Published in final edited form as:
Ann Stat. 2009 September 1; 36(3): 444–462. doi:10.1111/j.1467-9469.2008.00634.x.

EMPIRICAL LIKELIHOOD INFERENCE FOR THE COX MODEL WITH TIME-DEPENDENT COEFFICIENTS VIA LOCAL PARTIAL LIKELIHOOD

Yanqing Sun[*],
Department of Mathematics and Statistics, University of North Carolina at Charlotte, 9201 University City Boulevard, Charlotte, NC 28223, E-mail: yasun@uncc.edu

Rajeshwari Sundaram[†], and
Division of Epidemiology, Statistics and Prevention Research, National Institute of Child Health and Human Development, 6100 Executive Boulevard, Rockville, MD 20852, E-mail: sundaramr2@mail.nih.gov

Yichuan Zhao[‡]
Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, E-mail: yzhao@mathstat.gsu.edu

Abstract

The Cox model with time-dependent coefficients has been studied by a number of authors recently. In this paper, we develop empirical likelihood (EL) pointwise confidence regions for the time-dependent regression coefficients via local partial likelihood smoothing. The EL simultaneous confidence bands for a linear combination of the coefficients are also derived based on the strong approximation methods. The empirical likelihood ratio is formulated through the local partial log-likelihood for the regression coefficient functions. Our numerical studies indicate that the EL pointwise/simultaneous confidence regions/bands have satisfactory finite sample performances. Compared with the confidence regions derived directly based on the asymptotic normal distribution of the local constant estimator, the EL confidence regions are overall tighter and can better capture the curvature of the underlying regression coefficient functions. Two data sets, the gastric cancer data and the Mayo Clinic primary biliary cirrhosis data, are analyzed using the proposed method.

Keywords and phrases

Empirical likelihood; empirical processes; Gaussian multiplier calibration; local partial likelihood; pointwise and simultaneous confidence regions/bands; proportional hazards model; strong approximations

1. Introduction

Let T^* be the survival time of interest and $X = (X_1, \ldots, X_p)^T$ the possible explanatory variables or covariates, where the superscript denotes the transpose of a vector or matrix. We consider the situation where the survival time may be subject to possible random right censorship. Let

[*]Partially supported by NSF grants DMS-0304922 and DMS-0604576 and NIH grant 5 RO1 AI054165-05
[†]Supported by intramural research award of NIH
[‡]Partially supported by NSA grant H98230-07-1-0002

$T = \min(T^*, C)$ and $\delta = I(T^* \le C)$, where C is the censoring random variable that is independent of T^* conditional on X. We observe independent identically distributed (i.i.d.) copies of (T, X, δ): $(T_1, X_1, \delta_1), \ldots, (T_n, X_n, \delta_n)$. The survival time T_i^* is observed if $\delta_i = 1$ and censored at C_i if $\delta_i = 0$. The $X_i = (X_{i1}, \ldots, X_{ip})^T$ is the covariate for the ith subject. Let $\lambda(t|x)$ be the conditional hazard function of T^* given $X = x$, defined as

$$\lambda(t|x) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P\left\{t \le T^* \le t + \Delta t \middle| T^* \ge t, X = x\right\}.$$

The Cox proportional hazards model postulates the following form (Cox, 1972 and 1975):

$$\lambda(t|x) = \lambda_0(t) \exp\left\{\beta_0^T x\right\}, \tag{1.1}$$

where β_0 is a p-dimensional parameter and $\lambda_0(t)$ is an unspecified function. The function $\lambda_0(t)$ corresponds to the conditional hazard function of T^* given $X = 0$, thus also called the baseline hazard function. Under the Cox model (1.1), the conditional hazard rates associated with any two values of covariate X are proportional. The β_0 yields the log relative hazard rate between two sets of covariates. The proportional hazards model (1.1) is the most popular model for the analysis of survival data and has been studied extensively by many authors. The extension of this model to time dependent covariates has been done by Andersen and Gill (1982) using the counting process and the martingale approach.

In practice, the covariate effects may vary over time and the proportional hazards assumption may become questionable with longer follow-up. Different hazard regression models have been proposed to accommodate such situations. These include the general stratified Cox model by Dabroska (1997), additive hazards models by Aalen (1980) and Huffer and McKeague (1991), accelerated failure time models by Buckley and James (1979), Koul, Susarla and Van Ryzin (1981) and transformation models by Cheng, Wei and Ying (1995) and Chen, Jin and Ying (2002), among others. Because of wide applications and well established interpretations of the Cox model in practice, it is of great interest to allow the vector of coefficients β_0 to change over time on the basis of the Cox model (1.1). The Cox model with time-dependent coefficients specifies that

$$\lambda(t|x) = \lambda_0(t) \exp\left\{(\beta_0(t))^T x\right\}, \tag{1.2}$$

where $\lambda_0(t)$ is an unspecified baseline function as before and $\beta_0(t)$ is a p-dimensional vector of unspecified coefficient functions of t. Unless $\beta_0(t)$ is constant, this model represents a non-proportional hazards model. The model (1.2) has been investigated by Murphy and Sen (1991), Martinussen, Scheike and Skovgaard (2002), Cai and Sun (2003) and Tian, Zucker, and Wei (2005). Alternatively, Pons (2000) studied the extension of the model by letting β_0 depend on some covariate.

Fan, Gijbels and King (1997) suggested to estimate the risk factor in the hazard regression using a local partial likelihood estimation technique. Cai and Sun (2003) used a similar approach to estimate the time-dependent coefficients $\beta_0(t)$ for the model (1.2). The asymptotic consistency and asymptotic normality for the estimators $\hat{\beta}(t)$ of $\beta_0(t)$ were obtained at each time t. Pointwise Wald-type confidence intervals for $\beta_0(t)$ can thus be constructed. More recently, Tian, et al. (2005) further studied the local constant partial likelihood estimators. For any given p-vector κ, they obtained point-wise Wald-type confidence intervals for $\kappa^T \beta_0(t)$ based on the asymptotic normal distribution of $\hat{\beta}(t)$. They also constructed simultaneous confidence bands over a properly chosen time interval using the strong approximation technique presented by Bickel and Rosenblatt (1973) and Yandell (1983) and a resampling technique of Lin, Fleming and Wei (1994). Both these pointwise and simultaneous confidence intervals constructed for $\kappa^T \beta_0(t)$ are centered on the point estimate $\kappa^T \hat{\beta}(t)$.

In this paper, we present alternative constructions of confidence intervals for $\beta_0(t)$ and $\kappa^T \beta_0(t)$ based on the empirical likelihood (EL) approach of Owen (1988, 1990). Owen (1988, 1990) introduced EL confidence regions for the mean of a random vector based on i.i.d. complete data. Owen (1991, 1992) and Kolaczyk (1994), among others, have extended the EL methodology to a broad range of regression problems involving linear models, generalized linear models and project pursuit models. The EL methods to nonparametric regressions with local smoothing have been explored by Chen and Qin (2000) and Zhang and Liu (2003). The use of EL methods in survival analysis traces back to Thomas and Grunkemeier (1975) who derived pointwise confidence intervals for survival function with right censored data; see also Li (1995) and Murphy (1995). This approach has been used in the constructions of simultaneous confidence bands for survival function, Q-Q plot and the ratio of survival functions by Hollander, McKeague and Yang (1997), Einmahl and McKeague (1999), Li and Van Keilegom (2002), McKeague and Zhao (2002). Qin and Lawless (1994) linked the EL method to general estimating equations. Hjort, McKeague and Van Keilegom (2007) further extended the EL method to allow for plug-in estimates of nuisance parameters and slower than $\sqrt{n}$−rates of convergence. The EL method has some unique features, such as range respecting, transformation-preserving, asymmetric confidence interval and Bartlett correctability. A discussion of the advantages of the EL method over classical methods (based on normal approximation and bootstrap) can be found in Hall and La Scala (1990) and Owen (2001).

Qin and Jing (2001) considered using the EL approach for the Cox model (1.1). However, their method does not really apply to the Cox model in general since it assumes a known baseline function. In this paper, we develop EL pointwise confidence regions for the time-dependent regression coefficients $\beta_0(t)$ based on the local partial likelihood of Cai and Sun (2003). We also construct EL simultaneous confidence bands for $\kappa^T \beta_0(t)$ based on the strong approximation methods. Our main interest includes the development of EL simultaneous confidence bands coupled with local smoothing for the Cox model with time-dependent coefficients. An early work for constructing simultaneous confidence bands that used the EL with smoothing was considered by Hall and Owen (1993) for kernel density estimation.

The rest of the article is organized as follows. In Section 2, we construct the EL pointwise confidence regions for $\beta_0(t)$ and simultaneous confidence bands for $\kappa^T \beta_0(t)$, for any given p-vector κ. The numerical studies of the proposed method are reported in Section 3, which include a simulation study and applications to analyze the gastric cancer data (Stablein et al., 1981) and the Mayo Clinic primary biliary cirrhosis data (Fleming and Harrington 1991, Appendix D). All the proofs are collected in the Appendix.

2. Empirical likelihood confidence regions/bands for time-dependent coefficients

2.1. Local partial maximum likelihood estimator

Let $N_i(t) = I(T_i \le t, \delta_i = 1)$ be the counting process of observed failures for the ith individual, and $Y_i(t) = I(T_i \ge t)$ the at risk indicator process. For a fixed time point t, consider the following local partial log-likelihood function of the p-vector β to estimate $\beta_0(t)$:

$$l(\beta,t)=\sum_{i=1}^{n}\int_{0}^{\tau}K_h(u-t)\left\{\beta^T X_i(u)-\log\left[\sum_{k=1}^{n}Y_k(u)\exp(\beta^T X_k(u))\right]\right\}dN_i(u). \tag{2.1}$$

where $K_h(\cdot) = K(\cdot/h)/h$, $h = h_n > 0$ is the bandwidth that controls the size of a local neighborhood, $K(\cdot)$ is a kernel function that weighs smoothly down the contribution of remote data points and τ is a prespecified constant such that $P(T_i > \tau) > 0$. The local constant partial maximum likelihood estimator $\hat{\beta}(t)$ is obtained by maximizing (2.1) with respect to β (Cai and Sun, 2003).

Let $S^{(j)}(\beta,u)=n^{-1}\sum_{i=1}^{n}Y_i(u)\exp(\beta^T X_i(u))X_i(u)^{\otimes j}$ and $s^{(j)}(\beta, u) = ES^{(j)}(\beta, u)$, $j = 0, 1, 2$ where $X_i(u)^{\otimes 0} = 1$, $X_i(u)^{\otimes 1} = X_i(u)$ and $X_i(u)^{\otimes 2}=X_i(u)X_i^T(u)$. The score function is given by

$$U(\beta,t)=\sum_{i=1}^{n}\int_{0}^{\tau}K_h(u-t)\left[X_i(u)-\frac{S^{(1)}(\beta,u)}{S^{(0)}(\beta,u)}\right]dN_i(u). \tag{2.2}$$

Cai and Sun (2003) showed that $\widehat{\beta}(t)\xrightarrow{P}\beta_0(t)$ and that

$$\sqrt{nh}\left[\widehat{\beta}(t)-\beta_0(t)-\frac{1}{2}h^2(\text{bias}(t))\right]\quad\xrightarrow{D}\quad N(0,\nu_0\sum\nolimits^{-1}(t)), \tag{2.3}$$

for any fixed interior point t of $[0, \tau]$, where $\nu_0 = \int K^2(u)du$, $bias(t)$ is a function of t depending on the model specifications and

$$\sum(t)=\lambda_0(t)[s^{(2)}(\beta_0(t),t)-s^{(1)}(\beta_0(t),t)(s^{(1)}(\beta_0(t),t)^T/s^{(0)}(\beta_0(t),t)]. \tag{2.4}$$

The asymptotic covariance $\sum(t)$ can be consistently estimated by $\hat{\sum}(t) = \int K_h(u-t)V(\hat{\beta}(t), u)$ $dN(u)$, where $N(t)=n^{-1}\sum_{i=1}^{n}I(T_i\le t,\delta_i=1)$ and $V(\beta, u) = S^{(2)}(\beta, u)/S^{(0)}(\beta, u) - [S^{(1)}(\beta, u)]^{\otimes 2}/[S^{(0)}(\beta, u)]^2$.

Let $nh^5 \to 0$, which results in an under-smoothed estimator $\hat{\beta}(t)$. An asymptotic $100(1-\alpha)\%$ pointwise confidence region for $\beta_0(t)$, $0 < t < \tau$, based on the asymptotic normality (2.3), is given by

$$\mathscr{R}_1=\left\{\beta:\text{nh}\left[\widehat{\beta}(t)-\beta\right]^T\nu_0^{-1}\widehat{\sum}(t)\left[\widehat{\beta}(t)-\beta\right]\le\chi_p^2(\alpha)\right\},$$

where $\chi_p^2(\alpha)$ is the $1-\alpha$ quantile of the chi-square distribution with p degrees of freedom.

In the next subsection, we derive EL pointwise confidence regions for the time-dependent coefficients $\beta_0(t)$ based on the score function (2.2). We also construct EL simultaneous confidence bands for each component of $\beta_0(t)$ based on the strong approximation methods. The EL simultaneous confidence bands for $\kappa^T\beta_0(t)$ can then be obtained through an linear transformation. The following conditions are assumed in order to establish the asymptotic properties.

Let $\mathcal{B}$ be a compact set of the space $\mathbb{R}^p$ that includes a neighborhood of $\beta_0(t)$ for $t \in [0, \tau]$. We assume the following regularity conditions.

a. $X(t)$ is a bounded and predictable process on $[0, \tau]$.

b. $s^{(j)}(\beta, t)$, for $j = 0, 1, 2$, and their partial derivatives with respect to β are continuous in $(\beta, t) \in \mathcal{B} \times [0, \tau]$ with $s^{(0)}(\beta, t) > 0$ for $(\beta, t) \in \mathcal{B} \times [0, \tau]$. $\lambda_0(t)$, $\beta_0(t)$, $s^{(j)}(\beta_0(t), t)$ and $s^{(j)}(\beta, t)$, for $j = 0, 1$, are twice continuously differentiable in $t \in [0, \tau]$.

c. $\|S^{(j)}(\beta, t) - s^{(j)}(\beta, t)\| = O_p(n^{-1/2})$ uniformly in $(\beta, t) \in \mathcal{B} \times [0, \tau]$.

d. The matrix $\sum(t)$ is positive definite for all $t \in [0, \tau]$.

e. The kernel function $K(\cdot)$ is a symmetric density with the bounded support on $[-1, 1]$ and is continuously differentiable. The bandwidth $h = n^{-\upsilon}$, where $1/5 < \upsilon < 1/2$.

2.2. The EL pointwise confidence regions

The score function (2.2) can be written as

$$U(\beta, t)\sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)\left[X_i(u) - \frac{S^{(1)}(\beta, u)}{S^{(0)}(\beta, u)}\right](dN_i(u) - Y_i(u)\exp(\beta^T X_i(u))\, d\Lambda_0(u)). \tag{2.5}$$

Each term in the summation has mean close to zero at $\beta = \beta_0(t)$ by the martingale property and the fact that the difference $\beta_0(t) - \beta_0(u)$ is small for $|t - u| \le h$.

Let $\widehat{\Lambda}_0(\beta(\cdot), t) = \int_0^t (S^{(0)}(\beta(u), u))^{-1} dN(u)$. Replacing the cumulative baseline function $\Lambda_0(t) = \int_0^t \lambda_0(u)\,\mathrm{du}$ by its nonparametric estimator $\hat{\Lambda}_0(t) = \hat{\Lambda}_0(\hat{\beta}(\cdot), t)$, we define, for $1 \le i \le n$,

$$U_i(\beta, t) = \int_0^{\tau} K_h(u-t)\left[X_i(u) - \frac{S^{(1)}(\beta, u)}{S^{(0)}(\beta, u)}\right](dN_i(u) - Y_i(u)\exp(\beta^T X_i(u))\, d\widehat{\Lambda}_0(u)).$$

Then $U(\beta, t) = \sum_{i=1}^{n} U_i(\beta, t)$.

The empirical likelihood method was originally proposed by Owen (1988, 1990) for constructing confidence regions for the mean of a random vector based on i.i.d. complete data. Qin and Lawless (1994) linked the EL method to general estimating equations. Hjort, McKeague and Van Keilegom (2007) further extended the EL method to allow for plug-in estimates of nuisance parameters and slower than $\sqrt{n}$−rates of convergence under a given set of conditions. The basic idea of empirical likelihood is to regard the observations, e.g., $\{(T_i, X_i, \delta_i), i = 1, \dots, n\}$, as if they are i.i.d. from a fixed and unknown distribution P, and to model P by a multinomial distribution concentrated on the observations, with p_i as the probability mass at the ith observation. The empirical likelihood is then $L(P) = \prod_{i=1}^{n} p_i$ and the empirical likelihood ratio is of the form $L(P)/L(P_n)$, where P_n is the empirical distribution dividing the probability equally among n observations. The following empirical likelihood ratio function can be formulated to obtain confidence regions for $\beta_0(t)$ at each time t, by considering the constrain $E_P U_i(\beta, t) = 0$

$$R(\beta, t) = \sup\{\prod_{i=1}^{n} np_i : \sum_{i=1}^{n} p_i U_i(\beta, t) = 0,\ p_i \ge 0,\ \sum p_i = 1\}. \tag{2.6}$$

Our empirical likelihood ratio uses the plug-in estimate for the nonparametric base-line function $\Lambda_0(t)$ and needs to consider local smoothing which results in slower than $\sqrt{n}-\text{rate}$ of convergence the score $U(\beta, t)$. Unlike Hjort, et al. (2007), our $U_i(\beta, t)$, $i = 1, \ldots, n$, are not i.i.d. even for the true $\Lambda_0(t)$ since each term involves both $S^{(0)}(\beta, t)$ and $S^{(1)}(\beta, t)$. The asymptotic theorem for constructing the EL pointwise confidence regions for $\beta_0(t)$ can be proved similarly as in Hjort et al. (2007). This proof is based on a technical lemma given in the Appendix, Lemma A.1, which specify the terms similar to the conditions of Hjort et al. (2007). In fact, Lemma A.1 presents some results that are stronger than what are needed for the EL point-wise confidence regions. It is also the basis for the constructions of EL simultaneous confidence bands which present some new challenges and necessitate additional treatments, as derived in the next subsection.

For each β and t, the maximum of $\prod_{i=1}^{n} np_i$ in (2.6) under the constraints $\sum_{i=1}^{n} p_i\, U_i(\beta, t)=0,\ p_i \geq 0,$ and $\sum p_i = 1$ exists, provided that 0 is inside the convex hull of the points $U_1(\beta, t), \ldots, U_n(\beta, t)$, cf. Owen (1988,1990). It is clear that the local constant estimator $\hat{\beta}(t)$ is also the empirical likelihood ratio estimator since $R(\beta, t)$ attains the maximum at $\hat{\beta}(t)$ based on the fact that $\prod_{i=1}^{n} p_i$ maximized at $p_i = 1/n$. The maximum $R(\beta, t)$ may be obtained using the method of Lagrange multipliers. Similar to Qin and Lawless (1994), the maximum is attained when $p_i = n^{-1}\{1 + \lambda^T U_i(\beta, t)\}^{-1}$, where $\lambda = (\lambda_1, \ldots, \lambda_p)^T$ is a solution to

$$\sum_{i=1}^{n} \frac{U_i(\beta, t)}{1+\lambda^T U_i(\beta, t)}=0. \tag{2.7}$$

The empirical likelihood ratio statistic for testing $H_0 : \beta = \beta_0(t)$ at a fixed time $t \in (0, \tau)$ becomes

$$-2 \log R(\beta, t)=2\sum_{i=1}^{n} \log\left\{1+\lambda^T U_i(\beta, t)\right\}. \tag{2.8}$$

The next theorem derives its asymptotic distribution and leads to the construction of pointwise confidence regions for $\beta_0(t)$.

THEOREM 2.1—*Under Conditions (a)–(e),* $-2 \log R(\beta_0(t), t)$ *converges in distribution to a chi-square distribution with p degrees of freedom as* $n \to \infty$ *for each* $t \in (0, \tau)$.

For each $t \in (0, \tau)$, an asymptotic $100(1 - \alpha)\%$ confidence region for $\beta_0(t)$ is obtained by $\left\{\beta: -2 \log R(\beta, t) \leq \chi_p^2(\alpha)\right\}$.

As discussed by Owen (2001), the coverage error in empirical likelihood using χ^2 calibration is typically of order of $1/n$. A Bartlett correction reduces the coverage error to $O(1/n^2)$ and the bootstrap has been found to be effective in reducing the coverage error in some settings. Here for the same purpose we introduce a Gaussian multiplier calibration for the critical value $\chi_p^2(\alpha)$. The Gaussian multiplier method has been widely used and very effective in many semiparametric and nonparametric settings, cf. Lin et al. 1994. Let $\tilde{U}(\beta, t)=(h/n)^{1/2}\sum_{i=1}^{n} U_i(\beta, t)$. From (A.8) in the proof of Theorem 2.1 in the Appendix,

$$-2 \log R(\beta, t)=(\tilde{U}(\beta, t))^T(\Gamma(t))^{-1}\tilde{U}(\beta, t)+o_p(n^{-\epsilon}), \tag{2.9}$$

holds uniformly for β in a neighborhood of $\beta_0(t)$, $t \in [h, \tau - h]$ for some $\epsilon > 0$, where $\Gamma(t) = \nu_0 \Sigma(t)$ is the asymptotic variance of $\tilde{U}(\beta_0(t), t)$ as shown in Lemma A.1.

Let $\widehat{\Gamma}(t) = hn^{-1} \sum_{i=1}^{n} \widehat{U}_i(t) \widehat{U}_i^T(t)$ where $\hat{U}_i(t)$ is obtained by replacing β with $\hat{\beta}(t)$ in $U_i(\beta, t)$. Then $\hat{\Gamma}(t)$ is a consistent estimator for $\Gamma(t)$. Let $\xi_1, \ldots, \xi_n$ be i.i.d. standard normal random variables. Define $U^*(\widehat{\beta}(t), t) = (h/n)^{1/2} \sum_{i=1}^{n} \xi_i \widehat{U}_i(t)$ and

$$GR^*(t) = U^*(\widehat{\beta}(t), t))^T (\widehat{\Gamma}(t))^{-1} U^*(\widehat{\beta}(t), t). \tag{2.10}$$

The critical value $\chi_p^2(\alpha)$ in the EL confidence region construction can be calibrated using the $1 - \alpha$ empirical quantile of $GR_1^*, \ldots, GR_B^*$, which are B independent copies of GR^* obtained by repeatedly generating independent sets of $\{\xi_1, \ldots, \xi_n\}$ given the observed data sequence.

2.3. The EL Simultaneous confidence bands

Based on the pointwise confidence interval for a coefficient function, one can make inference for the covariate effect at a given time. However, to assess how the covariate effect changes over time, one would need to establish simultaneous confidence band for coefficient function over time. Tian, et al. (2005) has derived simultaneous confidence bands for the coefficient functions over time that are symmetric about the corresponding estimators. Their confidence bands for $\{\kappa^T \beta_0(t), t \in [t_1, t_2]\}$ are of the form $\{\kappa^T \hat{\beta}(t) \pm c_\alpha \hat{w}(t)^{-1}, t \in [t_1, t_2]\}$, where κ is a given p-vector. In this section, we develop EL simultaneous confidence bands for each component of the coefficients $\beta_0(t)$ and for $\kappa^T \beta_0(t)$ over time. Such confidence bands are easy to plot and interpolate. We use an approach that resembles the method of profile empirical likelihood ratio by plugging in the estimates of other components.

As we see from (2.9), the difficulty in the constructions of the simultaneous confidence bands lies in the fact that the process $(h/n)^{1/2} U(\beta, t)$ is not tight, thus it does not converge to a process in distribution; see Tian, et al. (2005). The strong approximation techniques presented by Bickel and Rosenblatt (1973) and Yandell (1983) are used to obtain the approximation to the distribution of the supremum of some appropriate process. Furthermore, since this type of approximation is known to be not very accurate (Hall, 1993), we propose a Gaussian multiplier simulation technique to calibrate the critical values.

Let $[t_1, t_2] \subset (0, \tau)$. Without loss of generality, we describe the method for obtaining pointwise and simultaneous confidence bands for $\beta_1(t)$, the first component of $\beta_0(t)$, over $t \in [t_1, t_2]$ in the following. Let $R(\beta_1, \hat{\beta}_{-1}(t), t)$ be the empirical likelihood ratio $R(\beta, t)$ evaluated at $\beta = (\beta_1, \hat{\beta}_{-1}(t))$, where β_1 is the first component of β and $\hat{\beta}_{-1}(t)$ is the vector consisting of the components of $\hat{\beta}(t)$ other than the estimate $\hat{\beta}_1(t)$. Let e_1 be the first column of a $p \times p$ identity matrix and

$$Q = \sup_{t_1 \le t \le t_2} \left(\frac{-2 \log R(\beta_1, \widehat{\beta}_{-1}(t), t)}{(e_1^T \widehat{\Gamma}(t) e_1)(e_1^T \widehat{\Gamma}^{-1}(t) e_1) \nu_0^{-1}} \right),$$

$$Q^* = \sup_{t_1 \le t \le t_2} \left(\frac{[(\widehat{\Gamma}^{-1}(t) e_1)^T U^*(\widehat{\beta}(t), t)]^2}{(e_1^T \widehat{\Gamma}^{-1}(t) e_1) \nu_0^{-1}} \right).$$

THEOREM 2.2—*Assume that Conditions (a)−(e) hold. Let* $nh^5 \to 0$. *Then*

$$-2\log R(\beta_1,\widehat{\beta}_{-1}(t),t)=[(e_1^T\Gamma(t)e_1)^{1/2}(\Gamma^{-1}(t)e_1)^T\tilde{U}(\beta,t)]^2+o_p(n^{-\epsilon}),$$

for some $\epsilon > 0$. *Further*,

i. *For each* $t \in [t_1, t_2]$,

$$\frac{-2\log R(\beta_1,\widehat{\beta}_{-1}(t),t)}{(e_0^T\widehat{\Gamma}(t)e_1(e_1^T\widehat{\Gamma}^{-1}(t)e_1)}\xrightarrow{\mathscr{D}}\chi_1^2;$$

ii. *For any* $M > 0$,

$$\sup_{|x|\le M}\left|P\left\{\frac{r_n}{2d_n}(Q-d_n^2)<x)\right\}-\exp(-2e^{-x})\right|=O((\log n)^{-1}), \tag{2.11}$$

and

$$\sup_{x\ge-\frac{1}{2}r_nd_n}\left|P\left\{\frac{r_n}{2d_n}(Q-d_n^2)<x)\right\}-P\left\{\frac{r_n}{2d_n}(Q^*-d_n^2)<x)\middle|\{T_i,X_i(\cdot),\delta_i\}\right\}\right|=O(n^{-\epsilon}), \tag{2.12}$$

where $r_n = [2\log((t_2 - t_1)/h)]^{1/2}$, $d_n=r_n+\log\{\int_{-1}^1(K\prime(s))^2\,ds/[4\pi^2\int_{-1}^1(K(s))^2\,ds]\}$ $(2r_n)^{-1}$ *and* $K'(s)$ *is the derivative of* $K(s)$.

From Theorem 2.2, an asymptotic 100(1−α)% EL pointwise confidence band for $\beta_1(t)$, $t_1 \le t \le t_2$ is obtained by

$$\{\beta_1:-2\log R(\beta_1,\widehat{\beta}_{-1}(t),t)\le\chi_1^2(\alpha)(e_1^T\widehat{\Gamma}(t)e_1)(e_1^T\widehat{\Gamma}^{-1}(t)e_1)\},$$

where $\chi_p^2(\alpha)$ is the $1-\alpha$ quantile of the distribution of χ_1^2 with 1 degree of freedom. An asymptotic $100(1-\alpha)$% EL simultaneous confidence band for $\beta_1(t)$, $t_1 \le \cdot t \le t_2$ can be constructed by

$$\{(t,\beta_1):-2\log R(\beta_1,\widehat{\beta}_{-1}(t),t)\le c_\alpha(e_1^T\widehat{\Gamma}(t)e_1)(e_1^T\widehat{\Gamma}^{-1}(t)e_1)v_0^{-1}\},$$

where c_α is the 1−α quantile of the asymptotic distribution of Q. By (2.11), c_α can be estimated by $d_n^2-(2d_n/r_n)\log[-.5\log(1-\alpha)]$. As discussed earlier and according to Theorem 2.2, a more accurate approximation for c_α can be obtained by the quantile of Q^* given the observed data sequence. Specifically, c_α can be estimated by the $1-\alpha$ empirical quantile of $Q_1^*,\ldots,Q_B^*$, which are independent copies of Q^* obtained by repeatedly generating independent sets of $\{\xi_1, \ldots, \xi_n\}$. Similarly, the critical value $\chi_1^2(\alpha)$ in the EL pointwise confidence band for $\beta_1(t)$, $t_1 \le t \le t_2$, can be calibrated using the $1-\alpha$ empirical quantile of Q^* after dropping the supremum over t.

To construct EL pointwise and simultaneous confidence bands for $\kappa^T\beta_0(t)$ where κ is a $p \times 1$-dimensional constant vector, we let ψ be a $p \times (p-1)$ matrix chosen such that the matrix $A =$

(κ, ψ) is nonsingular. Put $\gamma_0(t) = A^T \beta_0(t)$. Then $\beta_0(t) = (A^T)^{-1}\gamma_0(t)$ and $(\beta_0(t))^T x = (\gamma_0(t))^T A^{-1}x$. Hence, the EL confidence bands for $\kappa^T \beta_0(t)$ can be carried out by constructing the confidence bands for the first component of $\gamma_0(t)$ in the covariate transformed Cox model: $\lambda(t|z) = \lambda_0(t) \exp((\gamma_0(t))^T z)$, where $z = A^{-1}x$. In fact, the vector κ can be allowed to depend on t since the method developed here is applicable to time-dependent covariates. Based on the simultaneous confidence bands for $\kappa^T \beta_0(t)$, one can test the hypothesis $\kappa^T \beta_0(t) = 0$ by examining whether the confidence band encloses the x-axis, or identify the covariates whose effects are time-dependent by checking whether the corresponding confidence band contains a horizonal line.

3. Numerical studies

In this section we conduct a simulation study to check the finite sample performance of the proposed EL pointwise and simultaneous confidence bands. We compare the performance of the proposed EL pointwise confidence regions with the normal approximation based confidence regions in terms of coverage probability. The EL simultaneous confidence bands are also assessed. Moreover, The proposed method is applied to analyze the gastric cancer data (Stablein et al., 1981) and the Mayo Clinic primary biliary cirrhosis data (Fleming and Harrington 1991, Appendix D).

3.1. Simulation study

We consider two models for the simulation experiments. First, we simulate data from the following model:

$$\lambda(t|X)=\lambda_0(t)\exp(\beta_0(t)X) \tag{3.1}$$

where $\lambda_0(t) = t^{-1/2}/2$, $\beta_0(t) = t^{1/2}$ and the covariate X is a 0/1 treatment indicator, with each value having equal probability of occurring. The censoring variable C is generated from a uniform distribution on an interval $[0, \alpha]$, where α is chosen such that it results in either 10% or 30% censoring. Note that the time varying effect $\beta_0(t)$ in this case denotes the change in the relative risk between the treatment group and the control group on a logarithmic scale. Our simulation compares the EL confidence intervals for $\beta_0(t)$, $t \in [.25, 1.25]$ with that based on Cai and Sun (2003). Table 1 presents the coverage probabilities of the asymptotic 95% pointwise confidence intervals for $\beta_0(t)$, $t \in [.25, 1.25]$ for various sample sizes $n = 250$, 500 and $n = 750$, bandwidths $h = .2$ and $h = .3$, and for censoring percentages 10% and 30%, where ELSQ corresponds to the EL procedure using the chi-square critical values, ELGM is the EL procedure using Gaussian multiplier method for estimating the critical values and CS is the method based on Cai and Sun (2003). The coverage probabilities of the EL simultaneous confidence bands for $\beta_0(t)$, $t \in [.25, 1.25]$ are presented in Table 2. The Epanechinikov kernel function $K(x) = .75(1 - x^2)I(|x| \le 1)$ is used throughout. All the coverage probabilities are based on 1000 replications. The proposed EL pointwise confidence intervals performs very competitively with that of Cai and Sun (2003) approach in terms of coverage probability. The EL approach remains stable across various censoring percentages. The coverage probabilities of the EL simultaneous confidence band for $\beta_0(t)$, $t \in [.25, 1.25]$ reported in Table 2 shows that as sample size increases from $n = 250$ to $n = 750$, the performance of ELGM improves reaching very close to the nominal level and remains stable across varying censoring percentages, while ELEX (with c_α calculated directly from the extreme value distribution) performs very poorly.

Next, we investigate the performance of the proposed confidence regions based on the following model with one time-varying effect and one fixed effect:

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\lambda(t|X_1, X_2) = \exp(\beta_1(t)X_1 + \beta_2 X_2) \tag{3.2}$$

where $\beta_1(t) = t$ and $\beta_2 = 1/2$. The covariates X_1 is generated from a uniform distribution on the interval [0, 1] and X_2 is generated from a standard normal distribution. The baseline $\lambda_0(t)$ is taken to be 1: In Table 3, we present the coverage probabilities of the EL pointwise confidence interval for $\beta_1(t)$ for $t \in [.25, 1.25]$ for various sample sizes $n = 250$, 500 and 750, bandwidths $h = .5$ and $h = .6$, and censoring percentages 10% and 30%: The coverage probabilities of the pointwise confidence intervals based on Cai and Sun (2003) are not reported here. The findings on the comparison are similar to those from Table 1. The coverage probabilities of the EL simultaneous confidence band (ELGM) for $\beta_1(t)$ for $t \in [.25, 1.25]$ are given in Table 4, while those of ELEX are not reported simply because the method does not seem to work for practical sample sizes due to its slow convergence. Again, the numbers reported in the tables are based on 1000 replications.

Note that the performance of the EL pointwise confidence band is very good even for the sample size $n = 250$ and remains stable across varying censoring percentages. Table 4 also indicates good finite sample behavior of the EL simultaneous confidence bands.

Overall, based on the extensive simulation studies that we have conducted (including the ones not reported here), the empirical likelihood based approach seems to have very good finite sample properties. In fact, all three approaches (ELSQ, ELGM and CS) for the pointwise confidence intervals seem to yield similar coverage probabilities. The advantage of the EL confidence bands is that the produced confidence bands need not be symmetric about the estimates, and are narrower particularly in the situation of higher dimensional covariates, as we shall see from the two application studies next. This property translates to greater power in the hypothesis testing with the EL confidence bands. These findings are in line with the simulation outcomes of Qin and Lawless (1994). In addition, our data analysis experience indicates that the EL procedure using Gaussian multiplier method (ELGM) for estimating the critical values is more stable than the EL procedures using the chi-square critical values (ELSQ), probably due to the fact that the ELGM is based on a better approximation to the finite sample distribution especially for small sample sizes. Further, the EL simultaneous confidence bands using the critical values calculated directly from the extreme value distribution (ELEX) do not perform well for practical sample sizes due to its slow convergence.

To compute EL pointwise/simultaneous confidence bands, one can use a combination of grid search and Newton Raphson type of root solving algorithms with a modification for checking constraints; see the discussions in Hall and La Scala (1990). In practice, the appropriate bandwidth selection can be based on a cross-validation method. This approach is widely used in nonparametric function estimation literature and has been investigated by Tian et al. (2005) in this scenario. They used a κ-fold cross-validation approach for the bandwidth selection with the optimum choice based on minimizing "prediction error". This procedure appears to be robust.

3.2. Gastric cancer study

In this section, we illustrate our method by first analyzing gastric cancer patients study (Stablein et al., 1981). The data consists of survival times of ninety patients divided between two treatments, one group on chemotherapy and the other group on combined treatment of both chemotherapy and radiation. We analyze this data using the Cox model with time-varying treatment effect $\beta_0(t)$. In Figure 1, we present a comparison of the proposed EL confidence interval/band with the confidence interval based on the Cai and Sun approach, using the

NIH-PA Author Manuscript
NIH-PA Author Manuscript

bandwidth $h = 600$ days. The critical values of the EL confidence intervals/bands are estimated using the Guassian multiplier approach.

As can be seen from Figure 1, the treatment effect is linear in agreement with the findings of Carter et al. (1983). However, the effect changes from positive to negative, i.e., $\hat{\beta}(t) > 0$ roughly up to 15 months and then $\hat{\beta}(t)$ becomes negative. The empirical likelihood based confidence interval is narrower than that based on the Cai and Sun approach in the right tail and is asymmetric.

3.3. Mayo Clinic primary biliary cirrhosis study

We further illustrate our proposed EL interval estimates by analyzing the well-known Mayo Clinic primary biliary cirrhosis data (Fleming and Harrington 1991, Appendix D). The data consists of survival times of 418 patients and various potential prognostic factors with two subjects missing covariate values. As in previous literature, we focus our attention on the 416 patients with complete covariate information and include the following five covariates: age, log(albumin), log(bilirubin), log(prothrombin time) and edema. These covariates have been selected as important predictors for Cox's regression model with time-invariant regression parameters by Fleimng and Harrington (1991, page 195) and by Tian et al. (2005) for time-varying parameters. Tian et al. (2005) showed that the optimal bandwidth using on the prediction error based on the minus logarithm of partial likelihood function is 690 days or 1.89 years. We shall use this bandwidth for all the interval estimates calculations.

We compare the 95% EL pointwise confidence intervals for the time-dependent effects of the five covariates with those based on the Cai and Sun approach. As seen in Figure 2, the EL pointwise confidence intervals (dashed lines) are overall tighter than the Cai and Sun based approach (grey lines). We also report the 95% EL confidence bands for the time-dependent effects of the five covariates (dotted lines) in the time interval [1:78; 8:22] years. Again, the critical values of the EL confidence intervals/bands are estimated using the Guassian multiplier approach. Our findings are in agreement with the analysis of Tian et al. (2005) that the effect of log(prothrombin time) is significant initially with the effect decreasing as time progresses. Also with the introduction of time-dependent effects, edema and log(prothrombin time) can be introduced into the model without violation of proportionality assumption (Fleming and Harrington, 1991). Overall, the EL pointwise confidence intervals are narrower than the Cai and Sun approach based pointwise confidence intervals and can better capture the curvature of the underlying treatment effects.

Appendix

Here we give proofs for Theorem 2.1 and Theorem 2.2. The following lemma provides the critical results needed for the proofs. Let $\|A\|$ be the Euclidean modulus of a matrix A. According to the Appendix A and B of Tian et al. (2005), there exists a $\alpha > 0$ such that $\sup_{h\le t\le \tau-h} \|\hat{\beta}(t) - \beta_0(t)\| = (nh)^{-1/2} O_p(n^{-\alpha})$. Let $\delta_n = (nh)^{-1/2} n^{-\delta}$ with $0 < \delta < \alpha$, and $\mathcal{B}_t$ be a neighborhood of $\beta_0(t)$ such that $\|\beta - \beta_0(t)\| \le \delta_n$, for $0 \le t \le \tau$.

LEMMA A.1

Under Conditions (a)–(e), *for* $0 < t < \tau$,

i. $\sqrt{nh}[n^{-1} \sum_{i=1}^n U_i(\beta_0(t), t)] \xrightarrow{\mathscr{D}} N(0, \Gamma(t))$;

ii. $n^{-1} \| \sum_{i=1}^n U_i(\beta, t) \| = O_p((nh)^{-1/2} + h^2 + \delta_n)$, *uniformly in* $\beta \in \mathcal{B}_t$, $t \in (h, \tau - h)$;

iii. $\tilde{\Gamma}(\beta,t)=hn^{-1}\sum_{i=1}^{n}U_i(\beta,t)U_i^T(\beta,t)\xrightarrow{P}\Gamma(t)$, *uniformly in* $\beta\in\mathcal{B}_t$, $t\in(h,\tau-h)$, *where* $\Gamma(t)=\nu_0\Sigma(t)$.

Proof

Let $M_i(t)=N_i(t)-\int_0^t Y_i(u)\exp((\beta_0(u))^T X_i(u))\,d\Lambda_0(u)$ be the martingale associated with the ith individual. Then for each fixed t,

$$\begin{aligned}V_n(t) &= \sqrt{nh}[n^{-1}\sum_{i=1}^{n}U_i(\beta_0(t),t)]\\ &= \sqrt{nh}n^{-1}\int_0^\tau K_h(u-t)\sum_{i=1}^{n}\left[X_i(u)-\frac{S^{(1)}(\beta_0(t),u)}{S^{(0)}(\beta_0(t),u)}\right]dM_i(u)+O_p((nh^5)^{1/2}),\end{aligned}$$

where the first term is a locally square integrable martingale in τ with the predictable variation process

$$\begin{aligned}<V_n,V_n>(t) &=n^{-1}h\sum_{i=1}^{n}\int_0^\tau K_h^2(u-t)\left(X_i(u)-\frac{S^{(1)}(\beta_0(t),u}{S^{(0)}(\beta_0(t),u)}\right)^{\otimes 2}Y_i(u)\lambda(u|X_i(u))\,du\\ &\xrightarrow{P}\Gamma(t)\end{aligned}$$

by Lemma 1 of Cai and Sun (2003). Let $H_{n,i,l}(u)$ be the lth element of $X_i(u)-S^{(1)}(u,\beta_0(t))/S^{(0)}(u,\beta_0(t))$. The Lindeberg condition is satisfied since

$$\sum_{i=1}^{n}\int_0^\tau n^{-1}hK_h^2(u-t)H_{n,i,l}^2(u)I\{\sqrt{h/n}K_h(u-t)H_{n,i,l}(u)|>\epsilon\}Y_i(u)\lambda(uX_i(u))du\xrightarrow{P}0$$

for all $\epsilon>0$. Thus $V_n(t)\xrightarrow{\mathcal{D}}N(0,\Gamma(t))$ by applying the martingale central limit theorem [cf. Theorem 5.3.5 of Fleming and Harrington (1991)].

The proof of (ii) follows from Appendix A of Tian et al. (2005) by noting that the rate $o_p(1)$ in the last line of its page 180 can be substituted by $O_p((nh)^{-1/2}+h^2)$ and that the derivatives of $s^{(j)}(\beta,t)$, $j=0,1$, are uniformly continuous in the neighborhood $\mathcal{B}_t\times[0,\tau]$.

To prove the assertion (iii), we let

$$\begin{aligned}W_i(\beta,t)&=\int_0^\tau K_h(u-t)\left(X_i(u)-\frac{S^{(1)}(\beta,u)}{S^{(0)}(\beta,u)}\right)dM_i(u)\\ w_i(\beta,t)&=\int_0^\tau K_h(u-t)\left(X_i(u)-\frac{s^{(1)}(\beta,u)}{s^{(0)}(\beta,u)}\right)dM_i(u).\end{aligned}$$

Then,

$$
\begin{aligned}
U_i(\beta,t) &= W_i(\beta,t)+\int_0^\tau K_h(u-t)\left(X_i(u)-\frac{S^{(1)}(\beta,u)}{S^{(0)}(\beta,u)}\right)Y_i(u)\\
&\quad\times\left(\exp\{(\beta_0(u))^T X_i(u)\}\,d\Lambda_0(u)-\exp\{\beta^T X_i(u)\}\,d\hat{\Lambda}_0(u)\right)\\
&= W_i(\beta,t)+A_{1i}(t)-A_{2i}(t)-A_{3i}(t),
\end{aligned}
$$

where

$$
\begin{aligned}
A_{1i}(t) &= \int_0^\tau K_h(u-t)\left(X_i(u)-\frac{S^{(1)}(\beta,u)}{S^{(0)}(\beta,u)}\right)Y_i(u)\\
&\quad\times(\exp\{(\beta_0(u))^T X_i(u)\}-\exp\{\beta^T X_i(u)\})\,d\Lambda_0(u)\\
A_{2i}(t) &= \int_0^\tau K_h(u-t)\left(X_i(u)-\frac{S^{(1)}(\beta,u)}{S^{(0)}(\beta,u)}\right)Y_i(u)\exp\{\beta^T X_i(u)\}\\
&\quad\times d(\widehat{\Lambda}_0(\widehat{\beta}(\cdot),u)-\widehat{\Lambda}_0(\beta_0(\cdot),u))\\
A_{3i}(t) &= \int_0^\tau K_h(u-t)\left(X_i(u)-\frac{S^{(1)}(\beta,u)}{S^{(0)}(\beta,u)}\right)Y_i(u)\exp\{\beta^T X_i(u)\}\\
&\quad\times d(\widehat{\Lambda}_0(\beta_0(\cdot),u)-\Lambda_0(u)).
\end{aligned}
$$

Since the covariate processes are bounded, $\beta_0(u)$ is continuously differentiable and $S^{(1)}(\beta, u)/S^{(0)}(\beta, u) = O_p(1)$ holds uniformly in $\beta \in \mathscr{B}_u$, $u \in [0, \tau]$, we have

$$
\begin{aligned}
\| A_{1i}(t) \| &= O_p(1)\int_0^\tau K_h(u-t)|(\beta_0(u)-\beta)^T X_i(u)|\,d\Lambda_0(u)\\
&= O_p(\delta_n+h),\quad \text{uniformly in } \beta\in\mathscr{B}_t,\, t\in(h,\tau-h),\, i\in\{1,\ldots,n\}.
\end{aligned}
$$

Since $\sup_{\beta\in\mathscr{B}_t, t\in[0,\tau]} |S^{(0)}(\beta, t) - s^{(0)}(\beta, t)| = O_p(n^{-1/2})$ and by the uniform consistency of $\hat{\beta}(\cdot)$ given in Appendix A (Tian, et al., 2005),

$$
\begin{aligned}
\| A_{2i}(t) \| &= O_p(1)\sup_{t-h\le u\le t+h}|(S^{(0)}(\widehat{\beta}(u),u))^{-1}-S^{(0)}(\beta_0(u),u))^{-1}|\\
&\quad\times\int_0^\tau K_h(u-t)\,dN(u)\\
&= O_p(\delta),\quad \text{uniformly in } \beta\in\mathscr{B}_t,\, t\in(h,\tau-h),\, i\in\{1,\ldots,n\}.
\end{aligned}
$$

Let

$$
B_i(t)=\sum_{j=1}^n\int_0^t\left(X_i(u)-\frac{S^{(1)}(\beta,u)}{S^{(0)}(\beta,u)}\right)Y_i(u)\exp\{\beta^T X_i(u)\}\frac{dM_j(u)}{nS^{(0)}(\beta_0(u),u)}.
$$

Then

$$
\begin{aligned}
\| A_{3i}(t) \| &= \left\| \int_0^\tau K_h(u-t)\,dB_i(u) \right\| + o_p(h)\\
&= \left\| \int_{t-h}^{t+h} n^{1/2}B_i(u)n^{-1/2}h^{-2}K'((u-t/h)\,du \right\| + o_p(h)\\
&\le \left\| \sup_{0\le u\le\tau}|n^{1/2}B_i(u)|O((nh^2)^{-1/2}) \right\| + o_p(h)
\end{aligned}
$$

uniformly in $\beta \in \mathcal{B}_t$, $t \in (h, \tau - h)$ and $i \in \{1, \ldots, n\}$.

Since $\|W_i(\beta, t)\| = O_p(h^{-1})$, $\|w_i(\beta, t)\| = O_p(h^{-1})$, $A_{1i}(t)$ and $A_{2i}(t)$ converge to zero in probability uniformly in $\beta \in \mathcal{B}_t$, $t \in (h, \tau - h)$ and $i \in \{1, \ldots, n\}$, it follows that

$$hn^{-1}\sum_{i=1}^{n}U_i(\beta,t)(U_i(\beta,t))^T = hn^{-1}\sum_{i=1}^{n}W_i(\beta,t)(W_i(\beta,t))^T \\ -hn^{-1}\sum_{i=1}^{n}W_i(\beta,t)(A_{3i}(t))^T - hn^{-1}\sum_{i=1}^{n}A_{3i}(t)(W_i(\beta,t))^T + O_p(\delta_n+h).$$

Further,

$$\begin{aligned} &E\sup_{\beta\in\mathcal{B}_t,0\le t\le\tau}\| hn^{-1}\sum_{i=1}^{n}W_i(\beta,t)(A_{3i}(t))^T + hn^{-1}\sum_{i=1}^{n}A_{3i}(t)W_i(\beta,t)^T \| \\ &\le 2hn^{-1}\sum_{i=1}^{n}E\sup_{\beta\in\mathcal{B}_t,0\le t\le\tau}\| A_{3i}(t)W_i(\beta,t)^T \| \\ &\le 2hn^{-1}h^{-1}\sum_{i=1}^{n}E\sup_{\beta\in\mathcal{B}_t,0\le t\le\tau}\| A_{3i}(t) \| \\ &\le 2n^{-1}O((nh^2)^{-1/2})\sum_{i=1}^{n}E\sup_{\beta\in\mathcal{B}_t,0\le u\le\tau}|n^{1/2}B_i(u)| + o(h) \\ &= O((nh^2)^{-1/2})E\sup_{\beta\in\mathcal{B}_t,0\le u\le\tau}|n^{1/2}\beta_i(u)| + o(h), \end{aligned}$$

since $B_i(\cdot)$'s have the same distribution for $i = 1, \ldots, n$. By Lemma 2 of Gilbert, McKeague and Sun (2007, manuscript) and the Donsker theorem (cf. van der Vaart, 1998), one can show that $n^{1/2}B_i(u)$ converges weakly in $(\beta, u) \in \mathcal{B} \times [0, \tau]$. This is followed by $E\sup_{\beta\in\mathcal{B}_t,0\le u\le\tau}|n^{1/2}B_1(u)| = O(1)$. Thus, we have

$$hn^{-1}\sum_{i=1}^{n}U_i(\beta,t)U_i^T(\beta,t) = hn^{-1}\sum_{i=1}^{n}W_i(\beta,t)W_i^T(\beta,t) + O((nh^2)^{-1/2}) + O_p(\delta_n+h). \tag{A.1}$$

Note that

$$\begin{aligned} \| W_i(\beta,t) - w_i(\beta,t) \| &= \| \int_0^{\tau}K_h(u-t)\left(\frac{S^{(1)}(\beta,u)}{S^{(0)}(\beta,u)} - \frac{s^{(1)}(\beta,u)}{s^{(0)}(\beta,u)}\right)dM_i(u) \| \\ &= O_p(n^{-1/2}h^{-1}), \end{aligned}$$

uniformly in $\beta \in \mathcal{B}_t$, $t \in (0, \tau)$ and $i \in 1, \ldots, n\}$. We have

$$\begin{aligned} & hn^{-1} \parallel \sum_{i=1}^{n} W_i(\beta,t)W_i^T(\beta,t) - \sum_{i=1}^{n} w_i(\beta,t)w_i^T(\beta,t) \parallel \\ & = hn^{-1} \parallel \sum_{i=1}^{n} (W_i(\beta,t) - w_i(\beta,t))(W_i(\beta,t) - w_i(\beta,t))^T \\ & \quad + (W_i(\beta,t) - w_i(\beta,t))(w_i(\beta,t))^T + w_i(\beta,t)(W_i(\beta,t) - w_i(\beta,t))^T \parallel \\ & \leq hn^{-1} \sum_{i=1}^{n} \parallel W_i(\beta,t) - w_i(\beta,t) \parallel_2 + O_p(1)n^{-1} \sum_{i=1}^{n} \parallel W_i(\beta,t) - w_i(\beta,t) \parallel \\ & = O_p(n^{-1/2}h^{-1}), \quad \text{uniformly in } \beta \in \mathscr{B}_t, t \in (0,\tau). \end{aligned} \tag{A.2}$$

For each $j, k \in \{1, \ldots, p\}$, let

$$G_n(\beta,t) \equiv hn^{-1} \sum_{i=1}^{n} [w_{ij}(\beta,t)w_{ik}(\beta,t) - E(w_{ij}(\beta,t)w_{ik}(\beta,t))], \tag{A.3}$$

where $w_{ij}(\beta, t)$ is the jth element of $w_i(\beta, t)$. We next show that $|G_n(\beta, t)| = o_p((nh^2)^{-1/2})$ uniformly in $\beta \in \mathscr{B}_t$, $t \in (h, \tau - h)$ by applying Theorem 19.28 of van der Vaart (1998) for changing classes.

Let $l^\infty([0, \tau]$ be the space of all bounded functions on $[0, \tau]$ equipped with uniform topology. Consider the empirical process

$$Q_n(\beta,t) = \sqrt{n}\left[n^{-1}\sum_{i=1}^{n} f_{n,\beta,t}(X_i, M_i) - E f_{n,\beta,t}(X_i, M_i)\right],$$

where, for each $\beta \in \mathscr{B}$ and $t \in (h, \tau - h)$, $f_{n,\beta,t}(X_i, M_i) = \int_0^\tau hK_h(u-t)(X_i(u) - \bar{x}(\beta,u))\, dM_i(u)$ is a mapping from $l^\infty([0, \tau] \times l^\infty([0, \tau]$ to $\mathbb{R}^p$. Let $\mathscr{F}_n = \{f_{n,\beta,t} : \beta \in \mathscr{B}, t \in (h, \tau - h)\}$ be a sequence of the classes of functions. Then the classes $\mathscr{F}_n$ possess a uniform bounded envelop function independent of n, under the conditions (a), (b) and (e). Simple calculation shows that $E f_{n,\beta_1,t_1}(X_i, M_i)(f_{n,\beta_2,t_2}(X_i, M_i))^T - E f_{n,\beta_1,t_1}(X_i, M_i)(E f_{n,\beta_2,t_2}(X_i, M_i))^T \to 0$ for (β_1, t_1), $(\beta_2, t_2) \in \mathscr{B} \times [0, \tau]$. Further,

$$\begin{aligned} & E\| f_{n,\beta_1,t_1}(X_i, M_i) - E f_{n,\beta_2,t_2}(X_i, M_i) \|^2 \\ & \leq 2E\| f_{n,\beta_1,t_1}(X_i, M_i) - E f_{n,\beta_1,t_2}(X_i, M_i) \|^2 \\ & \quad + 2E\| f_{n,\beta_1,t_2}(X_i, M_i) - E f_{n,\beta_2,t_2}(X_i, M_i) \|^2 \\ & = K_1 \int_0^\tau (h^2(K_h(u-t_1) - K_h(u-t_2))^2 du + K_2 \int_0^\tau \| \bar{x}(\beta_1,u) - \bar{x}(\beta_2,u) \|^2 du \\ & = K_1'|t_1 - t_2| + K_2' \geq \| \beta_1 - \beta_2 \|, \end{aligned}$$

for some constants K_1, K_1', K_2, K_2'. Thus, the condition (19.27) of van der Vaart (1998) is satisfied. Since

$$\| f_{n,\beta_1,t_1}(X,M) - f_{n,\beta_2,t2}(X,M) \| \leq m(|t_1 - t_2| + \| \beta_1 - \beta_2 \|)$$

for some constants $m > 0$ under the conditions (a), (b) and (e), we have, by Example 19.7 (van der Vaart, 1998), the bracketing integral $J_{[]}(\delta_n, \mathscr{F}n, L_2(P)) \to 0$ for every $\delta_n \downarrow 0$. Hence, the processes $Q_n(\beta,t)=\sqrt{nh}[n^{-1}\sum_{i=1}^n w_i(\beta,t)-E(w_i(\beta,t))]$, $(\beta, t) \in \mathcal{B} \times [0, \tau]$, converges weakly to zero (the degenerated Gaussion process) by Theorem 19.28 of van der Vaart (1998). Since the classes $\mathcal{F}_n$ is uniformly bounded independent of n, we have, by Example 19.20 (van der Vaart, 1998), $\{\sqrt{nh}G_n(\beta,t),(\beta,t)\in\mathcal{B}\times[0,\tau]\}$ converges weakly to zero. Hence $\sup_{\beta\in\mathcal{B},t\in(h,\tau-h)}|G_n(\beta, t)| = o_p((nh^2)^{-1/2})$.

Based on (A.1), (A.2) and the rate for (A.3), and by the condition (e), there exists some $\epsilon > 0$, such that

$$\begin{aligned}
&hn^{-1}\sum_{i=1}^n U_i(\beta,t)(U_i(\beta,t))^T=hn^{-1}\sum_{i=1}^n w_i(\beta,t)(w_i(\beta,t))^T+o_p(n^{-\epsilon})\\
&=hn^{-1}\sum_{i=1}^n E(w_i(\beta,t)(w_i(\beta,t))^T+o_p(n^{-\epsilon})\\
&=hn^{-1}\sum_{i=1}^n\int_0^\tau K_h^2(u-t)E\left[\left(X_i(u)-\frac{s^{(1)}(\beta,u)}{s^{(0)}(\beta,u)}\right)^{\otimes 2}Y_i(u)\exp\{(\beta_0(u))^TX_i(u)\}\right]\\
&\quad\times\lambda_0(u)\,du+o_p(n^{-\epsilon})\\
&=h\int_0^\tau K_h^2(u-t)\left[s^{(2)}(\beta_0(u),u)-\frac{s^{(1)}(\beta_0(u),u)(s^{(1)}(\beta,u))^T}{s^{(0)}(\beta,u)}\right.\\
&\quad\left.-\frac{s^{(1)}(\beta,u)(s^{(1)}(\beta_0(u),u))^T}{s^{(0)}(\beta,u)}+\left(\frac{s^{(1)}(\beta,u)}{s^{(0)}(\beta,u)}\right)^{\otimes 2}s^{(0)}(\beta,u)\right]\lambda_0(u)\,du+o_p(n^{-\epsilon})\\
&=\nu_0\Sigma(t)+o_p(n^{-\epsilon})=\Gamma(t)+o_p(n^{-\epsilon})
\end{aligned}$$

uniformly in $\beta \in \mathcal{B}_t$, $t \in (h, \tau - h)$ as $nh^2 \to \infty$, where the last line is obtained under the condition (b).

Proof of Theorem 2.1.

Consider λ defined in Section 2.2 satisfying (2.7) and let $\rho = \|\lambda\|$ and $\theta = \lambda/\rho$. Then $|\theta| = 1$. Let $g(\lambda)=n^{-1}\sum_{i=1}^n\frac{U_i(\beta,t)}{1+\lambda^TU_i(\beta,t)}$. By (2.7),

$$\begin{aligned}
0\;&=\;\|g(\rho\theta)\|\geq|\theta^Tg(\rho\theta)|=n^{-1}\left|\theta^T\sum_{i=1}^n U_i(\beta,t)\left(1-\frac{\rho\theta^TU_i(\beta,t)}{1+\rho\theta^TU_i(\beta,t)}\right)\right|\\
&=\;n^{-1}\left|\theta^T\left(\sum_{i=1}^n U_i(\beta,t)-\rho\sum_{i=1}^n\frac{U_i(\beta,t)(U_i(\beta,t))^T\theta}{1+\rho\theta^TU_i(\beta,t)}\right)\right|\\
&\geq\;\frac{\rho\theta^T\tilde\Gamma(\beta,t)\theta}{h(1+\rho Z_n(\beta,t))}-n^{-1}\left|\theta^T\sum_{i=1}^n U_i(\beta,t)\right|,
\end{aligned}\tag{A.4}$$

where $Z_n(\beta, t) = \max_{1\le i\le n} \|U_i(\beta, t)\| = O_p(h^{-1})$. By Lemma A.1, $n^{-1}\|\sum_{i=1}^n U_i(\beta,t)\|=O_p((nh)^{-1/2}+h^2+\delta_n)$ and $\tilde\Gamma(\beta,t)=nh^{-1}\sum_{i=1}^n U_i(\beta,t)U_i^T(\beta,t)=\Gamma(t)+o_p(1)$. Let $\gamma(t) > 0$ be the smallest eigenvalue of $\Gamma(t)$. By the continuity of $\Gamma(t)$, there exits $c > 0$ such that such that $\gamma(t) \ge c$ for $t \in [0, \tau]$. Then, $\theta^T\tilde\Gamma(\beta, t)\theta \ge \gamma + o_p(1) \ge c+o_p(1)$. It follows from (A.4) that

$$\frac{\rho(c+o_p(1))}{h(1+\rho Z_n(\beta,t))}\leq\frac{\rho\theta^T\tilde\Gamma(\beta,t)\theta}{h(1+\rho Z_n(\beta,t))}\leq O_p((nh)^{-1/2}+h^2+\delta_n).$$

Hence $\alpha_n \equiv \rho/(1+\rho Z_n(\beta,t)) = O_p((n/h)^{-1/2}+h^3+h\delta_n)$ and $\rho = \alpha_n(1-\alpha_n Z_n(\beta,t))$. Since $\alpha_n Z_n(\beta,t) \to 0$, we have $\|\lambda\| = \rho = O_p((n/h)^{-1/2}+h^3+h\delta_n)$. This holds uniformly for $\beta \in \mathcal{B}_t$, $t \in (h, \tau-h)$ by Lemma A.1.

Since $\lambda^T U_i(\beta,t) = O_p((nh)^{-1/2}+h^2+\delta_n) = o_p(1)$, uniform in $\beta \in \mathcal{B}_t$, $t \in (h, \tau-h)$, applying the second order Taylor expansion to (2.8), we have

$$-2\log R(\beta,t) = 2\sum_{i=1}^{n}\left(\lambda^T U_i(\beta,t) - \frac{1}{2}(\lambda^T U_i(\beta,t))^2\right)+r_n, \tag{A.5}$$

where

$$\begin{aligned}|r_n| &= O_p(1)\sum_{i=1}^{n}|\lambda^T U_i(\beta,t)|^3\\ &=O_p(1)\|\lambda\|^3\max_{1\le i\le n}\|U_i(\beta,t)\|\cdot\|\sum_{i=1}^{n}U_i(\beta,t)U_i^T(\beta,t)\|\\ &=O_p(1)((n/h)^{-1/2}+h^3+h\delta_n)^3 h^{-1}(n/h)\\ &=O_p(1)[(nh)^{-1/6}+(nh^7)^{1/3}+(nh)^{1/3}\delta_n]^3,\end{aligned}$$

uniform in $\beta \in \mathcal{B}_t$, $t \in (h, \tau-h)$. By the equation (2.7),

$$\begin{aligned}0 &=\sum_{i=1}^{n}\frac{\lambda^T U_i(\beta,t)}{1+\lambda^T U_i(\beta,t)}\\ &=\sum_{i=1}^{n}(\lambda^T U_i(\beta,t))-\sum_{i=1}^{n}(\lambda^T U_i(\beta,t))^2+\sum_{i=1}^{n}\frac{(\lambda^T U_i(\beta,t))^3}{1+\lambda^T U_i(\beta,t)}\\ &=\sum_{i=1}^{n}(\lambda^T U_i(\beta,t))-\sum_{i=1}^{n}(\lambda^T U_i(\beta,t))^2+O_p(1)\left[(nh)^{-1/6}+(nh^7)^{1/3}+(nh)^{1/3}\delta_n\right]^3,\end{aligned}$$

we have $\sum_{i=1}^{n}(\lambda^T U_i(\beta,t))^2=\sum_{i=1}^{n}(\lambda^T U_i(\beta,t)+O_p(1)[(nh)^{-1/6}+(nh^7)^{1/3}+(nh)^{1/3}\delta_n]^3$. Thus

$$-2\log R(\beta,t)=\sum_{i=1}^{n}\lambda^T U_i(\beta,t)+O_p(1)[(nh)^{-1/6}+(nh^7)^{1/3}+(nh)^{1/3}\delta_n]^3 \tag{A.6}$$

Now, note that

$$\begin{aligned}0 &=n^{-1}\sum_{i=1}^{n}\frac{U_i(\beta,t)}{1+\lambda^T U_i(\beta,t)}=n^{-1}\sum_{i=1}^{n}U_i(\beta,t)\left(1-\lambda^T U_i(\beta,t)+\frac{(\lambda^T U_i(\beta,t))^2}{1+\lambda^T U_i(\beta,t)}\right)\\ &=n^{-1}\sum_{i=1}^{n}U_i(\beta,t)-\left(n^{-1}\sum_{i=1}^{n}U_i(\beta,t)U_i(\beta,t))^T\right)\lambda+n^{-1}\sum_{i=1}^{n}\frac{U_i(\beta,t)(\lambda^T U_i(\beta,t))^2}{1+\lambda^T U_i(\beta,t)}\\ &=n^{-1}\sum_{i=1}^{n}U_i(\beta,t)-\left(n^{-1}\sum_{i=1}^{n}U_i(\beta,t)(U_i(\beta,t))^T\right)\lambda+O_p((nh)^{-1/2}+h^2+\delta_n)^2,\end{aligned}$$

NIH-PA Author Manuscript
NIH-PA Author Manuscript

where the last term is obtained since

$n^{-1} \| \sum_{i=1}^{n} U_i(\beta,t)\,(\lambda^T U_i(\beta,t))/(1+\lambda^T U_i(\beta,t)) \| \le O_p(1) n^{-1} \| \lambda \|^2 \max_{1\le i\le n} \| U_i(\beta,t) \| \| \| \sum_{i=1}^{n} U_i(\beta,t) U_i^T(\beta,t) \| = O_p(($

It follows that

$$\lambda = \left(n^{-1}\sum_{i=1}^{n} U_i(\beta,t)(U_i(\beta,t))^T\right)^{-1}\left(n^{-1}\sum_{i=1}^{n} U_i(\beta,t)\right) + O_p[h((nh)^{-1/2}+h^2+\delta_n)^2]. \tag{A.7}$$

By (A.4), (A.6), (A.7) and Lemma A.1 we have, for some $\epsilon > 0$,

$$\begin{aligned} -2\log R(\beta,t) &= \left(n^{-1}\sum_{i=1}^{n} U_i(\beta,t)\right)^T \left(n^{-1}\sum_{i=1}^{n} U_i(\beta,t)(U_i(\beta,t))^T\right)^{-1} \sum_{i=1}^{n} U_i(\beta,t) \\ &\quad + O_p(1)[(nh)^{-1/6}+(nh^7)^{1/3}+(nh)^{1/3}\delta_n]^3 \\ &= \left((nh)^{1/2} n^{-1}\sum_{i=1}^{n} U_i(\beta,t)\right)^T (\Gamma(t))^{-1} \left((nh)^{1/2} n^{-1}\sum_{i=1}^{n} U_i(\beta,t)\right) \\ &\quad + o_p(n^{-\epsilon}) \end{aligned} \tag{A.8}$$

uniformly in $\beta \in \mathcal{B}_t$, $t \in (h, \tau - h)$. Hence, $-2\log R(\beta_0(t), t)$ converges in distribution to a chi-square distribution with p degrees of freedom. This completes the proof of Theorem 2.1.

Proof of Theorem 2.2

Let $\tilde{U}(\beta,t)=(h/n)^{1/2}\sum_{i=1}^{n} U_i(\beta,t)$. By (A.8),

$$\begin{aligned} -2\log R(\beta,t) &= \left((h/n)^{1/2}\sum_{i=1}^{n} U_i(\beta,t)\right)^T (\Gamma(t))^{-1} \left((h/n)^{1/2}\sum_{i=1}^{n} U_i(\beta,t)\right) \\ &\quad + o_p(n^{-\epsilon}) \end{aligned} \tag{A.9}$$

holds uniformly in $\beta \in \mathcal{B}_t$, $t \in (h, \tau - h)$. Let $\Gamma^{-1}(t) = D^T(t)D(t)$ and $D(t) = (d_1(t), \ldots, d_p(t))^T$, where $d_k(t)$ is the kth column of $D^T(t)$. Then

$$\begin{aligned} -2\log R(\beta,t) &= (D(t)\tilde{U}(\beta,t))^T D(t)\tilde{U}(\beta,t) + o_p(n^{-\epsilon}) \\ &= \sum_{k=1}^{P} (d_k^T(t)\tilde{U}(\beta,t))^2 + o_p(n^{-\epsilon}), \end{aligned} \tag{A.10}$$

uniformly in $\beta \in \mathcal{B}_t$, $t \in (h, \tau - h)$. Since $\sup_{h\le t\le \tau-h} \|\hat{\beta}(t) - \beta_0(t)\| = (nh)^{-1/2} O_p(n^{-\alpha})$, $\delta_n = (nh)^{-1/2} n^{-\delta}$, $0 < \delta < \alpha$, it follows that

$$-2\log R(\beta_1, \widehat{\beta}_{-1}(t), t) = \sum_{k=1}^{p} (d_k^T(t)\tilde{U}(\beta_1, \widehat{\beta}_{-1}(t), t))^2 + o_p(n^{-\epsilon}), \tag{A.11}$$

where β_1 is the first component of $\beta_0(t)$ and $\hat{\beta}^{-1}(t)$ is obtained by dropping the first component $\hat{\beta}_1(t)$ from $\hat{\beta}(t)$.

Note that $\tilde{U}(\hat{\beta}(t), t) = 0$ and $\Gamma(t) = \nu_0\Sigma(t) = -\nu_0(\partial\tilde{U}(\beta, t)/\partial\beta_1, \ldots, \partial\tilde{U}(\beta, t)/\partial\beta_p + o_p(n^{-\epsilon}))$, uniformly in $\beta \in \mathcal{B}_t$, $t \in (h, \tau - h)$. Let e_k be the kth column of a $p \times p$ identity matrix. It follows $\partial\tilde{U}(\beta,t)/\partial\beta_i = -\nu_0^{-1}\Gamma(t)e_i + o_p(n^{-\epsilon}) = -\nu_0^{-1}D^{-1}(t)(D^T(t))^{-1}e_1 + o_p(n^{-\epsilon})$. For $k = 1, \ldots, p$,

$$\begin{aligned} d_k^T(t)\tilde{U}(\beta_1,\widehat{\beta}_{-1}(t),t) &= d_k^T(t)\frac{\partial\tilde{U}(\beta,t)}{\partial\beta_1}(\widehat{\beta}_1(t) - \beta_1 + o_p(n^{-\epsilon}) \\ &= \nu^{-1}d_k^T(t)D^{-1}(t)(D^T(t))^{-1}e_1(\widehat{\beta}_1(t) + \beta_1) + o_p(n^{-\epsilon}) \\ &= \nu_0^{-1}e_k^T(D^T(t))^{-1}e_1(\widehat{\beta}_1(t) - \beta_1) + o_p(n^{-\epsilon}). \end{aligned}$$

It follows from (A.11) that

$$\begin{aligned} -2\log R(\beta_1\widehat{\beta}_{-1}(t),t) &= \nu_0^{-2}\sum_{k=1}^{p}[e_k^T(D^T(t))^{-1}e_1]^2(\widehat{\beta}_1(t) - \beta_1)^2 + o_p(n^{-\epsilon}) \\ &= \nu_0^{-2}\sum_{k=1}^{p}[e_1^T(D(t))^{-1}e_k][e_k^T(D^T(t))^{-1}e_1](\widehat{\beta}_1(t) - \beta_1)^2 \\ &\quad + o_p(n^{-\epsilon}) \\ &= \nu_0^{-2}[e_1^T(D(t))^{-1}(D^T(t))^{-1}e_1](\widehat{\beta}_1(t) - \beta_1)^2 + o_p(n^{-\epsilon}) \\ &= \nu_0^{-2}[e_1^T\Gamma(t)e_1](\widehat{\beta}_1(t) - \beta_1)^2 + o_p(n^{-\epsilon}). \end{aligned}$$

Since $\widehat{\beta}_1(t) - \beta_1 = -\nu_0 e_1^T\Gamma^{-1}(t)\tilde{U}(\beta,t) + o_p(n^{-\epsilon})$, we have

$$-2\log R(\beta_1,\widehat{\beta}_{-1}(t),t) = [(e_1^T\Gamma(t)e_1)^{1/2}(\Gamma^{-1}(t)e_1)^T\tilde{U}(\beta,t)]^2 + o_p(n^{-\epsilon}). \quad \text{(A.12)}$$

Let $p(t) = e_1^T\Gamma(t)e_1)(\Gamma^{-1}(t)e_1)^T(\nu_0\Gamma(t))(\Gamma^{-1}(t)e_1) = \nu_0^{-1}(e_1^T\Gamma(t)e_1)(e_1^T\Gamma^{-1}(t)e_1)$. It follows from Appendix B of Tian et al. (2005) that the distribution of the process $-2\log R(\beta_1, \hat{\beta}_{-1}(t), t)$ can be represented by the distribution of

$$[p^{1/2}(t)h^{1/2}\int_0^\tau K_h(t-s)\,dW(s)]^2 o_p(n^{-\epsilon}),$$

where $W(t)$ is a standard Wiener process. Let $\widehat{p}(t) = \nu_0^{-1}(e_1^T\widehat{\Gamma}(t)e_1)(e_1^T\widehat{\Gamma}^{-1}(t)e_1)$. Then $\hat{p}(t) = p(t) + o_p(n^{-\epsilon})$ uniformly in $\beta \in \mathcal{B}_t$, $t \in (h, \tau - h)$. Hence, the distribution of the process $-2\log R(\beta_1, \hat{\beta}_{-1}(t), t)/\hat{p}(t)$ can be represented by the distribution of

$$[h^{1/2}\int_0^\tau K_h(t-s)\,dW(s)]^2 + o_p(n^{-\epsilon}). \quad \text{(A.13)}$$

The assertion (i) follows since $h^{1/2}\int_0^\tau K_h(t-s)\,dw(s) \xrightarrow{\mathscr{D}} N(0,\nu_0)$.

Let $Z(t) = h^{1/2}\int_0^\tau K_h(t-s)\,dw(s)$. By Bickel and Rosenblatt (1973),

$$\sup_x \left| P\left\{ r_n\left(\sup_{t_1 \le t \le t_2}|Z(t)| - d_n\right) \le x \right\} - \exp(-2e^{-x}) \right| = O((\log n)^{-1}).$$

Note that for $x \ge -r_n d_n$, $P\{\sup_{t_1 \le t \le t_2} |Z(t)| \le d_n + x/r_n\} = P\{\sup_{t_1 \le t \le t_2} |Z(t)|^2 \le (d_n + x/r_n)^2\}$. We have

$$\sup_{x \ge -r_n d_n} \left| P\left\{ \frac{r_n}{2d_n} (\sup_{t_1 \le t \le t_2} |Z(t)|^2 - d_n^2) \le x + \frac{x^2}{2r_n d_n} \right\} - \exp(-2e^{-x}) \right| - O((\log n)^{-1}).$$

Let $y = x + (2r_n d_n)^{-1} x^2$. We have, for any $y > -2^{-1} r_n d_n$, $x = r_n d_n[-1 + (1 + 2y(r_n d_n)^{-1})^{1/2}] = 2y[1 + (1 + 2y(r_n d_n)^{-1})^{1/2}]^{-1}$, which converges to y since $r_n d_n \to \infty$ for each y. Further, $y - x = 2y^2 (r_n d_n)^{-1} [1 + (1 + 2y(r_n d_n)^{-1})^{1/2}]^{-2} = O((r_n d_n)^{-1}) = O((\log n)^{-1})$, uniformly in $|y| \le M$ with $M > 0$. Hence $\sup_{|y| \le M} |\exp(-2e^{-y}) - \exp(-2e-x)| = O((\log n)^{-1})$. It follows

$$\sup_{|y| \le M} \left| P\left\{ \frac{r_n}{2d_n} [\sup_{t_1 \le t \le t_2} |Z(t)|^2 - d_n^2] \le y) \right\} - \exp(-2e^{-y}) \right| - O((\log n)^{-1}). \tag{A.14}$$

The (2.11) follows from (A.13) and (A.14). Let

$$\begin{aligned} Q_0 &= \sup_{t_1 \le t \le t_2} \frac{|(\hat{\Gamma}^{-1}(t) e_1)^T \tilde{U}(\hat{\beta}, t)|}{[(e_1^T \hat{\Gamma}^{-1}(t) e_1) v_0^{-1}]^{1/2}} \\ Q_0^* &= \sup_{t_1 \le t \le t_2} \frac{|(\hat{\Gamma}^{-1}(t) e_1)^T U*(\hat{\beta}, t)|}{[(e_1^T \hat{\Gamma}^{-1}(t) e_1) v_0^{-1}]^{1/2}}. \end{aligned}$$

By Appendix B of Tian et al. (2005), we have

$$\sup_x \left| P\{r_n(Q_0 - d_n) < x)\} - P\left\{ r_n(Q_0^* - d_n) < x) \middle| \{(T_i, X_i(\cdot), \delta_i\} \right\} \right| = O(n^{-\epsilon}).$$

Following the previous arguments in proving (2.11) and by (A.12), we have

$$\begin{aligned} \sup_{x \ge -r_n d_n} \Bigg| & P\left\{ \tfrac{r_n}{2d_n} (Q - d_n^2) < x + \tfrac{x^2}{2r_n d_n}) \right\} \\ & - P\left\{ \tfrac{r_n}{2d_n} (Q^* - d_n^2) < x + \tfrac{x^2}{2r_n d_n}) \middle| \{(T_i, X_i(\cdot), \delta_i\} \right\} \Bigg| = O(n^{-\epsilon}). \end{aligned}$$

Hence,

$$\sup_{y \ge -\frac{1}{2} r_n d_n} \left| P\left\{ \frac{r_n}{2d_n} (Q - d_n^2) < y) \right\} - P\left\{ \frac{r_n}{2d_n} (Q^* - d_n^2) < y) \middle| \{(T_i, X_i(\cdot), \delta_i\} \right\} \right| = O(n^{-\epsilon}).$$

The assertion (ii) is hence proved.

Acknowledgments

The first author's research work was partially supported by NSF grants DMS-0304922 and DMS-0604576 and NIH grant 5 RO1 AI054165-05. The second author's research was supported by Intramural research award of NIH. The third author's research work was partially supported by NSA grant H98230-07-1-0002.

References

Aalen, OO. Lecture Notes in Statistics-2: Mathematical statistics and probability theory. New York: Springer-Verlag; 1980. A model for nonparametric regression analysis of counting processes; p. 1-25.

Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. Annals of Statistics 1982;10:1100–1120.

3. Bickel PJ, Rosenblatt M. On some global measures of the deviation of density function estimates. Annals of Statistics 1973;1:1071–1095.Corr. 3, 1370.

Buckley B, James IR. Linear regression with censored data. Biometrika 1979;66:429–436.

Cai Z, Sun Y. Local linear estimation for time-dependent coefficients in Cox's regression models. Scandinavian Journal of Statistics 2003;30:93–111.

Carter, WH.; Wampler, GL.; Stablein, DM. Regression analysis of survival data in cancer chemotherapy. New York: Dekker; 1983.

Chen K, Jin Z, Ying Z. Semiparametric analysis of transformation models with censored data. Biometrika 2002;89:659–668.

Cheng SC, Wei LJ, Ying Z. Analysis of transformation models with censored data. Biometrika 1995;82:835–845.

Chen SX, Qin YS. Empirical likelihood confidence intervals for local linear smoothers. Biometrika 2000;87:946–953.

Cox DR. Regression models and life tables (with discussion). Journal of the Royal Statistical Society, Series B 1972;34:187–220.

Cox DR. Partial likelihood. Biometrika 1975;62:269–276.

Dabrowska DM. Smoothed Cox regression. Annals of Statistics 1997;25:1510–1540.

Einmahl JH, McKeague IW. Confidence tubes for multiple quantile plots via empirical likelihood. Annals of Statistics 1999;27:1348–1367.

Fan J, Gijbels I, King M. Local likelihood and local partial likelihood in hazard regression. Annals of Statistics 1997;25:1661–1690.

Fleming, TR.; Harrington, DP. Counting Processes and Survival Analysis. New York: Wiley; 1991.

Gilbert PB, McKeague IW, Sun Y. The two-sample problem for failure rates depending on a continuous mark: an application to vaccine efficacy. 2007submitted

Hall P. Pseudo-likelihood theory for empirical likelihood. Annals of Statistics 1990;18:121–140.

Hall P. On edgeworth expansion and bootstrap confidence bands in non-parametric curve estimation. Journal of the Royal Statistical Society, Series B 1993;55:291–304.

19. Hall P, La Scala B. Methodology and algorithms of empirical likelihood. International Statistical Review 1990;58:109–127.

Hall P, Owen AB. Empirical likelihood confidence bands in density estimation. Journal of Computational and Graphical Statistics 1993;2:273–289.

Hjort NL, McKeague IW, Van Keilegom I. Extending the scope of empirical likelihood. Annals of Statistics. 2007In press

Hollander M, McKeague IW, Yang J. Likelihood ratio-based confidence bands for survival functions. Journal of the American Statistical Association 1997;92:215–226.

Huffer FW, McKeague IW. Weighted least squares estimation for Aalen's additive risk model. Journal of the American Statistical Association 1991;86:114–129.

Koul H, Susarla V, Van Ryzin J. Regression analysis with randomly right censored data. Annals of Statistics 1981;9:1276–1288.

Li G. On nonparametric likelihood ratio estimation of survival probabilities for censored data. Statistics and Probability Letters 1995;25:95–104.confidence bands 640

Li G, Van Keilegom I. Likelihood ratio confidence bands in nonparametric regression with censored data. Scandinavian Journal of Statistics 2002;29:547–562.

Lin D, Fleming T, Wei LJ. Confidence bands for survival curves under the proportional hazards model. Biometrika 1994;81:61–71.

Martinussen T, Scheike TH, Skovgaard IM. Efficient estimation of fixed and time- varying covariates effects in multiplicative intensity models. Scandinavian Journal of Statistics 2002;29:59–77.

McKeague IW, Zhao Y. Simultaneous confidence bands for ratios of survival functions via empirical likelihood. Statistics and Probability Letters 2002;60:405–415.

Murphy SA. Likelihood ratio-based confidences in survival analysis. Journal of the American Statistical Association 1995;90:1399–1405.

Murphy SA, Sen PK. Time-dependent coefficients in a Cox-type regression model. Stochastic Processes and their Applications 1991;39:153–180.

Owen AB. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 1988;75:237–249.

Owen AB. Empirical likelihood and confidence regions. Annals of Statistics 1990;18:90–120.

Owen AB. Empirical likelihood for linear models. Annals of Statistics 1991;19:1725–1747.

Owen, AB. Empirical likelihood. London: Chapman & Hall/Crc; 2001.

Pons O. Nonparametric estimation in a varying-coefficient Cox model. Mathematical Methods of Statistics 2000;9:376–398.

Qin G, Jing BY. Empirical likelihood for Cox regression model under random censorship. Communications in Statistics: Simulation and Computation 2001;30:79–90.

Qin J, Lawless JF. Empirical likelihood and general estimating equations. Annals of Statistics 1994;22:300–325.

Stablein DM, Carter WH, Novak JW. Analysis of survival data with non-proportional hazard functions. Controlled Clinical Trials 1981;2:149–159. [PubMed: 7273792]

Thomas DR, Grunkemeier GL. Confidence interval estimation for survival probabilities for censored data. Journal of the American Statistical Association 1975;70:865–871.

Tian L, Zucker D, Wei LJ. On the Cox model with time-varying regression coefficients. Journal of the American Statistical Association 2005;100:172–183.

Van der Vaart, AW. Asymptotic Statistics. Cambridge: Cambridge University Press; 1998.

Yandell B. Nonparametric inference for rates with censored survival data. Annals of Statistics 1983;11:1119–1135.

Zhang J, Liu A. Local polynomial fitting based on empirical likelihood. Bernoulli 2003;9:579–605.

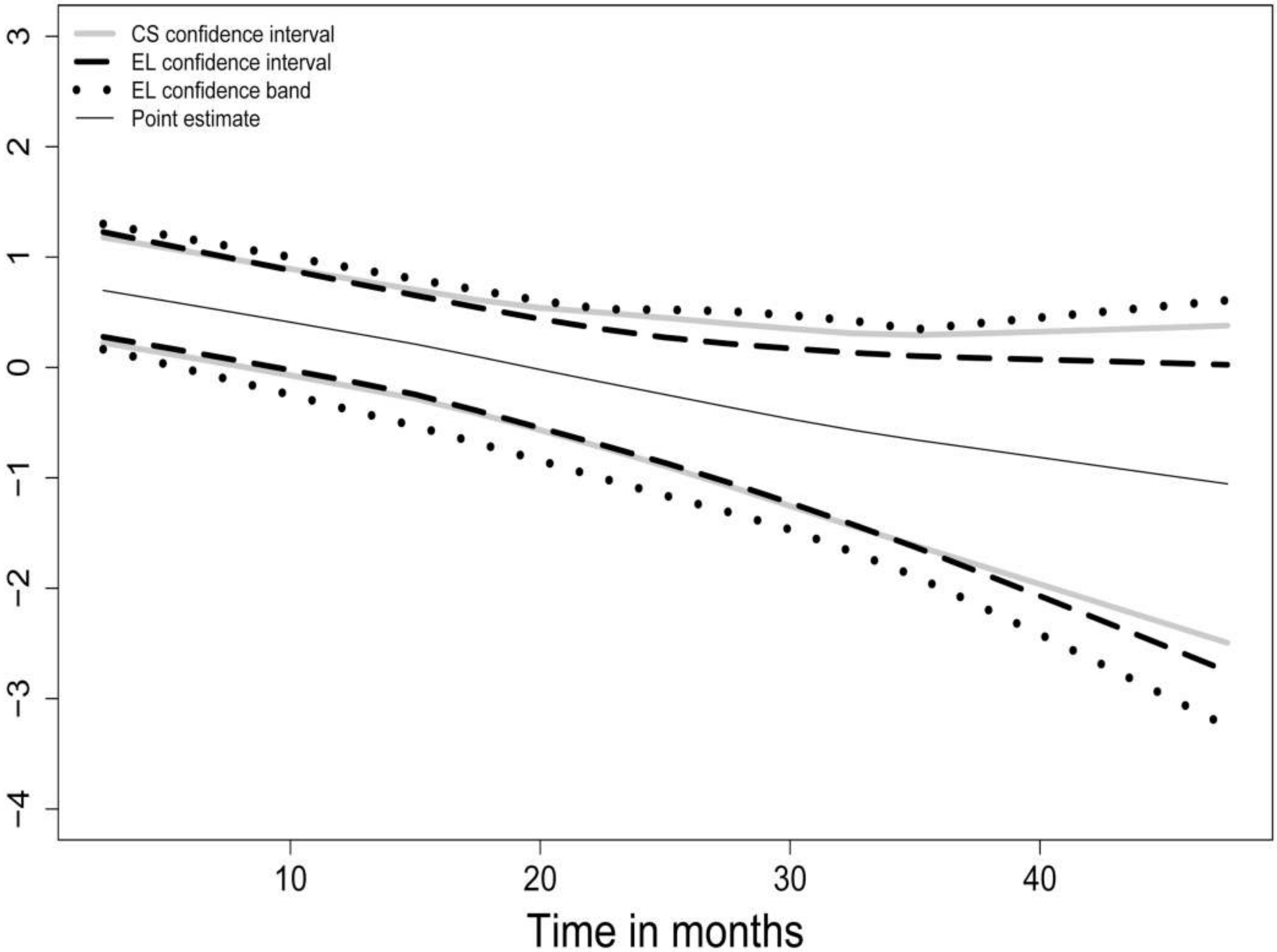

FIG 1.
Comparison of 95% confidence intervals for time-dependent treatment effect based on gastric cancer data.

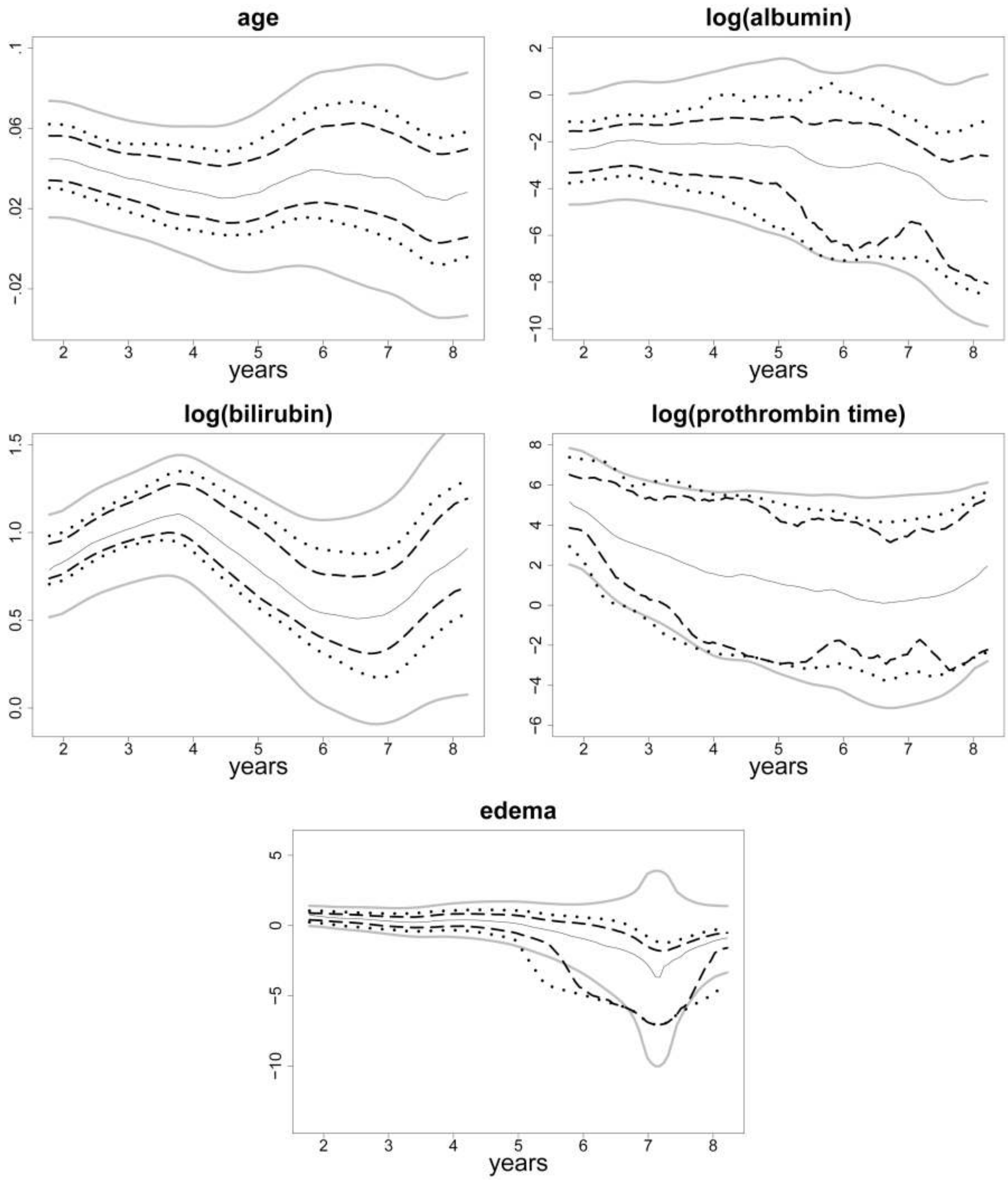

FIG 2.

Comparisons of 95% confidence intervals for time-dependent coefficients corresponding to Age, Log(albumin), Log(bilirubin), Log(prothrombin time) and Edema based on Mayo Primary Cirrhosis Data (center solid black line{point estimate; dashed lines–EL confidence interval; dotted lines–EL confidence band; solid grey lines–confidence interval based on Cai an Sun (2003)).

TABLE 1

Coverage probabilities for 95% pointwise confidence intervals for $\beta_0(t)$ based on the model (3.1).

		t	.35	.55	.75	.95	1.15	.35	.55	.75	.95	1.15
n	***h***	**Method**			**Censoring=10%**					**Censoring=30%**		
250	.2	ELSQ	94.9	94.5	94.6	93.6	93.9	94.3	94.8	94.2	93.6	92.1
		ELGM	95.4	94.8	94.5	94.1	93.2	94.6	94.6	94.7	93.4	92.6
		CS	95.0	95.6	95.6	95.4	96.0	94.9	96.9	96.3	97.1	98.6
	.3	ELSQ	94.9	95.1	95.1	94.5	96.1	94.2	94.8	94.9	94.1	93.0
		ELGM	94.3	94.3	94.7	94.8	95.2	94.3	94.5	95.2	94.3	93.3
		CS	95.1	93.9	95.1	95.5	95.6	94.4	95.0	95.5	95.3	96.2
500	.2	ELSQ	94.2	93.8	95.1	93.8	93.8	94.7	94.9	94.2	94.4	94.6
		ELGM	94.6	94.8	96.0	94.5	95.3	95.8	95.3	95.8	95.6	94.6
		CS	95.1	94.4	94.6	95.4	95.1	94.5	95.1	94.9	94.0	94.5
	.3	ELSQ	93.3	94.7	95.2	96.4	94.0	94.6	94.3	94.4	94.6	95.4
		ELGM	94.0	95.1	94.8	95.4	94.6	94.4	94.9	94.7	94.9	94.9
		CS	93.8	94.2	94.9	95.8	95.7	95.2	95.5	96.6	96.1	95.6
750	.2	ELSQ	94.2	94.9	95.2	94.9	95.0	94.2	93.7	95.0	96.2	94.9
		ELGM	94.3	94.0	95.4	95.1	94.9	94.3	95.2	94.2	94.8	95.4
		CS	94.4	93.7	95.5	94.4	95.4	94.0	95.6	94.5	95.2	94.4
	.3	ELSQ	93.1	93.5	94.9	95.8	94.7	93.3	95.3	95.0	95.0	94.8
		ELGM	93.7	94.1	95.2	96.1	94.2	93.9	94.6	94.6	95.4	95.6
		CS	92.9	94.6	93.8	94.9	94.9	93.1	94.9	94.2	94.9	95.6

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TABLE 2

Coverage probabilities for 95% simultaneous confidence bands for $\beta_0(t)$; $t \in [.25; 1.25]$ based on the model (3.1).

h	Method		Censoring=10%			Censoring=30%	
		n= 250	**n= 500**	**n = 750**	**n= 250**	**n= 500**	**n = 750**
.2	ELGM	93.6	95.1	94.9	95.0	95.5	94.8
	ELEX	76.8	82.8	85.1	58.9	82.1	84.8
.3	ELGM	92.1	94.4	94.8	93.3	93.5	94.6
	ELEX	72.9	70.9	71.9	66.6	68.8	68.9

TABLE 3

Coverage probabilities for 95% pointwise confidence intervals for $\beta_1(t)$ based on the model (3.2).

		t	**.35**	**.55**	**.75**	**.95**	**1.15**	**.35**	**.55**	**.75**	**.95**	**1.15**
n	***h***	**Method**			**Censoring=10%**					**Censoring=30%**		
250	.5	ELSQ	95.4	94.6	95.7	94.8	95.3	94.0	93.7	93.7	94.4	94.3
		ELGM	95.0	94.0	94.3	94.7	94.4	94.5	95.0	95.3	94.7	93.5
	.6	ELSQ	93.9	93.5	94.5	94.9	95.0	95.0	93.9	94.7	93.7	93.3
		ELGM	94.8	92.7	94.0	95.4	93.7	95.1	94.6	94.2	92.9	93.8
500	.5	ELSQ	94.1	94.4	93.5	94.2	94.2	95.9	94.8	93.3	94.0	93.4
		ELGM	94.8	94.2	94.9	95.8	94.8	93.9	93.8	93.5	94.4	94.2
	.6	ELSQ	95.2	94.6	94.0	94.6	94.7	94.6	92.3	93.5	93.9	93.1
		ELGM	94.7	93.6	94.4	94.5	94.3	94.0	93.4	93.4	94.3	93.2
750	.5	ELSQ	94.4	93.4	94.3	93.0	94.0	94.8	94.5	93.9	94.0	93.8
		ELGM	97.1	94.9	94.8	93.8	93.4	95.4	93.5	93.4	93.7	94.4
	.6	ELSQ	94.8	94.7	94.1	94.0	93.4	94.6	93.3	93.9	93.8	93.9
		ELGM	96.8	93.7	93.2	93.0	93.3	94.7	93.0	93.2	93.0	93.6

TABLE 4

Coverage probabilities for 95% simultaneous confidence bands for $\beta_1(t)$; $t \in [.25; 1.25]$ based on the model (3.2).

		Censoring=10%			Censoring=30%		
h	**Method**	**n= 250**	**n= 500**	**n = 750**	**n= 250**	**n= 500**	**n = 750**
.5	ELGM	95.1	95.6	96.3	94.3	95.8	95.3
.6	ELGM	96.0	95.2	94.1	95.6	95.0	94.4