Harvard University
Harvard University Biostatistics Working Paper Series

Year 2003 *Paper* 4

On the Cox Model with Time-Varying Regression Coefficients

Lu Tian[*] David Zucker[†]
L. J. Wei[‡]

[*]Harvard School of Public Health, ltian@hsph.harvard.edu
[†]Hebrew University
[‡]Harvard School of Public Health, wei@hsph.harvard.edu

ON THE COX MODEL WITH TIME-VARYING REGRESSION COEFFICIENTS

Lu Tian

Department of Biostatistics, Harvard University

677 Huntington Ave., Boston, MA 02115

David Zucker

Department of Statistics, Hebrew University

Mount Scopus, Jerusalem 91905, Israel

L.J. Wei

Department of Biostatistics, Harvard University

677 Huntington Ave., Boston, MA 02115

SUMMARY

In the analysis of censored failure time observations, the standard Cox proportional hazards model assumes that the regression coefficients are time-invariant. Often, these parameters vary over time, and the temporal covariate effects on the failure time are of great interest. In this article, along the lines of Cai and Sun (2003) we propose a simple estimation procedure for the Cox model with time-varying coefficients based on a kernel-weighted partial likelihood approach. Point-wise and simultaneous confidence intervals for the regression parameters over a properly chosen time interval are constructed via a resampling technique. A prediction method for future patients' survival with any specific set of covariates is derived. Building on the estimates for the time-varying coefficients, we also consider the mixed case and present an estimation procedure for time-independent parameters in the model. Furthermore, we show how to utilize an integrated function of the estimate for a specific regression coefficient to examine the adequacy of proportional hazards assumption for the corresponding covariate graphically and numerically. All the proposals are illustrated extensively with a well-known study from the Mayo Clinic.

KEY WORDS: *Confidence band; Kernel estimation; Martingale; Model checking and selection; Partial likelihood; Prediction; Survival analysis.*

1. INTRODUCTION

The most popular semi-parametric regression model for analyzing survival data is the proportional hazards (PH) model (Cox, 1972). This model relates an individual subject's hazard function to its covariates multiplicatively without assuming a parametric underlying hazard function. The large sample properties of the inference procedures for the PH model have been justified elegantly via martingale theory (Andersen and Gill, 1982). The standard PH model assumes that the regression coefficients are constant over time. Often, however, the regression parameters may vary over time, and it is important to know the temporal effects of the covariates on the failure time. For example, in an HIV-AIDS study comparing a new treatment with an active control, suppose that the time to a clinical or virologic event is the primary endpoint. The new drug may work well in the initial treatment period, but may gradually lose its potency due to mutation of the virus. If the drug does have the potential to lose its efficacy, it is crucial to know when and how fast the drug becomes ineffective. The Cox model with a time-varying coefficient for the treatment difference may shed light on these issues and help us to design future studies to explore optimal treatment strategies for HIV-infected patients. Moreover, even when the physical interpretation of individual covariate effects over time is not straightforward due to, for example, the presence of highly correlated covariates, the Cox model with time-varying coefficients is much more flexible than the conventional PH model for making prediction of future patients' survival.

The theoretical properties of certain inference procedures for the Cox model with time-varying regression coefficients have been studied, for example, by Zucker and Karr (1990) using a penalized partial likelihood approach, and by Murphy and Sen (1991) using the sieve method. Although Zucker and Karr (1990) proved consistency of their estimator for a general covariate vector, its asymptotic distribution is fully established only for the case with a single covariate. Recently Verweij and Houwelingen (1995) recommended a practical choice of the penalty parameter for the method studied by Zucker and Karr in a discrete time setup. The Zucker-Karr and Murphy-Sen estimation procedures involve maximization of functions over a parameter space whose dimension increases with the sample size. This optimization problem can be rather complex. Furthermore, there are no guidelines available for choosing, for instance, the "segments" of the sieve method (Murphy, 1993). With an initial consistent estimator, for example the sieve estimator by Murphy and Sen (1991), Martinussen and Scheike (2002) and Martinussen, Scheike and Skovgaard (2002) proposed a novel one-step estimation procedure for the *cumulative* parameter function of the semi-parametric hazard model with time-varying regression coefficients. Recently Winnett and Sasieni (2003) studied non-parametric estimates for the time-varying coefficients based on Schoenfeld residuals derived from the PH model (Schoenfeld, 1982).

In this paper, we use a novel, kernel-weighted partial likelihood technique considered by Valsecchi, Silvestri and Sasieni (1996) and Cai and Sun (2003) to construct a simple estimation procedure for the Cox model with time-varying coefficients. At each time point, the estimate is obtained by maximizing a smooth, convex function of a $p \times 1$ vector of parameters, where p is the dimension of the vector of covariates. The point-wise consistency and asymptotic normality of the resulting estimator have been established by Cai and Sun (2003). When one is interested in making inferences about a function over time, however, the point-wise distribution of the estimator for the function is of rather limited use. Valid inferences about the temporal covariate effects cannot be drawn based on such a distribution theory. For instance, a naive band constructed from point-wise confidence intervals for a time-dependent coefficient does not have the correct coverage probability, and conclusions about the covariate effect over time based on the point-wise intervals can be quite misleading. With general nonparametric density or rate function estimates, the above estimate does not converge weakly to a non-degenerate process, even after standardization. On the other hand, using the so-called strong approximation technique along with a novel resampling method, we are able to construct confidence bands for a time-dependent regression coefficient over a properly chosen time interval. These bands are quite informative for examining the temporal effects of a covariate over the entire time-span of interest with a certain degree of confidence. With this function estimate, we propose a prediction method for the survival function for future subjects with a specific set of covariates. Furthermore, we present an inference procedure for the mixed case, that is, a portion of the parameter vector in the above Cox model is time-independent. Lastly, we demonstrate how to use an integrated function of the estimate for a specific regression coefficient to check the proportional hazards assumption for the corresponding covariate graphically and numerically. All the proposals presented here are illustrated with the well-known Mayo primary biliary cirrhosis data.

Recently, Gilbert et al. (2002) derived nonparametric inference procedures for the ratio of two hazard functions, a very special case of the Cox model with time-dependent regression coefficients. The kernel-weighted likelihood-based approach has also been used in a different context for regression analysis (Staniswalis, 1989; Hunsberger, 1994). Models with varying regression coefficients other than the Cox model have been studied, for example, by Hastie and Tibshirani (1993), Cai, Fan and Li (2000), Fan and Zhang (2000) and Huang, Wu and Zhou (2002).

2. POINT ESTIMATION FOR TIME-DEPENDENT REGRESSION COEFFICIENTS

Let T be the failure time and C be the corresponding censoring variable. Also, let $\mathbf{Z}(t)$ be a possibly time dependent, p-vector-valued covariate process, which is bounded and predictable. Conditional on $\mathbf{Z}(\cdot)$, T and C are assumed to be independent. Let $\{(T_i, \mathbf{Z}_i(\cdot), C_i), i = 1, \cdots, n\}$ be n independent copies of $\{(T, \mathbf{Z}(\cdot), C)\}$. For the ith subject, one can only observe $\{(X_i, \mathbf{Z}_i(\cdot), \Delta_i)\}$, where $X_i = \min(T_i, C_i)$, and Δ_i equals 1 if $X_i = T_i$ and 0 otherwise. For the ith subject, the Cox proportional hazards function (Cox, 1972; Andersen and Gill, 1982) with time-dependent regression coefficients is

$$\lambda_i(t) = \lambda_0(t) e^{\boldsymbol{\beta}_0(t)'\mathbf{Z}_i(t)}, \tag{2.1}$$

where $\boldsymbol{\beta}_0(t)$ is a "smooth" function of t, and $\lambda_0(\cdot)$ is a completely unspecified underlying hazard function. We are interested in estimating the regression coefficient function

$\{\boldsymbol{\beta}_0(t), t > 0\}$ based on $\{(X_i, \mathbf{Z}_i(\cdot), \Delta_i), i = 1, \cdots, n\}$.

For a fixed time point t, let us consider a weighted "local" log partial likelihood function of the p-vector $\boldsymbol{\beta}$ to estimate $\boldsymbol{\beta}_0(t)$:

$$\mathcal{L}(\boldsymbol{\beta}, t) = (nh_n)^{-1} \sum_{i=1}^{n} \int_0^{\tau} K\left(\frac{s-t}{h_n}\right) \left[\boldsymbol{\beta}'\mathbf{Z}_i(s) - \log\left(\sum_{j=1}^{n} Y_j(s) e^{\boldsymbol{\beta}' Z_j(s)}\right)\right] dN_i(s), \quad (2.2)$$

where the kernel function $K(\cdot)$ is a symmetric probability density function with support $[-1, 1]$, mean 0, and bounded first derivative, $h_n = O(n^{-v})$ with $v > 0$, $Y_j(t) = I(X_j \geq t)$, $I(\cdot)$ is the indicator function, τ is a pre-specified constant such that $\text{pr}(X_i > \tau) > 0$, and $N_i(t) = I(X_i \leq t, \Delta_i = 1)$. The function (2.2) is convex in $\boldsymbol{\beta}$. For $t \in [h_n, \tau - h_n]$, let $\hat{\boldsymbol{\beta}}(t)$ be the maximum of (2.2) with respect to $\boldsymbol{\beta}$ (Cai and Sun, 2003). For $t < h_n$ and $t > \tau - h_n$, we let $\hat{\boldsymbol{\beta}}(t) = \hat{\beta}(h_n)$ and $\hat{\beta}(\tau - h_n)$, respectively. Note that the summation in $\mathcal{L}(\boldsymbol{\beta}, t)$ involves only subjects whose observed failure times are in a small neighborhood of t.

For $t \in [h_n, \tau - h_n]$, the maximum local partial likelihood estimator $\hat{\boldsymbol{\beta}}(t)$ is a root of the score equation $\mathbf{U}(\boldsymbol{\beta}, t) = 0$, where

$$\mathbf{U}(\boldsymbol{\beta}, t) = (nh_n)^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} (\mathbf{Z}_i(s) - \mathbf{E}(\boldsymbol{\beta}, s)) K\left(\frac{s-t}{h_n}\right) dN_i(s),$$

$$\mathbf{E}(\boldsymbol{\beta}, t) = S^{(1)}(\boldsymbol{\beta}, t) / S^{(0)}(\boldsymbol{\beta}, t),$$

$$S^{(r)}(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^{n} Y_i(t) \mathbf{Z}_i(t)^{\otimes r} e^{\boldsymbol{\beta}' \mathbf{Z}_i(t)}, r = 0, 1, 2.$$

Cai and Sun (2003) showed the point-wise consistency of $\hat{\boldsymbol{\beta}}(t)$ for each fixed time t. Under the mild regularity conditions A.1-A.3 stated in Appendix A, which are slightly stronger than those given in Andersen and Gill (1982) for the Cox model with fixed regression coefficients, one can show that if the kernel smoothing parameter $h_n = O(n^{-v})$ with $0 < v < 1$, $\hat{\boldsymbol{\beta}}(t)$ is *uniformly* consistent in the sense that $\sup_{t \in [0,\tau]} \|\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\| = o_p(1)$. A detailed proof of consistency is given in Appendix A. This stronger version of consistency is needed for establishing simultaneous inference procedures for $\boldsymbol{\beta}_0(\cdot)$ and the cumulative hazard function of $\lambda_0(\cdot)$.

One of the most challenging problems in the area of nonparametric function estimation is how to choose the bandwidth h_n in practice. In this paper, we used the $\mathcal{K}$-fold cross validation procedure for bandwidth selection commonly employed in the literature (Efron and Tibshirani, 1993, p. 240). In Sections 4, 5 and 6, we show empirically that the choice of the smoothing parameter can be quite flexible.

3. POINT-WISE AND SIMULTANEOUS INTERVAL ESTIMATION FOR REGRESSION COEFFICIENTS

First, suppose that we are interested in constructing confidence intervals for a contrast $\mathbf{a}'\boldsymbol{\beta}_0(t)$ at a fixed time point t, where $\mathbf{a}$ is a $p-$vector of known constants. By Taylor expansion of the score function $\mathbf{U}(\hat{\boldsymbol{\beta}}(t), t)$ around $\boldsymbol{\beta}_0(t)$, if $1/5 < v < 1$, then

$$(nh_n)^{1/2}(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)) \approx \mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}(t), t)\mathbf{U}(\boldsymbol{\beta}_0(t), t). \tag{3.1}$$

Here, $\mathbf{I}(\boldsymbol{\beta}, t) = -\frac{\partial^2 \mathcal{L}(\boldsymbol{\beta}, t)}{\partial \boldsymbol{\beta}^2} = (nh_n)^{-1} \sum_{i=1}^n \int_0^\tau \mathbf{V}(\boldsymbol{\beta}, s) K\left(\frac{s-t}{h_n}\right) dN_i(s)$, where

$$\mathbf{V}(\boldsymbol{\beta}, t) = \frac{S^{(2)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} - \left(\frac{S^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)}\right)^{\otimes 2}.$$

Using an argument similar to that of Cai and Sun (2003), one can show that if $1/5 < v < 1$,

$$\mathbf{U}(\boldsymbol{\beta}_0(t), t) \approx (nh_n)^{-1/2} \sum_{i=1}^n \int_0^\tau (\mathbf{Z}_i(s) - \mathbf{E}(\boldsymbol{\beta}_0(t), s)) K\left(\frac{s-t}{h_n}\right) dM_i(s), \tag{3.2}$$

where $M_i(s) = N_i(s) - \int_0^s Y_i(t)\lambda_i(t)dt$. Furthermore, for any fixed $t \in [h_n, \tau - h_n]$, the distribution of $(nh_n)^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t))$ is approximately normal with mean 0 and covariance matrix $\mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}(t), t) \int_{-1}^1 K^2(s)ds$. For any given p-vector $\mathbf{a}$, point-wise confidence intervals for the contrast $\mathbf{a}'\boldsymbol{\beta}_0(t)$ can be constructed using this large sample approximation to the distribution of $\hat{\boldsymbol{\beta}}(t)$.

Since we are interested in the temporal behavior of $\{\mathbf{a}'\boldsymbol{\beta}_0(t)\}$, inferences based on confidence bands for this function over a properly chosen time interval, say, $[b_1, b_2] \subset [0, \tau]$, are more informative than those based on point-wise intervals. To obtain such confidence bands for $\{\mathbf{a}'\boldsymbol{\beta}_0(t), t \in [b_1, b_2]\}$, a standard approach is to derive a large sample approximation to the distribution of

$$\mathcal{S} = \sup_{t\in[b_1,b_2]} \hat{w}(t)|\mathbf{a}'(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t))|, \tag{3.3}$$

where $\hat{w}(t)$ is a possibly data-dependent, positive weight function that converges uniformly to a deterministic function. Let c_α be the $100(1-\alpha)$th percentile of this approximate distribution, where $0 < \alpha < 1$. Then, a $1-\alpha$ confidence band for $\{\mathbf{a}'\boldsymbol{\beta}_0(t), t \in [b_1, b_2]\}$ is simply

$$\{\mathbf{a}'\hat{\boldsymbol{\beta}}(t) \pm c_\alpha \hat{w}(t)^{-1}, b_1 \le t \le b_2\}. \tag{3.4}$$

Like standard nonparametric kernel density function estimates, the process $\mathbf{U}(\boldsymbol{\beta}_0(t), t)$ in (3.2) is not tight. It follows from (3.1) that $\{\hat{w}(t)(nh_n)^{\frac{1}{2}}\mathbf{a}'(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)), t \in [b_1, b_2]\}$ does not converge to a process, and one cannot apply the continuous mapping theorem to obtain a large sample approximation to the distribution of (3.3). On the other hand, we may utilize the so-called strong approximation technique presented in Bickel and Rosenblatt (1973) and Yandell (1983) to obtain an approximation to the distribution of a standardized (3.3). Specifically, in Appendix B, we show that if $1/5 < v < 1$, the distribution of a standardized version of $\mathcal{S}$ in (3.3) can be approximated by an extreme value distribution. This approximation with the standardization parameters is given explicitly in (9.4) of Appendix B. It is well known, however, that this type of the analytic approximation is not very accurate (Hall, 1993).

Here, we propose a simulation technique to obtain a more accurate approximation to the distribution of the standardized $\mathcal{S}$. To this end, consider a stochastic perturbation of (3.2) defined by

$$\tilde{\mathbf{U}}(t) = (nh_n)^{-1/2} \sum_{i=1}^{n} \int_0^\tau (\mathbf{Z}_i(s) - \mathbf{E}(\hat{\boldsymbol{\beta}}(t), s)) K\left(\frac{s-t}{h_n}\right) dN_i(s) G_i, \tag{3.5}$$

where $\{G_i, i = 1, \cdots, n\}$ is a random sample from the standard normal distribution, and is independent of the data $\{(X_i, \mathbf{Z}_i(\cdot), \Delta_i), i = 1, \cdots, n\}$. Note that (3.5) is obtained by replacing $\boldsymbol{\beta}_0(t)$ and $M_i(s)$ on the right-hand side of (3.2) with $\hat{\boldsymbol{\beta}}(t)$ and $N_i(s)G_i$. It follows from the argument of Lin, Fleming and Wei (1994) that, for any fixed t, conditional on the data, the limiting distribution of $\tilde{\mathbf{U}}(t)$ in (3.5) is the same as the unconditional limiting distribution of $\mathbf{U}(\boldsymbol{\beta}_0(t), t)$. It is important to note that this equivalence relation does not hold if we treat these two functions as processes in t. On the other hand, in Appendix B we show that, conditional on the data $\{(X_i, \mathbf{Z}_i(\cdot), \Delta_i), i = 1, \cdots, n\}$, the distribution of the standardized

$$\tilde{\mathcal{S}} = \sup_{t \in [b_1, b_2]} (nh_n)^{-1/2} |\hat{w}(t) \mathbf{a}' \mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}(t), t) \tilde{\mathbf{U}}(t)| \tag{3.6}$$

can be used to approximate the unconditional distribution of its counterpart $\mathcal{S}$ defined by (3.3). The advantage of using such an approximation over the analytic one (9.4) is shown in (9.5).

In practice, to obtain the conditional distribution of (3.6), we replace all the random quantities in (3.5) by their observed counterparts, except for the random variables $\{G_i, i = 1, \cdots, n\}$. Subsequently, for each generated $\{G_i, i = 1, \cdots, n\}$, we compute a realized (3.5) and then (3.6). With a large number, say M, of such realizations of (3.6), the resulting $100(1-\alpha)$th empirical percentile can be used to approximate the cutoff point c_α for the confidence band (3.4).

We now use the Mayo primary biliary cirrhosis data (Fleming and Harrington, 1991, Appendix D) to illustrate our proposals. This data set consists of 418 patient records, each of which contains the patient's survival time in days and seventeen potential prognostic factors. In the original data set, there were two patients whose covariate values were incomplete. We deleted these two records in our analysis. Furthermore, to simplify the illustration we only considered five covariates in Model (2.1): age, log(albumin), log(bilirubin), edema, and log(prothrombin time), which were selected as the important predictors for the "final" Cox regression model with time-invariant regression parameters for prediction (Dickson et al., 1989; Fleming and Harrington, 1991, p. 195). However, Fleming and Harrington (1991, p. 191) indicated that log(prothrombin time) and edema do not satisfy the proportional hazards assumption, and the standard Cox model with time-independent regression coefficients can be improved. Here, we show that the effect from log(prothrombin time) on survival appears to diminish over time.

To estimate $\boldsymbol{\beta}_0(t)$ via (2.2), we use the kernel function $K(x) = 3(1-x^2)/4$, $-1 \le x \le 1$, the so-called Epanechnikov kernel (Andersen et al., 1993, p. 233). Also, we let $\tau = 4000$ (days). To choose the bandwidth $h = h_n$, we used a $\mathcal{K}$-fold cross-validation method, which is commonly used in the nonparametric function estimation literature (Hoover et al., 1988; Cai et al., 2000; Efron and Tibshirani, 1993). For the present case, we tried various scenarios to investigate the robustness of this cross validation procedure. For example, in one case, we split the data into $\mathcal{K} = 13$ equal-sized parts based on the patient's ID. For a fixed h, we deleted the kth part, $k = 1, \cdots, \mathcal{K}$, and fitted Model (2.1) to the other $\mathcal{K} - 1$ parts of the data. We then calculated the "prediction error", $\mathrm{PE}_k(h)$, of the fitted model when predicting the kth part of the data. We repeated this process and obtained the total prediction error, $\mathrm{PE}(h) = \sum_{k=1}^{13} \mathrm{PE}_k(h)$. The "optimal" bandwidth was chosen by minimizing $\mathrm{PE}(h)$ with respect to h. Two types of prediction error criteria were used in our analysis. For the first one, $\mathrm{PE}_k(h)$ is the minus logarithm of the standard partial likelihood function, which is

$$-\sum_{\{m \in D_k\}} \int_0^\tau \left[\hat{\boldsymbol{\beta}}(s)' \mathbf{Z}_m(s) - \log \left(\sum_{\{d \in D_k\}} Y_d(s) e^{\hat{\boldsymbol{\beta}}(s)' \mathbf{Z}_d(s)} \right) \right] dN_m(s),$$

where D_k is the index set for the kth part of the data set, and $\hat{\boldsymbol{\beta}}(s)$ is the maximum local partial likelihood estimate via (2.2) based on the other $\mathcal{K} - 1$ parts of the data. There is a unique minimizer for $\mathrm{PE}(h)$ at $h = 690$ (days). The second prediction error criterion we used for selecting h is based on the martingale residuals, that is,

$$\mathrm{PE}_k(h) = \sum_{\{m \in D_k\}} \int_0^\tau \left(N_m(t) - \int_0^t Y_m(s) e^{\hat{\boldsymbol{\beta}}(s)' \mathbf{Z}_m} d\hat{\Lambda}(\hat{\boldsymbol{\beta}}(\cdot), s) \right)^2 d\{ \sum_{\{d \in D_k\}} N_d(t) \},$$

where $\hat{\Lambda}(\hat{\boldsymbol{\beta}}(\cdot), s)$ is a generalized Breslow estimator given in (4.1) for the cumulative hazard function $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$ based on the data from the study patients who are not in D_k. This results in an optimal h of 590 (days).

We also repeated the above process fifty times, and at each time we split the data into $\mathcal{K}$ parts in a random fashion. Almost all of the optimal choices of h fall into the interval

[500, 800], and give us practically identical estimates $\hat{\boldsymbol{\beta}}(\cdot)$. In Figure 1, we present the time-varying regression parameter estimates (solid curves) for the above five covariates, and their corresponding 0.90 point-wise intervals (dotted curves) and simultaneous confidence bands (dashed curves) with $h_n = 650$ (days). Here, all the intervals were constructed based on $M = 5000$ realizations of $\{G_i, i = 1, \cdots, 416\}$ over the time interval $[b_1, b_2] = [650, 3000]$. Furthermore, we let $\{\hat{w}(t)\}^{-1}$ be the corresponding estimated standard error of $(\mathbf{a}'(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t))$. The horizontal lines in Figure 1 indicate the maximum partial likelihood estimates for the standard Cox model with time-independent regression parameters. Our results suggest that there is a strong effect of log(prothrombin time) on the patient's hazard function for $t < 1200$ days, but it gradually diminishes over time.

4. PREDICTION OF SUBJECT-SPECIFIC SURVIVAL FUNCTION

In this section, we are interested in predicting the survival function for future patients with a specific set of time-invariant covariates $\mathbf{Z}_0$. To this end, we first consider a generalized Breslow estimate $\hat{\Lambda}(\hat{\boldsymbol{\beta}}(\cdot); t)$ for the underlying cumulative hazard function $\Lambda_0(t)$, where

$$\hat{\Lambda}(\boldsymbol{\beta}(\cdot), t) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(s)}{\sum_{j=1}^{n} Y_j(s) e^{\boldsymbol{\beta}(s)'\mathbf{Z}_j(s)}}. \tag{4.1}$$

Here, we take $h_n = O(n^{-v})$ with $1/4 < v < 1/2$. The process $\mathcal{V}(t) = n^{1/2}(\hat{\Lambda}(\hat{\boldsymbol{\beta}}(\cdot), t) - \Lambda_0(t))$ can be written as

$$n^{1/2}(\hat{\Lambda}(\hat{\boldsymbol{\beta}}(\cdot), t) - \hat{\Lambda}(\boldsymbol{\beta}_0(\cdot), t)) + n^{1/2}(\hat{\Lambda}(\boldsymbol{\beta}_0(\cdot), t) - \Lambda_0(t)). \tag{4.2}$$

Note that the second term of (4.2) is

$$n^{-1/2} \sum_{i=1}^{n} \int_0^t \frac{dM_i(s)}{S^{(0)}(\boldsymbol{\beta}_0(s), s)}, \tag{4.3}$$

which is a martingale in t. It follows from the uniform consistency of $\hat{\boldsymbol{\beta}}(\cdot)$ that the first term of (4.2) approximately equals

$$-n^{-1/2} \sum_{i=1}^{n} \int_0^t \frac{\mathbf{E}(\boldsymbol{\beta}_0(s), s)'}{S^{(0)}(\boldsymbol{\beta}_0(s), s)} (\hat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}_0(s)) dN_i(s). \tag{4.4}$$

From (4.3) and (4.4), one can easily show that $\hat{\Lambda}(\hat{\boldsymbol{\beta}}(\cdot), t)$ is uniformly consistent for $t < \tau$. To show that $\mathcal{V}(t) = n^{1/2}(\hat{\Lambda}(\hat{\boldsymbol{\beta}}(\cdot), t) - \Lambda_0(t))$ converges weakly to a mean-zero Gaussian process, we make use of two basic facts. First, as shown in Appendix B, for $1/4 < v < 1/2$, $\sup_{s\in[h_n,\tau-h_n]}||n^{1/2}(\hat{\boldsymbol{\beta}}(s)-\boldsymbol{\beta}_0(s))-\mathbf{I}^{-1}(\boldsymbol{\beta}_0(s),s)h_n^{-1/2}\mathbf{U}(\boldsymbol{\beta}_0(s),s)|| \to 0$, as $n \to \infty$. Second, since $\sup_{\{s\le h_n\}}||\hat{\boldsymbol{\beta}}(s)-\boldsymbol{\beta}_0(s)|| = O_p(n^{-v})$, the expression (4.4) with the integral restricted to $[0, h_n]$ tends in probability to zero. It follows that (4.4) approximately equals

$$-n^{-1/2}\sum_{i=1}^{n}\Big\{(nh_n)^{-1}\sum_{j=1}^{n}\int_0^{\tau}\int_{h_n}^{t}\frac{\mathbf{E}(\boldsymbol{\beta}_0(s),s)'}{S^{(0)}(\boldsymbol{\beta}_0(s),s)}\mathbf{I}^{-1}(\boldsymbol{\beta}_0(s),s)(\mathbf{Z}_i(x)-\mathbf{E}(\boldsymbol{\beta}_0(s),x)) K(\frac{x-s}{h_n})dN_j(s)\Big\}dM_i(x). \tag{4.5}$$

By replacing the counting processes $N_j(s)$ in (4.5) with their compensators, one can show that (4.5) approximately equals

$$-n^{-1/2}\sum_{i=1}^{n}\int_0^{t}\mathbf{E}(\boldsymbol{\beta}_0(s),s)'\mathbf{V}^{-1}(\boldsymbol{\beta}_0(s),s)(\mathbf{Z}_i(s)-\mathbf{E}(\boldsymbol{\beta}_0(s),s))\frac{dM_i(s)}{S^{(0)}(\boldsymbol{\beta}_0(s),s)}. \tag{4.6}$$

Now, by the Martingale Central Limit Theorem, (4.6)+(4.3) converges weakly to a mean-zero Gaussian process. This proves that $\mathcal{V}(\cdot)$ converges weakly to this Gaussian process over the interval $[h_n, \tau - h_n]$, where $h_n = O(n^{-v}), 1/4 < v < 1/2$. Note that as a process in t, $\hat{\Lambda}(\hat{\boldsymbol{\beta}}(\cdot), t)$ is $n^{1/2}$-consistent.

To approximate the distribution of $\{\mathcal{V}(t)\}$, one may use the same simulation technique as described in Section 3. Specifically, we replace the integrators $M_i(s)$ in (4.3) and (4.6) by $N_i(s)G_i, i = 1, \cdots, n$, and then replace all the unknown parameters by their estimates. This creates two processes in t

$$n^{-1/2}\sum_{i=1}^{n}G_i\int_0^{t}\frac{dN_i(s)}{S^{(0)}(\hat{\boldsymbol{\beta}}(s),s)} \tag{4.7}$$

and

$$-n^{-1/2}\sum_{i=1}^{n}G_i\int_0^{t}\mathbf{E}(\hat{\boldsymbol{\beta}}(s),s)'\mathbf{V}^{-1}(\hat{\boldsymbol{\beta}}(s),s)(\mathbf{Z}_i(s)-\mathbf{E}(\hat{\boldsymbol{\beta}}(s),s))\frac{dN_i(s)}{S^{(0)}(\hat{\boldsymbol{\beta}}(s),s)}. \tag{4.8}$$

Let $\mathcal{V}^*(t) = (4.8) + (4.7)$. Using the argument of Section 3, the distribution of $\mathcal{V}(\cdot)$ can be approximated by the conditional distribution of $\mathcal{V}^*(\cdot)$ given the data.

In general, for a differentiable, known function $g(\cdot)$, by the functional δ-method,

$$\mathcal{V}_g(t) = n^{1/2}\hat{v}(t)(g(\hat{\Lambda}(\hat{\boldsymbol{\beta}}(\cdot), t)) - g(\Lambda_0(t))) \approx \hat{v}(t)\dot{g}(\hat{\Lambda}(\hat{\beta}(\cdot), t))\mathcal{V}(t),$$

where $\hat{v}(t)$ is a possibly data-dependent weight function that converges uniformly to a deterministic bounded function and $\dot{g}$ is the derivative of g. The distribution of $\mathcal{V}_g(t)$ can be approximated by the conditional distribution of $\mathcal{V}_g^*(t) = \hat{v}(t)\dot{g}(\hat{\Lambda}(\hat{\beta}(\cdot), t))\mathcal{V}^*(t)$. In practice, we generate M realizations of $\{G_i, i = 1, \cdots, n\}$ and then obtain M realized $\{\mathcal{V}_g^*(t), t \in [h_n, \tau - h_n]\}$, which can be used collectively to approximate the distribution of any continuous function of $\mathcal{V}_g(\cdot)$.

Now, suppose that we are interested in constructing point-wise confidence intervals for $\Lambda_0(t)$ and the corresponding survival function $S_0(t)$ for a fixed time point t. First, consider the parameter $\log(\Lambda_0(t))$, and the corresponding $\mathcal{V}_g(t)$ with $g(t) = \log(t)$ and $\hat{v}(t) = 1$. The standard error of $\mathcal{V}_g(t)$ can be estimated by $\hat{\sigma}(t)$, the sample standard deviation based on the above M realizations of $\mathcal{V}^*(t)$. Let $z_{\alpha/2}$ be the upper $100\alpha/2$ percentage point of the standard normal distribution. A $(1-\alpha)$ interval for $\log(\Lambda_0(t))$ is simply $\log(\hat{\Lambda}(\hat{\boldsymbol{\beta}}(\cdot), t)) \pm z_{\alpha/2}\hat{\sigma}(t)$. Then, the corresponding intervals for $\Lambda_0(t)$ and $S_0(t)$ are $\hat{\Lambda}(\hat{\boldsymbol{\beta}}(\cdot), t)\exp(\pm z_{\alpha/2}\hat{\sigma}(t))$ and

$$\hat{S}(t)^{\exp(\mp z_{\alpha/2}\hat{\sigma}(t))}, \tag{4.9}$$

respectively, where $\hat{S}(t) = \exp(-\hat{\Lambda}(\hat{\boldsymbol{\beta}}(\cdot), t))$.

To construct a $1-\alpha$ confidence band for $\log(\Lambda_0(\cdot))$ over a properly selected interval $[b_1, b_2] \subset [0, \tau]$, we approximate the distribution of $\sup_{t\in[b_1,b_2]} |\mathcal{V}_g(t)|$ by the conditional distribution of its counterpart based on $\mathcal{V}_g^*(t)$. Letting d_α be the upper 100α percentile of this approximating distribution, we obtain the confidence band $\log(\hat{\Lambda}(\hat{\boldsymbol{\beta}}(\cdot), t)) \pm d_\alpha n^{-1/2}\hat{v}(t)^{-1}$. The corresponding bands for $\Lambda_0(\cdot)$ and $S_0(\cdot)$ are $\{\hat{\Lambda}(\hat{\boldsymbol{\beta}}(\cdot), t)\exp(\pm d_\alpha \hat{v}^{-1}(t)n^{-1/2}), b_1 \le t \le b_2\}$ and

$$\{\hat{S}(t)^{\exp(\mp d_\alpha \hat{v}^{-1}(t)n^{-1/2})}, b_1 \le t \le b_2\}. \tag{4.10}$$

To make inferences about the cumulative hazard and survival functions with a given covariate vector $\mathbf{Z}_0$, one can simply replace $\mathbf{Z}_i(\cdot)$ by $\mathbf{Z}_i(\cdot) - \mathbf{Z}_0, i = 1, \cdots, n$, in the original data set and obtain the intervals and bands for the underlying cumulative hazard and survival functions with this modified data set.

Now, we use the Mayo primary biliary cirrhosis survival data to illustrate the prediction method with the Cox model consisting of five covariates utilized for illustration in Section 3. Suppose that we are interested in predicting the survival function of a patient with 51 years of age, a serum albumin of 3.5 gm/dl, a serum bilirubin of 1.4 mg/lpl, a prothrombin time of 10.6 seconds, and no edema. Again, we let $h_n = 650$ (days) be the smoothing parameter for estimating $\boldsymbol{\beta}_0(t)$. In Figure 2, we present the point estimate with the corresponding 0.90 confidence intervals (4.9) and bands (4.10). Here, the band is obtained by choosing the weight $\{\hat{v}(t)\}^{-1}$ to be the estimated standard error of $\log(\hat{\Lambda}(\hat{\boldsymbol{\beta}}(\cdot), t))$, which corresponds to the so-called equal precision band for $\log(\Lambda_0(\cdot))$. The plot in Figure 2 is quite useful. For instance, for this particular type of patient, a 0.9 point-wise interval for the probability that he/she would survive more than 3000 days is (0.60,0.72) . The corresponding counterpart from the simultaneous interval estimation is (0.57,0.76).

Note that empirically we find that the prediction procedure is quite stable with respect to the choice of the bandwidth parameter h_n. For instance, the predicted survival probability at 3000 days for the aforementioned patient varies from 0.67 to 0.68 for $h_n \in [500, 800]$, and from 0.65 to 0.68 for $h_n \in [400, 1000]$. More specifically, in Figure 3, we present the point estimates of the entire survival function with various h_n. The dark region in the Figure consists of all the survival function estimates with $h_n \in [500, 800]$, the gray zone is associated with $h_n \in [400, 500] \cup [800, 1000]$, and the light gray is with $h \in [250, 400] \cup [1000, 1500]$, which consists of quite under-smoothed and over-smoothed survival function estimates. Empirically we also find that the corresponding confidence intervals and bands for the survival function are practically identical to each other for $h_n \in [400, 1000]$.

5. INFERENCES ABOUT THE TIME INVARIANT REGRESSION PARAMETERS IN THE MIXED CASE

The Cox model (2.1) with time varying regression coefficients is more flexible than the standard proportional hazards model for analyzing survival observations. On the other hand, if we can identify possible time-independent coefficients, a mixed model with both time-varying and time-invariant coefficients is more desirable. The mixed model is less restrictive than the standard proportional hazards model and simpler than Model (2.1). Moreover, it is possible to obtain a $n^{1/2}$-consistent estimator for the time-invariant coefficient in the mixed model. For a general mixed Cox model, we let $\boldsymbol{\beta}_0'(\cdot) = (\boldsymbol{\eta}_0', \boldsymbol{\beta}_{20}'(\cdot))$ in Model (2.1), where $\boldsymbol{\eta}_0$ is a r-vector of time-invariant regression coefficients. Now, let $\hat{\boldsymbol{\beta}}_1(\cdot)$ be the time-varying estimate from $\hat{\boldsymbol{\beta}}(\cdot)$ obtained in Section 2 for the first r components of $\boldsymbol{\beta}_0(\cdot)$. To estimate $\boldsymbol{\eta}_0$, consider a class of estimators

$$\hat{\boldsymbol{\eta}}_{\mathbf{w}} = \int_{h_n}^{\tau - h_n} \mathbf{w}(s)\hat{\boldsymbol{\beta}}_1(s)ds, \tag{5.1}$$

where $\mathbf{w}(\cdot)$ is a weight function, which converges to a deterministic matrix, and $\int_{h_n}^{\tau-h_n} \mathbf{w}(s)ds$ is the $r \times r$ identity matrix. Since for $s \neq t$, as $n \to \infty$, $\hat{\boldsymbol{\beta}}_1(s)$ and $\hat{\boldsymbol{\beta}}_1(t)$ are independent, a natural choice of $\mathbf{w}(t)$ is $\mathbf{w}_{op} = \{\int_{h_n}^{\tau-h_n} \mathbf{J}(u)du\}^{-1}\mathbf{J}(t)$, where $\mathbf{J}(t)$ is the inverse of the upper left $r \times r$ sub-matrix of $\mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}(t), t)$, which is the asymptotic covariance matrix of $n^{1/2}\hat{\boldsymbol{\beta}}_1(t)$. Let the corresponding estimator for $\boldsymbol{\eta}_0$ be denoted by $\hat{\boldsymbol{\eta}}$.

Now for $1/4 < v < 1/2$, it follows from (9.1) that

$$\begin{aligned} n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) &\approx \int_0^\tau \mathbf{w}_{op}(s) n^{1/2}(\hat{\boldsymbol{\beta}}_1(s) - \boldsymbol{\eta}_0)ds \\ &\approx n^{-1/2}\sum_{i=1}^n \int_0^\tau \int_0^\tau h_n^{-1}\mathbf{w}_{op}(s)\mathbf{Q}_r\mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}(s), s)(\mathbf{Z}_i(u) - \mathbf{E}(\boldsymbol{\beta}_0(s), u))K(\frac{u-s}{h_n})dsdM_i(u) \end{aligned} \tag{5.2}$$

where $\mathbf{Q}_r$ is a $r \times p$ matrix whose left $r \times r$ submatrix is the identity matrix, and the right $r \times (p-r)$ submatrix consists of all zeros. Then

$$(5.2) \approx n^{-1/2}\sum_{i=1}^n \int_0^\tau \mathbf{w}_{op}(u)\mathbf{Q}_r\mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}(u), u)(\mathbf{Z}_i(u) - \mathbf{E}(\boldsymbol{\beta}_0(u), u))dM_i(u).$$

By the Martingale Central Limit Theorem $n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ converges weakly to a normal distribution with mean 0, and the limiting covariance matrix can be consistently estimated by $\{\int_{h_n}^{\tau-h_n} \mathbf{J}(s)ds\}^{-1}$. Inferences about $\boldsymbol{\eta}_0$ can then be made via this large sample approximation.

It is interesting to note that the above estimate $\hat{\boldsymbol{\eta}}$ is asymptotically equivalent to a solution for $\boldsymbol{\eta}$ to the following "partial likelihood score equations"

$$\sum_{i=1}^{n} \int_0^{\tau} (\mathbf{Z}_{i1}(t) - \mathbf{E}_1((\boldsymbol{\eta}', \boldsymbol{\beta}_2'(t))', t)) dN_i(t) = 0,$$

and

$$\sum_{i=1}^{n} \int_0^{\tau} (\mathbf{Z}_{i2}(t) - \mathbf{E}_2((\boldsymbol{\eta}', \boldsymbol{\beta}_2'(s))', t)) K(\frac{t-s}{h_n}) dN_i(t) = 0,$$

where $\mathbf{Z}_i'(t) = (\mathbf{Z}_{i1}'(t), \mathbf{Z}_{i2}'(t))$, $\mathbf{E}'(\boldsymbol{\beta}, t) = (\mathbf{E}_1'(\boldsymbol{\beta}, t), \mathbf{E}_2'(\boldsymbol{\beta}, t))$, $\mathbf{Z}_{i1}(t)$ and $\mathbf{E}_1(\boldsymbol{\beta}, t)$ are the $r-$dimensional vectors corresponding to $\boldsymbol{\eta}_0$, and $\mathbf{Z}_{i2}(t)$ and $\mathbf{E}_2(\boldsymbol{\beta}, t)$ are the analogous quantities corresponding to $\boldsymbol{\beta}_{20}(\cdot)$. It is straightforward to show that the asymptotic covariance matrix of $\hat{\boldsymbol{\eta}}$ is identical to that of the semi-parametric efficient estimator for $\boldsymbol{\eta}_0$ discussed in Section 3 of Martinussen, Scheike and Skovgaard (2002).

Now, for the Mayo liver disease example, assume that age and log(albumin) have constant covariate effects on the patient's survival time (this is justified empirically in Section 6), but assume that the other three regression coefficients are potentially time-dependent. With $h_n = 650$ (days), the observed $\hat{\boldsymbol{\eta}} = (0.037, -2.488)'$. The estimated standard errors are 0.009 and 0.799, respectively. The corresponding estimates for these two regression parameters based on the standard PH model are 0.040 and -2.507 with estimated standard errors of 0.008 and 0.653. We find that these estimates are quite stable with respect to the choice of the smoothing parameter value h_n. For example, for $h_n \in [450, 800]$, the observed $\hat{\boldsymbol{\eta}}$ is in a quite small region: $[0.035, 0.039] \times [-2.488, -2.429]$.

6. IDENTIFYING TIME INVARIANT REGRESSION COEFFICIENTS

In the previous section, we present an efficient estimation procedure for the time independent regression parameters in the mixed case. Here, we propose various numerical and graphical methods for identifying covariates whose effects on the survival time can be reasonably assumed to be constant over time. First, one may use the confidence bands presented in Figure 1 based on $\hat{\boldsymbol{\beta}}(\cdot)$ to examine if individual regression coefficients are time varying, for example, by checking whether a horizontal line is enclosed by the confidence band in each plot. However, this procedure may not be sensitive enough to identify the time varying covariate effects. Next, we may use the estimated cumulative function $\hat{\mathbf{B}}(t) = \int_0^t \hat{\boldsymbol{\beta}}(s)ds$ to examine the adequacy of the PH assumption for each covariate. Specifically, let $\mathbf{B}_0(t) = \int_0^t \boldsymbol{\beta}_0(s)ds$, it follows from an argument similar to that in Section 4 that, for $1/4 < v < 1/2$, the process $n^{1/2}(\hat{\mathbf{B}}(t) - \mathbf{B}_0(t))$ converges weakly to a mean-zero, independent increments, p-dimensional Gaussian process where variance function can be consistently estimated by $\int_0^t \mathbf{I}(\hat{\boldsymbol{\beta}}(s), s)^{-1}ds$. Note that this approximation is independent of the convergence rate of $\hat{\boldsymbol{\beta}}(t)$. Furthermore, $\hat{\mathbf{B}}(t)$ has the same limiting distribution as that of the sieve estimator proposed by Murphy and Sen (1981), which is semi-parametric efficient for $\mathbf{B}_0(t)$ (Martinussen, Scheike and Skovgaard, 2002, Remark 1). To check the PH assumption for a specific covariate, one can then construct a confidence band based on $\hat{\mathbf{B}}(\cdot)$ for the corresponding regression parameter over t. If we cannot find a straight line which goes through the origin $(0, 0)$ and is contained in the band, the PH assumption for this covariate is violated.

We now explore a more objective and efficient way to check the PH assumption based on $\hat{\mathbf{B}}(t)$. Assume that the first component, say, $\boldsymbol{\eta}_0$, of $\boldsymbol{\beta}_0(\cdot)$ is constant. Then, with $r = 1$ in Section 5, the estimator $\hat{\boldsymbol{\eta}}$ is an efficient estimator for $\boldsymbol{\eta}_0$. Consider the process $\Gamma(t) = n^{1/2} \int_{h_n}^t (\hat{\boldsymbol{\beta}}_1(s) - \hat{\boldsymbol{\eta}})ds$, where $t \in [h_n, \tau - h_n]$ and $\hat{\boldsymbol{\beta}}_1(\cdot)$ is defined in Section 5 for estimating the first component of $\boldsymbol{\beta}_0(\cdot)$. If the first covariate has a constant effect over time, $\Gamma(t)$ converges weakly to a mean zero Gaussian process. Furthermore, operationally, the distribution of the limiting process can be approximated by the distribution of the process

$$\frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \int_0^\tau (I(s \le t) - \mathbf{w}_{op}(s)) \mathbf{Q}_1 \mathbf{I}^{-1}((\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}_2'(s))', s)(\mathbf{Z}_i - \mathbf{E}((\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}_2'(s))', s)) K(\frac{t-s}{h_n}) dN_i(s) G_i, \tag{6.1}$$

where $\hat{\boldsymbol{\beta}}_2(\cdot)$ is the time-varying estimate from $\hat{\boldsymbol{\beta}}(\cdot)$ in Section 2 for $\boldsymbol{\beta}_{20}(\cdot)$. The observed value of $\{\Gamma(t)/\hat{\sigma}_\gamma(t), t > h_n\}$, where $\hat{\sigma}_\gamma(t)$ is the estimated standard error for $\Gamma(t)$, can then be compared graphically and numerically with a number of realizations from (6.1) by generating the random sample $\{G_i, i = 1, \cdots, n\}$ repeatedly to examine if the first coefficient is time varying. A general discussion on this type of model checking and selection techniques is given by Lin, Wei and Ying (2002).

We now apply the foregoing technique to the Mayo Clinic data to identify which of those five covariates: age, log(albumin), log(bilirubin), edema and log(prothrombin time), are likely to have constant effects on the survival time. In Figure 4, we present five plots corresponding to these five covariates. Each plot in the Figure consists of a curve from the observed $\Gamma(t)/\hat{\sigma}_\gamma(t)$ for a specific covariate along with ten realized counterparts generated from (6.1) with $h_n = 650$ (days). Compared with those background curves, visually the observed curves for log(prothrombin time) and edema seem rather atypical, and the observed curve for log(bilirubin) also appears to have an unusual pattern. To quantify how "unusual" each observed curve is, we randomly generated 1000 corresponding curves from (6.1) and tallied how many of these realized curves whose maxima are greater than the maximum of the observed curve or whose minima are less than its corresponding observed counterpart. This creates the so-called empirical p-value for testing the PH assumption for a specific covariate. The empirical p-values for age, log(albumin), log(bilirubin), edema and log(prothrombin time) are 0.652, 0.677, 0.038, 0.090, and 0.031, respectively, indicating that log(bilirubin), log(prothrombin time) and edema have non-constant covariate effects on the survival time. On the other hand, the other two covariates seem to have constant effects. This observation is consistent with that by Lin et al. (1993) using a quite different model checking technique for the same data set. However, it is important to note that the distribution theory for the procedure proposed by Lin et al. was derived under the assumption that the entire set of covariates satisfies the PH assumption. Therefore, a small marginal p-value for testing the PH assumption for a specific covariate by Lin et al. may be due to the violation of the PH assumption from other covariates.

On the other hand, our proposal is designed for checking the PH assumption of individual covariates. Furthermore, our procedure is quite stable with respect to the choice of the smoothing parameter h_n. For instance, when $h_n \in [500, 800]$, the p-value of the proposed test corresponding to log(prothrombin time) is between 0.01 and 0.04.

7. REMARKS

In this paper, we develop various relatively simple inference procedures with theoretical justification for the Cox model with time-varying regression coefficients or a mixture of time-varying and time-independent parameters. As in general nonparametric function estimation problems, the smoothing parameter plays a crucial role in practice. In this article, we suggest choosing the smoothing parameter by cross validation. Empirically we find that this approach works well in the sense that the bandwidth selection can be quite flexible, especially for predicting the survival or cumulative hazard function of future patients, making inferences about the constant covariate effects in the mixed case, and checking the proportional hazards assumption for individual covariates.

We have restricted the confidence bands for the various functions considered in this paper to the time interval $[h_n, \tau - h_n]$. The method of Gasser and Muller (1979) probably can be used to extend the bands to the boundary regions. Gilbert et al. (2002) explored such extension for the simple two-sample problem with censored survival data.

8. APPENDIX A
UNIFORM CONSISTENCY OF $\hat{\boldsymbol{\beta}}(\cdot)$

Let $\mathbf{u}(\boldsymbol{\beta}, t)$ be the limit of $(nh_n)^{-1/2}\mathbf{U}(\boldsymbol{\beta}, t)$, namely

$$\mathbf{u}(\boldsymbol{\beta}, t) = (s^{(1)}(\boldsymbol{\beta}_0(t), t) - \mathbf{e}(\boldsymbol{\beta}, t)s^{(0)}(\boldsymbol{\beta}_0(t), t))\lambda_0(t),$$

where $s^{(r)}(\boldsymbol{\beta}, t)$ and $\mathbf{e}(\boldsymbol{\beta}, t)$ are the limits of $S^{(r)}(\boldsymbol{\beta}, t)$ and $\mathbf{E}(\boldsymbol{\beta}, t)$, respectively, $r = 0, 1, 2$. Also, let $\mathbf{g}(\boldsymbol{\beta}, t) = -\partial \mathbf{u}(\boldsymbol{\beta}, t)/\partial \boldsymbol{\beta}$ and let $\mathcal{B}$ be a compact set of R^p that includes a neighborhood of $\boldsymbol{\beta}_0(t)$ for $t \in [0, \tau]$. Assume that

(A.1) $\boldsymbol{\beta}_0(t)$ has continuous second derivatives for $t \in [0, \tau]$;

(A.2) $s^{(r)}(\boldsymbol{\beta}, t)$ is uniformly continuous with respect to $(\boldsymbol{\beta}', t)' \in \mathcal{B}\times [0, \tau]$, for $r = 0, 1, 2, 3$;

(A.3) $\mathbf{g}(\boldsymbol{\beta}_0(t), t)$ is nonsingular for $t \in [0, \tau]$.

To show the uniform consistency of $\hat{\boldsymbol{\beta}}(t)$, first it follows from the definition of $\mathbf{u}(\boldsymbol{\beta}, t)$ that $\mathbf{u}(\boldsymbol{\beta}_0(t), t) = 0$. Secondly, it follows from Condition (A.3) and the fact that $\mathbf{g}(\boldsymbol{\beta}, t)$ is semi-negative definite for any $\boldsymbol{\beta}$, that $\boldsymbol{\beta}_0(t)$ is the unique root to the equation $\mathbf{u}(\boldsymbol{\beta}, t) = 0$. Lastly, one needs to show that $(nh_n)^{-1/2}\mathbf{U}(\boldsymbol{\beta}, t) \to \mathbf{u}(\boldsymbol{\beta}, t)$, uniformly for $\boldsymbol{\beta} \in \mathcal{B}$ and $t \in [h_n, \tau - h_n]$. Given this uniform convergence, the fact that the equation $\mathbf{u}(\boldsymbol{\beta}, t) = 0$ has a unique root at $\boldsymbol{\beta}_0(t)$, and the fact that the derivative of function $\hat{\boldsymbol{\beta}}(t)$ is uniformly bounded, it follows from the Arzela-Ascoli theorem and a subsequence argument that $\hat{\boldsymbol{\beta}}(t)$ converges uniformly to $\boldsymbol{\beta}_0(t)$ over $t \in [0, \tau]$. Now, to prove the uniform convergence for $(nh_n)^{-1/2}\mathbf{U}(\boldsymbol{\beta}, t)$, let

$$(nh_n)^{-1/2}\mathbf{U}(\boldsymbol{\beta}, t) = (nh_n)^{-1/2}\mathbf{U}_1(\boldsymbol{\beta}, t) + (nh_n)^{-1/2}\mathbf{U}_2(\boldsymbol{\beta}, t), \tag{8.1}$$

where

$$\mathbf{U}_1(\boldsymbol{\beta}, t) = (nh_n)^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} (\mathbf{Z}_i(s) - \mathbf{E}(\boldsymbol{\beta}, s)) K\left(\frac{s-t}{h_n}\right) dM_i(s),$$

$$\mathbf{U}_2(\boldsymbol{\beta}, t) = n^{1/2} {h_n}^{-1/2} \int_0^{\tau} K\left(\frac{s-t}{h_n}\right) (\mathbf{E}(\boldsymbol{\beta}_0(s), s) - \mathbf{E}(\boldsymbol{\beta}, s)) S^{(0)}(\boldsymbol{\beta}_0(s), s)\lambda_0(s) ds, \tag{8.2}$$

and $M_i(s) = N_i(s) - \int_0^s Y_i(u)\lambda_i(u)du$. Using the strong approximation argument similar to that given in Yandell (1983), it is not difficult to show that for any $\epsilon > 0$, the first term of (8.1) is bounded by $(nh_n)^{-1/2}O_p(n^{-\epsilon})$, which is $o_p(1)$, uniformly in $t \in [0, \tau]$, for any fixed $\boldsymbol{\beta}$ and $v < 1$. For the second term of (8.1), one can replace $(\mathbf{E}(\boldsymbol{\beta}_0(s), s) - \mathbf{E}(\boldsymbol{\beta}, s))S^{(0)}(\boldsymbol{\beta}_0(s), s)$ by its limit $(\mathbf{e}(\boldsymbol{\beta}_0(s), s) - (\mathbf{e}(\boldsymbol{\beta}, s))s^{(0)}(\boldsymbol{\beta}_0(s), s)$ based on the fact that

$$\sup_{t \in [0,\tau], \boldsymbol{\beta} \in \mathcal{B}} \|S^{(r)}(\boldsymbol{\beta}, t) - s^{(r)}(\boldsymbol{\beta}, t)\| = O_p(n^{-\frac{1}{2}}). \tag{8.3}$$

Note that (8.3) can be justified using the Central Limit Theorem for Banach Space (Ledoux and Talagrand, 1991). It follows that the second term on the right hand side of

(8.1) is asymptotically equivalent to

$$h_n^{-1}\int_0^\tau K\left(\frac{s-t}{h_n}\right)(\mathbf{e}(\boldsymbol{\beta}_0(s),s)-\mathbf{e}(\boldsymbol{\beta},s))s^{(0)}(\boldsymbol{\beta}_0(s),s)\lambda_0(s)ds. \tag{8.4}$$

Note that if $\boldsymbol{\omega}(\cdot)$ is a generic function whose second derivatives are bounded by a constant c, then by Taylor series expansion, it can be shown (Eubank, 1988, p. 128) that

$$||\int h_n^{-1}K\left(\frac{s-t}{h_n}\right)(\boldsymbol{\omega}(s)-\boldsymbol{\omega}(t))ds|| \le \frac{1}{2}ch_n^2, \tag{8.5}$$

where $t\in[h_n,\tau-h_n]$. It follows from (8.5) that (8.4) can be uniformly approximated by $\mathbf{u}(\boldsymbol{\beta},t)$. This implies that for any fixed $\boldsymbol{\beta}$, $\sup_{t\in[h_n,\tau-h_n]}\|(nh_n)^{-1/2}\mathbf{U}(\boldsymbol{\beta},t)-\mathbf{u}(\boldsymbol{\beta},t)\| = o_p(1)$. Since $\mathbf{U}(\boldsymbol{\beta},t)$ is monotone in $\boldsymbol{\beta}$ and $\mathbf{u}(\boldsymbol{\beta},t)$ is continuous in $\boldsymbol{\beta}$ and t, it follows from Appendix II of Andersen and Gill (1982) that the foregoing convergence also holds uniformly in $\boldsymbol{\beta}\in\mathcal{B}$.

9. APPENDIX B

APPROXIMATIONS TO THE DISTRIBUTION OF STANDARDIZED $\sup_{t\in[b_1,b_2]}\{\hat{w}(t)|\mathbf{a}'(\hat{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}_0(t))|\}$

First, we derive an analytic approximation to the distribution of a standardized version of the sup-statistic $\mathcal{S}$ in (3.3). For a generic function $A(t)$, $\sup_{t\in[h_n,\tau-h_n]}|A(t)|$ is denoted by $\|A(t)\|$.

Proposition 1. The random matrix $\mathbf{I}(\boldsymbol{\beta}^*(t),t)$ converges to a deterministic matrix uniformly in $t\in[h_n,\tau-h_n]$, where $\boldsymbol{\beta}^*(t)$ is between $\boldsymbol{\beta}_0(t)$ and $\hat{\boldsymbol{\beta}}(t)$. That is, if $1/5<v<1$, then for any $\epsilon>0$,

$$\|\mathbf{I}(\boldsymbol{\beta}^*(t),t)-\mathbf{g}(\boldsymbol{\beta}_0(t),t)\| = O_p(||\hat{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}_0(t)\| + n^{\frac{v-1+\epsilon}{2}}),$$

uniformly for $t\in[h_n,\tau-h_n]$, where $\mathbf{g}(\boldsymbol{\beta}_0(t),t)$ is given in Appendix A.

Proof. Let $\mathbf{v}(\boldsymbol{\beta}, t)$ be the limit of $\mathbf{V}(\boldsymbol{\beta}, t)$. Then, $\|\mathbf{I}(\boldsymbol{\beta}^*(t), t) - \mathbf{g}(\boldsymbol{\beta}_0(t), t)\|$ is bounded by

$$
\begin{aligned}
&\|(nh_n)^{-1}\sum_{i=1}^{n}\int_0^{\tau}(\mathbf{V}(\boldsymbol{\beta}^*(t), s) - \mathbf{v}(\boldsymbol{\beta}^*(t), s))K\left(\frac{s-t}{h_n}\right)dN_i(s)\| \\
&+\|(nh_n)^{-1}\sum_{i=1}^{n}\int_0^{\tau}(\mathbf{v}(\boldsymbol{\beta}^*(t), s) - \mathbf{v}(\boldsymbol{\beta}_0(t), s))K\left(\frac{s-t}{h_n}\right)dN_i(s)\| \\
&+\|(nh_n)^{-1}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{v}(\boldsymbol{\beta}_0(t), s)K\left(\frac{s-t}{h_n}\right)dM_i(s)\| \\
&+\|h_n^{-1}\int_0^{\tau}\mathbf{v}(\boldsymbol{\beta}_0(t), s)K\left(\frac{s-t}{h_n}\right)\{S^{(0)}(\boldsymbol{\beta}_0(t), s) - s^{(0)}(\boldsymbol{\beta}_0(t), s)\}\lambda_0(s)ds\| \\
&+\|h_n^{-1}\int_0^{\tau}\mathbf{v}(\boldsymbol{\beta}_0(t), s)s^{(0)}(\boldsymbol{\beta}_0(t), s)\lambda_0(s)K\left(\frac{s-t}{h_n}\right)ds - \mathbf{v}(\boldsymbol{\beta}_0(t), t)s^{(0)}(\boldsymbol{\beta}_0(t), t)\lambda_0(t)\| \\
&=I_1 + I_2 + I_3 + I_4 + I_5.
\end{aligned}
$$

By (8.3) and Theorem 3.1 of Bickel and Rosenblatt (1973), one can show that

$$
\begin{aligned}
I_1 \le O_p(n^{-\frac{1}{2}})(\|h_n^{-1}\int_0^{\tau}K\left(\frac{s-t}{h_n}\right)d\left[n^{-1}\sum_{i=1}^{n}(N_i(s) - E(N_i(s))\right]\| \\
+\|h_n^{-1}\int_0^{\tau}K\left(\frac{s-t}{h_n}\right)dE(N_1(s))\|) = O_p(n^{-\frac{1}{2}}).
\end{aligned}
$$

Since $\mathbf{v}(\beta, \cdot)$ is differentiable,

$$
I_2 \le \|\int_0^{\tau}h_n^{-1}K\left(\frac{s-t}{h_n}\right)d\left[n^{-1}\sum_{i=1}^{n}N_i(s)\right]\|O_p(\|\boldsymbol{\beta}^*(t) - \boldsymbol{\beta}_0(t)\|) = O_p(\|\boldsymbol{\beta}^*(t) - \boldsymbol{\beta}_0(t)\|).
$$

Similar to the way we handled the approximation to the first term of (8.1), $I_3 = O_p(n^{\frac{v-1+\epsilon}{2}})$ for any $\epsilon > 0$. Again, by (8.3), $I_4 = O_p(n^{-1/2})$. Lastly, by (8.5), $I_5 = O_p(n^{-2v})$.

□

Proposition 2. The approximation of (3.2) is uniform in $t \in [h_n, \tau - h_n]$, that is, if $1/5 < v < 1$,

$$
\mathbf{U}(\boldsymbol{\beta}_0(t), t) = (nh_n)^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^{\tau}(\mathbf{Z}_i(s) - \mathbf{E}(\boldsymbol{\beta}_0(t), s))K\left(\frac{s-t}{h_n}\right)dM_i(s) + O_p(n^{\frac{1-5v}{2}}),
$$

where the $O_p(\cdot)$ is free of t.

Proof. It is equivalent to show that $\|(nh_n)^{-\frac{1}{2}}\mathbf{U}_2(\boldsymbol{\beta}_0(t),t)\| = O_p(n^{-2v})$, where $\mathbf{U}_2$ is defined in (8.2). To this end, note that

$$\|\mathbf{E}(\boldsymbol{\beta}_0(s),s) - \mathbf{E}(\boldsymbol{\beta}_0(t),s))S^{(0)}(\boldsymbol{\beta}_0(s),s) - \mathbf{v}(\boldsymbol{\beta}_0(s),s)(\boldsymbol{\beta}_0(s)-\boldsymbol{\beta}_0(t))s^{(0)}(\boldsymbol{\beta}_0(s),s)\|$$
$$\leq \|(\mathbf{V}(\boldsymbol{\beta}^{**},s) - \mathbf{v}(\boldsymbol{\beta}^{**},s))S^{(0)}(\boldsymbol{\beta}_0(s),s)(\boldsymbol{\beta}_0(s)-\boldsymbol{\beta}_0(t))\|$$
$$+ \|\mathbf{v}(\boldsymbol{\beta}^{**},s)(S^{(0)}(\boldsymbol{\beta}_0(s),s) - s^{(0)}(\boldsymbol{\beta}_0(s),s))(\boldsymbol{\beta}_0(s)-\boldsymbol{\beta}_0(t))\|$$
$$+ \|s^{(0)}(\boldsymbol{\beta}_0(s),s)(\mathbf{v}(\boldsymbol{\beta}^{**},s) - \mathbf{v}(\boldsymbol{\beta}_0(s),s))(\boldsymbol{\beta}_0(s)-\boldsymbol{\beta}_0(t))\|$$

where $\boldsymbol{\beta}^{**}$ is between $\boldsymbol{\beta}_0(s)$ and $\boldsymbol{\beta}_0(t)$. By (8.3) and Condition (A.2), the above quantities are uniformly bounded by $O_p(n^{-\frac{1}{2}}|s-t| + |s-t|^2)$. It follows that

$$\|(nh_n)^{-\frac{1}{2}}\mathbf{U}_2(\boldsymbol{\beta}_0(t),t) - \int_0^\tau h_n^{-1}K\left(\frac{s-t}{h_n}\right)\mathbf{v}(\boldsymbol{\beta}_0(s),s)(\boldsymbol{\beta}_0(s)-\boldsymbol{\beta}_0(t))s^{(0)}(\boldsymbol{\beta}_0(s),s)\lambda_0(s)ds\|$$
$$= O_p(n^{-2v}).$$

Again, by (8.5), the second integral inside the above $\|\cdot\|$ is of $O_p(n^{-2v})$. This implies that $\|(nh_n)^{-\frac{1}{2}}\mathbf{U}_2(\boldsymbol{\beta}_0(t),t)\| = O_p(n^{-2v})$.

□

It follows from Propositions 1 and 2 that for $1/5 < v < 1, t \in [h_n, \tau - h_n]$, and any $\epsilon > 0$,

$$(nh_n)^{1/2}(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)) = \mathbf{g}^{-1}(\boldsymbol{\beta}_0(t),t)(nh_n)^{-\frac{1}{2}}\sum_{i=1}^n \int_0^\tau (\mathbf{Z}_i(s) - \mathbf{E}(\boldsymbol{\beta}_0(t),s))K\left(\frac{s-t}{h_n}\right)dM_i(s)$$
$$+ O_p(\|\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\|) + O_p(n^{\frac{1-5v}{2}}) + O_p(n^{\frac{v-1+\epsilon}{2}}). \quad (9.1)$$

Note that the implicit bounding constants associated with the O_p terms above are all uniform in t.

Now, let $\mathbf{d}(t) = (d_1(t), \cdots, d_p(t))'$, be a given deterministic function of t. Also, let

$$\mathcal{Q} = \|\{(\int_{-1}^1 K^2(s)ds)\mathbf{d}(t)'\mathbf{g}(\boldsymbol{\beta}_0(t),t)\mathbf{d}(t)\}^{-\frac{1}{2}}\mathbf{d}(t)'\mathbf{U}(\boldsymbol{\beta}_0(t),t)\|.$$

Here and in the sequel, for a generic function $A(t)$, $\sup_{t\in[b_1,b_2]}|A(t)|$ is denoted by $\|A(t)\|$, where $[b_1, b_2] \subseteq [h_n, \tau - h_n]$.

Proposition 3. For $1/5 < v < 1$,

$$\mathrm{pr}(r_n(\mathcal{Q} - d_n) < x) \to \exp(-2e^{-x}), \tag{9.2}$$

where $r_n = (2\log((b_2 - b_1)/h_n))^{1/2}$, $d_n = r_n + (2r_n)^{-1}\log(\int_{-1}^{1} \dot{K}(s)^2 ds / \int_{-1}^{1} K^2(s)ds)/4\pi^2)$, and $\dot{K}(s)$ is the derivative of $K(s)$.

Proof. Let $U_k(t)$ be the kth component of $\mathbf{U}(\boldsymbol{\beta}_0(t), t)$, then It is easy to show that for $v > 1/5$,

$$U_k(t) = h_n^{-\frac{1}{2}} \int_z \int_s (z - E_k(\boldsymbol{\beta}_0(t), s)) K\left(\frac{s-t}{h_n}\right) d\left[n^{\frac{1}{2}}(\hat{F}_k(z,s) - F_k(z,s))\right] + o_p(n^{-\epsilon}),$$

where $\epsilon > 0$, $Z_k(\cdot)$ and $E_k(\boldsymbol{\beta}_0(t), s))$ are the kth components of $\mathbf{Z}(\cdot)$ and $\mathbf{E}(\boldsymbol{\beta}_0(t), s)$, and $\hat{F}_k(z,s)$ and $F_k(z,s)$ are the empirical and cumulative distribution functions for $(Z_k(X^*), X^*)$, with $X^* = X$ if $\Delta = 1$; ∞, otherwise.

Next, it can be shown via (8.3) and integration by parts that $\|\mathbf{d}(t)'\mathbf{U}(\boldsymbol{\beta}_0(t), t) - D_0(t)\| = o_p(n^{-\epsilon})$, where

$$D_0(t) = \sum_{k=1}^{p} h_n^{-\frac{1}{2}} \int_z \int_s d_k(s)(z - e_k(t,s)) K(\frac{s-t}{h_n}) d\left[n^{\frac{1}{2}}(\hat{F}_k(z,s) - F_k(z,s))\right],$$

where $e_k(t,s)$ is the limit of $E_k(\boldsymbol{\beta}_0(t), s))$. Note that $D_0(t)$ can be rewritten as

$$h_n^{-\frac{1}{2}} \int_z \int_s (z - \bar{e}(t,s)) K\left(\frac{s-t}{h_n}\right) d\left[n^{\frac{1}{2}}(\bar{F}_n(z,s) - \bar{F}(z,s))\right], \tag{9.3}$$

where $\bar{e}(t,s) = \sum_{k=1}^{p} d_k(s) e_k(\boldsymbol{\beta}_0(t), s))$, $\bar{F}_n(z,s) = n^{-1}\sum_{i=1}^{n} I(\sum_{k=1}^{p} d_k(X_i^*) Z_{ik}(X_i^*) \le z, X_i^* \le s)$ with the expected value $\bar{F}(z,s)$, and $Z_{ik}(\cdot)$ is the kth component of $\mathbf{Z}_i(\cdot)$. Then, there exists a sequence of independent two dimensional Brownian bridges $\{B_n(z,t)\}$ (Tusnady, 1977), such that with probability one,

$$\sup_{z,s} |n^{\frac{1}{2}}(\bar{F}_n(z,s) - \bar{F}(z,s)) - B_n(R(z,s))| = O(n^{-\frac{1}{2}}\log(n)^2),$$

where $R(z,s)$ is a mapping from $R^2 \to R^2$, which transforms the bivariate random variable $(\sum_{k=1}^{p} d_k(X_1^*) Z_{1k}(X_1^*), X_1^*)$ to $U(0,1) \times U(0,1)$ (Rosenblatt, 1976). Furthermore, there exists a sequence of two dimensional Wiener processes $\{W_n\}$ on the same probability space such that $B_n(z,s) = W_n(z,s) - zsW_n(1,1)$.

Using integration by parts, the integrator of (9.3) can be replaced by $W_n(R(z,s))$, and one can show that $\|D_0(t) - D_1(t)\| = o_p(n^{-\epsilon})$ for $v < 1$, where

$$D_1(t) = h_n^{-\frac{1}{2}} \int_z \int_s (z - \bar{e}(s,s)) K(\frac{s-t}{h_n}) dW_n(R(z,t)).$$

Let

$$D_2(t) = h_n^{-\frac{1}{2}} \int_0^\tau p(s)^{1/2} K\left(\frac{s-t}{h_n}\right) dW(s),$$

where $W(s)$ is the standard univariate Wiener process and $p(s) = \int_z (z - \bar{e}(s,s))^2 |\frac{\partial T(z,s)}{\partial (z,s)}| dz$. The processes $D_1(\cdot)$ and $D_2(\cdot)$ have the same covariance function. It follows that these two processes have the same distribution.

Furthermore, it follows from the same arguments in Yandell (1983) that $D_2(t)$ can be approximated by $D_3(t) = p(t)^{1/2} h_n^{-\frac{1}{2}} \int_0^\tau K\left(\frac{s-t}{h_n}\right) dW(s)$. By comparing the variances of $D_3(t)$ and $\mathbf{d}(t)'\mathbf{U}(\boldsymbol{\beta}_0(t), t)$, we observe that $p(t) = \mathbf{d}(t)'\mathbf{g}(\boldsymbol{\beta}_0(t), t)\mathbf{d}(t) \int_{-1}^1 K(s)^2 ds$. One can then apply Theorem (3.1) of Bickel and Rosenblatt (1973) to $\|p^{-1/2}(t) D_3(t)\|$ and obtain (9.2). Moreover, the rate of convergence for $\|p^{-1/2}(t) D_3(t)\|$ to the standard extreme value distribution is of order $O((\log n)^{-1})$ (Hall, 1979).

□

Now, if $v < 1$, it follows from Proposition 1 that for some $\epsilon > 0$,

$$\|(nh_n)^{1/2}(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)) - \mathbf{g}^{-1}(\boldsymbol{\beta}_0(t), t)\mathbf{U}(\boldsymbol{\beta}_0(t), t)\| = o_p(n^{-\epsilon}).$$

Therefore,

$$\|\hat{w}(t)\mathbf{a}'\{(nh_n)^{1/2}(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t))\}\| - \|\hat{w}(t)\mathbf{a}'\mathbf{g}^{-1}(\boldsymbol{\beta}_0(t), t)\mathbf{U}(\boldsymbol{\beta}_0(t), t)\| = o_p(n^{-\epsilon}).$$

This implies that for $1/5 < v < 1$,

$$\sup_x \left\|\mathrm{pr}\big(r_n(\|\frac{\mathbf{a}'\hat{w}(t)\{(nh_n)^{1/2}(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t))\}}{\sigma_n(t)}\| - d_n) < x\big) - \exp(-2e^{-x})\right\| = O((\log n)^{-1}), \tag{9.4}$$

where $\sigma_n^2(t) = (\int_{-1}^1 K^2(s) ds)\hat{w}^2(t)\mathbf{a}'\mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}(t), t))\mathbf{a}$.

Next, to justify the approximation to $\mathcal{S}$ in (3.3) based on the simulation technique (3.5), let us consider

$$\tilde{\mathcal{Q}} = \|\{(\int_{-1}^{1} K^2(s)ds)\mathbf{d}(t)'\mathbf{I}(\hat{\boldsymbol{\beta}}(t), t)\mathbf{d}(t)\}^{-\frac{1}{2}}\mathbf{d}(t)'\tilde{\mathbf{U}}(t)\|.$$

Define $\mathbf{U}^*(t)$ as in (3.5) but with $\hat{\boldsymbol{\beta}}(s)$ in the integrand in place of $\hat{\boldsymbol{\beta}}(t)$. With some elementary algebraic manipulations, one can show that conditional on the data, the distribution of the process $\mathbf{d}(t)'\mathbf{U}^*(t)$ is identical to that of process $h_n^{-1/2}\int_0^\tau K(\frac{s-t}{h_n})dW(H_n(s))$, where $H_n(s) = \int_0^\tau n^{-1}\sum_{i=1}^n[\mathbf{d}(s)'(\mathbf{Z}_i(s) - \mathbf{E}(\hat{\boldsymbol{\beta}}(s), s))]^2 dN_i(s)$. Furthermore, it can be shown that $\|\mathbf{U}^*(t) - \tilde{\mathbf{U}}(t)\| = o_p(n^{-\epsilon})$ and that $\|H_n(s) - H(s)\| = o_p(n^{-\epsilon})$, where $\epsilon > 0$, $H(s) = \mathrm{E}\left[\int_0^\tau[\mathbf{d}(s)'(\mathbf{Z}_1(s) - \mathbf{E}(\boldsymbol{\beta}_0(s), s))]^2 dN_1(s)\right]$. Using integration by parts, the process $h_n^{-1/2}\int_0^\tau K(\frac{s-t}{h_n})dW(H_n(s))$ can be uniformly approximated by the process $h_n^{-1/2}\int_0^\tau K(\frac{s-t}{h_n})dW(H(s)) = D_2(t)$, where the equality holds because of $dH(s)/ds = p(s)$. Therefore, conditional on the data $\{(X_i, \Delta_i, \mathbf{Z}_i(\cdot)), i = 1, \cdots, n\}$, $\mathbf{d}(t)'\tilde{\mathbf{U}}(t) = p(t)^{1/2}h_n^{-\frac{1}{2}}$ $\int_0^\tau K(\frac{s-t}{h_n})dW(s) + o_p(n^{-\epsilon})$, and if $1/5 < v < 1$,

$$\sup_x \left\|\mathrm{pr}(r_n(\tilde{\mathcal{Q}} - d_n) < x|\{(X_i, \Delta_i, \mathbf{Z}_i(\cdot))\}) - \mathrm{pr}(r_n(\mathcal{Q} - d_n) < x)\right\| = o_p(n^{-\epsilon}). \qquad (9.5)$$

Figure 1. Point and interval estimates for time-varying coefficients for Mayo primary biliary cirrhosis data

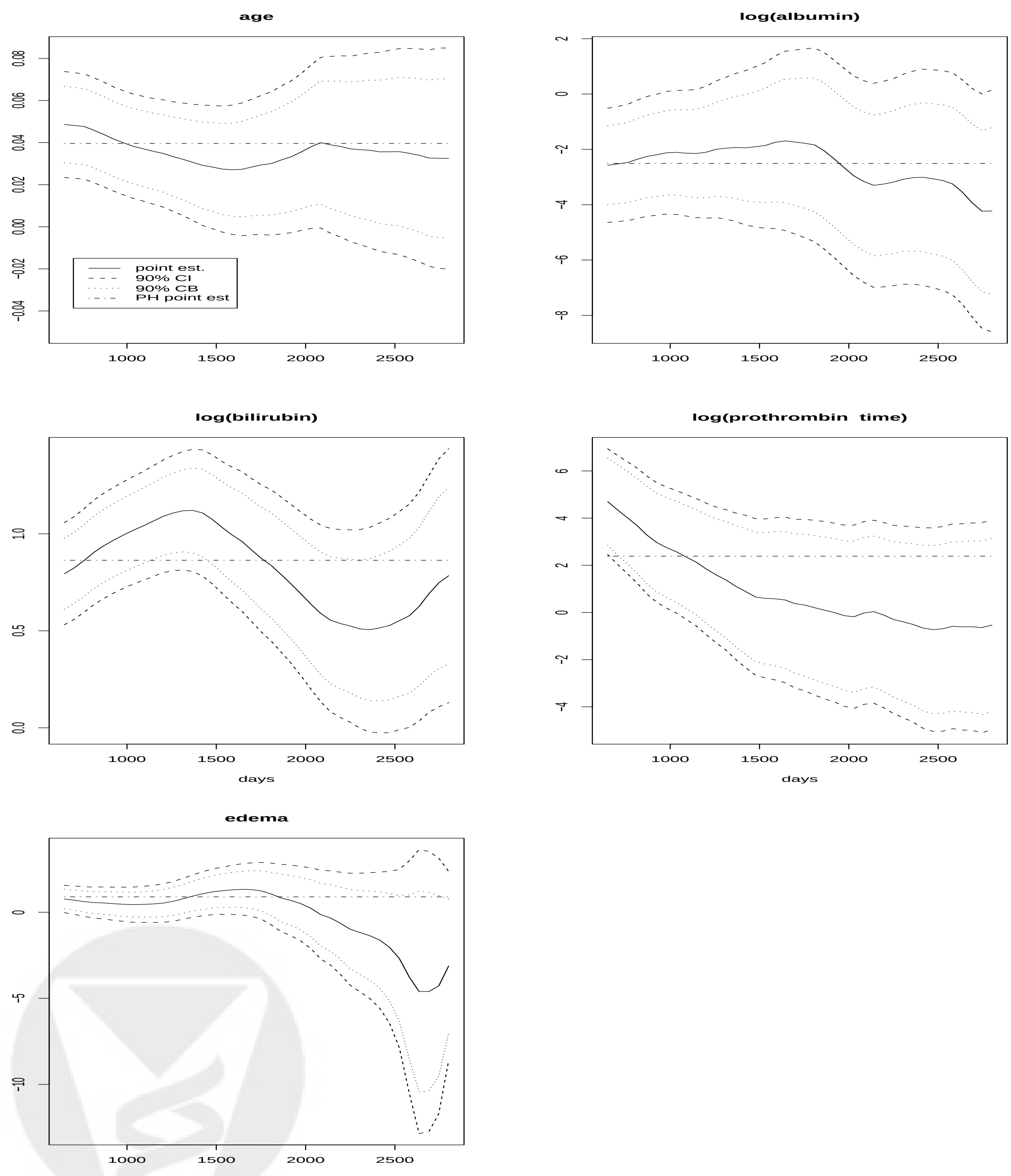


Figure 2. Predicting survival function for a patient with 51 years of age, 3.5 gm/dl albumin, 1.4 mg/pl bilirubin, 10.6 seconds of prothrombin time and no edema

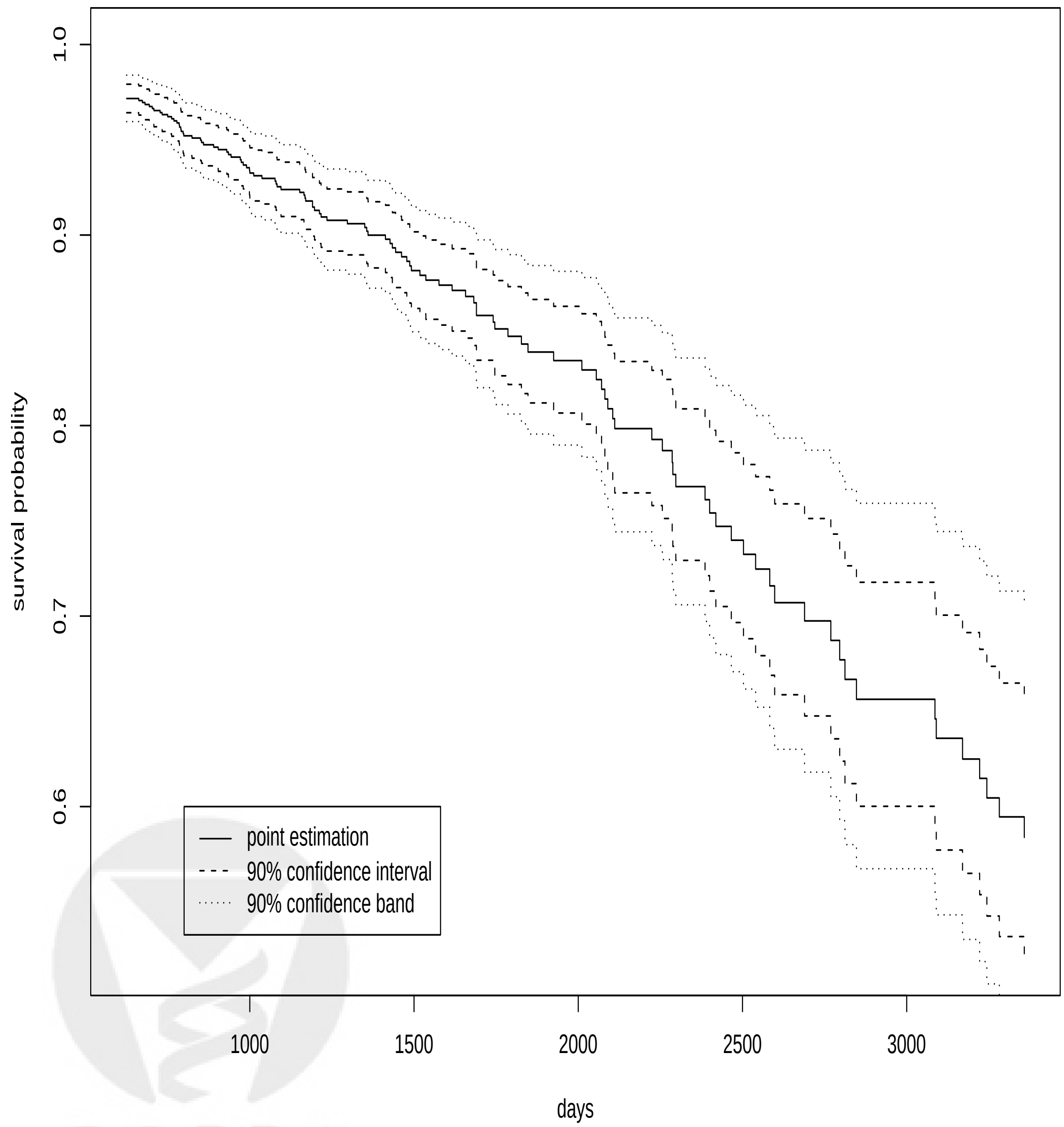


Figure 3. Point estimates of the survival function for the patient in Figure 2 with various smoothing parameter values

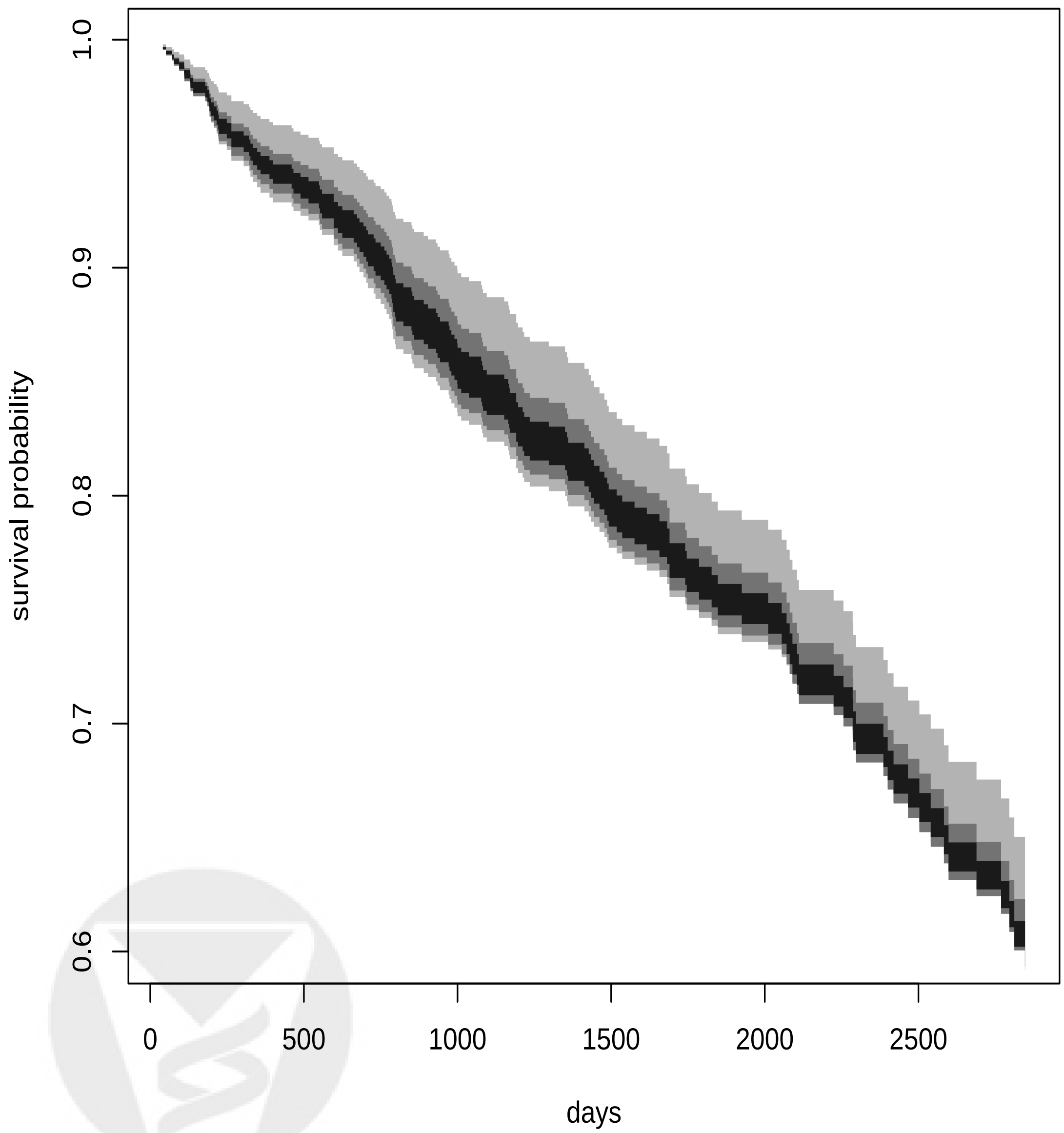
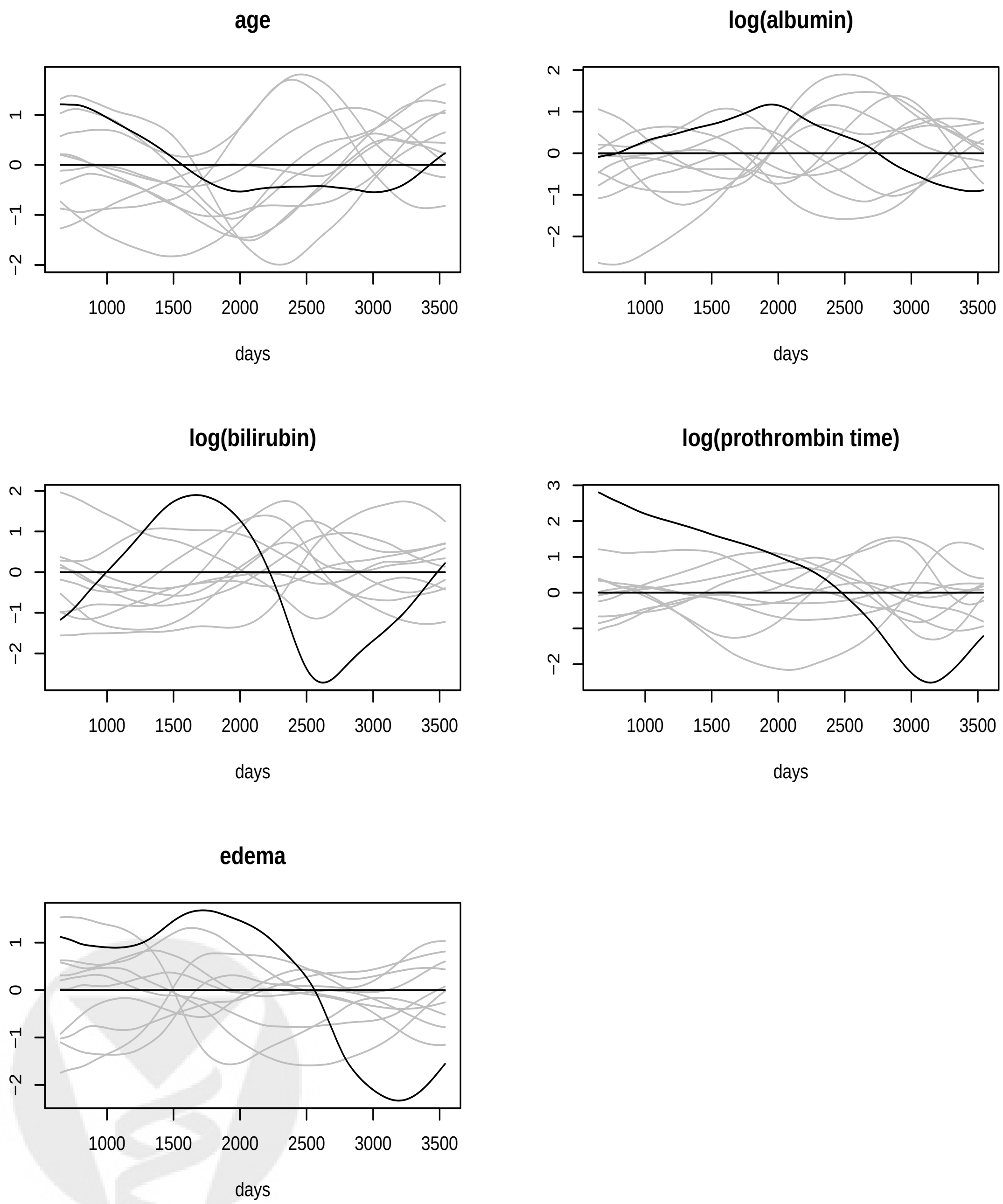


Figure 4. Graphic methods for identifying time-varying regression parameters with the Mayo data



References

Andersen, P., Borgan, O., Gill, R., and Keiding, N. (1993), " Statistical Models Based on Counting Processes", New York: Springer-Verlag.

Andersen, P. and Gill, R. (1982), " Cox's Regression Models for Counting Processes: A Large Sample Study", *Annal of Statistics*, **10**, 1100-1120.

Bickel, P. and Rosenblatt, R. (1973), " On some Global Measures of the Deviations of Density Function Estimates", *Annal of Statistics*, **1**, 1171-1095.

Cai, Z., Fan, J., and Li, R. (2000), " Efficient Estimation and Inference for Varying-Coefficient Models", *Journal of American Statistical Association* , **95**, 888-902.

Cai, Z., and Sun, Y. (2003), " Local Linear Estimation for Time-Dependent Coefficients in Cox's Regression Models", *Scandinavian Journal of Statistics* , **30**, 93-111.

Cox, D. R. (1972), " Regression Models and Life-Tables (with discussion)", *Journal of the Royal Statistical Society, Series B (Methodological)*, **34**, 187-220.

Dickson, E., Fleming, T., Grambsch, P., Fisher, L. and Langworthy, A. (1989), " Prognosis in Primary Biliary Cirrhosis: Model for Decision Making.", *Hepatology* , **10**, 1-7.

Efron, B. and R.J. Tibshirani (1993), " An Introduction to the Bootstrap", Chapman and Hall, New York, 239-241.

Eubank, R.L. (1988), " Spline Smoothing and Nonparametric Regression", Marcel Dekker, New York.

Fan, J. and Zhang, W. (2000), " Simultaneous Confidence Bands and Hypothesis Testing in Varying-coefficients Models", *Scandinavian Journal of Statistics*, **27**, 715-731.

FLEMING, T. AND HARRINGTON, D. (1991), " Counting Processes and Survival Analysis", John Wiley and Sons, Inc., New York.

GASSER, T. AND MULLER, H. (1979), " Kernel Estimation of Regression Functions." In *Smoothing Techniques for Curve Estimation*, Lecture Notes in Mathematics 757, 23-68. Berlin: Springer-Verlag.

GILBERT, P., WEI, L.J., KOSOROK, M. AND CLEMENS, J. (2002), " Simultaneous Inferences on the Contrast of Two Hazard Functions with Censored Observations" *Biometrics*, **58**, 773-780.

HALL, P. (1979), "The Rate of Convergence of Normal Extremes", *Journal of Applied Probability*, **16**, 433-439.

HALL, P. (1993), "On Edgeworth Expansion and Bootstrap Confidence Bands in Nonparametric Curve Estimation", *Journal of the Royal Statistical Society B (Methodological)* **55**, 291-304.

HASTIE, T. AND TIBSHIRANI, R. (1993), "Varying-Coefficient Models", *Journal of the Royal Statistical Society. Series B (Methodological)*, **55**, 757-796.

HOOVER, D., RICE, J., WU, C., AND YANG, L. (1998), "Nonparametric Smoothing Estimates of Time-Varying Coefficient Models With Longitudinal Data ", *Biometrika*, **85**, 809-822.

HUANG, J., WU, C., AND ZHOU, L. (2002), "Varying Coefficient Models and Basis Function Approximations for the Analysis of Repeated Measurements ", *Biometrika*, **89**, 111-128.

HUNSBERGER, S. (1994), 'Semiparametric regression in likelihood-based models", *Journal of the American Statistical Association*, **89**, 1354-1365.

LEDOUX, M. AND TALAGRAND, M. (1991), *Probability in Banach Spaces: Isoperimetry and Processes* , New York: Springer-Verlag Press.

Lin, D., Fleming, T., and Wei, L.J. (1994), "Confidence Bands for Survival Curves under Proportional Hazards Model ", *Biometrika*, **81**, 61-71.

Lin, D., Wei, L.J., and Ying, Z. (1993), "Checking the Cox model with Cumulative Sums of Martingale-based Residuals ", *Biometrika*, **80**, 557-572.

Lin, D., Wei, L.J., and Ying, Z. (2002), "Model-Checking Techniques Based on Cumulative Residuals ", *Biometrics*, **58**, 1-12.

Martinussen, T. and Scheike, T.H. (2002) "A flexible additive multiplicative hazard model",*Biometrika*, **89**, 283-298.

Martinussen, T., Scheike and Skovgaard, I. M. (2002) "Efficient estimation of fixed and time-varying covariates effects in multiplicative intensity models", *Scandinavian Journal of Statistics*, **29** , 57-74.

Murphy, S. and Sen, P. (1991), "Time-dependent Coefficients in a Cox-type Regression Model", *Stochastic Processes and their Applications*, **39**, 153-180.

Murphy, S. (1993), "Testing for Time-dependent Coefficients in a Cox's Regression Model", *Scandinavian Journal of Statistics*, **20**, 35-50.

Rosenblatt, M. (1977), "On the Maximal Deviation of k-dimensional Density Estimates", *The Annals of Probability*, **4**, 1008-1015.

Schoenfeld, D. (1982), "Partial Residuals for the Proportional Hazards Regression Model " *Biometrika*, **69**, 239-241.

Staniswalis, J. (1989), "The Kernel Estimate of a Regression Function in Likelihood-based Models " *Journal of the American Statistical Association,*, **84**, 276-283.

Tusnady, G. (1977), "A Remark on the Approximation of the Sample df in the Multi-dimensional Case", *Period. Math. Hungar.*, **8**, 53-55.

VALSECCHI, M., SILVESTRI, D., AND SASIENI, P. (1996), "Evaluation of Long-term Survival: Use of Diagnostics and Robust Estimators with Cox's Proportional Hazards Model", *Statistics in Medicine*, **15**, 2763-2780.

VERWEIJ, P. AND VAN HOUWELINGEN, H. (1995), "Time-dependent Effects of Fixed Covariates in Cox Regression", *Biometrics*, **51**, 1550-1556.

WINNETT, A. AND SASIENI, P. (2003), "Iterated Residuals and Time-varying Covariate Effect in Cox Regression", *Journal of the Royal Statistical Society, Series B (Methodological)*, **65**, 473-488.

YANDELL, B. (1983), "Nonparametric Inference for Rates with Censored Survival Data" *The Annals of Statistics*,, **11**, 1119-1135.

ZUCKER, D. AND KARR, A. (1990), "Nonparametric Survival Analysis with Time-dependent Covariate Effects: A Penalized Partial Likelihood Approach", *The Annals of Statistics*, **18**, 329-353.