Partially Linear Hazard Regression for Multivariate Survival Data

Jianwen CAI, Jianqing FAN, Jiancheng JIANG, and Haibo ZHOU

This article 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 the linear part is accomplished by maximization of a pseudo–partial likelihood profiled over the nonparametric part. This enables us 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 through 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 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.

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 epidemiologic cohort studies, in which the ages of disease occurrence are recorded for members of families; animal experiments, in which 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 in which 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; Clegg, Cai, and Sen 1999).

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, e.g., Fleming and Harrington 1991; Andersen, Borgan, Gill, and Keiding 1993). These assumptions are chosen mainly for their mathematical convenience, however. 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) in 2,336 men and 2,873 women in the well-known Framingham Heart Study (Dawber 1980), the investigators are interested in identifying a nonlinear cholesterol effect. A nonparametric method is desired for providing a continuous trend of the cholesterol effect that is sufficiently flexible to indicate local changes in this trend. Nonparametric modeling of such a trend has less restriction than the parametric approach and thus 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 method for attenuating this difficulty is to model the covariate effects through 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.,

Jianwen Cai is Professor, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 (E-mail: *cai@bios.unc.edu*). Jianqing Fan is Frederick L. Moore Professor of Finance, Department of Operation Research and Financial Engineering, Princeton University, Princeton, NJ 08544 (E-mail: *jqfan@princeton.edu*). Jiancheng Jiang is Assistant Professor, Department of Mathematics and Statistics, University of North Carolina at Charlotte, NC 28223 (E-mail: *jjiang1@uncc.edu*) and Associate Research Scholar, Department of Operation Research and Financial Engineering, Princeton University, Princeton, NJ 08544 (E-mail: *jjiang@princeton.edu*). Haibo Zhou is Associate Professor, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 (E-mail: *zhou@bios.unc.edu*). This research was supported in part by National Institutes of Health (NIH) grant R01 HL69720. Additional support was also provided by National Science Foundation grant DMS-03-54223 and NIH grant R01 GM072611 for Fan, National Science Foundation of China grant 10471006 for Jiang, and NIH grant R01 CA 79949 for Zhou. The authors gratefully acknowledge the constructive suggestions and comments from the associate editor and two anonymous referees that greatly improved this article.

Journal of the American Statistical Association
June 2007, Vol. 102, No. 478, Theory and Methods
DOI 10.1198/016214506000001374

Wahba 1984; Speckman 1988; Cuzick 1992; Carroll, Fan, Gijbels, and Wand 1997; Lin and Carroll 2001; Liang, Härdle, and Carroll 1999). An overview of the partially linear model has been given by 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 models have been studied by Huang (1999) 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 article, no formal methodology to address nonlinear covariate effects has been elaborated in the literature. In this article 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 hazard regression function.

We consider two general setups in which 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 prespecified number based on the goal of the study, and it does not vary across subjects. In the other setup, we assume that there are n clusters and that in each cluster there are J different types of members, for example, father and sons in a family. Because not all members are necessarily available in a cluster, the cluster size can vary. 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 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 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

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

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

To model partly nonlinear covariates effects, we assume the model

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

where $Z_{ij}(\cdot)$ is a main exposure variable of interest whose effect on the logarithm of the hazard might be nonlinear, $\mathbf{W}_{ij}(\cdot) = (W_{ij1}(\cdot), \ldots, W_{ijq}(\cdot))^{\tau}$ is a vector of covariates that have linear effect, $Y_{ij}(\cdot)$ is an at-risk indicator process [i.e., $Y_{ij}(t) = 1(X_{ij} \ge t)$], $\lambda_{0j}(\cdot)$ 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 failure. A related class of marginal model is given by restricting the baseline hazard functions in (1) to be common for all of the failure types within a subject, that is,

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

Although 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 of the failure types within a subject. Model (2) is very useful for modeling clustered failure time data in which subjects within clusters are exchangeable.

In this article we focus on statistical inference for model (1). We propose a profile pseudo–partial likelihood estimation to estimate $\boldsymbol{\beta}$. We show that the proposed estimator of $\boldsymbol{\beta}$ is root-n consistent. The asymptotic 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 must be estimated from its derivative. Thus the estimator of nonparametric function $g(\cdot)$ uses all observed information, and the score function of $\boldsymbol{\beta}$ cannot 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. Furthermore, the cost due to estimation of the nonparametric component g in the Cox model is very different than 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 of Huang (1999), the asymptotic covariance matrix of our estimators for the parametric part admits a sandwich formula that furnishes a consistent covariance matrix estimator using the plug-in method, whereas Huang's estimator achieves the semiparametric information bound but lacks a consistent covariance estimator; on the other hand, our estimator for the nonparametric part not only is optimal in convergence rate, but also has asymptotic normality, which is unavailable for Huang's estimation for the nonparametric part.

This article 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 through a real data analysis. We give proofs of the theorems in the Appendix.

2. MAXIMUM PSEUDO–PARTIAL LIKELIHOOD ESTIMATION

Let $\mathcal{R}_j(t) = \{i : X_{ij} \ge t\}$ denote the set of subjects at risk just before 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

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

where

$$R_{ij}(\boldsymbol{\beta}, g) = \log\Bigg(\sum_{l\in\mathcal{R}_j(X_{ij})} \xi_{lj}\exp\big[\boldsymbol{\beta}^{\tau}\mathbf{W}_{lj}(X_{ij}) + g(Z_{lj}(X_{ij}))\big]\Bigg).$$

Because failure times from the same subject are dependent, the foregoing function is referred to as pseudo–partial likelihood. We use this pseudo–partial likelihood for our estimation. However, we do not require that the failure times be independent or specify a dependence structure among failure times. This furnishes robustness of our estimation method against the misspecification 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 sec. 3.1 of Fan et al. 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,

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

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

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

where

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

$\tilde{Z}_{ij}(u) = \{Z_{ij}(u) - z_0, \ldots, (Z_{ij}(u) - z_0)^p\}^{\tau}$, $K_h(\cdot) = K(\cdot/h)/h$, 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 reflect the fact that the local model (4) is applied only to the data around z_0. It gives greater weight to the data closer to the point z_0. For the univariate case, the local pseudo–partial likelihood was derived by Fan et al. (1997) from a local maximum likelihood standpoint.

Let $(\hat{\boldsymbol{\beta}}(z_0), \hat{\boldsymbol{\gamma}}(z_0))$ maximize the local pseudo–partial likelihood (5). Then an estimator of $g'(\cdot)$ at the point z_0 is simply the first component of $\hat{\boldsymbol{\gamma}}(z_0)$, namely $\hat{g}'(z_0) = \hat{\gamma}_1(z_0)$. The curve $\hat{g}$ can be estimated by integration on the function $\hat{g}'(z_0)$ using the method of Hastie and Tibshirani (1990). To ensure 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) showed 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. Because only the local data are used in the estimation of $\boldsymbol{\beta}$, the resulting estimator for $\boldsymbol{\beta}$ cannot be root-n consistent. We call $(\hat{\boldsymbol{\beta}}(z_0), \hat{\boldsymbol{\gamma}}(z_0))$ 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 $\hat{g}^{(k)}(\cdot, \boldsymbol{\beta})$ of $g^{(k)}(\cdot)$, and hence $\hat{g}(\cdot, \boldsymbol{\beta})$, by maximizing (5) with respect to $\boldsymbol{\gamma}$. Denote by $\hat{\boldsymbol{\gamma}}(z_0, \boldsymbol{\beta})$ the maximizer. Substituting the estimator $\hat{g}(\cdot, \boldsymbol{\beta})$ into (3), we can obtain the logarithm of the profile pseudo–partial likelihood,

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

Here and hereinafter, for ease of presentation, we sometimes drop the dependence of covariates on time, with the understanding that the methods developed in this article are applicable to external time-dependent covariates (Kalbfleisch and Prentice 2002). Let $\hat{\boldsymbol{\beta}}$ maximize (6) and $\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\gamma}}(z_0, \hat{\boldsymbol{\beta}})$. Our proposed estimator for the parametric component is simply $\hat{\boldsymbol{\beta}}$, and that for the nonparametric component is $\hat{g}(\cdot) = \hat{g}(\cdot, \hat{\boldsymbol{\beta}})$.

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 defined implicitly. Let z_j $(j = 1, \ldots, n_g)$ 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 $\bar{\boldsymbol{\beta}} = n_g^{-1}\sum_{j=1}^{n_g}\hat{\boldsymbol{\beta}}(z_j)$ as the initial value. Set $\hat{\boldsymbol{\beta}} = \bar{\boldsymbol{\beta}}$.
2. *Estimation of the nonparametric component.* Maximize the local pseudo–partial likelihood $\ell(\hat{\beta}, \boldsymbol{\gamma})$ at each grid point z_j and obtain the nonparametric estimator $\hat{g}(\cdot, \hat{\boldsymbol{\beta}})$ at these grid points. Obtain the nonparametric estimator at points $\{Z_{ij}\}$ using the linear interpolation. We take the bandwidth h suitable for estimating $\boldsymbol{\beta}$. One example for such a suitable bandwidth is the ad hoc bandwidth in (7).
3. *Estimation of the parametric component.* With the estimator $\hat{g}(\cdot, \hat{\boldsymbol{\beta}})$, maximize the profile estimator $\ell_p(\boldsymbol{\beta})$ with $g(\cdot, \boldsymbol{\beta}) = \hat{g}(\cdot, \hat{\boldsymbol{\beta}})$, using the Newton–Raphson algorithm and the initial value $\hat{\boldsymbol{\beta}}$ from the previous step.
4. *Iteration.* Iterate between steps 2 and 3 until convergence.
5. *Reestimation of the nonparametric component.* Fix $\boldsymbol{\beta}$ at its estimated value from step 4. The final estimate of $\hat{g}(\cdot)$ is $\hat{g}(\cdot, \hat{\boldsymbol{\beta}})$. At this final step, take the bandwidth h suitable for estimating $g(\cdot)$, such as the estimated optimal bandwidth $\hat{h}_{opt}$ based on (10) in Section 3.

Because the initial estimator $\bar{\boldsymbol{\beta}}$ is consistent, we do not expect many iterations in step 4. Because 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 Chen (1999) and Fan and Jiang (1999) in nonparametric models. In fact, according to Robinson (1988), if an initial parametric estimator has rate $O(n^{-a})$, then 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 foregoing theory.

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

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

where $\hat{h}_{opt}$ is the estimated optimal bandwidth for $g'(\cdot)$ based on (10). In 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 using an even p to avoid boundary effects in estimation of $g(\cdot)$.

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

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

where $Y_{ij}(u) = 1(X_{ij} \ge u)$ is the at-risk indicator and $N_{ij}(u) = 1(X_{ij} \le u, \Delta_{ij} = 1)$ 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 $\hat{g}(\cdot, \boldsymbol{\beta})$ that uses all observed information. 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 we relegate to Appendix A for ease of exposition. The following theorems demonstrate that our estimators are consistent and asymptotically normal.

Theorem 1. Under conditions (a)–(h) in Appendix A, with probability tending to 1, there exists an estimator, $\hat{\boldsymbol{\beta}}$, which maximizes the profile pseudo–partial likelihood $\ell_p(\boldsymbol{\beta})$, such that $\hat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}_0$.

Theorem 2. Under conditions (a)–(h) in Appendix A, 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 $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges to a Gaussian distribution with mean 0 and covariance matrix $\boldsymbol{\Omega} = \mathbf{I}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)\mathbf{I}(\boldsymbol{\beta}_0)^{-1}$.

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

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

Note that $\hat{\mathbf{I}}$ and $\hat{\boldsymbol{\Sigma}}$ can be shown to be consistent for $\mathbf{I}(\boldsymbol{\beta}_0)$ and $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)$. Thus $\hat{\boldsymbol{\Omega}}$ is a consistent estimator of $\boldsymbol{\Omega}$ under the conditions of Theorem 2. Then for the semiparametric testing problem

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

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

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

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 $\boldsymbol{\beta}$.

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

Theorem 4. Assume that conditions (a)–(e) hold. If $\hat{\boldsymbol{\beta}}$ is $\sqrt{n}$-consistent and nh^{2p+3} is bounded, then

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

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

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

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, we 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 insight into this phenomena, see 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

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

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

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

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

is given by

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

Using the foregoing formula and the widely used plug-in technique or the preasymptotic substitution method (see, e.g., Fan and Yao 2003, p. 245), we can develop a data-driven approach to the selection of the bandwidth h, but this is beyond 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 the approach of Huang (1999) by simulations. First, we consider the case in which 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 are generated from a multivariate extension of the model of Clayton and Cuzick (1985) in which the joint survival function for $(T_1, \dots, T_J)$ given $(Z_1, \dots, Z_J)$ and $(\mathbf{W}_1, \dots, \mathbf{W}_J)$ is

$$S(t_1, \dots, t_J; Z_1, \dots, Z_J, \mathbf{W}_1, \dots, \mathbf{W}_J) = \left\{ \sum_{j=1}^{J} S_j(t_j)^{-1/\theta} - (J-1) \right\}^{-\theta}, \quad (11)$$

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

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

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

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

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 $\boldsymbol{\beta} = (.6, .4)^\tau$. We first simulate $Z_{ij} \overset{\text{iid}}{\sim} \mathrm{U}(0, 1)$, and $\mathbf{W}_{ij} = (W_{ij}^{(1)}, W_{ij}^{(2)})^\tau$ with $W_{ij}^{(1)}$ independently generated from a binomial distribution (taking 1 or 0 each with probability .5) and $(W_{i1}^{(2)}, W_{i2}^{(2)})$ from a bivariate normal distribution with the correlation coefficient .5 and the marginal distributions $\mathrm{N}(0, 1)$. For given (Z_{i1}, Z_{i2}) and $(\mathbf{W}_{i1}, \mathbf{W}_{i2})$, we generated (T_{i1}, T_{i2}) using the algorithm of Cai and Shen (2000, p. 2967). The censoring time distribution was generated from exponential distribution with mean chosen to produce a certain amount of censoring.

We used $p = 2$ and used the Epanechnikov kernel function for smoothing. The parameter θ was set as 100 and .01, which correspond to weak and strong correlation within each subject. By the argument in Section 2, the bandwidth h was taken as $.3n^{-1/3}$ for estimation of $\boldsymbol{\beta}$ in steps 2–3, and $.3n^{-1/7}$ for estimation of $g'(\cdot)$ in step 5 of the algorithm. We assess the sensitivity of the estimation methods as the bandwidth changes over a large range by using half of and double the foregoing bandwidths.

The estimators and their standard deviations (SDs) for the parameters were evaluated along with the average of the estimated standard error ($\hat{se}$) for the estimators. The coverage rate (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 $\hat{\boldsymbol{\beta}}$ (and hence the SE of its elements) is to use $\hat{\mathbf{I}}^{-1}$. The corresponding coverage rate (CP_{na}) of the 95% confidence intervals based only on $\hat{\mathbf{I}}^{-1}$ is also computed. We include the naive method here for comparison.

Tables 1 and 2 report the simulation results for the settings with no censoring and 40% censoring. It is evident that the proposed estimation performs well because the bias is small, the estimated standard error is close to the sample standard deviation, and the coverage rate of the constructed intervals is close to the nominal level. The naive method fails when there is nonignorable correlation between survival times (small θ). This is evidenced by the fact that the estimated SEs [column "Mean($\hat{\mathbf{I}}^{-1}$)"] 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 censoring percentage increases and gets smaller when the sample size increases.

We now report the performance of the estimated functions. The typical estimated functions with performance at 10th, 50th

Table 1. Summary of Simulation Results ($\beta_1 = .6$ and $\beta_2 = .4$)

Size	Model		No censoring					
(n, J)	θ	β	Mean($\hat{\beta}$)	SD($\hat{\beta}$)	Mean($\hat{se}$)	Mean($\hat{\mathbf{I}}^{-1}$)	95% CP_{se}	95% CP_{na}
(100, 2)	.01	β_1	.6137	.2107	.1987	.1515	.938	.848
		β_2	.4146	.1015	.0890	.0782	.914	.872
(100, 2)	100	β_1	.5959	.1558	.1460	.1512	.948	.956
		β_2	.4052	.0818	.0753	.0782	.920	.938
(200, 2)	.01	β_1	.6151	.1444	.1408	.1050	.952	.832
		β_2	.4050	.0706	.0633	.0537	.930	.876
(200, 2)	100	β_1	.6034	.1068	.1028	.1047	.938	.936
		β_2	.4011	.0559	.0531	.0537	.932	.940

Table 2. Summary of Simulation Results ($\beta_1 = .6$ and $\beta_2 = .4$)

Size (n, J)	Model θ	$\boldsymbol{\beta}$	40% censoring Mean($\hat{\boldsymbol{\beta}}$)	SD($\hat{\boldsymbol{\beta}}$)	Mean($\hat{se}$)	Mean($\hat{\mathbf{I}}^{-1}$)	95% CP_{se}	95% CP_{na}
(100, 2)	.01	β_1	.6119	.2578	.2413	.1919	.934	.854
		β_2	.4175	.1233	.1200	.0994	.934	.900
(100, 2)	100	β_1	.5920	.2041	.2092	.1918	.948	.934
		β_2	.4013	.1075	.1109	.0993	.938	.934
(200, 2)	.01	β_1	.6138	.1694	.1708	.1340	.952	.890
		β_2	.4043	.0844	.0866	.0685	.958	.908
(200, 2)	100	β_1	.6121	.1304	.1267	.1337	.940	.960
		β_2	.4015	.0706	.0702	.0684	.956	.934

(median), and 90th percentiles of the mean integrated squared errors (MISEs) among the 500 simulations are presented to assess the quality of estimated functions. We presented only the case where $n = 200$ in Figure 1 to save space. It is seen that the typical estimated curves in Figure 1 capture the form of the true curve well, reflecting 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 indicates the variability of the estimated functions at each point. Again, to save space, we present only 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.

Comparing the estimated curves in Figures 1 and 2 for $\theta = 100$ and .01, which correspond to weak and strong correlations within each subject, shows that the estimators of the nonparametric part do not depend heavily on the correlation. This exemplifies our statement in Remark 2.

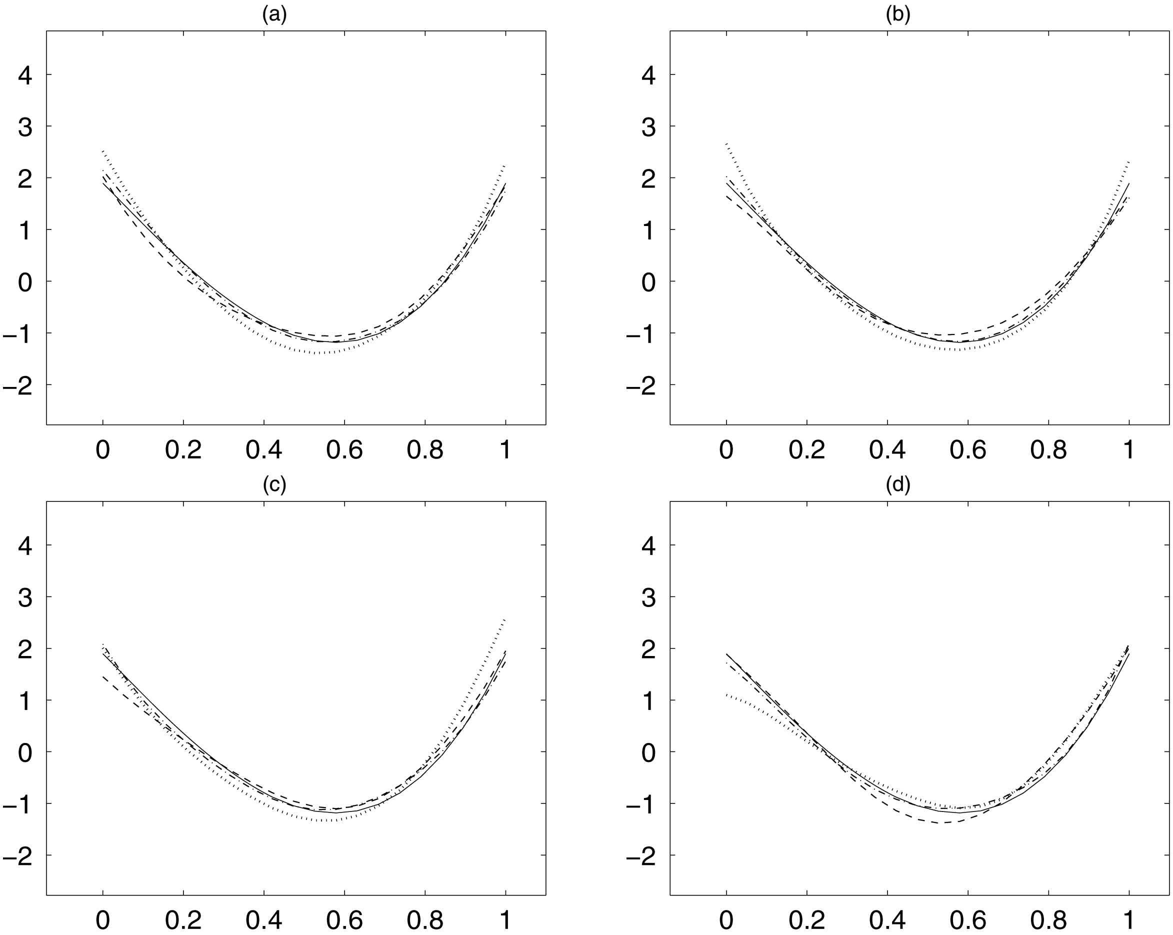

Figure 1. Typical Estimated Curves in Terms of Percentiles of MISEs Among the 500 Simulations With $n = 200$ and $J = 2$. (a) $\theta = .01$, no censoring; (b) $\theta = 100$, no censoring; (c) $\theta = .01$, 40% censoring; (d) $\theta = 100$, 40% censoring. (——, true curve; – · –·, the 10th percentile; – – –, the 50th percentile; · · · ·, the 90th percentile.)

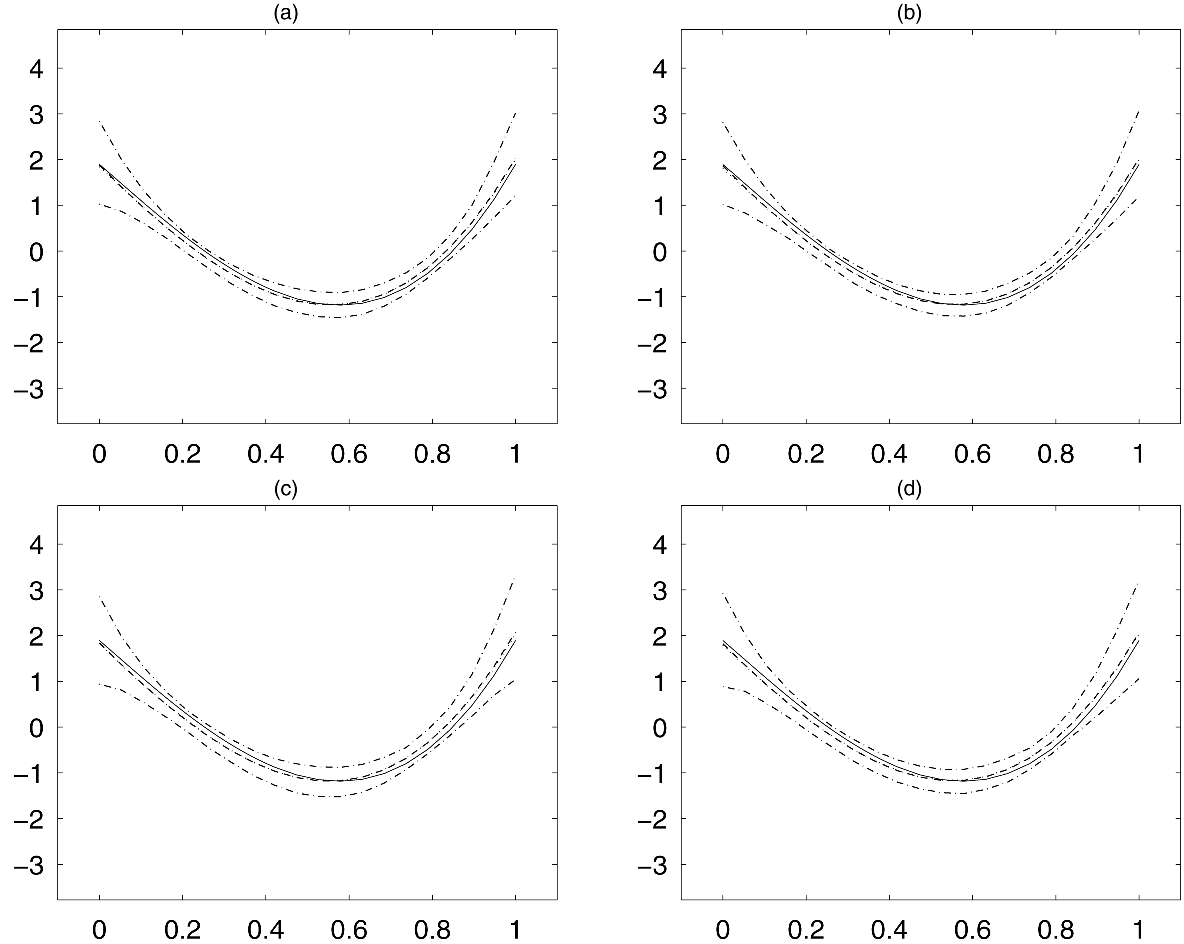

Figure 2. Sampling Variability of Estimated Functions at Each Point Based on the 500 Simulations With $n = 100$ and $J = 2$. (a) $\theta = .01$, no censoring; (b) $\theta = 100$, no censoring; (c) $\theta = .01$, 40% censoring; (d) $\theta = 100$, 40% censoring. (– – –, mean; · · · ·, median; – · – ·, 95% envelopes formed by the 2.5th and 97.5th percentiles in 500 simulations; ——, true curve.)

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

To assess the performance of the proposed method under the 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 $j = 1, \dots, 6$. For this setup, the maximum cluster size is 6. For cluster i ($i = 1, \dots, n$), J_i correlated failure times are generated from model (11). We consider $n = 100$ and $\lambda_{0j} = j^2$, and set the true parameter $\boldsymbol{\beta}$ 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. Similar conclusions as before can be drawn for the varying cluster size example.

Finally, we compare the proposed profile pseudo–partial likelihood method with the efficient estimation method of Huang (1999). Because Huang's method deals with univariate failure time data, we focus on model (11) with $J \equiv 1$. We again set the true parameter $\boldsymbol{\beta}$ 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 as 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 variance (results not shown). From the table, we see that both estimators perform similarly.

4.2 Applications to the Framingham Heart Study Dataset

In this section we apply our proposed procedure to analyze data from the well-known Framingham Heart Study (Dawber 1980), which began in 1948. The cohort consists of 2,336 men and 2,873 women. The participants were between age 30 and 62 years at the first examination, and they were recalled and examined every 2 years after 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 underwent examination at age 44 or 45 and were disease-free at that examination, in the sense that there was no history of hypertension or glucose intolerance and no previous CHD or CVA. There was a total of 1,571 disease-free subjects. The percentage of censoring was about 90.42%. The

Table 3. Summary of Simulation Results Under the Varying Cluster Size Situation ($\beta_1 = .6$ and $\beta_2 = .4$)

Size	Model		57% censoring					
(n, J)	θ	$\boldsymbol{\beta}$	Mean($\hat{\boldsymbol{\beta}}$)	SD($\hat{\boldsymbol{\beta}}$)	Mean($\hat{se}$)	Mean($\hat{\mathbf{I}}^{-1}$)	95% CP_{se}	95% CP_{na}
(100, 1–6)	.01	β_1	.5881	.1484	.1405	.1286	.918	.908
		β_2	.4062	.0775	.0727	.0647	.920	.908
(100, 1–6)	100	β_1	.5879	.1378	.1352	.1283	.928	.928
		β_2	.3963	.0717	.0668	.0645	.918	.916

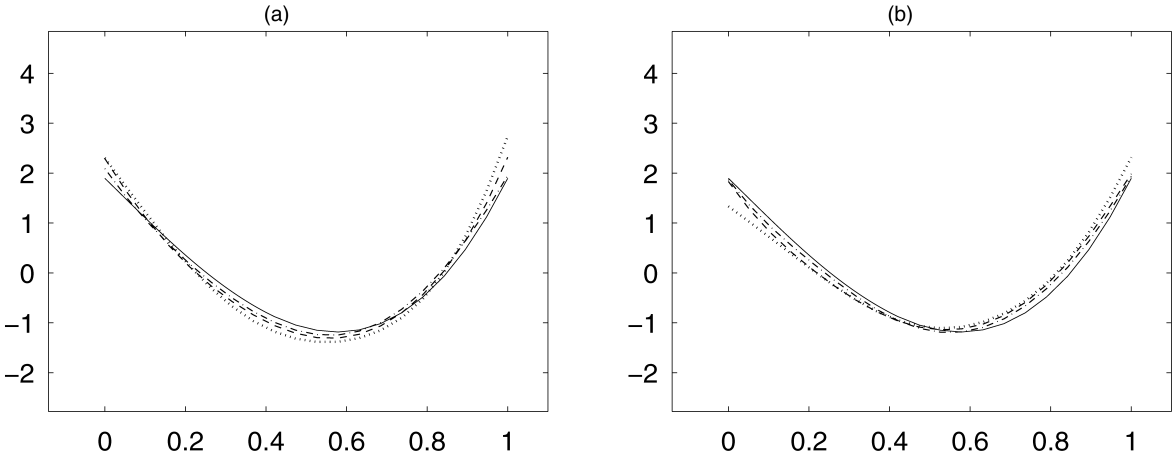

Figure 3. Typical Estimated Curves in Terms of Percentiles of MISEs Among the 500 Simulations With $n = 100$ and Varying Cluster Size. (a) $\theta = .01$, 57% censoring; (b) $\theta = 100$, 57% censoring. (——, true curve; $-\cdot-\cdot$, the 10th percentile; $- - -$, the 50th percentile; $\cdots\cdot$, the 90th percentile.)

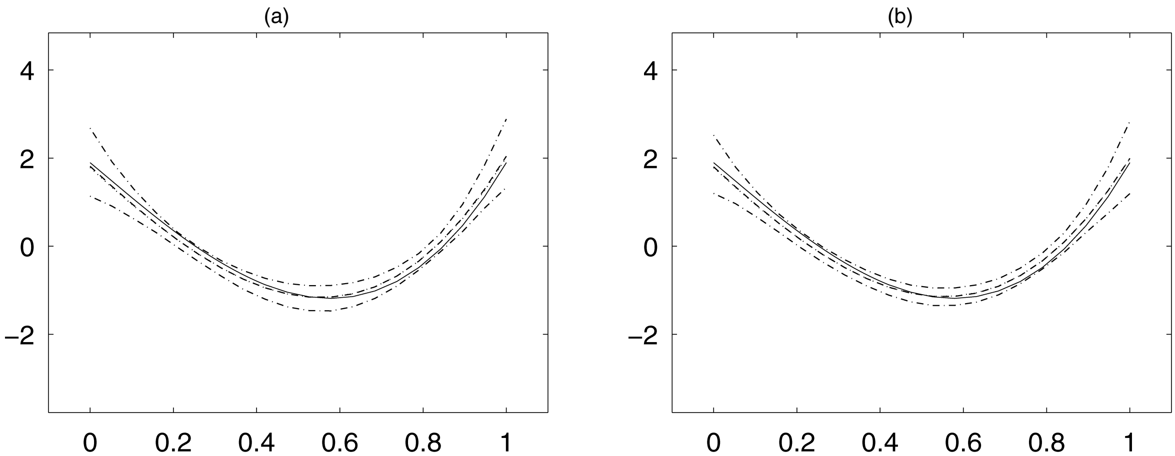

Figure 4. Sampling Variability of Estimated Functions at Each Point Based on the 500 Simulations With $n = 100$ and Varying Cluster Size. (a) $\theta = .01$, 57% censoring; (b) $\theta = 100$, 57% censoring. ($- - -$, mean; $\cdots\cdot$, median; $-\cdot-\cdot$, 95% envelopes formed by the 2.5 and 97.5th percentiles in 500 simulations; ——, true curve.)

Table 4. Comparison Between the Proposed and Huang's Estimators ($\beta_1 = .6$ and $\beta_2 = .4$)

Size (n, J)	Parameter $\boldsymbol{\beta}$	Method	Mean($\hat{\boldsymbol{\beta}}$)	SD($\hat{\boldsymbol{\beta}}$)	Mean($\hat{se}$)	95% CP_{se}
(100, 1)	β_1	PPL	.6144	.2256	.2093	.934
		HS	.6292	.2280	*	*
	β_2	PPL	.4016	.1153	.1081	.924
		HS	.4132	.1174	*	*
(200, 1)	β_1	PPL	.6021	.1526	.1463	.944
		HS	.6136	.1537	*	*
	β_2	PPL	.4041	.0811	.0752	.930
		HS	.4124	.0821	*	*

NOTE: PPL; the proposed method; HS; Huang's global spline method.
* Not available for estimation.

Table 5. Estimated Parameters for the FHS Data

Effect	$\hat{\beta}$	$\hat{se}$	p value
Age at "age 45"	.0304	.0887	.7322
Body mass index, kg/m^2	.0371	.0137	.0065
Systolic blood pressure, mm Hg	.0171	.0044	.0001
Smoking status: yes = 1; no = 0	.3578	.1186	.0026
Sex: female = 1; male = 0	−.5730	.0993	<.0001
Waiting time, year	.0031	.0162	.8465

NOTE: $\hat{\beta}$ represents the estimated parameters; $\hat{se}$, the standard error of $\hat{\beta}$.

risk factors of interest were sex, systolic blood pressure, body mass index, cholesterol level, cigarette smoking, and 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 or 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 used 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

$$W_{ij} = (\text{Age at "age 45", Smoking, BMI, SBP, Waiting time, Sex})^\tau.$$

Table 5 reports the estimated parameters and their estimated standard errors, 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 the .01 significance 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. The nonlinear form of g is evidenced by the confidence intervals of its derivative estimator, because the derivative function is not a constant. It reveals that the effect of cholesterol is lowest around the normal levels (160 ~ 170 mg/dL) and is monotone-increasing as the cholesterol level moves out of the normal range. Because there are only six participants with cholesterol levels >360, the estimator of the nonparametric function is unreliable on the sparse data region. Figure 5 displays only the estimated functions in the region with cholesterol <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 models with linear and nonlinear risk factors in the marginal hazard regression. This article 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 demonstrates that the finite parameters can be estimated at a root-n rate, whereas 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. We illustrated the methodology with an application to the Framingham Heart Study data.

Variable selections based on the nonconcave penalized likelihood also can be developed following the framework of Fan and Li (2004). Ongoing research is focusing 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 using the marginal hazard model.

Our model (1) allows us to explore the nonlinear effect of the one-dimensional covariate Z, but the proposed methodology can be extended in at least two directions when Z is a multivariate covariate vector. One direction is to use the partly linear additive structure, as was done by Huang (1999), and estimate the nonparametric part through 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 direction is to use the partially linear single-index structure of Carroll et al. (1997), as pointed out by a referee. For the first direction,

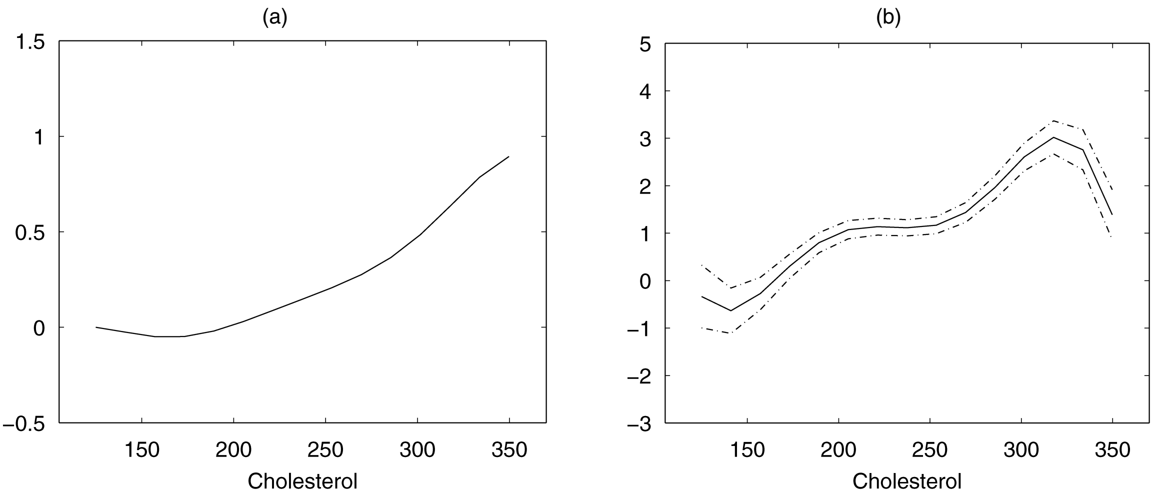

Figure 5. Estimated Function g (a) and Its Derivative (b) With 95% Confidence Intervals. (——, estimated curve; – · – ·, 95% pointwise confidence intervals.)

we anticipate that the results will be similar to ours with an oracle property in the sense that one nonparametric component can be estimated as if the others were known. For the second direction, our methodology in this article continuously applies, but with many more techniques involved.

APPENDIX A: NOTATIONS AND ASSUMPTIONS

Let $(\Omega, F, \mathcal{P})$ be a family of complete probability spaces 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 $n_j(u) = P(X_{1j} \le u, \Delta_{1j} = 1)$. Let $\mathcal{F}_{t,ij} = \sigma\{I(X_{ij} < u, \Delta_{ij} = 1), I(X_{ij}(u) < u, \Delta_{ij} = 0), \mathbf{W}_{ij}(u), Z_{ij}(u), Y_{ij}(u^-), 0 \le u \le t\}$ be information received up to time t for each (i,j), and let $M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(u)\lambda_{ij}(u)\,du$, for $i = 1, \dots, n, j = 1, \dots, J$. Assume that $N_{ij}(t)$ is $\mathcal{F}$-adapted and that the observation period is $[0, \tau_0]$, where τ_0 is the study end time. Then $M_{ij}(t)$ is a martingale with respect to the marginal filtration $\mathcal{F}_{t,ij}$ and the σ-field generated by $\bigcup_{i=1}^n \mathcal{F}_{t,ij}$, 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 developed similarly. The following notations and conditions are needed for the proofs of our theoretical results. For any vector $\mathbf{a}$, define $\mathbf{a}^{\otimes k} = 1, \mathbf{a}, \text{and } \mathbf{a}\mathbf{a}^\tau$ for $k = 0, 1, 2$, where $\mathbf{a}^\tau$ denotes the transpose of $\mathbf{a}$. Let $\tilde{\mathbf{u}} = (u, \dots, u^p)^\tau$, $\boldsymbol{\nu}_k = \int \tilde{\mathbf{u}}^{\otimes k} K(u)\,du$, and $\boldsymbol{\nu}_k^* = \int u\tilde{\mathbf{u}}^{\otimes k} K(u)\,du$ for $k = 0, 1, 2$. Put $K_1^*(t) = tK(t)/\int u^2 K(u)\,du$, $\mathbf{A} = \boldsymbol{\nu}_2 - \boldsymbol{\nu}_1^{\otimes 2}$, $\mathbf{B} = \int K^2(u)(\tilde{\mathbf{u}} - \boldsymbol{\nu}_1)^{\otimes 2}\,du$, and $\mathbf{b}_k = \int u^k(\tilde{\mathbf{u}} - \boldsymbol{\nu}_1)K(u)\,du$ for $k = 1$, $p+1$, and $p+2$. Write $c(K) = \mathbf{e}_1^\tau \mathbf{A}^{-1}\mathbf{b}_1$ and $d(K) = \mathbf{e}_1^\tau \mathbf{A}^{-1}(\boldsymbol{\nu}_1^* - \boldsymbol{\nu}_1\boldsymbol{\nu}_0^*)$, with $\mathbf{e}_1$ as a vector with a 1 in the first position and 0's elsewhere. Assume that the following conditions hold:

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

(b) $nh \to \infty$ and $h \to 0$ as $n \to \infty$. Let $\mathbf{H} = \operatorname{diag}(h, \dots, h^p)$ and $\tilde{Z}_{ij}^* = \mathbf{H}^{-1}\tilde{Z}_{ij}$.

(c) The density $f_j(\cdot)$ of Z_{1j} is of compact support and has a bounded second derivative for $j = 1, \dots, J$, 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 $\{X_{ij}, \Delta_{ij}, \mathbf{W}_{ij}, Z_{ij}\}$. Let $p_j = P(\xi_{1j} = 1)$, $j = 1, \dots, J$.

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

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

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

are equicontinuous in z, for $j = 1, \dots, J$ and $k = 0, 1$, where $s_{ij}(u, \boldsymbol{\beta}_0) = Y_{ij}(u)\exp(\boldsymbol{\beta}_0^\tau \mathbf{W}_{ij}(u) + g(Z_{ij}))$ 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 $z \in \bigcup_{j=1}^J \operatorname{supp}(f_j)$. 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 $\bigcup_{j=1}^J \operatorname{supp}(f_j)$.

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

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

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

(h) 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 $r_{jk}(\boldsymbol{\beta}, u) = E\{s_{1j}(u, \boldsymbol{\beta})(\mathbf{W}_{1j}(u) + \boldsymbol{\chi}(Z_{1j}))^{\otimes k}\}$ 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 0 on $\mathcal{B} \times [0, \tau_0]$; and r_{j1} and r_{j2} are bounded on $\mathcal{B} \times [0, \tau_0]$. The matrix $\mathbf{I}(\boldsymbol{\beta}_0)$ is finite positive definite, where

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

These conditions are similar to those of Andersen and Gill (1982) and Fan et al. (1997). Conditions (a)–(e) are standard for nonparametric component estimation using local partial likelihood. Conditions (f)–(h) guarantee the local asymptotic quadratic properties for the partial likelihood function, and thus the asymptotic normality of the estimators (see Andersen and Gill 1982; Murphy and van der Vaart 2000 for details).

Write

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

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

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

$$G_{ij}(\boldsymbol{\beta}) = \int_0^{\tau_0} H_{ij}(u)\,dM_{ij}(u),$$

and

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}) = E\left\{\sum_{j=1}^J \xi_{1j}G_{1j}(\boldsymbol{\beta})\right\}^{\otimes 2},$$

where

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

and

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

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

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

where $\hat{s}_{ij}(u) = Y_{ij}(u)\exp(\hat{\boldsymbol{\beta}}^\tau \mathbf{W}_{ij}(u) + \hat{g}(Z_{ij}))$ is the estimated risk corresponding to $s_{ij}(u, \boldsymbol{\beta}_0)$ and $\hat{E}(\cdot|\cdot)$ denotes a consistent estimator of

$E(\cdot|\cdot)$, such as the Nadaraya–Watson estimator or the local linear estimator in nonparametric regression. Put $\hat{\alpha}(z) = \int_0^{\tau_0} \mathcal{D}_z[\hat{\rho}_{j1}(u|z)/\hat{\rho}_{j0}(u|z)]\hat{\rho}_{j0}(u|z)\, d\hat{\Lambda}_{0j}(u)$. Then the plug-in estimator of $\boldsymbol{\chi}(z)$ is

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

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

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

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

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

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

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

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

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

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

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

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

APPENDIX B: PROOFS OF THEOREMS

These proofs involve the martingale theory, the theory of empirical processes, and the techniques commonly used in nonparametric literature. Given the identifiability condition $\hat{g}(0, \boldsymbol{\beta}) = 0$, we have that $\hat{g}(z_0, \boldsymbol{\beta}) = \int_0^{z_0} \hat{g}'(z, \boldsymbol{\beta})\, dz$. Let $\boldsymbol{\chi}_n(z_0) = \frac{\partial\hat{g}(z_0,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = \int_0^{z_0} \frac{\partial\hat{g}'(z,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\, dz$ and

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

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

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

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

$$\ell_p(\boldsymbol{\beta}) \equiv \sum_{j=1}^J\sum_{i=1}^n \xi_{ij}\int_0^{\tau_0}\left\{\boldsymbol{\beta}^\tau\mathbf{W}_{ij}(u) + \hat{g}(Z_{ij}, \boldsymbol{\beta}) - \log\left(\sum_{\ell=1}^n Y_{\ell j}(u)\exp[\boldsymbol{\beta}^\tau\mathbf{W}_{\ell j}(u) + \hat{g}(Z_{\ell j}, \boldsymbol{\beta})]\right)\right\} dN_{ij}(u). \tag{B.1}$$

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

$$\ell_p(\boldsymbol{\beta}) = \ell_p(\boldsymbol{\beta}_0) + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\tau\frac{\partial\ell_p(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\tau\frac{\partial^2\ell_p(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}^\tau}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + R_n(\boldsymbol{\beta}^*), \tag{B.2}$$

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

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

with β_j and β_{0j} the jth elements of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$. 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 thus $n^{-1}R_n(\boldsymbol{\beta}) = O_P(\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^3)$ for $\boldsymbol{\beta} \in \mathcal{B}$.

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

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

where $\tilde{s}_{\ell j}(u; \boldsymbol{\beta}, \boldsymbol{\gamma}^*) = Y_{\ell j}(u)\exp[\boldsymbol{\beta}^\tau\mathbf{W}_{\ell j}(u) + \boldsymbol{\gamma}^{*\tau}\tilde{Z}_{\ell j}^*]$. 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 $k = 0, 1$,

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

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

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

and

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

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

In what follows, we 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), $\hat{\boldsymbol{\gamma}}^* \equiv \hat{\boldsymbol{\gamma}}^*(z_0, \hat{\boldsymbol{\beta}})$ satisfies

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

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

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

Thus $\hat{\boldsymbol{\gamma}}^*$ satisfies

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

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

$$\hat{\mathbf{U}}(\boldsymbol{\gamma}^*,z_0)+\frac{\partial\hat{U}(\tilde{\boldsymbol{\gamma}}^*,z_0)}{\partial\boldsymbol{\gamma}^*}(\hat{\boldsymbol{\gamma}}^*-\boldsymbol{\gamma}^*)=o_p(1/\sqrt{nh}),\qquad\text{(B.8)}$$

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

$$-\frac{\partial\hat{\mathbf{U}}(\boldsymbol{\gamma}^*,z_0)}{\partial\boldsymbol{\gamma}^*}=n^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n}\xi_{ij}\int_0^{\tau_0}K_h(Z_{ij}-z_0)\times\frac{\Phi_{nj0}(u,\boldsymbol{\beta}_0,\boldsymbol{\gamma}^*)\Phi_{nj2}(u,\boldsymbol{\beta}_0,\boldsymbol{\gamma}^*)-\Phi_{nj1}^{\otimes2}(u,\boldsymbol{\beta}_0,\boldsymbol{\gamma}^*)}{\Phi_{nj0}^2(u,\boldsymbol{\beta}_0,\boldsymbol{\gamma}^*)}dN_{ij}(u).$$

It follows from (B.7) that

$$-\frac{\partial\hat{\mathbf{U}}(\boldsymbol{\gamma}^*,z_0)}{\partial\boldsymbol{\gamma}^*}=\mathbf{A}\sigma^{-1}(z_0)+o_p(1).\qquad\text{(B.9)}$$

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 $\hat{\mathbf{U}}(\boldsymbol{\gamma}^*,z_0)$ as

$$\hat{\mathbf{U}}(\boldsymbol{\gamma}^*,z_0)=\mathbf{d}_n(\tau_0)+\mathbf{q}_n(\tau_0),\qquad\text{(B.10)}$$

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

$$\mathbf{q}_n(\tau_0)=n^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n}\xi_{ij}\int_0^{\tau_0}K_h(Z_{ij}-z_0)\left\{\tilde{Z}_{ij}^*-\frac{\Phi_{nj1}(u,\boldsymbol{\beta}_0,\boldsymbol{\gamma}^*)}{\Phi_{nj0}(u,\boldsymbol{\beta}_0,\boldsymbol{\gamma}^*)}\right\}\times Y_{ij}(u)\exp[\boldsymbol{\beta}_0^\tau\mathbf{W}_{ij}(u)]\times\{\exp(g(Z_{ij}))-\exp(g(z_0)+\boldsymbol{\gamma}^{*T}\tilde{Z}_{ij}^*)\}\lambda_{0j}(u)\,du.\qquad\text{(B.11)}$$

Because 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 (B.6), we have

$$\mathbf{q}_n(\tau_0)=\frac{h^{p+1}}{(p+1)!}g^{(p+1)}(z_0)\mathbf{b}_{p+1}\sigma^{-1}(z_0)+\frac{h^{p+2}}{(p+2)!}g^{(p+2)}(z_0)\mathbf{b}_{p+2}\sigma^{-1}(z_0)+o_p(h^{p+2}).\qquad\text{(B.12)}$$

Rewrite $\mathbf{d}_n(\tau_0)$ in (B.10) as

$$\mathbf{d}_n(\tau_0)=\frac{1}{n}\sum_{j=1}^{J}\sum_{i=1}^{n}\xi_{ij}\int_0^{\tau_0}K_h(Z_{ij}-z_0)(\tilde{Z}_{ij}^*-\boldsymbol{\nu}_1)\,dM_{ij}(u)+\frac{1}{n}\sum_{j=1}^{J}\sum_{i=1}^{n}\xi_{ij}\int_0^{\tau_0}K_h(Z_{ij}-z_0)\times\left[\boldsymbol{\nu}_1-\frac{\Phi_{nj1}(u,\boldsymbol{\beta}_0,\boldsymbol{\gamma}^*)}{\Phi_{nj0}(u,\boldsymbol{\beta}_0,\boldsymbol{\gamma}^*)}\right]dM_{ij}(u).\qquad\text{(B.13)}$$

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 (B.6) and the dominated convergence theorem, ensures that the second of the foregoing terms is of mean 0 and variance $o(\frac{1}{nh})$. Thus

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

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

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

where

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

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

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

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

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

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 $\varepsilon>0$:

$$n^{-1}\sum_{i=1}^{n}hE\big[\|\mathbf{D}_i^*(t,h)\|^2\mathbf{I}(\sqrt{h/n}\,\mathbf{D}_i^*(t,h)>\varepsilon)\big]\to0.$$

This establishes the asymptotic normality of $\mathbf{d}_n^*(\tau_0)$ and thus $\sqrt{nh}\times\mathbf{d}_n(\tau_0)$, which, together with (B.8)–(B.10) and (B.12), yields the result of the theorem.

The following lemmas are needed to prove Theorems 1 and 2. Recall the expressions given in Appendix A.

Lemma B.1. Under conditions (a)–(h), if $nh^2\to\infty$, then the following items hold uniformly for $z\in\bigcup_{j=1}^J\operatorname{supp}(f_j)$:

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

$$\rho_j^*(z_0)=\int_0^{\tau_0}\mathcal{D}_z\left[\frac{\rho_{j2}(u|z_0)}{\rho_{j0}(u|z_0)}-\left(\frac{\rho_{j1}(u|z_0)}{\rho_{j0}(u|z_0)}\right)^{\otimes2}\right]\rho_{j0}(u|z_0)\,d\Lambda_{0j}(u).$$

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

Lemma B.2. Let $Q(u,Z_{ij})=c(K)\mathcal{D}_z(\log\sigma(Z_{ij}))+d(K)\times\mathcal{D}_z(\log\eta_{j0}(u|Z_{ij}))$. Assume that the conditions (a)–(h) hold. If $n\times h^{5/2}\to\infty$ and $nh^{2p}\to0$ for an even p, then

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

The proofs of Lemmas B.1 and B.2 are tedious. Detailed proofs are provided in a related technical report from the University of North Carolina at Chapel Hill (Cai, Fan, Jiang, and Zhou 2006).

Lemma B.3. Suppose that the conditions (a)–(h) hold. Then

$$n^{-1}\frac{\partial^2 \ell_p(\boldsymbol\beta)}{\partial\boldsymbol\beta\,\partial\boldsymbol\beta^\tau}\bigg|_{\boldsymbol\beta=\boldsymbol\beta_0} \xrightarrow{p} -\mathbf{I}(\boldsymbol\beta_0).$$

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

$$\begin{aligned} n^{-1}&\frac{\partial^2 \ell_p(\boldsymbol\beta)}{\partial\boldsymbol\beta\,\partial\boldsymbol\beta^\tau}\bigg|_{\boldsymbol\beta=\boldsymbol\beta_0} \\ &= -n^{-1}\sum_{j=1}^J\sum_{i=1}^n \xi_{ij}\int_0^{\tau_0}\left[\frac{R_{nj2}(\boldsymbol\beta_0,u)}{R_{nj0}(\boldsymbol\beta_0,u)} - \frac{R_{nj1}^{\otimes 2}(\boldsymbol\beta_0,u)}{R_{nj0}^2(\boldsymbol\beta_0,u)}\right]dN_{ij}(u) \\ &\quad - n^{-1}\sum_{j=1}^J\sum_{i=1}^n \xi_{ij}\int_0^{\tau_0}\left[\boldsymbol\kappa_n(Z_{ij}) - \frac{K_{nj1}(\boldsymbol\beta_0,u)}{K_{nj0}(\boldsymbol\beta_0,u)}\right]dN_{ij}(u), \end{aligned}$$

where $K_{njm}(\boldsymbol\beta_0,u) = \sum_{\ell=1}^n \xi_{\ell j}\hat s_{\ell j}(u,\boldsymbol\beta_0)(\boldsymbol\kappa_n(Z_{\ell j}))^m/\sum_{\ell=1}^n \xi_{\ell j}$, for $m=0,1$. By Lemma B.1 and $\hat g(z,\boldsymbol\beta_0) = g(z) + o_p(1)$ uniformly for $z \in \bigcup_{j=1}^J \operatorname{supp}[f_j(\cdot)]$,

$$\begin{aligned} n^{-1}&\frac{\partial^2 \ell_p(\boldsymbol\beta)}{\partial\boldsymbol\beta\,\partial\boldsymbol\beta^\tau}\bigg|_{\boldsymbol\beta=\boldsymbol\beta_0} \\ &= -n^{-1}\sum_{j=1}^J\sum_{i=1}^n \xi_{ij}\int_0^{\tau_0}\left[\frac{r_{j2}(\boldsymbol\beta_0,u)}{r_{j0}(\boldsymbol\beta_0,u)} - \left(\frac{r_{j1}(\boldsymbol\beta_0,u)}{r_{j0}(\boldsymbol\beta_0,u)}\right)^{\otimes 2}\right]dN_{ij}(u) \\ &\quad - n^{-1}\sum_{j=1}^J\sum_{i=1}^n \xi_{ij}\int_0^{\tau_0}\left[\boldsymbol\kappa(Z_{ij}) - \frac{K_{nj1}^*(\boldsymbol\beta_0,u)}{K_{nj0}^*(\boldsymbol\beta_0,u)}\right]dN_{ij}(u) + o_p(1), \end{aligned}$$

where $K_{njm}^*(\boldsymbol\beta_0,u)$ is defined similarly to $K_{njm}(\boldsymbol\beta_0,u)$ except with $\boldsymbol\kappa_n(Z_{\ell j})$ and $\hat g(Z_{\ell j},\boldsymbol\beta_0)$ replaced by $\boldsymbol\kappa(Z_{\ell j})$ and $g(Z_{\ell j})$. The second term equals

$$-n^{-1}\sum_{j=1}^J\sum_{i=1}^n \xi_{ij}\int_0^{\tau_0}\left[\boldsymbol\kappa(Z_{ij}) - \frac{K_{nj1}^*(\boldsymbol\beta_0,u)}{K_{nj0}^*(\boldsymbol\beta_0,u)}\right]dM_{ij}(u) = o_p(1);$$

therefore,

$$\begin{aligned} n^{-1}\frac{\partial^2 \ell_p(\boldsymbol\beta)}{\partial\boldsymbol\beta\,\partial\boldsymbol\beta^\tau}\bigg|_{\boldsymbol\beta=\boldsymbol\beta_0} &= -\sum_{j=1}^J p_j\int_0^{\tau_0}\left[\frac{r_{j2}(\boldsymbol\beta_0,u)}{r_{j0}(\boldsymbol\beta_0,u)} - \left(\frac{r_{j1}(\boldsymbol\beta_0,u)}{r_{j0}(\boldsymbol\beta_0,u)}\right)^{\otimes 2}\right] \\ &\qquad \times r_{j0}(\boldsymbol\beta_0,u)\,d\Lambda_{j0}(\boldsymbol\beta_0,u) + o_p(1) \\ &\equiv -\mathbf{I}(\boldsymbol\beta_0) + o_p(1). \end{aligned}$$

Proof of Theorem 1

By Lemma B.2,

$$n^{-1}\frac{\partial \ell_p(\boldsymbol\beta)}{\partial\boldsymbol\beta}\bigg|_{\boldsymbol\beta=\boldsymbol\beta_0} \xrightarrow{P} 0.$$

Thus, with probability tending to 1, for any small given $\varepsilon > 0$, if $\boldsymbol\beta \in S_\varepsilon \equiv \{\boldsymbol\beta : \|\boldsymbol\beta - \boldsymbol\beta_0\| \le \varepsilon\}$, then

$$\left|(\boldsymbol\beta-\boldsymbol\beta_0)^\tau\left[n^{-1}\frac{\partial \ell_p(\boldsymbol\beta)}{\partial\boldsymbol\beta}\bigg|_{\boldsymbol\beta=\boldsymbol\beta_0}\right]\right| \le \varepsilon^3. \tag{B.14}$$

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

$$(\boldsymbol\beta-\boldsymbol\beta_0)^\tau\left[n^{-1}\frac{\partial^2 \ell_p(\boldsymbol\beta)}{\partial\boldsymbol\beta\,\partial\boldsymbol\beta^\tau}\bigg|_{\boldsymbol\beta=\boldsymbol\beta_0}\right](\boldsymbol\beta-\boldsymbol\beta_0) \le -a\varepsilon^2, \tag{B.15}$$

with probability tending to 1. By the argument immediately after (B.3), with probability tending to 1, there is a constant $C > 0$ such that

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

Then substituting (B.14)–(B.16) into (B.2), we conclude with probability tending to 1 that when ε is small enough,

$$n^{-1}\ell_p(\boldsymbol\beta) - n^{-1}\ell_p(\boldsymbol\beta_0) \le 0. \tag{B.17}$$

Therefore, $\ell_p(\boldsymbol\beta)$ has a local maximum in the interior of S_ε, and, with probability tending to 1, there exists a consistent estimator sequence $\hat{\boldsymbol\beta}$ for $\boldsymbol\beta_0$ that maximizes the global profile pseudo–partial likelihood $\ell_p(\boldsymbol\beta)$.

Proof of Theorem 2

Lemma B.3 entails $n^{-1}\frac{\partial^2 \ell_p(\boldsymbol\beta)}{\partial\boldsymbol\beta\,\partial\boldsymbol\beta^\tau}|_{\boldsymbol\beta=\boldsymbol\beta_0} \xrightarrow{p} -\mathbf{I}(\boldsymbol\beta_0)$. Note that $\hat{\boldsymbol\beta}$ is consistent. Plugging the foregoing expression into (B.2), we obtain

$$\begin{aligned} \ell_p(\hat{\boldsymbol\beta}) &= \ell_p(\boldsymbol\beta_0) + (\hat{\boldsymbol\beta}-\boldsymbol\beta_0)^\tau\frac{\partial \ell_p(\boldsymbol\beta)}{\partial\boldsymbol\beta}\bigg|_{\boldsymbol\beta=\boldsymbol\beta_0} \\ &\quad - \frac{n}{2}(\hat{\boldsymbol\beta}-\boldsymbol\beta_0)^\tau\mathbf{I}(\boldsymbol\beta_0)(\hat{\boldsymbol\beta}-\boldsymbol\beta_0) \\ &\quad + o_p\{(\sqrt{n}\|\hat{\boldsymbol\beta}-\boldsymbol\beta_0\|+1)^2\}. \end{aligned} \tag{B.18}$$

Using corollary 1 of Murphy and van der Vaart (2000) and Lemma B.2, we obtain

$$\begin{aligned} \sqrt{n}&(\hat{\boldsymbol\beta}-\boldsymbol\beta_0) \\ &= \mathbf{I}(\boldsymbol\beta_0)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ell_p(\boldsymbol\beta)}{\partial\boldsymbol\beta}\bigg|_{\boldsymbol\beta=\boldsymbol\beta_0} + o_p(1) \\ &= \mathbf{I}(\boldsymbol\beta_0)^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^J\sum_{i=1}^n \xi_{ij}\int_0^{\tau_0}\bigg\{\mathbf{W}_{ij}(u) + \boldsymbol\chi(Z_{ij}) - \frac{r_{j1}(\boldsymbol\beta_0,u)}{r_{j0}(\boldsymbol\beta_0,u)} \\ &\qquad - \sigma(Z_{ij})\mathbf{s}(Z_{ij})Q(u,Z_{ij})\bigg\}\,dM_{ij}(u) + o_p(1+\sqrt{n}\|\hat{\boldsymbol\beta}-\boldsymbol\beta_0\|). \end{aligned}$$

Then by martingale central limit theorem and Slutsky's theorem,

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

[Received May 2005. Revised September 2006.]

REFERENCES

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993), *Statistical Models Based on Counting Processes*, New York: Springer-Verlag.
Andersen, P. K., and Gill, R. D. (1982), "Cox's Regression Model for Counting Processes: A Large-Sample Study," *The Annals of Statistics*, 10, 1100–1120.
Bickel, P. J. (1975), "One-Step Huber Estimates in Linear Models," *Journal of the American Statistical Association*, 70, 428–433.
Cai, J., Fan, J., Jiang, J., and Zhou, H. (2006), "Partially Linear Hazard Regression for Multivariate Survival Data," Mimeo Series 2235, Institute of Statistics, University of North Carolina.
Cai, J., and Prentice, R. L. (1995), "Estimating Equations for Hazard Ratio Parameters Based on Correlated Failure Time Data," *Biometrika*, 82, 151–164.
——— (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 the American Statistical Association*, 92, 477–489.
Clayton, D., and Cuzick, J. (1985), "Multivariate Generalizations of the Proportional Hazards Model" (with discussion), *Journal of the Royal Statistical Society*, Ser. 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.

Cuzick, J. (1992), "Semiparametric Additive Regression," *Journal of the Royal Statistical Society*, Ser. B, 54, 831–843.

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

Fan, J., and Chen, J. (1999), "One-Step Local Quasi-Likelihood Estimation," *Journal of the Royal Statistical Society*, Ser. B, 61, 927–943.

Fan, J., Gijbels, I., and King, M. (1997), "Local Likelihood and Local Partial Likelihood in Hazard Regression," *The Annals of Statistics*, 25, 1661–1690.

Fan, J., and Jiang, J. (1999), "Variable Bandwidth and One-Step Local M-Estimator," *Science in China*, Ser. A, 29, 1–15 (in Chinese).

Fan, J., and Li, R. (2004), "New Estimation and Model Selection Procedures for Semiparametric Modeling in Longitudinal Data Analysis," *Journal of the American Statistical Association*, 99, 710–723.

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

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

Gentleman, R., and Crowley, J. (1991), "Local Full Likelihood Estimation for the Proportional Hazards Model," *Biometrics*, 47, 1283–1296.

Härdle, W., Liang, H., and Gao, J. (2004), *Partially Linear Models*, Heidelberg: Springer-Verlag.

Hastie, T. J., and Tibshirani, R. J. (1990), "Exploring the Nature of Covariate Effects in the Proportional Hazards Model," *Biometrics*, 46, 1005–1016.

——— (1993), "Varying-Coefficient Models," *Journal of the Royal Statistical Society*, Ser. B, 55, 757–796.

Horowitz, J. L., and Mammen, E. (2004), "Nonparametric Estimation of an Additive Model With a Link Function," *The Annals of Statistics*, 32, 2412–2443.

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

Huang, J. (1999), "Efficient Estimation of the Partly Linear Additive Cox Model," *The Annals of Statistics*, 27, 1536–1563.

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*, New York: Wiley.

Liang, H., Härdle, W., and Carroll, R. J. (1999), "Estimation in a Semiparametric Partially Linear Errors-in-Variables Model," *The Annals of Stattistics*, 27, 1519–1535.

Liang, K. Y., Self, S. G., and Chang, Y. (1993), "Modeling Marginal Hazards in Multivariate Failure Time Data," *Journal of the Royal Statistical Society*, Ser. B, 55, 441–453.

Lin, D. Y. (1994), "Cox Regression Analysis of Multivariate Failure Time Data: The Marginal Approach," *Statistics in Medicine*, 13, 2233–2247.

Lin, X., and Carroll, R. J. (2001), "Semiparametric Regression for Clustered Data," *Biometrica*, 88, 1179–1185.

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*, eds. J. P. Klein and P. K. Goel, Dordrecht: Kluwer Academic, pp. 237–247.

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 the 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," *Biometrika*, 84, 349–363.

Robinson, P. M. (1988), "The Stochastic Difference Between Econometric and Statistics," *Econometrica*, 56, 531–547.

Speckman, P. (1988), "Kernel Smoothing in Partial Linear Models," *Journal of the Royal Statistical Society*, Ser. B, 50, 413–436.

Spiekerman, C. F., and Lin, D. Y. (1998), "Marginal Regression Models for Multivariate Failure Time Data," *Journal of the American Statistical Association*, 93, 1164–1175.

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, pp. 319–329.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989), "Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions," *Journal of the American Statistical Association*, 84, 1065–1073.