Partially Linear Hazard Regression for Multivariate Survival Data [*]

Jianwen Cai, Jianqing Fan, Jiancheng Jiang and Haibo Zhou

Abstract

This paper studies estimation of partially linear hazard regression models for multivariate survival data. A profile pseudo-partial likelihood estimation method is proposed under the marginal hazard model framework. The estimation on the parameters for linear part is accomplished via maximization of a pseudo-partial likelihood profiled over the nonparametric part. This enables one to obtain $\sqrt{n}$-consistent estimators of the parametric component. Asymptotic normality is obtained for the estimates of both the linear and nonlinear parts. The new technical challenge is that the nonparametric component is indirectly estimated via

[*] Jianwen Cai is Professor, Department of Biostatistics, University of North Carolina at Chapel Hill, NC 27599 (email: cai@bios.unc.edu); Jianqing Fan is Professor, Department of Operation Research and Financial Engineering, Princeton University, Princeton, NJ 08544 (email: jqfan@princeton.edu); Jiancheng Jiang is Assistant Professor, Department of Mathematics and Statistics, University of North Carolina at Charlotte, NC 28223 (email: jjiang1@uncc.edu), and Associate Research Staff, Department of Operation Research and Financial Engineering, Princeton University, Princeton, NJ 08544 (email: jjiang@Princeton.edu); Haibo Zhou is Associate Professor, Department of Biostatistics, University of North Carolina at Chapel Hill, NC 27599 (email: zhou@bios.unc.edu). This research was supported in part by NIH grant R01 HL69720. Additional partial support was also provided by NSF grant DMS-0354223 and NIH grant R01 GM072611 for Fan, NSFC grant 10471006 for Jiang, and NIH grant R01 CA 79949 for Zhou. The authors acknowledge gratefully on the constructive suggestions and comments from the associate editor and two anonymous referees that greatly improved this article.

its integrated derivative function from a local polynomial fit. An algorithm of fast implementation of our proposed method is presented. Consistent standard error estimates using sandwich-type of ideas are also developed, which facilitates inferences for the model. It is shown that the nonparametric component 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 method. A real dataset is analyzed to illustrate the proposed methodology.

Abbreviated Title. Partially Linear Hazard Regression.

KEY WORDS: Local pseudo-partial likelihood; Marginal hazard model; Multivariate failure time; Partially linear; Profile pseudo-partial likelihood.

1 Introduction

Multivariate survival data arise from many contexts. Some examples are epidemiological cohort studies in which the ages of disease occurrence are recorded for members of families, animal experiments where treatments are applied to samples of littermates, clinical trials in which individual study subject are followed for the occurrence of multiple events, and intervention trials involving group randomization. A common feature of the data in these examples is that the failure times are correlated. For example, in animal experiments, the failure times of animals within a litter may be correlated because they share common genetic traits and environmental factors. Similarly, in clinical trials where the patients are followed for repeated recurrent events, the times between recurrences for a given patient may be correlated.

In general, there are three types of models in the multivariate failure time literature: overall intensity process models, frailty models, and marginal hazard models. The overall hazard models deal with the overall intensity, which is defined as the hazard rate given the history of the entire cluster (Andersen and Gill 1982). Interpretation of the parameters in an overall hazard model is conditioned on the failure and censoring information of every individual in the cluster. The frailty model considers the conditional hazard given the unobservable frailty random variables, which is particularly useful when the association of failure types within a subject is of interest (see Hougaard 2000). However, such models tend to be restrictive with respect to the types of dependence that can be modeled and model fitting is usually cumbersome. When the correlation among the observations is not of interest, the marginal hazard model approach which models the "population-averaged" covariate effects has been widely used (e.g. 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, and Clegg, Cai, and Sen 1999 among others).

Most statistical methods developed for failure time data assume that the covariate effects on the logarithm of the hazard function are linear and the regression coefficients are constants (see for example Fleming and Harrington 1991 and Andersen *et al.* 1993). These assumptions, however, are mainly chosen for their mathematical convenience. True covariate effects can be more complex than the log-linear effect and new analytic challenges arise in assessing nonlinear effects. As an example, in studying the effect of cholesterol on the time to coronary heart disease (CHD) and cerebrovascular accident (CVA) among $2,336$ men and $2,873$ women in the well-known Framingham Heart Study (Dawber 1980), the investigators are interested in identifying a non-linear cholesterol effect. A nonparametric method is desired for providing a continuous trend of the cholesterol effect that is flexible enough to indicate local changes in this trend. Nonparametric modeling of such trend has less restriction than the parametric approach and therefore it is less likely to distort the underlying relationship between the failure time and the covariate.

In developing nonparametric methods for analyzing multivariate censored survival data, high dimensional covariates may cause the so-called "curse of dimensionality" problem. One of the methods for attenuating this difficulty is to model the covariate effects via a partially linear structure, a combination of linear and nonparametric parts in the marginal hazard model. It allows one to explore nonlinearity of certain covariates when the covariate effects are unknown and avoids the "curse-of-dimensionality" problem inherent in the saturated multivariate nonparametric regression model. It also allows the statistical model to retain the nice interpretability of the traditional linear structure. The partial linear structure has been systematically studied in the multivariate regression setting by many authors (e.g., Wahba 1984, Speckman 1988, Cuzick 1992, Carroll *et al.* 1997, Lin and Carroll 2001, Liang, Härdle and Carroll 1999). An overview on the partially linear model can be found in Härdle, Liang, and Gao (2004). Some authors have considered modeling nonlinear covariate effects for the univariate failure time data under the Cox proportional hazards model (e.g., Hastie and Tibshirani 1993, Gentleman and Crowley 1991, Fan, Gijbels, and King 1997). The partially linear covariate effects for the univariate failure time data in the framework of Cox-type of models have also been studied in Huang (1999) by using polynomial splines, where the estimator of parametric part achieves root-n consistency and the semiparametric information bound but lacks a consistent estimator for its asymptotic covariance matrix.

For the multivariate failure time data analyzed in this paper, no formal methodology has been elaborated in the literature to address nonlinear covariate effects. In this paper, we develop a nonparametric approach for the nonlinear covariate effects under the Cox-type marginal hazards model. We consider a semiparametric structure by allowing parametric as well as nonparametric components to be included in the hazards regression function.

We consider two general setups where multivariate failure time data commonly arise. In one setup, we assume that there is a random sample of n subjects from an underlying population and that we are interested in J different types of failures. In this setup, J is a pre-specified number based on the goal of the study and it does not vary across subjects. In the other setup, we assume there are n clusters and in each cluster there are J different types of members, for example, father and sons in a family. Since not all members are necessarily available in a cluster, the cluster size can vary. In order to incorporate varying cluster size, we define an indicator variable ξ_{ij} to be 1 if the jth member of the ith cluster is available and 0 otherwise. Let J be the maximum cluster size, then the size of the ith cluster is $J_i = \sum_{j=1}^{J} \xi_{ij}$. We will use (i, j) to denote the jth failure type of the ith subject or the jth member of the ith cluster. Without loss of generality, we will refer to failure types of subjects and keep in mind that the model and the results also apply to members-in-clusters-type of setup.

Let T_{ij} denote the potential failure time, C_{ij} the potential censoring time, and $X_{ij} = min(T_{ij}, C_{ij})$ the observed time for (i, j). Let Δ_{ij} be the 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 for the jth failure type as well as the covariate information for the other failure types of the ith subject up to time t. The marginal hazard function is defined as

$$\lambda_{ij}(t) = h^{-1} \lim_{h \downarrow 0} P[t < T_{ij} \leq t + h | T_{ij} > t, \mathcal{F}_{t,ij}].$$

The censoring time is assumed to be independent of the failure time conditioning on the covariates (that is the so-called "independent censoring scheme").

To model partly nonlinear covariates effects, we assume the following model

$$\lambda_{ij}(t) = Y_{ij}(t)\lambda_{0j}(t)\exp[\boldsymbol{\beta}^\tau \boldsymbol{W}_{ij}(t) + g(Z_{ij}(t))], \tag{1}$$

where $Z_{ij}(\cdot)$ is a main exposure variable of interest whose effect on the logarithm of the hazard might be non-linear; $\boldsymbol{W}_{ij}(\cdot) = (W_{ij1}(\cdot), \cdots, W_{ijq}(\cdot))^\tau$ is a vector of covariates that have linear effects; $Y_{ij}(\cdot)$ is an at risk indicator process, i.e. $Y_{ij}(t) = 1(X_{ij} \geq t)$; $\lambda_{0j}(\cdot)$ is an unspecified baseline hazard function; and $g(\cdot)$ is an unspecified smooth function.

Model (1) allows for a different set of covariates for different failure types of the subject. It also allows for a different baseline hazard function for different failure types of the subject. It is useful when the failure types in a subject have different susceptibilities to failures. A related class of marginal model is given by restricting the baseline hazard functions in (1) to be common for all the failure types within a subject, i.e.,

$$\lambda_{ij}(t) = Y_{ij}(t)\lambda_0(t)\exp[\boldsymbol{\beta}^\tau \boldsymbol{W}_{ij}(t) + g(Z_{ij}(t))]. \tag{2}$$

While this model is more restrictive, the common baseline hazard model (2) leads to more efficient estimation when the baseline hazards are indeed the same for all the failure types within a subject. Model (2) is very useful for modeling clustered failure time data where subjects within clusters are exchangeable.

In this article, we focus on statistical inference for model (1). A profile pseudo-partial likelihood estimation will be proposed to estimate $\boldsymbol{\beta}$. We show that the proposed estimator of $\boldsymbol{\beta}$ is root-n consistent. The asympotic normality is obtained for the parameters of the linear and nonlinear parts. Consistent estimators of the asymptotic variances are provided. New technical challenges arise from the fact that the function g is not directly estimable from the local pseudo-partial likelihood and has to be estimated from its derivative. Hence, the estimator of nonparametric function $g(\cdot)$ uses all observed information, and the score function of $\boldsymbol{\beta}$ can not be expressed asymptotically as an integral of a predictable process with respect to a martingale. Obtaining the asymptotic properties of the estimators is very challenging. Further, the cost due to the estimation of the nonparametric component g in the Cox model is shown to be very different from that in the least-squares regression model (Speckman 1988; Carroll *et al.* 1997). Indeed, even in the univariate case $J = 1$, the results are new.

Compared with the polynomial spline estimators for univariate failure time data in Huang (1999), the asymptotic covariance matrix of our estimators for the parametric part admits a sandwich formula which furnishes a consistent covariance matrix estimator using the plug-in method, while Huang's estimator achieves the semiparametric information bound but lacks a consistent covariance estimator; on the other hand, our estimator for the nonparametric part is not only optimal in convergence rate, but also possesses asymptotic normality which is unavailable for Huang's estimation for the nonparametric part.

This paper is organized as follows. In Section 2, we describe the procedure for estimating the coefficient $\boldsymbol{\beta}$ and the nonparametric component $g(\cdot)$ from model (1). In Section 3, we focus on the asymptotic properties of the proposed estimators along with some technical conditions. In Section 4, we conduct intensive simulations and illustrate the proposed estimation via a real data analysis. Proofs of the theorems are given in Appendix.

2 Maximum Pseudo-partial Likelihood Estimation

Let $\mathcal{R}_j(t) = \{i : X_{ij} \geq t\}$ denote the set of subjects at risk just prior to time t for failure type j. If failure times from the same subject were independent, then the logarithm of the partial likelihood for (1) is

$$\ell(\boldsymbol{\beta}, g(\cdot)) = \sum_{j=1}^{J} \sum_{i=1}^{n} \xi_{ij} \Delta_{ij} \{\boldsymbol{\beta}^{\tau} \boldsymbol{W}_{ij}(X_{ij}) + g(Z_{ij}(X_{ij})) - R_{ij}(\boldsymbol{\beta}, g)\}, \tag{3}$$

where $R_{ij}(\boldsymbol{\beta}, g) = \log\Big(\sum_{l \in \mathcal{R}_j(X_{ij})} \xi_{lj} \exp[\boldsymbol{\beta}^{\tau} \boldsymbol{W}_{lj}(X_{ij}) + g(Z_{lj}(X_{ij}))]\Big)$. Since failure times from the same subject are dependent, the above function 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 dependence structure among failure times. This furnishes robustness of our estimation method against the mis-specification of correlations among failure times. For the univariate case with $J = 1$, the partial likelihood in (3) is equivalent to the full likelihood if the least informative baseline is used (see Section 3.1 of Fan, Gijbels and King 1997).

Assume that $g(\cdot)$ is smooth so that it can be approximated locally by a polynomial of order p. For any given point z_0, by Taylor's expansion,

$$g(z) \approx g(z_0) + \sum_{k=1}^{p} \frac{g^{(k)}(z_0)}{k!}(z - z_0)^k \equiv \alpha + \boldsymbol{\gamma}^\tau \tilde{Z}, \tag{4}$$

where $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_p)^\tau$ and $\tilde{Z} = \{z - z_0, \cdots, (z - z_0)^p\}^\tau$. Using the local model (4) for the data around z_0, noting that the local intercept α cancels in (3), we obtain the logarithm of the local pseudo-partial likelihood:

$$\ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{j=1}^{J}\sum_{i=1}^{n} \xi_{ij} K_h(Z_{ij}(X_{ij}) - z_0)\Delta_{ij}[\boldsymbol{\beta}^\tau \boldsymbol{W}_{ij}(X_{ij}) + \boldsymbol{\gamma}^\tau \tilde{Z}_{ij}(X_{ij}) - R^*_{ij}(\boldsymbol{\beta}, \boldsymbol{\gamma})], \tag{5}$$

where

$$R^*_{ij}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \log\Big(\sum_{l \in \mathcal{R}_j(X_{ij})} \xi_{lj} \exp[\boldsymbol{\beta}^\tau \boldsymbol{W}_{lj}(X_{ij}) + \boldsymbol{\gamma}^\tau \tilde{Z}_{lj}(X_{ij})] K_h(Z_{lj}(X_{ij}) - z_0)\Big),$$

$\tilde{Z}_{ij}(u) = \{Z_{ij}(u) - z_0, \cdots, (Z_{ij}(u) - z_0)^p\}^\tau$, $K_h(\cdot) = K(\cdot/h)/h$, 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 confine the fact that the local model (4) is only applied to the data around z_0. It gives a larger weight to the data closer to the point z_0. For the univariate case, the local pseudo-partial likelihood was derived by Fan, Gijbels and King (1997) from a local maximum likelihood point of view.

Let $(\widehat{\boldsymbol{\beta}}(z_0), \widehat{\boldsymbol{\gamma}}(z_0))$ maximize the local pseudo-partial likelihood (5). Then, an estimator of $g'(\cdot)$ at the point z_0 is simply the first component of $\widehat{\boldsymbol{\gamma}}(z_0)$, namely $\widehat{g'}(z_0) = \widehat{\gamma_1}(z_0)$. The curve $\widehat{g}$ can be estimated by integration on the function $\widehat{g'}(z_0)$ using the method by Hastie and Tibshirani (1990). To assure the identifiability of $g(\cdot)$, we set $g(0) = 0$ without loss of generality.

In the context of the generalized linear models, Carroll *et al.* (1997) show that such a naive method produces an estimator for g that achieves the optimal rate of convergence. However, the asymptotic variance for estimating g has been inflated. Since only the local data are used in the estimation of $\boldsymbol{\beta}$, the resulting estimator for $\boldsymbol{\beta}$ cannot be root-n consistent. We refer to $(\widehat{\boldsymbol{\beta}}(z_0), \widehat{\gamma}(z_0))$ as the naive estimator. To fix the drawbacks of the naive estimator, we next propose a new estimator for $\boldsymbol{\beta}$ that is root-n consistent.

Our proposed estimator is profile likelihood based. Specifically, for a given $\boldsymbol{\beta}$, we obtain an estimator $\widehat{g}^{(k)}(\cdot, \boldsymbol{\beta})$ of $g^{(k)}(\cdot)$, and hence $\widehat{g}(\cdot, \boldsymbol{\beta})$, by maximizing (5) with respect to $\boldsymbol{\gamma}$. Denote by $\widehat{\boldsymbol{\gamma}}(z_0, \boldsymbol{\beta})$ the maximizer. Substituting the estimator $\widehat{g}(\cdot, \boldsymbol{\beta})$ into (3), we can obtain the logarithm of the profile pseudo-partial likelihood:

$$\begin{aligned} \ell_p(\boldsymbol{\beta}) \quad = \quad & \sum_{j=1}^{J} \sum_{i=1}^{n} \xi_{ij} \Delta_{ij} \Big\{ \beta^\tau \boldsymbol{W}_{ij} + \widehat{g}(Z_{ij}, \boldsymbol{\beta}) \\ & - \log \Big(\sum_{l \in \mathcal{R}_j(X_{ij})} \xi_{lj} \exp[\boldsymbol{\beta}^\tau \boldsymbol{W}_{lj} + \widehat{g}(Z_{lj}, \boldsymbol{\beta})] \Big) \Big\}. \qquad (6) \end{aligned}$$

Here and hereafter, for 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 $\widehat{\boldsymbol{\beta}}$ maximize (6) and $\widehat{\boldsymbol{\gamma}} = \widehat{\boldsymbol{\gamma}}(z_0, \widehat{\boldsymbol{\beta}})$. Our proposed estimator for the parametric component is simply $\widehat{\boldsymbol{\beta}}$ and for the nonparametric component is $\widehat{g}(\cdot) = \widehat{g}(\cdot, \widehat{\boldsymbol{\beta}})$.

The proposed profile likelihood estimator can be computed by the following backfitting algorithm. The algorithm takes care of the fact that $g(\cdot, \boldsymbol{\beta})$ is implicitly defined. Let $z_j (j = 1, \cdots, n_g)$ be a grid of points on the range of the exposure variable Z. Our algorithm proceeds as follows.

1. *Initialization.* Use the average of the naive estimator $\bar{\boldsymbol{\beta}} = n_g^{-1} \sum_{j=1}^{n_g} \widehat{\boldsymbol{\beta}}(z_j)$ as the initial value. Set $\widehat{\boldsymbol{\beta}} = \bar{\boldsymbol{\beta}}$.

2. *Estimation of nonparametric component.* Maximize the local pseudo-partial likelihood $\ell(\widehat{\beta}, \boldsymbol{\gamma})$ at each grid point z_j and obtain the nonparametric estimator $\widehat{g}(\cdot, \widehat{\boldsymbol{\beta}})$ at these grid points. Obtain the nonparametric estimator at points $\{Z_{ij}\}$ by using the linear interpolation. We take the bandwidth h suitable for estimation of $\boldsymbol{\beta}$. One example for such a suitable bandwidth is the ad hoc bandwidth in (7) below.

3. *Estimation of parametric component.* With the estimator $\widehat{g}(\cdot, \widehat{\boldsymbol{\beta}})$, maximize

the profile estimator $\ell_p(\boldsymbol{\beta})$ with $g(\cdot,\boldsymbol{\beta}) = \widehat{g}(\cdot,\widehat{\boldsymbol{\beta}})$, using the Newton-Raphson algorithm and the initial value $\widehat{\boldsymbol{\beta}}$ from previous step.

4. *Iteration.* Iterate between the steps 2 and 3 until convergence.

5. *Re-estimating the nonparametric component.* Fix $\boldsymbol{\beta}$ at its estimated value from step 4. The final estimate of $\widehat{g}(\cdot)$ is $\widehat{g}(\cdot,\widehat{\boldsymbol{\beta}})$. At this final step we take the bandwidth h suitable for estimating $g(\cdot)$, such as the estimated optimal bandwidth $\widehat{h}_{opt}$ based on (10) below.

Since the initial estimator $\bar{\beta}$ is consistent, we do not expect many iterations in step 4. Since the initial estimator in step 3 has at least the nonparametric rate $O_p(n^{-(p+1)/(2p+3)})$, two 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 Jiang (1999) and Fan and Chen (2000) in nonparametric models. In fact, according to Robinson (1988), if an initial parametric estimator has 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 $k = 2$, the order of error is $o(n^{-1/2})$. Our experience in simulations shows that the results are consistent with the above theory.

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

$$\widehat{h}_{opt} \times n^{\frac{1}{7}} \times n^{-\frac{1}{3}} = \widehat{h}_{opt} \times n^{-4/21}, \tag{7}$$

where $\widehat{h}_{opt}$ is the estimated optimal bandwidth for $g'(\cdot)$ based on (10) below. In the step 5, however, the goal is to estimate the nonparametric component $g'(\cdot)$, and hence the bandwidth h should be optimal in this respect. In addition, we suggest an even p be used to avoid boundary effects in estimation of $g(\cdot)$.

With the estimators of $\boldsymbol{\beta}$ and $g(\cdot)$, 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:

$$\widehat{\Lambda}_{0j}(t) = \int_0^t [\sum_{i=1}^n \xi_{ij} Y_{ij}(u) \exp\{\widehat{\boldsymbol{\beta}}^T \boldsymbol{W}_{ij}(u) + \widehat{g}(Z_{ij}(u))\}]^{-1} \sum_{i=1}^n \xi_{ij} dN_{ij}(u), \tag{8}$$

where $Y_{ij}(u) = 1(X_{ij} \geq u)$ is the at-risk indicator and $N_{ij}(u) = 1(X_{ij} \leq u, \Delta_{ij} = 1)$ is the associated counting process.

3 Asymptotic Properties

The technical challenges of studying the property of the profile pseudo-likelihood estimator, maximizing (6), arise from the implicit estimate of $\widehat{g}(\cdot, \boldsymbol{\beta})$ which utilizes all observed information. Hence commonly used martingale methods can not 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 an estimator* $\widehat{\boldsymbol{\beta}}$ *which maximizes the profile pseudo-partial likelihood* $\ell_p(\boldsymbol{\beta})$ *such that* $\widehat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}_0$.

Theorem 2 *Under Conditions* (i)-(viii) *in Appendix I, if* $nh^{5/2} \to \infty$ *and* $nh^{2p} \to 0$ *for an even* p*, then the sequence of estimators in Theorem 1 satisfies that* $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ *converges to a Gaussian distribution with mean zero and covariance matrix* $\boldsymbol{\Omega} = \boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)\boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}$.

Remark 1 From the proof of Theorem 2, the second term inside the bracket in $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)$ arises because of the estimation of nonparametric component $g(\cdot)$. The contribution to the covariance matrix due to estimating $g(\cdot)$ is very different from those in the partial linear model (Speckman 1988, Caroll *et al.* 1997), as the current model studies the estimation in the risk domain.

From Theorem 2, the asymptotic covariance matrix of $\widehat{\boldsymbol{\beta}}$ is of sandwich form. This can be estimated by $\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{I}}^{-1}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{I}}^{-1}$, where $\widehat{\boldsymbol{I}}$ and $\widehat{\boldsymbol{\Sigma}}$ are empirical plug-in estimators of $\boldsymbol{I}(\boldsymbol{\beta}_0)$ and $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)$, respectively, which are defined in Appendix I.

Note that $\widehat{\boldsymbol{I}}$ and $\widehat{\boldsymbol{\Sigma}}$ can be shown to be consistent for $\boldsymbol{I}(\boldsymbol{\beta}_0)$ and $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)$, respectively. Hence, $\widehat{\boldsymbol{\Omega}}$ is a consistent estimator of $\boldsymbol{\Omega}$ under the conditions of Theorem 2. Then for the following semiparametric testing problem:

$$H_0: \ \boldsymbol{\beta} = \boldsymbol{\beta}_0 \leftrightarrow H_1: \ \boldsymbol{\beta} \neq \boldsymbol{\beta}_0,$$

where $g(\cdot)$ is a nuisance function, a generalized Wald test statistic W_n can be defined as

$$W_n = n(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\tau \widehat{\boldsymbol{\Omega}}^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0). \tag{9}$$

In particular, this can be applied for testing whether a set of variables are 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 $\boldsymbol{\beta}$.*

Theorem 3 can easily be extended for testing a subset of the coefficient of $\boldsymbol{\beta}$. The nonparametric component possesses the following result.

Theorem 4 *Assume that Conditions* (i)-(v) *hold. If $\widehat{\boldsymbol{\beta}}$ is $\sqrt{n}$-consistent and nh^{2p+3} is bounded, then*

$$\sqrt{nh}\Big[\boldsymbol{H}(\widehat{\gamma} - \gamma) - \boldsymbol{b}_n(z_0)\Big] \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{V}(z_0)),$$

where $\boldsymbol{b}_n(z_0) = \frac{g^{(p+1)}(z_0)}{(p+1)!}\boldsymbol{A}^{-1}\boldsymbol{b}_{p+1}h^{p+1}$, $\boldsymbol{V}(z_0) = \sigma(z_0)\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}$. Furthermore, if $g(\cdot)$ has a continuous $(p+2)$-th derivative, then the asymptotic bias term can be expressed as

$$\boldsymbol{b}_n(z_0) = \frac{g^{(p+1)}(z_0)}{(p+1)!}\boldsymbol{A}^{-1}\boldsymbol{b}_{p+1}h^{p+1} + \frac{g^{(p+2)}(z_0)}{(p+2)!}\boldsymbol{A}^{-1}\boldsymbol{b}_{p+2}h^{p+2}.$$

Remark 2 It is interesting to note that if the failure types within each subject are from the same population, then the asymptotic property of the proposed estimator for $g'(\cdot)$ reduces to that of Fan *et al.* (1997). From the proof of the theorem, one can see that the asymptotic distribution does not depend on the correlation of the failure types within each subject and the estimator of nonparametric component performs as well as if the failure types were independent. For an insight for this phenomena, one can refer to the work of Masry and Fan (1997) and Jiang and Mack (2001).

Corollary 1 *Under the conditions of Theorem 4 with $p = 2$, if K is symmetric, then*

$$\sqrt{nh^3}\Big[\widehat{g}'(z_0) - g'(z_0) - \frac{1}{6}\int u^3 K_1^*(u)\, du\, g^{(3)}(z_0)h^2\Big] \xrightarrow{\mathcal{D}} \mathcal{N}(0, v^2(z_0)),$$

where $v^2(z_0) = \sigma(z_0)\int K_1^*(t)^2\, dt$.

As a result of Corollary 1, the theoretical optimal bandwidth, which minimizes the asymptotic weighted mean integrated squared error

$$\int \Big[\Big\{\frac{1}{6}\int u^3 K_1^*(u)\,du\; g^{(3)}(z)h^2\Big\}^2 + \frac{1}{nh^3}v^2(z)\Big]w(z)\,dz$$

is given by

$$h_{opt} = \Big[\frac{27\int v^2(z)w(z)\,dz}{\int\{g^{(3)}(z)\}^2 w(z)\,dz(\int u^3 K_1^*(u)\,du)^2}\Big]^{1/7} n^{-1/7}. \tag{10}$$

Using the above formula and the widely used plug-in technique or the pre-asymptotic substitution method (see e.g. page 245 of Fan and Yao 2003), one can develop a data-driven approach to the selection of the bandwidth h, but this is out of the scope of current study.

4 Numerical Studies

4.1 Simulations

In this section, we evaluate the finite sample performance of the proposed estimation approach and compare it with Huang's (Huang 1999) by simulations.

First, we consider the case where J failure types are considered for each subject, or, in the clustered failure time data setup, there are J members within each cluster (fixed cluster size). Multivariate failure times will be 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 $(Z_1, \cdots, Z_J)$ and $(\boldsymbol{W}_1, \cdots, \boldsymbol{W}_J)$ is:

$$S(t_1, \cdots, t_J; Z_1, \cdots, Z_J, \boldsymbol{W}_1, \cdots, \boldsymbol{W}_J) = \{\sum_{j=1}^{J} S_j(t_j)^{-1/\theta} - (J-1)\}^{-\theta}, \tag{11}$$

where $S_j(t)$ is the marginal survival probability for the jth failure type. Note that θ is a parameter which represents the degree of dependence within a subject. The relationship between Kendall's tau and θ is $\tau = 1/(2\theta+1)$. The marginal distribution of T_{1j} is taken to be exponential with failure rate

$$\lambda_{0j}\exp\{\boldsymbol{\beta}^\tau \boldsymbol{W}_j + g(Z_j)\}$$

for $g(z) = -8z(1-z^2)$. Then the marginal survival function is

$$S_j(t) = \exp\Big\{-t\lambda_{0j}\exp[\boldsymbol{\beta}_0^\tau \boldsymbol{W}_j + g(Z_j)]\Big\}.$$

We consider the settings with $n = 100$, 200 and $J = 2$ (fixed cluster size: $J_i \equiv J$). The baselines $\lambda_{01} = 1$ and $\lambda_{02} = 4$ are used. The true parameter is set as $\boldsymbol{\beta} = (0.6, 0.4)^\tau$. We first simulate $Z_{ij} \overset{iid}{\sim} U(0,1)$, and $\boldsymbol{W}_{ij} = (W_{ij}^{(1)}, W_{ij}^{(2)})^\tau$ with $W_{ij}^{(1)}$ independently generated from a binomial distribution (taking 1 or 0 each with probability 0.5) and $(W_{i1}^{(2)}, W_{i2}^{(2)})$ from a bivariate normal distribution with the correlation coefficient 0.5 and the marginal distributions $N(0,1)$. For given (Z_{i1}, Z_{i2}) and $(\boldsymbol{W}_{i1}, \boldsymbol{W}_{i2})$, we generated (T_{i1}, T_{i2}) by employing the algorithm on page 2967 in Cai and Shen (2000). The censoring time distribution was generated from exponential distribution with mean chosen to produce a certain amount of censoring.

We used $p = 2$ and employed the Epanechnikov kernel function for smoothing. The parameter θ was set as 100 and 0.01 which correspond to weak and strong correlation within each subject. By the argument in Section 2, the bandwidth h was taken as $0.3n^{-1/3}$ for estimation of $\boldsymbol{\beta}$ in the steps 2-3, and $0.3n^{-1/7}$ for estimation of $g'(\cdot)$ in the step 5 of the algorithm. We assess the sensitivity of the estimation methods as the bandwidth changes over a large range by using a half and twice of the above bandwidths.

The estimators and their standard deviations (SD) for the parameters were evaluated along with the average of the estimated standard error ($\widehat{se}$) for the estimators. The coverage probability (CP$_{se}$) of the 95% confidence intervals for $\boldsymbol{\beta}$ was also calculated based on the normal approximation in Theorem 2. A naive but simple method for estimating the covariance of $\widehat{\boldsymbol{\beta}}$ (and hence the SE of its elements) is to use $\widehat{\boldsymbol{I}}^{-1}$. The corresponding coverage probability (CP_{na}) of the 95% confidence intervals based on only $\widehat{\boldsymbol{I}}^{-1}$ is also computed. We include the naive method here for comparisons.

Tables 1 and 2 report the simulation results for the setting with no censoring and 40% censoring, respectively. It is evident that the proposed estimation performs well since the bias is small, the estimated standard error is close to the sample standard deviation, and the coverage probability of the constructed intervals is close to the nominal level. The naive method fails when there is non-ignorable correlation between survival times (small θ). This is evidenced by the fact that the estimated SEs (columns Mean($\widehat{\boldsymbol{I}}^{-1}$)) are too small and the CP$_{na}$'s are much lower than the nominal level. When the within subject dependence is weak, the naive method works reasonably as expected. In addition, it is seen that the variance of the parameter estimator gets larger as the percent of censoring increases and gets smaller when the sample size

Table 1: *Summary of Simulation Results* ($\beta_1 = 0.6$ *and* $\beta_2 = 0.4$)*.*

Size	Model		No Censoring					
(n, J)	θ	$\boldsymbol{\beta}$	Mean($\widehat{\boldsymbol{\beta}}$)	SD($\widehat{\boldsymbol{\beta}}$)	Mean($\widehat{se}$)	Mean($\widehat{\boldsymbol{I}}^{-1}$)	95% CP$_{se}$	95% CP$_{na}$
(100,2)	0.01	β_1	0.6137	0.2107	0.1987	0.1515	0.938	0.848
		β_2	0.4146	0.1015	0.0890	0.0782	0.914	0.872
(100,2)	100	β_1	0.5959	0.1558	0.1460	0.1512	0.948	0.956
		β_2	0.4052	0.0818	0.0753	0.0782	0.920	0.938
(200,2)	0.01	β_1	0.6151	0.1444	0.1408	0.1050	0.952	0.832
		β_2	0.4050	0.0706	0.0633	0.0537	0.930	0.876
(200,2)	100	β_1	0.6034	0.1068	0.1028	0.1047	0.938	0.936
		β_2	0.4011	0.0559	0.0531	0.0537	0.932	0.940

Table 2: *Summary of Simulation Results* ($\beta_1 = 0.6$ *and* $\beta_2 = 0.4$)*.*

Size	Model		40% Censoring					
(n, J)	θ	$\boldsymbol{\beta}$	Mean($\widehat{\boldsymbol{\beta}}$)	SD($\widehat{\boldsymbol{\beta}}$)	Mean($\widehat{se}$)	Mean($\widehat{\boldsymbol{I}}^{-1}$)	95% CP$_{se}$	95% CP$_{na}$
(100,2)	0.01	β_1	0.6119	0.2578	0.2413	0.1919	0.934	0.854
		β_2	0.4175	0.1233	0.1200	0.0994	0.934	0.900
(100,2)	100	β_1	0.5920	0.2041	0.2092	0.1918	0.948	0.934
		β_2	0.4013	0.1075	0.1109	0.0993	0.938	0.934
(200,2)	0.01	β_1	0.6138	0.1694	0.1708	0.1340	0.952	0.890
		β_2	0.4043	0.0844	0.0866	0.0685	0.958	0.908
(200,2)	100	β_1	0.6121	0.1304	0.1267	0.1337	0.940	0.960
		β_2	0.4015	0.0706	0.0702	0.0684	0.956	0.934

increases.

We now report the performance of the estimated functions. The typical estimated functions with performance at 10th, 50th (median) and 90th percentiles of the mean integrated squared errors (MISE) among the 500 simulations are presented to assess

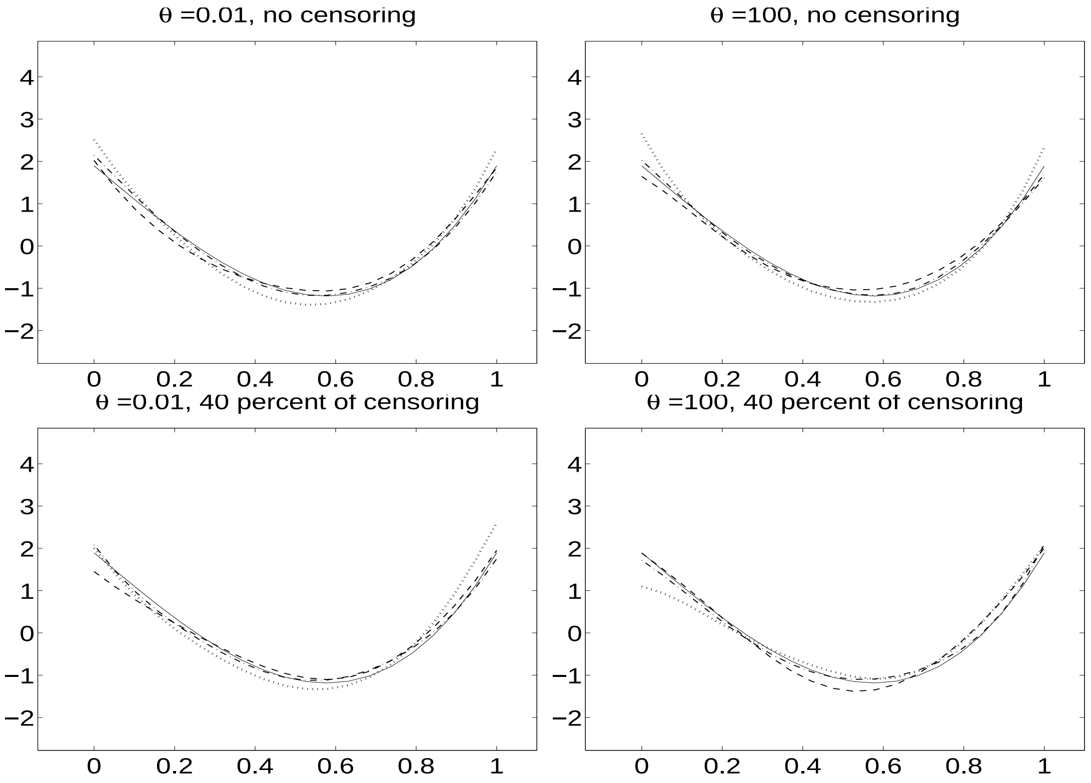

Figure 1: *Typical estimated curves in terms of percentiles of MISEs among the* 500 *simulations with* $n = 200$ *and* $J = 2$*. Solid – true curve, dashed dotted – the 10th percentile, dashed – the 50th percentile, dotted – the 90th percentile.*

the quality of estimated functions. We only presented the case $n = 200$ in Figure 1 to save space. It is seen that the typical estimated curves in Figure 1 capture well the form of the true curve, which reflects the effectiveness of the proposed estimation method.

To appreciate the sampling variability of the estimated nonparametric functions at each point, we present the 2.5th, 50th (median) and 97.5th percentiles of the estimated functions at each grid points among the 500 simulations. The 2.5th and 97.5th percentiles form a 95% pointwise confidence interval for the nonparametric function. This assesses the variability of the estimated functions at each point. Again, to save space, we only present the case with $n = 100$ in Figure 2. The results show that the function is estimated with reasonably good accuracy. The shape of function is captured well.

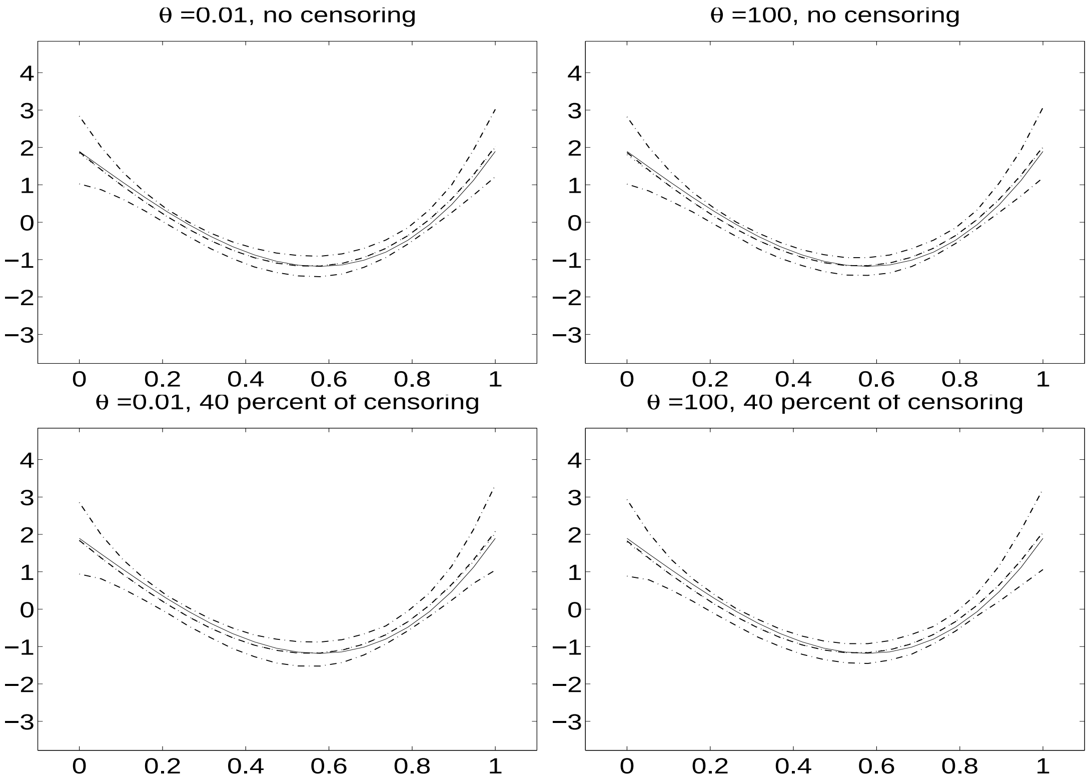

Figure 2: *Sampling variability of estimated functions at each point based on the* 500 *simulations with* $n = 100$ *and* $J = 2$. *Dashed – mean, dotted – median, dash-dotted –* 95% *envelopes formed by the* 2.5*th and* 97.5*th percentiles in* 500 *simulations, solid – true curve.*

Comparing the estimated curves in Figures 1 and 2 for $\theta = 100$ and 0.01 which correspond to weak and strong correlations within each subject, we find that the estimators of the nonparametric part do not heavily depend on the correlation. This exemplifies our statement in Remark 2.

By setting the bandwidths used in simulations twice or a half of those used above, we found that the estimators of the parametric part are very similar to the results above, which reflects that the estimation of finite parameters is robust against the bandwidth over a large range. The results were omitted to save space.

In order to assess the performance of the proposed method under varying cluster size situation, we generate the random cluster size J_i for the ith cluster such that $P(J_i = j) = 1/6$ for $j = 1, \ldots, 6$. For this setup, the maximum cluster size is 6. For cluster i ($i = 1, \cdots, n$), J_i correlated failure times are generated from model (11). We consider $n = 100$, $\lambda_{0j} = j^2$, and the true parameter $\boldsymbol{\beta}$ is set as before.

Table 3 reports the simulation results for the settings with 57% censoring and different correlation structures. It can be seen that for the parametric part the proposed method works well under the varying cluster size situation. Figures 3 and 4 display the typical estimated curves and percentiles of the estimated functions among 500 simulations, respectively. Similar conclusions as before can be drawn for the varying cluster size example.

Table 3: *Summary of Simulation Results Under Varying Cluster Size Situation* ($\beta_1 =$ 0.6 *and* $\beta_2 = 0.4$).

Size	Model		57% Censoring					
(n, J_i)	θ	$\boldsymbol{\beta}$	Mean($\widehat{\boldsymbol{\beta}}$)	SD($\widehat{\boldsymbol{\beta}}$)	Mean($\widehat{se}$)	Mean($\widehat{\boldsymbol{I}}^{-1}$)	95% CP$_{se}$	95% CP$_{na}$
(100,1-6)	0.01	β_1	0.5881	0.1484	0.1405	0.1286	0.918	0.908
		β_2	0.4062	0.0775	0.0727	0.0647	0.920	0.908
(100,1-6)	100	β_1	0.5879	0.1378	0.1352	0.1283	0.928	0.928
		β_2	0.3963	0.0717	0.0668	0.0645	0.918	0.916

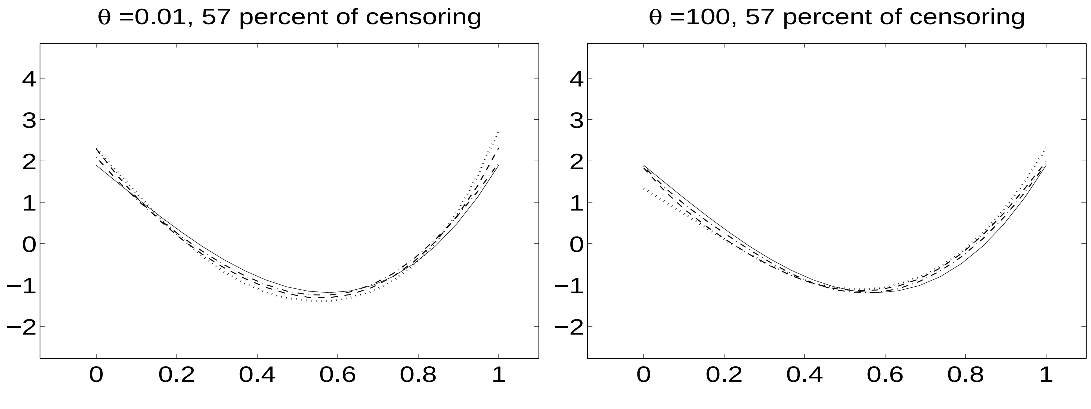

Figure 3: *Typical estimated curves in terms of percentiles of MISEs among the* 500 *simulations with* $n = 100$ *and varying cluster size. Solid – true curve, dashed dotted – the 10th percentile, dashed – the 50th percentile, dotted – the 90th percentile.*

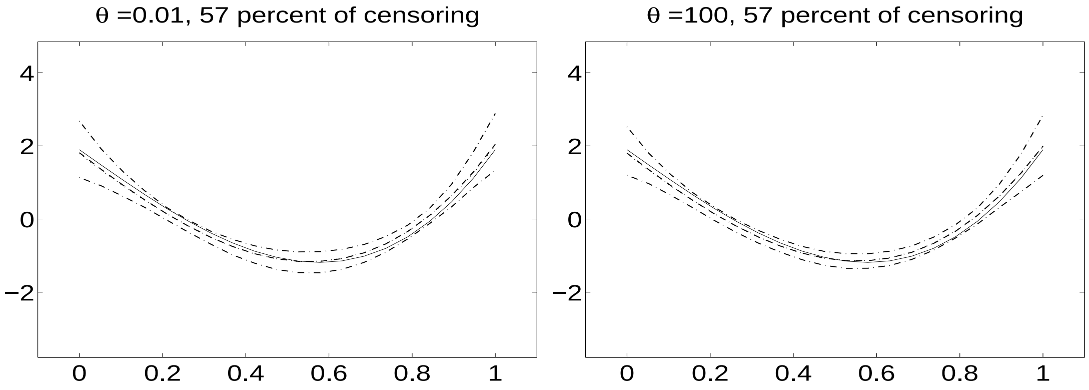

Figure 4: *Sampling variability of estimated functions at each point based on the* 500 *simulations with* $n = 100$ *and varying cluster size. Dashed – mean, dotted – median, dash-dotted –* 95% *envelopes formed by the* 2.5*th and* 97.5*th percentiles in* 500 *simulations, solid – true curve.*

Finally, we compare the proposed profile pseudo-partial likelihood method with the efficient estimation method of Huang (1999). Since Huang's method deals with univariate failure time data, we focus on the model (11) with $J \equiv 1$. The true parameter $\boldsymbol{\beta}$ was again set as before. For Huang's estimation, we need to choose the number and locations of knots used in spline approximation. We follow Huang's suggestion and specify the degrees of freedom to be 3. Table 4 gives the simulation results from both estimation methods. The results with larger degrees of freedom for Huang's method are similar but with increased variances (results not shown). From the table, we see that both estimators perform similarly.

4.2 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 consists 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 and 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 and

Table 4: *Comparison between the proposed and Huang's estimators ($\beta_1 = 0.6$ and $\beta_2 = 0.4$).* PPL - the proposed method, HS - Huang's global spline method. $*$ - not available for estimation.

Size (n, J)	Parameter $\boldsymbol{\beta}$	Method	Mean($\widehat{\boldsymbol{\beta}}$)	SD($\widehat{\boldsymbol{\beta}}$)	Mean($\widehat{se}$)	95% CP_{se}
(100,1)	β_1	PPL	0.6144	0.2256	0.2093	0.934
		HS	0.6292	0.2280	$*$	$*$
	β_2	PPL	0.4016	0.1153	0.1081	0.924
		HS	0.4132	0.1174	$*$	$*$
(200,1)	β_1	PPL	0.6021	0.1526	0.1463	0.944
		HS	0.6136	0.1537	$*$	$*$
	β_2	PPL	0.4041	0.0811	0.0752	0.930
		HS	0.4124	0.0821	$*$	$*$

were disease-free at that examination in the sense that there exists no history of hypertension or glucose intolerance and no previous experiences of a CHD and a CVA. There are a total of 1571 disease-free subjects. The percentage of censoring is about 90.42%. The risk factors of interest were sex, systolic blood pressure, body mass index, cholesterol level, cigarette smoking, and the waiting time. 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. However, there is no evidence in theory and practice validating the linear effects of covariates. To explore the possible nonlinear effects of some covariate (e.g. total cholesterol), we used the proposed method to assess the association between these risk factors on the times to CHD and CVA. Specifically, we employed the following hazards model:

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

where Z_{ij} = cholesterol, and

$$W_{ij} = (\text{Age at ``age 45''}, \text{Smoking}, \text{BMI}, \text{SBP}, \text{Waiting Time}, \text{Gender})^{\tau}.$$

Table 5 reports the estimated parameters and their estimated standard errors

Table 5: *Estimated Parameters for the FHS data.* $\widehat{\boldsymbol{\beta}}$ – the estimated parameters, $\widehat{se}$ – the standard error of $\widehat{\boldsymbol{\beta}}$.

Effect	$\widehat{\boldsymbol{\beta}}$	$\widehat{se}$	P-value
Age at "age 45"	0.0304	0.0887	0.7322
Body mass index, kg/m^2	0.0371	0.0137	0.0065
Systolic blood pressure, mm Hg	0.0171	0.0044	0.0001
Smoking status: yes=1,no=0	0.3578	0.1186	0.0026
Gender: female=1, male=0	-0.5730	0.0993	0.0000
Waiting time, year	0.0031	0.0162	0.8465

along with their p-values from the Wald test in (9). It is evident that all of the selected risk factors are statistically significant at 0.01 significant level except for the confounding factors, Age and Waiting Time. Figure 5 shows the estimated function g and its derivative with 95% confidence intervals based on the normal approximation in Corollary 1. Nonlinear form of g is evidenced by the confidence intervals of its derivative estimator, since the derivative function is not a constant. It reveals that the effect of cholesterol achieves its lowest around the normal levels (160 $\sim$ 170 mg/dl) and is monotone increasing as the cholesterol level gets away from the normal values. Since there are only 6 participants with cholesterol greater than 360, the estimator of the nonparametric function is unreliable on the sparse data region. We displayed in Figure 5 only the estimated functions in the region with Cholesterol less than 360.

5 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-type of models with linear and nonlinear risk factors in the marginal hazard regression. This paper fills in the gap in this area. Without specifying the correlation structure among failure types within each subject, we suggest a profile pseudo-partial likelihood estimation approach to fit the partial linear hazard regression model. Our theory demonstrated that the finite parameters can be estimated at rate of root-n, while

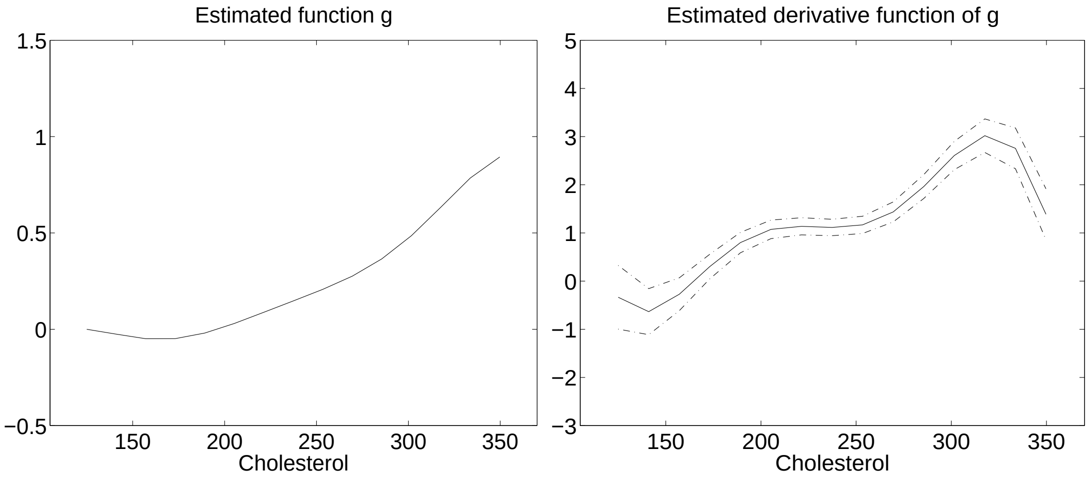

Figure 5: *Estimated function g and its derivative with 95% confidence intervals. Solid – estimated curve; dash-dotted –* 95% *pointwise confidence intervals.*

the nonparametric part can be estimated with the optimal rate independent of the parametric part. We also derived consistent estimates for the covariance matrix of the estimators, which facilitates the inference for the parameters. The methodology is illustrated via an application to the FHS data.

Variable selections based on the non-concave penalized likelihood can also be developed along the framework of Fan and Li (2004). An ongoing research will focus on the testing problem of significance of the nonparametric component. This together with our current work will provide a practical inference tool for the analysis of multivariate survival data by employing the marginal hazard model.

Our model (1) allows one to explore the nonlinear effect of the one-dimensional covariate Z, but the proposed methodology can be extended at least in two directions when Z is a multivariate covariate vector. One is to use the partly linear additive structure as in Huang (1999) and to estimate the nonparametric part via a two-stage procedure with series estimator in the first stage (e.g. Horowitz and Mammen 2004) and our estimation method in the second stage. The other is to employ the partially linear single-index structure in Carroll *et al.* (1997) as pointed out by a referee. For the first topic, we anticipate that the results are similar to ours with an oracle property in the sense that one nonparametric component can be estimated as well as if the others were known. For the second topic, our methodology in this paper continuously applies but with much more techniques involved.

Appendix I: Notations and Assumptions

Let $(\Omega, F, \mathcal{P})$ 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 $\bar{N}_{\cdot j}(u) = \sum_{i=1}^n \xi_{ij} N_{ij}(u) / \sum_{i=1}^n \xi_{ij}$ and $n_j(u) = P(X_{1j} \leq u, \Delta_{1j} = 1)$. Let $\mathcal{F}_{t,ij} = \sigma\{X_{ij}(u^-), \Delta_{ij}, \boldsymbol{W}_{ij}(u), Z_{ij}(u), Y_{ij}(u^-), 0 \leq u \leq t\}$ be 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 $N_{ij}(t)$ is $\mathcal{F}$-adapted, and the observation period is $[0, \tau_0]$, where τ_0 is the study ending time. Then $M_{ij}(t)$ is a martingale with respect to the marginal 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.

For ease of exposition, we consider only the model with time-independent covariate Z. The time-dependent covariate model can be similarly developed. The following notations and conditions are needed for the proofs of our theoretical results. For any vector $\boldsymbol{a}$, define $\boldsymbol{a}^{\otimes k} = 1, \boldsymbol{a}$, and $\boldsymbol{a}\boldsymbol{a}^\tau$, respectively, for $k = 0, 1, 2$, where $\boldsymbol{a}^\tau$ denotes the transpose of $\boldsymbol{a}$. Let $\tilde{\boldsymbol{u}} = (u, \cdots, u^p)^\tau$, $\boldsymbol{\nu}_k = \int \tilde{\boldsymbol{u}}^{\otimes k} K(u)\,du$ and $\boldsymbol{\nu}_k^* = \int u\tilde{\boldsymbol{u}}^{\otimes k} K(u)\,du$ for $k = 0, 1$. Put $K_1^*(t) = tK(t)/\int u^2 K(u)\,du$, $\boldsymbol{A} = \boldsymbol{\nu}_2 - \boldsymbol{\nu}_1^{\otimes 2}$, $\boldsymbol{B} = \int K^2(u)(\tilde{\boldsymbol{u}} - \boldsymbol{\nu}_1)^{\otimes 2}\,du$, and $\boldsymbol{b}_k = \int u^k(\tilde{\boldsymbol{u}} - \boldsymbol{\nu}_1)K(u)\,du$, for $k = 1$, $p+1$, and $p+2$. Denote by $c(K) = \boldsymbol{e}_1^\tau \boldsymbol{A}^{-1}\boldsymbol{b}_1$ and $d(K) = \boldsymbol{e}_1^\tau \boldsymbol{A}^{-1}(\boldsymbol{\nu}_1^* - \boldsymbol{\nu}_1\boldsymbol{\nu}_0^*)$, with $\boldsymbol{e}_1$ as a vector with a 1 in the first position and 0's elsewhere.

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

(ii) $nh \to \infty$ and $h \to 0$, as $n \to \infty$. Let $\boldsymbol{H} = \text{diag}(h, \cdots, h^p)$ and $\tilde{Z}_{ij}^* = \boldsymbol{H}^{-1}\tilde{Z}_{ij}$.

(iii) The density $f_j(\cdot)$ of Z_{1j} is of compact support and has a bounded second derivative for $j = 1, \cdots, J$, where $J < \infty$. Assume that for each j, $\{\xi_{ij}\}_{i=1}^n$ are independent and identically distributed. Suppose that ξ_{ij} is independent of $\{X_{ij}, \Delta_{ij}, \boldsymbol{W}_{ij}, Z_{ij}\}$. Let $p_j = P(\xi_{1j} = 1)$, $j = 1, \cdots, J$.

(iv) The function $g(\cdot)$ has a continuous $(p+1)$-th derivative with $g(0) = 0$.

(v) Let $\boldsymbol{\beta}_0$ be the true value of the parameter $\boldsymbol{\beta}$. The conditional expectations

$$\rho_{jk}(u|z) = E[s_{1j}(u, \boldsymbol{\beta}_0)(\boldsymbol{W}_{1j}(u))^{\otimes k} | Z_{1j} = z]$$

are equi-continuous in z, for $j = 1, \cdots, J$ and $k = 0, 1$, where $s_{ij}(u, \boldsymbol{\beta}_0) = Y_{ij}(u) \exp(\boldsymbol{\beta}_0^\tau \boldsymbol{W}_{ij}(u) + g(Z_{ij}))$ is the risk function for the jth failure type in the ith subject. The conditional expectation $\rho_{j0}(u|z)$ has a continuous second derivative with respect to z. Let $\eta_{jk}(u|z) = \rho_{jk}(u|z) f_j(z)$ for $k = 0, 1, 2$. Put $\Lambda_j(t, z) = \int_0^t \rho_{j0}(u|z) \lambda_{0j}(u)\, du$. Assume that

$$\sigma^{-1}(z) = \sum_{j=1}^{J} p_j f_j(z) \Lambda_j(\tau_0, z) > 0$$

for $z \in \cup_{j=1}^J \mathrm{supp}(f_j)$. It can be shown that

$$\sigma^{-1}(z) = \sum_{j=1}^{J} p_j f_j(z) E[\Delta_{1j} | Z_{1j} = z].$$

Assume that $\sigma(z)$ has a bounded second derivative in $\cup_{j=1}^J \mathrm{supp}(f_j)$.

(vi) $\int_0^{\tau_0} \lambda_{0j}(t)\, dt < \infty$ for each $j \in \{1, 2, \cdots, J\}$.

(vii) There exists a neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}_0$ such that for $k = 0, 1, 2, 3$,

$$E\{ \sup_{(\boldsymbol{\beta}, t) \in \mathcal{B} \times [0, \tau_0]} Y_{ij}(t) ||\boldsymbol{W}_{ij}(t)||^k \exp[\boldsymbol{\beta}^\tau \boldsymbol{W}_{ij}(t) + g(Z_{ij})]\} < \infty.$$

(viii) Let $\alpha(z) = \int_0^{\tau_0} \mathcal{D}_z[\frac{\rho_{j1}(u|z)}{\rho_{j0}(u|z)}] \rho_{j0}(u|z) \lambda_{0j}(u)\, du$,

$$\boldsymbol{\chi}(z) = -d(K) \sum_{j=1}^{J} p_j \int_0^z \sigma(z^*) f_j(z^*) \alpha(z^*)\, dz^*,$$

and $r_{jk}(\boldsymbol{\beta}, u) = E\{s_{1j}(u, \boldsymbol{\beta})(\boldsymbol{W}_{1j}(u) + \boldsymbol{\chi}(Z_{1j}))^{\otimes k}\}$ for $k = 0, 1, 2$, where $\mathcal{D}_z$ denotes the derivative with respect to z. The functions $r_{j0}(\cdot, u)$, $r_{j1}(\cdot, u)$ and $r_{j2}(\cdot, u)$ are continuous in $\boldsymbol{\beta} \in \mathcal{B}$, uniformly in $u \in [0, \tau_0]$; r_{j0} is bounded away from zero on $\mathcal{B} \times [0, \tau_0]$; r_{j1} and r_{j2} are bounded on $\mathcal{B} \times [0, \tau_0]$. The matrix $\boldsymbol{I}(\boldsymbol{\beta}_0)$ is finite positive definite, where

$$\boldsymbol{I}(\boldsymbol{\beta}) = \sum_{j=1}^{J} p_j \int_0^{\tau_0} \Big[\frac{r_{j2}(\boldsymbol{\beta}, u)}{r_{j0}(\boldsymbol{\beta}, u)} - \Big(\frac{r_{j1}(\boldsymbol{\beta}, u)}{r_{j0}(\boldsymbol{\beta}, u)}\Big)^{\otimes 2}\Big] r_{j0}(\boldsymbol{\beta}_0, u) \lambda_{0j}(u)\, du.$$

The above conditions are similar to those in Andersen and Gill (1982) and Fan *et al.* (1997). Conditions (i)-(v) are standard for nonparametric component estimation using local partial likelihood. Conditions (vi)-(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.

Denote by

$$\varphi(u,z;\boldsymbol{\beta}_0)=\rho_{j1}(u|z)+\boldsymbol{\chi}(z)\rho_{j0}(u|z)-\rho_{j0}(u|z)\frac{r_{j1}(\boldsymbol{\beta}_0,u)}{r_{j0}(\boldsymbol{\beta}_0,u)},$$

$$\boldsymbol{a}_j(z)=\int_0^{\tau_0}\varphi(u,z;\boldsymbol{\beta}_0)r_{j0}^{-1}(\boldsymbol{\beta}_0,u)\,dn_j(u),$$

$$\boldsymbol{s}(z)\;=\;\sum_{j=1}^{J}p_j\int_{-\infty}^{z}\boldsymbol{a}_j(z^*)f_j(z^*)\,dz^*,$$

$$G_{ij}(\boldsymbol{\beta})=\int_0^{\tau_0}H_{ij}(u)dM_{ij}(u),\quad\text{and}\quad\boldsymbol{\Sigma}(\boldsymbol{\beta})\;=\;E\Big\{\sum_{j=1}^{J}\xi_{1j}G_{1j}(\boldsymbol{\beta})\Big\}^{\otimes 2},$$

where

$$H_{ij}(u)=\boldsymbol{W}_{ij}(u)+\boldsymbol{\chi}(Z_{ij})-\frac{r_{j1}(\boldsymbol{\beta},u)}{r_{j0}(\boldsymbol{\beta},u)}-V(u,Z_{ij}),$$

$$V(u,Z_{ij})=\sigma(Z_{ij})\boldsymbol{s}(Z_{ij})\{c(K)\mathcal{D}_z[\log\sigma(Z_{ij})]+d(K)\mathcal{D}_z[\log\eta_{j0}(u|Z_{ij})]\}.$$

Let $\widehat{F}_j(z^*)$ be the empirical distribution function for Z_{1j} based on the observed $\{Z_{ij}\}_{i=1}^n$, that is, $\widehat{F}_j(z^*)=\sum_{i=1}^n\xi_{ij}1(Z_{ij}\le z^*)/\sum_{i=1}^n\xi_{ij}$. Denote by

$$\widehat{\rho}_{jk}(u|z)=\widehat{E}[\widehat{s}_{1j}(u)(\boldsymbol{W}_{1j}(u))^{\otimes k}|Z_{1j}=z],$$

where $\widehat{s}_{ij}(u)=Y_{ij}(u)\exp(\widehat{\boldsymbol{\beta}}^{\tau}\boldsymbol{W}_{ij}(u)+\widehat{g}(Z_{ij}))$ is the estimated risk corresponding to $s_{ij}(u,\boldsymbol{\beta}_0)$, and $\widehat{E}(\cdot|\cdot)$ denotes a consistent estimator of $E(\cdot|\cdot)$ such as the Nadaraya-Watson estimator or the local linear estimator in nonparametric regression. Put $\widehat{\alpha}(z)=\int_0^{\tau_0}\mathcal{D}_z\left[\frac{\widehat{\rho}_{j1}(u|z)}{\widehat{\rho}_{j0}(u|z)}\right]\widehat{\rho}_{j0}(u|z)\,d\widehat{\Lambda}_{0j}(u)$. Then the plug-in estimator of $\boldsymbol{\chi}(z)$ is

$$\widehat{\boldsymbol{\chi}}(z)=-d(K)\sum_{j=1}^{J}\widehat{p}_j\int_0^{z}\widehat{\sigma}(z^*)\widehat{\alpha}(z^*)\,d\widehat{F}_j(z^*),$$

where $\widehat{p}_j=n^{-1}\sum_{i=1}^n1(\xi_{ij}=1)$, $\widehat{\sigma}(z)=\Big\{\sum_{j=1}^J\widehat{p}_j\widehat{f}_j(z)\widehat{E}[\Delta_{1j}|Z_{1j}=z]\Big\}^{-1}$. Let the empirical estimator of $r_{jk}(t)$ be

$$\widehat{r}_{jk}(t)=\sum_{i=1}^{n}\xi_{ij}\widehat{s}_{ij}(t)(\boldsymbol{W}_{ij}(t)+\widehat{\boldsymbol{\chi}}(Z_{ij}))^{\otimes k}/\sum_{i=1}^{n}\xi_{ij}.$$

Then the empirical estimator of $\boldsymbol{I}$ is

$$\widehat{\boldsymbol{I}} = \sum_{j=1}^{J} \widehat{p}_j \sum_{i=1}^{n} \xi_{ij} \Delta_{ij} \{ \frac{\widehat{r}_{j2}(X_{ij})}{\widehat{r}_{j0}(X_{ij})} - \Big(\frac{\widehat{r}_{j1}(X_{ij})}{\widehat{r}_{j0}(X_{ij})} \Big)^{\otimes 2} \} / \sum_{i=1}^{n} \xi_{ij}.$$

The matrix $\widehat{\boldsymbol{\Sigma}}$ is defined as follows. Let $\widehat{\eta}_{j0}(u|z)$ and $\widehat{\boldsymbol{s}}(z)$ be the plug-in estimators of $\eta_{j0}(u|z)$ and $\boldsymbol{s}(z)$ respectively, that is, $\widehat{\eta}_{j0}(u|z) = \widehat{f}_j(z)\widehat{\rho}_{j0}(u|z)$ and $\widehat{\boldsymbol{s}}(z) = \sum_{j=1}^{J} \widehat{p}_j \int_{-\infty}^{z} \widehat{\boldsymbol{a}}_j(z^*)\, d\widehat{F}_j(z^*)$, respectively, where

$$\widehat{\boldsymbol{a}}_j(z) = \int_0^{\tau_0} [\widehat{\rho}_{j1}(u|z) + \widehat{\boldsymbol{\chi}}(z)\widehat{\rho}_{j0}(u|z) - \widehat{\rho}_{j0}(u|z) \frac{\widehat{r}_{j1}(u)}{\widehat{r}_{j0}(u)}] \widehat{r}_{j0}^{-1}(u)\, d\bar{N}_{\cdot j}(u)$$

is the plug-in estimator of $\boldsymbol{a}_j(z)$. Set the empirical plug-in estimator of $\boldsymbol{G}_{ij}$ as

$$\widehat{\boldsymbol{G}}_{ij} \quad = \quad \Delta_{ij}\widehat{H}_{ij}(X_{ij}) - \sum_{m=1}^{n} \xi_{mj}\Delta_{mj}\widehat{s}_{ij}(X_{mj})\widehat{r}_{j0}^{-1}(X_{mj})\widehat{H}_{ij}(X_{mj}) \Big/ \sum_{m=1}^{n} \xi_{mj},$$

where $\widehat{H}_{ij}(u) = \boldsymbol{W}_{ij}(u) + \widehat{\boldsymbol{\chi}}(Z_{ij}) - \frac{\widehat{r}_{j1}(u)}{\widehat{r}_{j0}(u)} - \widehat{V}(u, Z_{ij})$ with $\widehat{V}(u, Z_{ij})$ being the plug-in estimator of $V(u, Z_{ij})$, that is,

$$\widehat{V}(u, Z_{ij}) = \widehat{\sigma}(Z_{ij})\widehat{\boldsymbol{s}}(Z_{ij})[c(K)\mathcal{D}_z(\log \widehat{\sigma}(Z_{ij})) + d(K)\mathcal{D}_z(\log \widehat{\eta}_{j0}(u|Z_{ij}))].$$

Then the empirical estimator of $\boldsymbol{\Sigma}$ is $\widehat{\boldsymbol{\Sigma}} = n^{-1} \sum_{i=1}^{n} [\sum_{j=1}^{J} \xi_{ij} \widehat{G}_{ij}]^{\otimes 2}$.

Appendix II: Proofs of Theorems

The proofs involve the martingale theory, the theory of empirical processes and the techniques commonly used in nonparametric literature.

Given the identifiability condition $\widehat{g}(0, \boldsymbol{\beta}) = 0$, we have $\widehat{g}(z_0, \boldsymbol{\beta}) = \int_0^{z_0} \widehat{g}'(z, \boldsymbol{\beta})\, dz$. Let $\boldsymbol{\chi}_n(z_0) = \frac{\partial \widehat{g}(z_0, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = \int_0^{z_0} \frac{\partial \widehat{g}'(z, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} dz$ and

$$\boldsymbol{\kappa}_n(z_0) = \frac{\partial^2 \widehat{g}(z, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\tau} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = \int_0^{z_0} \frac{\partial^2 \widehat{g}'(z, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\tau} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} dz_0.$$

Then for any $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}_0$, using Taylor expansion we have

$$\widehat{g}(z_0, \boldsymbol{\beta}) \quad \approx \quad \widehat{g}(z_0, \boldsymbol{\beta}_0) + \boldsymbol{\chi}_n(z_0)^\tau (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\tau \boldsymbol{\kappa}_n(z_0) (\boldsymbol{\beta} - \boldsymbol{\beta}_0).$$

Recall that the global profile pseudo-partial likelihood is (6), which can be written as

$$\begin{aligned} \ell_p(\boldsymbol{\beta}) \quad \equiv \quad & \sum_{j=1}^{J} \sum_{i=1}^{n} \xi_{ij} \int_0^{\tau_0} \Big\{ \boldsymbol{\beta}^\tau \boldsymbol{W}_{ij}(u) + \widehat{g}(Z_{ij}, \boldsymbol{\beta}) \\ & - \log \Big(\sum_{\ell=1}^{n} Y_{\ell j}(u) \exp[\boldsymbol{\beta}^\tau \boldsymbol{W}_{\ell j}(u) + \widehat{g}(Z_{\ell j}, \boldsymbol{\beta})] \Big) \Big\} dN_{ij}(u). \qquad \text{(A.1)} \end{aligned}$$

By Taylor expansion around point $\boldsymbol{\beta}_0$,

$$\begin{aligned} \ell_p(\boldsymbol{\beta}) &= \ell_p(\boldsymbol{\beta}_0) + (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^\tau \frac{\partial \ell_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \\ &\quad + \frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^\tau \frac{\partial^2 \ell_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\tau}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} (\boldsymbol{\beta}-\boldsymbol{\beta}_0) + R_n(\boldsymbol{\beta}^*), \end{aligned} \tag{A.2}$$

where $\boldsymbol{\beta}^*$ lies between $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$ and

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

with β_j and β_{0j} being the j-th elements of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$ respectively. It can be shown that $n^{-1}\frac{\partial^3 \ell_p(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_k \partial \beta_\ell}$ is bounded in probability, and hence $n^{-1}R_n(\boldsymbol{\beta}) = O_p(||\boldsymbol{\beta}-\boldsymbol{\beta}_0||^3)$ for $\boldsymbol{\beta} \in \mathcal{B}$.

Let $\boldsymbol{\gamma}^* = \boldsymbol{H}\boldsymbol{\gamma}$, and $\widehat{\boldsymbol{\gamma}}^*(\boldsymbol{\beta}) \equiv \widehat{\boldsymbol{\gamma}}^*(z_0,\boldsymbol{\beta}) = \boldsymbol{H}\widehat{\boldsymbol{\gamma}}(z_0,\boldsymbol{\beta})$. Define for $k=0,1,2$

$$\Phi_{njk}(u,\boldsymbol{\beta},\boldsymbol{\gamma}^*) = \sum_{\ell=1}^{n} \xi_{\ell j}\tilde{s}_{\ell j}(u;\boldsymbol{\beta},\boldsymbol{\gamma}^*)\tilde{Z}_{\ell j}^{*\otimes k} K_h(Z_{\ell j}-z_0)/\sum_{\ell=1}^{n}\xi_{\ell j}, \tag{A.4}$$

where $\tilde{s}_{\ell j}(u;\boldsymbol{\beta},\boldsymbol{\gamma}^*) = Y_{\ell j}(u)\exp[\boldsymbol{\beta}^\tau \boldsymbol{W}_{\ell j}(u) + \boldsymbol{\gamma}^{*\tau}\tilde{Z}_{\ell j}^*]$. Note that $\boldsymbol{\gamma}^{*\tau}\tilde{Z}_{\ell j}^* = g(Z_{\ell j}) - g(z_0) + O(h^{p+1})$. Simple algebra gives that for $k=0,1$,

$$E[\Phi_{njk}(u,\boldsymbol{\beta}_0,\boldsymbol{\gamma}^*)] = e^{-g(z_0)}\{\eta_{j0}(u|z_0)\boldsymbol{\nu}_k + h\boldsymbol{\nu}_k^*\mathcal{D}_z[\eta_{j0}(u|z_0)]\} + O(h^2) \tag{A.5}$$

and $\mathrm{var}[\Phi_{njk}(u,\boldsymbol{\beta}_0,\boldsymbol{\gamma}^*)] = O(\frac{1}{nh})$, uniformly for $u \in [0,\tau_0]$. Then using the same argument as for Lemma 1 of Fan *et al.* (1997), we get

$$\sup_{0\le u\le \tau_0} ||\frac{\Phi_{nj1}(u,\boldsymbol{\beta}_0,\boldsymbol{\gamma}^*)}{\Phi_{nj0}(u,\boldsymbol{\beta}_0,\boldsymbol{\gamma}^*)} - \boldsymbol{\nu}_1|| \xrightarrow{P} 0, \tag{A.6}$$

and

$$\frac{\Phi_{nj0}(u,\boldsymbol{\beta}_0,\boldsymbol{\gamma}^*)\Phi_{nj2}(u,\boldsymbol{\beta}_0,\boldsymbol{\gamma}^*) - \Phi_{nj1}^{\otimes 2}(u,\boldsymbol{\beta}_0,\boldsymbol{\gamma}^*)}{\Phi_{nj0}^2(u,\boldsymbol{\beta}_0,\boldsymbol{\gamma}^*)} = \boldsymbol{A} + o_p(1), \tag{A.7}$$

uniformly for $u \in [0,\tau_0]$.

In the following, we will first give the proof of Theorem 4, and then introduce some lemmas for the proofs of Theorems 1 and 2.

Proof of Theorem 4. By (5), $\widehat{\boldsymbol{\gamma}}^* \equiv \widehat{\boldsymbol{\gamma}}^*(z_0,\widehat{\boldsymbol{\beta}})$ satisfies the equation

$$n^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n}\xi_{ij}\int_0^{\tau_0} K_h(Z_{ij}-z_0)\Big\{\tilde{Z}_{ij}^* - \frac{\Phi_{nj1}(u,\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\gamma}}^*)}{\Phi_{nj0}(u,\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\gamma}}^*)}\Big\}dN_{ij}(u) = 0,$$

where Φ_{njk} is defined in (A.4). It can be shown from the assumption $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p(n^{-1/2})$ that

$$\sup_{u\in[0,\tau_0]} \Big\| \frac{\Phi_{nj1}(u, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}^*)}{\Phi_{nj0}(u, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}^*)} - \frac{\Phi_{nj1}(u, \boldsymbol{\beta}_0, \widehat{\boldsymbol{\gamma}}^*)}{\Phi_{nj0}(u, \boldsymbol{\beta}_0, \widehat{\boldsymbol{\gamma}}^*)} \Big\| = O_p(n^{-1/2}).$$

Thus, $\widehat{\boldsymbol{\gamma}}^*$ satisfies

$$n^{-1} \sum_{j=1}^{J} \sum_{i=1}^{n} \xi_{ij} \int_0^{\tau_0} K_h(Z_{ij} - z_0) \Big\{ \tilde{Z}_{ij}^* - \frac{\Phi_{nj1}(u, \boldsymbol{\beta}_0, \widehat{\boldsymbol{\gamma}}^*)}{\Phi_{nj0}(u, \boldsymbol{\beta}_0, \widehat{\boldsymbol{\gamma}}^*)} \Big\} \, dN_{ij}(u) = O_p(n^{-1/2}).$$

As before, we denote by $\widehat{\boldsymbol{U}}(\widehat{\boldsymbol{\gamma}}^*, z_0)$ the left-hand side of the above equation. Then $\widehat{\boldsymbol{U}}(\widehat{\boldsymbol{\gamma}}^*, z_0) = o_p(1/\sqrt{nh})$. The consistency of $\widehat{\boldsymbol{\gamma}}^*$ can be derived by using the same argument as in Fan *et al.* (1997). Then by Taylor's expansion, we obtain

$$\widehat{\boldsymbol{U}}(\boldsymbol{\gamma}^*, z_0) + \frac{\partial \widehat{U}(\tilde{\boldsymbol{\gamma}}^*, z_0)}{\partial \boldsymbol{\gamma}^*} (\widehat{\boldsymbol{\gamma}}^* - \boldsymbol{\gamma}^*) = o_p(1/\sqrt{nh}), \tag{A.8}$$

where $\tilde{\boldsymbol{\gamma}}^*$ lies between $\widehat{\boldsymbol{\gamma}}^*$ and $\boldsymbol{\gamma}^*$ and hence $\tilde{\boldsymbol{\gamma}}^* \to \boldsymbol{\gamma}^*$ in probability. Simple algebra gives that

$$\begin{aligned} -\frac{\partial \widehat{\boldsymbol{U}}(\boldsymbol{\gamma}^*, z_0)}{\partial \boldsymbol{\gamma}^*} &= n^{-1} \sum_{j=1}^{J} \sum_{i=1}^{n} \xi_{ij} \int_0^{\tau_0} K_h(Z_{ij} - z_0) \\ &\quad \times \frac{\Phi_{nj0}(u, \boldsymbol{\beta}_0, \boldsymbol{\gamma}^*) \Phi_{nj2}(u, \boldsymbol{\beta}_0, \boldsymbol{\gamma}^*) - \Phi_{nj1}^{\otimes 2}(u, \boldsymbol{\beta}_0, \boldsymbol{\gamma}^*)}{\Phi_{nj0}^2(u, \boldsymbol{\beta}_0, \boldsymbol{\gamma}^*)} \, dN_{ij}(u). \end{aligned}$$

It follows from (A.7) that

$$-\frac{\partial \widehat{\boldsymbol{U}}(\boldsymbol{\gamma}^*, z_0)}{\partial \boldsymbol{\gamma}^*} = \boldsymbol{A} \sigma^{-1}(z_0) + o_p(1). \tag{A.9}$$

In addition, using the Doob-Meyer Decomposition $N_{ij}(u) = M_{ij}(u) + \int_0^u Y_{ij}(s) \lambda_{ij}(s) \, ds$, we can express $\widehat{\boldsymbol{U}}(\boldsymbol{\gamma}^*, z_0)$ as

$$\widehat{\boldsymbol{U}}(\boldsymbol{\gamma}^*, z_0) = \boldsymbol{d}_n(\tau_0) + \boldsymbol{q}_n(\tau_0), \tag{A.10}$$

where $d_n(\tau_0)$ and $q_n(\tau_0)$ are defined similarly to $\widehat{\boldsymbol{U}}(\boldsymbol{\gamma}^*, z_0)$ except that $dN_{ij}(u)$ is replaced by $Y_{ij}(u)\lambda_{ij}(u)du$. Note that

$$\begin{aligned} \boldsymbol{q}_n(\tau_0) &= n^{-1} \sum_{j=1}^{J} \sum_{i=1}^{n} \xi_{ij} \int_0^{\tau_0} K_h(Z_{ij} - z_0) \Big\{ \tilde{Z}_{ij}^* - \frac{\Phi_{nj1}(u, \boldsymbol{\beta}_0, \boldsymbol{\gamma}^*)}{\Phi_{nj0}(u, \boldsymbol{\beta}_0, \boldsymbol{\gamma}^*)} \Big\} Y_{ij}(u) \exp[\boldsymbol{\beta}_0^\tau \boldsymbol{W}_{ij}(u)] \\ &\quad \times \{\exp(g(Z_{ij})) - \exp(g(z_0) + \boldsymbol{\gamma}^{*T} \tilde{Z}_{ij}^*)\} \lambda_{0j}(u) \, du. \end{aligned} \tag{A.11}$$

Since the kernel function $K(\cdot)$ is of compact support, it suffices to consider only $Z_{ij} - z_0 = O(h)$ in the asymptotic analysis. By Taylor's expansion of $\exp[g(Z_{ij})]$ around z_0 and (A.6), we have

$$\begin{aligned} \boldsymbol{q}_n(\tau_0) &= \frac{h^{p+1}}{(p+1)!} g^{(p+1)}(z_0) \boldsymbol{b}_{p+1} \sigma^{-1}(z_0) \\ &\quad + \frac{h^{p+2}}{(p+2)!} g^{(p+2)}(z_0) \boldsymbol{b}_{p+2} \sigma^{-1}(z_0) + o_p(h^{p+2}). \qquad \text{(A.12)} \end{aligned}$$

Rewrite $\boldsymbol{d}_n(\tau_0)$ in (A.10) as

$$\begin{aligned} \boldsymbol{d}_n(\tau_0) &= \frac{1}{n} \sum_{j=1}^{J} \sum_{i=1}^{n} \xi_{ij} \int_0^{\tau_0} K_h(Z_{ij} - z_0)(\tilde{Z}_{ij}^* - \boldsymbol{\nu}_1)\, dM_{ij}(u) \\ &\quad + \frac{1}{n} \sum_{j=1}^{J} \sum_{i=1}^{n} \xi_{ij} \int_0^{\tau_0} K_h(Z_{ij} - z_0) \Big[\boldsymbol{\nu}_1 - \frac{\Phi_{nj1}(u, \boldsymbol{\beta}_0, \boldsymbol{\gamma}^*)}{\Phi_{nj0}(u, \boldsymbol{\beta}_0, \boldsymbol{\gamma}^*)} \Big]\, dM_{ij}(u)\text{(A.13)} \end{aligned}$$

Note that $\boldsymbol{\nu}_1 - \frac{\Phi_{nj1}(u, \boldsymbol{\beta}_0, \boldsymbol{\gamma}^*)}{\Phi_{nj0}(u, \boldsymbol{\beta}_0, \boldsymbol{\gamma}^*)}$ is a bounded $\mathcal{F}_{u,ij}$-predictable process, which combined with (A.6) and the dominated convergence theorem ensures that the second term above is of mean zero and variance $o(\frac{1}{nh})$. Hence,

$$\boldsymbol{d}_n(\tau_0) = \boldsymbol{d}_{n1}(\tau_0) + o_p(1/\sqrt{nh}),$$

where $\boldsymbol{d}_{n1}(\tau_0)$ is the first term in (A.13). We now treat the process $\boldsymbol{d}_{n1}(t)$, employing the martingale central limit theorem (see Theorem 5.35 of Fleming and Harrington (1991)). It can be shown that the asymptotic variance of $\boldsymbol{d}_n^*(\tau_0) = \sqrt{nh}\boldsymbol{d}_{n1}(\tau_0)$ is

$$\text{var}(\boldsymbol{d}_n^*(\tau_0)) = \boldsymbol{B}\sigma^{-1}(z_0) + \boldsymbol{D}_{12}(z_0) + o(1),$$

where

$$\begin{aligned} \boldsymbol{D}_{12}(z_0) &= \lim_{n \to \infty} E\Big\{ \sum_{j=1}^{J} \sum_{k=1, \neq j}^{J} \xi_{1j} \xi_{1k} h \int_0^{\tau_0} K_h(Z_{1j} - z_0)(\tilde{Z}_{1j}^* - \boldsymbol{\nu}_1)\, dM_{1j}(u) \\ &\quad \times \int_0^{\tau_0} K_h(Z_{1k} - z_0)(\tilde{Z}_{1k}^* - \boldsymbol{\nu}_1)^\tau\, dM_{1k}(u) \Big\}. \end{aligned}$$

Using the boundedness of $E[M_{1j}(\tau_0) M_{1k}(\tau_0) | Z_{1j}, Z_{1k}]$, we obtain

$$\boldsymbol{D}_{12}(z_0) = O(h\, E[K_h(Z_{1j} - z_0)(\tilde{Z}_{1j}^* - \boldsymbol{\nu}_1) K_h(Z_{1k} - z_0)(\tilde{Z}_{1k}^* - \boldsymbol{\nu}_1)^\tau]) = O(h).$$

Write $\boldsymbol{d}_n^*(t)$ as

$$\frac{\sqrt{nh}}{n} \sum_{i=1}^{n} \sum_{j=1}^{J} \xi_{ij} \int_0^t K_h(Z_{ij} - z_0) H_{ij}^*(u)\, dM_{ij}(u) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sqrt{h} \boldsymbol{D}_i^*(t, h).$$

Then for $|Z_{ij} - z_0| = O(h)$, $H^*_{ij}(u)$ is a bounded random variable. It can be shown that the following Lindeberg condition holds for any given $\varepsilon > 0$:

$$n^{-1} \sum_{i=1}^{n} hE\Big[||\boldsymbol{D}^*_i(t,h)||^2 \boldsymbol{I}(\sqrt{h/n}\, \boldsymbol{D}^*_i(t,h) > \varepsilon)\Big] \to 0.$$

This establishes the asymptotic normality of $\boldsymbol{d}^*_n(\tau_0)$ and hence $\sqrt{nh}\boldsymbol{d}_n(\tau_0)$, which together with (A.8)-(A.10) and (A.12) yield the result of the theorem.

The following lemmas are needed for proving Theorems 1 and 2. Recall the expressions at the beginning of the appendix.

Lemma 1 *Under Conditions (i)-(viii), if* $nh^2 \to \infty$, *then the following items hold uniformly for* $z \in \cup_{j=1}^J \text{supp}(f_j)$,

(i) $\boldsymbol{\kappa}_n(z) = \boldsymbol{\kappa}(z) + o_p(1)$, *where* $\boldsymbol{\kappa}(z) = -d(K)\sum_{j=1}^J p_j \int_0^z \sigma(z_0) f_j(z_0)\rho^*_j(z_0)\, dz_0$ *with*

$$\rho^*_j(z_0) = \int_0^{\tau_0} \mathcal{D}_z\Big[\frac{\rho_{j2}(u|z_0)}{\rho_{j0}(u|z_0)} - \Big(\frac{\rho_{j1}(u|z_0)}{\rho_{j0}(u|z_0)}\Big)^{\otimes 2}\Big]\rho_{j0}(u|z_0)\, d\Lambda_{0j}(u).$$

(ii) $\frac{\partial^3 \widehat{g}'(z,\boldsymbol{\beta})}{\partial\beta_j\partial\beta_k\partial\beta_\ell} = O_p(1)$, *for* $\boldsymbol{\beta} \in \mathcal{B}$.

Lemma 2 *Let* $Q(u, Z_{ij}) = c(K)\mathcal{D}_z(\log\sigma(Z_{ij})) + d(K)\mathcal{D}_z(\log\eta_{j0}(u|Z_{ij}))$. *Assume the conditions (i)-(viii) hold. If* $nh^{5/2} \to \infty$ *and* $nh^{2p} \to 0$ *for an even* p *, then*

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial \ell_p(\boldsymbol{\beta}))}{\partial\boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \quad = \quad & \frac{1}{\sqrt{n}}\sum_{j=1}^J\sum_{i=1}^n \xi_{ij}\int_0^{\tau_0}\Big\{\boldsymbol{W}_{ij}(u) + \boldsymbol{\chi}(Z_{ij}) \\ & -\frac{r_{j1}(\boldsymbol{\beta}_0,u)}{r_{j0}(\boldsymbol{\beta}_0,u)} - \sigma(Z_{ij})\boldsymbol{s}(Z_{ij})Q(u,Z_{ij})\Big\}dM_{ij}(u) + o_p(1).\end{aligned}$$

The proofs of above lemmas are tedious. Detailed proofs are provided in the technical report (Cai, Fan, Jiang, and Zhou, 2006) from the University of North Carolina at Chapel Hill.

Lemma 3 *Suppose the conditions (i)-(viii) hold. Then*

$$n^{-1}\frac{\partial^2 \ell_p(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\tau}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \xrightarrow{p} -\boldsymbol{I}(\boldsymbol{\beta}_0).$$

Proof. By (A.1), simple algebra gives that

$$n^{-1}\frac{\partial^2 \ell_p(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\tau}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = -n^{-1}\sum_{j=1}^J\sum_{i=1}^n \xi_{ij}\int_0^{\tau_0}[\frac{R_{nj2}(\boldsymbol{\beta}_0,u)}{R_{nj0}(\boldsymbol{\beta}_0,u)} - \frac{R_{nj1}^{\otimes 2}(\boldsymbol{\beta}_0,u)}{R_{nj0}^2(\boldsymbol{\beta}_0,u)}]\,dN_{ij}(u) -n^{-1}\sum_{j=1}^J\sum_{i=1}^n \xi_{ij}\int_0^{\tau_0}[\boldsymbol{\kappa}_n(Z_{ij}) - \frac{K_{nj1}(\boldsymbol{\beta}_0,u)}{K_{nj0}(\boldsymbol{\beta}_0,u)}]\,dN_{ij}(u),$$

where $K_{njm}(\boldsymbol{\beta}_0,u) = \sum_{\ell=1}^n \xi_{\ell j}\widehat{s}_{\ell j}(u,\boldsymbol{\beta}_0)(\boldsymbol{\kappa}_n(Z_{\ell j}))^m/\sum_{\ell=1}^n\xi_{\ell j}$, for $m = 0,1$. By Lemma 1 and $\widehat{g}(z,\boldsymbol{\beta}_0) = g(z) + o_p(1)$ uniformly for $z \in \cup_{j=1}^J\text{supp}[f_j(\cdot)]$,

$$n^{-1}\frac{\partial^2 \ell_p(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\tau}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = -n^{-1}\sum_{j=1}^J\sum_{i=1}^n \xi_{ij}\int_0^{\tau_0}\Big[\frac{r_{j2}(\boldsymbol{\beta}_0,u)}{r_{j0}(\boldsymbol{\beta}_0,u)} - \Big(\frac{r_{j1}(\boldsymbol{\beta}_0,u)}{r_{j0}(\boldsymbol{\beta}_0,u)}\Big)^{\otimes 2}\Big]\,dN_{ij}(u) -n^{-1}\sum_{j=1}^J\sum_{i=1}^n \xi_{ij}\int_0^{\tau_0}[\boldsymbol{\kappa}(Z_{ij}) - \frac{K^*_{nj1}(\boldsymbol{\beta}_0,u)}{K^*_{nj0}(\boldsymbol{\beta}_0,u)}]\,dN_{ij}(u) + o_p(1),$$

where $K^*_{njm}(\boldsymbol{\beta}_0,u)$ is defined similarly to $K_{njm}(\boldsymbol{\beta}_0,u)$ except $\boldsymbol{\kappa}_n(Z_{\ell j})$ and $\widehat{g}(Z_{\ell j},\boldsymbol{\beta}_0)$ replaced by $\boldsymbol{\kappa}(Z_{\ell j})$ and $g(Z_{\ell j})$. The second term above equals

$$-n^{-1}\sum_{j=1}^J\sum_{i=1}^n \xi_{ij}\int_0^{\tau_0}[\boldsymbol{\kappa}(Z_{ij}) - \frac{K^*_{nj1}(\boldsymbol{\beta}_0,u)}{K^*_{nj0}(\boldsymbol{\beta}_0,u)}]\,dM_{ij}(u) = o_p(1).$$

Therefore

$$\begin{aligned} n^{-1}\frac{\partial^2 \ell_p(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\tau}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} &= -\sum_{j=1}^J p_j\int_0^{\tau_0}\Big[\frac{r_{j2}(\boldsymbol{\beta}_0,u)}{r_{j0}(\boldsymbol{\beta}_0,u)} - \Big(\frac{r_{j1}(\boldsymbol{\beta}_0,u)}{r_{j0}(\boldsymbol{\beta}_0,u)}\Big)^{\otimes 2}\Big] \\ &\quad \times r_{j0}(\boldsymbol{\beta}_0,u)\,d\Lambda_{j0}(\boldsymbol{\beta}_0,u) + o_p(1) \\ &\equiv -\boldsymbol{I}(\boldsymbol{\beta}_0) + o_p(1). \end{aligned}$$

Proof of Theorem 1. By Lemma 2,

$$n^{-1}\partial\ell_p(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \xrightarrow{P} 0.$$

Thus, with probability tending to one, for any small given $\varepsilon > 0$, if $\boldsymbol{\beta} \in S_\varepsilon \equiv \{\boldsymbol{\beta} : ||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \le \varepsilon\}$,

$$\Big|(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^\tau\Big[n^{-1}\frac{\partial\ell_p(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\Big]\Big| \le \varepsilon^3. \tag{A.14}$$

Let a be the minimum eigenvalue of positive definitive matrix $I(\boldsymbol{\beta}_0)$. By Lemma 3, we conclude that for all $\boldsymbol{\beta} \in S_\varepsilon$

$$(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^\tau\Big[n^{-1}\frac{\partial^2\ell_p(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\tau}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\Big](\boldsymbol{\beta}-\boldsymbol{\beta}_0) \le -a\varepsilon^2, \tag{A.15}$$

with probability tending to one. By the argument right after (A.3), with probability tending to one that there is a constant $C > 0$ such that

$$|n^{-1}R_n(\boldsymbol{\beta})| \leq C\varepsilon^3. \tag{A.16}$$

Then substituting (A.14)-(A.16) into (A.2), we conclude with probability tending to one that, when ε is small enough,

$$n^{-1}\ell_p(\boldsymbol{\beta}) - n^{-1}\ell_p(\boldsymbol{\beta}_0) \leq 0. \tag{A.17}$$

Therefore, $\ell_p(\boldsymbol{\beta})$ has a local maximum in the interior of S_ε, and with probability tending to one, there exists a consistent estimator sequence $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}_0$ which maximizes the global profile pseudo-partial likelihood $\ell_p(\boldsymbol{\beta})$.

Proof of Theorem 2. Lemma 3 entails $n^{-1}\frac{\partial^2\ell_p(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\tau}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \xrightarrow{p} -\boldsymbol{I}(\boldsymbol{\beta}_0)$. Note that $\widehat{\boldsymbol{\beta}}$ is consistent. Plugging the above expression into (A.2), we obtain

$$\begin{aligned}\ell_p(\widehat{\boldsymbol{\beta}}) &= \ell_p(\boldsymbol{\beta}_0) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\tau \frac{\partial\ell_p(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \\ &\quad -\frac{n}{2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\tau\boldsymbol{I}(\boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p\Big\{(\sqrt{n}||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|| + 1)^2\Big\},\end{aligned} \tag{A.18}$$

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

$$\begin{aligned}\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &= \boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}\frac{1}{\sqrt{n}}\frac{\partial\ell_p(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} + o_p(1) \\ &= \boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{J}\sum_{i=1}^{n}\xi_{ij}\int_0^{\tau_0}\Big\{\boldsymbol{W}_{ij}(u) + \boldsymbol{\chi}(Z_{ij}) - \frac{r_{j1}(\boldsymbol{\beta}_0, u)}{r_{j0}(\boldsymbol{\beta}_0, u)} \\ &\quad -\sigma(Z_{ij})\boldsymbol{s}(Z_{ij})Q(u, Z_{ij})\Big\}\, dM_{ij}(u) + o_p(1 + \sqrt{n}||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||).\end{aligned}$$

Then by martingale central limit theorem and Slutsky Theorem,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)\boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}).$$

References

[1] Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

[2] 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.

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

[4] Cai, J., Fan, J., Jiang, J., and Zhou, H. (2006). Partially linear hazard regression for multivariate survival data. Institute of Statistics, Mimeo Series No. 2235, University of North Carolina at Chapel Hill.

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

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

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

[8] 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

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

[10] 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.

[11] Cuzick, J. (1992). Semiparametric additive regression, *Journal Royal Statistical Society B*, **54**, 831–843.

[12] Dawber, T. R. (1980). The Framingham Study, The Epidemilogy of Atherosclerotic Disease. Cambridge, MA: Harvard University Press.

[13] Fan, J. and Chen, J. (2000). One-step local quasi-likelihood estimation. *Journal Royal Statistical Society B*, **61**, 927-943.

[14] Fan, J. and Gijbels, I. (1996). Local Polynomial Modelling and its Applications. London: Chapman and Hall.

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

[16] Fan, J. and Jiang, J. (1999). Variable bandwidth and one-step local M-estimator. *Science in China, (Series A)*, **29**, 1-15.

[17] 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.

[18] Fan, J. and Yao, Q. (2003). Nonlinear Time Series: Nonparametric and Parametric Methods. New York: Springer.

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

[20] Gentleman, R. and Crowley, J. (1991). Local full likelihood estimation for the proportional hazards model, *Biometrics*, **47**, 1283 – 1296.

[21] Härdle, W., Liang, H. and Gao, J. (2004). *Partially linear models*, Springer-Verlag: Heidelberg.

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

[23] Horowitz, J. L. and Mammen, E. (2004). Nonparametric estimation of an additive model with a link function. *Annals of Statistics*, **32**, 2412-2443.

[24] Hougaard, P. (2000). *Analysis of Multivariate Survival Data.* New York: Springer.

[25] Huang, J. (1999). Efficient estimation of the partly linear additive Cox model, *Annals of Statistics*, **27**, 1536-1563.

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

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

[28] Liang, H., Härdle, W. and Carroll, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model, *Annals of Statistics*, **27**, 1519–1535.

[29] 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.

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

[31] Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data, *Biomtrka*, **88**, 1179-1185.

[32] 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, *Survival Analysis: State of the Art. J. P. Klein and P. K. Goel (eds.)*, Kluwer Academic Publishers, 237-247.

[33] Masry, E. and Fan, J. (1997). Local Polynomial Estimation of Regression functions for mixing processes. *Scandinavian Journal of Statistics*, **24**, 165-179.

[34] Murphy, S. A. and van der Vaart, A. W. (2000). On Profile Likelihood (with discussion), *Journal of American Statistical Association*, **95**, 449-485.

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

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

[37] Sandler, D. P., Weinberg, C. R., Archer, V. E., Rothney-Kozlak, L., Bishop, M., and Stolwijk, J. (1999). Indoor radon and lung caner risk: a case-control study in Connecticut and Utah, *Radiation Research*, **151**, 103-104.

[38] Speckman, P. (1988). Kernel smoothing in partial linear models, *Journal Royal Statistical Society B*, **50**, 413–436.

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

[40] Wahba, G. (1984). Partial spline models for semiparametric estimation of functions of several variables. In *Statistical Analysis of Time Series*, Proceedings of

the Japan U.S. Joint Seminar, Tokyo, 319–329. Institute of Statistical Mathematics, Tokyo.

[41] 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.