

Efficient Estimation of the Partly Linear Additive Hazards Model with Current Status Data

XUEWEN LU

Department of Mathematics and Statistics, University of Calgary

PETER X.-K. SONG

Department of Biostatistics, University of Michigan

ABSTRACT. This paper focuses on efficient estimation, optimal rates of convergence and effective algorithms in the partly linear additive hazards regression model with current status data. We use polynomial splines to estimate both cumulative baseline hazard function with monotonicity constraint and nonparametric regression functions with no such constraint. We propose a simultaneous sieve maximum likelihood estimation for regression parameters and nuisance parameters and show that the resultant estimator of regression parameter vector is asymptotically normal and achieves the semiparametric information bound. In addition, we show that rates of convergence for the estimators of nonparametric functions are optimal. We implement the proposed estimation through a backfitting algorithm on generalized linear models. We conduct simulation studies to examine the finite-sample performance of the proposed estimation method and present an analysis of renal function recovery data for illustration.

Key words: backfitting, rate of convergence, semiparametric information bound, sieve method, splines

1. Introduction

In many applications, the response variable of interest is the time to occurrence of an event. When current status instead of exact timing of an event is observed, it results in the so-called current status data. For example, in a clinical study, suppose that T is time to occurrence of a certain event like tumour onset and $\mathbf{X}(t) = (X_1(t), \dots, X_d(t))'$ is a d -dimensional covariate process, which may be dependent on time t and observable up to t . Here, \mathbf{a}' denotes the transpose of a column vector \mathbf{a} . Owing to different practical reasons, such as study budget constraint, a subject may be examined at a random time C so that the actual response variable T is not observed directly; instead, a pair of observations (C, Δ) is yielded, where $\Delta = I[C \leq T]$ is an indicator of subjects' current status. Therefore, the available data for a subject are given by $(C, \Delta, \mathbf{X}(t) (t \leq C))$.

Lin & Ying (1994) considered an additive hazards regression (AHR) model in that the hazard function at time t , given the covariate value $\mathbf{x}(t)$ up to t , is assumed to take the following form:

$$\lambda(t|\mathbf{x}(t)) = \lambda_0(t) + \boldsymbol{\beta}'\mathbf{x}(t), \quad (1)$$

where $\boldsymbol{\beta}$ is a d -dimensional regression parameter vector and $\lambda_0(t)$ is an unknown baseline hazard function. To estimate $\boldsymbol{\beta}$, the covariate process $\mathbf{x}(\cdot)$ is usually restricted to only external covariates; see Kalbfleisch & Prentice (2002, section 6.3) for details.

Current status data are frequently encountered in practice; some real-world examples are given in Shiboski (1998) and Shiboski & Jewell (1992) and in one recent study on renal function recovery (Heung *et al.*, 2012). In the use of regression model to address the relationships

between event time and covariates, the (Cox, 1972) proportional hazards (PH) model is commonly the method of choice. For instance, for the Cox PH model with current status data, Huang (1996) studied the maximum likelihood estimation using a profile likelihood approach. Being formed in a different way, the AHR model (1) enables the characterization of different types of associations between event time and covariates, which sometimes are quite appealing for practitioners. It has been shown in some circumstances that the AHR model can be more plausible and interpretable than the Cox PH model (Lin & Ying, 1994; McKeague & Sasieni, 1994). However, such a model has not yet been widely used in practice largely because of its numerical difficulty until the breakthrough work of Lin *et al.* (1998). Their work showed a novel connection of an AHR model to an ordinary Cox PH model based on the following fact. That is, when a Cox PH model governs over the monitoring intensity, techniques of fitting the Cox PH model with right-censored data are readily applicable to the AHR model. This finding effectively alleviates the numerical difficulty in the AHR model, where any standard software with available numerical recipe for the Cox PH model may be directly employed. Later, Martinussen & Scheike (2002) developed a more efficient estimation method based on the efficient score estimating equations where a more flexible nonparametric model is proposed for the monitoring time C . Under some regularity conditions, their efficient estimator is verified to achieve the semiparametric information bound.

In this paper, we consider an extension of the AHR model to allow greater flexibility with nonparametric functions of covariates. In addition to the covariate process $\mathbf{X}(t)$, denote a J -dimensional vector by $\mathbf{W} = (W_1, \dots, W_J)'$ of baseline covariates. The conditional hazard function of event time given covariate values $(\mathbf{x}(\cdot)', \mathbf{w}') \in \mathbb{R}^d \times \mathbb{R}^J$ specified in a partly linear additive form is modelled as follows:

$$\lambda(t|\mathbf{x}(s), 0 \leq s \leq t, \mathbf{w}) = \lambda_0(t) + \boldsymbol{\beta}'\mathbf{x}(t) + \phi_1(w_1) + \dots + \phi_J(w_J), \tag{2}$$

where $\mathbf{x}(\cdot) = (x_1(\cdot), \dots, x_d(\cdot))'$, $\mathbf{w} = (w_1, \dots, w_J)'$, λ_0 is the unknown baseline hazard function, $\boldsymbol{\beta}$ is the d -dimensional regression parameter vector and ϕ_1, \dots, ϕ_J are unknown smooth functions. Alternatively, this model can be expressed in terms of the cumulative hazard function,

$$\Lambda(t|\mathbf{z}(s), 0 \leq s \leq t, \mathbf{w}) = \Lambda_0(t) + \boldsymbol{\beta}'\mathbf{z}(t) + t \left\{ \phi_1(w_1) + \dots + \phi_J(w_J) \right\}, \tag{3}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$ is the cumulative baseline hazard function and $\mathbf{z}(t) = \int_0^t \mathbf{x}(s) ds$.

In real applications, it is desirable for the model to accommodate some covariates, such as drug treatment or other categorical variables, in parametric forms and others in nonparametric forms when they are potentially nonlinearly associated with the hazard function. Then parameter $\boldsymbol{\beta}$ may be interpreted as covariate effects, adjusting for w_j 's effects nonparametrically. To the best of our knowledge, the proposed model (2) has not been systematically studied in the current literature. Our goals are to find efficient and optimal estimations for both the parametric and nonparametric components in model (2) and to provide a feasible computing algorithm to obtain numerical results.

According to Huang & Rossini (1997), in dealing with interval-censored data (including current status data as a special case), two features regarding the likelihood function make the related estimation different from estimations with other types of censored data. First, the likelihood of model (2) is a function of finite-dimensional parameter $\boldsymbol{\beta}$ and infinite-dimensional nuisance parameters $\Lambda_0(\cdot)$ and $\phi_j(\cdot)$, $j = 1, \dots, J$. Second, the parameter of interest $\boldsymbol{\beta}$ and the nuisance parameters $\Lambda_0(\cdot)$ and $\phi_j(\cdot)$'s need to be estimated simultaneously in order to achieve semiparametric efficiency for the estimator of $\boldsymbol{\beta}$. Because of interval censoring, we cannot

take the route of classical partial likelihood theory or as such to derive desirable asymptotic properties of the proposed simultaneous estimators. To overcome the difficulty, two promising approaches suggested in the literature include semiparametric maximum likelihood estimation (SPMLE) (e.g. Huang, 1996) and sieve maximum likelihood estimation (SMLE) (e.g. Rabinowitz *et al.*, 1995; Huang & Rossini, 1997). In particular, Huang & Rossini (1997) made a comparison between SMLE and SPMLE methods in the proportional odds model and showed that the SMLE of the baseline distribution converged faster than the SPMLE under certain smoothness conditions of the baseline distribution. For example, when the baseline distribution function has a second-order bounded derivative, the convergence rate of the SMLE can achieve an order of $O(n^{2/5})$, which is faster than the SPMLE's cubic rate $O(n^{1/3})$. In addition, the SMLE is also computationally easier than the SPMLE, because the sieve likelihood function typically involves fewer number of parameters than the semiparametric likelihood function.

With current status data, Xue *et al.* (2004) studied a partly linear accelerated failure time model, which contains only one nonparametric component, where the SMLE method was used for estimation. Regarding model (2), one similar but not directly related work is the partly linear transformation model considered by Ma & Kosorok (2005a). Another work by Ma (2011) focused on the partly linear AHR cure model, which included only one effectively nonparametric covariate. Ma & Kosorok (2005a) considered penalized log-likelihood estimation for the partly linear transformation (PLT) model. In their PLT model, there are two unknown functions: one is the nondecreasing transformation function, which is estimated by the cumulative sum diagram approach of Groeneboom & Wellner (1992), and the other is the smooth regression function, which is estimated by a sieve method. It is shown that convergence rates of these two nonparametric estimators are dominated by the rate of the SPMLE, namely $O(n^{1/3})$. Ma (2011) obtained estimators with similar properties.

According to Ma & Kosorok (2005a) and Ma (2011), both estimators $\hat{\Lambda}_n(t)$ and $\hat{\phi}_n(\mathbf{w})$ (see (8) and (7)) of the cumulative baseline hazard function and the nonparametric function of covariate can only achieve the cubic convergence rate, which is not optimal. This is regarded as a consequence of not using the sieve approach to estimate $\Lambda_0(t)$, instead using the nonparametric maximum likelihood estimation approach based on isotonic regression, which involves more unknown parameters than the sieve approach and hence requires a larger sample size. In addition, most of the existing models are limited to include only one nonparametric function of covariate. Including more than one nonlinear function of covariate in the model is clearly needed. Moreover, to understand theoretical properties of estimators and to implement them in computation are our primary tasks in this paper.

Our overarching goal concerning estimation and inference in model (2) with current status data is threefold: first, to establish an efficient estimator of β , the regression parameter of possibly time-dependent linear covariates; second, to develop nonparametric estimators for functions Λ_0 and ϕ_j 's with faster convergence rates; and third, to provide and test effective algorithms that allow fast computation of the proposed estimators. We adopt the sieve likelihood and empirical process approaches to achieve the preceding three objectives. Our contributions are beneficial to both methodological and practical studies, because modelling flexibility and improved efficiencies in estimation and computation are very appealing. Furthermore, smoothed monotonic estimator of $\Lambda_0(t)$ gives rise to a better interpretation than a non-smoothed piecewise constant function obtained by the SPMLE.

Our strategy of deriving efficient estimation and faster convergence rates is to invoke the SMLE method for the estimation of nuisance nonparametric parameters in model (2). To do so, we plan to construct two sieve spaces. One sieve space for Λ_0 is constructed on the basis of nonnegative monotone polynomial splines, and the other space for individual ϕ_j is generated by the basis of unconstrained polynomial splines. The SMLE method is then applied to

simultaneously estimate all parameters. We show that the resulting estimator of β is asymptotically normal and semiparametrically efficient. We also show that the convergence rates of the resulting sieve estimators for the nonparametric functions are of the order $p/(1 + 2p)$, provided that these unknown functions have continuous and bounded p th-order derivatives on certain finite intervals.

The paper is organized as follows. In Section 2, we introduce polynomial splines and articulate model assumptions. In Section 3, we present our main results. Section 4 concerns computing algorithms to implement the proposed estimation procedure, and Section 5 discusses inference methods. Section 6 presents a simulation study to evaluate finite-sample performance of the proposed methods. Section 7 is devoted to an application regarding the analysis of renal function recovery data. We conclude the paper with a discussion in Section 8. All the proofs are relegated to the appendix and the additional supporting information in the supplementary material.

2. Definitions and assumptions

Let $\phi(\mathbf{w}) = \phi_1(w_1) + \dots + \phi_J(w_J)$ and $\mathbf{v} = (c, \mathbf{z}(c), \mathbf{w})$. Denote the regression function by $g(\mathbf{v}) = \Lambda(c) + \beta' \mathbf{z}(c) + c\phi(\mathbf{w})$. Let $(\beta_0, \Lambda_0, \phi_0)$ represent the set of true parameter values, where $\phi_0(\mathbf{w}) = \phi_{01}(w_1) + \dots + \phi_{0J}(w_J)$. Denote $g_0(\mathbf{v}) = \Lambda_0(c) + \beta_0' \mathbf{z}(c) + c\phi_0(\mathbf{w})$. The observable random variables are $(\Delta, \mathbf{V}) \in \{0, 1\} \times \mathbb{R}^+ \times \mathbb{R}^{d+J}$, where $\mathbf{V} = (C, \mathbf{Z}(C), \mathbf{W})$. Assume that C and T are conditionally independent given the covariate history $(\mathbf{Z}(\cdot), \mathbf{W})$ up to time C . Let (Δ_i, \mathbf{V}_i) , $i = 1, \dots, n$, be an independent and identically distributed (i.i.d.) random sample drawn from the distribution of (Δ, \mathbf{V}) .

We construct two sieve spaces of polynomial splines, one for ϕ and the other for Λ , respectively. Assume that \mathbf{W} takes values in $[a, b]^J$, where a and b are finite numbers. Let $a = \xi_0 < \xi_1 < \dots < \xi_K < \xi_{K+1} = b$ be a set of knots that partition $[a, b]$ into $(K + 1)$ sub-intervals, with $I_{K_i} = [\xi_i, \xi_{i+1})$, $i = 0, \dots, K - 1$ and $I_{K_K} = [\xi_K, \xi_{K+1}]$, where $K \equiv K_n = O(n^\nu)$ is a sample size-dependent positive integer such that $\max_{1 \leq i \leq K+1} (\xi_i - \xi_{i-1}) = O(n^{-\nu})$, $0 < \nu < 0.5$. The reason for postulating an upper bound of 0.5 on ν is that, according to Stone (1985), the optimal rate of convergence of a nonparametric estimator in an L^2 -norm is typically of the form $n^{-p/(2p+1)}$, $p \geq 1$, which is achieved at $\nu = 1/(2p + 1)$. This implies that $0 < \nu \leq 1/(2p + 1) < 0.5$, namely the upper bound of ν will not exceed 0.5. Let \mathcal{S}_n be the space of polynomial splines of order $\rho \geq 1$, in which a functional element s satisfies the following: (i) s is a ρ th-order polynomial on interval I_{K_i} for $0 \leq i \leq K$, and (ii) s is r times continuously differentiable on $[a, b]$ for $\rho \geq 2$ and $0 \leq r \leq \rho - 2$.

Let Φ_n be a collection of functions ϕ on $[a, b]^J$ with additive form $\phi(\mathbf{w}) = \phi_1(w_1) + \dots + \phi_J(w_J)$, with $\phi_j \in \mathcal{S}_n$, $j = 1, \dots, J$. According to Schumaker (1981, p. 117, Corollary 4.10), there exists a local basis $\{B_k, 1 \leq k \leq q_n\}$ for \mathcal{S}_n , where B_k are normalized B-spline basis functions and $q_n = K_n + \rho$. Thus, for any $\phi_j \in \mathcal{S}_n$, we can write

$$\phi_j(w_j) = \sum_{k=1}^{q_n} \gamma_{jk} B_k(w_j), \quad 1 \leq j \leq J. \tag{4}$$

Suppose that the support of variable C is a finite interval, say $[l_c, u_c]$. Then a partition of this interval can be made such that $\max_{1 \leq k \leq N+1} (t_k - t_{k-1}) = O(n^{-\nu})$, where t_k are knots, $k = 0, 1, \dots, N + 1$, with $t_0 = l_c$ and $t_{N+1} = u_c$, $N \equiv N_n = O(n^\nu)$. Let \mathcal{L}_n be the space of polynomial splines similar to \mathcal{S}_n , on the basis of knots t_0, t_1, \dots, t_N and sub-intervals, $I_{N_k} = [t_k, t_{k+1})$, $k = 0, \dots, N - 1$ and $I_{N_N} = [t_N, t_{N+1}]$.

Let \mathbb{L}_n be the collection of nonnegative and nondecreasing functions $\Lambda \in \mathcal{L}_n$ on $[l_c, u_c]$. Then for any $\Lambda \in \mathbb{L}_n$, we can write

$$\Lambda(t) = \sum_{k=1}^{r_n} \alpha_k L_k(t), \tag{5}$$

where $L_k(t)$ are normalized B-spline basis functions and $r_n = N_n + \rho$. Because $\Lambda(t)$ is non-negative and monotonically nondecreasing, the coefficients in (5) are subject to a constraint $0 \leq \alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_{r_n}$. By theorem 5.9 of Schumaker (1981), both nonnegativity and monotonicity of $\Lambda(t)$ are guaranteed by such a constraint on the coefficients. This class of monotone B-splines has been used by Lu *et al.* (2007) to approximate the logarithm of baseline mean function. This approach renders an SMLE method analytically and computationally simpler. Because the number of basis B-splines, r_n , is often taken much smaller than the sample size n , the curse of dimensionality is greatly alleviated, and the resulting computational burden is less than that of SPMLE.

Using the preceding two sieve spaces, we can proceed with the sieve-based maximum likelihood estimators of $(\boldsymbol{\beta}, \Lambda, \phi)$ as follows. Denote $\boldsymbol{\kappa} = (\boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\alpha}')'$. Let $\hat{\boldsymbol{\kappa}}_n = (\hat{\boldsymbol{\beta}}_n', \hat{\boldsymbol{\gamma}}_n', \hat{\boldsymbol{\alpha}}_n')'$ with $\hat{\boldsymbol{\gamma}}_n = \{\hat{\gamma}_{jk} : 1 \leq j \leq J, 1 \leq k \leq q_n\}$ and $\hat{\boldsymbol{\alpha}}_n = \{\hat{\alpha}_i : 1 \leq i \leq r_n, 0 \leq \hat{\alpha}_1 \leq \hat{\alpha}_2 \leq \dots \leq \hat{\alpha}_{r_n}\}$ be the values that maximize

$$l_n(\boldsymbol{\beta}, \phi, \Lambda) = \sum_{i=1}^n [\Delta_i \log \{\exp \{-\Lambda(C_i) - \boldsymbol{\beta}'\mathbf{Z}_i(C_i) - C_i \phi(\mathbf{W}_i)\}\} + (1 - \Delta_i) \log \{1 - \exp \{-\Lambda(C_i) - \boldsymbol{\beta}'\mathbf{Z}_i(C_i) - C_i \phi(\mathbf{W}_i)\}\}], \tag{6}$$

with $\phi(\mathbf{W}_i) = \sum_{j=1}^J \phi_j(W_{ji}), 1 \leq i \leq n$, where ϕ_j and Λ are given in (4) and (5), respectively. For $\mathbf{w} = (w_1, \dots, w_J)'$, the resulting estimators of ϕ and Λ are given by

$$\begin{aligned} \hat{\phi}_n(\mathbf{w}) &= \sum_{j=1}^J \hat{\phi}_{jn}(w_j) \\ &= \sum_{j=1}^J \left[\sum_{k=1}^{q_n} \hat{\gamma}_{jk} B_k(w_j) - \left\{ n^{-1} \sum_{i=1}^n \sum_{k=1}^{q_n} \hat{\gamma}_{jk} B_k(W_{ji}) \right\} \right] \end{aligned} \tag{7}$$

and

$$\hat{\Lambda}_n(t) = \sum_{k=1}^{r_n} \hat{\alpha}_k L_k(t), \tag{8}$$

where each $\hat{\phi}_{jn}$ is centred such that $n^{-1} \sum_{i=1}^n \hat{\phi}_{jn}(W_{ji}) = 0$. The following conditions are assumed in order to establish large-sample properties of the preceding estimators.

- (B1) Suppose the support of C is interval $S[C] = [l_c, u_c]$ and $0 < l_c < u_c < \infty$. Let p be an integer, $p \geq 1$. Let \mathcal{A} and \mathcal{L} be the classes of functions ϕ on $[a, b]$ and Λ on $[l_c, u_c]$, respectively, whose p th derivatives $\phi^{(p)}$ and $\Lambda^{(p)}$ exist and are continuous. Assume the following: (i) the true regression parameter value $\boldsymbol{\beta}_0$ is an interior point of a bounded parameter space $\mathcal{B} \subset \mathbb{R}^d$, the true functions $\Lambda_0 \in \mathcal{L}$ and each $\phi_{0j} \in \mathcal{A}$ for $1 \leq j \leq J$; and (ii) in the rest of the paper, $\mathbf{Z} = \mathbf{Z}(C)$, unless otherwise noticed. For any $c \in [l_c, u_c]$, $E(\mathbf{Z}|C = c) = 0$ and $E[\phi_{0j}(W_j)] = 0, 1 \leq j \leq J$. Moreover, for any $\boldsymbol{\beta} \in \mathcal{B}$ and $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, the probability $P(\boldsymbol{\beta}'\mathbf{Z} \neq \boldsymbol{\beta}_0'\mathbf{Z}|C) > 0$.

- (B2) Denote the conditional hazard of C given \mathbf{W} and $\mathbf{X}(\cdot)$ by $\sigma(t) = \sigma(t, \mathbf{X}(t), \mathbf{W})$. Assume the following: (i) the distribution of C is absolutely continuous; and (ii) event time T and censoring time C are conditionally independent given $(\mathbf{X}(t), \mathbf{W})$ or $(\mathbf{Z}(t), \mathbf{W})$ ($t \leq C$), the covariate history up to C .
- (B3) (i) Assume that the distribution of T, F_0 , is continuous. The support of $C, S[C]$, is strictly contained in the support of F_0 , that is, $0 < l_c < u_c < \tau_{F_0}, \tau_{F_0} = \inf\{t : F_0(t) = 1\}$. (ii) The covariate $\mathbf{Z}(t)$ takes values in a bounded subset of \mathbb{R}^d for t on $S(C)$, and the covariate \mathbf{W} takes values in $[a, b]^J$.
- (B4) Suppose there exists a small positive constant ε such that $P(T < l_c | C, \mathbf{Z}, \mathbf{W}) > \varepsilon$ and $P(T > u_c | C, \mathbf{Z}, \mathbf{W}) > \varepsilon$ in probability 1 with respect to the probability measure of $(C, \mathbf{Z}, \mathbf{W})$.
- (B5) Suppose the underlying parameter function of $(\beta_0, \Lambda_0, \phi_0)$ satisfies $g_0(\mathbf{V}) = \Lambda_0(C) + \beta_0' \mathbf{Z} + C\phi_0(\mathbf{W}) > 0$ with probability 1.
- (B6) Let $q \geq 1$ be a positive integer. For $1 \leq j \leq J$, the q th partial derivative of the joint density $f(c, \mathbf{z}, \mathbf{w})$ of $(C, \mathbf{Z}, \mathbf{W})$ with respect to c or w_j exists and is bounded.

Similar conditions to those listed in the preceding text have been considered in the literature (e.g. Huang, 1996), and they are in fact quite reasonable. Condition (B1)(i) requires certain smoothness characteristics on Λ and ϕ_j 's in order to apply splines-based smoothing techniques. (B1)(ii) ensures identifiability of the parameters. Condition (B2) implies noninformative censoring mechanism. (B3)(i) bounds likelihood and score functions away from infinity at the boundaries of the support of the observed event time. The boundedness condition on the covariates in (B3)(ii) is an unpleasant restriction, but in many practical situations, continuous covariates may be typically rescaled to fall between 0 and 1. Condition (B4) ensures that the probability of being either left censored or right censored is positive and bounded away from zero regardless of the covariate values. Condition (B5) requires the true cumulative hazard function to be a positive function. Condition (B6), a key assumption in Huang (1999) as well, implies that the partial score functions (or partial derivatives) of the nonparametric components in the least favourable direction are nearly zero, so that the root- n convergence rate and asymptotic normality of the finite-dimensional estimator can be obtained.

3. Theory of estimation

Consider the following two counting processes,

$$N_1(t) = \Delta I[C \leq t] \text{ and } N_2(t) = (1 - \Delta)I[C \leq t].$$

Let $N(t) = N_1(t) + N_2(t)$, $Y(t) = I[t \leq C]$ and $p(t) = p(t, \mathbf{z}(t), \mathbf{w}) = P(T \geq C | C = t, \mathbf{z}(t), \mathbf{w})$. By (3), we have

$$p(t) = \exp[-\{\Lambda_0(t) + \beta' \mathbf{z}(t) + t \{\phi_1(w_1) + \dots + \phi_J(w_J)\}\}]. \tag{9}$$

The two compensated counting processes are given by

$$M_1(t) = N_1(t) - \int_0^t Y(s)\sigma(s)p(s) ds, \tag{10}$$

$$M_2(t) = N_2(t) - \int_0^t Y(s)\sigma(s)(1 - p(s)) ds,$$

both of which are known as martingales by Martinussen & Scheike (2002).

Let $\|\cdot\|$ denote the Euclidean norm; let $\|\cdot\|_2$ denote the L_2 -norm with respect to a probability measure; and let $\|\cdot\|_\infty$ denote the supremum norm. For example, $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| = \sqrt{\sum_{i=1}^d (\beta_i - \beta_{0i})^2}$ for $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)'$ and $\boldsymbol{\beta}_0 = (\beta_{01}, \dots, \beta_{0d})'$; $\|\Lambda(C) - \Lambda_0(C)\| = |\Lambda(C) - \Lambda_0(C)|$; $\|\Lambda(C) - \Lambda_0(C)\|_2 = \sqrt{\int_{l_c}^{u_c} \{\Lambda(c) - \Lambda_0(c)\}^2 dF_C(c)}$, where $F_C(\cdot)$ is the distribution function of C ; and $\|\phi_j(w_j) - \phi_{0j}(w_j)\|_\infty = \sup_{w_j \in [a, b]} |\phi_j(w_j) - \phi_{0j}(w_j)|$. Our main results are stated in the following three theorems. The proofs of these theorems are presented in detail in the Appendix. Several technical lemmas are needed to prove these theorems; their proofs are relegated to the supplementary material.

Theorem 1. Under conditions (B1)–(B5), the efficient score for estimation of $\boldsymbol{\beta}$ in model (2) is

$$l_{\boldsymbol{\beta}}^*(\Delta, \mathbf{V}) = \int_{l_c}^{u_c} t \{ \tilde{\mathbf{Z}}(t) - \mathbf{a}^*(t) - \mathbf{h}^*(\mathbf{W}) \} \left\{ \frac{p(t, \mathbf{Z}(t), \mathbf{W})}{1 - p(t, \mathbf{Z}(t), \mathbf{W})} dM_2(t) - dM_1(t) \right\},$$

where $\tilde{\mathbf{Z}}(t) = \mathbf{Z}(t)/t$, $\mathbf{h}^*(\mathbf{w}) = \mathbf{h}_1^*(w_1) + \dots + \mathbf{h}_J^*(w_J)$ and components of $(\mathbf{a}^*, \mathbf{h}_1^*, \dots, \mathbf{h}_J^*)$ are the L_2 -norm functions that minimize

$$E \left[C^2 \left\| \tilde{\mathbf{Z}}(C) - \mathbf{a}^*(C) - \mathbf{h}^*(\mathbf{W}) \right\|^2 \frac{p(C, \mathbf{Z}(C), \mathbf{W})}{1 - p(C, \mathbf{Z}(C), \mathbf{W})} \right].$$

The information bound for estimation of $\boldsymbol{\beta}$ is

$$\begin{aligned} I(\boldsymbol{\beta}) &= E \left[l_{\boldsymbol{\beta}}^*(\Delta, \mathbf{V})^{\otimes 2} \right] \\ &= E \left[\left\langle \int_{l_c}^{u_c} t \{ \tilde{\mathbf{Z}}(t) - \mathbf{a}^*(t) - \mathbf{h}^*(\mathbf{W}) \} \left\{ \frac{p(t, \mathbf{Z}(t), \mathbf{W})}{1 - p(t, \mathbf{Z}(t), \mathbf{W})} dM_2(t) - dM_1(t) \right\} \right\rangle \right] \\ &= E \left[\int_{l_c}^{u_c} t^2 \{ \tilde{\mathbf{Z}}(t) - \mathbf{a}^*(t) - \mathbf{h}^*(\mathbf{W}) \}^{\otimes 2} \frac{p(t, \mathbf{Z}(t), \mathbf{W})}{1 - p(t, \mathbf{Z}(t), \mathbf{W})} Y(t) \sigma(t) dt \right] \\ &= E \left[C^2 \{ \tilde{\mathbf{Z}}(C) - \mathbf{a}^*(C) - \mathbf{h}^*(\mathbf{W}) \}^{\otimes 2} \frac{p(C, \mathbf{Z}(C), \mathbf{W})}{1 - p(C, \mathbf{Z}(C), \mathbf{W})} \right], \end{aligned} \tag{11}$$

where $\mathbf{a}^{\otimes 2} = \mathbf{a} \otimes \mathbf{a}'$ denotes the out-product of a column vector \mathbf{a} and $\langle M \rangle$ denotes the predictable variation process of a martingale $M(t)$.

Theorem 2. Suppose that conditions (B1)–(B5) hold and $0 < \nu < 0.5$. For the estimators defined in (6)–(8), let $\hat{\mathbf{g}}_n(\mathbf{v}) = \hat{\Lambda}_n(c) + \hat{\boldsymbol{\beta}}_n' \mathbf{z}(c) + c \hat{\phi}_n(\mathbf{w})$. Then

$$\begin{aligned} \|\hat{\mathbf{g}}_n - \mathbf{g}_0\|_2^2 &= E \left[\hat{\Lambda}_n(C) + \hat{\boldsymbol{\beta}}_n' \mathbf{Z}(C) + C \hat{\phi}_n(\mathbf{W}) - \{ \Lambda_0(C) + \boldsymbol{\beta}_0' \mathbf{Z}(C) + C \phi_0(\mathbf{W}) \} \right]^2 \\ &= O_p \left(n^{-(1-\nu)} + n^{-2\nu p} \right), \end{aligned}$$

where the expectation is taken under the distribution of $\mathbf{V} = (C, \mathbf{Z}(C), \mathbf{W})$ conditional on the given sample. Furthermore, if $I(\boldsymbol{\beta}_0)$ in (11) is nonsingular, then

$$\begin{aligned} \|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|^2 &= O_p \left(n^{-(1-\nu)} + n^{-2\nu p} \right), \\ \|\hat{\Lambda}_n - \Lambda_0\|_2^2 &= O_p \left(n^{-(1-\nu)} + n^{-2\nu p} \right), \\ \|\hat{\phi}_{jn} - \phi_{0j}\|_2^2 &= O_p \left(n^{-(1-\nu)} + n^{-2\nu p} \right), \quad 1 \leq j \leq J. \end{aligned}$$

Theorem 3. Suppose that conditions (B1)–(B6) hold and $I(\beta_0)$ in (11) is nonsingular. If ν satisfies the restrictions of $0.25/p < \nu < 0.5$ and $\nu(p + q) > 0.5$, where p and q are orders of differentiation postulated in (B1) and (B6), respectively, then

$$\begin{aligned} \sqrt{n}(\hat{\beta}_n - \beta_0) &= n^{-1/2} I^{-1}(\beta_0) \sum_{i=1}^n I_{\beta_0}^*(\Delta_i, \mathbf{V}_i) + o_p(1) \\ &\xrightarrow{d} N(0, \Sigma(\beta_0)), \end{aligned}$$

where \xrightarrow{d} denotes convergence in distribution $\Sigma(\beta_0) = I^{-1}(\beta_0)$.

In the estimation of Λ_0 and ϕ_{0j} 's, the optimal choice of ν satisfying the restrictions stated in theorem 3 is $\nu = 1/(1 + 2p)$, which leads to $n^{\nu/(1+2p)}$ as the convergence rate of $\hat{\Lambda}_n$ and $\hat{\phi}_n$ by theorem 2. Note that this is indeed the optimal rate in nonparametric regression and regression splines (Stone, 1982; Claeskens *et al.*, 2009). Especially for Λ_0 , when $p \geq 2$, the convergence rate of $\hat{\Lambda}_n$ is faster than the order of $n^{1/3}$, the rate of the SPMLE obtained by Huang (1996), for the estimation of the cumulative hazard function in the Cox model with current status data as well as by Ma & Kosorok (2005a) for the estimation of the transformation function in the partly linear transformation models. Our proposed estimator $\hat{\beta}_n$ achieves both optimal parametric convergence rate $n^{1/2}$ and semiparametric efficiency bound.

4. Implementation

Noticing that from (9)

$$\log\{p(C)\} = -[\Lambda(C) + \beta'Z(C) + C\{\phi_1(W_1) + \dots + \phi_J(W_J)\}],$$

we may regard model (2) with current status data as a generalized partly linear additive model for binary responses with log-link and predictor $-\zeta$, where $\zeta = \Lambda(C) + \beta'Z(C) + C\{\phi_1(W_1) + \dots + \phi_J(W_J)\}$.

In the log-likelihood (6), let $\mathbf{L}_i = (L_1(C_i), \dots, L_{r_n}(C_i))'$, $\mathbf{B}_{ji} = (B_1(W_{ji}), \dots, B_{q_n}(W_{ji}))'$, $\mathbf{B}_i = (\mathbf{B}'_{1i}, \dots, \mathbf{B}'_{Ji})'$, $\mathbf{Z}_i = \mathbf{Z}_i(C_i)$, $\alpha = (\alpha_1, \dots, \alpha_{r_n})'$, $\gamma_j = (\gamma_{j1}, \dots, \gamma_{jq_n})'$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}'_1, \dots, \boldsymbol{\gamma}'_J)'$, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}'_J)'$ and $\mathbf{ZB}_i = (\mathbf{Z}'_i, (C_i * \mathbf{B}_i)')'$. Through the B-spline smoothing technique, the predictor becomes a linear predictor of the following form,

$$\begin{aligned} \zeta_i &= \sum_{k=1}^{r_n} \alpha_k L_k(C_i) + \boldsymbol{\beta}'\mathbf{Z}_i + C_i \sum_{j=1}^J \sum_{k=1}^{q_n} \gamma_{jk} B_k(W_{ji}) \\ &= \boldsymbol{\alpha}'\mathbf{L}_i + \boldsymbol{\beta}'\mathbf{Z}_i + C_i \sum_{j=1}^J \boldsymbol{\gamma}'_j \mathbf{B}_{ji} \\ &= \mathbf{L}'_i \boldsymbol{\alpha} + \mathbf{Z}'_i \boldsymbol{\beta} + C_i \mathbf{B}'_i \boldsymbol{\gamma} \\ &= \mathbf{L}'_i \boldsymbol{\alpha} + \mathbf{ZB}'_i \boldsymbol{\theta}, \end{aligned} \tag{12}$$

and $p_i = p_i(C_i) = \exp(-\zeta_i)$. It follows that the score equations for $(\boldsymbol{\alpha}, \boldsymbol{\theta})$ can be written as

$$\mathbf{L}'\mathbf{u} = 0 \text{ and } \mathbf{ZB}'\mathbf{u} = 0, \tag{13}$$

where $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_n)'$, $\mathbf{L} = \partial\boldsymbol{\zeta}/\partial\boldsymbol{\alpha} = (\mathbf{L}_1, \mathbf{L}_2, \dots, \mathbf{L}_n)'$ and $\mathbf{ZB} = \partial\boldsymbol{\zeta}/\partial\boldsymbol{\theta} = (\mathbf{ZB}_1, \dots, \mathbf{ZB}_n)'$, respectively, are $n \times r_n$ and $n \times (d + J * q_n)$ matrices and $\mathbf{u} = \partial l_n / \partial \boldsymbol{\zeta}$ is an $n \times 1$ vector with elements

$$u_i = \frac{(\Delta_i - p_i)dp_i/d\xi_i}{p_i(1 - p_i)} = \frac{-(\Delta_i - p_i)}{1 - p_i}.$$

The score equations (13) can be solved separately for α and θ using a modified version of the Newton scoring algorithm for generalized linear models. Note that the two equations in (13) are subject to the monotonicity constraint given in (5). The algorithm iterates the following steps, starting from $k = 0$ with the initial estimates $\alpha_{(0)}$ and $\theta_{(0)}$ of α and θ :

- Step k . Given estimates $\alpha_{(k)}$ and $\theta_{(k)}$, calculate \mathbf{u} and the weight matrix $\Omega = \text{Var}(\mathbf{u}) = \text{diag}(\omega_1, \dots, \omega_n)$, with weights

$$\omega_i = \frac{(dp_i/d\xi_i)^2}{p_i(1 - p_i)} = \frac{p_i}{1 - p_i},$$

then construct the adjusted dependent variable:

$$\mathbf{y} = \mathbf{L}\alpha_{(k)} + \mathbf{ZB}\theta_{(k)} + \Omega^{-1}\mathbf{u}.$$

- Step $(k + 1)$. Update $\theta_{(k)}$ by optimizing the following weighted least squares function:

$$\min_{\theta} \left\{ (\mathbf{y} - \mathbf{L}\alpha_{(k)} - \mathbf{ZB}\theta)' \Omega (\mathbf{y} - \mathbf{L}\alpha_{(k)} - \mathbf{ZB}\theta) \right\}.$$

Denote the minimizer by $\theta_{(k+1)}$, update $\alpha_{(k)}$ by solving the following weighted isotonic regression:

$$\min_{\alpha} \left\{ (\mathbf{y} - \mathbf{L}\alpha - \mathbf{ZB}\theta_{(k+1)})' \Omega (\mathbf{y} - \mathbf{L}\alpha - \mathbf{ZB}\theta_{(k+1)}) \right\},$$

subject to $0 \leq \alpha_1 \leq \dots \leq \alpha_{r_n}$. Denote the minimizer by $\alpha_{(k+1)}$.

Cycle these two steps until a certain convergence criterion is satisfied. This procedure is known as backfitting (Hastie *et al.*, 2008, Chapter 9), a modification of the Gauss–Seidel procedure for solving simultaneous linear equations (Thisted, 1988). Our model can be viewed as a special case of the generalized additive models (GAM) discussed by (Hastie *et al.*, 2008), in which the outcome is a binary response Δ and covariates are $(C, \mathbf{Z}(C), \mathbf{W})$. Different from a classical GAM, here, we treat the monitoring time C as a nonlinear covariate, its functional form is subject to a monotonic non-decreasing constraint. Hence, we cannot apply the classical GAM directly. Instead we consider the approach of backfitting. That is, in step $(k + 1)$, our algorithm proceeds to optimize the weighted least squares function for unconstrained parameters and to run the weighted isotonic regression for constrained parameters iteratively. The idea of utilizing adjusted dependent variable \mathbf{y} in step k originates from the numerical recipe in generalized linear models. Using the B-splines, we approximate a nonlinear predictor by a linear predictor ζ_i in (12), so that techniques from the generalized linear models can be readily applied, including the iterative weighted least squares method with \mathbf{y} as a working response. To be precise, in the expression of \mathbf{y} , we treat $\mathbf{L}\alpha_{(k)} + \mathbf{ZB}\theta_{(k)}$ as a linear predictor with a design matrix $(\mathbf{L}, \mathbf{ZB})$ and a regression parameter vector $(\alpha'_{(k)}, \theta'_{(k)})'$ and $\Omega^{-1}\mathbf{u}$ as an error term, whose variance–covariance is given by

$$\text{Var}(\Omega^{-1}\mathbf{u}) = \text{Var} \left(\left(\omega_1^{-1}u_1, \omega_2^{-1}u_2, \dots, \omega_n^{-1}u_n \right)' \right) = \Omega^{-1}.$$

Therefore, the estimation procedure becomes solving a standard iterative weighted least squares problem. The weighted isotonic regression implemented in step $(k + 1)$ is an extension of the theory of isotonic regression discussed by Robertson *et al.* (1988, Chapter 1), where a

nonparametric estimate in the form of a step function is presented. Numerically, such an estimate is obtained by using the pool adjacent violators algorithm, which essentially minimizes a weighted sum of squares. In our approach, we reparameterize α_k in (5) to $\tilde{\alpha}_k$ and then rewrite $\Lambda(t)$ as

$$\Lambda(t) = \sum_{k=1}^{r_n} \tilde{\alpha}_k \tilde{L}_k(t),$$

where the coefficients are subject to the constraints $\tilde{\alpha}_1 = \alpha_1 \geq 0$, $\tilde{\alpha}_k = \alpha_k - \alpha_{k-1} \geq 0$, for $k = 2, \dots, r_n$, and $\tilde{L}_k(t) = \sum_{i=k}^{r_n} L_i(t)$, for $k = 1, 2, \dots, r_n$. This reparametrization makes computation much easier in any standard computing environment such as R software. The convergence properties of the algorithm are similar to those of the Gauss–Seidel method. Depending on the matrix $\mathbf{H} = (\mathbf{L}, \mathbf{ZB})' \Omega(\mathbf{L}, \mathbf{ZB}) = (h_{ij})_{i,j=1,\dots,d+r_n+J * q_n}$, the algorithm will converge if either \mathbf{H} is positive definite or strictly or irreducibly diagonally dominant (i.e. $|h_{ii}| \geq \sum_{j \neq i} |h_{ij}|$). When it converges, the global convexity of the log-likelihood function guarantees that we can reach the maximum.

Clearly, the proposed method needs to select orders and knots of B-splines. In practice, we often use cubic B-splines for every unknown function. In terms of the choice of knots for B-splines, for simplicity, we assume $r_n = q_n$ in (4) and (5), that is, $N_n + 4 = K_n + 4$ so that the numbers of knots for all the $(J + 1)$ nonparametric functions are the same; they are equal to $K'_n = K_n + 2$, where K_n is the number of internal knots in a finite interval. For example, let ξ_0 and ξ_{K_n+1} be the two endpoints of interval $[a, b]$, the support of a covariate; the knots ξ_j , $j = 0, 1, \dots, K_n + 1$, may be chosen as the $\{j/(K_n + 1) \times 100\%$ th quantiles of observations on this covariate. We proposed to determine q_n or K_n by Bayesian information criterion (BIC), a model selection criterion proposed by Schwarz (1978). Specifically, suppose $\hat{\boldsymbol{\kappa}}_{q_n} = (\hat{\boldsymbol{\beta}}'_{q_n}, \hat{\boldsymbol{\gamma}}'_{q_n}, \hat{\boldsymbol{\alpha}}'_{q_n})'$ is the maximum likelihood estimator from (6) under two q_n -dimensional spline spaces \mathcal{S}_n and \mathbb{L}_n . Define

$$\text{BIC}(q_n) = -2l_n(\hat{\boldsymbol{\kappa}}_{q_n}) + \log(n) \{d + (J + 1)q_n\}. \tag{14}$$

A large value of BIC implies lack of fit. Various forms of BIC have been proposed and tested for knots selection in semiparametric models. For instance, He *et al.* (2002) applied the BIC in a semiparametric longitudinal data model. From both our theoretical and empirical results, we suggest search for the first local minimum of BIC over the range $\max\{n^{1/5}, 4\} \leq q_n \leq 10 + 2n^{1/5}$. There usually exists only one local minimum within this range. The reason for q_n starting at 4 at minimum is that for cubic splines, the smallest possible space without any internal knots is of four dimensions.

One potential limitation for the additive hazards model (2) is that the estimated conditional cumulative hazard function given by (3) may occasionally result in negative values or very large positive values, which consequently generates improper conditional probabilities in (9); that is, for a C value, $P(C)$ is bigger than 1 or has small probabilities close to zero. To avoid these two extremes in numerical calculation, we propose to confine $P(C)$ within an interval $[10^{-6}, 1 - 10^{-6}]$ and the resulting algorithm worked numerically quite stably.

5. Inference

Standard inference method for $\boldsymbol{\beta}_0$ is conducted with the utility of the estimated information matrix $I(\hat{\boldsymbol{\beta}}_n)$ via the plug-in method, which provides a consistent estimator of the limiting variance–covariance matrix of $\hat{\boldsymbol{\beta}}_n$. In our case, because the closed form of $I(\boldsymbol{\beta}_0)$ is hard to derive analytically, this plug-in method is not available in general here. Although there are

many alternative methods in the literature, such as the nonparametric bootstrap and the m out of n subsampling bootstrap, computational burden is a major concern. We found that the weighted bootstrap procedure suggested by Ma & Kosorok (2005b) is easy to implement and works reasonably well in our setting. We just need to modify the algorithm in Section 4 by replacing u_i and ω_i by $u_i = -\tilde{W}_i(\Delta_i - p_i)/(1 - p_i)$ and $\omega_i = 2p_i/(1 - p_i)$, respectively, where $\tilde{W}_i, i = 1, \dots, n$ are i.i.d. positive random weights drawn from \tilde{W} , for example, exponential distribution with a rate equal to 1, satisfying $E(\tilde{W}) = 1$ and $\text{var}(\tilde{W}) = 1$ and independent of $(C, \Delta, \mathbf{Z}(C), \mathbf{W})$. More details concerning the weighted bootstrap can be found in their paper.

6. Simulation experiment

To evaluate finite-sample performances of the proposed estimators, we conduct a simulation study. Event times are generated from an exponential distribution with a hazard rate given as follows:

$$\lambda(t|\mathbf{x}, \mathbf{w}) = \lambda_0(t) + \boldsymbol{\beta}'\mathbf{x} + \phi_1(w_1) + \phi_2(w_2), \tag{15}$$

where $\lambda_0(t) \equiv 0.1t + 3.3$, $\boldsymbol{\beta} = (\beta_1, \beta_2)' = (0.3, 0.5)'$ and covariates $X_1 \sim \text{Uniform}[-1.5, 1.5]$ and $X_2 \sim \text{Bernoulli}(prob = 0.5) - 0.5$. The two nonlinear functions are $\phi_1(w_1) = \sin\{\pi(w_1/3 - 1)\}$ and $\phi_2(w_2) = 0.3\{(w_2 - 6)^2 - 3\}$, with both W_1 and $W_2 \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}[3, 9]$. Note that these two covariates satisfy the condition in (B1)(ii), as $E(X_1) = E(X_2) = E\{\phi_1(W_1)\} = E\{\phi_2(W_2)\} = 0$. Censoring times are generated from an exponential distribution with a hazard rate $\lambda_c(t|\mathbf{x}, \mathbf{w}) = 1 + |x_1 + x_2 + 0.5w_1 - 0.5w_2|$, confined in interval $[0.04, 1.8]$. In the simulation, we use cubic B-splines to fit the three unknown curves, where the number of knots is determined by the BIC given in (14). The simulation is replicated 1000 times for sample sizes equal to $n = 200, 400$ and 800 .

Table 1 summarizes the results of estimation and coverage probability. We observe that the averages of parameter estimates are close to the true values, so the estimation consistency is evident. The weighted bootstrap method tends to slightly underestimate the standard deviations of the parameter estimators, which may be attributed to lower coverage probabilities when the sample size is 200 or smaller. However, when sample size increases to 400 or larger, the sample standard deviations and the averaged estimated standard deviations are becoming close, and the coverage probabilities of 95% confidence intervals are very near the nominal level.

Table 1. Summary results of the simulation study for the partly linear AHR model with current status data. Mean: mean of the parameter estimates; SSD: sample standard deviation of the parameter estimates; AESD: average estimated standard deviation of the parameter estimates based on 100 weighted bootstrap samples; CP: coverage probability of the 95% confidence interval; n: sample size. Results are based on 1000 simulation replicates

n	True $\boldsymbol{\beta}$	Mean	SSD	AESD	CP(%)
200	$\beta_1 = 0.3$	0.276	0.325	0.314	93.5
	$\beta_2 = 0.5$	0.517	0.598	0.576	94.5
400	$\beta_1 = 0.3$	0.296	0.233	0.227	94.1
	$\beta_2 = 0.5$	0.483	0.416	0.409	95.1
800	$\beta_1 = 0.3$	0.293	0.172	0.170	94.7
	$\beta_2 = 0.5$	0.501	0.291	0.299	95.1

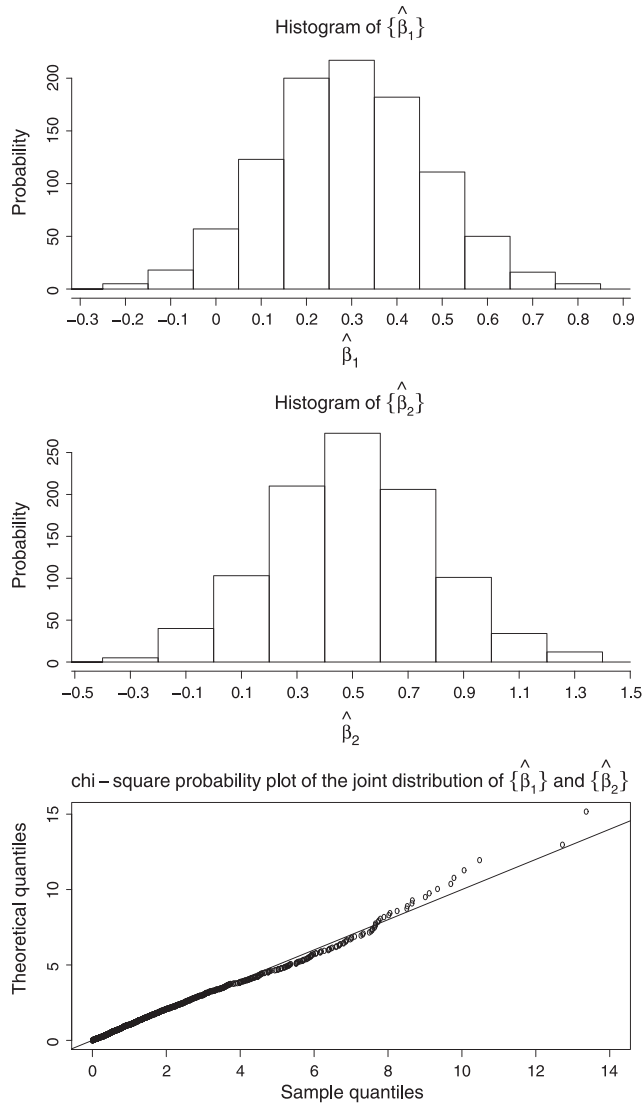


Fig. 1. Histograms of $\hat{\beta}_1$ and $\hat{\beta}_2$ and chi-squared probability plot of the joint distribution of $\hat{\beta}_1$ and $\hat{\beta}_2$, based on 1000 replicates with sample size 800.

Histograms of $\hat{\beta}_1$ and $\hat{\beta}_2$ and chi-square probability plot for the joint distribution of $\hat{\beta}_1$ and $\hat{\beta}_2$ are shown in Figure 1 for $n = 800$. It is apparent that the marginal distributions and the joint distribution of $\hat{\beta}_1$ and $\hat{\beta}_2$ are approximately normal and bivariate-normal, respectively. The fitted curves and pointwise 95% confidence intervals for $\phi_1(\cdot)$, $\phi_2(\cdot)$ and $\Lambda_0(\cdot)$ are shown in Figure 2. It is evident that average estimated curves capture the true curves very well and that the true curves lie in the pointwise 95% confidence intervals. Because of the sparsity of data near the right limit of interval $[0.04, 1.8]$, the estimated cumulative baseline hazard curve $\hat{\Lambda}_n(t)$ deviates from the true curve slightly starting at $t = 1.0$, and the 95% confidence interval becomes wider. In summary, the proposed method has shown satisfactory empirical evidence in good alignment with our theoretical results.

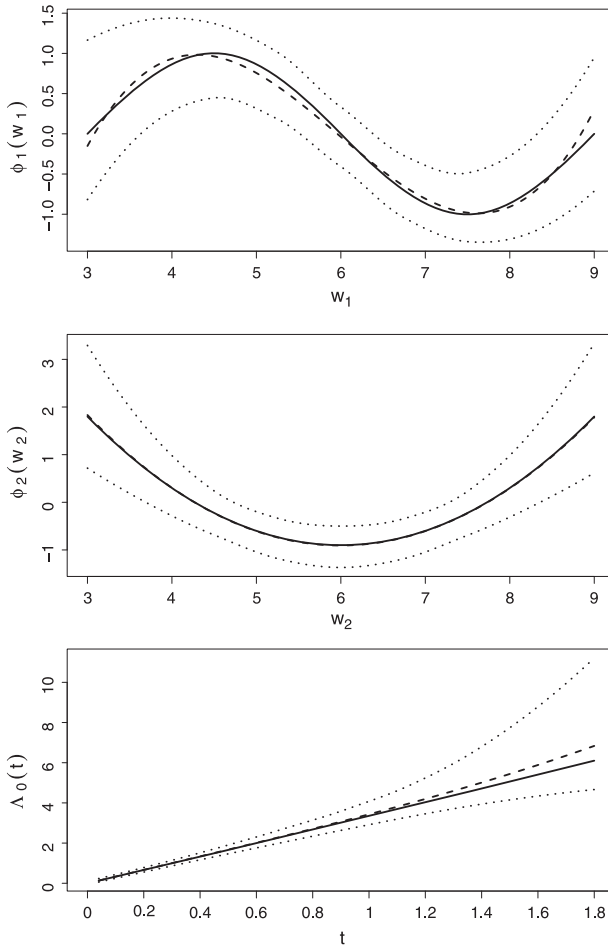


Fig. 2. Sieve estimates and 95% confidence intervals for functions $\phi_1(w_1)$, $\phi_2(w_2)$ and $\Lambda_0(t)$, based on 1000 replicates with sample size 800. The solid lines stand for the true curves. The dashed lines are the average estimated curves. The dotted lines represent the pointwise 95% confidence intervals.

7. Analysis of renal function recovery

Among hospitalized renal patients, acute kidney injury (AKI) is a typical kidney disease syndrome with substantial impact on both short-term and long-term clinical outcomes. Identifying risk factors associated with renal recovery in patients requiring renal replacement therapy (RRT) can help clinicians develop strategies to prevent non-recovery and improve patient's quality of life. At the University of Michigan Hospital, a retrospective study was conducted to understand the relationship between the risk of renal function recovery and clinical covariates of interest. The study involved 170 consecutive hospitalized patients who developed AKI requiring RRT. For each of the patients, his or her time of the inception of dialysis was recorded along with time of hospital discharge, which may be regarded as a monitoring time. In this study, the investigators only observed patient's current status of renal recovery at discharge time but did not know exactly when renal function recovery occurred. More details concerning the study background and preliminary findings can be found in Heung *et al.* (2012).

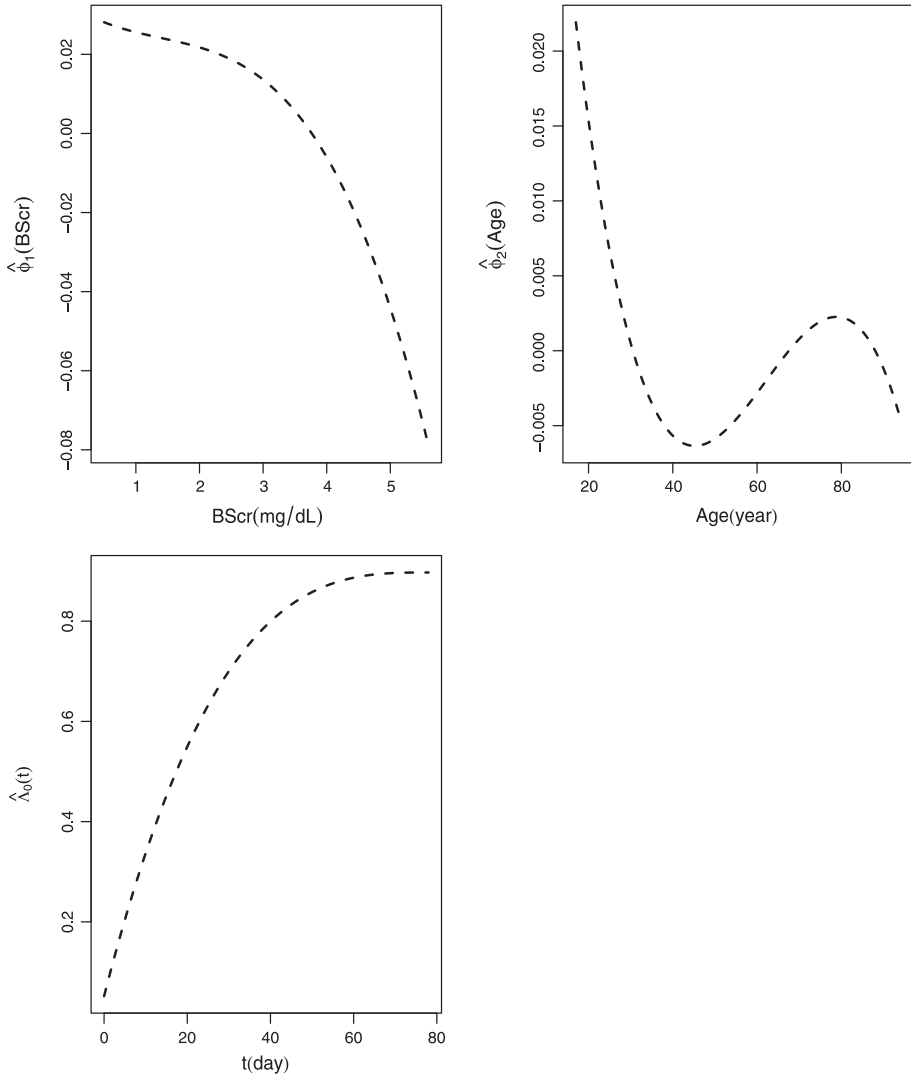


Fig. 3. Estimated curves $\phi_1(\text{BScr})$, $\phi_2(\text{Age})$ and $\Lambda_0(t)$ in the analysis of renal function recovery data.

In this analysis, we are particularly interested in applying the partly linear AHR model to assess the relationship between the hazard of occurrence of renal recovery and several clinical factors, including baseline serum creatinine level, use of vasopressor, age and gender. Measuring baseline serum creatinine is a useful and inexpensive method of evaluating renal dysfunction. It is known that creatinine is a non-protein waste product of creatine phosphate metabolism by skeletal muscle tissue.

To begin our analysis, let T be time in days from the time of starting dialysis to the date of renal function recovery, and let C be the monitoring time given as of the time of hospital discharge. In our analysis, to obtain a more stable estimation of the cumulative hazard function on the right boundary, we removed two outliers with the largest observed monitoring times, 100 and 367 days, so that the remaining 168 observations were used and the values of C range from 0 to 78 days with a mean of 17.7 days. Through the preliminary analysis of Heung *et al.*

(2012), two baseline covariates, baseline serum creatinine (BScr, varying between 0.5 and 5.6) and use of vasopressor (VP, binary) are important clinical predictors, as well as age (varying between 17 and 94 years) and gender (binary). VP is coded as 1 for yes and 0 for no. Gender is coded as 1 for male and 0 for female.

Denote the indicator of renal recovery at the time of discharge by Δ , 1 for 'no' and 0 for 'yes'. Clearly, $\Delta = 1$ corresponds to the case of being right censored by C , whereas $\Delta = 0$ corresponds to the case of being left censored by C . Let $\lambda(t, \mathbf{X}, \mathbf{W})$ be the hazard function of recovery time T , where covariate vectors $\mathbf{X} = (X_1, X_2)'$ and $\mathbf{W} = (W_1, W_2)'$ include $X_1 = \text{VP}$, $X_2 = \text{Gender}$, $W_1 = \text{BScr}$ and $W_2 = \text{Age}$. Model (2) is applied to establish a relationship between the hazard function and the four covariates. The parameter estimates are obtained as $\hat{\beta}_1 = 0.00093$ and $\hat{\beta}_2 = -0.01489$. To test the effects of VP and gender on the hazards rate, we use the weighted bootstrap procedure to estimate standard errors, which are obtained as 0.03122 and 0.02838, respectively. The corresponding Z-test statistic values equal to 0.030 and -0.525 , and the p -values equal to 0.976 and 0.600, implying that both adjusted effects of VP and gender are not statistically significant in the AHR model. The three estimated curves for $\phi_2(\text{BScr})$, $\phi_2(\text{Age})$ and $\Lambda_0(t)$ are shown in Figure 3. They indicate nonlinear effects of BScr and age on the hazards rate. The hazards rate decreases with an increase in BScr or age or both, which suggests that it will take a longer time to recover from renal injury for older patients with higher BScr, owing to probably more severe renal dysfunction at baseline and/or older age. These results are consistent with those findings in the preliminary analysis reported by Heung *et al.* (2012). But our analysis provides more detailed nonlinear profiles regarding the effects of BScr and age, both of which are of clinical importance.

8. Concluding remarks

In this paper, we have established an efficient sieve estimation in the partly linear AHR model with current status data and an effective backfitting algorithm for implementation. Our model allows including multiple nonparametric functions of covariates and time-dependent linear covariates, which is more appealing and useful than Ma & Kosorok (2005a)'s model and Ma (2011)'s model with only a single nonparametric function of covariate. As clearly illustrated in our data example of renal function recovery, two covariates have both exhibited nonlinear effects to predict the hazard rate. We use polynomial splines to model all the nonparametric regression components. The splines for the cumulative hazard function are restricted to non-negative and monotone polynomial splines. By maximizing the log-likelihood function over the spline spanned sieve spaces, we jointly estimate both nonparametric functions and the parametric vector of regression coefficients. We show that the estimator for the parameter vector of regression coefficients is semiparametrically efficient and the sieve estimators of the unknown functions, including the cumulative baseline hazard function, achieve the optimal convergence rates, which are faster than the cubic rate existing in the current literature (e.g. Ma & Kosorok, 2005a; Ma, 2011). The theory pertaining to the monotonicity-constrained sieve estimation method based on polynomial splines with arbitrary orders is technically and numerically new in the context of current status data analysis regardless of the AHR model or the Cox PH model. The proposed method can also handle time-dependent linear covariates, which have led us to make use of the theory of counting processes and martingales combined with the theory of empirical processes. The utility of the martingale theory in our proofs is a noticeably new approach in the analysis of 'case 1' interval-censored or current status data. For example, in comparison with Huang (1999), who considered the partly linear additive Cox PH model with right-censored data, we have seemingly greater technical challenges because in our setting of current status data, it is not possible to profile out the baseline hazard function, so that

we need to simultaneously estimate the baseline and nonparametric functions subject to the monotonicity constraint.

To facilitate implementation of the proposed method, we provide the backfitting algorithm via a generalized linear model for binary outcomes in the estimation of the parametric vector of regression coefficients and multiple nonparametric functions of covariates with a monotonicity constraint. Because it is difficult to analytically yield a closed form for the asymptotic variance of the parameter estimator, statistical inference is based on the weighted bootstrap method. A simulation study shows that the proposed estimation and inference methods work well. In comparison with other semiparametric models such as the partly linear additive Cox model and partly linear transformation models, the partly linear AHR model offers a very useful alternative to analyse current status data.

If one is interested in estimating the survival function for a given covariate process $\mathbf{z}(t)$ and auxiliary covariate vector \mathbf{w} , as suggested by Zeng *et al.* (2006), a natural estimator would take the form $\hat{S}_n(t) = \exp[-\hat{\Lambda}_n(t) - \hat{\boldsymbol{\beta}}_n' \mathbf{z}(t) - t\{\hat{\phi}_{1n}(w_1) + \dots + \hat{\phi}_{Jn}(w_J)\}]$. Because $\hat{\boldsymbol{\beta}}_n' \mathbf{z}(t) + t\{\hat{\phi}_{1n}(w_1) + \dots + \hat{\phi}_{Jn}(w_J)\}$ varies with time t , the monotonicity of $\hat{S}_n(t)$ might not hold. A modified estimator satisfying the monotonicity constraint may be given by the maximal decreasing function below $\hat{S}_n(t)$, denoted by $\hat{S}_n^*(t)$. Using the established large-sample theory for $(\hat{\Lambda}_n, \hat{\boldsymbol{\beta}}_n, \hat{\phi}_n)$ in this paper, we can show with little effort that $\hat{S}_n^*(t)$ is also consistent.

It is worth mentioning that Kosorok (2008) provided a comprehensive survey of the theory for semiparametric models. Demonstrated in this monograph, empirical processes have become standard tools to study semiparametric estimation and inferences. We have adopted these tools combined with the theory of martingales in this paper to deal with constrained joint estimation and simultaneous inference in the partly linear AHR model with current status data. We envision that the techniques developed in this paper can also be applied to other important survival analysis models, such as partly linear additive Cox models and partly linear transformation models, so that the convergence rate of nonparametric estimators may be improved to the optimal level. Moreover, it is of interest to apply our techniques in other types of censored data, such as 'case 2' interval-censored data studied by Zeng *et al.* (2006) in a semiparametric linear additive risk model.

In this paper, we have used regression splines; like all spline-based estimation approaches, it may have a potential issue of numerical stability in the functional estimation when available data are sparse. To overcome this, some authors (e.g. Zucker & Karr, 1990; Gray, 1992) suggested to add a certain penalty term in the estimation, so that the resulting estimated functions can become smoother and avoid overfitting. This idea may be adopted here by subtracting a penalty function in (6) to improve the estimates. Another issue of practical importance is how to determine which covariate has a linear form and which has a nonlinear form. Although desirable, a rigorous procedure remains an unsolved problem in our paper and is worth further exploration. In practice, one may gather some preliminary evidence about the functional form of a covariate by performing hypothesis testing or simple univariate screening analysis. For partially linear additive models with uncensored data, Zhang *et al.* (2011) and Lian *et al.* (2012) have proposed some procedures for automatic detection of functional forms, and it is worth investigating in a future work concerning whether their methods may be extended in our proposed model.

Acknowledgements

The authors would like to thank the associate editor and two anonymous referees for their constructive comments that have greatly improved an earlier version of this paper. Lu is grateful to the hospitality of the Department of Biostatistics at the University of Michigan, where a draft

of this work was commenced during his sabbatical leave in 2009. Particularly, both authors thank Profs. John Kalbfleisch and Bin Nan for many helpful discussions during the preparation of the manuscript. Lu’s research was supported by an NSERC Discovery Grant from Canada, and Song’s research was supported in part by an NSF grant from the USA.

Appendix: Proofs of theorems 1–3

A.1. Proof of theorem 1

We follow Sasieni (1992) to derive the information bound. Note that Huang (1999) uses the same approach to obtain the information bound in the partly linear additive Cox model with right-censored data, where a projection onto a sum space of $J + 1$ non-orthogonal L_2 spaces is the key. In the following, we take similar steps to derive the information bound for the estimator of β_0 .

The log-likelihood function for one observation in terms of counting processes is given by, up to an additive constant independent of β , Λ and ϕ ,

$$l(\beta, \phi, \Lambda) = \Delta \log [\exp \{-\Lambda(C) - \beta'Z(C) - C\phi(W)\}] + (1 - \Delta) \log [1 - \exp \{-\Lambda(C) - \beta'Z(C) - C\phi(W)\}] = \int \log\{p(t)\} dN_1(t) + \int \log\{1 - p(t)\} dN_2(t),$$

where $\Lambda(t) = \int_0^t \lambda(s) ds$, $p(t) = p(t, Z(t), W) = \exp\{-\Lambda(t) - \beta'Z(t) - t\phi(W)\}$ and $N_1(t)$ and $N_2(t)$ are given in Section 3. Consider a parametric smooth submodel $\{\Lambda(\varpi) : \varpi \in \mathbb{R}^d\}$ and $\{\phi_j(\varpi_j) : \varpi_j \in \mathbb{R}^d, 1 \leq j \leq J\}$ in which $\Lambda_{(0)} = \Lambda$, $\phi_{j(0)} = \phi_j$. Let

$$a_1(t) = \left. \frac{\partial \Lambda(\varpi)}{\partial \varpi} (t) \right|_{\varpi=0},$$

and

$$h_j(w_j) = \left. \frac{\partial \phi_j(\varpi_j)}{\partial \varpi_j} (w_j) \right|_{\varpi_j=0}, \quad 1 \leq j \leq J.$$

The score operators for β , Λ and ϕ_j are the partial derivatives of the log-likelihood $l(\beta, \phi_1(\varpi_1), \dots, \phi_J(\varpi_J), \Lambda(\varpi))$ with respect to β and $(\varpi, \varpi_1, \dots, \varpi_J)$ evaluated at $\varpi = 0, \varpi_1 = 0, \dots, \varpi_J = 0$. Recall that both M_1 and M_2 defined in (10) are the martingales. Let $a(t) = a_1(t)/t$. We obtain

$$i_\beta \equiv \int_0^\infty \tilde{Z}(t)t \left\{ \frac{p(t)}{1 - p(t)} dM_2(t) - dM_1(t) \right\},$$

$$i_\Lambda a \equiv \int_0^\infty a(t)t \left\{ \frac{p(t)}{1 - p(t)} dM_2(t) - dM_1(t) \right\},$$

$$i_{\phi_j} h_j \equiv \int_0^\infty h_j(W_j)t \left\{ \frac{p(t)}{1 - p(t)} dM_2(t) - dM_1(t) \right\}, \quad 1 \leq j \leq J.$$

Let $Q(t, Z(t), W) = t^2 p(t, Z(t), W) / \{1 - p(t, Z(t), W)\}$. Define $L_2(P_C^Q) \equiv \{a : E\{ \|a(C)\|^2 Q(C, Z(C), W) \} < \infty\}$ and $L_2^0(P_{W_j}^Q) \equiv \{h_j : E\{h_j(W_j)\} = 0, E\{ \|h_j(W_j)\|^2 Q(C, Z(C), W) \} < \infty\}, 1 \leq j \leq J$. Let

$$A_\Lambda = \{ \dot{l}_\Lambda \mathbf{a} : \mathbf{a} \in L_2(P_C^Q) \}$$

and

$$H_j = \{ \dot{l}_{\phi_j} \mathbf{h}_j : \mathbf{h}_j \in L_2^0(P_{W_j}^Q) \}, \quad 1 \leq j \leq J.$$

Let $\mathbf{h} = (\mathbf{h}_1, \dots, \mathbf{h}_J)$ and $\dot{l}_\phi \mathbf{h} = \dot{l}_{\phi_1} \mathbf{h}_1 + \dots + \dot{l}_{\phi_J} \mathbf{h}_J$. To obtain the information bound for estimation of β_0 , we apply the non-orthogonal projection approach given by Sasieni (1992) or the alternating projections theorem of Bickel *et al.* (1993, Theorem A.4.2).

We first project \dot{l}_β onto the space A_Λ . Finding function $\tilde{\mathbf{a}}(t) \in L_2(P_C^Q)$ so that $\dot{l}_\beta - \dot{l}_\Lambda(\tilde{\mathbf{a}}) \perp \dot{l}_\Lambda(\mathbf{a})$ is equivalent to finding $\tilde{\mathbf{a}}(t) \in L_2(P_C^Q)$ such that

$$E \left\{ (\dot{l}_\beta - \dot{l}_\Lambda(\tilde{\mathbf{a}}))' \dot{l}_\Lambda(\mathbf{a}) \right\} = 0.$$

It yields

$$\tilde{\mathbf{a}}(t) = \frac{E \{ \tilde{\mathbf{Z}}(t) Y(t) \sigma(t) Q(t, \mathbf{Z}(t), \mathbf{W}) \}}{E \{ Y(t) \sigma(t) Q(t, \mathbf{Z}(t), \mathbf{W}) \}}.$$

Hence, we have

$$\dot{l}_\beta - \dot{l}_\Lambda(\tilde{\mathbf{a}}) = \int_0^\infty \{ \tilde{\mathbf{Z}}(t) - \tilde{\mathbf{a}}(t) \} t \left\{ \frac{p(t, \mathbf{Z}(t), \mathbf{W})}{1 - p(t, \mathbf{Z}(t), \mathbf{W})} dM_2(t) - dM_1(t) \right\}.$$

Next, we project $\dot{l}_{\phi_j} \mathbf{h}_j$ onto A_Λ . Using similar calculations, we obtain

$$\mathbf{b}_{h_j}(t) = \frac{E \{ \mathbf{h}_j(W_j) Y(t) \sigma(t) Q(t, \mathbf{Z}(t), \mathbf{W}) \}}{E \{ Y(t) \sigma(t) Q(t, \mathbf{Z}(t), \mathbf{W}) \}},$$

and

$$\dot{l}_{\phi_j} \mathbf{h}_j - \dot{l}_\Lambda(\mathbf{b}_{h_j}) = \int_0^\infty \{ \mathbf{h}_j(W_j) - \mathbf{b}_{h_j}(t) \} t \left\{ \frac{p(t, \mathbf{Z}(t), \mathbf{W})}{1 - p(t, \mathbf{Z}(t), \mathbf{W})} dM_2(t) - dM_1(t) \right\}.$$

Finally, we project the space generated by $\dot{l}_\beta - \dot{l}_\Lambda(\tilde{\mathbf{a}})$ onto the sum space generated by $\dot{l}_{\phi_j}(\mathbf{h}_j) - \dot{l}_\Lambda(\mathbf{b}_{h_j}^*)$. Denote $K\mathbf{h} = K\mathbf{h}_1 + \dots + K\mathbf{h}_J$, where $K\mathbf{h}_j = \dot{l}_{\phi_j} \mathbf{h}_j - \dot{l}_\Lambda(\mathbf{b}_{h_j})$, $1 \leq j \leq J$. Then the least favourable direction is given by $\mathbf{h}^* = (\mathbf{h}_1^*, \dots, \mathbf{h}_J^*)$ that minimizes

$$E \left[\int_0^\infty \| \tilde{\mathbf{Z}}(t) - \tilde{\mathbf{a}}(t) - \mathbf{h}^*(\mathbf{W}) \|^2 Q(t, \mathbf{Z}(t), \mathbf{W}) Y(t) \sigma(t) \right] dt,$$

which is equivalent to finding \mathbf{h}^* for every $\mathbf{h}_j \in L_2^0(P_{W_j}^Q)$ such that

$$\{ \dot{l}_\beta - \dot{l}_\Lambda(\tilde{\mathbf{a}}) \} - K\mathbf{h}^* \perp K\mathbf{h}_j,$$

where $\mathbf{h}^*(\mathbf{W}) = \mathbf{h}_1^*(W_1) + \dots + \mathbf{h}_J^*(W_J)$. This leads to

$$E \int_0^\infty \{ (\tilde{\mathbf{Z}}(t) - \tilde{\mathbf{a}}(t)) - K\mathbf{h}^* \}' (K\mathbf{h}_j) Q(t, \mathbf{Z}(t), \mathbf{W}) Y(t) \sigma(t) dt = 0. \tag{16}$$

Let $\mathbf{h}(\mathbf{W}) = \mathbf{h}_1(W_1) + \dots + \mathbf{h}_J(W_J)$. Note that

$$K\mathbf{h} = \int_0^\infty \{\mathbf{h}(\mathbf{W}) - \mathbf{b}_h(t)\} t \left\{ \frac{p(t, \mathbf{Z}(t), \mathbf{W})}{1 - p(t, \mathbf{Z}(t), \mathbf{W})} dM_2(t) - dM_1(t) \right\},$$

where

$$\mathbf{b}_h(t) = \frac{E \{\mathbf{h}(\mathbf{W})Y(t)\sigma(t)Q(t, \mathbf{Z}(t), \mathbf{W})\}}{E \{Y(t)\sigma(t)Q(t, \mathbf{Z}(t), \mathbf{W})\}}.$$

Let $\mathbf{a}^*(t) = \tilde{\mathbf{a}}(t) - \mathbf{b}_{h^*}(t)$. Because for any $\mathbf{a}(t) \in L_2(P_C^Q)$,

$$E \int_0^\infty \left[\{(\tilde{\mathbf{Z}}(t) - \tilde{\mathbf{a}}(t)) - (\mathbf{h}^*(\mathbf{W}) - \mathbf{b}_{h^*}(t))\}' \mathbf{a}(t) Q(t, \mathbf{Z}(t), \mathbf{W}) Y(t) \sigma(t) dt \right] = 0,$$

and for any $\mathbf{a}(t) \in L_2(P_C^Q)$ and $\mathbf{h}(\mathbf{W})$ belonging to the sum space $L_2^0(P_{W_1}^Q) + \dots + L_2^0(P_{W_J}^Q)$,

$$E \int_0^\infty \{\tilde{\mathbf{Z}}(t) - (\mathbf{a}^*(t) + \mathbf{h}^*(\mathbf{W}))\}' \{\mathbf{a}(t) + \mathbf{h}(\mathbf{W})\} Q(t, \mathbf{Z}(t), \mathbf{W}) Y(t) \sigma(t) dt = 0,$$

we can equivalently obtain $(\mathbf{a}^*, \mathbf{h}_1^*, \dots, \mathbf{h}_J^*)$ as the solution to the following equations:

$$E [\{\tilde{\mathbf{Z}}(C) - \mathbf{a}^* - \mathbf{h}_1^* - \dots - \mathbf{h}_J^*\} Q(C, \mathbf{Z}(C), \mathbf{W}) | C = c] = 0 \quad \text{a.s. } P_C, \tag{17}$$

and for $1 \leq j \leq J$,

$$E [\{\tilde{\mathbf{Z}}(C) - \mathbf{a}^* - \mathbf{h}_1^* - \dots - \mathbf{h}_j^*\} Q(C, \mathbf{Z}(C), \mathbf{W}) | W_j = w_j] = 0 \quad \text{a.s. } P_{W_j}. \tag{18}$$

It is clear that these \mathbf{a}^* and \mathbf{h}^* give the desired least favourable directions for the information bound.

We note that $\mathbf{a}^* + \mathbf{h}_1^* + \dots + \mathbf{h}_J^*$ is actually the projection of $\tilde{\mathbf{Z}}$ onto the sum space $\mathbf{S} = L_2(P_C^Q) + L_2^0(P_{W_1}^Q) + \dots + L_2^0(P_{W_J}^Q)$. We now show that the sum space \mathbf{S} is closed, in which the project exists. According to proposition 2 of Bickel *et al.* (1993, pp. 440–441), it suffices to show that for $\mathbf{a} \in L_2(P_C^Q)$ and $\mathbf{h}_j \in L_2^0(P_{W_j}^Q)$, $1 \leq j \leq J$,

$$E \left(\|\mathbf{a} + \mathbf{h}_1 + \dots + \mathbf{h}_J\|_Q^2 \right) \geq mE \left(\|\mathbf{a}\|_Q^2 + \|\mathbf{h}_1\|_Q^2 + \dots + \|\mathbf{h}_J\|_Q^2 \right), \tag{19}$$

where $\|\cdot\|_Q^2 = \|\cdot\|^2 Q(C, \mathbf{Z}(C), \mathbf{W})$. It is easy to show that a Hilbert space endowed with the inner product $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle_Q = E(\mathbf{e}_1 \mathbf{e}_2 Q)$ and the norm $\|\mathbf{e}\|_Q = \sqrt{\langle \mathbf{e}, \mathbf{e} \rangle_Q}$ is still a Hilbert space. Under conditions (B3)–(B5), we obtain

$$\begin{aligned} E \left(\|\mathbf{a} + \mathbf{h}_1 + \dots + \mathbf{h}_J\|_Q^2 \right) &\geq m_1 E \left(\|\mathbf{a} + \mathbf{h}_1 + \dots + \mathbf{h}_J\|^2 \right) \\ &\geq m_2 E \left(\|\mathbf{a}\|^2 + \|\mathbf{h}_1\|^2 + \dots + \|\mathbf{h}_J\|^2 \right) \\ &\geq mE \left(\|\mathbf{a}\|_Q^2 + \|\mathbf{h}_1\|_Q^2 + \dots + \|\mathbf{h}_J\|_Q^2 \right). \end{aligned}$$

Thus, (19) holds, where the second inequality follows from lemma 1 of Stone (1985). □

A.2. Proof of theorem 2

To prove theorem 2, we need lemmas L.1–L.5, the detail of which is included in the supplementary material. For l_0 defined in lemma L.4 and any g satisfying $P l_0(\cdot, g_n) - P l_0(\cdot, g) \geq 0$, we define a distance d_n to be

$$d^2(g, g_n) = Pl_0(\cdot, g_n) - Pl_0(\cdot, g).$$

Let $M_0(g) = Pl_0(\cdot, g)$, $\eta > 0$; by lemma L.2 and lemma 3.4.2 of Van der Vaart & Wellner (1996), we have

$$\begin{aligned} & E \sup_{n/2 \leq d(g, g_n) \leq n} |\mathbb{P}_n(l_0(\cdot, g)) - \mathbb{P}_n(l_0(\cdot, g_n)) - (M_0(g) - M_0(g_n))| \\ &= E \sup_{n/2 \leq d(g, g_n) \leq n} |(\mathbb{P}_n - P)\{l_0(\cdot, g) - l_0(\cdot, g_n)\}| \\ &\leq n^{-1/2} \eta q_n^{1/2}. \end{aligned}$$

Because \hat{g}_n is a consistent estimator of g_n shown by lemma L.5, taking $\eta = r_n^{-1}$, by theorem 3.4.1 of Van der Vaart & Wellner (1996), we obtain

$$r_n^2 d^2(\hat{g}_n, g_n) = O_p(1),$$

where r_n satisfies

$$r_n^2 (r_n^{-1} q_n^{1/2}) = O(n^{1/2}).$$

It follows that $r_n = q_n^{-1/2} n^{1/2} = n^{(1-\nu)/2}$. Therefore, assuming $g = \hat{g}_n$ in lemma L.4, by lemmas L.4 and L.5,

$$\|\hat{g}_n - g_n\|_2^2 = O_p(n^{-(1-\nu)} + n^{-2\nu p}).$$

Because $\|g_n - g_0\|_2^2 = O_p(n^{-(1-\nu)} + n^{-2\nu p})$ by lemma L.3, we have

$$\|\hat{g}_n - g_0\|_2^2 = O_p(n^{-(1-\nu)} + n^{-2\nu p}).$$

Observe that

$$\|\hat{g}_n - g_0\|_2^2 = E \left[\hat{\Lambda}_n(C) + \hat{\beta}'_n \mathbf{Z}(C) + C \hat{\phi}_n(\mathbf{W}) - \{\Lambda_0(C) + \beta'_0 \mathbf{Z}(C) + C \phi_0(\mathbf{W})\} \right]^2,$$

where the expectation is taken with respect to $\mathbf{V} = (C, \mathbf{Z}(C), \mathbf{W})$. Let $q(\mathbf{V}) = p(\mathbf{V})/(1-p(\mathbf{V}))$, there exists $m_2 > m_1 > 0$ such that $0 \leq m_1 \leq q(\mathbf{V}) \leq m_2$. For the projections \mathbf{a}^* and \mathbf{h}^* defined in theorem 1,

$$\begin{aligned} & E \left[\left\{ \hat{\Lambda}_n(C) + \hat{\beta}'_n \mathbf{Z}(C) + C \hat{\phi}_n(\mathbf{W}) - (\Lambda_0(C) + \beta'_0 \mathbf{Z}(C) + C \phi_0(\mathbf{W})) \right\}^2 q(\mathbf{V}) \right] \\ &= E \left[\left\{ (\hat{\beta}_n - \beta_0)' \tilde{\mathbf{Z}}(C) + (\hat{\Lambda}_n(C) - \Lambda_0(C)) / C + (\hat{\phi}_n(\mathbf{W}) - \phi_0(\mathbf{W})) \right\}^2 C^2 q(\mathbf{V}) \right] \\ &= E \left[\left\{ (\hat{\beta}_n - \beta_0)' (\tilde{\mathbf{Z}}(C) - \mathbf{a}^*(C) - \mathbf{h}^*(\mathbf{W})) + (\hat{\beta}_n - \beta_0)' (\mathbf{a}^*(C) + \mathbf{h}^*(\mathbf{W})) \right. \right. \\ &\quad \left. \left. + (\hat{\Lambda}_n(C) - \Lambda_0(C)) / C + (\hat{\phi}_n(\mathbf{W}) - \phi_0(\mathbf{W})) \right\}^2 C^2 q(\mathbf{V}) \right] \\ &= E \left[\left\{ (\hat{\beta}_n - \beta_0)' (\tilde{\mathbf{Z}}(C) - \mathbf{a}^*(C) - \mathbf{h}^*(\mathbf{W})) \right\}^2 C^2 q(\mathbf{V}) \right] \\ &\quad + E \left[\left\{ (\hat{\beta}_n - \beta_0)' (\mathbf{a}^*(C) + \mathbf{h}^*(\mathbf{W})) + (\hat{\Lambda}_n(C) - \Lambda_0(C)) / C \right. \right. \\ &\quad \left. \left. + (\hat{\phi}_n(\mathbf{W}) - \phi_0(\mathbf{W})) \right\}^2 C^2 q(\mathbf{V}) \right]. \end{aligned}$$

The last equality follows the orthogonality given in the proof of theorem 1. Therefore, the first term on the right-hand side of the last equality equals

$$\begin{aligned}
 & E \left[\left\{ \left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 \right)' \left(\tilde{\mathbf{Z}}(C) - \mathbf{a}^*(C) - \mathbf{h}^*(\mathbf{W}) \right) \right\}^2 C^2 q(\mathbf{V}) \right] \\
 &= \left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 \right)' E \left[\left\{ \tilde{\mathbf{Z}}(C) - \mathbf{a}^*(C) - \mathbf{h}^*(\mathbf{W}) \right\}^{\otimes 2} C^2 q(\mathbf{V}) \right] \left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 \right) \\
 &= \left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 \right)' I(\boldsymbol{\beta}_0) \left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 \right) \\
 &\leq E \left[\left\{ \hat{\Lambda}_n(C) + \hat{\boldsymbol{\beta}}_n' \mathbf{Z}(C) + C \hat{\phi}_n(\mathbf{W}) - (\Lambda_0(C) + \boldsymbol{\beta}'_0 \mathbf{Z}(C) + C \phi_0(\mathbf{W})) \right\}^2 q(\mathbf{V}) \right] \\
 &\leq m_2 E \left[\left\{ \hat{\Lambda}_n(C) + \hat{\boldsymbol{\beta}}_n' \mathbf{Z}(C) + C \hat{\phi}_n(\mathbf{W}) - (\Lambda_0(C) + \boldsymbol{\beta}'_0 \mathbf{Z}(C) + C \phi_0(\mathbf{W})) \right\}^2 \right] \\
 &= m_2 \|\hat{\mathbf{g}}_n - \mathbf{g}_0\|_2^2 \\
 &\leq O_p \left(n^{-(1-\nu)} + n^{-2\nu p} \right).
 \end{aligned}$$

Because the information matrix $I(\boldsymbol{\beta}_0)$ is assumed to be nonsingular, it follows that

$$\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|^2 = O_p \left(n^{-(1-\nu)} + n^{-2\nu p} \right).$$

This in turn implies that

$$E \left[\hat{\Lambda}_n(C) - \Lambda_0(C) + C \left\{ \hat{\phi}_n(\mathbf{W}) - \phi_0(\mathbf{W}) \right\} \right]^2 = O_p \left(n^{-(1-\nu)} + n^{-2\nu p} \right)$$

and

$$E \left[\left\{ \hat{\Lambda}_n(C) - \Lambda_0(C) \right\} + \left\{ \hat{\phi}_n(\mathbf{W}) - \phi_0(\mathbf{W}) \right\} \right]^2 = O_p \left(n^{-(1-\nu)} + n^{-2\nu p} \right).$$

Thus, by lemma 1 of Stone (1985),

$$E \left\{ \hat{\Lambda}_n(C) - \Lambda_0(C) \right\}^2 = O_p \left(n^{-(1-\nu)} + n^{-2\nu p} \right)$$

and

$$E \left\{ \hat{\phi}_{jn}(\mathbf{W}) - \phi_{0j}(\mathbf{W}) \right\}^2 = O_p \left(n^{-(1-\nu)} + n^{-2\nu p} \right), \quad 1 \leq j \leq J.$$

□

A.3. Proof of theorem 3

To prove theorem 3, we need lemma L.6, the detail of which is included in the supplementary material. Because $\mathbb{P}_{nS}(\cdot, \hat{\mathbf{g}}_n)[\mathbf{Z}] = 0$, by (C1) and (C2) of lemma L.6, we have

$$P \left\{ s(\cdot, \hat{\mathbf{g}}_n) [\mathbf{Z} - \mathbf{U}^*] - s(\cdot, \mathbf{g}_0) [\mathbf{Z} - \mathbf{U}^*] \right\} = \mathbb{P}_{nS}(\cdot, \mathbf{g}_0) [\mathbf{Z} - \mathbf{U}^*] + o_p \left(n^{-1/2} \right).$$

Hence, by (C3) of lemma L.6,

$$I(\boldsymbol{\beta}_0) \left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 \right) = \mathbb{P}_{nS}(\cdot, \mathbf{g}_0) [\mathbf{Z} - \mathbf{U}^*] + o_p \left(n^{-1/2} \right) = \mathbb{P}_{nJ}^*_{\boldsymbol{\beta}_0} (\Delta, \mathbf{V}) + o_p \left(n^{-1/2} \right).$$

Finally, by the central limit theorem, $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ has a limiting normal distribution with the limiting variance matrix equal to $I^{-1}(\boldsymbol{\beta}_0)$; hence, $\hat{\boldsymbol{\beta}}_n$ is semiparametric efficient. The proof of theorem 3 is completed. □

References

- Bickel, P. J., Klaassen, C. A., Ritov, Y. & Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*, The Johns Hopkins University Press, Baltimore.
- Claeskens, G., Krivobokova, T. & Opsomer, J. D. (2009). Asymptotic properties of penalized spline estimators. *Biometrika* **96**, 529–544.
- Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Stat. Soc. B Met.* **34**, 187–220.
- Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *J. Am. Stat. Assoc.* **87**, 942–951.
- Groeneboom, P. & Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Verlag Basel, Birkhäuser.
- Hastie, T., Tibshirani, R. & Friedman, J. (2008). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, (2nd edn.), Springer-Verlag, New York.
- He, X., Zhu, Z. Y. & Fung, W. K. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika* **89**, 579–590.
- Heung, M., Wolfgram, D. F., Kommareddi, M., Hu, Y., Song, P. X. & Ojo, A. O. (2012). Fluid overload at initiation of renal replacement therapy is associated with lack of renal recovery in patients with acute kidney injury. *Nephrol. Dial. Transpl.* **27**, 956–961.
- Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Stat.* **24**, 540–568.
- Huang, J. (1999). Efficient estimation of the partly linear additive Cox. model. *Ann. Stat.* **27**, 1536–1563.
- Huang, J. & Rossini, A. J. (1997). Sieve estimation for the proportional-odds failure-time regression model with interval censoring. *J. Am. Stat. Assoc.* **92**, 960–967.
- Kalbfleisch, J. D. & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, (2nd edn.), John Wiley & Sons, Inc., New York.
- Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*, Springer, New York.
- Lian, H., Chen, X. & Yang, J. Y. (2012). Identification of partially linear structure in additive models with an application to gene expression prediction from sequences. *Biometrics* **68**, 437–445.
- Lin, D. Y., Oakes, D. & Ying, Z. (1998). Additive hazards regression with current status data. *Biometrika* **85**, 289–298.
- Lin, D. Y. & Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–71.
- Lu, M., Zhang, Y. & Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika* **94**, 705–718.
- Ma, S. (2011). Additive risk model for current status data with a cured subgroup. *Ann. I. Stat. Math.* **63**, 117–134.
- Ma, S. & Kosorok, M. R. (2005a). Penalized log-likelihood estimation for partly linear transformation models with current status data. *Ann. Stat.* **33**, 2256–2290.
- Ma, S. & Kosorok, M. R. (2005b). Robust semiparametric m-estimation and the weighted bootstrap. *J. Multivariate Anal.* **96**, 190–217.
- Martinussen, T. & Scheike, T. H. (2002). Efficient estimation in additive hazards regression with current status data. *Biometrika* **89**, 649–658.
- McKeague, I. W. & Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika* **81**, 501–514.
- Rabinowitz, D., Tsiatis, A. & Aragon, J. (1995). Regression with interval-censored data. *Biometrika* **82**, 501–513.
- Robertson, T., Wright, F. T., Dykstra, R. L. & Robertson, T. (1988). *Order Restricted Statistical Inference*, Wiley, New York.
- Sasieni, P. (1992). Non-orthogonal projections and their application to calculating the information in a partly linear Cox. model. *Scand. J. Stat.* **9**, 215–233.
- Schumaker, L. (1981). *Spline Functions: Basic Theory*, Wiley, New York.
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464.
- Shiboski, S. C. (1998). Generalized additive models for current status data. *Lifetime Data Anal.* **4**, 29–50.
- Shiboski, S. C. & Jewell, N. P. (1992). Statistical analysis of the time dependence of HIV infectivity based on partner study data. *J. Am. Stat. Assoc.* **87**, 360–372.
- Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Stat.* **10**, 1040–1053.
- Stone, C. J. (1985). Additive regression and other nonparametric models. *Ann. Stat.* **13**, 689–705.
- Thisted, R. A. (1988). *Elements of Statistical Computing: Numerical Computation*, Chapman & Hall/CRC, New York.

- Van der Vaart, A. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics*, Springer, New York.
- Xue, H., Lam, K. F. & Li, G. (2004). Sieve maximum likelihood estimator for semiparametric regression models with current status data. *J. Am. Stat. Assoc.* **99**, 346–356.
- Zeng, D., Cai, J. & Shen, Y. (2006). Semiparametric additive risks model for interval-censored data. *Stat. Sinica* **16**, 287–302.
- Zhang, H. H., Cheng, G. & Liu, Y. (2011). Linear or nonlinear? automatic structure discovery for partially linear models. *J. Am. Stat. Assoc.* **106**, 1099–1109.
- Zucker, D. M. & Karr, A. F. (1990). Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *Ann. Stat.* **18**, 329–353.

Received July 2013, in final form June 2014

Xuwen Lu, Department of Mathematics and Statistics, University of Calgary, 2500 University Drive NW, Calgary, AB, T2N 1N4, Canada.
E-mail: lux@math.ucalgary.ca

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.