}}\boldsymbol{X}_i$. By the Taylor expansion for $\boldsymbol{\beta}_{n,u_2}$ at $\boldsymbol{\beta}_0$,

$$\log\left\{\frac{S_n^{(0)}(t;\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_{n,u_2})}{S_n^{(0)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}\right\}=\frac{S_n^{(0)}(t;\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_0)-S_n^{(0)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}{S_n^{(0)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}$$
$$-n^{-1/2}\boldsymbol{u}_2^{\mathrm{T}}\frac{\boldsymbol{S}_n^{(1)}(t;\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_0)}{S_n^{(0)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}$$
$$+\frac{n^{-1}}{2}\boldsymbol{u}_2^{\mathrm{T}}\boldsymbol{V}_n(t;\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_0)\boldsymbol{u}_2+o_p(n^{-1}),$$

where $\boldsymbol{V}_n(t;\boldsymbol{\eta},\boldsymbol{\beta})=\boldsymbol{S}_n^{(2)}(t;\boldsymbol{\eta},\boldsymbol{\beta})/S_n^{(0)}(t;\boldsymbol{\eta},\boldsymbol{\beta})-\{\boldsymbol{S}_n^{(1)}(t;\boldsymbol{\eta},\boldsymbol{\beta})/S_n^{(0)}(t;\boldsymbol{\eta},\boldsymbol{\beta})\}^{\otimes 2}$ and $o_p(\cdot)$, here and in the sequel, denotes the sequence of random variables converging uniformly in $\boldsymbol{u}_1,\boldsymbol{u}_2$ in any bounded set. Thus, we have

$$l_n(\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_{n,u_2})-l_n(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)=Q_n(\boldsymbol{u}_1)-\boldsymbol{u}_2^{\mathrm{T}}\boldsymbol{C}_n(\boldsymbol{u}_1)-\frac{1}{2}\boldsymbol{u}_2^{\mathrm{T}}\boldsymbol{I}(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)\boldsymbol{u}_2+o_p(n^{-1}),$$

where

$$Q_n(\boldsymbol{u}_1)=\sum_{i=1}^n\Delta_i\Bigg[\left\{\beta_{20}+\boldsymbol{\beta}_{30}^{\mathrm{T}}\boldsymbol{Z}_i(T_i)\right\}\times\left\{I\left(0\ge W_{i0}>n^{-1}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{u}_1,\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{u}_1<0\right)\right.$$
$$\left.-I\left(0<W_{i0}\le n^{-1}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{u}_1,\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{u}_1\ge 0\right)\right\}$$
$$-\frac{S_n^{(0)}(T_i;\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_0)-S_n^{(0)}(T_i;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}{S_n^{(0)}(T_i;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}\Bigg],$$

and

$$\boldsymbol{C}_n(\boldsymbol{u}_1)=n^{-1/2}\sum_{i=1}^n\int_0^\tau\left\{\widetilde{\boldsymbol{Z}}_i(t;\boldsymbol{\eta}_{n,u_1})-\frac{\boldsymbol{S}_n^{(1)}(t;\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_0)}{S_n^{(0)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}\right\}dM_i(t)$$
$$+n^{-1/2}\int_0^\tau\sum_{i=1}^n\widetilde{\boldsymbol{Z}}_i(t;\boldsymbol{\eta}_{n,u_1})\left(\exp\left[r_{\boldsymbol{\eta}_0,\boldsymbol{\beta}_0}\left\{\boldsymbol{W}_i(t)\right\}\right]\right.$$
$$\left.-\exp\left[r_{\eta_{n,u_1},\beta_0}\left\{\boldsymbol{W}_i(t)\right\}\right]\right)d\Lambda_0(t).$$

Using the uniform convergence property for the martingale process and noting

$$n^{-1/2}\sum_{i=1}^n\widetilde{\boldsymbol{Z}}_i(t;\boldsymbol{\eta}_{n,u_1})\left(\exp\left[r_{\boldsymbol{\eta}_0,\boldsymbol{\beta}_0}\left\{\boldsymbol{W}_i(t)\right\}\right]-\exp\left[r_{\eta_{n,u_1},\beta_0}\left\{\boldsymbol{W}_i(t)\right\}\right]\right)$$

converges to 0 uniformly in t, we obtain that $\boldsymbol{C}_n(\boldsymbol{u}_1)$ is asymptotically equivalent to

$$\widetilde{\boldsymbol{l}}_n = n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{\widetilde{\boldsymbol{Z}}_i(t;\boldsymbol{\eta}_0) - \frac{\boldsymbol{S}_n^{(1)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}{S_n^{(0)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}\right\} dM_i(t)$$

in probability, uniformly for $\boldsymbol{u}_1$ with $\|\boldsymbol{u}_1\| \le A$ for the given constant A. Then we have

$$l_n(\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_{n,u_2}) - l_n(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0) = Q_n(\boldsymbol{u}_1) - \boldsymbol{u}_2^{\mathrm{T}}\widetilde{\boldsymbol{l}}_n - \frac{1}{2}\boldsymbol{u}_2^{\mathrm{T}}\boldsymbol{I}(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)\boldsymbol{u}_2 + o_p(1).$$

Next, we derive the asymptotic distributions of $Q_n(\boldsymbol{u}_1)$ and $\widetilde{\boldsymbol{l}}_n$. Clearly, the variable $-\widetilde{\boldsymbol{l}}_n$ converges weakly to a Gaussian variable following the normal distribution $\mathcal{Z} = N(\boldsymbol{0},\boldsymbol{I}(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)^{-1})$. Thus, if we can prove that $Q_n(\boldsymbol{u}_1)$ converges to a tight process, say, $Q(\boldsymbol{u}_1)$, then the argmax mapping theorem gives that the maximizer for $l_n(\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_{n,u_2}) - l_n(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)$, i.e., $\{n(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta}), \sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\}$ converges in distribution to the maximizer for the limiting process, $Q(\boldsymbol{u}_1) + \boldsymbol{u}_2^{\mathrm{T}}\mathcal{Z} - (1/2)\boldsymbol{u}_2^{\mathrm{T}}\boldsymbol{I}(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)\boldsymbol{u}_2$, which is

$$\{\operatorname{argmax} Q(\boldsymbol{u}_1), \boldsymbol{I}(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)^{-1}\mathcal{Z}\}.$$

Furthermore, it is clear that the latter two random variables are independent. We then obtain the theorem.

It remains to show that $Q_n(\boldsymbol{u}_1)$ converges weakly to $Q(\boldsymbol{u}_1)$ in the Skorohod space in $\boldsymbol{u}_1$. First,

$$\int\left\{S_n^{(0)}(t;\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_0) - S_n^{(0)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)\right\}\left\{\frac{d\bar{N}_n(t)}{nS_n^{(0)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)} - d\Lambda_0(t)\right\} = o_p(1)$$

and

$$\begin{aligned}
&S_n^{(0)}(t;\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_0) - S_n^{(0)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)\\
&= -n^{-1}\sum_{i=1}^{n} Y_i(t)\exp\left\{\boldsymbol{\beta}_{10}^{\mathrm{T}}\boldsymbol{Z}_i(t)\right\}\left[1-\exp\left\{\beta_{20}+\boldsymbol{\beta}_{30}^{\mathrm{T}}\boldsymbol{Z}_i(t)\right\}\right]\\
&\quad\times I\left(0 \ge W_{i0} > n^{-1}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{u}_1, \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{u}_1 < 0\right)\\
&\quad + n^{-1}\sum_{i=1}^{n} Y_i(t)\exp\left\{\boldsymbol{\beta}_{10}^{\mathrm{T}}\boldsymbol{Z}_i(t)\right\}\left[1-\exp\left\{\beta_{20}+\boldsymbol{\beta}_{30}^{\mathrm{T}}\boldsymbol{Z}_i(t)\right\}\right]\\
&\quad\times I\left(0 < W_{i0} \le n^{-1}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{u}_1, \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{u}_1 \ge 0\right).
\end{aligned}$$

We then obtain

$$Q_n(\boldsymbol{u}_1) = Q_n^{+}(\boldsymbol{u}_1) - Q_n^{-}(\boldsymbol{u}_1) + o_p(1),$$

where

$$\begin{aligned}Q_n^-(\boldsymbol{u}_1) =& \sum_{i=1}^{n}\left(\Delta_i\left[\left\{\beta_{20}+\boldsymbol{\beta}_{30}^{\mathrm{T}}\boldsymbol{Z}_i(T_i)\right\}\right]+\int_0^{\tau}\phi_i(t)d\Lambda_0(t)\right)\\ &\times I\left(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{u}_1 \geq 0, 0 < W_{i0} \leq n^{-1}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{u}_1\right),\\ Q_n^+(\boldsymbol{u}_1) =& \sum_{i=1}^{n}\left(\Delta_i\left[\left\{\beta_{20}+\boldsymbol{\beta}_{30}^{\mathrm{T}}\boldsymbol{Z}_i(T_i)\right\}\right]+\int_0^{\tau}\phi_i(t)d\Lambda_0(t)\right)\\ &\times I\left(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{u}_1 < 0, 0 \geq W_{i0} > n^{-1}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{u}_1\right)\end{aligned}$$

with

$$\phi_i(t) = Y_i(t)\exp\left\{\boldsymbol{\beta}_{10}^{\mathrm{T}}\boldsymbol{Z}_i(t)\right\}\left[1-\exp\left\{\beta_{20}+\boldsymbol{\beta}_{30}^{\mathrm{T}}\boldsymbol{Z}_i(t)\right\}\right].$$

Next, we aim to determine the asymptotic process for $Q_n(\boldsymbol{u}_1)$, which can be viewed as a random process on the Skorohod space in R^{p_1}. To this end, we first show that the finite dimensional convergence holds for $Q_n^-(\boldsymbol{u}_1)$ (the same holds for $Q_n^+(\boldsymbol{u}_1)$), and we will identify its limit process based on this finite dimensional convergence. Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_S$ be a sequence of vectors then we wish to obtain the limit distribution of any linear combination $\sum_{s=1}^{S} q_s Q_n^-(\boldsymbol{v}_s)$, where $q_1, q_2, \ldots, q_S$ are any fixed constants. Let

$$\eta = \Delta\left(\left\{\beta_{20}+\boldsymbol{\beta}_{30}^{\mathrm{T}}\boldsymbol{Z}(T)\right\}\right)+\int_0^{\tau}\phi(t)d\Lambda_0(t).$$

We let $H^{(1)}, \ldots, H^{(S)}$ be the ordered statistic of $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{v}_1, \ldots, \boldsymbol{X}^{\mathrm{T}}\boldsymbol{v}_S$, i.e., $H^{(s)} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{v}^{(s)}$. Correspondingly, we let $q_{(1)}, \ldots, q_{(S)}$ be the corresponding sequence of $q_1, \ldots, q_S$. We then define set $A_s = \left\{H^{(s-1)} < 0 < H^{(s)}\right\}$ for $1 \leq s \leq S$ and let A_0 be the set of $H^{(S)} \leq 0$. We have that the characteristic function for $\sum_{s=1}^{S} q_s Q_n^-(\boldsymbol{v}_s)$ is given by

$$\begin{aligned}&\mathrm{E}\left\{\exp\left(i\tilde{t}\sum_{s=1}^{S}q_sQ_n^-(\boldsymbol{v}_s)\right)\right\}\\ &= \left(P(A_0)+\sum_{s=1}^{S}P(A_s)\left[E\left\{I\left(0<W<\frac{H^{(s)}}{n}\right)e^{(q_{(s)}+\cdots+q_{(S)})i\tilde{t}\eta}\middle|A_s\right\}\right.\right.\\ &\quad +E\left\{I\left(\frac{H^{(s)}}{n}\leq W<\frac{H^{(s+1)}}{n}\right)e^{(q_{(s+1)}+\cdots+q_{(S)})i\tilde{t}\eta}\middle|A_s\right\}+\cdots\\ &\quad \left.\left.+E\left\{I\left(\frac{H^{(S-1)}}{n}\leq W<\frac{H^{(S)}}{n}\right)e^{q_{(S)}i\tilde{t}\eta}\middle|A_s\right\}\right]\right)^n\end{aligned}$$

$$= \left\{1 + \sum_{s=1}^{S} P(A_s)\left(E\left[I\left(0 < W < \frac{H^{(s)}}{n}\right)\{e^{(q_{(s)}+\cdots+q_{(S)})i\tilde{t}\eta} - 1\}\Big| A_s\right]\right.\right.$$
$$+E\left[I\left(\frac{H^{(s)}}{n} \le W < \frac{H^{(s+1)}}{n}\right)\{e^{(q_{(s+1)}+\cdots+q_{(S)})i\tilde{t}\eta} - 1\}\Big| A_s\right] + \cdots$$
$$\left.\left.+E\left[I\left(\frac{H^{(S-1)}}{n} \le W < \frac{H^{(S)}}{n}\right)\{e^{q_{(S)}i\tilde{t}\eta} - 1\}\Big| A_s\right]\right)\right\}^n.$$

Since

$$P(A_s)E\left[I\left(\frac{H^{(s)}}{n} \le W < \frac{H^{(s+1)}}{n}\right)\{e^{(q_{(s+1)}+\cdots+q_{(S)})i\tilde{t}\eta} - 1\}\Big| A_s\right]$$
$$= n^{-1}E\left[(H^{(s+1)} - H^{(s)})I(A_s)\{e^{(q_{(s+1)}+\cdots+q_{(S)})i\tilde{t}\eta} - 1\}\Big| W = 0\right] f_W(0) + O(n^{-2}),$$

we conclude that

$$\mathrm{E}\left[\exp\left\{i\tilde{t}\sum_{s=1}^{S} q_s Q_n^-(\boldsymbol{v}_s)\right\}\right]$$
$$= \left\{1 + n^{-1} f_W(0)\sum_{s=1}^{S}\left(E\left[H^{(s)}I(A_s)\{e^{(q_{(s)}+\cdots+q_{(S)})i\tilde{t}\eta} - 1\}\Big| W = 0\right]\right.\right.$$
$$+E\left[(H^{(s+1)} - H^{(s)})I(A_s)\{e^{(q_{(s+1)}+\cdots+q_{(S)})i\tilde{t}\eta} - 1\}\Big| W = 0\right] + \cdots$$
$$\left.\left.+E\left[(H^{(S)} - H^{(S-1)})I(A_s)\{e^{q_{(S)}i\tilde{t}\eta} - 1\}\Big| W = 0\right]\right) + O(n^{-2})\right\}^n$$

so it converges to

$$\exp\left\{f_W(0)\sum_{s=1}^{S}\sum_{k=s}^{S}\left(E\left[(H^{(k)} - H^{(k-1)})I(A_s)\{e^{(q_{(k)}+\cdots+q_{(S)})i\tilde{t}\eta} - 1\}\Big| W = 0\right]\right)\right\}.$$

We want to show that the limit distribution of $\sum_{s=1}^{S} q_s Q_n^-(\boldsymbol{v}_s)$ is the same as $\sum_{s=1}^{S} q_s Q^-(\boldsymbol{v}_s)$. Similarly, let $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}_{(k)}, k = 1, \ldots, S$ denote the ordered value for $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}_k, k = 1, \ldots, S$ and A_s denotes the set of $\boldsymbol{x}$ for which 0 is between $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}_{(s-1)}$ and $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}_{(s)}$. To this end, we note

$$E\left[\exp\{i\tilde{t}\sum_{s=1}^{S} q_s Q^-(\boldsymbol{v}_s)\}\right]$$
$$= E\left\{\exp\left(i\tilde{t}\sum_{s=1}^{S} q_s \int_{\Omega}\left[I\{\boldsymbol{X}(\omega)^{\mathrm{T}}\boldsymbol{v}_s > 0\}\int_0^{\boldsymbol{X}(\omega)^{\mathrm{T}}\boldsymbol{v}_s}\Gamma(\boldsymbol{X}(\omega), t)v(d\omega, dt)\right]\right)\right\}$$

$$
\begin{aligned}
&= E\left(\exp\left[i\widetilde{t}\sum_{s=1}^{S}\int_{\Omega} I\{\boldsymbol{X}(\omega)\in A_s\}\right.\right.\\
&\quad\left.\left.\times\sum_{k=s}^{S} q_{(k)}\int_{0}^{\boldsymbol{X}(\omega)^{\mathrm{T}}\boldsymbol{v}_{(k)}}\Gamma\{\boldsymbol{X}(\omega),t\}v(d\omega,dt)\right]\right)\\
&= \prod_{s=1}^{S}\prod_{k=s}^{S} E\left(\exp\left[i\widetilde{t}\int_{\Omega} I\{\boldsymbol{X}(\omega)\in A_s\}\right.\right.\\
&\quad\left.\left.\times\ (q_{(k)}+\cdots+q_{(S)})\int_{\boldsymbol{X}(\omega)^{\mathrm{T}}\boldsymbol{v}_{(k-1)}}^{\boldsymbol{X}(\omega)^{\mathrm{T}}\boldsymbol{v}_{(k)}}\Gamma\{\boldsymbol{X}(\omega),t\}v(d\omega,dt)\right]\right).
\end{aligned}
$$

Note that the integration inside the above expectation is essentially the discrete summation over ω and t where $v(\omega,t)$ has jumps. Since conditional on that $v(\omega,t)$ has jumps at $(\omega_j,t_j), j=1,\ldots,m$, $\Gamma\{\boldsymbol{X}(\omega),t\}$ is independent for any ω and t, we have

$$
\begin{aligned}
&E\left(\exp\left[i\widetilde{t}\int_{\Omega} I\{\boldsymbol{X}(\omega)\in A_s\}(q_{(k)}+\cdots+q_{(S)})\right.\right.\\
&\quad\left.\left.\int_{\boldsymbol{X}(\omega)^{\mathrm{T}}\boldsymbol{v}_{(k-1)}}^{\boldsymbol{X}(\omega)^{\mathrm{T}}\boldsymbol{v}_{(k)}}\Gamma\{\boldsymbol{X}(\omega),t\}v(d\omega,dt)\right]\right)\\
&= E\left\{\exp\left(i\widetilde{t}\sum_{j} I\{\boldsymbol{X}(\omega_j)\in A_s\}(q_{(k)}+\cdots+q_{(S)})\right.\right.\\
&\quad\left.\left.\times I[t_j\in\{\boldsymbol{X}(\omega_j)^{\mathrm{T}}\boldsymbol{v}_{(k-1)},\boldsymbol{X}(\omega_j)^{\mathrm{T}}\boldsymbol{v}_{(k)}\}]\Gamma\{\boldsymbol{X}(\omega_j),t_j\}\right)\right\}\\
&= E\left\{\prod_{j}\exp\left(i\widetilde{t}I\{\boldsymbol{X}(\omega_j)\in A_s\}\right.\right.\\
&\quad\left.\left.\times\ I[t_j\in\{\boldsymbol{X}(\omega_j)^{\mathrm{T}}\boldsymbol{v}_{(k-1)},\boldsymbol{X}(\omega_j)^{\mathrm{T}}\boldsymbol{v}_{(k)}\}](q_{(k)}+\cdots+q_{(S)})\Gamma\{\boldsymbol{X}(\omega_j),t_j\}\right)\right\}\\
&= E\left[\exp\left\{\sum_{j} I\{\boldsymbol{X}(\omega_j)\in A_s\}\times\ I[t_j\in\{\boldsymbol{X}(\omega_j)^{\mathrm{T}}\boldsymbol{v}_{(k-1)},\boldsymbol{X}(\omega_j)^{\mathrm{T}}\boldsymbol{v}_{(k)}\}]\right.\right.\\
&\quad\left.\left.\times\log E\left(\exp\left[i\widetilde{t}(q_{(k)}+\cdots+q_{(S)})\Gamma\{\boldsymbol{X}(\omega_j),t_j\}\right]\,\middle|\,\boldsymbol{X},w_j,t_j\right)\right\}\right]\\
&= E\left\{\exp\left(\int_{\Omega} I\{\boldsymbol{X}(\omega)\in A_s\}\right.\right.
\end{aligned}
$$

$$\times \int_{\boldsymbol{X}(\omega)^{\mathrm{T}}\boldsymbol{v}_{(k-1)}}^{\boldsymbol{X}(\omega)^{\mathrm{T}}\boldsymbol{v}_{(k)}} \log E\left[e^{i\widetilde{t}(q_{(k)}+\cdots+q_{(S)})\Gamma\{\boldsymbol{X}(\omega),t\}}\right] v(d\omega,dt)\Bigg)\Bigg\}$$

$$= \exp\left[\int_{\Omega} I\{\boldsymbol{X}(\omega)\in A_s\}\int_{\boldsymbol{X}(\omega)^{\mathrm{T}}\boldsymbol{v}_{(k-1)}}^{\boldsymbol{X}(\omega)^{\mathrm{T}}\boldsymbol{v}_{(k)}} \left(E\left[e^{i\widetilde{t}(q_{(k)}+\cdots+q_{(S)})\Gamma\{\boldsymbol{X}(\omega),t\}}\right]-1\right) dP(\omega)dt\right],$$

where the last equality uses the fact that $v(d\omega, dt)$ is independent Poisson with rate $dP(w)dt$. Consequently, since the characteristics function for $\Gamma(\boldsymbol{x}, t)$ is independent of t, we obtain

$$E\left[\exp\left\{i\widetilde{t}\sum_{s=1}^{S} q_s Q^-(\boldsymbol{v}_s)\right\}\right]$$

$$= \prod_{s=1}^{S}\prod_{k=s}^{S} \exp\left[\int_{\Omega} I\{\boldsymbol{X}(\omega)\in A_s\}\right.$$

$$\left.\int_{\boldsymbol{X}(\omega)^{\mathrm{T}}\boldsymbol{v}_{(k-1)}}^{\boldsymbol{X}(\omega)^{\mathrm{T}}\boldsymbol{v}_{(k)}} \left(E\left[e^{i\widetilde{t}(q_{(k)}+\cdots+q_{(S)})\Gamma\{\boldsymbol{X}(\omega),t\}}\Big|\boldsymbol{X}\right]-1\right) dP(\omega)dt\right]$$

$$= \exp\left\{f_W(0)\sum_{s=1}^{S}\sum_{k=s}^{S} E\Big(\{H^{(s)}-H^{(s-1)}\}I(\boldsymbol{X}\in A_s)\right.$$

$$\left.\times\left[E\left\{e^{i\widetilde{t}(q_{(k)}+\cdots+q_{(S)})\Gamma(\boldsymbol{X},t)}\Big|\boldsymbol{X}\right\}-1\right]\Big|W=0\Big)\right\},$$

which is the same as the characteristic function for the limit distribution of $\sum_{s=1}^{S} q_s Q_n^-(\boldsymbol{v}_s)$. Similarly, we apply the same proof to $Q_n^+(\boldsymbol{u}_1)$ (by changing W_{i0} to $-W_{i0}$ and $\boldsymbol{X}_i$ to $-\boldsymbol{X}_i$) to obtain the finite dimensional distribution of $Q_n^+(\boldsymbol{u}_1)$ to the the finite dimensional distribution of $Q^+(\boldsymbol{u}_1)$.

Finally, we can easily show $E[|Q_n^-(\boldsymbol{v}_2)-Q_n^-(\boldsymbol{v}_1)||Q_n^-(\boldsymbol{v}_2)-Q_n^-(\boldsymbol{v}_1)|]$ is bounded by $\|\boldsymbol{v}_2-\boldsymbol{v}_1\|$ times a constant. Thus, the processes Q_n^- is tight so converge weakly to Q^-, using the D-tightness criterion (Billingsley (2009)). Similarly, we can prove that Q_n^+ converges weakly to Q^+ in the Skorohod space. Therefore, $Q_n(\boldsymbol{u}_1)$ converges weakly to $Q(\boldsymbol{u}_1)$. We have completed the proof.

References

Bickel, P. J., Götze, F. and van Zwet, W. R. (2012). *Resampling Fewer Than n Observations: Gains, Losses, and Remedies for Losses*. Springer.

Bickel, P. J. and Sakov, A. (2005). On the choice of m in the m out of n bootstrap and its application to confidence bounds for extreme percentiles. *Unpublished manuscript.*

Bickel, P. J. and Sakov, A. (2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica* **18**, 967–985.

Billingsley, P. (2009). *Convergence of Probability Measures.* John Wiley & Sons.

Cheung, K., Lee, S. M. and Young, G. A. (2005). Iterating the m out of n bootstrap in nonregular smooth function models. *Statistica Sinica* **15**, 945–967.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**, 247–254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74**, 33–43.

Gandy, A. and Jensen, U. (2005). On goodness-of-fit tests for Aalen's additive risk model. *Scandinavian Journal of Statistics* **32**, 425–445.

Gandy, A., Jensen, U. and Lütkebohmert, C. (2005). A Cox model with a change-point applied to an actuarial problem. *Brazilian Journal of Probability and Statistics* **19**, 93–109.

Hall, P., Horowitz, J. L. and Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82**, 561–574.

Huang, J. Z. and Liu, L. (2006). Polynomial spline estimation and inference of proportional hazards regression models with flexible relative risk form. *Biometrics* **62**, 793–802.

Jensen, U. and Lütkebohmert, C. (2008). A Cox-type regression model with change-points in the covariates. *Lifetime Data Analysis* **14**, 267–285.

Kosorok, M. R. and Song, R. (2007). Inference under right censoring for transformation models with a change-point based on a covariate threshold. *The Annals of Statistics* **35**, 957–989.

Lee, S. M. (1999). On a class of m out of n bootstrap confidence intervals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61**, 901–911.

Liang, K. Y., Self, S. G. and Liu, X. H. (1990). The Cox proportional hazards model with change point: An epidemiologic application. *Biometrics* **46**, 783–93.

Luo, X. (1996). The asymptotic distribution of MLE of treatment lag threshold. *Journal of Statistical Planning and Inference* **53**, 33–61.

Luo, X. and Boyett, J. M. (1997). Estimations of a threshold parameter in Cox regression. *Communications in Statistics - Theory and Methods* **26**, 2329–2346.

Marquis, K., Debigaré, R., Lacasse, Y., LeBlanc, P., Jobin, J., Carrier, G. et al.(2002). Midthigh muscle cross-sectional area is a better predictor of mortality than body mass index in patients with chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine* **166**, 809–813.

Politis, D. and Romano, J. (1999). *Subsampling.* Springer, New York.

Pons, O. (2002). Estimation in a Cox regression model with a change-point at an unknown time. *Statistics: A Journal of Theoretical and Applied Statistics* **36**, 101–124.

Pons, O. (2003). Estimation in a Cox regression model with a change-point according to a threshold in a covariate. *The Annals of Statistics* **31**, 442–463.

Sekhon, J. S. and Mebane, W. R. (1998). Genetic optimization using derivatives. *Political Analysis* **7**, 187–210.

Shao, J. (1994). Bootstrap sample size in nonregular cases. *Proceedings of the American Mathematical Society* **122**, 1251–1262.

Tapp, R., Zimmet, P., Harper, C., de Courten, M., McCarty, D., Balkau, B. et al. (2006). Diagnostic thresholds for diabetes: The association of retinopathy and albuminuria with glycaemia. *Diabetes Research and Clinical Practice* **73**, 315–321.

van der Vaart, A. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: With*

Applications to Statistics. Springer-Verlag New York, New York.

Wang, W. (2004). Proportional hazards regression models with unknown link function and time-dependent covariates. *Statistica Sinica* **14**, 885–906.

Xu, G., Sen, B. and Ying, Z. (2014). Bootstrapping a change-point Cox model for survival data. *Electronic Journal of Statistics* **8**, 1345–1379.

Zhao, J., Zhu, Y., Lin, J., Matsuguchi, T., Blackburn, E., Zhang, Y. et al. (2014). Short leukocyte telomere length predicts risk of diabetes in American indians: The strong heart family study. *Diabetes* **63**, 354–62.

Yu Deng

Genentech, Inc., South San Francisco, CA 94080, USA.

E-mail: deng.yu@gene.com

Jianwen Cai

Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, USA.

E-mail: cai@bios.unc.edu

Donglin Zeng

Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, USA.

E-mail: dzeng@email.unc.edu

(Received April 2020; accepted September 2020)