Estimation in Partially Linear Single-index Additive Hazards Models with Current Status Data

Pooneh Pordeli * Xuewen Lu † Murray Burke ‡ Peter X.-K. Song §

Abstract
Current status data arise in such areas as demography, economics, epidemiology and survival models. We propose a partially linear single-index additive hazards regression model for current status data. The proposed model can model both linear and nonlinear covariate effects on the hazard and it is a parsimonious model, since it does not use too many parameters. This is particularly important for high dimensional data, which might suffer from "the curse of dimensionality". Our model reduces the dimension of the data through a single-index term. For the estimation, we use B-splines to model the unknown cumulative baseline hazard function and the nonparametric covariate functions. Asymptotic properties of the estimators are derived using the theory of counting processes. Simulation studies are presented to evaluate our method. A renal recovery data set is analyzed to illustrate the usefulness of our proposed model.

Key Words: Current status data, Single index model, Spline, Additive risk model, Counting process

1. Introduction

In biostatistical applications usually the variable of interest is failure time ,that is, the time of occurrence of some event for a sample of individuals. However, in many situations there is limited information for a single observation about the event of interest. With current status data, each subject is observed only once and the only information that we have is that the failure of interest has occurred before or after the examination time. The failure time is either left- or right-censored instead of being observed exactly. For example, events such as the time to onset of nonlethal tumours, time to develop HIV, age at weaning or age at incidence of non-fatal human diseases cannot be known exactly; the only information that is available is a time interval that the event has happened in that period. Among others, Turnbull [1976], Groeneboom and Wellner [1992], Jewell and van der Laan [1996] and Sun [2006], have studied different methods to analyze current status data.

In analysis of survival data a popular choice is the Cox proportional hazards (PH) model. For instance, in a study of current status data, Huang [1996] estimated the parameters of the PH model using profile likelihood approach. However, when we are interested in the absolute hazards change instead of hazards ratio, or when the proportional hazards assumption is violated, an additive hazards (AH) regression model may be more practical. Similar to other models, the AH regression model enables to characterize different types of relations between covariates and event time, which sometimes are quite demanding for practitioners. It has been shown in some situations the AH model can be more plausible and interpretable than the Cox PH model (Lin and Ying [1994]; McKeague and Sasieni [1994]).

Suppose T is time to occurrence of a certain event like tumor onset, C is the random examination time, and $Z(t) = (V(t)^T, X^T)^T$ is the $(p+q)$ dimensional covariate vector

*Ph.D. student, University of Calgary, 2500 University Drive NW, Calgary, AB, Canada, T2N 1N4
†Professor, University of Calgary, 2500 University Drive NW, Calgary, AB, Canada, T2N 1N4
‡Professor, University of Calgary, 2500 University Drive NW, Calgary, AB, Canada, T2N 1N4
§Professor, Department of Biostatistics, University of Michigan, School of Public Health, Ann Arbor, MI

where $V = (V_1, \cdots, V_p)^T$ is the possibly time dependent p-dimensional linear component up to time t, and $X = (X_1, \cdots, X_q)^T$ is the q-dimensional non-linear component. To analyze current status data, Lin et al. [1998] proposed the linear AH model where the hazard function at time t, given the covariate value $V(t)$ up to t, is assumed to take the following form:

$$\lambda(t|V(t)) = \lambda_0(t) + \alpha^T V(t), \tag{1}$$

where $\lambda_0(\cdot)$ is the unknown baseline hazard function and α is the q-dimensional regression coefficient vector. They assumed a semiparametric PH model for the monitoring time and made inference about the regression parameters of model (1) with current status data by using the familiar asymptotic theory for the PH model with right censored data. However, this condition on monitoring time may not be accurate in some applications and the analysis does not make efficient use of data. Martinussen and Scheike [2002] studied model (1) with current status data utilizing the semiparametric efficient score function in estimation. Their approach has two advantages. First, the estimators are efficient and reach the semiparametric information bound; second, it does not involve any structured model assumption for the monitoring time. Although their method provides efficient estimators, it is difficult to use in practice due to numerical instability and the required considerable effort in implementation. Later, Lu and Song [2012b] suggested a simple method to estimate parameters of model (1) with current status data in the context of Lin et al. [1998] by assuming that the intensity of monitoring times could follow a PH model. They showed that by using two counting processes instead of one as suggested by Martinussen and Scheike [2002], not only the proposed estimator achieves the semiparametric information bound, but also its implementation can be done easily using existing statistical software.

When a large number of covariates are considered, covariates often display more complex effects than the linear format and there may exist interactions between them. In this case, flexible models which could handle potential nonlinear effects of covariates with high dimensionality are greatly desired. To incorporate possible nonlinear covariate effects, Lu and Song [2012a] used polynomial splines and sieve maximum likelihood estimation methods to estimate parameters of the partly linear additive hazards (PLAH) model:

$$\lambda(t|V(s), 0 \le s \le t, X) = \lambda_0(t) + \alpha^T V(t) + \phi_1(X_1) + \cdots + \phi_q(X_q), \tag{2}$$

where $\phi_j(\cdot)$'s are unknown smooth functions, for each $j = 1, \cdots, q$. They focused on efficient estimation in model (2) and used polynomial splines to estimate both the cumulative baseline hazard function with monotonicity constraint and the nonparametric regression functions with no such constraint. They proposed a simultaneous sieve maximum likelihood estimation for regression parameters and nuisance parameters, and showed that the resultant estimator of regression parameters vector is asymptotically normal and achieves the semiparametric information bound. Ma [2011] considered current status data with a cured subgroup where subjects in this subgroup are not susceptible to the event of interest, and assumed that the cure probability satisfies a generalized linear model with a known link function. For subjects sensitive to the event, he used a PLAH model and investigated penalized maximum likelihood estimation.

Another approach in handling high dimensional nonlinear covariate effects and avoiding the "curse of dimensionality" is to use partially linear single-index models. In single-index models time to the event of interest only depends on the nonparametric covariate vector through an unknown smooth function $\psi(\cdot)$ which is called link function. For example, Sun et al. [2008] suggested a partially linear single-index proportional hazards (PLSI-PH) model with right-censored data and Shang et al. [2013] proposed PLSI-PH model to analyze nested case-control (NCC) data. They used a B-spline method to estimate the unknown link function of the single-index term. Here, we propose a partially linear single-index additive hazards (PLSI-AH) regression model to analyze current status data which is a modification of Lu and Song [2012a] by assuming a semiparametric baseline hazard function that depends on X through a single-index $\beta^T X$. The linear covariates are time-dependent and the nonlinear covariates are time-independent. The proposed model can model both linear and nonlinear covariate effects on the hazard and it is a parsimonious model, since it does not use too many parameters. This is particularly important for high dimensional data modelled by a nonparametric function, which might suffer from the curse of dimensionality. Our model reduces the dimension of data through a single-index term. Unlike the PH model for right-censored data, the AH model involves the baseline hazard function in estimation. In our PLSI-AH model, we use B-splines to model the cumulative hazard function and the nonparametric covariate function. Asymptotic properties of the estimators are derived using the theory of counting processes. Another nice feature of our proposed model is that the relative importance of the components of X can be fully characterized by the orientation vector β since the derivative of $\lambda(t|Z)$ with respect to X_i, the ith component of the nonlinear covariate vector X, is proportional to β_i, thus β_i characterizes how fast $\lambda(t|Z)$ changes with X_i, for each $i = 1, \cdots, n$ (Ding et al. [2013]).

The paper is organized as follows. In Section 2, we introduce polynomial splines and indicate required conditions and model assumptions. In Section 3, we present the computing algorithms to implement the proposed estimation procedure. Section 4 concerns the use of counting processes and martingales in order to obtain efficient estimators and achieve their asymptotic properties. Section 5 presents simulation studies. In Section 6, we discuss the application of our proposed model regarding the analysis of the renal function recovery data.

2. Model Description and Estimation

Given the covariate vector $Z(t) = (V(t)^T, X^T)^T$, the PLSI-AHM, in terms of the hazard function of T conditional on the covariate history up to time t is defined as follows

$$\lambda(t|Z(s), 0 \leq s \leq t) = \lambda_0(t) + \alpha^T V(t) + \psi(\beta^T X), \tag{3}$$

where $\alpha = (\alpha_1, \cdots, \alpha_q)^T$ and $\beta = (\beta_1, \cdots, \beta_p)^T$ are q- and p-dimensional regression coefficient vectors, respectively. The parameter β is also called "orientation vector". Based on Huang and Liu [2006] for the purpose of identifiability we have to put some constraints on the orientation vector. First, since the sign of the regression coefficients should be identified, we assume the first component of β to be positive, (i.e. $\beta_1 > 0$) otherwise, $\psi(\beta^T X) = \psi(-(-\beta^T X))$ and we cannot distinguish the two functions $\psi(\cdot)$ and $\psi(-\cdot)$. Second, because any constant scale can be absorbed in $\psi(\cdot)$ we can only estimate the direction of β and the scale of it is not identifiable, so it is required that $\|\beta\| = 1$, where $\|a\| = (a^T a)^{1/2}$ is the Euclidean norm. The function $\lambda_0(t)$ is the baseline hazard function corresponding to $V(t) = 0$ and $X = 0$ which is an unknown and unspecified nonnegative function and $\psi(\cdot)$ is an unknown and smooth link function for the single-index term. On the other hand, since any constant in $\psi(\cdot)$ can be assimilated in the baseline hazard function, $\psi(\cdot)$ is not identifiable. Thus we impose $\psi(0) = 0$. In terms of cumulative hazard function we can write model (3) as follows

$$\Lambda(t|Z(s), 0 \leq s \leq t) = \Lambda_0(t) + \alpha^T V^*(t) + t\psi(\beta^T X), \tag{4}$$

where $\Lambda_0(t) = \int_0^t \lambda(s)ds$ is the cumulative baseline hazard function, and $V^*(t) = \int_0^t V(s)ds$. In the setting of current status data, we do not observe the values of T directly. Our observed data consist of independent samples of $\{C_i, \delta_i, V_i(t), X_i, 0 \leq t \leq C_i\}_{i=1}^n$, drawn from the population $\{C, \delta, V(t), X, 0 \leq t \leq C\}$, where C is the monitoring time that is continuous with the hazard function $\lambda_c(t)$, and $\delta = I(C \leq T)$ is the censoring indicator. We have $\delta = 1$ if the event of interest has not occurred by time C otherwise $\delta = 0$. T and C are independent given the covariate vectors $Z(t)$.

Here we assume that the baseline hazard function is unspecified, so in model (4) we have two unknown functions which are ,in fact, our infinite-dimensional nuisance parameters to be estimated. One of them is the cumulative baseline hazard function, $\Lambda_0(\cdot)$, and the other is the link function related to the single-index term, $\psi(\cdot)$. When we are dealing with infinite-dimensional nuisance parameters, one solution is to use the sieve method. Based on Sun [2006], the key idea behind this method is to approximate the infinite-dimensional nuisance parameter by a sequence of finite-dimensional parameters, that is, the original parameter space is approximated by a sequence of increasing finite-dimensional subspaces (sieves). In our case, the original parameter spaces related to $\Lambda_0(\cdot)$ can be the collection of all nondecreasing functions, and the sieves can be, for instance, collections of nondecreasing and continuous piecewise linear functions. A similar sieve space without the monotonicity constraint can be assumed for $\psi(\cdot)$. For any given finite sample, estimation of the finite-dimensional parameters α and β, along with infinite-dimensional parameters $\Lambda_0(\cdot)$ and $\psi(\cdot)$ can be carried out by maximizing the likelihood function over the product of the parameter spaces for α, β and the sieves. In other words, one only needs to work with a finite-dimensional parameter space by utilizing the sieve method. Considering the idea of Lu and Song [2012a], we use the sieve method along with the B-spline smoothing procedure to estimate $\Lambda_0(\cdot)$ and $\psi(\cdot)$. In the following, for each function we define a sieve space.

Before starting the estimation procedure, we assume the following conditions which are needed in order to establish large sample properties of the estimators.

(A1) The finite-dimentional parameter spaces Θ_1 for α and Θ_2 for β are bounded subsets of $\mathbb{R}^q$ and $\mathbb{R}^p$, respectively. For any $\alpha_0 \neq \alpha$ and $\beta_0 \neq \beta$ we have $P(\alpha_0^T V^* \neq \alpha^T V^* | C) > 0$ and $P(\beta_0^T X \neq \beta^T X) > 0$.

(A2) For $b \geq 1$, Λ_0 and ψ have positive and continuous bth order derivatives on their supports, and (i) If C has the support $[a_c, b_c]$ such that $0 < a_c < b_c < \infty$, then for any $c \in [a_c, b_c]$, $E(V^* | C = c) = 0$. (ii) for the true parameter β_0 and the true function $\psi_0(\beta_0^T X)$, $E(\psi_0(\beta_0^T X)) = 0$.

(A3) (i) Covariates $V^*(t)$ and X have bounded supports which are subsets of $\mathbb{R}^q$ (for $t \in [a_c, b_c]$) and $\mathbb{R}^p$, respectively. (ii) If we denote the distribution of T by F_0 such that $F_0(0) = 0$, then the support of C is strictly contained in F_0, that is for $t_{F_0} = \inf\{t : F_0(t) = 1\}$, $0 < a_c < b_c < t_{F_0}$.

(A4) $\Lambda_0(C) + \alpha_0 V^*(C) + C\psi_0(\beta_0^T X) > 0$ for underlying parameter values α_0, β_0, Λ_0, ψ_0.

(A5) For a small $\varepsilon > 0$ we have $P(T < a_c | C, V^*, X) > \varepsilon$ and $P(T > b_c | C, V^*, X) > \varepsilon$ with probability one.

(A6) For $r \geq 1$, there exists the rth partial derivative of the joint density $f(c, v^*, x)$ of (C, V^*, X) with respect to c or x and it is bounded.

Condition (A1) is to ensure identifiability of the parameters, (A2) implies certain characteristics in order to apply spline smoothing techniques, and (A3) bounds likelihood and score functions away from infinity at the boundaries of the support of the observed event time. Condition (A4) is required for the cumulative hazard function to be positive, and (A5) 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 (A6) requires for the partial score functions (or partial derivatives) of the nonparametric components in the least favorable direction to be close to zero, so that the root-n convergence rate and asymptotic normality of the finite-dimensional estimator can be obtained.

To estimate $\Lambda_0(\cdot)$ at censoring time points C, we assume that the observed values of C are defined in the finite interval $[a_c, b_c]$, then we define the set of B-spline knots as $u_0 < u_1 < \cdots < u_{K_L+1}$ where $u_0 = a_c$, $u_{K_L+1} = b_c$ and K_L is a positive integer denotes the number of B-spline basis functions such that $K_L = O(n^\kappa)$ with $0 < \kappa < 0.5$. The reason of considering 0.5 as the upper limit for κ is that according to Stone [1980], the optimal rate of convergence of a nonparametric estimator in an L_2-norm typically has the form $n^{-p/(2p+1)}$, $p \leq 1$, thus to achieve that we have $\kappa = 1/(2p+1)$ which implies that $0 < \kappa \leq 1/(2p+1) < 0.5$. Then we make a partition of $[a_c, b_c]$ using the B-spline knots as follows: $[u_0, u_1), [u_1, u_2), \cdots, [u_{K_L-1}, u_{K_L}), [u_{K_L}, u_{K_L+1}]$. We restrict the choice of partitions by letting $\max_{1\leq i\leq K_L+1}(u_i - u_{i-1}) = O(n^{-\kappa})$. Let $\mathcal{L}_n$ be the space of polynomial splines of order $\rho_L \geq 1$, where each functional element of this space is a polynomial of order ρ_L on each sub interval of our partition, and for $\rho_L \geq 2$ and $0 \leq r \leq \rho_L - 2$, each functional element of this space is r times continuously differentiable on $[a_c, b_c]$. Suppose $\mathbb{L}_n$ is the collection of nonnegative and nondecreasing functions $\Lambda_0(\cdot)$ on $[a_c, b_c]$, then for any ,$\Lambda_0 \in \mathbb{L}_n$, we can write

$$\Lambda_0(C) = \sum_{k=1}^{df_L} \tau_k L_k(C) = \tau^T L(C), \tag{5}$$

where $L(C) = (L_1(C), \cdots, L_{df_L}(C))^T$ is the vector of B-spline basis functions with $L_k(C) \in \mathcal{L}_n$ for each $k = 1, \cdots, df_L$, $\tau = (\tau_1, \cdots, \tau_{df_L})^T$ is the vector of B-spline coefficients and $df_L = K_L + \rho_L$ is the degree of freedom for B-spline. In order to $\Lambda_0(C)$ be nonnegative and nondecreasing we have to put a constraint on the coefficients of the basis functions, that is $0 \leq \tau_1 \leq \tau_2 \leq \cdots \leq \tau_{df_L}$. Based on Schumaker [1981] , both non-negativity and monotonicity of $\Lambda_0(C)$ are guaranteed by this constraint.

Then, to estimate the other unknown function, $\psi(\cdot)$, first we consider an initial value for the orientation vector β, say $\beta_{(0)}$, then assume that the interval $[a_{xb}, b_{xb}]$ is the support for $\beta_0^T X$ values. In the same procedure as for Λ_0, we define the set of knots as $a_{xb} = w_0 < w_1 < \cdots < w_{K_B+1} = b_{xb}$ where K_B shows the number of B-spline basis functions which is a positive integer and $K_B = O(n^\kappa)$ and $\max_{1\leq i\leq K_B+1}(w_i - w_{i-1}) = O(n^{-\kappa})$. Suppose $[w_0, w_1), [w_1, w_2), \cdots, [w_{K_B-1}, w_{K_B}), [w_{K_B}, w_{K_B+1}]$ is the partition of $[a_{xb}, b_{xb}]$ and $\mathcal{B}_n$ be the space of polynomial splines of order $\rho_B \geq 1$, with the same properties as $\mathcal{L}_n$.

Let Ψ_n be the collection of $\psi(\cdot)$ functions on $[a_{xb}, b_{xb}]$, so we can write

$$\psi(\beta^T X) = \sum_{l=1}^{df_B} \gamma_l B_l(\beta^T X) = \gamma^T B(\beta^T X), \tag{6}$$

where $B(\beta^T X) = (B_1(\beta^T X), \cdots, B_{df_B}(\beta^T X))^T$ is the vector of local normalized B-spline basis functions, $B_l(\beta^T X) \in \mathcal{B}_n$, $\gamma = (\gamma_1, \cdots, \gamma_{df_B})^T$ is the vector of B-spline coefficients and $df_B = K_B + \rho_B$ is the degree of freedom for this B-spline. Here we also have to fulfill the constraints on β that we mentioned for identifiability purposes (i.e. $\beta_1 > 0$ and $\|\beta\| = 1$).

Since our main purpose is to estimate α and β, then any reasonable choice of partitions should work well. Under suitable smoothness assumptions $\Lambda_0(\cdot)$ and $\psi(\cdot)$ can be well approximated by functions in $\mathcal{L}_n$ and $\mathcal{B}_n$, respectively. Therefore, we have to find a member of $\mathcal{L}_n$ and $\mathcal{B}_n$ along with values for α and β that maximize the semiparametric log-likelihood function.

3. Implementation

Regarding the relation between the cumulative hazard and the survival functions, i.e. $\Lambda(t|Z) = -\log\{S(t|Z)\}$, we can rewrite model (4) at observation time C as follows

$$-\log(p) = \Lambda_0(C) + \alpha^T V^*(C) + C\psi(\beta^T X),$$

and through B-spline estimators for $\Lambda_0(C)$ and $\psi(\beta^T X)$ it is equivalent to

$$\log(p) = -\tau^T L(C) - \alpha^T V^*(C) - C\gamma^T B(\beta^T X), \tag{7}$$

where $p = p(C) = S(C|Z) = S(C|Z(s), 0 \le s \le t)$. Since for subject i, $i = 1, \cdots, n$, we have $S(C_i|Z_i) = P(C_i \le T_i) = E[I(C_i \le T_i)|Z_i] = E(\delta_i|Z_i)$. By assuming δ_i as a binary response, we can consider model (7) as a generalized linear model (GLM) with linear predictor $-\xi = -\{\tau^T L(C) + \alpha^T V^*(C) + C\gamma^T B(\beta^T X)\}$ and log-link. Then we use GLM methods, available in various computer software packages, to estimat parameters α, β, τ and γ. We used "glm" function in the R package to do that. The estimated values obtained in this step are considered as the initial values of the parameters of our model for the next step which is maximizing the semiparametric log-likelihood function subject to the constraints.

For given $Z_i(t)$'s, $i = 1, \cdots, n$, the likelihood function is proportional to

$$L(\alpha, \beta, \Lambda_0, \psi) = \prod_{i=1}^{n} [S\{C_i|Z_i(C_i)\}]^{\delta_i} [1 - S\{C_i|Z_i(C_i)\}]^{1-\delta_i},$$

where $S(t|Z_i(t)) = \exp\{-\Lambda(t|Z_i(t))\}$ is the survival function of failure time T conditional on covariates $Z_i(t) = (V_i^T(t), X_i^T)^T$. Based on (4) we have $S(t|Z_i(t)) = \exp\{-\Lambda_0(t) - \alpha^T V_i^*(t) - t\psi(\beta^T X_i)\}$. Thus we can write the semiparametric log-likelihood function as follows

$$\begin{aligned}\ell(\alpha, \beta, \Lambda_0, \psi) = \sum_{i=1}^{n} [&\delta_i \log\{\exp(-\Lambda_0(C_i) - \alpha^T V_i^*(C_i) - t\psi(\beta^T X_i))\} \\ &+ (1 - \delta_i) \log\{1 - \exp(-\Lambda_0(C_i) - \alpha^T V_i^*(C_i) - C_i\psi(\beta^T X_i))\}].\end{aligned}$$

Since we can not observe the exact failure times and only the values of $S(t)$ at the observation times C_i affect the likelihood function, without loss of generality, we can focus only on the maximization of $l(\alpha, \beta, \Lambda_0, \psi)$ over all non-increasing step functions with jumps only at the C_i. By the sieve method, we plug in the B-spline approximations of $\Lambda_0(\cdot)$ and $\psi(\cdot)$ obtained from (5) and (6) into the semiparametric log-likelihood function. So we have the log likelihood function as follows

$$\begin{aligned}\ell(\alpha, \beta, \tau, \gamma) = \sum_{i=1}^{n} [&\delta_i \log\{ \exp(-\tau^T L(C_i) - \alpha^T V_i^*(C_i) - C_i\gamma^T B(\beta^T X_i))\} \\ &+ (1 - \delta_i) \log\{ 1 - \exp(-\tau^T L(C_i) - \alpha^T V_i^*(C_i) - C_i\gamma^T B(\beta^T X_i))]. \end{aligned} \tag{8}$$

Now we can estimate the parameters of our regression model, $(\alpha, \beta, \tau, \gamma)$, by maximizing the log-likelihood function given in (8) which has a parametric form after using B-spline approximated values of the infinite-dimensional nuisance parameters. To maximize (8) we used sieve method through an iterative algorithm subject to some constraints on coefficients τ and β which we mentioned in Section 2, i.e. $0 \le \tau_1 \le \tau_2 \le \cdots \le \tau_{df_L}$ for non-negativity and monotonicity of $\Lambda_0(C)$ and $\beta_1 > 0$ and $\|\beta\| = 1$ for the purpose of identifiability in $\psi(X^T \beta)$.

Suppose we want to maximize the log-likelihood function $\ell(\theta)$ as our objective function subject to some constraints $g_u(\theta)$ and obtain an estimator for θ. For this respect, we have to use one of the methods for solving a constrained optimization problem. The method we used here is known as the adaptive barrier algorithm which is defined as follows.

Definition: (Barrier algorithm) In the case of barrier methods, a term is added that favors points in the interior of the feasible region over those near the boundary. Consider the problem of minimizing $\{f(\theta) : \theta \in S_\theta\}$, where f is a continuous function on $\mathcal{R}^n$ and S_θ is a constraint set in $\mathcal{R}^n$. Here, and in most applications, S_θ is defined explicitly by a number of functional constraints such that $S_\theta = \{\theta : g_u(\theta) \in \mathcal{G}_\theta, u = 1, \cdots, U\}$. In this case, we define an augmented objective function as follows

$$f_a(m, \theta) = f(\theta) + m \sum_{u=1}^{M} \frac{1}{g_u(\theta)},$$

where m is the barrier parameter. f_a is only valid for those interior points that all constraints are strictly satisfied $g_u(\theta) \geq 0$ for all $u = 1, \cdots, M$. $f_a(m, \theta)$ is called barrier function.

The algorithm we used to maximize ℓ iterates the following steps.

- Step 0: Start with initial values $\beta^{(0)}$, $\alpha^{(0)}$, $\tau^{(0)}$, $\gamma^{(0)}$, where α_0, τ_0, γ_0 are obtained from the GLM method.

- Step 1. Given current values $\alpha^{(k)}$, $\tau^{(k)}$, $\gamma^{(k)}$ update the value of $\beta^{(k)}$ by maximizing the log-likelihood function given in (8) subject to $1 - \sum_{l=2}^{p} \beta_l^2 > 0$ which satisfies the constraints $\beta_1 > 0$ and $\|\beta\| = \sqrt{\sum_{l=1}^{p} \beta_l^2} = 1$. We use a non-linear constrained optimization algorithm named Barrier method which is implemented by "constrOptim.nl" function in R package "alabama", to maximize ℓ with respect to β. So, we obtain $\beta^{(k+1)}$ in this step.

- Step 2. Having $\beta^{(k+1)}$, update the values of $\alpha^{(k)}$, $\tau^{(k)}$, $\gamma^{(k)}$ simultaneously through GLM with the binary response δ, log link and linear predictor $\xi = -{\tau^{(k)}}^T L(C) - {\alpha^{(k)}}^T V^*(C) - C{\gamma^{(k)}}^T B({\beta^{(k+1)}}^T X)$. Then by letting $\omega = (\tau^T, \alpha^T, \gamma^T)^T$, we use Newton-Raphson method to obtain $\omega^{(k+1)} = ({\tau^{(k+1)}}^T, {\alpha^{(k+1)}}^T, {\gamma^{(k+1)}}^T)^T$ which is implemented by "nlminb" function in R. "nlminb" uses a quasi-Newton algorithm that fills the same niche as the "L-BFGS-B" method in "optim". It seems a bit more robust than "optim" in that it is more likely to return a solution in marginal cases where "optim" will fail to converge.

- Step 3. In the same procedure as Step 1, considering $0 \leq \tau_1 \leq \tau_2 \leq \cdots \leq \tau_{df_L}$ as the constraint on τ, we update the value of $\tau^{(k+1)}$ using "constrOptim.nl" and obtain $\tau^{(k+2)}$.

- Step 4. Update $\alpha^{(k+1)}$ based on the estimated increasing $\Lambda(C) = {\tau^{(k+2)}}^T L(C)$ and $\psi({\beta^{(k+1)}}^T)X) = {\gamma^{(k+1)}}^T B({\beta^{(k+1)}}^T)X)$. Again we use the Newton-Raphson method implemented by "nlminb" in R to obtain $\alpha^{(k+2)}$.

- Step 5. Repeat Step 1 to 4 until a certain convergence criterion is met.

Finally, we consider $\beta^{(k+1)}$, $\tau^{(k+2)}$, $\alpha^{(k+2)}$ and $\gamma^{(k+1)}$ as the estimated values for β, τ, α and γ, respectively.

4. Inference

4.1 Efficient Estimation

Efficient score and information bounds for the additive hazards model with current status data have been studied by Martinussen and Scheike [2002]. In their model there is a linear component plus the unknown cumulative baseline hazard function as the nuisance parameter. In our paper, we are faced with two infinite-dimensional components along with the linear part.

In our semiparametric model, the observed data $C_1, \cdots, C_n$ are iid random vectors which have density belonging to the class $\mathcal{P} = \{P(c, \alpha, \beta, \Lambda, \psi), \alpha \in \Theta_1, \beta \in \Theta_2, \Lambda \in \mathbb{L}, \psi \in \Psi\}$ where $\Theta_1 \subseteq \mathbb{R}^q$ and $\Theta_2 \subseteq \mathbb{R}^p$, Ψ is some set of smooth real-valued functions ψ of $\beta^T X$, satisfying $E[\delta\psi(\beta^T X)] = 0$ and $E[\delta\psi^2(\beta^T X)] < \infty$ and $\mathbb{L}$ is the collection of absolutely continuous increasing functions on $\mathbb{R}^+$. One way to handle two nuisance parameters is projecting the score of the finite-dimensional component onto the space orthogonal to the sumspace of the tangent spaces of the two nuisance parameters. But since we have a nonparametric component with a coefficient vector inside, the projection is not guaranteed and thus this method is not practical in our case. Although it is possible to obtain the efficient score for our model more directly, we consider $\psi(\cdot)$ as a B-spline function for a reasonable approximation, and take $\Lambda(\cdot)$ as the only infinite-dimensional parameter of the model. So by replacing $\psi(\beta^T X)$ with the B-spline function obtained in (6), our model belongs to the class $\mathcal{P} = \{P(c, \alpha, \beta, \gamma, \Lambda), \alpha \in \Theta_1, \beta \in \Theta_2, \gamma \in \Theta_3, \Lambda \in \mathbb{L}\}$ where $\Theta_3 \subseteq \mathbb{R}^{df_B}$, Λ is the only infinite-dimensional parameter. Using the projection theory, it is not hard to derive the efficient score and the information bound for (α, β, γ).

Based on observations $\{C_i, \delta_i, V_i(t), X_i, 0 \le t \le C_i\}_{i=1}^n$ we define two counting processes $N_{1i}(t) = \delta_i I(C_i \le t)$ and $N_{2i}(t) = (1-\delta_i) I(C_i \le t)$ which are step functions, zero at time zero, with jumps of size one only and no two component processes jumping at the same time. The process $N_{1i}(t)$ jumps when subject i is monitored at time t and and found to be failure-free, and $N_{2i}(t)$ jumps when subject i is monitored at time t and found that it has experienced failure. Then let $N_i(t) = N_{1i} + N_{2i} = \delta_i I(C_i \le t) + (1 - \delta_i) I(C_i \le t)$, and also let $Y_i(t) = I(t \le C_i)$ be the at risk process for time point t. Here we assume the distribution of C is continuous with hazard function
$nu_i = nu_i(t) = \lambda_c(t|Z_i)$ at time t conditional on covariate vector $Z_i = Z_i(t)$ for each $i = 1, \cdots, n$.

Noted in Martinussen and Scheike [2002], the intensities of N_{1i} and N_{2i} are as follows

$$\lambda_{N_{1i}}(t) = Y_i(t)\nu_i(t)p_i(t) = Y_i\nu_i p_i$$
$$\lambda_{N_{2i}}(t) = Y_i(t)\nu_i(t)(1 - p_i(t)) = Y_i\nu_i(1 - p_i),$$

where, as we mentioned before, $p_i = p_i(t; \Lambda, \theta) = S(t|Z_i) = e^{-\Lambda(t|Z_i)}$ for $0 < t \le C_i$, $\theta = (\beta^T, \alpha^T, \gamma^T)^T$ and $i = 1, \cdots, n$. If we replace $\psi(\beta^T X)$ by the B-spline obtained from (6), we have $p_i = \exp\{-\Lambda_0(t) - \alpha^T V_i^*(t) - t\gamma^T B(\beta^T X_i)\}$.

Using $\lambda_{N_{1i}}$ and $\lambda_{N_{2i}}$, we define M_{1i} and M_{2i} as their corresponding compensated counting processes, respectively as:

$$M_{1i}(t) = N_{1i}(t) - \int_0^t Y_i(s)\nu_i(s)p_i(s)ds$$
$$M_{2i}(t) = N_{2i}(t) - \int_0^t Y_i(s)\nu_i(s)(1 - p_i(s))ds.$$

Based on Martinussen and Scheike [2002], M_1 and M_2 are martingales.

The log-likelihood functions can be written as follows

$$\ell = \ell(\alpha, \beta, \Lambda, \psi) = \sum_{i=1}^{n} \{ \int (\log p_i)\, dN_{1i} + \int (\log (1 - p_i))\, dN_{2i} \}. \tag{9}$$

Parametric submodel for the nuisance parameter Λ is a mapping of the form $\eta \to \Lambda_{(\eta)}$ where $\{\Lambda_{(\eta)} : \eta \in \mathbb{R}\}$ characterizes Λ by a finite-dimensional parameter η in which $\Lambda = \Lambda_{(0)}$ and

$$\frac{\partial \Lambda_{(\eta)}}{\partial \eta}(t) = a(t) = a. \tag{10}$$

Having censoring time C observed, the marginal score vector for θ is obtained by partially differentiating $\ell(\theta, \Lambda_{(\eta)})$ given in (9) with respect to θ such that

$$S_\theta = \frac{\partial \ell}{\partial \theta} = (S_\alpha^T, S_\beta^T, S_\gamma^T)^T,$$

where $S_\alpha = \frac{\partial \ell}{\partial \alpha}$, $S_\beta = \frac{\partial \ell}{\partial \beta}$ and $S_\gamma = \frac{\partial \ell}{\partial \gamma}$. So by knowing that $dN_{1i}(t) = dM_{1i}(t) + Y_i \nu_i p_i dt$ and $dN_{2i}(t) = dM_{2i}(t) + Y_i \nu_i (1 - p_i) dt$ we have

$$S_\theta = \int U^* \left(\frac{p}{1-p} dM_2 - dM_1 \right),$$

where $p = p(t; \Lambda, \theta) = S(t|Z) = e^{-\Lambda(t|Z)}$ for $0 < t \le C$ and

$$U^* = U^*(t) = ((V^*(t))^T, (tX\gamma^T B'(\beta^T X))^T, (tB(\beta^T X))^T)^T$$

is a $(q + p + df_B) \times 1$ vector.

Then for Λ we have $S_\Lambda(a) = \frac{\partial l}{\partial \Lambda} \times \frac{\partial \Lambda_\eta}{\partial \eta}$, so considering (10) we have the score operator associated with the cumulative hazard function Λ as follows

$$S_\Lambda(a) = \int a(t) \left(\frac{p}{1-p} dM_2 - dM_1 \right).$$

Under conditions (A1) to (A6), the efficient score for the finite-dimensional parameter θ is the difference between its score vector, S_θ, and the score for a particular submodel of the nuisance parameter, $S_\Lambda(a)$. The particular submodel is the one with the property that the difference is uncorrelated with the scores for all submodels of the nuisance parameters. Therefore the efficient scores for θ is as follows

$$S_\theta^* = S_\theta - S_\Lambda(a).$$

Define $L_2(P_C) = \{a : E[\|a(C)\|^2 p(C)(1 - p(C))^{-1}] < \infty\}$, and let $A_\Lambda = \{S_\Lambda(a) : a \in L_2(P_C)\}$. Then to obtain the efficient score and information bound, we project S_θ onto the space A_Λ and try to find from all functions $a(t) \in L_2(P_C)$ the one for which $S_\Lambda(a)$ is the "closest" to S_θ. Call this a-function $a^*(t) = a^*$ with score $S_\Lambda(a^*)$. By the L_2 theory, this closest score is the one for which $S_\theta - S_\Lambda(a^*)$ is orthogonal to any other $S_\Lambda(a)$. Thus,

$$E(S_\theta^* S_\Lambda) = E\{[S_\theta - S_\Lambda(a^*)] S_\Lambda(a)\} \tag{11}$$

should equal 0, for every a. (11) is the orthogonality equation for θ and it is equivalent to

$$E\{\int (U^* - a^*)(\frac{p}{1-p} dM_2 - dM_1) \int a(\frac{p}{1-p} dM_2 - dM_1)\} = 0, \tag{12}$$

then equation (12) is equivalent to

$$E\left[\int (U^* - a^*)a\frac{p^2}{(1-p)^2}Y\nu(1-p)dt + \int (U^* - a^*)aY\nu p dt\right]$$
$$= \int \left(a\left\{E\left[U^*Y\nu(\frac{p}{1-p})\right] - a^*E\left[Y\nu(\frac{p}{1-p})\right]\right\}\right)dt$$
$$= 0,$$

so we obtain

$$a^* = E\left[U^*Y\nu\frac{p}{1-p}\right]E^{-1}\left[Y\nu\frac{p}{1-p}\right].$$

By plugging in a^* into (12) we have the desired efficient score as follows

$$S^*_\theta = \int \left\{U^* - E\left[U^*Y\nu\frac{p}{1-p}\right]E^{-1}\left[Y\nu\frac{p}{1-p})\right]\right\}\left\{\frac{p}{1-p}dM_2 - dM_1\right\}. \quad (13)$$

The empirical version of (13) gives us the score function of interest, namely

$$S(\theta, \Lambda) = \sum_{i=1}^{n}\int \left\{U_i^* - \frac{S_1^{(\theta)}}{S_0^{(\theta)}}\right\}\left\{\frac{p_i}{1-p_i}dM_{2i} - dM_{1i}\right\},$$

where

$$S_u^{(\theta)} = S_u^{(\theta)}(t) = S_u^{(\theta)}(t) = \sum_i \frac{p_i}{1-p_i}Y_i\nu_i[U_i^*]^{\otimes u}, \text{ for } u = 0, 1,$$

with $\otimes$ denotes Kronecker operation, defined as $b^{\otimes 0} = 1$, $b^{\otimes 1} = b$ and $b^{\otimes 2} = bb^T$. Since $\nu_i = \nu_i[t|Z_i(t)]$ is an unknown function of the covariate vector, $Z_i(t) = (V_i^T(t), X_i^T)^T$, so $S_u^{(\theta)}$ has to be estimated. In the same way as Martinussen and Scheike [2002], we suggest the simple kernel estimator for ν_i as follows

$$\hat{S}_u^{(\theta)} = \sum_{i=1}^{n}\int \frac{\hat{p}_i(s)}{1-\hat{p}_i(s)}Y_i(s)[U_i^*]^{\otimes u}\hat{\nu}_i(s|Z_i(s))ds, \text{ for } u = 0, 1,$$

where $\hat{p}_i(s) = \exp\{-\hat{\Lambda}(s) - \alpha^T V_i^*(s) - s\gamma^T B(\beta^T X_i)\}$ and $\hat{\nu}_i(s|Z_i(s))ds = K_b(s - t)dN_i(s)$, $K_b(\cdot) = (1/b)K(\cdot/b)$, and $b > 0$ is the bandwidth of the kernel estimator. We assume that $\int K_b(u)dt = 1$, $\int uK_b(u)dt = 0$ and the kernel has compact support. Our sieve estimator of θ is equivalent to the solution of the estimated empirical efficient score equation

$$S(\theta, \hat{\Lambda}) = 0, \quad (14)$$

where $\hat{\Lambda} = \hat{\Lambda}(t)$ is assumed to be a predictable estimator of Λ such that $\Lambda - \hat{\Lambda} = o_p(n^{-\frac{1}{4}})$.

The information for $\theta_0 = (\beta_0^T, \alpha_0^T, \gamma_0^T)^T$ is given as follows

$$I(\theta_0) = E(S^*_{\theta_0})^{\otimes 2} = E[\langle \int \left\{U^* - E(U^*Y\nu\frac{p}{1-p})E^{-1}\left(Y\nu\frac{p}{1-p}\right)\right\}\{\frac{p}{1-p}dM_2 - dM_1\}\rangle]$$
$$= E[\int \left\{U^* - E(U^*Y\nu\frac{p}{1-p})E^{-1}\left(Y\nu\frac{p}{1-p}\right)\right\}^{\otimes 2}\frac{p}{1-p}Y\nu dt].$$

We can estimate the semiparametric efficient information bound as follows be taking as $\theta = \hat{\theta}$,

$$\hat{I}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\int D_\theta\left(U_i^* - \frac{\hat{S}_1^{(\theta)}}{\hat{S}_0^{(\theta)}}\right)\{\frac{\hat{p}_i}{1-\hat{p}_i}dN_{2i} - dN_{1i}\}$$
$$- \frac{1}{n}\sum_{i=1}^{n}\int\left\{U_i^* - \frac{\hat{S}_1^{(\theta)}}{\hat{S}_0^{(\theta)}}\right\}U_i^{*T}\frac{\hat{p}_i}{(1-\hat{p}_i)^2}dN_{2i},$$

where

$$D_\theta\left(U_i^* - \frac{\hat{S}_1^{(\theta)}}{\hat{S}_0^{(\theta)}}\right) = -\frac{\hat{S}_0\bar{S}_1 - \bar{S}_0\hat{S}_1}{\hat{S}_0^{\ 2}}, \quad \text{and}$$
$$\bar{S}_u = -\sum_i \frac{\hat{p}_i}{(1-\hat{p}_i)^2}Y_i[U_i^*]^{\otimes(u+1)}\hat{\nu}_i, \quad u = 0, 1.$$

4.2 Asymptotic Properties

Since $S(\theta_0, \hat{\Lambda})$ is a martingale so it still has asymptotic mean zero, the score equation in (14) will produce consistent estimators even when $\hat{\Lambda}$ is not a consistent estimator of Λ. Under regularity conditions (A1) to (A6) and with a consistent estimator $\hat{\Lambda}$ of Λ, If κ satisfies the restrictions of $0.25/s < \kappa < 0.5$ and $\kappa(s+r) > 0.5$, where s and r are orders of differentiation postulated in (A2) and (A6) respectively, then using the central limit theorem for martingales we can say that $n^{1/2}S(\theta_0, \hat{\Lambda})$ converges in distribution to a normal distribution with mean zero and covariance matrix Σ_1 which is consistently estimated by

$$\hat{\Sigma}_1 = \frac{1}{n}\sum_{i=1}^{n}\int\left(U_i^* - \frac{\hat{S}_1^{(\theta)}}{\hat{S}_0^{(\theta)}}\right)\left(U_i^* - \frac{\hat{S}_1^{(\theta)}}{\hat{S}_0^{(\theta)}}\right)^T\left\{\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right)^2 dN_{2i} + dN_{1i}\right\},$$

where $\hat{\Sigma}_1$ converges in probability to Σ_1 and therefore, we have $n^{1/2}(\hat{\theta} - \theta_0)$ converges in distribution to a mean zero normal distribution with covariance matrix $\Sigma = I^{-1}(\theta_0)\Sigma_1 I^{-1}(\theta_0)$. The robust sandwich estimator of the variance is given by $\hat{\Sigma} = \hat{I}^{-1}(\hat{\theta})\hat{\Sigma}_1\hat{I}^{-1}(\hat{\theta})$.

With the consistent estimator of Λ we can conclude that $\hat{\Sigma}_1 = \hat{I}(\hat{\theta}) + o_p(1)$, thus, $n^{1/2}(\hat{\theta} - \theta_0)$ converges in distribution to a mean zero random vector with covariance matrix $I^{-1}(\theta_0)$ estimated by $\hat{I}^{-1}(\hat{\theta})$. Therefore, our obtained estimators are efficient.

5. Simulation Studies

To assess the finite-sample performance of the methods we explained in the previous sections, we present a simulation study for the estimation of parameters of our model. The failure time, T, is generated from model (3) that is from exponential distribution with rate $\lambda(t|V, X) = \lambda_0(t) + \alpha_0^T V + \psi(\beta_0^T X)$ where $\lambda_0(t)$ is assumed to be a constant equals $\lambda_0 = 7$, $\alpha_0 = (0.5, -1)^T$, $\beta_0 = (2, -1, -1)^T/\sqrt{6}$ and the $p = 2$ dimensional linear covariate vector $V = (V_1, V_2)^T$ and the $q = 3$ nonlinear covariate vector $X = (X_1, X_2, X_3)^T$ where $X_1, X_2, X_3 \sim \text{Uniform}(-4, 4)$, $V_1 \sim \text{Uniform}(1, 2) - 1.5$ and $V_2 \sim \text{Bernoulli}(0.5) - 0.5$. The single-index function is defined as $\psi(\beta_0^T X) = 5 \times \sin(\beta_0^T X)$. The covariates satisfy conditions (A1) and (A2) since $E(V_1) = E(V_2) = E(\psi(\beta_0^T X)) = 0$. The censoring time, C, is generated from exponential distribution with rate $\lambda_{c0} = 4$, confined in interval $[0.1, 1.2]$. As explained before, we used B-splines to approximate the unknown curves. Here we considered the number of knots equals 6 for Λ and 8 for $\psi(\cdot)$. The simulation is replicated 500 and 1000 times for sample sizes equal to n = 400 and 800, respectively.

Table 1: Simulation Results for PLSI-AHM

	Simulation times	S=500		S=1000	
True values	Sample size	n=400	n=800	n=400	n=800
β_1= 0.816	$\hat{\beta}_1$	0.804	0.805	0.813	0.813
	(sd)	(0.043)	(0.043)	(0.025)	(0.025)
β_2= -0.408	$\hat{\beta}_2$	-0.409	-0.409	-0.410	-0.410
	(sd)	(0.062)	(0.062)	(0.038)	(0.038)
β_3= -0.408	$\hat{\beta}_3$	-0.419	-0.418	-0.408	-0.408
	(sd)	(0.063)	(0.062)	(0.036)	(0.037)
α_1= 0.5	$\hat{\alpha}_1$	0.396	0.410	0.439	0.456
	(sd)	(1.361)	(1.345)	(0.885)	(0.898)
α_2= -1	$\hat{\alpha}_2$	-0.910	-0.912	-0.897	-0.896
	(sd)	(0.888)	(0.839)	(0.555)	(0.568)

Table 1 summarizes the resulted estimates for α and β with the standard deviations in the brackets. As it is shown in the table, the estimated values are very close to the true parameter values.

6. Real Data Analysis

Acute kidney injury (AKI) is a typical kidney disease syndrome with substantial impact on both short and long-term clinical outcomes. Identifying risk factors associated with renal recovery in patients requiring renal replacement therapy (RRT) can help clinicians to develop strategies to prevent non-recovery and improve patient's quality of life. At University of Michigan Hospital, a study was conducted to characterize survival and renal outcomes of hospitalized patients with AKI requiring RRT, both during hospitalization and up to 1 year following RRT initiation. The primary outcome of interest was the recovery of renal function to the point of no longer necessitating maintenance dialysis in patients who initially required RRT due to AKI during the index hospitalization. In this study they conducted a single-center, retrospective analysis of 170 hospitalized adult patients with AKI attributed to acute tubular necrosis who required inpatient initiation of RRT. Data collection included patient characteristics, laboratory data, details of hospital course and degree of fluid overload at RRT initiation. The primary outcome was recovery of renal function to dialysis independence. For each of the patients, his/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]. Lu and Song [2012a] analyzed this data using PL-AH model.

Here we applied our proposed PLSI-AH model to assess the relationship between the hazard of occurrence of renal recovery and the clinical factors, including baseline serum creatinine level, use of vasopressor, age and gender. Let T be time which is the number of days from the time of starting dialysis to the date of renal function recovery, and let C be monitoring time given as of the time of hospital discharge. 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 (Age, varying between 17 and 94 years) and gender (Gender, binary). VP is coded as 1 for Yes and 0 for No. Gender is coded as 1 for male and 0 for female.

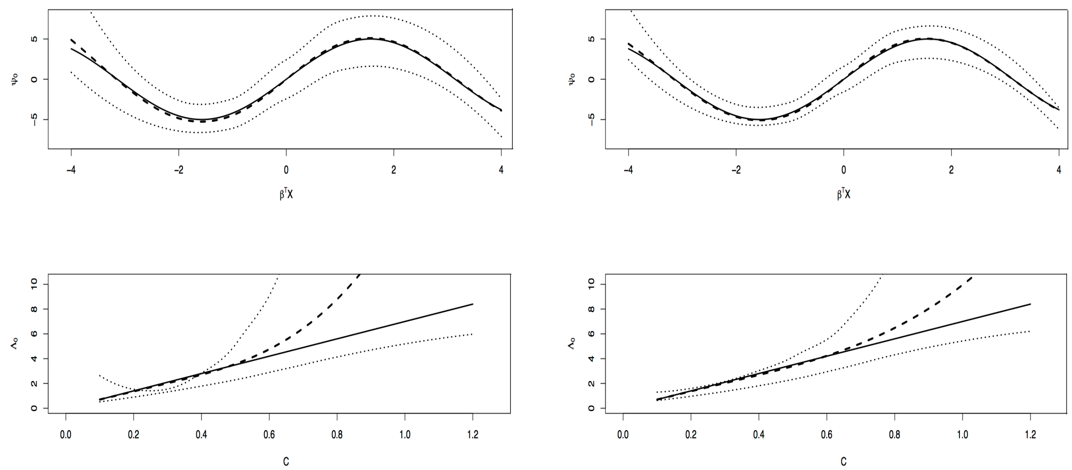

Figure 1: Sieve B-spline estimation of two curves Λ and ψ. The solid lines stand for the true curves, the dashed lines illustrate the estimated curves and the doted lines show the 95% confidence intervals. The left figure depicts the resulted curves with sample size $n = 400$ and simulation times $S = 500$ and the right one shows the results with $n = 800$ and $S = 1000$

Denote the time of starting dialysis to the date of renal recovery by T, Time of hospital discharge by C, Use of vasopressor by V, Gender by V_2, Baseline serum creatinine level by X_1, Age by X_2 and the indicator of renal recovery at the time of discharge by $\delta = I(C \leq T)$. Let $\lambda(t; V, X)$ be the hazard function of recovery time, T, where covariate vectors $V = (V_1, V_2)^T$ and $X = (X_1, X_2)^T$. Model (3) is applied to establish a relationship between the hazard function of T and the four covariates. The parameter estimates are $\hat{\alpha}_1 = -0.085$, $\hat{\alpha}_2 = 0.0104$, $\hat{\beta}_1 = 0.856$ and $\hat{\beta}_2 = -0.517$ which are close to what we expected from previous analysis. More analysis of this data set is postponed to our future work.

References

Ding, K., Kosorok, M. R., and Zeng, D. (2013). On the local and stratified likelihood approaches in single-index hazards model. *Communications in Mathematics and Statistics*, 1:115–132.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*, volume 19. Springer, Boston.

Heung, M., Wolfgram, D. F., Kommareddi, M., Hu, Y., Song, P. X.-K., and 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. *Nephrology Dialysis Transplantation*, 27(3):956–961.

Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics*, 24(2):540–568.

Huang, J. Z. and Liu, L. (2006). Polynomial spline estimation and inference of proportional hazards regression models with flexible relative risk form. *Biometrics*, 62(3):793–802.

Jewell, N. P. and van der Laan, M. (1996). Generalizations of current status data with applications. In *Lifetime Data: Models in Reliability and Survival Analysis*, pages 141–148. Springer, New York.

Lin, D., Oakes, D., and Ying, Z. (1998). Additive hazards regression with current status data. *Biometrika*, 85(2):289–298.

Lin, D. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika*, 81(1):61–71.

Lu, X. and Song, P. X.-K. (2012a). Efficient sieve estimation of the partly linear additive hazards regression model with current status data. *Unpublished paper*.

Lu, X. and Song, P. X.-K. (2012b). On efficient estimation in additive hazards regression with current status data. *Computational Statistics* & *Data Analysis*, 56(6):2051–2058.

Ma, S. (2011). Additive risk model for current status data with a cured subgroup. *Annals of the Institute of Statistical Mathematics*, 63(1):117–134.

Martinussen, T. and Scheike, T. H. (2002). Efficient estimation in additive hazards regression with current status data. *Biometrika*, 89(3):649–658.

McKeague, I. W. and Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika*, 81(3):501–514.

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

Shang, S., Liu, M., Zeleniuch-Jacquotte, A., Clendenen, T. V., Krogh, V., Hallmans, G., and Lu, W. (2013). Partially linear single index cox regression model in nested case-control studies. *Computational Statistics & Data Analysis*, 67:199–212.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, 10(4):1348–1360.

Sun, J. (2006). *The Statistical Analysis of Interval-Censored Failure Time Data*. Springer, New York.

Sun, J., Kopciuk, K. A., and Lu, X. (2008). Polynomial spline estimation of partially linear single-index proportional hazards regression models. *Computational Statistics* & *Data Analysis*, 53(1):176–188.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 38(3):290–295.