Open access • Journal Article • DOI:10.1111/J.1467-9868.2007.00630.X

Partially linear hazard regression with varying coefficients for multivariate survival data — Source link

Jianwen Cai, Jianqing Fan, Jiancheng Jiang, Haibo Zhou

Institutions: University of North Carolina at Chapel Hill, Princeton University, University of North Carolina at Charlotte

Published on: 01 Feb 2008 - Journal of The Royal Statistical Society Series B-statistical Methodology (Blackwell Publishing Ltd)

Topics: M-estimator, Asymptotic distribution, Estimator, Linear model and Linear regression

Partially Linear Hazard Regression with Varying-coefficients for Multivariate Survival Data

Jianwen Cai[1], Jianqing Fan[2], Jiancheng Jiang † [2,3] and Haibo Zhou[1]

[1] University of North Carolina at Chapel Hill, NC 27599

[2] Princeton University, NJ 08544

[3] University of North Carolina at Charlotte, NC 28223

Summary. This paper studies estimation of partially linear hazard regression models with varying coefficients for multivariate survival data. A profile pseudo-partial likelihood estimation method is proposed. The estimation of the parameters of the linear part is accomplished via maximization of the profile pseudo-partial likelihood, while the varying-coefficient functions are considered as nuisance parameters profiled out of the likelihood. It is shown that the estimators of the parameters are $\sqrt{n}$-consistent and the estimators of the nonparametric coefficient functions achieve optimal convergence rates. Asymptotic normality is obtained for the estimators of the finite parameters and varying-coefficient functions. Consistent estimators of the asymptotic variances are derived and empirically tested, which facilitate inference for the model. We prove that the varying-coefficient functions can be estimated as well as if the parametric components were known and the failure times within each subject were independent. Simulations are conducted to demonstrate the performance of the proposed estimators. A real dataset is analysed to illustrate the proposed methodology.

Keywords: Local pseudo-partial likelihood, Marginal hazard model, Multivariate failure time, Partially linear, Profile pseudo-partial likelihood, Varying-coefficients.

Running Short Title: Partially Linear Hazard Regression.

†Address for correspondence: Department of Mathematics and Statistics, University of North Carolina at Charlotte, NC 28223, USA. E-mail: jjiang1@uncc.edu

1. Introduction

Multivariate survival data are frequently encountered in data analysis. A key feature of this type of data is that the failure times might be correlated. For example, in animal experiments, the failure times of animals within a litter may be correlated because they share common genetic and environmental traits; in clinical trials where the patients are followed for repeated recurrent events, the times between recurrences for a given patient may be correlated. Usually, the structures of correlation are unknown. Modeling the multivariate failure times without specifying a correlation structure has been an active field of research in statistical literature.

A popular approach for modeling multivariate failure data is the so-called marginal hazard model approach which models the "population-averaged" covariate effects. This model is attractive especially when the correlation among observations is not of interest. The model also is linked with the Cox model in the univariate case because of its semiparametric structures. It has received much attention in the literature. See for example, Wei, Lin and Weissfeld 1989, Lee, Wei and Amato 1992, Liang, Self and Chang 1993, Lin 1994, Cai and Prentice 1995, 1997, Prentice and Hsu 1997, Spiekerman and Lin 1998, Cai 1999, and Clegg, Cai, and Sen 1999 among others.

Most statistical methods developed for handling the failure time data typically assume that the covariate effects on the logarithm of the hazard function are linear and the regression coefficients are constant. These assumptions, however, are mainly chosen for their mathematical convenience. True associations in practical studies are usually more complex than a simple linear relationship. An important extension of the constant coefficient model is the varying coefficient model, which addresses an issue frequently encountered by investigators in practical studies. For instance, the effect of an exposure variable on the hazard function may change with the level of a confounding covariate. This can be traditionally modelled by including an interaction term in the model for simplicity, but when the effect of the exposure on the hazard function changes nonlinearly with the confounding variable this approach may introduce a large modeling bias. An illustrative example is the well-known Framingham Heart Study (Dawber 1980). There were totally 2,336 men and 2,873 women in the study. The investigators were interested in the effect of the body mass index (BMI) on the time to coronary heart disease (CHD) and cerebrovascular accident (CVA), where the effect could vary over different birth cohorts. To model possible birth cohort effects of the BMI on the failure time (the times to CHD and CVA), one needs to use a varying-coefficient model with the coefficient for the BMI being an unknown function of the year of birth. The varying-coefficient structure allows one to model possible complex interaction between the BMI and the birth cohort. In general, there may be several exposure variables which interacts with a confounding covariate. This leads to a multivariate varying-coefficient model with the coefficients of variables changing nonlinearly over the level of the confounding variable.

Varying-coefficient models have received much attention in the analysis of non-failure time data. Related work appears in the literature on multivariate nonparametric regression, generalized linear models, analysis of longitudinal data, and nonlinear time series, etc. See, for example, Hastie and Tibshirani (1993), Brumback and Rice (1998), Carrol *et al.* (1998), Hoover *et al.* (1998), Fan and Zhang (1999), and Cai, Fan and Yao (2000), among others. For univariate failure time data, Fan, Lin and Zhou (2006) studied the estimation of varying-coefficient hazard model based on nonparametric smoothing techniques. This approach was extended to model multivariate failure data by Cai, Fan, Zhou and Zhou (2007) using a local pseudo-partial likelihood procedure. While this approach seems appealing in addressing the interactions among covariates, it ignores possible linear structure in the hazard regression and hence would suffer from the loss of efficiency when some coefficients are indeed constant. Therefore, for modelling the multivariate failure time data without specifying a correlation structure, there is a genuine need to consider a partially linear hazard regression model with varying-coefficients, under the marginal hazard model framework.

To our knowledge there is no formal work elaborating this problem in the literature. It is important to develop an effective estimation methodology for the partially linear model. This paper addresses this problem by using the idea of profile likelihood. We develop a profile local pseudo-likelihood-based approach for estimating the varying-coefficient functions $\alpha(\cdot)$ and a global profile pseudo-likelihood-based method for estimating the finite parameter vector β as specified in the model (1.1) below.

A recent article of Cai, Fan, Jiang and Zhou (2007) considered the partially linear hazard regression model with a one dimensional nonlinear component for modelling multivariate failure data, under the marginal hazard model framework. This model is useful to model nonlinear covariate effects, but it cannot deal with possible interaction among covariates, such as the BMI and birth cohort covariates mentioned above.

Suppose that there is a random sample of n subjects from an underlying population and that there are J failure types in each subject. Let i indicate subject and (i, j) denote the jth failure type in the ith subject. Let T_{ij} $(i = 1, \cdots, n,\ j = 1, \cdots, J)$ denote the failure time, C_{ij} $(i = 1, \cdots, n,\ j = 1, \cdots, J)$ the censoring time, and $X_{ij} = min(T_{ij}, C_{ij})$ the observed time. Let Δ_{ij} be an indicator which equals 1 if X_{ij} is a failure time and 0 otherwise. Let $\mathcal{F}_{t,ij}$ represent the failure, censoring and covariate information up to time t for the (i, j) failure type as well as the covariate information of the other failure types in the ith subject up to time t. The marginal hazard function is defined as $\lambda_{ij}(t) = \lim_{h\downarrow 0} h^{-1}\mathcal{P}[T_{ij} \le t + h | T_{ij} > t, \mathcal{F}_{t,ij}]$. The censoring time is assumed to be independent of the failure time conditional on the covariates (that is the so-called "independent censoring scheme"). Throughout this paper, for any vector b we use notation b^T to denote the transpose of b.

The partially linear hazard regression model we consider is

$$\lambda_{ij}(t) = \lambda_{0j}(t)\exp\{\beta^T W_{ij}(t) + \alpha(V_{ij}(t))^T Z_{ij}(t)\}, \tag{1.1}$$

where $W_{ij}(\cdot) = (W_{ij1}(\cdot), \cdots, W_{ijq}(\cdot))^T$ is a vector of covariates that has linear effects on the logarithm of the hazard, $Z_{ij}(\cdot) = (Z_{ij1}(\cdot), \cdots, Z_{ijp}(\cdot))^T$ is a vector of covariates that may interact with some exposure covariate $V_{ij}(\cdot)$; $\lambda_{0j}(\cdot)$ is an unspecified baseline hazard function; and $\alpha(\cdot)$ is a vector of unspecified coefficient functions. For example, in the aforementioned FHS study, V would represent the calendar year of birthdate, W would consist of confounding variables such as gender, blood pressure, cholesterol level and smoking status, etc, and Z would contain covariates possibly interacting with V such as the BMI. Our interest centers around the efficient estimation of parameter vector β and coefficient functions $\alpha(\cdot)$.

We will show that the proposed estimator of β is root-n consistent and that the estimators of α achieves optimal convergence rates. The asymptotic normality will be established for both β and $\alpha(\cdot)$. We will also provide consistent estimators for the asymptotic variances of the proposed estimators, in order to facilitate the inference for the parameters. One particular challenge in developing those large sample properties lies in the fact that the profile pseudo-partial likelihood involves the estimation of coefficient functions $\alpha(\cdot)$ which uses all observed information from the data, hence the score function of β cannot be expressed asymptotically as an integration of a predictable process with respect to a martingale. Consequently, the commonly used martingale tools cannot be straightforwardly applied. Obtaining the asymptotic properties of the estimators is a major challenge, and a determined effort has been made in this article to derive the asymptotic distributions of the proposed estimators.

This paper is organized as follows. In Section 2, we describe the procedure for estimating the parameters β and the coefficient functions $\alpha(\cdot)$ from model (1.1). Section 3 focuses on the asymptotic properties of the proposed estimators. Issues on the implementation of the proposed methods are discussed in Section 4. In Section 5, we conduct intensive simulations and illustrate the proposed estimation via a real data analysis. Some technical conditions are put in Appendix I. Proofs of the theorems are given in Appendix II.

2. Maximum Pseudo-partial Likelihood Estimation

We propose the following pseudo-partial likelihood function for estimating the regression coefficient vector β and the coefficient functions $\alpha(.)$. Let $\mathcal{R}_j(t) = \{i : X_{ij} \geq t\}$ denote the set of the individuals at risk just prior to time t for failure type j. If failure times from the same subject were independent, then the partial likelihood for (1.1) is

$$L(\beta, \alpha) = \prod_{j=1}^{J} \prod_{i=1}^{n} \left\{ \frac{\exp\{\beta^T W_{ij}(X_{ij}) + \alpha(V_{ij}(X_{ij}))^T Z_{ij}(X_{ij})\}}{\sum_{l \in \mathcal{R}_j(X_{ij})} \exp\{\beta^T W_{lj}(X_{ij}) + \alpha(V_{lj}(X_{ij}))^T Z_{lj}(X_{ij})\}} \right\}^{\Delta_{ij}}. \quad (2.1)$$

For the case with $J = 1$, if the coefficient functions are constant, the partial likelihood above is just the one in Cox's model (Cox 1972).

Since failure times from the same subject are dependent, the above partial likelihood is referred to as pseudo-partial likelihood. We will use this pseudo-partial likelihood for our estimation. However, we neither require that the failure times are independent nor specify a dependent structure among failure times. This furnishes robustness of our estimation method against the misspecification of correlations among failure times.

If $\alpha(\cdot)$ has been parameterized, one can obtain the maximum pseudo-partial likelihood estimator by maximizing (2.1) with respect to β and the parameters in $\alpha(\cdot)$. Since the form of α has not been specified in our nonparametric method, we can only use its qualitative traits.

Next we suggest a naive nonparametric estimator. Assume that $\alpha(\cdot)$ is smooth so that it can be approximated locally by a linear function. Denote by $f_j(\cdot)$ the density of V_{1j}. For any given point $v_0 \in \sum_{j=1}^{J} \text{supp}(f_j)$, where $\text{supp}(f_j)$ denotes the support of $f_j(\cdot)$, by Taylor's expansion,

$$\alpha(v) \approx \alpha(v_0) + \alpha'(v_0)(v - v_0) \equiv \delta + \eta(v - v_0), \tag{2.2}$$

where $\delta = (\delta_1, \cdots, \delta_p)^T$. Using the local model (2.2) for the data around v_0, we obtain the logarithm of the local pseudo-partial likelihood:

$$\ell(\beta, \gamma) = \sum_{j=1}^{J} \sum_{i=1}^{n} K_h(V_{ij}(X_{ij}) - v_0)\Delta_{ij}\big\{\beta^T W_{ij}(X_{ij}) + \gamma^T U_{ij}(X_{ij}, v_0) - R^*_{ij}(\beta, \gamma)\big\}, \tag{2.3}$$

where $U_{ij}(u, v_0) = \{Z_{ij}(u)^T, Z_{ij}(u)^T(V_{ij}(u) - v_0)\}^T$, $\gamma = (\delta^T, \eta^T)^T$, $K_h(\cdot) = K(\cdot/h)/h$, and

$$R^*_{ij}(\beta, \gamma) = \log\Big(\sum_{l \in \mathcal{R}_j(X_{ij})} \exp[\beta^T W_{lj}(X_{ij}) + \gamma^T U_{lj}(X_{ij}, v_0)]K_h(V_{lj}(X_{ij}) - v_0)\Big).$$

Here K is a probability density called a kernel function, and h represents the size of the local neighborhood called a bandwidth. The kernel function is introduced to reflect the fact that the local model (2.2) is only applied to the data around v_0. It gives a larger weight to the data closer to the point v_0. For the univariate case, the local pseudo-partial likelihood was derived by Fan *et al.* (1997) from a local maximum likelihood point of view.

Let $(\hat{\beta}(v_0), \hat{\gamma}(v_0))$ maximize the local pseudo-partial likelihood (2.3). Then, an estimator of $\alpha(\cdot)$ at the point v_0 is simply the local intercept $\hat{\delta}(v_0)$, namely $\hat{\alpha}(v_0) = \hat{\delta}(v_0)$. When v_0 varies over a grid of prescribed points, the estimates of the functions are obtained.

In the context of the generalized linear models, Carroll *et al.* (1997) show that such a naive method produces an estimator for α that achieves the optimal rate of convergence. However, the asymptotic variance for estimating α has been inflated. Since only the local data are used in the estimation of β, the resulting estimator for β cannot be root-n consistent. We refer to $(\hat{\beta}(v_0), \hat{\alpha}(v_0))$ as the naive estimator. To fix the drawbacks of the naive estimator, we next propose a new estimator for β that is root-n consistent.

Our proposed estimator is based on a profile likelihood. Specifically, for a given β, we obtain an estimator of $\hat{\alpha}(\cdot, \beta)$ by maximizing (2.3) with respect to γ. Substituting the estimator $\hat{\alpha}(\cdot, \beta)$ into (2.1), we obtain the logarithm of the profile pseudo-partial likelihood:

$$\begin{aligned} \ell_p(\beta) \quad = \quad & \sum_{j=1}^{J} \sum_{i=1}^{n} \Delta_{ij} \Big\{ \beta^T W_{ij} + \hat{\alpha}(V_{ij}, \beta)^T Z_{ij} \\ & - \log\Big(\sum_{l \in \mathcal{R}_j(X_{ij})} \exp[\beta^T W_{lj} + \hat{\alpha}(V_{lj}, \beta)^T Z_{lj}] \Big) \Big\}. \end{aligned} \tag{2.4}$$

Here and hereafter, for the ease of presentation, we sometimes drop the dependence of covariates on time, with the understanding that the methods developed in this paper are applicable to external time dependent covariates (Kalbfleisch and Prentice, 2002). Let $\hat{\beta}$ maximize (2.4). Our proposed estimator for the parametric component is simply $\hat{\beta}$ and for the coefficient function is $\hat{\alpha}(\cdot) = \hat{\alpha}(\cdot, \hat{\beta})$.

With the estimators of β and $\alpha(\cdot)$, one can estimate the cumulative baseline hazard function $\Lambda_{0j}(t) = \int_0^t \lambda_{0j}(u) du$ under mild conditions by a consistent estimator:

$$\hat{\Lambda}_{0j}(t) = \int_0^t [\sum_{i=1}^{n} Y_{ij}(u) \exp\{\hat{\beta}^T W_{ij}(u) + \hat{\alpha}(V_{ij}(u))^T Z_{ij}(u)\}]^{-1} \sum_{i=1}^{n} dN_{ij}(u),$$

where $Y_{ij}(\cdot)$ is an at risk indicator process for the jth failure type of subject i, i.e. $Y_{ij}(t) = I(X_{ij} \geq t)$, and $N_{ij}(u) = 1(X_{ij} \leq u, \Delta_{ij} = 1)$.

3. Asymptotic Properties

The technical challenges of deriving the property of the proposed estimator $\hat{\beta}$ arise from the fact that the logarithm of the profile pseudo-partial likelihood $\ell_p(\beta)$ in (2.4) involves $\hat{\alpha}(\cdot, \beta)$ which utilizes all observed information. Consequently, the score function of β cannot be expressed asymptotically as a sum of integrations of predictable processes with respect to martingales. In other words, the score function cannot be expressed asymptotically as a sum of martingales. Hence commonly-used martingale methods cannot be directly applied.

To derive the asymptotic properties of our estimators, we need some notations and technical conditions which are relegated to Appendix I for ease of exposition. The following theorems demonstrate that our estimators are consistent and asymptotically normal.

Theorem 1. Under Conditions (i)-(viii) in Appendix I, with probability tending to one there exists a sequence of estimators $\hat{\beta}$ which maximizes the global profile pseudo-partial likelihood $\ell_p(\beta)$ such that $\hat{\beta} \xrightarrow{P} \beta_0$.

Theorem 2. Under Conditions (i)-(viii) in Appendix I, if $nh^2 \to \infty$ and $nh^4 \to 0$ then the estimation sequence in Theorem 1 satisfies that $\sqrt{n}(\hat{\beta} - \beta_0)$ converges to a Gaussian distribution with mean zero and covariance matrix $\Omega = I^{-1}\Sigma I^{-1}$, specified in Appendix I.

From Theorem 2, the asymptotic covariance matrix of $\hat{\beta}$ is of sandwich form, which can be estimated by $\hat{\Omega} = \hat{I}^{-1}\hat{\Sigma}\hat{I}^{-1}$, where $\hat{I}$ and $\hat{\Sigma}$ are empirical plug-in estimators of I and Σ, respectively, which are defined in Appendix I.

Note that $\hat{I}$ and $\hat{\Sigma}$ can be shown to be consistent for I and Σ, respectively. Hence, $\hat{\Omega}$ is a consistent estimator of Ω under the conditions of Theorem 2. Then for the following semiparametric testing problem:

$$H_0 : \beta = \beta_0 \text{ versus } H_1 : \beta \neq \beta_0,$$

where $\alpha(\cdot)$ is a vector of nuisance functions, a generalized Wald test statistic W_n can be defined as

$$W_n = n(\hat{\beta} - \beta_0)^T \hat{\Omega}^{-1} (\hat{\beta} - \beta_0). \tag{3.1}$$

In particular, this can be applied for testing whether a set of variables is statistically significant in the semiparametric model. By Theorem 2, we have the following results.

Theorem 3. Under Conditions of Theorem 2, the asymptotic null distribution of W_n is $\chi^2(q)$, where q is the dimension of β.

Theorem 4. Assume that Conditions (i)-(v) in Appendix I hold. If $\hat{\beta}$ is $\sqrt{n}$-consistent and nh^5 is bounded, then

$$\sqrt{nh}\big[H(\hat{\gamma} - \gamma) - b_n(v_0)\big] \xrightarrow{\mathcal{L}} \mathcal{N}(0, Q(v_0)),$$

where $b_n(v_0) = \frac{1}{2}h^2\mu_2(\alpha''(v_0)^T, 0_p^T)^T + o_p(h^2)$ with 0_p being a $p \times 1$ vector of all elements zeros, $Q(v_0) = \text{blockdiag}\{\nu_0 \tilde{A}_1^{-1}(v_0), \nu_2\mu_2^{-2}\tilde{A}_2^{-1}(v_0)\}$, and H is a $p \times p$ diagonal matrix whose first p diagonal elements are ones and others are h's.

Corollary 1. Under the assumptions of Theorem 4,

$$\sqrt{nh}\Big[\hat{\alpha}(v_0) - \alpha(v_0) - h^2\alpha''(v_0)\mu_2/2 + o_p(h^2)\Big] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu_0 \tilde{A}_1^{-1}(v_0)).$$

Remark 1. It is interesting to note that if the marginal distributions of different failure types within each subject are the same, then the asymptotic property of the proposed estimator for $\alpha(\cdot)$ reduces to that of Fan, Lin and Zhou (2006). That is, even though the failure types within subjects are correlated, our estimators of the coefficient functions perform as well as if they were independent. For an insight into this phenomena, see the work of Masry and Fan (1997) and Jiang and Mack (2001).

Remark 2. Note that when β_0 is known, one would maximize (2.4) with respect to γ. The resulting estimator has the same asymptotic normality as in Corollary 1. In other words, $\hat{\alpha}(v)$ is adaptive in the sense that it estimates $\alpha(v)$ as well as if β_0 were known.

As a result of Corollary 1, the theoretic optimal bandwidth for estimating $\alpha_k(\cdot)$, in the sense of minimizing the asymptotic weighted mean integrated squared error, is

$$h_{opt} = \Big[\int \Gamma_k(v)w(v)\,dv\Big]^{1/5}\Big[\mu_2^2\int[\alpha_k''(v)]^2 w(v)\,dv\Big]^{-1/5} n^{-1/5}, \tag{3.2}$$

where Γ_k and α_k are the kth components of $\nu_0\tilde{A}_1^{-1}$ and α, respectively.

4. Implementation of the Proposed Methods

The proposed profile likelihood estimator involves maximization in (2.3) and (2.4). It can be computed by an algorithm similar to that in Cai, Fan, Jiang and Zhou (2007). The algorithm takes cares of the fact that $\hat{\alpha}(\cdot,\beta)$ is implicitly defined. Let v_k $(k = 1,\cdots,m)$ be a grid of points on the range of the exposure variable V. Then the algorithm proceeds as follows.

1. *Initialization.* Use the average of the naive estimator $\bar{\beta} = m^{-1}\sum_{k=1}^m \hat{\beta}(v_k)$ as the initial value. Set $\hat{\beta} = \bar{\beta}$. This step can be undersmoothed for reducing biases, as the average reduces the variances.
2. *Estimation of coefficient function.* Maximize the local pseudo-partial likelihood $\ell(\hat{\beta},\gamma)$ at each grid point v_k and obtain the nonparametric estimator $\hat{\alpha}(\cdot,\hat{\beta})$ at these grid points. Obtain the nonparametric estimator at points $\{V_{ij}\}$ by using the linear interpolation. We take h suitable for estimation of β. One example for such a suitable bandwidth is the ad hoc bandwidth in (4.1) below.
3. *Estimation of parametric component.* With the estimator $\hat{\alpha}(\cdot,\hat{\beta})$, maximize the profile estimator $\ell_p(\beta)$ with $\alpha(\cdot,\beta) = \hat{\alpha}(\cdot,\hat{\beta})$, using the Newton-Raphson algorithm and the initial value $\hat{\beta}$ from the previous step.
4. *Iteration.* Iterate between steps 2 and 3 until convergence.
5. *Re-estimating the coefficient function.* Fix β at its estimated value from step 4. The final estimate of $\hat{\alpha}(\cdot)$ is $\hat{\alpha}(\cdot,\hat{\beta})$. At this final step we take bandwidth h suitable for estimating $\alpha(\cdot)$ such as the estimated optimal bandwidth $\hat{h}_{opt}$ based on (3.2).

In step 1, the estimators $\hat{\beta}(v_k)$'s are nonparametric ones from (2.3) using only those local data around the point v_k, and hence they and their average are consistent estimators of β in a large range of bandwidth. In (2.3), we used local linear smoothing, and hence the biases are of order $O(h^2)$ and variance are of order $O(1/\sqrt{nh})$. Using an optimal bandwidth ($h \propto n^{-1/5}$) in smoothing will result in the biases and variances of order $O(n^{-2/5})$. By undersmoothing i.e. using a bandwidth less than the optimal one, the bias of each of estimators $\hat{\beta}(v_j)$ is reduced. Even though the variance of each estimator increases, the average offsets this variance increase.

Since the initial estimator $\bar{\beta}$ is consistent, we do not expect many iterations in step 4. In implementation of the estimation method, an undersmoothing bandwidth $h = O(n^{-1/3})$ is used in step 1. Then the initial estimator in step 3 has at least the nonparametric rate $O_p(h^2 + 1/\sqrt{nh}) = O_p(n^{-1/3})$, one-step or two-step iterations in the Newton-Raphson algorithm suffices. This is backed by the theoretical work of Bickel (1975) and Robinson (1988) in parametric models and by Fan and Chen (1999) and Fan and Jiang (2000) in nonparametric models. In fact, according to Robinson (1988), if an initial parametric estimator has a rate $O(n^{-a})$, the difference between the k-step Newton-Raphson estimator and the maximum likelihood estimator is only of order $O_p(n^{-ak})$. With $a = 1/3$ and $k = 2$, the order of error is $o(n^{-1/2})$. Therefore, two iterations in the above algorithm would produce an estimator as efficiently as the fully iterated estimator. Our experience in simulations shows that the results are quite promising and in line with the above theory.

The estimation procedure involves the choice of a smoothing parameter h on two quite different levels. In steps 2-3 of the algorithm the aim is to estimate β, and hence the bandwidth h should be suitable for this task. From our theoretic result in Section 3, a large range of bandwidths satisfies those theoretical requirements. For example, one can employ the following ad hoc bandwidth

$$\hat{h}_{opt} \times n^{\frac{1}{5}} \times n^{-\frac{1}{3}} = \hat{h}_{opt} \times n^{-2/15}, \tag{4.1}$$

where $\hat{h}_{opt}$ is the optimal bandwidth estimated for $\alpha(\cdot)$ based on (3.2). This guarantees that the required bandwidth has correct order of magnitude. In the step 5, however, the goal is to estimate the nonparametric component $\alpha(\cdot)$, and hence the bandwidth h should be optimal in this respect.

5. Numerical Studies

This section illustrate the performance of proposed method by using simulated example and a real data example. The results were obtained using Matlab and Splus softwares, which are available at *http://www.princeton.edu/~jqfan/papers/06/PLVC/*??????

5.1. Simulations

In this section, we evaluate the performance of the proposed estimation procedure in finite samples. The aim of the simulations is three-fold: to demonstrate that our method correctly captures the forms of nonparametric coefficient functions and accurately estimates the parametric components, to assess if the proposed variance estimation is consistent, and to illustrate that the partly linear hazard regression model with varying coefficients can reveal interaction among covariates while usual Cox's models with interactive effects of factors fail.

Multivariate failure times were generated from a multivariate extension of the model of Clayton and Cuzick (1985) in which the joint survival function for $(T_1, \cdots, T_J)$ given $(V_1, \cdots, V_J)$, $(Z_1, \cdots, Z_J)$ and $(W_1, \cdots, W_J)$ is

$$F(t_1, \cdots, t_J; V_1, \cdots, V_J, Z_1, \cdots, Z_J, W_1, \cdots, W_J) = \{\sum_{j=1}^{J} S_j(t_j)^{-1/\theta} - (J-1)\}^{-\theta},$$

where $S_j(t)$ is the marginal survival probability for the jth failure type, depending on the covariates V_j, Z_j and W_j to be specified below. Note that θ is a parameter which represents the degree of dependence of T_j and $T_{j'}$. Specifically, a small value of θ represents strong positive dependence with $\theta \to 0$ giving maximal positive dependence, while a very large θ gives nearly independence. The relationship between Kendall's tau and θ is $\tau = 1/(2\theta+1)$.

The marginal distribution of T_{1j} was taken to be exponential with failure rate

$$\lambda_{0j} \exp\{\beta^T W_j + \alpha(V_j)^T Z_j\},$$

entailing the marginal survival function

$$S_j(t) = \exp\Big\{-t\lambda_{0j} \exp[\beta_0^T W_j + \alpha(V_j)^T Z_j]\Big\}.$$

We considered the settings with $n = 100$, 200, $J = 2$ and $(p, q) = (1, 3)$. The baselines $\lambda_{01} = 1$ and $\lambda_{02} = 2$ were used. The true parameter was set as $\beta_0 = (0.8, 0.6, 1.0)^T$. We first simulated $Z_{ij} \overset{iid}{\sim} N(0, 1)$ and $W_{ij} = (W_{ij}^{(1)}, W_{ij}^{(2)}, W_{ij}^{(3)})^T$ with components being independently sampled from the Bernoulli distribution (taking 1 or 0 each with probability 0.5), $N(0, 1)$ and $U(0, 1)$, respectively. The covariate V_{ij} was set as $W_{ij}^{(3)}$. The coefficient function was taken as $\alpha(v) = 0.5 + 3v$ or $\alpha(v) = 2 - 3\cos((v - 0.5)\pi/2)$. The censoring time distribution was generated from an exponential distribution with mean chosen to produce a certain percentage of censoring.

To gauge the relative performance of partly linear hazard model with varying coefficients, we compare it with Cox's model with the first-order interaction. Specifically, we consider fitting the data to the following model for the marginal distribution of T_j:

$$\lambda_{0j} \exp\{\beta_1 W_{1j}^{(1)} + \beta_2 W_{1j}^{(2)} + \beta_3 V_{1j} + \alpha_0 Z_{1j} + \alpha_1 V_{1j} Z_{1j}\}. \tag{5.1}$$

This specification of the model is correct when $\alpha(v) = 0.5 + 3v$ but incorrect when $\alpha(v) = 2 - 3\cos((v - 0.5)\pi/2)$. By using the Splus function **c**$oxph$ with **s**$trata(\cdot)$ and **c**$luster(id)$ options, it is easy to fit the above model. We denote by $\tilde{\beta}$ the fitted parameters' value for β. This estimator can serve as a benchmark when $\alpha(v) = 0.5 + 3v$.

The number of replications was 500. The parameter θ was set as 10 and 0.1 which correspond to weak and strong correlation within each subject. The Epanechnikov kernel function was employed for smoothing. By the argument in Section 4, the bandwidth h was taken as $\hat{h} = cn^{-1/3}$ for estimation of β in steps 2-3, and $cn^{-1/5}$ for estimation of $\alpha(\cdot)$ in step 5 of the algorithm, with $c = 0.4$ and 0.3 respectively for the cases of $\alpha(v) = 0.5 + 3v$ and $2 - 3\cos((v - 0.5)\pi/2)$.

Table 1. Summary of Results in Simulations with $\alpha(v) = 0.5 + 3v$.

Size	Model		41% Censoring							
(n, J)	θ	β	Mean($\hat\beta$)	Mean($\tilde\beta$)	SD($\hat\beta$)	SD($\tilde\beta$)	se($\hat\beta$)	se($\tilde\beta$)	$CP(\hat\beta)$	$CP(\tilde\beta)$
		β_1	0.83	0.82	0.22	0.21	0.25	0.20	0.952	0.928
(100,2)	10	β_2	0.62	0.62	0.11	0.11	0.13	0.11	0.966	0.938
		β_3	1.01	1.02	0.41	0.39	0.50	0.38	0.960	0.948
		β_1	0.82	0.81	0.22	0.21	0.25	0.20	0.962	0.936
(100,2)	0.1	β_2	0.62	0.62	0.12	0.12	0.14	0.11	0.960	0.920
		β_3	0.99	1.00	0.42	0.40	0.48	0.38	0.960	0.928

The estimators and their standard deviations (SD) for the finite parameters were evaluated along with the average of the estimated standard error (se) for the estimators. The coverage probability (CP) of the 95% confidence intervels for β was also calculated based on the normal approximation.

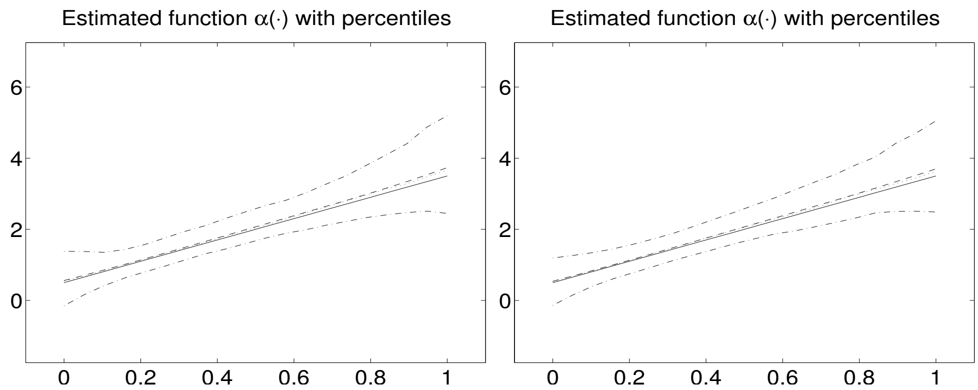

Fig. 1. *Simulation Results for* $\alpha = 0.5 + 3v$. Estimated curves with confidence intervals. Left panel: $\theta = 10$; right panel: $\theta = 0.1$. Dash-dotted: the 2.5th and 97.5th percentiles; dashed: the mean; dotted: the median.

Tables 1 and 2 report the simulation results with 41% and 37.4% censoring, respectively. It is evident from Table 1 that the proposed estimator performs similarly to the parametric partial likelihood estimator for Cox's model (5.1), except for a little larger SD. This demonstrates that our estimators for parametric components are quite efficient and are indeed

Table 2. Summary of Results in Simulations with $\alpha(v) = 2 - 3\cos((v - 0.5)\pi/2)$.

Size	Model		37.4% Censoring							
(n, J)	θ	β	Mean($\hat{\beta}$)	Mean($\tilde{\beta}$)	SD($\hat{\beta}$)	SD($\tilde{\beta}$)	se($\hat{\beta}$)	se($\tilde{\beta}$)	$CP(\hat{\beta})$	$CP(\tilde{\beta})$
		β_1	0.81	0.78	0.14	0.20	0.14	0.16	0.944	0.912
(200,2)	10	β_2	0.61	0.59	0.07	0.08	0.08	0.10	0.962	0.926
		β_3	1.02	0.92	0.24	1.58	0.27	2.08	0.946	0.886
		β_1	0.81	0.77	0.14	0.23	0.15	11.61	0.960	0.930
(200,2)	0.1	β_2	0.60	0.57	0.08	0.13	0.08	11.79	0.956	0.904
		β_3	1.01	0.61	0.24	3.00	0.24	326.55	0.956	0.898

comparable to the estimators for the true parametric model. From Table 2, we see that the first-order interactive model (5.1) gave very biased estimators of the parameters, but our method had very little bias and gave reasonable variance estimation and covarage probability. It demonstrates that our method continues to work while the first-order interactive model fails due to its mis-specification. Some values of $SD(\tilde{\beta})$ and $se(\tilde{\beta})$ in Table 2 are very large because of the presence of outliers. This shows that our partly linear hazard model can avoid possible model mis-specification errors.

To appreciate the sampling properties of the estimated coefficient functions, we present in Figures 1 and 2 the 2.5th, 50th (median) and 97.5th percentiles of the estimated curves for the coefficient function $\alpha(\cdot)$ under different settings. The 2.5th and 97.5th percentiles form a 95% pointwise confidence interval for the coefficient function. This assesses the variability of the estimated functions at each point. It is evident that the estimator of nonparametric component does not heavily depend on the correlation within each subject. This is consistent with our claim in Remark 1.

By varying the bandwidths used above over the range of $[\frac{2}{3}\hat{h}, \frac{3}{2}\hat{h}]$, we found that the results for the parametric part are very similar. This reflects that the estimation of finite parameters is robust against the choice of the bandwidth over a reasonable range. The results are omitted to save space.

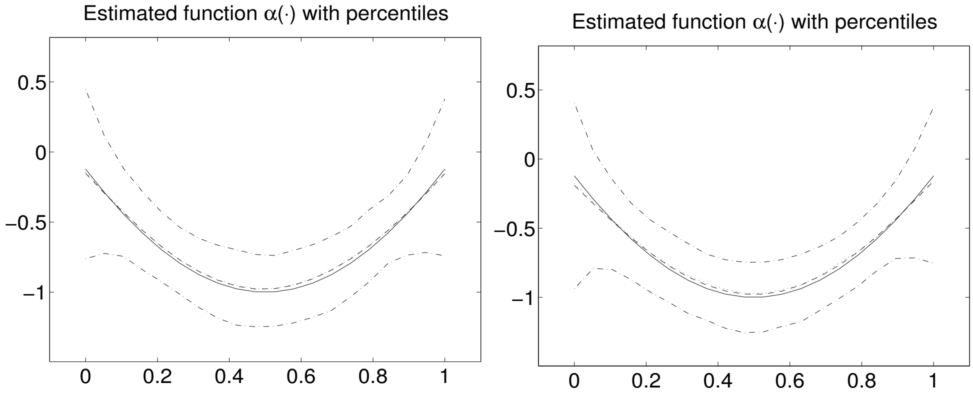

Fig. 2. *Simulation Results for* $\alpha(v) = 2 - 3\cos((v - 0.5)\pi/2)$. Estimated curves with confidence intervals. Left panel: $\theta = 10$; right panel: $\theta = 0.1$. Dash-dotted: the 2.5th and 97.5th percentiles; dashed: the mean; dotted: the median.

5.2. Applications to FHS dataset

In this section, we apply our proposed procedure to analyze data from the well-known Framingham Heart Study (Dawber 1980). The Framingham Heart Study began in 1948. The cohort consisted of 2,336 men and 2,873 women. At the first examination, the participants were between 30 and 62 years of age, and they were recalled and examined every two years after their entry into the study. Times until coronary heart disease (CHD) and cerebrovascular accident (CVA) were recorded. Those times recorded from the same individual might be correlated. The dataset used here included all participants in the study who had an examination at age 44 or 45, which we refer to as the "age 45" exam, and were disease-free at that examination in the sense that there existed no history of hypertension or glucose intolerance and no previous experiences of a CHD or CVA. There were a total of 1571 disease-free individuals. The percentage of censoring was about 90.42%. The risk factors of interest were age, gender, systolic blood pressure (SBP), body mass index (BMI), cholesterol level, and cigarette smoking. Clegg *et al.* (1999) previously analyzed the dataset based on a marginal mixed baseline hazards model, where the effects of all of the covariates were specified as linear in the marginal regression. To explore the possible interaction among covariates (e.g. BMI and age), we used the proposed method to assess the asso-

Table 3. Estimated Parameters for the Framingham Heart Study data. $\hat{\beta}$ - the estimated parameters, $\hat{se}$ - the standard error of $\hat{\beta}$.

Effect	$\hat{\beta}$	$\hat{se}$	P-value
Age at the "age 45" exam	0.0312	0.1528	0.8383
Cholesterol, mg/dl	0.0041	0.0018	0.0217
Systolic blood pressure, mm Hg	0.0161	0.0065	0.0132
Smoking status: yes=1,no=0	0.3732	0.2856	0.1912
Gender: female=1, male=0	-0.5654	0.2342	0.0158
Waiting time, year	0.1687	0.0832	0.0425

ciation between these risk factors and the times to CHD and CVA. Specifically, times to CHD and CVA were measured from the time at the "age 45" exam to the occurrences of the corresponding diseases. We employed the following hazards model:

$$\lambda_{ij}(t; W_{ij}, Z_{ij}) = \lambda_{0j}(t) \exp[\beta^\tau W_{ij} + \alpha(V_{ij}) Z_{ij}],$$

where

$$W_{ij} = (\text{Age at the "age 45" exam}, \text{SBP}, \text{Cholesterol}, \text{Smoking status}, \text{Gender}, \text{Waiting Time})^\tau,$$

V_{ij} = Waiting time, Z_{ij} = BMI. The variables BMI, SBP, cholesterol level, and smoking status were measured at the "age 45" exam. "Waiting time" was the time elapsed from the initial entry into the FHS study to the "age 45" exam. It is included in the model to adjust for birth cohort effect. Note that the year of birth equals to (1948 + waiting time - 45).

Table 3 reports the estimated parameters and their estimated standard errors along with their P-values from the Wald test in (3.1). It is evident that all of the selected risk factors are statistically significant at the 0.05 significance level except for Smoking status and Age at the "age 45" exam. The nonsignificance result for Age at the "age 45" exam is expected since, by definition, the values for Age at the "age 45" exam do not vary much among subjects. Figure 5.2(a) shows the estimated coefficient function $\alpha(\cdot)$ along with the 95% confidence intervals. It reveals that the effect of BMI is smaller for participants who had longer waiting time. Notice that age at the "age 45" exam equals age at the initial entry

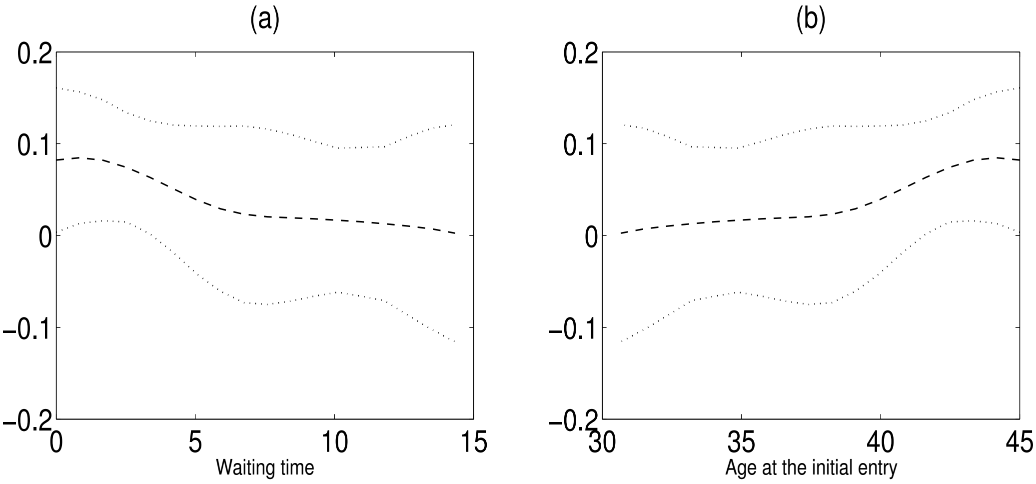

Fig. 3. **Estimated coefficient function $\alpha(\cdot)$ with confidence intervals for FHS Dataset.** Dashed: estimated coefficient functions; dotted: 95% pointwise confidence intervals.

into the FHS study plus waiting time, and waiting time of 0-15 years corresponds to the age at the initial entry into the FHS study from 45 (born in 1903) to 30 (born in 1918) years old. Figure 5.2(b) is a mirror image of Figure 5.2(a). The x-axis in the figure is the age at the initial entry into the FHS study which was calculated as 45 minus waiting time, so the x-axis now starts at 30 and goes on to 45. Therefore, Figure 5.2(b) demonstrates the birth cohort effect of BMI: It is bigger for participants who were older at the initial entry into the FHS study. It is interesting to note that there seems to be a turning point around waiting time 7 for the estimated function $\hat{\alpha}(v)$. On the left region $[0, 7]$ of the turning point, the slope of $\hat{\alpha}(v)$ seems to be bigger than that on the right region. However, the 95% confidence intervals for $\hat{\alpha}(v)$ only excludes the zero point for v (waiting time) in region $[0, 4]$, which corresponds to age at the initial entry of $[41, 45]$. These results suggest that the effect of BMI could increase at a higher rate with the age at the initial entry into the FHS study for those participants who were older than 40 years old at the initial entry.

6. Discussion

Marginal hazard models have been shown to be useful for analyzing multivariate survival data. However, no formal work in the literature is available for Cox's type of models with linear and nonlinear interactions in the multivariate hazard regression. This paper fills in the gap in this area. Without specifying the dependent structure among failure types within each subject, we propose a profile pseudo-partial likelihood estimation approach to fitting the partially linear hazard regression model. Our theory demonstrates that the finite parameters can be estimated at the rate of root-n, while the coefficient functions can be estimated with optimal rates independent of the parametric part. We also derive consistent estimators for the covariance matrix of the estimators of the finite parameters, which facilitate the inference for the parameters.

Variable selections based on the non-concave penalty likelihood can also be developed for the partly linear hazard model along the framework of Fan and Li (2004). An ongoing research will focus on testing whether the coefficient functions are constants or of certain parametric forms. This together with our current work will provide a practical inference tool for exploring possible interaction among risk factors in the analysis of multivariate survival data by employing the marginal hazard model.

Acknowledgements. The work was partially supported by NIH grant R01 HL69720. Fan's research was also partially supported by NSF grant DMS-0354223. Jiang's work was also partially supported by NSFC grant 10471006. Zhou's research was partially supported by NIH grant R01 CA79949. The authors acknowledge gratefully the constructive suggestions and comments from the editor, AE and two anonymous referees that improve significantly the presentation of this article.

Appendix I: Notations and Conditions

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a family of complete probability space with a history $\mathcal{F}$ for an increasing right-continuous filtration $\mathcal{F}_t \subset \mathcal{F}$. Put $\bar{N}_{\cdot j}(u) = n^{-1}\sum_{i=1}^n N_{ij}(u)$ and $n_j(u) = P(X_{1j} \le u, \Delta_{1j} = 1)$. Let $\mathcal{F}_{t,ij} = \sigma\{X_{ij} < u, W_{ij}(u), V_{ij}(u), Z_{ij}(u), Y_{ij}(u),\ 0 \le u \le t]\}$ be the information received up to time t for each (i,j), and $M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(u)\lambda_{ij}(u)\,du$, for $i = 1, \cdots, n;\ j = 1, \cdots, J$. Assume that $N_{ij}(u)$ is $\mathcal{F}$-adapted, and the observation period is $[0, \tau]$, where τ is the study ending time. Then $M_{ij}(t)$ is a martingale with respect to the marginal failure filtration $\mathcal{F}_{t,ij}$ and the σ-field generated by $\cup_{i=1}^n \mathcal{F}_{t,ij}$ respectively, under the independent censoring scheme.

The following notations are needed for our theoretical results. For any vector a, define $a^{\otimes k} = 1, a, \text{and}\, aa^T$, respectively, for $k = 0, 1, 2$. Let β_0 be the true value of the parameter β. Denote by $s_{1j}(\beta, u)$ the risk function for the jth failure type in the ith subject, i.e. $s_{1j}(\beta, u) = Y_{1j}(u)\exp[\beta^T W_{1j}(u) + \alpha(V_{1j})^T Z_{1j}(u)]$. Let

$$\rho_j(u, v, w, z) = E\{s_{1j}(\beta_0, u)|V_{1j} = v, W_{1j} = w, Z_{1j} = z\},$$

be the conditional expectation of the risk function. For $k = 0, 1, 2$, set

$$a_{jk}(u, v) = E[s_{1j}(\beta_0, u)(Z_{1j}(u))^{\otimes k}|V_{1j} = v],$$

$$b_{jk}(u, v) = E[s_{1j}(\beta_0, u)(Z_{1j}(u))^{\otimes k} \otimes W_{1j}(u)|V_{1j} = v],$$

where $\otimes$ denotes the Kronecker product. It is seen that a_{j0} is a positive function, a_{j1} is a $p \times 1$ vector-valued function, and a_{j2} is a $p \times p$ matrix-valued function; b_{j0} is a $q \times 1$ vector-valued function, b_{j1} is a $p \times q$ matrix-valued function, and b_{j2} is a $p \times p \times q$ array-valued function. Let $A_{j1}(u, v) = a_{j2}(u, v)/a_{j0}(u, v) - (a_{j1}(u, v)/a_{j0}(u, v))^{\otimes 2}$, $A_{j2}(u, v) = a_{j2}(u, v)/a_{j0}(u, v)$, and $\tilde{A}_k(v_0) = \sum_{j=1}^J f_j(v_0) \int_0^\tau A_{jk}(u, v_0)a_{j0}(u, v_0)\lambda_{0j}(u)\,du$ (for $k = 1, 2$). We construct matrices

$$B_j(u, v_0) = \frac{b_{j1}(u, v_0)}{a_{j0}(u, v_0)} - \frac{a_{j1}(u, v_0)}{a_{j0}(u, v_0)} \otimes \frac{b_{j0}(u, v_0)}{a_{j0}(u, v_0)},$$

and $\tilde{B}(v_0) = \sum_{j=1}^J f_j(v_0) \int_0^\tau B_j(u, v_0)a_{j0}(u, v_0)\lambda_{0j}(u)\,du$. Let

$$\chi(v_0) = -\tilde{A}_1^{-1}(v_0)\tilde{B}(v_0) \text{ and } \theta_{ij}(u) = W_{ij}(u) + \chi(V_{ij})^T Z_{ij}(u).$$

For $k = 0, 1, 2$, let $r_{jk}(\beta, u) = E\{s_{1j}(\beta, u)(\theta_{1j}(u))^{\otimes k}\}$ and $r_{jk}(u) = r_{jk}(\beta_0, u)$. For $k = 0, 1$, put $\tilde{r}_{jk}(u, v) = E\{s_{1j}(\beta_0, u)\theta_{1j}(u)^{\otimes k} Z_{1j}(u)^T \big| V_{1j} = v\}$. Let

$$\eta_j(u, v) = f_j(v)[\tilde{r}_{j1}(u, v) - [r_{j1}(u)/r_{j0}(u)]\tilde{r}_{j0}(u, v)]\tilde{A}_1^{-1}(v),$$

$\xi(v) = \sum_{j=1}^{J} \int_0^\tau r_{j0}^{-1}(u)\eta_j(u,v)\, dn_j(u)$, and

$$\Sigma = E\Big\{\sum_{j=1}^{J} \int_0^\tau \Big[\theta_{1j}(u) - \frac{r_{j1}(u)}{r_{j0}(u)} - \xi(V_{1j})\Big(Z_{1j}(u) - \frac{a_{j1}(u,V_{1j})}{a_{j0}(u,V_{1j})}\Big)\Big]\, dM_{1j}(u)\Big\}^{\otimes 2}.$$

The following conditions are needed for the proofs of our theoretical results.

(i) The kernel function $K(\cdot)$ is a bounded and symmetric density with a compact support $[-1,1]$, say.

(ii) The density $f_j(\cdot)$ of V_j is of compact support and has a bounded second derivative for $j = 1,\cdots,J$.

(iii) $nh \to \infty$ and $h \to 0$, as $n \to \infty$. Let $\mu_2 = \int v^2 K(v)\, dv$, $\nu_0 = \int K^2(v)\, dv$, $\nu_2 = \int v^2 K^2(v)\, dv$, and H be a $p \times p$ order diagonal matrix whose first p elements are ones and others h's.

(iv) The function $\alpha(\cdot)$ has a continuous second derivative.

(v) The conditional probability

$$P_j(u,v,w,z) = E\{Y_{1j}(u)|V_{1j} = v, W_{1j}(u) = w, Z_{1j}(u) = z\}$$

is equi-continuous in the argument (u,v) on $[0,\tau] \times \cup_{j=1}^{J}\text{supp}[f_j(\cdot)]$. The conditional expectations $a_{jk}(u,v)$ and $b_{jk}(u,v)$ are equi-continuous in $v \in \text{supp}[f_j(\cdot)]$, for $j = 1,\cdots,J$ and $k = 0,1,2$.

(vi) $\int_0^\tau \lambda_{0j}(t)\, dt < \infty$ for each $j \in \{1,2,\cdots,J\}$.

(vii) There exists a neighborhood $\mathcal{B}$ of β_0 such that for $k = 0,1,2,3$

$$E\Big\{\sup_{(\beta,t)\in\mathcal{B}\times[0,\tau]} Y_{ij}(t)||Z_{ij}(t)||^k \exp[\beta^T W_{ij}(t) + \alpha(V_{1j}(t))^T Z_{ij}(t)]\Big\} < \infty.$$

(viii) The functions $r_{j0}(\cdot,u)$, $r_{j1}(\cdot,u)$ and $r_{j2}(\cdot,u)$ are continuous in $\beta \in \mathcal{B}$, uniformly in $u \in [0,\tau]$; $r_{j0}(\cdot,\cdot)$ is bounded away from zero on $\mathcal{B} \times [0,\tau]$; r_{j1} and r_{j2} are bounded on $\mathcal{B} \times [0, \tau]$; $\tilde{r}_{jk}(u, v)$ is equi-continuous in $v \in \mathrm{supp}[f_j(\cdot)]$. The matrix I is finite positive definite, where

$$I = \sum_{j=1}^{J} \int_0^{\tau} \Big[\frac{r_{j2}(u)}{r_{j0}(u)} - \Big(\frac{r_{j1}(u)}{r_{j0}(u)}\Big)^{\otimes 2}\Big] r_{j0}(u)\lambda_{0j}(u)\, du.$$

The above conditions are similar to those in Andersen and Gill (1982) and Fan *et al.* (1997). Conditions (i)-(iv) are standard for nonparametric component estimation using local partial likelihood (see Fan *et al.* (1997)); Conditions (v)-(viii) guarantee the local asymptotic quadratic properties for the partial likelihood function, and hence the asymptotic normality of the estimators. See Andersen and Gill (1982) and Murphy and van der Vaart (2000) for details.

To derive consistent estimator of the covariance matrix Ω of $\hat{\beta}$, we need to consistently estimate I and Σ. Let the empirical estimator of $r_{jk}(t)$ be

$$\hat{r}_{jk}(t) = n^{-1} \sum_{i=1}^{n} Y_{ij}(t)(\hat{\theta}_{ij}(t))^{\otimes k} \exp\big[\hat{\beta}^T W_{ij}(t) + \hat{\alpha}(V_{ij})^T Z_{ij}(t)\big].$$

Then the empirical estimator of I,

$$\hat{I} = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{J} \Delta_{ij} \{\hat{r}_{j2}(X_{ij})/\hat{r}_{j0}(X_{ij}) - [\hat{r}_{j1}(X_{ij})/\hat{r}_{j0}(X_{ij})]^{\otimes 2}\},$$

and the empirical estimator of Σ, $\hat{\Sigma} = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{k=1}^{J} \hat{G}_{ij} \hat{G}_{ik}^T$, are consistent, where

$$\begin{aligned} \hat{G}_{ij} \quad = \quad & \Delta_{ij}\Big\{\hat{\theta}_{ij}(X_{ij}) - \frac{\hat{r}_{j1}(X_{ij})}{\hat{r}_{j0}(X_{ij})} - \hat{\xi}(V_{ij})\big[Z_{ij}(X_{ij}) - \frac{\hat{a}_{j1}(X_{ij}, V_{ij})}{\hat{a}_{j0}(X_{ij}, V_{ij})}\big]\Big\} \\ & - n^{-1} \sum_{m=1}^{n} \Delta_{mj} Y_{ij}(X_{mj}) \exp\{\hat{\beta}^T W_{ij}(X_{mj}) + \hat{\alpha}(V_{ij})^T Z_{ij}(X_{mj})\} \hat{r}_{j0}^{-1}(X_{mj}) \\ & \times \Big\{\hat{\theta}_{ij}(X_{mj}) - \frac{\hat{r}_{j1}(X_{mj})}{\hat{r}_{j0}(X_{mj})} - \hat{\xi}(V_{ij})\big[Z_{ij}(X_{mj}) - \frac{\hat{a}_{j1}(X_{mj}, V_{ij})}{\hat{a}_{j0}(X_{mj}, V_{ij})}\big]\Big\}, \end{aligned}$$

is the natural substitution estimator of G.

Appendix II: Proofs of Theorems

The proofs will involve the martingale theory, the theory of empirical processes and the techniques commonly used in nonparametric literature. For ease of exposition, we consider only the model with time-independent covariate V. The time-dependent covariate model can similarly be developed.

Let $\gamma^* = H\gamma$ and $U^*_{ij}(u, v_0) = H^{-1}U_{ij}(u, v_0)$. Then the local partial likelihood in (2.3) can be rewritten as

$$\begin{aligned}\ell(\beta, \gamma^*) &= \sum_{j=1}^{J}\sum_{i=1}^{n} K_h(V_{ij} - v_0)\Delta_{ij}\Big\{\beta^\tau W_{ij} + \gamma^{*T} U^*_{ij} \\ &\quad - \log\Big(\sum_{\ell\in\mathcal{R}_j(X_{ij})} \exp[\beta^\tau W_{\ell j} + \gamma^{*T} U^*_{\ell j}] K_h(V_{\ell j} - v_0)\Big)\Big\}. \qquad (A.1)\end{aligned}$$

Note that the global profile pseudo-partial likelihood in (2.4) is

$$\begin{aligned}\ell_p(\beta) &= \sum_{j=1}^{J}\sum_{i=1}^{n}\int_0^\tau \Big\{\beta^\tau W_{ij}(u) + \hat\alpha(V_{ij}, \beta)^\tau Z_{ij}(u) \\ &\quad - \log\Big(\sum_{\ell=1}^{n} Y_{\ell j}(u)\exp[\beta^\tau W_{\ell j}(u) + \hat\alpha(V_{\ell j}, \beta)^\tau Z_{\ell j}(u)]\Big)\Big\}\, dN_{ij}(u). \qquad (A.2)\end{aligned}$$

Denote by β_j and β_{0j} the j-th elements of β and β_0 respectively. It follows from Taylor's expansion that for any β in a neighborhood of β_0

$$\begin{aligned}\ell_p(\beta) &= \ell_p(\beta_0) + (\beta - \beta_0)^\tau\, D_\beta\{\ell_p(\beta_0)\} \\ &\quad + \frac{1}{2}(\beta - \beta_0)^\tau\, D^2_\beta\{\ell_p(\beta_0)\}(\beta - \beta_0) + R_n(\beta^*), \qquad (A.3)\end{aligned}$$

where β^* is between β and β_0, and

$$R_n(\beta^*) = \frac{1}{6}\sum_{j,k,\ell}(\beta_j - \beta_{0j})(\beta_k - \beta_{0k})(\beta_\ell - \beta_{0\ell})\Big[\frac{\partial^3 \ell_p(\beta)}{\partial\beta_j\partial\beta_k\partial\beta_\ell}\Big|_{\beta=\beta^*}\Big]. \qquad (A.4)$$

It can be shown that $n^{-1}\frac{\partial^3 \ell_p(\beta)}{\partial\beta_j\partial\beta_k\partial\beta_\ell}$ is bounded in probability, and hence $n^{-1}R_n(\beta) = O_p(||\beta - \beta_0||^3)$ for $\beta \in \mathcal{B}$.

Let $\hat\gamma^* = H\hat\gamma$. Given β, $\hat\gamma^*(\beta)$ satisfies the equation $D_{\gamma^*}\{\ell(\beta, \gamma^*)\} = 0$, that is

$$\sum_{j=1}^{J}\sum_{i=1}^{n} K_h(V_{ij} - v_0)\Delta_{ij}\Big\{U^*_{ij} - \frac{\sum_{\ell\in\mathcal{R}_j(X_{ij})}\exp[\beta^\tau W_{\ell j} + \hat\gamma^{*T} U^*_{\ell j}]U^*_{\ell j} K_h(V_{\ell j} - v_0)}{\sum_{\ell\in\mathcal{R}_j(X_{ij})}\exp[\beta^\tau W_{\ell j} + \hat\gamma^{*T} U^*_{\ell j}] K_h(V_{\ell j} - v_0)}\Big\} = 0,$$

which is equivalent to

$$\sum_{j=1}^{J}\sum_{i=1}^{n}\int_0^\tau K_h(V_{ij} - v_0)\Big\{U^*_{ij} - \frac{\Phi_{nj1}(u, v_0, \beta, \hat\gamma^*)}{\Phi_{nj0}(u, v_0, \beta, \hat\gamma^*)}\Big\} dN_{ij}(u) = 0,$$

where and thereafter for $k = 0, 1, 2$

$$\Phi_{njk}(u, v_0, \beta, \gamma^*) = n^{-1} \sum_{\ell=1}^{n} Y_{\ell j}(u) \exp[\beta^\tau W_{\ell j}(u) + \gamma^{*\tau} U_{\ell j}^*] U_{\ell j}^{*\otimes k} K_h(V_{\ell j} - v_0). \quad (A.5)$$

Next, we introduce several technical lemmas. The profile pseudo-partial likelihood $\ell_p(\beta)$ in (2.4) involves in $\hat{\alpha}(V_{ij}, \beta)$ which uses all observational information and leads to unpredictability in the analysis of score function and hence results in the infeasibility of direct use of the common martingale method. To overcome this difficulty, we introduce a leave-one-out argument in the proofs of these lemmas. For details, one can refer to the technical report of Cai, Fan, Jiang and Zhou (2005) at *http://www.princeton.edu/~jqfan/papers/06/PLVC/* ????.

Lemma 1. Let $\nu_j(u, v_0) = (a_{j1}(u, v_0)^\tau / a_{j0}(u, v_0), \mathbf{0}_p^\tau)^\tau$ and $c_n = O_p(h^2 + \frac{1}{\sqrt{nh}})$, where $\mathbf{0}_p$ denotes a $p \times 1$ vector whose elements are all zeros. Under Conditions (i)-(v),

(i) $\Phi_{nj0}(u, v_0, \beta_0, \gamma^*) = f_j(v_0) a_{j0}(u, v_0) + c_n$, $\Phi_{nj1}(u, v_0, \beta_0, \gamma^*) = f_j(v_0)(a_{j1}(u, v_0)^\tau, \mathbf{0}_p^\tau)^\tau + c_n$, and $\Phi_{nj2}(u, v_0, \beta_0, \gamma^*) = \text{blockdiag}\{a_{j2}(u, v_0), \mu_2 a_{j2}(u, v_0)\} f_j(v_0) + c_n$.

(ii) If $||\hat{\beta} - \beta_0|| = O_p(1/\sqrt{n})$, then

$$\sup_{u \in [0,\tau]} ||\Phi_{njk}(u, v_0, \hat{\beta}, \hat{\gamma}^*) - \Phi_{njk}(u, v_0, \beta_0, \hat{\gamma}^*)|| = O_p(1/\sqrt{n}).$$

(iii) $\sup_{u \in [0,\tau]} \left|\left| \frac{\Phi_{nj1}(u, v_0, \beta_0, \gamma^*)}{\Phi_{nj0}(u, v_0, \beta_0, \gamma^*)} - \nu_j(u, v_0) \right|\right| = o_p(1)$.

(iv) $\sup_{u \in [0,\tau]} \left|\left| \Phi_{nj}(u, v_0, \beta_0, \gamma^*) - A_j(u, v_0) \right|\right| = o_p(1)$,
where $A_j(u, v_0) = \text{blockdiag}\{A_{j1}(u, v_0), \mu_2 A_{j2}(u, v_0)\}$, and

$$\Phi_{nj}(u, v_0, \beta_0, \gamma^*) = \frac{\Phi_{nj0}(u, v_0, \beta_0 \gamma^*) \Phi_{nj2}(u, v_0, \beta_0, \gamma^*) - \Phi_{nj1}^{\otimes 2}(u, v_0, \beta_0, \gamma^*)}{\Phi_{nj0}^2(u, v_0, \beta_0, \gamma^*)}.$$

Lemma 2. Assume Conditions (i)-(viii) hold. If $nh^2 \to \infty$ and $nh^4 \to 0$, then

$$\begin{aligned} \frac{1}{\sqrt{n}} D_\beta\{\ell_p(\beta_0)\} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{J} \int_0^\tau \Big\{ \theta_{ij}(u) - \frac{r_{j1}(\beta_0, u)}{r_{j0}(\beta_0, u)} \\ &\quad - \xi(V_{ij}) \Big[Z_{ij}(u) - \frac{a_{j1}(u, V_{ij})}{a_{j0}(u, V_{ij})} \Big] \Big\} \, dM_{ij}(u) + o_p(1). \end{aligned}$$

Lemma 3. Suppose Conditions (i)-(x) hold. Then $n^{-1}D^2_\beta\{\ell_p(\beta_0)\} \xrightarrow{p} -I$.

Proof of Theorem 1. The proof is given under the framework of Fan *et al.* (1997). By Lemma 2, $n^{-1}D_\beta\{\ell_p(\beta_0)\} = O_p(1/\sqrt{n})$. Thus, for any small given $\varepsilon > 0$, if $\beta \in S_\varepsilon \equiv \{\beta : ||\beta - \beta_0|| \le \varepsilon\}$,

$$\left|(\beta - \beta_0)^\tau n^{-1}D_\beta\{\ell_p(\beta_0)\}\right| \le \varepsilon^3, \tag{A.6}$$

with probability tending to one. Let a be the minimum eigenvalue of positive definitive matrix $I(\beta_0)$. By Lemma 3, we conclude that for all $\beta \in S_\varepsilon$

$$(\beta - \beta_0)^\tau n^{-1}D^2_\beta\{\ell_p(\beta_0)\}(\beta - \beta_0) < -a\varepsilon^2, \tag{A.7}$$

with probability tending to one. By (A.4), with probability tending to one that there is a constant $C > 0$ such that

$$|n^{-1}R_n(\beta)| \le C\varepsilon^3. \tag{A.8}$$

Then by (A.3), with probability tending to one that when ε is small enough,

$$n^{-1}[\ell_p(\beta) - \ell_p(\beta_0)] \le 0, \qquad \text{for all } \beta \in S_\varepsilon. \tag{A.9}$$

Therefore, $\ell_p(\beta)$ has a local maximum in the interior of S_ε, and with probability tending to one, there exists a consistent estimator sequence $\hat\beta$ for β_0 which maximizes the global profile pseudo-partial likelihood $\ell_p(\beta)$.

Proof of Theorem 2. By Lemma 3, $n^{-1}D^2_\beta\{\ell_p(\beta_0)\} \xrightarrow{p} -I$. Note that $\hat\beta$ is consistent. Plugging the above expression into (A.3), we establish that

$$\begin{aligned}\ell_p(\hat\beta) &= \ell_p(\beta_0) + (\hat\beta - \beta_0)^\tau D_\beta\{\ell_p(\beta_0)\} \\ &\quad -\frac{n}{2}(\hat\beta - \beta_0)^\tau I(\hat\beta - \beta_0) + o_p\{(1 + \sqrt{n}||\hat\beta - \beta_0||)^2\}.\end{aligned} \tag{A.10}$$

Using Corollary 1 in Murphy and van der Vaart (2000) and Lemma 2, we obtain that

$$\begin{aligned}\sqrt{n}(\hat\beta - \beta_0) &= I^{-1}\frac{1}{\sqrt{n}}D_\beta\{\ell_p(\beta_0)\} + o_p(1) \\ &= I^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{j=1}^J\int_0^\tau \Big\{\theta_{ij}(u) - \frac{r_{j1}(\beta_0, u)}{r_{j0}(\beta_0, u)} \\ &\quad -\xi(V_{ij})\Big[Z_{ij}(u) - \frac{a_{j1}(u, V_{ij})}{a_{j0}(u, V_{ij})}\Big]\Big\}\, dM_{ij}(u) + o_p(1 + \sqrt{n}||\hat\beta - \beta_0||).\end{aligned}$$

Then by the martingale central limit theorem and the Slutsky Theorem,

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I^{-1}\Sigma I^{-1}).$$

Proof of Theorem 4. It is easy to show that $\hat{\gamma}^*$ is consistent and satisfies

$$n^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n}\int_0^\tau K_h(V_{ij} - v_0)\Big\{U_{ij}^* - \frac{\Phi_{nj1}(u, v_0, \beta_0, \hat{\gamma}^*)}{\Phi_{nj0}(u, v_0, \beta_0, \hat{\gamma}^*)}\Big\}\, dN_{ij}(u) = O_p(n^{-1/2}).$$

Denote by $\hat{U}(\beta_0, \hat{\gamma}^*, v_0)$ the left-hand side of the above equation. Then by Taylor's expansion, we have

$$\hat{U}(\beta_0, \gamma^*, v_0) + D_{\gamma^*}\{\hat{U}(\beta_0, \tilde{\gamma}^*, v_0)\}(\hat{\gamma}^* - \gamma^*) = o_p(1/\sqrt{nh}), \tag{A.11}$$

where $\tilde{\gamma}^*$ is between $\hat{\gamma}^*$ and γ^*, and hence $\tilde{\gamma}^* \to \gamma^*$ in probability. Using Lemma 1(iv), we obtain that

$$\begin{aligned} -D_{\gamma^*}\{\hat{U}(\beta_0, \gamma^*, v_0)\} &= n^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n}\int_0^\tau K_h(V_{ij} - v_0)\Phi_{nj}(u, v_0, \beta_0, \gamma^*)\, dN_{ij}(u) \\ &= \tilde{A}(v_0) + o_p(1), \end{aligned} \tag{A.12}$$

uniformly for $v_0 \in \cup_{j=1}^J \text{supp}[f_j(\cdot)]$, where $\tilde{A}(v_0) = \text{blockdiag}\{\tilde{A}_1(v_0), \mu_2\tilde{A}_2(v_0)\}$. We can decompose $\hat{U}(\beta_0, \gamma^*, v_0)$ as

$$\hat{U}(\beta_0, \gamma^*, v_0) \equiv \quad d_n(\tau) + q_n(\tau), \tag{A.13}$$

where

$$\begin{aligned} q_n(\tau) &= n^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n}\int_0^\tau K_h(V_{ij} - v_0)\Big\{U_{ij}^* - \frac{\Phi_{nj1}(u, v_0, \beta_0, \gamma^*)}{\Phi_{nj0}(u, v_0, \beta_0, \gamma^*)}\Big\}\, Y_{ij}(u)\lambda_{ij}(u)du \\ &= -\frac{1}{2}h^2\mu_2\tilde{A}(v_0)(\alpha''(v_0)^\tau, \mathbf{0}_p^\tau)^\tau + o_p(h^2), \end{aligned} \tag{A.14}$$

and $d_n(\tau) = n^{-1}\sum_{j=1}^J \sum_{i=1}^n \int_0^\tau K_h(V_{ij} - v_0)\Big\{U_{ij}^* - \frac{\Phi_{nj1}(u,v_0,\beta_0,\gamma^*)}{\Phi_{nj0}(u,v_0,\beta_0,\gamma^*)}\Big\}\, dM_{ij}(u)$. Using Lemma 1, we can show that

$$d_n(\tau) = n^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n}\int_0^\tau K_h(V_{ij} - v_0)[U_{ij}^* - \nu_j(u, v_0)]\, dM_{ij}(u) + o_p(1/\sqrt{nh}),$$

which combined with (A.11)-(A.14) and the martingale central limit theorem (see Theorem 5.35 of Fleming and Harrington (1991)) leads to the result of the theorem.

REFERENCES

Andersen, P. K. and Gill, R. D. (1982) Cox's regression model for counting processes: A large sample study. *Annals of Statistics*, **10**, 1100–1120.

Bickel, P. J. (1975) One-step Huber estimates in linear models. *Journal of American Statistical Association*, **70**, 428-433.

Brumback, B. and Rice, J. A. (1998) Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). *Journal of American Statistical Association*, **93**, 961-994.

Cai, J. (1999) Hypothesis testing of hazard ratio parameters in marginal models for multivariate failure time data. *Lifetime Data Analysis*, **5**, 39-53.

Cai, J., Fan, J., Jiang, J. and Zhou, H. (2005) Partially linear hazard regression with varying-coefficients for multivariate survival data. Technical Report, http://www.math.uncc.edu/˜jjiang1/...

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

Cai, J., Fan, J, Zhou, H. and Zhou, Y. (2007). Marginal hazard models with varying-coefficients for multivariate failure time data. *The Annals of Statistics*, **35**, 324-354.

Cai, Z., Fan, J. and Yao, Q. (2000) Functional-coefficient regression models for nonlinear time series. *Journal of American Statistical Association*, **95**, 941-956.

Cai, J. and Prentice, R.L (1995) Estimating equations for hazard ratio parameters based on correlated failure time data. *Biometrika*, **82**, 151-164.

Cai, J. and Prentice, R.L. (1997) Regression analysis for correlated failure time data. *Lifetime Data Analysis*, **3**, 197-213.

Cai, J. and Shen, Y. (2000) Permutation tests for comparing marginal survival functions with clustered failure time data. *Statistics in Medicine*, **19**, 2963-2973.

Carroll, R.J., Fan, J., Gijbels, I, and Wand, M.P. (1997) Generalized partially linear single-index models. *Journal of American Statistical Association*, **92**, 477-489

Carroll, R.J., Ruppert, D. and Welsh, A.H. (1998) Nonparametric estimation via local estimating equations. *Journal of American Statistical Association*, **93**, 214-227.

Clayton, D. and Cuzick, J. (1985) Multivariate generalizations of the proportional hazards model (with discussion). *Journal Royal Statistical Society A*, **148**, 82-117.

Clegg, L. X. Cai, J. and Sen, P. K. (1999) A marginal mixed baseline hazards model for multivariate failure time data. *Biometrics*, **55**, 805-812.

Cox, D. R. (1972). Regression models and life-tables. *Journal Royal Statistical Society B*, **34**, 187-220.

Dawber, T. R. (1980) *The Framingham Study, The Epidemiology of Atherosclerotic Disease.* Cambridge, MA: Harvard University Press.

Fan, J. and Chen, J. (1999) One-step local quasi-likelihood estimation. *Journal Royal Statistical Society B*, **61**, 927-943.

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

Fan, J., Lin, H. and Zhou, Y. (2006) Local partial likelihood estimation for life time data. *Annals of Statistics*, **34**, 290-325.

Fan, J. and Jiang, J. (2000) Variable bandwidth and one-step local M-estimator. *Science in China, (Series A)*, **43**, 65-81.

Fan, J. and Li, R. (2004) New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of American Statistical Association*, **99**, 710-723.

Fan, J. and Zhang, W. (1999) Statistical estimation in varying coefficient models. *Annals of Statistics*, **27**, 1491-1518.

Hastie, T. J. and Tibshirani, R. J. (1993) Varying-coefficient models. *J. Roy. Statist. Soc. B*, **55**, 757-796.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L.-P. (1998) Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrics*, **85**, 809-822.

Jiang, J. and Mack. Y. P. (2001) Robust local polynomial regression for dependent data. *Statistica Sinica*, **11**, 705-722.

Kalbfleisch, J. D. and Prentice, R. L. (2002) *The Statistical Analysis of Failure Time Data, 2nd edition.* New York: Wiley.

Liang, K. Y., Self, S. G. and Chang, Y (1993) Modeling marginal hazards in multivariate failure time data. *Journal Royal Statistical Society B*, **55**, 2, 441-453.

Lin, D. Y. (1994) Cox regression analysis of multivariate failure time data: the marginal approach. *Statistics in Medicine*, **13**: 2233-2247.

Lee, E. W., Wei, L. J., and Amato, D. A. (1992) Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In *Survival Analysis: State of the Art. (ed. J. P. Klein and P. K. Goel)*, pp. 237-247. Kluwer Academic Publishers.

Masry, E. and Fan, J. (1997) Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics*, **24**, 165-179.

Murphy, S. A. and van der Vaart, A. W. (2000) On profile likelihood (with discussion). *Journal of American Statistical Association*, **95**, 449-485.

Prentice, R. L. and Hsu, L. (1997) Regression on hazard ratios and cross ratios in multivariate failure time analysis. *Biomtrka*, **84**, 349-363.

Robinson, P.M. (1988) The stochastic difference between econometric and statistics. *Econometrica*, **56**, 531-547.

Spiekerman, C. F. and Lin, D. Y. (1998) Marginal regression models for multivariate failure time data. *Journal of American Statistical Association*, **93**, 1164-1175.

Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989) Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of American Statistical Association*, **84**, 1065-1073.