Statistica Sinica **14**(2004), 885-905

PROPORTIONAL HAZARDS REGRESSION MODELS WITH UNKNOWN LINK FUNCTION AND TIME-DEPENDENT COVARIATES

Wei Wang

Harvard University

Abstract: Proportional hazards regression models assume that the covariates affect the survival time through a link function and an index which is a linear function of the covariates. We study the situation when the link is unspecified and some covariates are time-dependent. Due to the nature of irregular designs, oftentimes the history of the time-dependent covariates is not observable. We propose a two-stage approach to account for the missingness. In the first stage, we impute the missing time-dependent covariates using functional data analysis methods. In the second stage, we perform a two-step iterative algorithm to estimate the unknown link function. Asymptotic properties are derived for the non-parametric estimated link function when time-dependent covariates history is observable. The approach is illustrated through several simulations and a data set of a prostate cancer clinical trial.

Key words and phrases: Functional data analysis, non-parametric smoothing, proportional hazards regression, time-dependent covariates.

1. Introduction

In clinical studies, one is often interested in the relationship between a survival outcome and some predictive covariates. Proportional hazards regression models introduced by Cox (1972) have been widely used to examine this relationship for censored data. Let T be the failure time, C be the censoring time, $X = \min(T, C)$ be the observed event-time and $\Delta = I\{T \leq C\}$ be the censoring indicator on the study interval $[0, \tau]$. Let $Z(t)$ be a q-dimensional covariate vector measured at time t which may affect the survival distribution. Some covariates may be time-dependent, with values that change over time. We assume that the censoring is noninformative in that the failure time T and the censoring time C are conditionally independent, given the history of the covariate vector $Z^*(X) = \{Z(u) : 0 \leq u < X\}$. The data $\{(X_i, Z_i^*(X_i), \Delta_i) : i = 1, \ldots, n\}$ are an i.i.d. sample of $\{(X, Z^*(X), \Delta)\}$. Let $t_i < \cdots < t_N$ be the N distinct ordered failure times, (j) be the label of the item failing at time t_j, and $\mathcal{R}_j = \{i : X_i \geq t_j\}$

be the risk set at time t_j. The hazard function for T, is defined as

$$\begin{aligned}\lambda\{t \mid Z^*(t)\} &= \lim_{\triangle t \downarrow 0} \frac{1}{\triangle t} P\{t \le T < t + \triangle t \mid T \ge t, Z^*(t)\} \\ &= \lambda_0(t)\psi\{\beta^T z(t)\},\end{aligned} \tag{1}$$

where $z(t)$ is a realization of $Z(t)$, $\lambda_0(t)$ is an unknown baseline hazard function corresponding to $z(t) = (0, \ldots, 0)$, and $\psi(\cdot)$ is a link function. In this model, we assume that the hazard depends on the complete covariate history $Z^*(t)$ only through the current value of $Z(t)$ at t. Here we use the same notation $z(t)$ for both time-dependent and time-independent covariates. For a time-independent covariate, $z(t) \equiv z$. When ψ is specified, the partial likelihood method was introduced in Cox (1972) and Cox (1975) to estimate the regression parameters β with censored data.

The Cox model assumes that the covariates affect the risk through a log-linear link function when $\psi\{\beta^T z(t)\} = \exp\{\beta^T z(t)\}$ or another known link function. Such an assumption may not be realistic or, at the least, should be validated before a specific link function is applied. Several alternative approaches have been explored to reflect the nature of the covariate effect when the link function is unspecified with one-dimensional time-independent covariate (Tibshirani and Hastie (1987) and Fan, Gijbels and King (1997)). Our goal here is to develop methods to check the link function in (1) with more than one covariate, including some time-dependent covariates.

When dealing with time-dependent covariates, one has to be very careful since there are some complications with such data. Kalbfleisch and Prentice (2002) pointed out that there are two types of time-dependent covariates: external covariates and internal covariates. External time-dependent covariates are those whose values do not depend on the failure process. For example, when studying how long someone remains employed, the inflation rate is essentially external to the individual employment duration. Internal time-dependent covariates are often the measurements taken on the subjects. Since these time-dependent measurements can only be taken when the subjects are under observation, the distribution of these covariates usually carries information on the failure process. For example, a patient's daily blood pressure is a time-dependent internal covariate. The definition of the hazard function and the construction of the partial likelihood for the proportional hazards regression models apply to both internal and external time-dependent covariates. But it is not possible to estimate the conditional survival function when there are internal covariates, because $P\{T \ge t \mid Z^*(t)\} = 1$ (since if $Z^*(t)$ is known, the subject must be alive at time t and at risk of failure). Also the baseline survival function $F_0(t) = \exp\{-\int_0^t \lambda_0(u)du\}$ has no simple interpretation when internal covariates

are present. Therefore one needs to pay extra attention when interpreting the results for data with internal covariates.

Another problem naturally occurs when the whole history of these covariates is not available. For example, in a clinical trial to study the relationship between PSA levels and time to progression, patients are only scheduled to visit at certain time points and PSA levels may not be available for all patients at a particular event time. This may result in incomplete history of the covariates. This complication exists in both data with external and internal time-dependent covariates.

Considering the problems in (1) with unknown link function and time-dependent covariates, we propose some methods to estimate the link and handle intermittent time-dependent covariates. In Section 2 we present methods for the proposed model that are applicable to both external and internal time-dependent covariates when the entire history is available. Asymptotic results for these estimators are also presented in Section 2. In Section 3 we discuss the situation when the full history of the covariates is unobtainable so that covariate values are only available intermittently on each individual, and introduce some nonparametric imputation methods to handle this problem. Simulation studies to examine the finite sample performance are reported in Section 4. The methods are demonstrated on data from a prostate cancer clinical trial in Section 5.

2. Estimation Procedure

When the entire history of the covariates is available, both $\psi(\cdot)$ and β in (1) can be estimated simultaneously. To ensure identifiability, we set $\psi(0) = 0$ and fix $\|\beta\|=1$ (here $\|\cdot\|$ represents the Euclidean length) and β_1, the first element of β, is positive.

For a fixed parametric value β, we propose to estimate the link function $\psi(\cdot)$ by the local likelihood approach as set forth in Fan et al. (1997). More specifically, for a given point v, assume that the pth order derivative of $\psi(\cdot)$ at point v exists. Let $\gamma(v) = \{\psi'(v), \ldots, \psi^{(p)}(v)/p!\}^T$, be the vector associated with the derivatives of ψ, and $\beta^T\mathbf{Z}(t) = \{\beta^T Z(t) - v, \ldots, (\beta^T Z(t) - v)^p\}^T$, then, by Taylor expansion, for $\beta^T Z(t)$ in a neighborhood of v, $\psi\{\mathbf{Z}(t)\}$ can be written as

$$\psi\{\mathbf{Z}(t)\} \approx \psi(v) + [\beta^T\mathbf{Z}(t)]^T\gamma(v).$$

Letting $Y(t) = I(X \geq t)$, we can estimate $\gamma(v)$ through the local log partial likelihood defined as

$$\sum_{j=1}^{n}\Bigg[K_h\{\beta^T Z_{(j)}(X_j) - v\}\Bigg([\beta^T\mathbf{Z}_{(j)}(X_j)]^T\gamma(v) - \log\Bigg\{\sum_{i=1}^{n}\exp\{[\beta^T\mathbf{Z}_i(X_j)]^T\gamma(v)\}K_h\{\beta^T Z_i(X_j) - v\}Y_i(X_j)\Bigg\}\Bigg)\Bigg]^{\Delta_j}, \quad (2)$$

where $\beta^T \mathbf{Z}_i(t)$ and $\beta^T \mathbf{Z}_{(j)}(t)$ are $\beta^T \boldsymbol{Z}(t)$ with $\boldsymbol{Z}(t)$ replaced by $\boldsymbol{Z}_i(t)$ and $\boldsymbol{Z}_{(j)}(t)$, respectively, and the kernel function is $K_h(u) = (1/h)K(u/h)$, h being the bandwidth. It can be shown that the local partial likelihood is strictly concave with respect to $\gamma(\cdot)$, so it has a unique maximizer.

Since (2) involves only $\gamma(\cdot)$ and not $\psi(\cdot)$, $\psi(\cdot)$ can be estimated through an integration method. For computational simplicity, the trapezoidal rule was applied in the simulation study of this approach in Section 4, as suggested in Tibshirani and Hastie (1987), and it appears to be satisfactory.

Once an estimate of $\psi(\cdot)$ is obtained, one can estimate β through the global partial likelihood with the link being replaced by its estimate, i.e.,

$$l_G(\beta, \hat{\psi}) = \sum_{j=1}^{n} \Bigg[\hat{\psi}(\mathbf{Z}_{(j)}(X_j)) - \log \Bigg\{ \sum_{i=1}^{n} \exp\{\hat{\psi}(\mathbf{Z}_i(X_j))\} Y_i(X_j) \Bigg\} \Bigg]^{\Delta_j}. \tag{3}$$

The following two-step iterative algorithm is used to estimate β and $\gamma(\cdot)$ simultaneously.

Step 0. Assign an initial value $\hat{\beta}$;

Step 1. Plug $\hat{\beta}$ into (2), then for a given v, maximize (2) w.r.t $\gamma(v)$ to get the estimate $\hat{\gamma}(v)$. Obtain the values of $\hat{\gamma}(v)$, for $v = \hat{\beta}^T Z_i(t), i = 1, \ldots, n;\ t = X_1, \ldots, X_n$. Apply the trapezoidal rule to get $\{\hat{\psi}(\hat{\beta}^T Z_i(t)) : i = 1, \ldots, n, t = X_1, \ldots, X_n\}$.

Step 2. Plug $\hat{\psi}(\cdot)$ into (3) and maximize it w.r.t β to update the estimate $\hat{\beta}$.

Repeat these Steps 1 and 2, until some convergence criterion is met.

The initial value of β can be set in different ways. One method is to fit the Cox model and use the β estimate in the first step. For details of this estimation procedure, see Wang, Wang and Wang (2001) and Wang (2001).

Data with time-dependent covariates whose entire history is observable are not uncommon in socioeconomic studies. Usually, the history of external time-dependent covariates is more easily observed since it is independent of the study subject. An example of such data is given in Maples, Murphy and Axinn (2002) where distance, from each neighborhood a study subject lived at different time periods to the nearest school, was considered as an exogenous time-dependent covariate to link dynamic changes in socioeconomic context to individual-level life histories. There are also data whose internal time-dependent covariates are attainable throughout the study. Chiou, Mueller and Wang (2003) studied the relationship between reproductive behaviors and longevity of a cohort of medflies (Ceratitis capitata). The number of eggs laid daily by a medfly was treated as a time-dependent covariate.

2.1. Large-sample properties

Let f be the probability density of $\beta^T \mathbf{Z}(t)$ and, for a given v, let $P(t \mid v) = P(X \geq t \mid \beta^T \mathbf{Z}(t) = v)$, $H = \operatorname{diag}\{h, \ldots, h^p\}^T$ and $\mathbf{u} = \{u, \ldots, u^p\}$.

Now we develop the large-sample properties of the link $\psi(\cdot)$ estimator under the following regularity conditions.

(C1) $K \geq 0$ is a bounded density with compact support, and it has bounded first and second derivatives.

(C2) $\psi(\cdot)$ has a continuous $(p+1)$th derivative around v.

(C3) The density $f(\cdot)$ of $\beta^T \mathbf{Z}(t)$ is continuous at point v and $\inf_v f(v) > 0$.

(C4) The conditional probability $P(t|\cdot)$ is equicontinuous at v.

(C5) $\int_0^\tau \lambda_0(u)du < \infty$.

We follow the proofs in Wang (2001) with Z replaced by $Z(t)$. For any $\sqrt{n}$ consistent estimator $\hat{\beta}$ of the true parameter β_0, let $\hat{\gamma}\{\hat{\beta}^T z(t)\}$ be the corresponding estimator for the derivative vector $\gamma_0\{{\beta_0}^T z(t)\}$ of the true link ψ and $\hat{\psi}(v) = \int_0^v \hat{\psi}'(x)dx$, where $\hat{\psi}'(\cdot)$ is the first component of $\hat{\gamma}(\cdot)$. If $n \to \infty$, $h \to 0$, $nh \to \infty$, nh^{2p+3} is bounded, and $nh^4 \to \infty$ then $\sup_{|z(t)|\leq B} |\hat{\psi}\{\hat{\beta}^T z(t)\} - \psi\{{\beta_0}^T z(t)\}| \to_p 0$, and

$$\sqrt{nh}\left\{ H(\hat{\gamma}\{\hat{\beta}^T z(t)\} - \gamma_0\{{\beta_0}^T z(t)\}) - \frac{\psi^{(p+1)}\{{\beta_0}^T z(t)\}}{(p+1)!} A^{-1} b h^{p+1} \right\}$$
$$\to_D N\left\{ 0, \frac{\sigma^2\{{\beta_0}^T z(t)\}}{f\{{\beta_0}^T z(t)\}} A^{-1} D A^{-1} \right\},$$

where $\nu_1 = \int \mathbf{u} K(u)du$, $b = \int u^{p+1}(\mathbf{u} - \nu_1)K(u)du$, $A = \int \mathbf{u}\mathbf{u}^T K(u)du - \nu_1 \nu_1^T$, $D = \int K^2(u)(\mathbf{u} - \nu_1)^{\otimes 2} du$ and $\sigma^2\{{\beta_0}^T z(t)\} = E\{\delta | \mathbf{Z}(t) = {\beta_0}^T z(t)\}^{-1}$.

Although we can get an estimate for β from the proposed procedure, the main purpose is not to estimate β under the nonparametric link, but rather to use the two-step iterative procedure as a tool to check the proper form of the link function. A suitable parametric form of the link function to estimate the covariate effects can be chosen based on the point estimate and the interval estimate of the nonparametric link function.

3. Intermittent Time-dependent Covariates

The previous section considered an ideal situation in which the entire history of $Z(t)$ from the origin to the observed event time X is known for each individual. The reason that the covariate history needs to be known before one can either use a parametric link or estimate the link nonparametrically can be shown from the partial likelihood for the Cox model with time-dependent covariates.

Examine the partial likelihood

$$L(\beta) = \prod_{j=1}^{n} \left[\frac{\exp\{\beta^T Z_{(j)}(X_j)\}}{\sum_{i=1}^{n} \exp\{\beta^T Z_i(X_j)\} Y_i(X_j)} \right]^{\Delta_j}, \tag{4}$$

to see that, for each failure time X_j, the covariate value $Z_i(X_j)$ for each individual i at risk at time X_j needs to be present. But sometimes it is not practical to observe the entire history of a time-dependent covariate. For example in a clinical trial, patients are usually scheduled at preset time points and it is very common that some patients' covariates are not measured at other patients' event time. That is, even though one patient can make all of his visits, there could still be missingness in the sense that we may need to know his covariate value at an event time which does not fall on his scheduled visiting time points. This results in missing values for time-dependent covariates in (4). Here we assume that the "real missingness" that dictates the intermittent times at which we have covariate values is completely at random. In the rest of the paper we will use the word "missing" interchangeably with "intermittent".

They are some ways to handle this situation. If one is willing to make the assumption that the time-dependent covariates follow linear mixed effects models with or without measurement errors, common strategies can be adopted including regression calibration methods (Pawitan and Self (1993), Tsiatis, Degruttola and Wulfsohn (1995) and Dafni and Tsiatis (1998)), joint likelihood methods (Degruttola and Tu (1994), Faucett and Thomas (1996), Wulfsohn and Tsiatis (1997) and Henderson, Diggle and Dobson (2000)), and conditional score methods (Tsiatis and Dividian (2001)). But there are no existing methods which can be directly applied to estimate the unknown link function.

Alternatively, one can model the time-dependent covariates in a nonparametric way. An ad hoc approach commonly adopted in practice and in SAS and Splus, referred as the last value carried forward (LVCF) method, is to impute the missing covariates with the nearest observation from the same individual before the time at which the covariates were missing. This is equivalent to adopting a nearest 1-neighbor estimate for the missing data, or conceptually assuming that the covariates have not changed since the most recent measurement. Obviously, different smoothing methods based on each individual's observations can be used to estimate the unobserved part of the data. For example, one can apply kernel smoothing methods to the observed part of $Z_i(t)$ for individual i to get an estimated curve $\{\hat{Z}_i(t) : t_1 \le t \le X_i\}$, then replace the missing $Z_i(t_j)$ with $\hat{Z}_i(t_j)$ for $t_1 \le t_j \le X_i$. However, when the data are sparse, none of these methods works well as they only utilize information from the individual history. A more appealing approach for imputing missing covariates for a particular individual may be to utilize relevant information from other individuals. This is often referred to as "borrowing strength". When the time-dependent covariates are all external and are missing completely at random, "borrowing strength" methods will work well since one can utilize the time-dependent data beyond the failure time. But for internal time-dependent covariates, there is a major complication caused by the

fact that the longitudinal trajectories are truncated by death, which may induce informative censoring, as discussed by Wu and Carroll (1988) and many other authors. Failure to take into account this problem can lead to biased estimates of interest.

Here we first outline a general nonparametric method based on functional principal components analysis (FPCA) to impute time-dependent covariates by "borrowing strength" that does not take the informative censoring into account, then we modify the procedure by adopting an idea similar to one of Tsiatis et al. (1995) when there is informative censoring.

Although for each study only discrete data points are observed, one can assume that these data are realizations of a smooth L^2 stochastic process. Thus the data are regarded as sample curves of functions, and termed "curve" or "functional" data. The observed covariate process for an individual is often referred as a sample "curve", $Z(t)$, with t being the index of the function. In practice, t often represents observation time in a longitudinal study.

Assume $\{Z(t) : 0 \le t < \tau\}$ is an L^2-process with mean $\mu(t)$. Let $\Sigma(s,t)$ be the covariance function of $Z(t)$ at time (s,t), and $\{(\omega_j, \phi_j(t)), j = 1, 2, \ldots, 0 \le t < \tau\}$ be the corresponding eigenvalues and eigenfunctions.

By the *Karhunen-Loève Theorem* (Ash and Gardner (1975)), we can write

$$Z(t) = \mu(t) + \sum_{j=1}^{\infty} A_j \phi_j(t), \quad 0 \le t < \tau,$$

where $A_j = \int_0^\tau \phi_j(t) Z(t) dt \equiv \langle \phi_j, Z \rangle, j = 1, 2, \ldots$, are called functional principal components (see details in the book by Ramsay and Silverman (1997)). It is easy to show that A_j's, are uncorrelated random variables with $E(A_j) = 0$ and $E(|A_j|^2) = \omega_j$. The series $\mu(t) + \sum_{j=1}^{r} A_j \phi_j(t)$ converges in L^2 to $Z(t)$ as $r \to \infty$, uniformly in t.

If we calculate the sample covariance function of $\{Z_i(t) : i = 1, \ldots, n; 0 \le t < \tau\}$ and perform a FPCA to estimate the eigenvalues and eigenfunctions, $\{\hat{\omega}_j, \hat{\phi}_j(t) : j = 1, 2, \ldots; 0 \le t < \tau\}$, we can approximate $Z_i(t)$ with

$$\hat{Z}_i(t) = \hat{\mu}(t) + \sum_{j=1}^{r} \hat{A}_{ij} \hat{\phi}_j(t), \tag{5}$$

where $\hat{A}_{ij} = \int_0^\tau \hat{\phi}_j(t) Z_i(t) dt$.

A nice and simple example of FPCA is given by Rice and Silverman (1991). In most cases, the curve data are sampled at irregular time points, and some smoothing of the covariance function is required before we employ the function principal component analysis. Suppose we have data $\{Z_i(t_{ij}) : i = 1, \ldots, n;\ j = 1, \ldots, n_i\}$, the measurements for subject i observed at the time points t_{ij}. Note

that for each subject i, the observed time interval can be different. Let $\{S_1, \ldots, S_L\}$ be the distinct ordered time points of $\{t_{ij} : i = 1, \ldots, n,\ j = 1, \ldots, n_i\}$ observed from all individuals. Denote the sample mean function by $\bar{Z}(t) = (1/n_t)\sum_{i=1}^{n_t} Z_i(t)$ and the sample covariance function between the time points s and t by $\bar{\Sigma}(s,t) = (1/n_{st})\sum_{i=1}^{n_{st}}(Z_i(s) - \bar{Z}(s))(Z_i(t) - \bar{Z}(t))$, where n_t is the number of individuals without missing values at the time point t and n_{st} is the total number of available pairs at the time pair (s,t).

If there is sufficiently large number of observations n_t and n_{st} available at time points s and t, then we can use $\bar{Z}(t)$ and $\bar{\Sigma}(s,t)$ as the estimates for $\mu(t)$ and $\Sigma(s,t)$. Generally, we may need to apply smoothing techniques to $\bar{Z}(t)$ and $\bar{\Sigma}(s,t)$ to obtain $\hat{\mu}(t)$ and $\hat{\Sigma}(s,t)$ when individual's observed time points are different.

Smoothing procedures for the mean function and the covariance function can be found in Rice and Silverman (1991), Staniswalis and Lee (1998) and Diggle and Verbyla (1998). The details of how to smooth $\bar{Z}(t)$ and $\bar{\Sigma}(s,t)$ are as follows. Note that the choices of smoothers are subjective, here the local polynomial smoother is chosen.

First, one needs to estimate the mean function $\mu(t)$ and this can be achieved through a scatter-plot smoothing. Let K_1 be a kernel function satisfying $\int K_1(u)\,du = 1$, $\int uK_1(u)du = 0$. Define $K_{1,h}(u) = (1/h)K_1(1/h)$, where h is a bandwidth. Suppose the p_1th derivative of $\mu(t)$ exists at point t, and let the design matrix be

$$D_1 = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ S_1 - t & S_2 - t & \cdots & S_L - t \\ \cdots & \cdots & \cdots & \cdots \\ (S_1 - t)^{p_1} & (S_2 - t)^{p_1} & \cdots & (S_L - t)^{p_1} \end{pmatrix}^T,$$

$W_1 = \operatorname{diag}\{K_{1,h}(S_1 - t), K_{1,h}(S_2 - t), \ldots, K_{1,h}(S_L - t)\}$ and $\bar{\mathbf{Z}} = \{\bar{Z}(S_1), \ldots, \bar{Z}(S_L)\}$.

The estimator of the rth derivative of $\mu(t)$ is then given as $r!\mathbf{1}_r^T(D_1^T W_1 D_1)^{-1} D_1 W_1 \bar{\mathbf{Z}}$, where $\mathbf{1}_r$ is the $(p_1+1)\times 1$ vector having the value 1 in the rth entry and 0 elsewhere. The local linear ($p_1 = 1$) estimator of $\mu(t)$ can be written as $(1,0)(D_1^T W_1 D_1)^{-1} D_1 W_1 \bar{\mathbf{Z}}$.

Similarly, a local linear smoothing can be applied to the covariance function of $e(t) = Z(t) - \hat{\mu}(t)$. Let $\mathbf{u}^* = (u_1, u_2)$ and K_2 be a bivariate nonnegative kernel function satisfying $\int K_2(\mathbf{u}^*)d\mathbf{u}^* = 1$, $\int \mathbf{u}^* K_2(\mathbf{u}^*)d\mathbf{u}^* = 0$ and $\int u_i u_j K_2(\mathbf{u}^*)d\mathbf{u}^* = \delta_{ij}\mu_2(K_2)$ with $\mu_2(K_2) \geq 0$, where $i, j = 1, 2$ and $\delta_{ij} = 1$ if $i = j$, 0 otherwise. Define $K_{2,H^*}(\mathbf{u}^*) = (1/|H^*|)K_2([H^*]^{-1}\mathbf{u}^*)$, where H^* is a nonsingular 2×2 matrix.

For a given time pair (s,t), let the design matrix be

$$D_2 = \begin{pmatrix} 1 & \cdots & 1 & 1 & \cdots & 1 & \cdots & 1 & 1 \\ S_1-s & \cdots & S_1-s & S_2-s & \cdots & S_2-s & \cdots & S_{L-1}-s & S_L-s \\ S_1-t & \cdots & S_L-t & S_2-t & \cdots & S_L-t & \cdots & S_L-t & S_L-t \end{pmatrix}^T,$$

which is a $(L(L+1)/2) \times 3$ matrix.

Let $\Sigma^*(s,t) = (1/n_{st}) \sum_{i=1}^{n_{st}} (e_i(s) - \bar{e}(s))(e_i(t) - \bar{e}(t))$ for $s \le t$ with $\bar{e}(t)$ being defined similarly to $\bar{Z}(t)$. Let $W_2 = \text{diag}\{K_{2,H^*}(S_1 - s, S_1 - t), K_{2,H^*}(S_1 - s, S_2 - t), \ldots, K_{2,H^*}(S_1 - s, S_L - t), \ldots, K_{2,H^*}(S_{L-1} - s, S_L - t), K_{2,H^*}(S_L - s, S_L - t)\}$, which is a $(L(L+1)/2) \times 3$ matrix, and $V = \{\Sigma^*(S_1, S_1), \Sigma^*(S_1, S_2), \ldots, \Sigma^*(S_1, S_L), \ldots, \Sigma^*(S_{L-1}, S_L), \Sigma^*(S_L, S_L)\}$, a $(L(L+1)/2)$ vector of distinct element of the covariance matrix of $e(t)$.

The estimator $\hat{\Sigma}(s,t)$ of $\Sigma(s,t)$ is given by $\hat{\Sigma}(s,t) = (1,0,0)(D_2^T W_2 D_2)^{-1} D_2^T W_2 V$. For simplicity, one can choose the product kernel, so that the smoothed covariance function at time (s,t) borrows the information from the neighbor rectangle. The bandwidth matrix H^* is now of the form $\begin{pmatrix} h_1 & 0 \\ 0 & h_2 \end{pmatrix}$. This may be further simplified by choosing $h_1 = h_2$, as we are dealing with a covariance function.

Note that the aforementioned procedure can be easily modified to account for data with measurement errors. For data with measurement errors, suppose we observe $Z_\varepsilon(t) = Z(t) + \varepsilon(t)$, $0 \le t \le X$, where $\varepsilon(t)$ are the uncorrelated measurement errors with mean zero and variance σ^2, independent of the event time X, the measurement time points and the true covariate history $Z^*(X)$. We assume that one is interested in estimating the relationship between the survival and the true covariate history $Z^*(X)$ rather than the observed covariate history that are contaminated with errors, that is, we consider the same model (1), i.e., $\lambda\{t \mid Z^*(t)\} = \lambda_0(t)\psi\{\beta^T z(t)\}$.

To adjust for measurement errors, one may apply smoothing procedures for the variance part and the covariance part separately. Consider

$$\text{Cov}\,(Z_\epsilon(t), Z_\epsilon(s)) = \text{Cov}\,(Z(t), Z(s)) + \sigma^2 \delta(s,t), \tag{6}$$

where $\delta(s,t) = 1$ if $s = t$ and 0 otherwise. So the $\{\Sigma^*(s,t) : s = t\}$ part of the raw covariance of $Z(t)$ (or $e(t)$) should be taken out of the vector V along with the corresponding parts of matrix D_2 and W_2 in the procedure for smoothing the covariance matrix. Local linear and local quadratic forms are chosen to smooth the covariance matrix along the direction of diagonal and perpendicular to diagonal, respectively, due to the fact that the variance of $Z_\varepsilon(t)$ is maximized along the diagonal. Estimate of the variance function of $Z_\varepsilon(t)$ can be obtained

by any smoother using the data $\{\Sigma^*(s,t) : s = t\}$. Then the variance term of the measurement errors can be written as

$$\hat{\sigma}^2 = \frac{1}{\tau}\int_0^\tau (\hat{var}(Z_\epsilon(t)) - \hat{\Sigma}(t,t))dt. \tag{7}$$

More details of smoothing variance/covariance matrix for data with measurement errors can be found in Staniswalis and Lee (1998) and Yao et al. (2003).

By applying functional principal components analysis in the setting of survival data with longitudinally measured time-dependent covariates, one is able to handle missing time-dependent covariates by following the steps below.

Step A. Calculate the sample mean function $\bar{Z}(t), 0 \le t \le S_L$, and the sample covariance function $\bar{\Sigma}(s,t), S_1 \le s,t, \le S_L$ for $Z(t)$. Apply certain smoothing techniques to get a smoothed mean function $\hat{\mu}(t)$ and covariance function $\hat{\Sigma}(s,t), S_1 \le s,t, \le S_L$.

Step B. Apply functional principal components analysis to $\hat{\Sigma}(s,t)$ and get the estimates $\hat{Z}_i(t), t_1 \le t \le t_N$, as defined in (5) for each $i = 1, \ldots, n$.

Step C. Run the two-step iterative algorithm in Section 2 with $Z_i(t)$ replaced by $\hat{Z}_i(t)$.

In practice, the number of functional principal components, r, can be selected so that 90% of variation can be explained by the first r functional principal components.

For data with internal time-dependent covariates, one can modify this FPCA method to reduce the bias caused by informative censoring. Instead of using all data from all individuals during the smoothing, one can perform smoothing at each failure time point t using only the data up to that time t from those individuals at risk at time t, as done in the parametric context by Tsiatis et al. (1995). This approach is referred as the functional principal components analysis within risk set (FPCARS). Simulation studies in Section 4 indicate that this modified method leads to much smaller bias. One drawback of the FPCARS method is the computational complexity caused by applying smoothing procedures at each event time.

The estimation procedure of the FPCARS approach is essentially the same as the FPCA method except for some definition changes. Suppose at a failure time t_j, $\{S_1^*, \ldots, S_m^*\}$ is the distinct ordered time points of $\{t_{ik} : i \in \mathcal{R}_j, k = 1, \ldots, n_i\}$ observed from the individuals in the risk set $\mathcal{R}_j$. Denote the sample mean function by $\bar{Z}^{RS}(t) = (1/n_m)\sum_{k=1}^{n_m} Z_k(t)$ and the sample covariance function between time s and t by $\bar{\Sigma}^{RS}(s,t) = (1/n_{m1})\sum_{k=1}^{n_{m1}}(Z_k(s) - \bar{Z}^{RS}(s))(Z_k(t) - \bar{Z}^{RS}(t))$, where n_m is the number of individuals in the risk set $\mathcal{R}_j$ without missing values at the time point t, and n_{m1} is the total number of available pairs at the time or before pair (s,t) from the risk set. $D_1^{RS}, D_2^{RS}, W_1^{RS},, W_2^{RS}, \bar{\mathbf{Z}}^{RS}$ and V^{RS} can be defined accordingly.

The local linear ($p_1 = 1$) estimator of $\mu(t)$ can be written as $(1,0)[(D_1^{RS})^T W_1^{RS} D_1^{RS}]^{-1} D_1^{RS} W_1^{RS} \bar{\mathbf{Z}}^{RS}$, and the estimator $\hat{\Sigma}^{RS}(s,t)$ of $\Sigma^{RS}(s,t)$ is

$$\hat{\Sigma}^{RS}(s,t) = (1,0,0)[(D_2^{RS})^T W_2^{RS} D_2^{RS}]^{-1} (D_2^{RS})^T W_2^{RS} V^{RS}.$$

Then (5) is used to obtain the estimate of $\{Z_i(t_j) : i \in \mathcal{R}_j\}$. Repeat this process at each event time to get the estimated covariate history for each individual at risk at that time.

Note that there are limitations to this approach at both very early and very late failures. For very early failures, there may be lots of individuals in the risk set, but very few observed covariate values over time from each individual in the risk set. Similarly, for very late failures, there will be very few individuals in the risk set even though each of these individuals may provide lots of observations over time. Smoothing may not perform well in these cases.

Another drawback of this approach is that one needs to run the entire process N times, where N is the total number of failures. When there are lots of failures, it will be computationally expensive. One shortcut is to run the entire process only at a few selected failure times, and then use some kind of imputation or interpolation method to get the estimates at all failure times. A similar shortcut was mentioned in Tsiatis and Dividian (2001). In the simulation studies in Section 4, we selected the 25th, 50th, 75th and 100th percentiles of the ordered failure times, and then applied interpolation methods to get the estimates at all failure times. The results seem to be satisfactory.

4. Simulation Studies

Several simulations were carried out to examine the finite sample performance of the procedures discussed in the previous sections. A sample size of 100 was selected for all studies.

Simulation 1: To check the performance of the semiparametric procedures described in Section 2 when the time-dependent covariates are available throughout the study interval, we did two simulations. First, a linear link function $\psi(t) = Z_1(t) + Z_2(t)$ was examined on the fixed interval [0, 300] with $Z_1(t) \equiv Z_1$, a time-independent covariate sampled from $N(0,1)$, and $Z_2(t) = \theta_1 + \theta_2 t$, a time-dependent process where $\theta_1 \sim N(-100, 10)$ and $\theta_2 \sim N(0.5, 0.1)$. We set up the simulation so that subjects were only censored at the end of the study. The censoring rate was about 10%. Figure 1 shows the mean curve based on 100 simulated datasets and a few randomly selected samples of the estimated link functions. It is clear that the mean curve is almost a straight line. The performance of the estimate of β was also explored though it was not the main target of the new procedure. Due to the identifiability conditions discussed in Section 2, we can only estimate β subject to $||\beta|| = 1$. To see the difference between

the Cox estimate, which assumes the true link function, and the semiparametric estimate, we then fix the first component of both estimates of β to be the true value 1, and compare the second component only. The two methods yielded similar results − the mean difference between the proposed estimate and the true parameter was 0.052 with a standard deviation of 0.173 when the bandwidth was set to be half of the covariate range, and the mean difference between the Cox estimate and the true parameter was 0.039 with a standard deviation of 0.170.

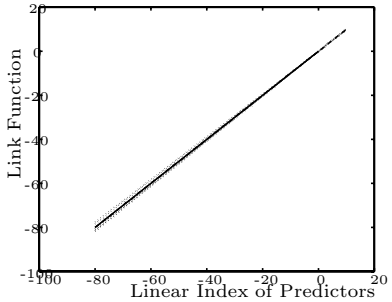

Figure 1. The estimated link function. The solid line is the true link function, the dashed line is the mean curve of the estimated link functions, and the dotted lines are some randomly selected samples of the estimated link functions.

Simulation 2: A quadratic link $\psi(t) = [Z(t)]^2$ was selected with $Z(t)$ sampled from $Z(t) = \theta_1 + \theta_2 t$, where $\theta_1 \sim U(0, 0.02)$ and $\theta_2 \sim N(-0.5, 1)$ on the interval [0,150]. Here we assume that we can observe the complete history of this time-dependent covariate. Censoring variables were generated independently from exponential distributions to ensure that the censoring rate is around 25%. Figure 2 shows the mean curve based on 100 simulated datasets and a few randomly selected samples of the estimated link functions. The difference between the adjusted second component of $\hat{\beta}$ and the true β is 0.008 with a standard deviation of 0.156.

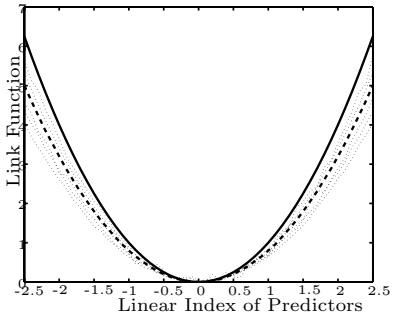

Figure 2. The estimated link functions. The solid curve is the true link function, the dashed curve is the mean function and the dotted curves are some randomly selected samples of the estimated link functions.

Simulation 3: The functional principal components analysis approach was applied in the longitudinal data setting to understand the imputation only, without the complication of considering the survival outcome.

A fixed time interval from 0 to 300 days with increment of 10 days was adopted as the scheduled visiting time points. Each subject followed one of the three schedules: $\{10, 40, 70, \ldots, 280\}$, $\{20, 50, 80, \ldots, 290\}$ or $\{30, 60, 90, \ldots, 300\}$. The process $Z_i(t)$ is sampled from $Z_i(t) = \mu(t) + \eta_i(t)$, where $\mu(t) = 0.01t + \sin(\pi t/150)$, $\eta_i(t) = -\sqrt{1/150}\cos(\pi t/300)a_i + \sqrt{1/150}\sin(\pi t/300)b_i$, a_i and b_i are independent $N(0, 4)$. A leave-one-curve-out cross-validation method was used to select the optimal bandwidths for smoothing the mean function and the covariance function. Random errors from distribution $N(0, 0.09)$ were then added to each of the individual process after the initial fitting. The "BIC" criterion was used to select the optimal bandwidths. Estimated curves from ten subjects, with or without random error terms, are shown in Figure 3. The two sets of estimated curves are very close to the true links, suggesting that the FPCA procedures worked well under this setting.

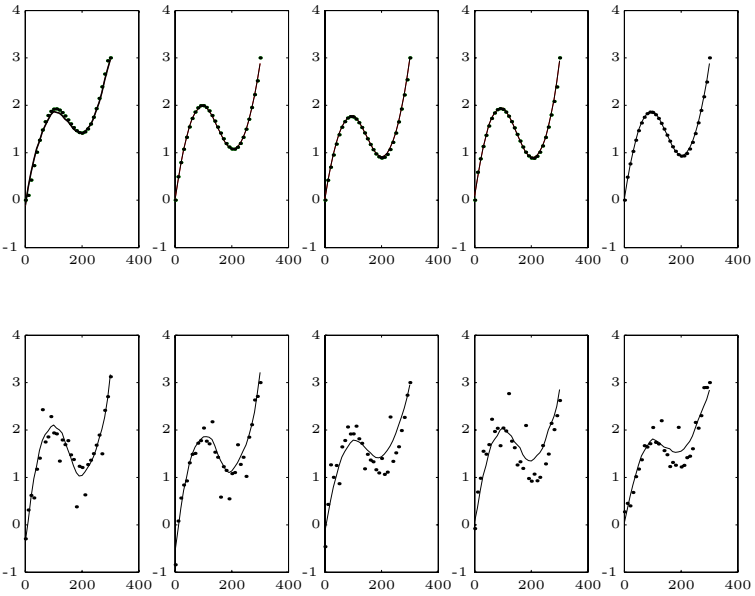

Figure 3. Top part: Estimated curves from five randomly selected datasets simulated from the model without random errors. Bottom part: Estimated curves from five randomly selected datasets simulated from the model with random errors.

Simulation 4: A nonparametric imputation method was applied to data with survival outcome and possibly missing time-dependent covariates. $Z(t)$

was sampled from $Z(t) = \theta_1 + \theta_2 t + \epsilon(t)$, where $\theta_1 \sim U(0, 2)$, $\theta_2 \sim U(0, 0.3)$ and $\epsilon(t) \sim N(0, 0.1)$. Each individual was randomly assigned to one of the three visit schedules: day $\{5, 20, \ldots, 140\}$, $\{10, 25, \ldots, 145\}$ or $\{15, 30, \ldots, 150\}$. Various bandwidths $25, 30, 45, 60, 75, 90$ were tried in covariance function smoothing. Considering the fact that the proposed method is in the framework of two-stage approaches which might introduce bias in the estimation of β, we first focused on the impact of the nonparametric imputation methods without estimating the link function. For computational simplicity, the identity link, i.e., the Cox model, was employed. A constant baseline hazard of 0.0005 was used in the simulation at all censoring rates. Censoring variables were generated independently from the exponential distributions. All these parameters were chosen to ensure that there are not too many very early deaths. Figure 4 shows the distribution of the event time and the trajectory of the time-dependent covariate from a random simulated dataset under each censoring rate.

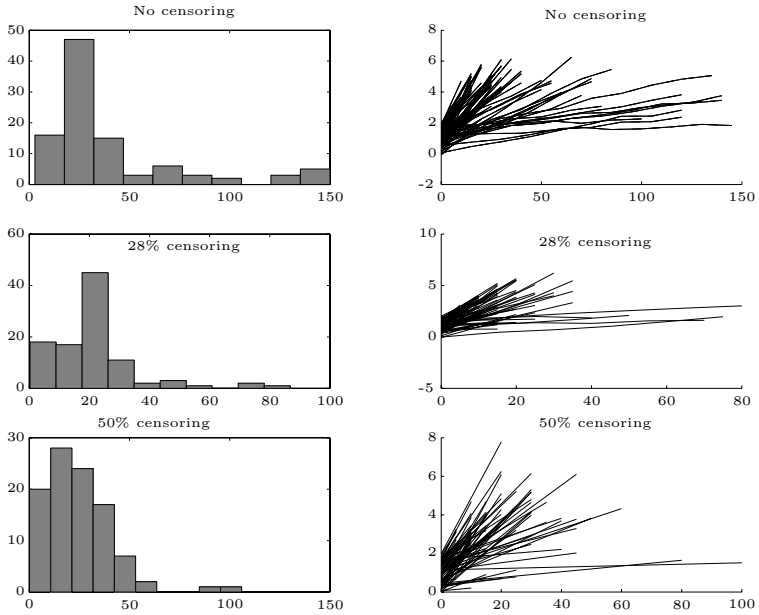

Figure 4. The histograms of the event time and the trajectories of the time-dependent covariate from a randomly simulated dataset under each censoring rate.

Four estimates of β were obtained by using the full data assuming no missing time-dependent covariates, and by replacing the missing covariates with the LVCF estimates, the linear mixed-effects estimates, and the FPCARS estimates. Here, since these time-dependent covariates are internal, the FPCA method will not provide good estimates.

Several bandwidths were applied, and the results of the best procedures in terms of the smallest MSE of the differences between estimated β and true β are listed in Table 1 based on 100 runs. The corresponding bandwidths for the FPCARS methods were 25, 60 and 50. Both the LME estimates and the FPCARS estimates performed better than the LVCF estimates. Since the true model is a linear mixed-effects model, it is not surprising that the LME estimators did the best. The FPCARS estimators were also very close to the complete estimators, suggesting that the bias can indeed be very small. This was quite remarkable since, even though we had around ten planned visits (i.e., the time-dependent covariates should be longitudinally measured ten times), the actual number of measurements for the time-dependent covariates per subject was much less due to the occurrence of event or censoring. The large variation might be reduced with larger samples.

Table 1. The differences between the true β and estimated β based on the best procedure with time-dependent covariates under the model: $\psi[Z(t)] = Z(t)$ where $Z(t) = \theta_1 + \theta_2 t + \epsilon(t)$, $\theta_1 \sim U(0,2)$, $\theta_2 \sim U(0,0.3)$, $\epsilon(t) \sim N(0,0.1)$ and $t = 5, 10, \ldots, 150$.

censoring rate		Complete	LVCF	Parametric	FPCARS
		estimate	estimate	estimate	estimate
0	Bias	-0.001	-0.128	-0.004	-0.008
	Std	0.107	0.094	0.110	0.118
	MSE	0.011	0.025	0.012	0.014
25%	Bias	0.006	-0.165	0.011	-0.016
	Std	0.146	0.118	0.150	0.155
	MSE	0.021	0.041	0.023	0.024
50%	Bias	0.030	-0.156	0.065	0.081
	Std	0.134	0.110	0.151	0.169
	MSE	0.019	0.041	0.027	0.035

Simulation 5: A similar setup to the previous one was chosen with a different covariate pattern $Z(t) = \theta_1\sqrt{t} + \theta_2 \sin(\sqrt{\pi t/75}) + \epsilon(t)$, $\theta_1 \sim U(0,0.2)$, $\theta_2 \sim N(-4,1)$ and $\epsilon(t) \sim N(0,0.1)$. Figure 5 shows the distribution of the event time and the trajectory of the time-dependent covariate from a random simulated dataset under each censoring rate. A cubic polynomial term was used in the parametric mixed-effects model. The results shown in Table 2 indicate that both the parametric method and the FPCARS method performed much better than the LVCF method. The FPCARS method performed a little bit better than the parametric model, suggesting that the FPCARS method is more data adaptive. The bandwidths for the FPCARS methods were 75, 75 and 90, respectively, under three censoring rates.

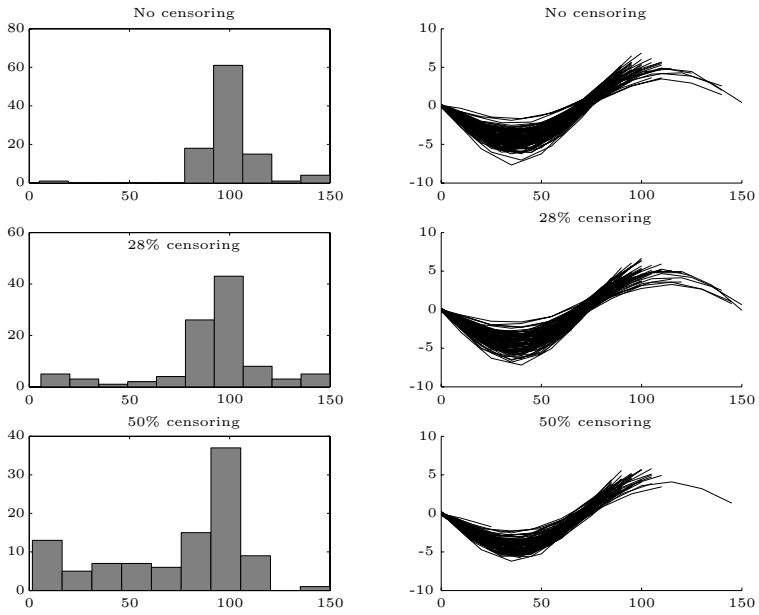

Figure 5. The histograms of the event time and the trajectories of the time-dependent covariate from a randomly simulated dataset under each censoring rate.

Table 2. The differences between the true β and estimated β based on the best procedure with time-dependent covariates under the model: $\psi[Z(t)] = Z(t)$ where $Z(t) = \theta_1\sqrt{t} + \theta_2 \sin(\sqrt{\pi t/75})$, $\theta_1 \sim U(0, 0.2)$, $\theta_2 \sim N(-4, 1)$ and $t = 5, 10, \ldots, 150$.

censoring rate		Complete	LVCF	Parametric	FPCARS
		estimate	estimate	estimate	estimate
0	Bias	-0.071	-0.496	-0.160	-0.094
	Std	0.133	0.093	0.140	0.151
	MSE	0.023	0.255	0.045	0.032
28%	Bias	0.068	-0.492	0.182	-0.021
	Std	0.143	0.117	0.195	0.239
	MSE	0.025	0.256	0.071	0.058
50%	Bias	0.046	-0.493	0.180	0.034
	Std	0.210	0.147	0.235	0.262
	MSE	0.047	0.265	0.088	0.069

Simulation 6: We did a full simulation combining the imputation procedure and the nonparametric link estimation procedure. We substituted the FPCARS estimates and estimated the link function under the same setting as in Simulation 4, i.e., the true link function is a linear function. The mean curve of the estimated

link functions is shown in Figure 6. Again it seems to be a straight line.

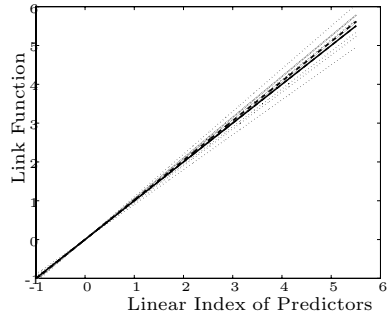

Figure 6. The estimated link functions. The solid curve is the true link function, the dashed curve is the mean function and the dotted curves are some randomly selected samples of the estimated link functions.

5. Data Analysis

Prostate cancer is the most common cancer among men in the United States behind skin cancer, and is second only to lung cancer as a cause of cancer-related death among men. The American Cancer Society estimated that 220,900 new cases of prostate cancer would be diagnosed and that approximately 28,900 men would die of the disease in 2003 in the United States. Approximately half of the estimated cases of prostate cancer are not curable by surgery or radiation. The majority of these men eventually develop clinical manifestations of distant metastases, and most will succumb within three years of the appearance of bone metastases. Although androgen deprivation therapies provide temporary control, metastatic disease usually progresses within 12−18 months. The failure of hormonal therapies is a consequence of the selective growth and dominance of androgen-independent cells. Progress in the systemic treatment of prostate cancer will depend on the ability to eradicate or restrain the growth of androgen-independent tumor cells.

The study was designed to examine the effect of weekly paclitaxel plus estramustine in patients with hormone-refractory prostate carcinoma. Seventy-eight patients were enrolled in the study from February 1999 to November 2000. The serum prostate specific antigen (PSA) levels were measured at the time of entry, weeks 3 and 6 of each 8-week course of treatment, and every 4 weeks after treatment until progression. The number of measurements ranges from 3 to 11 with a median of 8. There were 56 progressions among these patients. Literature has shown that PSA plays a very important role in the diagnosis and treatment of prostate cancer. We are interested in the relationship between PSA velocity or PSA history and time to progression. The profile of the PSA levels is shown in

Figure 7. It seems that linear mixed models may not be adequate to fit the PSA trajectories.

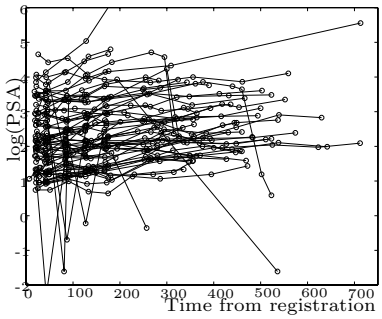

Figure 7. The PSA profile.

First a Cox model was fitted to study the impact of the baseline PSA levels (after log transformation) on the time to progression. The result suggested that the effect was not significant (p=0.92). Then a linear mixed effects model was attempted to find the patterns of these PSA levels measured over time. Quadratic term, cubic term and higher-order polynomials of the time since registration were also added to the model as random effects. To see whether these polynomial terms were needed in the final model, we used the parametric bootstrap approach described in Tsiatis et al. (1995) to derive the distribution of the likelihood ratio test since the null models occurred on the boundary of the covariance parameter space and the asymptotic distribution of the likelihood test statistic became a mixture of χ^2 distributions (Self and Liang (1987)). We randomly generated data from distributions corresponding to the model with only linear term of time and computed the maximized likelihood from fitting the model with only a linear time term and from the model with an added quadratic time term for each simulated dataset. In this way we obtained an empirical distribution of the likelihood ratio test. The likelihood ratio test for the actual dataset was compared to the empirical distribution of the likelihood ratio test to determine whether the quadratic time term was necessary. A significant increase in the maximized log-likelihood was found in the model with the quadratic term. This resampling procedure was repeated to test for higher-order terms and it turned out that the third, fourth and fifth order terms were all significant (we stopped the procedure after testing for the fifth order term). We felt it would be better to use a nonparametric form of the time variable in the proportional hazards regression model to describe the stochastic process of the PSA levels.

Since the PSA levels are internal covariates with measurement errors, the proposed FPCARS method for data with measurement errors was applied to estimate the PSA levels at each event time. We then fitted the Cox model again

with these estimated time-varying PSA levels, and the result suggested that, while the baseline PSA was of little prognostic value, the process of PSA was a significant predictor of time-to-progression ($p<0.01$) with a higher PSA level corresponding to a larger hazard rate. Finally, the nonparametric estimating procedures were used to check the link function. Figure 8 shows some departure from the identity link. Though it is not our intention to promote the nonparametric model, bootstrapping methods may be used to make inference on the β estimate with the estimated link function and the imputed time-dependent covariates. Since there are only 78 patients in the trial, further investigation with more data is needed for inference.

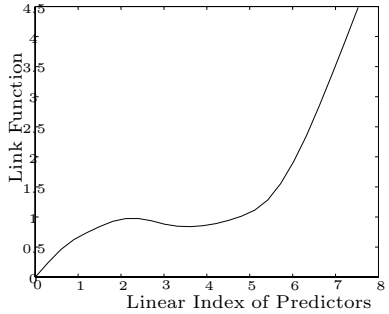

Figure 8. The estimated link functions of the prostate cancer clinical trial.

6. Discussion

A lot of studies have been done on checking the individual forms of covariates in the proportional hazards regression model with a pre-specified link function, but very little has been explored to check the parametric form of the link function. The proposed nonparametric method for data with multivariate covariates provides a graphic tool to suggest or examine a parametric form of the link function. The estimating procedures are easy to use and are distribution free. Simulation results and the asymptotic distribution for the non-parametric link have supported the approach when there are no missing time-dependent covariates.

Even though the proposed FPCARS method for handing intermittent time- dependent covariates are in the framework of the two-stage approach which usually induces bias, the simulation results suggest that the bias is relatively small.

The asymptotic property of the estimated link function with estimated time-dependent covariates is highly technical and an open problem.

Acknowledgements

I wish to thank Jane-Ling Wang for her generous help and comments on this manuscript. I also want to thank the referee and an associate editor for their

insightful commends and suggestions. The research was partially supported by National Science Foundation grant DMS-98-03627, National Institutes of Health grant R01-DK-45939, and National Institute of Allergy and Infectious Diseases grant AI24643.

References

Anderson, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-1120.

Ash, R. B. and Gardner, M. F. (1975). *Topics in Stochastic Processes*. Academic Press, New York.

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

Chiou, J. M., Müller, H. G. and Wang, J. L. (2003). Functional quasi-likelihood regression models with smooth random effects. *J. Roy. Statist. Soc. Ser. B* **65**, 405-423.

Cox, D. R. (1972). Regression models and life-table (with discussion). *J. Roy. Statist. Soc. Ser. B* **4**, 187-220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269-276.

Dafni, U. G. and Tsiatis, A. A. (1998). Evaluating surrogate markers of clinical outcome when measured with error. *Biometrics* **54**, 1445-1462.

DeGruttola, V. and Tu, X. M. (1994). Modeling progression of CD-4 lymphocyte count and its relationship to survival time. *Biometrics* **50**, 1003-1014.

Diggle, P. J. and Verbyla, A. P. (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics* **54**, 301-315.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman Hall, London.

Fan, J., Gijbels, I. and King, M. (1997). Local likelihood and local partial likelihood in hazard regression. *Ann. Statist.* **25**, 1661-1690.

Faucett, C. J. and Thomas, D. C. (1996). Simultaneously modeling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statist. Medicine* **15**, 1663-1685.

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

Henderson, R., Diggle, P. and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465-480.

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley, New York.

Maples, J., Murphy, S. and Axinn, W. (2002). Two-level proportional hazards models. *Biometrics* **58**, 754-763.

Pawitan, Y. and Self, S. (1993). Modeling disease marker processes in AIDS. *J. Amer. Statist. Assoc.* **83**, 719-726.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. Springer-Verlag, New York.

Rice, J. (2003). Functional and longitudinal data analysis perspectives on smoothing. Manuscript.

Rice, J. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Royal Statist. Soc. Ser. B* **53**, 233-243.

Self, S. and Liang, K. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio test under nonstandard conditions. *J. Amer. Statist. Assoc.* **82**, 605-610.

Staniswalis, J. G. and Lee, J. J. (1998). Nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **93**, 1403-1418.

Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation. *J. Amer. Statist. Assoc.* **82**, 559-567.

Tsiatis, A. A. and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88**, 447-558.

Tsiatis, A. A., Degruttola, V. and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error: applications to survival and CD4 counts in patients with AIDS. *J. Amer. Statist. Assoc.* **90**, 27-37.

Wang, W. (2001). Proportional hazards model with unknown link function and applications to time-to-event longitudinal data. Ph.D. Thesis, University of California, Davis.

Wang, W., Wang, J. and Wang, Q. (2001). Proportional hazards model with unknown link function. Technology report, University of California, Davis.

Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175-188.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **55**, 330-339.

Yao, F., Müller, H. G., Clifford, A. J., Dueker, S. R., Follett, J., Lin, Y., Buchholz, B. A. and Vogel, J. S. (2003). Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. *Biometrics* **59**, 676-685.

Department of Biostatistics, School of Public Health, Harvard University, Boston, MA 02115, U.S.A.

E-mail: wwang@jimmy.harvard.edu

(Received May 2003; accepted April 2004)