Statistica Sinica **21** (2011), 901-921

INFERENCE FOR PARTLY LINEAR ADDITIVE COX MODELS

Jiancheng Jiang[1] and Xuejun Jiang[2,3]

[1] *University of North Carolina, Charlotte,* [2] *Zhongnan University of Economics and Law and* [3] *The Chinese University of Hong Kong*

Abstract: The partly linear additive Cox model is a useful tool for modeling failure time data with multiple covariates. The global smoothing method based on polynomial splines has been demonstrated as an efficient estimation approach for this model in the sense that it achieves the semiparametric information bound. However, there is no method available for consistently estimating the asymptotic variance matrix of the resulting estimators of finite parameters, which hampers inference for the model. This motivates us to propose a bootstrap method for estimating the distributions of the estimators; it is shown to be consistent. Moreover, to test linear hypotheses on the finite parameters, we propose a new test statistic and obtain its asymptotic null distribution. We show that the test is consistent and can detect alternatives nearing the null hypothesis at a rate of $\sqrt{n}$. Our results enable inference about the model based on the efficient polynomial splines estimation. Simulations are conducted to demonstrate nice performance of the proposed method. A data example is also given.

Key words and phrases: Bootstrap, conditional hazard rate, hypothesis testing, partial likelihood, polynomial spline.

1. Introduction

As an extension of the Cox (1972) model, the partly linear Cox model (Sasieni (1992)) specifies the conditional hazard rate of failure time T given the covariate value $(\mathrm{x}, \mathrm{w}) \in R^d \times R^J$ as

$$\begin{aligned} \lambda\{t; \mathrm{x}, \mathrm{w}\} &= \lim_{\Delta t \downarrow 0} \left[\frac{1}{\Delta t} \Pr\{t \le T < t + \Delta t | T \ge t, \mathrm{x}, \mathrm{w}\} \right] \\ &= \lambda_0(t) \exp\{\beta' \mathrm{x} + \phi(\mathrm{w})\}, \end{aligned} \tag{1.1}$$

where $\lambda_0(\cdot) \ge 0$ is an unspecified base-line hazard, β is a d-vector of parameters, and $\phi(\mathrm{w})$ is an unknown function of w. This model provides flexibility to the covariate specification when some continuous covariates may have nonlinear effects on the risk function. There are many works in the literature concerning estimation of the model. For example, Grambsch, Therneau, and Fleming

(1990) and Fleming and Harrington (1991) explored the functional form of the covariate effect by smoothed martingale residuals; Sasieni (1992) calculated the information bound for estimating β and suggested using the partial likelihood based on a spline to estimate the model; Dabrowska (1997) considered the continuously stratified Cox model in Sasieni (1992), a variation of (1.1) where the covariate w was embedded into the baseline; Heller (2001) studied the profile partial likelihood estimation using a kernel function.

Model (1.1) is useful for modeling failure time data with multiple covariates, but for high-dimensional covariate w it suffers from the so-called "curse-of-dimensionality" problem in estimation. One of the methods attenuating this difficulty uses an additive structure for the function $\phi(\cdot)$, as in Huang (1999), which leads to the partly linear additive Cox model. It specifies the conditional hazard of the failure time T given the covariate value (x, w) as

$$\lambda\{t; \mathrm{x}, \mathrm{w}\} = \lambda_0(t) \exp\{\beta' \mathrm{x} + \phi(\mathrm{w})\}, \tag{1.2}$$

where $\phi(\mathrm{w}) = \phi_1(w_1) + \cdots + \phi_J(w_J)$. The parameters of interest are the finite parameter vector β and the unknown ϕ's. The former measures the effect of the treatment variable vector x, and the latter may be used to suggest a parametric structure of the risk. This model allows one to explore nonlinearity of certain covariates, avoids the "curse-of-dimensionality" problem inherent in the saturated multivariate semiparametric hazard regression model (1.1), and retains the nice interpretability of the traditional linear structure in Cox's model (Cox (1972)). See the discussions in Hastie and Tibshirani (1990).

The global smoothing method based on polynomial splines is efficient for estimating the finite parameters in (1.2), in the sense that it achieves the semiparametric information bound. See the seminal work of Huang (1999). This indicates that the estimator of β is asymptotically most efficient among all the regular estimators (see for example, van der Vaart (1991) and Chapter 3 of Bickel et al. (1993)). However, there is no consistent estimator available for the asymptotic variance matrix of the estimator of β. This is similar to parameter estimators for censored regression models or from nonsmooth estimation equations, where the asymptotic covariance matrix of the estimator involves unknown density functions. In general, resampling techniques are employed to deal with this problem. For example, Bilias, Chen, and Ying (2000) Zhou (2006), and Zeng and Lin (2008) used resampling to estimate the variance matrix of the estimators.

For the global smoothing method, Huang (1999, p.1540) empirically suggested using the inverse of the observed partial information matrix to estimate the variance matrix, but its consistency is unknown and seems difficult to establish. This hampers applications of the efficient polynomial spline estimation for model (1.2). It is desirable to develop a consistent estimate of the variance matrix.

In this paper we propose a bootstrap approach to estimating the distribution of the above estimator of β, which furnishes a consistent variance estimation method for the efficient estimator of β in Huang (1999). The result allows for a useful inference tool for model (1.2) based on the efficient polynomial spline estimation. Moreover, we introduce a new statistic to test some linear hypotheses about β, while leaving the nonparametric components as nuisance functions. The asymptotic null distribution and consistency of the test are established, and the limit rate for the detectable alternatives converging to the null hypothesis is $\sqrt{n}$. To derive our theoretic results, we draw a parallel between the estimator and its bootstrap analogue and obtain their Bahadur's representations which share a similar form, except that the representation for the latter is a random weighting sum while the former is a uniform weighting sum. This facilitates the proofs of theorems, and the parallel arguments are novel, representative, and applicable to other scenarios.

Note that model (1.2) was extended by Cai et al. (2007) to multivariate cases with one additive component. They proposed a profile pseudo-partial likelihood estimation approach. The resulting estimators of the finite parameters allowed for consistent variance estimation but they were not efficient, since the asymptotic variances of the estimators did not achieve the semiparametric information bound. Therefore, it is meaningful to study the efficient estimation method based on the polynomial splines approximation.

The paper is organized as follows. In Section 2, we introduce the partial likelihood of model (1.2) along with the polynomial splines based estimation. In Section 3, we propose a bootstrap method for estimating the distribution of the efficient estimator of β, and establish the consistency and asymptotic normality of the bootstrap estimation. In Section 4, we consider hypothesis testing for β and derive asymptotics for the proposed test. We present the details of implementation and conduct simulations in Section 5. Some concluding remarks are given in Section 6. Technical conditions and proofs are provided in the Appendix.

2. Partial Likelihood and Global Smoothing Based on Polynomial Splines

Suppose that there are n independent individuals in a study cohort. Let $\{\mathrm{X}_i, \mathrm{W}_i\}$ denote the covariate vector for the ith subject ($i = 1, \cdots, n$). Then observed data for the ith subject is $\{S_i, \delta_i, \mathrm{W}_i, \mathrm{X}_i\}$, where S_i is the observed event time for the ith subject, which is the minimum of the potential failure time

T_i and the censoring time C_i, and δ_i is the indicator of failure. Let $r_i(\beta, \phi) = \exp\{\beta' \mathrm{X}_i + \phi(\mathrm{W}_i)\}$. Then the partial likelihood function for model (1.2) is

$$L(\beta, \phi) = \prod_{i=1}^{n} \Big\{ \frac{r_i(\beta, \phi)}{\sum_{j \in \mathcal{R}_i} r_j(\beta, \phi)} \Big\}^{\delta_i}, \tag{2.1}$$

where $\mathcal{R}_i = \{j : S_j \geq S_i\}$ is the risk set at time S_i.

If $\phi(\cdot)$ has been parameterized, one can obtain the maximum partial likelihood estimator by maximizing the partial likelihood (2.1) with respect to β and the parameters in $\phi(\cdot)$. Since ϕ is unknown, the partial likelihood function in (2.1) cannot be maximized for estimating ϕ.

In the following, we consider the partial likelihood estimation for β and $\phi(\cdot)$ using polynomial splines approximation. The use of polynomial splines in estimating the fully nonparametric additive Cox model based on the partial likelihood was first proposed by Stone (1986). This method was extended to the partly linear additive Cox model by Huang (1999). In principle, the additive components can be approximated by polynomial splines with different numbers of knots. For easy exposition in the following, we employ the same number of knots for each component, and use some notation from Huang (1999) and Fan and Jiang (2009).

Without loss of generality, assume that W takes values in $\mathcal{W} = [0, 1]^J$. Let $\underline{\xi} = \{0 = \xi_0 < \xi_1 < \cdots < \xi_K < \xi_{K+1} = 1\}$ be a partition of $[0, 1]$ into K subintervals $I_{Ki} = [\xi_i, \xi_{i+1}),\ i = 0, \ldots, K-1,\ h_k = \xi_k - \xi_{k-1}$, and $I_{KK} = [\xi_K, \xi_{K+1}]$, where $K \equiv K_n = O(n^v)$ with $0 < v < 0.5$ being a positive integer such that $h \equiv \max_{1 \leq k \leq K+1} |h_k| = O(n^{-v})$. Let $\mathcal{S}(\ell, \underline{\xi})$ be the space of polynomial splines of degree $\ell \geq 1$ consisting of functions $s(\cdot)$ satisfying

(i) the restriction of $s(\cdot)$ to I_{Ki} is a polynomial of order $\ell - 1$ for $1 \leq i \leq K$,

(ii) for $\ell \geq 2$, $s(\cdot)$ is $\ell - 2$ times continuously differentiable on $[0, 1]$.

According to Schumaker (1981, p.124), there exists a local basis $B_i(\cdot),\ 1 \leq i \leq q$, for $\mathcal{S}(\ell, \underline{\xi})$ with $q = K_n + \ell$, such that for any $\phi_{nj}(\cdot) \in S(\ell, \underline{\xi})$,

$$\phi_{nj}(w_j) = \sum_{i=1}^{q} b_{ji} B_i(w_j), \quad 1 \leq j \leq J.$$

For convenience of exposition, let $\{B_i(\cdot)\}_{i=1}^q$ be the normalized B-spline basis (see de Boor (1978)), such that

$$\begin{cases} B_i(w) = 0 \ \text{ unless } \xi_i < w < \xi_{i+\ell}, \\ B_i(w) \geq 0, \\ \sum_{i=1}^q B_i(w) = 1. \end{cases}$$

A specific example for such a basis was given in Zhou, Shen and Wolfe (1998). Put $B(w) = (B_1(w), \ldots, B_q(w))'$, $\mathrm{B}(\mathrm{w}) = (B'(w_1), \ldots, B'(w_J))'$, $\mathrm{b}_j = (b_{j1}, \ldots, b_{jq})'$, and $\mathrm{b} = (\mathrm{b}_1', \ldots, \mathrm{b}_J')'$. Then $\phi_{nj}(w_j) = \mathrm{b}_j' B(w_j)$ and $\phi_n(\mathrm{w}) \equiv \sum_{j=1}^J \phi_{nj}(w_j) = \mathrm{b}'\mathrm{B}(\mathrm{w})$. Under regular smoothness assumptions, ϕ_j's can be well-approximated by functions in $\mathcal{S}(\ell, \underline{\xi})$. Therefore, by (2.1), the logarithm of an approximated partial likelihood is

$$\ell(\beta, \mathrm{b}) = \sum_{i=1}^n \delta_i \Big\{ \beta'\mathrm{X}_i + \phi_n(\mathrm{W}_i) - \log \sum_{k \in \mathcal{R}_i} \exp[\beta'\mathrm{X}_k + \phi_n(\mathrm{W}_k)] \Big\}, \qquad (2.2)$$

where $\phi_n(\mathrm{W}_i) = \sum_{j=1}^J \phi_{nj}(W_{ji})$, with W_{ji} being the jth component of W_i, for $i = 1 \ldots, n$. Let $(\hat{\beta}, \hat{b})$ maximize (2.2). Then an estimator of $\phi(\cdot)$ at point w is simply $\hat{\phi}(\mathrm{w}) = \sum_{i=1}^J \hat{\phi}_j(w_j)$ with $\hat{\phi}_j(w_j) = \hat{\mathrm{b}}_j' B(w_j)$. Computationally, the maximization problem in (2.2) can be solved via the existing Cox regression program, for example **coxph** and **bs** in the R software (for details, see Huang (1999)).

With the estimators of β and $\phi(\cdot)$, one can estimate the cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ under mild conditions by the following Breslow estimator (see Breslow (1972, 1974):

$$\hat{\Lambda}_0(t) = \int_0^t \big[\sum_{i=1}^n Y_i(u) \exp\{\hat{\beta}'\mathrm{X}_i + \hat{\phi}(\mathrm{W}_i)\}\big]^{-1} \sum_{i=1}^n dN_i(u),$$

where $Y_i(u) = 1(S_i \geq u)$ is the at-risk indicator and $N_i(u) = 1(S_i < u, \delta_i = 1)$ is the associated counting process.

The estimator $\hat{\beta}$ achieves $\sqrt{n}$-consistency. Especially, it achieves the semiparametric information lower bound (see Huang (1999, Remark 3.2)). This shows that the estimator is asymptotically efficient among all the regular estimators. However, the information lower bound cannot be consistently estimated, which makes inference for β difficult in practice. This motivates us to propose the bootstrap estimation for the distribution of $\hat{\beta}$.

3. Bootstrap Estimation for the Parameter Vector

The estimation of the variance of $\hat{\beta}$ allows one to construct confidence intervals for β. Since a direct consistent estimator is not available, it is useful to develop a bootstrap method for this task.

For the proportional hazards model of Cox (1972), Burr (1994) studied three methods of bootstrapping: *ordinary resampling from the empirical cumulative distribution function; resampling conditional on the covariates; resampling conditional on the covariates and the censoring pattern.* We opt for the first method, since its consistency is guaranteed (see Theorem 1 below), but the other two bootstrap methods are also interesting and worth investigation. Specifically, we resample from the empirical distribution F_n of the observations $\{S_i, \delta_i, \mathrm{W}_i, \mathrm{X}_i\}_{i=1}^n$. The bootstrap procedure is detailed as follows.

1. Resample a bootstrap sample $\{S_i^*, \delta_i^*, \mathrm{W}_i^*, \mathrm{X}_i^*\}_{i=1}^n$ from F_n.
2. Based on the sample, fit model (1.2) and obtain the estimated value $\hat{\beta}^*$ of β.
3. Repeat steps 1 and 2 to obtain a sample of $\hat{\beta}^*$'s, $\hat{\beta}^{*(k)}$, $k = 1, \ldots, B$, say. Compute $\hat{\beta}^{*(k)} - \hat{\beta}$.
4. Use the bootstrap sample $\{\hat{\beta}^{*(k)} - \hat{\beta}\}_{k=1}^B$ to determine the quantiles of $\hat{\beta} - \beta$.

As in Huang (1999, Thm. 3.1), let $\mathrm{Z} = (\mathrm{X}, \mathrm{W})$ and let $I(\beta) = E[l_\beta^*(S, \delta, \mathrm{Z})]^{\otimes 2}$ be the information bound for estimation of β, where $l_\beta^*(S, \delta, \mathrm{Z}) = \int_0^\tau (\mathrm{X} - a^*(t) - g^*(\mathrm{W}))\, dM(t)$ is the efficient score for estimating β in model (1.2), τ is defined in (A3) of Appendix I, $g^*(\mathrm{w}) = \sum_{j=1}^J g_j^*(w_j)$, and $a^*, g_1^*, \ldots, g_J^*$ are the unique L_2 functions that minimize $E[\delta\, \|\mathrm{X} - a(S) - \sum_{j=1}^J g_j(\mathrm{W}_j)\|^2]$ with $\|\cdot\|$ denoting the Euclidean norm.

The following theorem demonstrates that the conditional distribution of $(\hat{\beta}^* - \hat{\beta})$ given F_n is asymptotically the same as the unconditional distribution of $\hat{\beta} - \beta$; this establishes the consistency of the proposed bootstrap method.

Theorem 1. *Assume that the conditions in Appendix I hold. If* $0.25/p < v < 0.5$ *and* $v(q+p) > 0.5$, *where* p *is the measure of smoothness of* ϕ_j *defined in Condition* (A1), *and with* q *defined in Condition* (A6), *then*

$$\sup_t |P\{\sqrt{n}(\hat{\beta}^* - \hat{\beta}) < t | F_n\} - P\{\sqrt{n}(\hat{\beta} - \beta) < t\}| \to 0, \quad a.s..$$

From the theorem, we can estimate the variance matrix $n^{-1}I^{-1}(\beta)$ of the $\hat{\beta}$ by the bootstrap sample variance

$$S_B = (B-1)^{-1} \sum_{k=1}^B (\hat{\beta}^{*(k)} - \hat{\beta})(\hat{\beta}^{*(k)} - \hat{\beta})',$$

and approximate the $100(1-\alpha)\%$ confidence region of β by

$$(\hat{\beta} - \beta)' S_B^{-1} (\hat{\beta} - \beta) \le \chi_d^2(\alpha),$$

where $\chi_d^2(\alpha)$ is the $(1-\alpha)$th quantile of the chi-squared distribution with d degrees of freedom.

4. Testing Linear Hypotheses

Many problems of statistics can be reduced to the problem of testing a linear hypothesis. In particular, it is interesting to investigate whether β, lying in the d-dimensional linear space R^d, belongs to a linear subspace of dimension $d-m$, $m>0$. This reduces to testing,

$$H_0: A\beta = 0 \text{ versus } H_1: A\beta \neq 0,$$

where A is a known $m \times d$ matrix with rank m and d is the dimension of β. Since A is of full row rank, by the singular value decomposition there exists a unique $(d-m)\times d$ matrix C satisfying $CC' = I_{d-m}$ and $AC' = 0$. Then the null hypothesis becomes $H_0': \beta = C'\gamma$, where γ is a $(d-m)\times 1$ vector.

As the additive components are nuisance functions, the testing problem is semiparametric. If $V(\beta) = [I^{-1}(\beta) - C'\{CI(\beta)C'\}^{-1}C]$ and $\hat{\gamma}$ is the estimator of γ under H_0', then we have a Bahadur representation for the estimators.

Theorem 2. *If the conditions in Appendix I hold, then under* H_0,

(i) $C'(\hat{\gamma}-\gamma) = C'\{CI(\beta)C'\}^{-1}Cn^{-1}\sum_{i=1}^n \ell_\beta^*(S_i,\delta_i,\mathrm{Z}_i) + o_p(n^{-1/2})$,

(ii) $\sqrt{n}(\hat{\beta} - C'\hat{\gamma}) = V(\beta)n^{-1/2}\sum_{i=1}^n \ell_\beta^*(S_i,\delta_i,\mathrm{Z}_i) + o_p(1)$.

Note that $V(\beta)I(\beta)V(\beta) = V(\beta)$. It follows from Theorem 2 that, under H_0, $\sqrt{n}(\hat{\beta}-C'\hat{\gamma}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V(\beta))$. Let $V^*(\beta) = I_d - \{I(\beta)\}^{1/2}C'\{CI(\beta)C'\}^{-1} C\{I(\beta)\}^{1/2}$. Then $V(\beta) = \{I(\beta)\}^{-1/2}V^*(\beta)\{I(\beta)\}^{-1/2}$, $V^*(\beta)$ is a symmetric and idempotent matrix with rank m, and $V(\beta)$ is a $d\times d$ singular matrix with rank $m<d$. Therefore, the Wald-type test statistic, which takes the form $T_{n0} = (\hat{\beta}-C'\hat{\gamma})'V^{-1}(\beta)(\hat{\beta}-C'\hat{\gamma})$, cannot be employed, since the inverse of $V(\beta)$ does not exist.

An intuitive test method is to directly compare the difference between the estimators of parameters under the null and alternative hypotheses, which leads to the test statistic

$$T_n = ||\sqrt{n}(\hat{\beta}-C'\hat{\gamma})||^2,$$

where $\hat{\beta}$ and $\hat{\gamma}$ are the estimators of β and γ, respectively, under H_1 and H_0. In principle, a large value of T_n suggests rejecting H_0.

Since $V^*(\beta)$ is idempotent, $V^*(\beta)$ has m eigenvalues equal to 1 and $d-m$ eigenvalues equal to 0. Let D be a $d\times m$ matrix consisting of m normalized eigenvectors corresponding to the m nonzero eigenvalues of $V^*(\beta)$, and let $\{w_i, i=1,\ldots,m\}$ be the eigenvalues of $D\{I(\beta)\}^{-1}D'$. The next theorem describes the asymptotic null distribution of the test statistic T_n.

Theorem 3. *If the conditions in Appendix* I *hold, then under* H_0*,* $T_n \xrightarrow{\mathcal{D}} \sum_{i=1}^{m} w_i \chi^2_{1,i}$*, where the* $\chi^2_{1,i}$*'s are independent chi-square variables with one degree of freedom.*

Unfortunately, this result cannot be used to make statistical inference directly, because there is no consistent estimator available for the unknown weights w's. We propose a bootstrap method to estimate the null distribution. Specifically, given F_n, one can draw a bootstrap sample, $\{S_i^*, \delta_i^*, \mathrm{W}_i^*, \mathrm{X}_i^*\}_{i=1}^n$, from F_n as in Section 3, then fit the null model under H_0 and the alternative model (1.2). Denote by $\hat{\beta}^*$ and $\hat{\gamma}^*$ the resulting estimators of β and γ, respectively. We take the bootstrap version of T_n to be

$$T_n^* = ||\sqrt{n}(\hat{\beta}^* - C'\hat{\gamma}^*) - \sqrt{n}(\hat{\beta} - C'\hat{\gamma})||^2.$$

Theorem 4. *Assume that the conditions in Theorem* 3 *hold. Then under* H_0*,* $\sup_t |P(T_n^* < t|F_n) - P(T_n < t)| \to 0$ *a.s..*

Although the null distribution of T_n cannot be calculated, the null distribution of T_n^* can be obtained by resampling. Theorem 4 has the null distribution of T_n well-approximated by that of T_n^*.

To study the power of the proposed test, we consider the local (Pitman) alternatives of the form

$$H_{1n}:\ \beta = C'\gamma + n^{-r}\beta_n,$$

where β_n is a sequence of nonzero vectors in R^d such that $\beta_n \to \beta^*$. Assume that there does not exist a $\gamma_n \in R^{d-m}$ such that $\beta_n = C'\gamma_n$, otherwise H_{1n} coincides with H_0 and any test has no power to detect H_{1n}. Both the null and alternative are semiparametric, since the additive components $\{\phi_j(\cdot)\}_{j=1}^J$ are nuisance functions. From Theorem 3, an approximate level-α test can be constructed as $\psi_n = 1\{T_n \geq \chi_{n,1-\alpha}\}$, where $\chi_{n,1-\alpha}$ is the $100(1-\alpha)$th percentile of the limit distribution of T_n under H_0. By Theorem 4, one can estimate $\chi_{n,1-\alpha}$ by the $100(1-\alpha)$th percentile of distribution of T_n^*.

The probability of type II error at the alternative H_{1n} is then given by $\beta(\alpha, r) = P\{\psi_n = 0|H_{1n}\}$. The following theorems characterize how fast the test ψ_n detects the alternatives.

Theorem 5. *If the conditions in Appendix I hold, then for the testing problem* $H_0 \leftrightarrow H_{1n}$ *when* $r < 1/2$*, the test* ψ_n *can detect alternative* H_{1n} *asymptotically with probability one.*

We conclude this section by considering the limiting behavior of the test statistic under the local alternative H_{1n} with $r = 1/2$.

Theorem 6. *If the conditions in Appendix I hold, then under H_{1n} with $r = 1/2$,*

(i) $\sqrt{n}(\hat{\beta} - C'\hat{\gamma}) \xrightarrow{\mathcal{D}} \mathcal{N}(\beta^*, V(\beta))$,

(ii) *the test statistic T_n converges weakly to the random variable $\|\beta^* + \eta\|^2$, where η is a m-dimensional normal random variable with mean zero and variance matrix $V(\beta)$.*

The theorem shows that the test ψ_n can detect the local alternatives at a rate of $\sqrt{n}$, the optimal rate in all regular parametric tests.

5. Numerical Studies

5.1. Choice of knots

For implementation of the estimation method in (2.2), one needs to specify the location of knot sequence $\{\xi_k\}_{k=1}^{K_n}$, given the number of knots K_n. In practice, equally spaced and quantile knot methods are usually used. The latter places knots at the sample quantiles of the variable so that there are approximately the same number of observed values of the variable between any two adjacent knots. We choose it. The number of knots, K_n, acts as a smoothing parameter. It may be selected by visual trial and error to pick a value balancing smoothness against fidelity to the data. Experience from simulations suggest that stable and satisfactory results can be obtained by using three to ten knots.

More formal methods of selecting the number of knots are to minimize the mean squared errors of the fit, either by employing a formula approximating the mean square error (e.g. the plug-in estimators), or by a form of cross-validation (CV). For uncensored data, CV and generalized cross-validation (GCV) are commonly used for selection of K_n; see for example Hastie and Tibshirani (1990). For survival data, Ø'Sullivan (1988) proposed CV and GCV for choosing the smoothing parameter in smoothing spline estimation of relative risk; Nan et al. (2005) extended CV and GCV to choose the number of knots in regression spline estimation of a varying-coefficient Cox model. It is worth developing the CV and GCV methods to give a data-driven choice for the number of knots in the current setting, but we do not attempt this here.

Since our theoretic results hold for a large range of numbers of knots, we empirically decide the number of knots in simulations. Throughout this section, we employ the B-spline basis with three quantile knots. Even though the number of knots is not optimal, the simulation results are acceptable for evaluation of our methodology.

5.2. Simulations

We did simulations to demonstrate that the bootstrap method gives an accurate estimate for the distribution of $\hat{\beta}$, and to check the consistency and power of the proposed test. Specifically, we calculated the standard deviation of $\hat{\beta}$ in simulations, the bootstrap estimate of the standard error of $\hat{\beta}$, and the 95% coverage probability for β based on the bootstrap. We also compared the null distribution of the testing statistic T_n with its bootstrap version T_n^* and calculated the powers of the test under local contiguous alternatives.

In addition, we compared the proposed estimation method with a "naive method" using the inverse of the observed partial information matrix to estimate the variance matrix of $\hat{\beta}$ in Huang (1999). When there is only one additive component in the model, the profile pseudo-likelihood (PPL) method in Cai et al. (2007) can be used, and the Wald-type test based on this method can be constructed. We did comparison with the PPL method. We expected that our efficient-estimation-based test would be more powerful than the Wald-type test based on the PPL.

Example 1. Covariate X was generated from the bivariate normal distribution with marginal $N(0,1)$ and correlation 0.5, W_1 and W_2 were uniform on $[0,1]$ with correlation 0.5, and the failure time was from an exponential distribution with hazard function

$$\lambda(t; X, \mathrm{W}) = \lambda_0 \exp(\beta_1 X_1 + \beta_2 X_2 + \phi(\mathrm{W})), \tag{5.1}$$

where $\beta_1 = 0.6$, $\beta_2 = 0.4$, and $\phi(\mathrm{w}) = \phi_1(w_1) + \phi_2(w_2)$ with $\phi_1(w_1) = -8w_1(1 - w_1^2)$ and $\phi_2(w_2) = 2\log(200 + (w_2 - 1.2)^3)$. The censoring variable given (X, W) was uniformly distributed on $[0,6]$ and independent of the failure time. The baseline λ_0 was taken to be 1. There were about 47% censored data.

First, we assessed our estimation approach with 1,000 simulations. We considered different sample sizes $n =$400, 800, 1,600. In each simulation, we used the "naive method" mentioned above and 1,000 bootstrap replicates to calculate the standard error of $\hat{\beta}$. Let $\hat{se}_b$ and $\hat{se}_{na}$ represent the estimated standard error for the bootstrap and naive methods, respectively. The estimators and their standard deviations (SD) were evaluated along with the average of the estimated standard error for the estimators. The coverage probability (CP_b and CP_{na}) of the 95% confidence interval for β was also calculated based on the bootstrap and naive methods. For the naive method, the CP_{na} for β was calculated using the normal approximation and the estimated standard error $\hat{se}_{na}$.

Table 1 reports the simulation results. It is evident that the estimated standard errors were quite close to the corresponding standard deviations and the

Table 1. Summary of simulation results ($\beta_1 = 0.6$ and $\beta_2 = 0.4$).

Size	Parameters	Mean($\hat{\beta}$)	$SD(\hat{\beta})$	$\hat{se}_b$ ($\hat{se}_{na}$)	95% CP_b (CP_{na})
400	β_1	0.612	0.096	0.094 (0.092)	0.92 (0.93)
	β_2	0.405	0.089	0.092 (0.089)	0.93 (0.94)
800	β_1	0.607	0.063	0.065 (0.064)	0.94 (0.95)
	β_2	0.402	0.060	0.063 (0.062)	0.94 (0.96)
1,600	β_1	0.602	0.044	0.045 (0.045)	0.95 (0.96)
	β_2	0.403	0.045	0.044 (0.043)	0.95 (0.94)

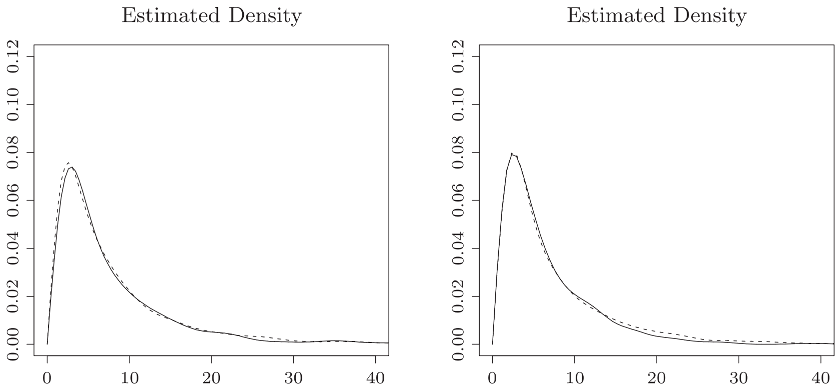

Figure 1. Estimated densities. Left panel: n=400; right panel: n=800. Solid — true, dotted — the bootstrap approximation.

coverage probability was satisfactory for both methods. The naive and bootstrap estimation methods for β performed well in this example.

Second, we checked the performance of the proposed test. To this end, we considered testing the null hypothesis, $H_0 : A\beta = 0$ versus $H_1 : A\beta \neq 0$,where $A = (2, -3)$. For each simulation, we obtained four bootstrap samples (this is merely for the reduction of computational cost; using more samples would not alternate our results) and computed the test statistic T_n^*. Pooling together the bootstrap samples from each simulation, we obtained 4,000 bootstrap statistics. Their sampling distribution, computed via the kernel density estimate, is the distribution of T_n^*. By using the kernel density estimation method, the distribution of the realized values of the test statistic T_n in simulations was the true distribution (except for the Monte Carlo errors). Figure 1 displays the estimated densities for T_n^* and the true densities of T_n. It is seen there that the bootstrapped distribution was good enough for approximating the true one. It is reasonable to use the bootstrap distribution to approximate the null distribution of the proposed test statistic T_n.

We next investigated the power of the test by considering the following al-

Table 2. Powers of the proposed test at significance level α

n	α	$\theta=0$	$\theta=0.2$	$\theta=0.4$	$\theta=0.6$	$\theta=0.8$	$\theta=1.0$
400	0.10	0.086	0.367	0.867	0.991	1.000	1.000
	0.05	0.040	0.239	0.756	0.982	1.000	1.000
800	0.10	0.094	0.606	0.989	1.000	1.000	1.000
	0.05	0.048	0.513	0.977	1.000	1.000	1.000
1,600	0.10	0.099	0.862	1.000	1.000	1.000	1.000
	0.05	0.042	0.788	1.000	1.000	1.000	1.000

ternative sequences indexed by $\theta = 0, 0.2, 0.4, 0.6, 0.8, 1.0$:

$$H_{1n,\theta} : \beta = (0.6, 0.4)' + \theta\sqrt{0.52}\,\beta_n,$$

where $\beta_n = (2/\sqrt{13}, -3/\sqrt{13})'$ is orthogonal to β and satisfies $\|\beta_n\| = 1$, and $\sqrt{0.52}$ is the length of β under H_0. Note that for all of the above alternatives, when θ is small, the null and alternative models are nearly impossible to differentiate, and especially when $\theta = 0$, the null and the alternative are the same. Therefore, it can be expected that, (i) when $\theta = 0$, the power of test should be close to the significance level; and (ii) the farther is θ away from 0, the greater is the power. These are consistent with the results in Table 2, where the power is calculated as the frequency of rejections in 1,000 simulations and in each simulation the null is rejected if the value of T_n^* is bigger than the critical value of the bootstrap distribution of T_n^*. The test sizes are reasonable and the power of the test increases with the sample size.

Example 2. Motivated by comments from an AE and a referee, we compared the power of the proposed test with the Wald-type test based on the PPL method. Since the PPL method deals with only the partly linear model with $J = 1$, we modified model (5.1) by taking

$$\lambda(t; X, W) = \lambda_0 \exp(\beta_1 X_1 + \beta_2 X_2 + \phi(W)), \tag{5.2}$$

where $\beta_1 = 0.6$, $\beta_2 = 0.4$, and $\phi(w) = \sin(4\pi w)$. X $= (X_1, X_2)'$ was generated as before, and W was uniform on $[0, 1]$. The censoring variable given (X$, W$) was uniformly distributed on $[0, 6]$ and independent of the failure time. The baseline λ_0 was taken to be 1.

For model (5.2), it is also interesting to compare the proposed bootstrap method with the naive method as in Example 1. Table 3 summarizes the simulation results for estimation. It seems that the proposed method is better than the naive method here since the estimated $\hat{se}$ is closer to the standard deviation and the coverage probability is comparable.

Table 3. Summary of simulation results ($\beta_1 = 0.6$ and $\beta_2 = 0.4$).

Size	Parameters	Mean($\hat{\beta}$)	$SD(\hat{\beta})$	$\hat{se}_b$ ($\hat{se}_{na}$)	95% CP_b (CP_{na})
400	β_1	0.595	0.082	0.079 (0.077)	0.95 (0.94)
	β_2	0.399	0.079	0.077 (0.075)	0.94 (0.94)
800	β_1	0.606	0.055	0.055 (0.054)	0.94 (0.94)
	β_2	0.404	0.054	0.053 (0.053)	0.95 (0.94)

Table 4. Comparison of powers at significance level α.

Test	α	$\theta = 0$	$\theta = 0.2$	$\theta = 0.4$	$\theta = 0.6$	$\theta = 0.8$	$\theta = 1.0$
T_n^*	0.05	0.044	0.415	0.940	0.998	1.000	1.000
	0.10	0.096	0.558	0.972	0.999	1.000	1.000
T_{n1}	0.05	0.047	0.217	0.742	0.977	0.997	1.000
	0.10	0.093	0.323	0.852	0.988	0.997	1.000

We also investigated whether the efficient global spline estimation leads to a more powerful test than the PPL method. To this end, we considered the Wald-type statistic

$$T_{n1} = n(\hat{\beta}_{PL} - \beta)'\hat{\Omega}^{-1}(\hat{\beta}_{PL} - \beta) \to \chi^2(2), \text{ under } H_0,$$

where $\hat{\beta}_{PL}$ is the PPL estimate of β and $\hat{\Omega}$ is the estimate of the asymptotic variance matrix of $\sqrt{n}(\hat{\beta}_{PL} - \beta)$ in Cai et al. (2007).

To investigate the powers of the tests, we considered the alternative sequence $H_{1n,\theta}$ in Example 1 with $\theta = 0$ corresponding to the null hypothesis. We set the sample size $n = 400$ and did 600 simulations for calculating the powers of the test statistic T_{n1}. For each simulation, we drew data from $H_{1n,\theta}$ and calculated the values of T_{n1}. If the value of T_{n1} was larger than the critical value of $\chi^2(2)$ for significance level α, then the null H_0 was rejected. The power of the test T_{n1} is just the relative frequency of rejections in the 600 simulations.

Table 4 reports the powers of the test statistics T_n^* and T_{n1}. It is obvious that the proposed test T_n^* was more powerful than the Wald-type test. This was to be expected, since T_n^* is constructed from the efficient estimate and T_{n1} is based on the less efficient PPL estimate.

5.3. A data example

We applied our proposed procedure to analyze data from the well-known "Framingham Heart Study (FHS)" (Dawber (1980)). The FHS study started in 1948. The dataset used here contains all participants in the study who had an examination at age 44 or 45, which we refer to as the "age 45" exam, and were disease-free at that examination in the sense that there existed no history of hypertension or glucose intolerance, and no previous experiences of a coronary

Table 5. Estimated parameters for the FHS data. $\hat{\beta}$ - the estimated parameters, $\hat{se}_b$ - the standard error of $\hat{\beta}$, LCI and UCI are respectively the lower and upper bounds of the 95% confidence interval.

Effect	$\hat{\beta}$	$\hat{se}_b$	LCI	UCI
Age (at the "age 45" exam)	0.0676	0.1168	-0.16544	0.28338
Cholesterol, mg/dl	0.0041	0.0016	0.00069	0.00717
Systolic blood pressure, mm Hg	0.0160	0.0052	0.00471	0.02602
Smoking status: yes=1,no=0	0.3254	0.1491	0.03786	0.62090
Gender: female=1, male=0	-0.6180	0.1418	-0.90471	-0.34171

heart disease (CHD). The sample size n is 1571. There are about 90.42% censoring in the dataset. The failure time, times to CHD, is our concern. The risk factors are age (at age "45" exam), gender, systolic blood pressure (SBP), body mass index (BMI), cholesterol level, waiting time, and cigarette smoking.

Times to CHD are measured from the time at the "age 45" exam to the occurrences of the corresponding diseases. The variables BMI, SBP, cholesterol level, and smoking status are measured at the "age 45" exam. We fit the data using model (1.2) with

$$\text{x} = (\text{Age at the "age 45" exam, SBP, Cholesterol,}\\ \text{Smoking status, Gender})',$$

w_1 = Waiting time, w_2 = MBI, and $w_3 = w_1 * w_2$, where w_3 is used to explore possible interaction between the BMI and Age, and "Waiting time" is the time elapsed from the initial entry into the FHS study to the "age 45" exam. Since the year of birth equals to (1948 + waiting time - 45), we include the waiting time in the model to adjust for birth cohort effect.

Table 5 reports the estimated parameters and their estimated standard errors along with their 95% confidence intervals based on the bootstrap method. It is evident that all of the selected risk factors except for Age are statistically significant at significance level 5%. The nonsignificance result for Age is expected since, by definition, the values for Age do not vary much among subjects.

Now we applied the test statistic T_n^* to test the null hypothesis $H_0 : A\beta = 0$ against $H_1 : A\beta \neq 0$, where $A = (1, 0, 0, 0)$. This tests whether the risk factor, Age, is significant while retaining the effects of other factors as nuisance parameters. The realized value of T_n^* for this dataset was 6.785, and the corresponding p-value was calculated as 0.44. There is no enough evidence for rejecting the null.

6. Conclusion

In this paper we studied the efficient polynomial-spline estimation for the partly linear additive Cox model. We developed the bootstrap method to estimate the distribution of $\hat{\beta}$ and proposed the test statistic T_n and its bootstrap version to test linear hypotheses on the finite parameters. This facilitates the inference about the finite parameters in the model. The proposed methodology is easy to implement using the software R and is endorsed by the mathematical theory and simulations.

Further work in progress for the model includes hypothesis tests for the nonparametric components using the generalized likelihood ratio test in Fan, Zhang, and Zhang (2001) and selection of variables in the parametric part via the nonconcave penalized likelihood in Fan and Li (2002) and Li and Liang (2008), based on the efficient polynomial splines estimation. This, together with the method in this paper, provides a useful semiparametric inference tool for the model.

Acknowledgement

This work was supported, in part, by the NSF grant DMS-0906482 and by funds provided by the University of North Carolina at Charlotte. The authors thank the AE and two referees for helpful comments which has led to improvement of the paper.

Appendix I: Notations and Conditions

In addition to the condition for the smoothing parameter K_n in Section 2, the following conditions used in Huang (1999) are also needed for the theorems.

(A1) (i) The regression parameter β belongs to an open subset (not necessarily bounded) of R^d, and each ϕ_j lies in $\mathcal{A}$ for $j = 1, \ldots, J$, where $\mathcal{A}$ is the class of functions ϕ on $[0,1]$ whose kth derivative exists and satisfies the Lipschitz condition of order α:

$$|\phi^{(k)}(s) - \phi^{(k)}(t)| \leq C|s-t|^{\alpha} \text{ for } s, t \in [0,1],$$

where $\alpha \in (0,1]$ satisfies $p = k + \alpha > 0.5$.
(ii) $E(\delta \mathrm{X}) = 0$ and $E[\delta\phi_j(W_j)] = 0$, $1 \leq j \leq J$.

(A2) The failure time T and the censoring time C are conditionally independent given the covariate (X, W).

(A3) (i) Only the observations for which the event time S_i $(1 \leq i \leq n)$ is in a finite interval $[0,\tau]$, say, are used in the partial likelihood. The baseline cumulative hazard function $\Lambda_0(\tau) = \int_0^{\tau} \lambda_0(s)\, ds < \infty$. (ii) The covariate X takes values in a bounded subset of R^d, and the covariate W takes values in $\mathcal{W}$.

(A4) There exists a small positive constant ε such that (i) $P(\delta = 1|\mathrm{X}, \mathrm{W}) > \varepsilon$ and (ii) $P(C > \tau|\mathrm{X}, \mathrm{W}) > \varepsilon$ almost surely with respect to the probability measure of (X, W).

(A5) Let $0 < c_1 < c_2 < \infty$ be constants. The joint density $f(t, \mathrm{w}, \delta = 1)$ of $(S, \mathrm{W}, \delta = 1)$ satisfies $c_1 \le f(t, \mathrm{w}, \delta = 1) < c_2$ for all $(t, \mathrm{w}) \in [0, \tau] \times \mathcal{W}$.

(A6) Assume that the q-th partial derivative of the joint density $f(t, \mathrm{x}, \mathrm{w}, \delta = 1)$ of $(S, \mathrm{X}, \mathrm{W}, \delta = 1)$ with respect to t or w exists and is bounded. (For discrete covariate X, $f(t, \mathrm{x}, \mathrm{w}, \delta = 1)$ is defined to be $(\partial^2/\partial t \partial \mathrm{w})P(S \le t, \mathrm{X} = \mathrm{x}, W_1 \le w_1, \ldots, W_J \le w_J, \delta = 1)$.)

(A7) The information matrix $I(\beta)$ is positive definite.

Appendix II: Proof of Theorem

Proof of Theorem 1. The results are proven by drawing a parallel between the approximated partial likelihood and its bootstrap analogue.

Let $\omega_{ki} = 1(S_k^* = S_i)$, for $k, i = 1, \ldots, n$. That is, ω_{ki} equals 1 if $S_k^* = S_i$, and is 0 otherwise. Then

$$P(\omega_{ki} = 1|F_n) = \frac{1}{n}, \quad \text{and} \quad P(\omega_{ki} = 0|F_n) = 1 - \frac{1}{n},$$

where F_n is the empirical distribution of $\{S_i, \delta_i, \mathrm{W}_i, \mathrm{X}_i\}_{i=1}^n$. Given the bootstrap sample $\{S_i^*, \delta_i^*, \mathrm{W}_i^*, \mathrm{X}_i^*\}$, $i = 1, \ldots, n$, the logarithm of the approximated partial likelihood is

$$\ell^*(\beta, \mathrm{b}) = \sum_{i=1}^n \delta_i^* \Big\{\beta' \mathrm{X}_i^* + \phi_n(\mathrm{W}_i^*) - \log \sum_{k \in \mathcal{R}_i} \exp[\beta' \mathrm{X}_k^* + \phi_n(\mathrm{W}_k^*)]\Big\},$$

which can be written as

$$\ell^*(\beta, \mathrm{b}) = \sum_{i=1}^n \omega_i \delta_i \Big\{\beta' \mathrm{X}_i + \phi_n(\mathrm{W}_i) - \log \sum_{k \in \mathcal{R}_i} \omega_k \exp[\beta' \mathrm{X}_k + \phi_n(\mathrm{W}_k)]\Big\}, \quad \text{(A.1)}$$

where $\omega_i = \sum_{k=1}^n \omega_{ki}$ for $i = 1, \ldots, n$. This is just a random weighted version of the approximated partial likelihood in (2.2). Note that the bootstrap estimators $(\hat{\beta}^*, \hat{\mathrm{b}}^*)$ for (β, b) maximize the likelihood in (A.1). Following the argument for Theorem 3.3 in Huang (1999), we can obtain the Bahadur representation

$$\sqrt{n}(\hat{\beta}^* - \beta) = n^{-1/2} I(\beta)^{-1} \sum_{i=1}^n \omega_i l_\beta^*(S_i, \delta_i, \mathrm{Z}_i) + o_p(1). \quad \text{(A.2)}$$

Recalling the result of Huang's Theorem 3.3,

$$\sqrt{n}(\hat{\beta}-\beta) = n^{-1/2}I(\beta)^{-1}\sum_{i=1}^{n} l_\beta^*(S_i,\delta_i,\mathrm{Z}_i) + o_p(1), \tag{A.3}$$

we obtain from (A.2) and (A.3) that

$$\begin{aligned}\sqrt{n}(\hat{\beta}^*-\hat{\beta}) &= n^{-1/2}I(\beta)^{-1}\sum_{i=1}^{n}(\omega_i-1)l_\beta^*(S_i,\delta_i,\mathrm{Z}_i) + o_p(1)\\ &\equiv L_n + o_p(1).\end{aligned}$$

It suffices to show that

$$P\{L_n < t|F_n\} \to N(0, I(\beta)^{-1}), \quad a.s.. \tag{A.4}$$

Note that

$$\begin{aligned}L_n &= n^{-1/2}I(\beta)^{-1}\sum_{i=1}^{n}[\sum_{k=1}^{n}(\omega_{ki}-\frac{1}{n})]l_\beta^*(S_i,\delta_i,\mathrm{Z}_i)\\ &= n^{-1/2}I(\beta)^{-1}\sum_{k=1}^{n}\eta_{kn}, \end{aligned}\tag{A.5}$$

where $\eta_{kn} = \sum_{i=1}^{n}[(\omega_{ki}-1/n)]l_\beta^*(S_i,\delta_i,\mathrm{Z}_i)$. It is easy to see that conditional on F_n, the η_{kn}'s are iid random vectors with mean zero. Note that $E[(\omega_{ki}-1/n)^2|F_n] = 1/n-1/n^2$, and for $i \neq j$, $E[(\omega_{ki}-1/n)(\omega_{kj}-1/n)|F_n] = -1/n^2$, almost surely. Simple algebra gives that

$$\begin{aligned}E[\eta_{kn}\eta_{kn}'|F_n] &= n^{-1}\sum_{i=1}^{n}[l_\beta^*(S_i,\delta_i,\mathrm{Z}_i)]^2 - [n^{-1}\sum_{i=1}^{n}l_\beta^*(S_i,\delta_i,\mathrm{Z}_i)]^2\\ &\to I(\beta), \quad \text{almost surely.}\end{aligned}$$

Further, for any $\mathrm{a} \in R^d$ such that $\|\ \mathrm{a}\ \| = 1$, $E[\mathrm{a}'\eta_{kn}\eta_{kn}'\mathrm{a}|F_n]$ converges to $\mathrm{a}'I(\beta)\mathrm{a}$ almost surely. It is easy to verify that the following Lindberg condition holds:

$$\sum_{k=1}^{n}E[(\mathrm{a}'\eta_{kn})^2 1(|\mathrm{a}'\eta_{kn}| \geq \varepsilon)|F_n] \to 0, \quad \text{almost surely for any } \varepsilon > 0.$$

Therefore, conditional on F_n, $\mathrm{a}'\eta_{kn}$ is asymptotically normal. By the Cramér-Wold device, $\eta_{kn}|F_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, I(\beta))$. This combined with (A.5) leads to (A.4).

Proof of Theorem 2. By Theorem 3.3 of Huang (1999),

$$I(\beta)(\hat{\beta}-\beta) = n^{-1}\sum_{i=1}^{n}\ell_\beta^*(S_i,\delta_i,\mathrm{Z}_i) + o_p(n^{-1/2}). \tag{A.6}$$

Following the argument for (A.6), one obtains that, under H_0,

$$CI(\beta)C'(\hat{\gamma}-\gamma) = n^{-1}C\sum_{i=1}^{n}\ell_\beta^*(S_i,\delta_i,\mathrm{Z}_i) + o_p(n^{-1/2}).$$

Note that $CC' = I_{d-m}$. It follows that

$$C'(\hat{\gamma}-\gamma) = C'\{CI(\beta)C'\}^{-1}Cn^{-1}\sum_{i=1}^{n}\ell_\beta^*(S_i,\delta_i,\mathrm{Z}_i) + o_p(n^{-1/2}). \tag{A.7}$$

Under H_0, combining (A.6) and (A.7) leads to

$$\begin{aligned}\sqrt{n}(\hat{\beta}-C'\hat{\gamma}) &= [I^{-1}(\beta) - C'\{CI(\beta)C'\}^{-1}C]n^{-1/2}\sum_{i=1}^{n}\ell_\beta^*(S_i,\delta_i,\mathrm{Z}_i) + o_p(1)\\ &= V(\beta)n^{-1/2}\sum_{i=1}^{n}\ell_\beta^*(S_i,\delta_i,\mathrm{Z}_i) + o_p(1).\end{aligned}$$

Proof of Theorem 3. Let $\Theta = I(\beta)$ and

$$\Phi_n = n^{-1/2}\sum_{i=1}^{n}\ell_\beta^*(S_i,\delta_i,\mathrm{Z}_i).$$

Then under H_0, by Theorem 2(ii),

$$\sqrt{n}(\hat{\beta}-C'\hat{\gamma}) = \Theta^{-1/2}[I_d - \Theta^{1/2}C'\{C\Theta C'\}^{-1}C\Theta^{1/2}]\Theta^{-1/2}\Phi_n + o_p(1).$$

Since the matrix $V^*(\beta) = I_d - \Theta^{1/2}C'\{C\Theta C'\}^{-1}C\Theta^{1/2}$ is symmetrical and idempotent with rank m, by condition (A7), it can be rewritten as $D'D$, where D is a $m\times d$ matrix satisfying $DD' = I_m$. Let $\xi_n \equiv D\Theta^{-1/2}\Phi_n$. Then $\xi_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, I_m)$, and under H_0 we have $\sqrt{n}(\hat{\beta}-C'\hat{\gamma}) = \Theta^{-1/2}D'\xi_n + o_p(1)$. Therefore,

$$T_n = \xi_n' D\Theta^{-1}D'\xi_n + o_p(1). \tag{A.8}$$

By (A.8) and the asymptotic normality of ξ_n, the result holds.

Proof of Theorem 4. Using the same argument as for (A.2), we obtain that, under H_0,

$$CI(\beta)C'(\hat{\gamma}^*-\gamma) = n^{-1}C\sum_{i=1}^{n}w_i\ell_\beta^*(S_i,\delta,\mathrm{Z}_i) + o_p(n^{-1/2}),$$

that is,

$$\sqrt{n}C'(\hat{\gamma}^*-\gamma) = C'\{CI(\beta)C'\}^{-1}Cn^{-1/2}\sum_{i=1}^{n}w_i\ell_\beta^*(S_i,\delta,\mathrm{Z}_i) + o_p(1).$$

This combined with (A.3) yields that, under H_0,

$$\sqrt{n}(\hat{\beta}^* - C'\hat{\gamma}^*) = n^{-1/2}V(\beta)\sum_{i=1}^{n} w_i \ell_\beta^*(S_i, \delta, \mathrm{Z}_i) + o_p(1), \tag{A.9}$$

where $V(\beta)=[I^{-1}(\beta)–C'\{CI(\beta)C'\}^{-1}C]$. Let $\Phi_n^*=n^{-1/2}\sum_{i=1}^n(w_i–1)\ell_\beta^*(S_i,\delta,\mathrm{Z}_i)$ and $\xi_n^* = D\Theta^{-1/2}\Phi_n^*$. Then by (A.8) and (A.9),

$$\begin{aligned}&\sqrt{n}(\hat{\beta}^* - C'\hat{\gamma}^*) - \sqrt{n}(\hat{\beta} - C'\hat{\gamma})\\ &\quad = [I^{-1}(\beta) - C'\{CI(\beta)C'\}^{-1}C]\Phi_n^* + o_p(1).\end{aligned}$$

Therefore, $T_n^* = \xi_n^{*\prime} D\Theta^{-1}D'\xi_n^* + o_p(1)$. Conditional on F_n, the first term of the righthand side is asymptotically distributed as $\sum_{i=1}^m w_i\chi_{1,i}^2$. Then by Slutsky's Theorem, the same is true for T_n^*. This combined with Theorem 3 and the Polya Theorem completes the proof of the theorem.

Proof of Theorem 5. The approximated partial likelihood for γ is

$$\ell(\gamma, \mathrm{b}) = \sum_{i=1}^{n} \delta_i\Big\{\gamma' C\mathrm{X}_i + \phi_n(\mathrm{W}_i) - \log\sum_{k\in\mathcal{R}_i}\exp[\gamma' C\mathrm{X}_k + \phi_n(\mathrm{W}_k)]\Big\}.$$

Let $\mathrm{Z}_i^* = (C\mathrm{X}_i, \mathrm{W}_i)$ and the efficient score be $\ell_\gamma^*(S_i, \delta_i, \mathrm{Z}_i^*)$. Similar to (A.3), we have

$$\sqrt{n}(\hat{\gamma} - \gamma) = n^{-1/2}I(\gamma)^{-1}\sum_{i=1}^{n}\ell_\gamma^*(S_i, \delta_i, \mathrm{Z}_i^*) + o_p(1).$$

Note that by definition of the efficient score, $\ell_\gamma^*(S_i, \delta_i, \mathrm{Z}_i^*) = C\ell_\beta^*(S_i, \delta_i, \mathrm{Z}_i)$, and hence $I(\gamma) = E[\ell_\gamma^*(S, \delta, \mathrm{Z}^*)]^{\otimes 2} = CI(\beta)C'$. It follows that

$$\sqrt{n}C'(\hat{\gamma} - \gamma) = n^{-1/2}C\{CI(\beta)C'\}^{-1}C\sum_{i=1}^{n}\ell_\beta^*(S_i, \delta_i, \mathrm{Z}_i) + o_p(1)$$

which, together with (A.2), leads to

$$\sqrt{n}(\hat{\beta} - C'\hat{\gamma}) = \sqrt{n}(\beta - C'\gamma) + n^{-1/2}V(\beta)\sum_{i=1}^{n}\ell_\beta^*(S_i, \delta_i, \mathrm{Z}_i) + o_p(1).$$

Thus, under H_{1n},

$$\sqrt{n}(\hat{\beta} - C'\hat{\gamma}) = n^{\frac{1}{2}-r}\beta_n + n^{-1/2}V(\beta)\sum_{i=1}^{n}\ell_\beta^*(S_i, \delta_i, \mathrm{Z}_i) + o_p(1). \tag{A.10}$$

Since the second term on the righthand side is asymptotically normal with mean zero and covariance $\Sigma(\beta) = V(\beta)I(\beta)V'(\beta) = V(\beta)$,

$$\beta(\alpha, r) = P\{T_n < \chi_{n,1-\alpha}|H_{1n}\} \to 0,$$

only if $||n^{1/2-r}\beta_n|| \to \infty$, that is, $r < 1/2$.

Proof of Theorem 6..

(i) Note that $V(\beta)I(\beta)V(\beta) = V(\beta)$ and $\beta_n \to \beta^*$. The result is obvious from (A.10).

(ii) By (A.10), under H_{1n} with $r = 1/2$,

$$\sqrt{n}(\hat{\beta} - C'\hat{\gamma}) = \beta_n + n^{-1/2} V(\beta) \sum_{i=1}^{n} \ell_\beta^*(S_i, \delta_i, \mathrm{Z}_i) + o_p(1).$$

Similar to (), we obtain that $\sqrt{n}(\hat{\beta} - C'\hat{\gamma}) = \beta_n + \Theta^{-1/2} D'\xi_n + o_p(1)$, which converges weakly to $\mathcal{N}(\beta_n, V(\beta))$. Let $\eta_n = \Theta^{-1/2} D'\xi_n$. Then η_n converges weakly to $\eta \sim \mathcal{N}(0, V(\beta))$. Therefore, under H_{1n}, $T_n = \|\beta_n + \eta_n\|^2 + o_p(1)$. Hence, by the Continuous Mapping Theorem, $T_n \xrightarrow{\mathcal{L}} \|\beta^* + \eta\|^2$.

References

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. John Hopkins University Press.

Bilias, Y., Chen, S. and Ying, Z. (2000). Simple resampling methods for censored regression quantiles. *J. Econometrics* **99**, 373-386.

Breslow, N. E. (1972). Contribution to the discussion on the paper by D. R. Cox, "Regression and life tables". *J. Roy. Statist. Soc. Ser. B* **34**, 216-217.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89-99.

Burr, D. (1994). A comparison of certain bootstrap confidence interbals in the Cox model. *J. Amer. Statist. Assoc.* **89**, 1290-1302.

Cai, J., Fan, J., Jiang, J. and Zhou, H. (2007). Partially linear hazard regression for multivariate survival data. Jour. *J. Amer. Statist. Assoc.* **102**, 538-551.

Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

Dabrowska, D. M. (1997). Smoothed Cox regression. *Ann. Statist.* **25**, 1510-1540.

Dawber, T. R. (1980). *The Framingham Study, The Epidemilogy of Atheroscle rotic Disease.* Harvard University Press, Cambridge, MA.

de Boor, C. (1978). *A Practical Guide to B-splines*. Springer-Verlag, New York.

Fan, J. and Jiang, J. (2009). Non- and semi- parametric modeling in survival analysis, In *New Developments in Biostatistics and Bioinformatics - Frontiers of Statistics*, **1** (Edited by Jianqing Fan, Xinhong Lin and Jun Liu), 3-33. World Scientific & Higher Education Press, New Jersey.

Fan, J. and Li, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. *Ann. Statist.* **30**, 74-99.

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29**, 153-193.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Process and Survival Analysis*, Wiley, New York.

Grambsch, P. M., Therneau, T. M. and Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* **77**, 147-160.

Hastie, T. and Tibshirani, R. (1990). Exploring the nature of covariate effects in the proportional hazards model. *Biometrics* **51**, 1005-1016.

Heller, G. (2001). The Cox proportional hazards model with a partly linear relative risk function. *Lifetime Data Anal.* **7**, 255-277.

Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *Ann. Statist.* **27**, 1536-1563.

Li, R. and Liang, H. (2008). Variable selection in semiparametric regression modeling. *Ann. Statist.* **36**, 261-286.

Nan, B., Lin, X., Lisabeth, L. D. and Harlow, S. D. (2005). A varying-coeffient Cox model for the effct of age at a marker event on age at menopause. *Biometrics* **61**, 576-583.

Ø'Sullivan, F. (1988). Nonparametric estimation of relative risk using splines and cross-validation. *Siam J. Sci. Stat. Comput.* **9**, 531-542.

Sasieni, P. (1992). Information bounds for the conditional hazard ratio in a nested family of regression models. *J. Roy. Statist. Soc. Ser. B* **54**, 617-635.

Schumaker, L. (1981). *Spline Functions: Basic Theory,* Wiley, New York.

Stone, C. J. (1986). Comment on "Generalized Additive Models" by Hastie and Tibshirani. *Statist. Sci.* **1**, 312-314.

van der Vaart, A. W. (1991). On differentiable functionals. *Ann. Statist.* **19**, 178-204.

Zeng, D. and Lin, D. Y. (2008). Efficient resampling methods for nonsmooth estimating functions. *Biostatistics* **9**, 355-363.

Zhou, L. (2006). A simple censored median regression estimator. *Statist. Sinica* **16**, 1043-1058.

Zhou, S., Shen, X. and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.* **26**, 1760-1782.

Department of Mathematics and Statistics, University of North Carolina, Charlotte, NC 28223, USA.

E-mail: jjiang1@uncc.edu

School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan 430073, China.

Department of Statistics, The Chinese University of Hong Kong, Hong Kong, China.

E-mail: s063598@sta.cuhk.edu.hk

(Received November 2009; accepted January 2010)