Ann Inst Stat Math (2014) 66:931–960
DOI 10.1007/s10463-013-0430-0

On estimation and inference in a partially linear hazard model with varying coefficients

Yunbei Ma · Alan T. K. Wan · Xuerong Chen · Yong Zhou

Received: 16 May 2012 / Revised: 14 August 2013 / Published online: 26 October 2013

Abstract We study estimation and inference in a marginal proportional hazards model that can handle (1) linear effects, (2) non-linear effects and (3) interactions between covariates. The model under consideration is an amalgamation of three existing marginal proportional hazards models studied in the literature. Developing an estimation and inference procedure with desirable properties for the amalgamated model is rather challenging due to the co-existence of all three effects listed above. Much of the existing literature has avoided the problem by considering narrow versions of the model. The object of this paper is to show that an estimation and inference procedure that

Electronic supplementary material The online version of this article (doi:10.1007/s10463-013-0430-0) contains supplementary material, which is available to authorized users.

Y. Ma
Center of Statistical Research and School of Statistics, Southwestern University of Finance and Economics, Chengdu 611130, China

A. T. K. Wan (✉)
Department of Management Sciences, City University of Hong Kong,
83 Tat Chee Ave, Kowloon Tong, Hong Kong
e-mail: Alan.Wan@cityu.edu.hk

X. Chen
Department of Statistics, George Washington University, Rome Hall,
Room 553, 801 22nd St. NW, Washington, DC 20052, USA

Y. Zhou
Academy of Mathematics and Systems Science, Chinese Academy of Sciences,
Beijing 100190, China

Y. Zhou
Department of Statistics, Shanghai University of Finance and Economics,
Shanghai 200433, China

accommodates all three effects is within reach. We present a profile partial-likelihood approach for estimating the unknowns in the amalgamated model with the resultant estimators of the unknown parameters being root-n consistent and the estimated functions achieving optimal convergence rates. Asymptotic normality is also established for the estimators.

Keywords Failure time · Hazard function · Profile local partial-likelihood · Root-n consistency

1 Introduction

The marginal proportional hazards model (Wei et al. 1989) is arguably the most popular model for analyzing multivariate failure time data. A common feature of multivariate failure time data is that failure times are often correlated (e.g., recurrences of a given disease in clinical trials). Consider a random sample of n subjects where in each subject there are m failure types. Let (i, j) denote the jth failure type in the ith subject, T_{ij} be the failure time, C_{ij} be the censoring time, $X_{ij} = \min(T_{ij}, C_{ij})$ the observed time, and Δ_{ij} an indicator that equals 1 if X_{ij} is a failure time and 0 otherwise, $i = 1, \ldots, n$, $j = 1, \ldots, m$. Further, let $\mathcal{F}_{t,ij}$ represent, up to time t, the failure, censoring and covariate information for the jth failure type in the ith subject, as well as the covariate information of other failure types in the ith subject. The marginal hazard function for the jth type of failure of the ith subject is defined as $\lambda_{ij}(t) = \lim_{h\downarrow 0}(h^{-1}P(T_{ij} \le t + h \mid T_{ij} > t, F_{t,i,j}))$. The censoring times are assumed to be independent of the failure times conditional on the covariates. The following marginal proportional hazards model examined by Wei et al. (1989) permits covariate effects on the hazard rate:

$$\lambda_{ij}(t) = \lambda_{0j}(t)\exp\{\alpha^{\mathrm{T}} V_{ij}(t)\}, \quad t \ge 0, \tag{1}$$

where $\lambda_{0j}(t)$ is an unspecified baseline hazard function, $V_{ij}(t)$ is a vector of covariates and α is the corresponding vector of regression coefficients to be estimated. Model (1) resembles the well-known Cox proportional hazards model for univariate failure time data, but allows for the dependence of related failure times. The latter dependence arises from the correlations between $V_{\cdot j}(t)$'s of different j's. The regression parameters in α are typically estimated by maximizing the failure specific partial likelihoods. During the past decades, there have been considerable research efforts extending and placing the marginal hazards model in more general contexts (e.g., Liang et al. 1993; Prentice and Hsu 1997; Spiekerman and Lin 1998).

Model (1) and its extensions typically assume that the variables in the covariate vector V enter the model linearly. The argument against the linear specification is that it is mainly chosen for convenience; in practice, the true covariate effect can be more complex than the log-linear effects. Non-parametric approaches are obvious alternatives, but the rate of convergence of non-parametric estimators typically decreases as the dimension of the model grows—this is the so-called curse of dimensionality problem. Cai et al. (2007a) proposed to model the covariate effects on the hazard rate through a partially linear model, where the logarithm of the hazard rate depends on the main exposure covariate of interest non-parametrically and other covariates linearly. That is,

$$\lambda_{ij}(t) = \lambda_{0j}(t)\exp\{\alpha^{\mathrm{T}} V_{ij}(t) + g(W_{ij}(t))\}, \quad t \geq 0, \tag{2}$$

where $V_{ij}(t)$ is a vector of covariates with linear effect on the logarithm of $\lambda_{ij}(t)$, $W_{ij}(t)$ is an exposure covariate and $g(\cdot)$ is an unspecified smooth function. Cai et al. (2007a) proposed a profile partial-likelihood estimation approach and established the asymptotic normality of the estimators of the parameters in both the linear and non-linear parts. When the (non-linear) interaction between the exposure variable and a confounding covariate is of interest, Cai et al. (2007b) (see also Fan et al. 2006) studied the varying-coefficient model

$$\lambda_{ij}(t) = \lambda_{0j}(t)\exp\{\beta(W_{ij}(t))^{\mathrm{T}} Z_{ij}(t) + g(W_{ij}(t))\}, \quad t \geq 0, \tag{3}$$

where $Z_{ij}(t)$ is a vector of covariates that interact with the main exposure covariate $W_{ij}(t)$ and $\beta(.)$ is an unknown varying coefficient function which may depend on $W_{ij}(t)$. Cai et al. (2007b) illustrated the relevance of model (3) by an example, whereby the exposure covariate is the patient's age, and factors such as body mass index, serum cholesterol level and smoking habit vary with the patient's age in determining the risk of death due to cardiovascular diseases. More recently, Cai et al. (2008) considered the model

$$\lambda_{ij}(t) = \lambda_{0j}(t)\exp\{\alpha^{\mathrm{T}} V_{ij}(t) + \beta(W_{ij}(t))^{\mathrm{T}} Z_{ij}(t)\}, \quad t \geq 0, \tag{4}$$

which includes possible interactions between the exposure and confounding covariates in the partially linear hazard regression. These authors further proposed a method based on profile partial-likelihood for estimating the unknown parameters in (4), and showed that the estimator of α is root-n consistent and the estimated function of β achieves optimal convergence rates.

This paper goes beyond those of Cai et al. (2007a, 2007b, 2008) by considering the model

$$\lambda_{ij}(t) = \lambda_{0j}(t)\exp\{\alpha^{\mathrm{T}} V_{ij}(t) + \beta(W_{ij}(t))^{\mathrm{T}} Z_{ij}(t) + g(W_{ij}(t))\}, \quad t \geq 0, \tag{5}$$

which combines (2), (3) and (4), and thus encompasses the merits of all three models, namely, it can model linear and non-linear covariate effects as well as the possibility of interactions between the covariates. An obvious theoretical appeal of (5) is that, because it accounts for all three effects it is less prone to misspecification than its narrow versions. From a practical standpoint, there is also a genuine need to consider (5) since there are many settings for which (5) is appropriate. One such example is given in Sect. 3.2, where a subset of data from the Busselton Population Health Surveys is used to study the hazard rate for coronary heart disease (CHD). There, age is taken to be W, the main exposure variable of interest; the covariates in Z are gender, body mass index, cholesterol level and smoking status, which are the risk factors that possibly interact with age; V, the vector of covariates exerting linear effect on the logarithm of hazard rate, is represented by a sole indicator variable of hypertension symptom. The non-linear form of $g(.)$ and the strong dependence of the time to CHD on age are clearly evident from the results, which also reveal the significance of the linear effect of hypertension symptom, and that gender, body mass index, cholesterol level and smoking status produce effects that vary with age.

Model (5) is clearly relevant, but previous studies only considered narrow versions of it with either $\alpha^T V_{ij}(t)$ or $g(W_{ij}(t))$ removed from the model. Model (5) is best analyzed using a profile local-partial likelihood estimation approach, but the inclusion of $g(.)$ and $\beta(.)$ combined with the interaction effects between $Z_{ij}(t)$ and $W_{ij}(t)$ makes it much harder to derive asymptotic properties for the resultant estimators of the unknowns. Although $g(W_{ij}(t))$ in (5) can be incorporated into (4) through a dummy variable with a column of ones in $Z_{ij}(t)$, the local intercept of $g(.)$ will cancel out in the local partial-likelihood function. The only way to estimate $g(.)$ is by integrating the estimate of its derivative function $g'(\cdot)$ using the trapezoidal rule (Hastie and Tibshirani 1993). This leads to a different estimation rule for $g(.)$ from that for $\beta(\cdot)$, making it extremely cumbersome to develop asymptotic results for the estimators of $\beta(.)$ and $g(.)$. The inclusion of $\alpha^T V_{ij}(t)$ and the need to derive asymptotic properties of the estimator of α pose further technical challenges. For this reason, the machinery and asymptotic results developed by Cai et al. (2008) are inappropriate for model (5). It is precisely because of these technical challenges that Cai et al. (2007a, 2007b, 2008) considered only narrow versions of model (5). Not surprisingly, the analytical tools developed in the above-mentioned papers *cannot* be straightforwardly applied to develop asymptotic theory in the present context.

The major scientific interest of this paper is to pursue an estimation and inference procedure with asymptotically valid properties for the amalgamated model (5). We demonstrate that such a procedure is in fact within reach in spite of the challenges involved in deriving the asymptotic results. In addition to technical innovations, the results developed in this paper also have important applications in view of the practical advantages offered by model (5) over the narrow models as mentioned above. The rest of this paper is organized as follows. Section 2 is devoted to a presentation of the main theoretical results of the paper. It begins by describing a local partial-likelihood framework for estimating the unknown coefficient functions and parameters in (5); the drawback of the local partial-likelihood approach is spelt out and a profile local partial-likelihood approach which enables the derivation of an estimator of α that is root-n consistent is described in detail along with a discussion of the proposed estimators' key asymptotic properties. Section 2 reports results of a simulation study that investigates the properties of the proposed estimator of α and estimated functions of $\beta(.)$ and $g(.)$ in finite samples. The proposed methodology is also illustrated in Sect. 2 using a subset of the data from the afore-mentioned Busselton Population Health Surveys. Section 4 concludes. Proofs of the major theorems, and the technical conditions and supporting lemmas that underline these theorems, are contained in a three-part Appendix. In addition, an online supplementary document available at http://personal.cb.cityu.edu.hk/msawan/research.htm provides the proofs to these supporting lemmas and an additional theorem.

2 Estimation and main results

As in Cai et al. (2007b), we drop the dependence of covariates on time with the understanding that the methods and proofs developed are applicable to external time-dependent covariates. Let $\mathcal{R}_j(t) = \{i : X_{ij} \geq t\}$ denote the set of the individuals at risk prior to time t. By assuming independence of failure times from the same subject, the partial-likelihood for model (5) may be written as

$$L(\beta(\cdot), \alpha, g(\cdot)) = \prod_{j=1}^{m}\prod_{i=1}^{n}\left\{\frac{\exp\{\beta(W_{ij})^{\mathrm{T}}Z_{ij} + \alpha^{\mathrm{T}}V_{ij} + g(W_{ij})\}}{\sum_{l\in\mathcal{R}_j(X_{ij})}\exp\{\beta(W_{lj})^{\mathrm{T}}Z_{lj} + \alpha^{\mathrm{T}}V_{lj} + g(W_{lj})\}}\right\}^{\Delta_{ij}}. \tag{6}$$

Now, for any given point w_0, by applying the Taylor series expansion, we have

$$\begin{aligned} &\beta(w) \approx \beta(w_0) + \beta'(w_0)(w - w_0) \equiv \delta + \gamma(w - w_0),\\ \text{and}\quad &g(w) \approx g(w_0) + g'(w_0)(w - w_0) \equiv a + b(w - w_0). \end{aligned} \tag{7}$$

Using (7) for the data around w_0 and introducing the kernel function, we obtain the following logarithm of the local partial-likelihood

$$\begin{aligned} &\ell_n(\delta, \gamma, \alpha, b)\\ &= \frac{1}{n}\sum_{j=1}^{m}\sum_{i=1}^{n} K_h(W_{ij} - w_0)\Delta_{ij} \times \Bigg\{\delta^{\mathrm{T}}Z_{ij} + \gamma^{\mathrm{T}}Z_{ij}(W_{ij} - w_0) + \alpha^{\mathrm{T}}V_{ij} + b(W_{ij} - w_0)\\ &\quad - \log\Bigg(\sum_{l\in\mathcal{R}_j(X_{ij})}\exp\{\delta^{\mathrm{T}}Z_{lj} + \gamma^{\mathrm{T}}Z_{lj}(W_{lj} - w_0) + \alpha^{\mathrm{T}}V_{lj} + b(W_{lj} - w_0)\}K_h(W_{lj} - w_0)\Bigg)\Bigg\}, \end{aligned} \tag{8}$$

where $K_h(\cdot) = K(\cdot/h)/h$ is a symmetric kernel function and h is a bandwidth. Equation (8) may be derived along the lines of Fan et al. (1997). Suppose (8) is maximized at $(\hat{\delta}(w_0), \hat{\gamma}(w_0), \hat{\alpha}(w_0), \hat{b}(w_0))$. Then, $\hat{\beta} = \hat{\delta}$ is a local linear estimator of the coefficient function $\beta(\cdot)$ at the point w_0. An estimator of $g'(\cdot)$ at ω_0 is simply the local slope $\hat{b}(\omega_0)$, i.e., $\hat{g}'(\omega_0) = \hat{b}(\omega_0)$. The curve $\hat{g}$ can be estimated by integrating $\hat{g}'(\omega_0)$, which may be approximated by the trapezoidal rule (Hastie and Tibshirani 1993). For the purpose of ensuring the identifiability of $g(\cdot)$, we set $g(0) = 0$ without loss of generality. As mentioned previously, although (5) may be treated as a special case of (4) by including a column of ones in $Z_{ij}(t)$ in (4), the local intercept for $g(.)$ will cancel out in the local partial-likelihood (8), rendering the technique and results of Cai et al. (2008) inapplicable.

One can show that although the local partial-likelihood estimation (LPLE) approach results in an estimator of $\beta(\cdot)$ that accomplishes the optimal rate of convergence, it leads to an estimator of $g(\cdot)$ with an asymptotic variance that is inflated since $g(\cdot)$ is estimated indirectly. Moreover, since only local data are used, the estimator of α

cannot converge to the true parameter at a rate faster than the non-parametric regression optimal rate of $n^{-2/5}$. However, if one is to test $\alpha = 0$ in (5), an estimator of α that converges to α faster than the rate of $n^{-2/5}$ is required to construct the test statistic. Cai et al. (2007b) encountered similar difficulties when estimating $g(\cdot)$ in the context of the narrow model (3), and similar problems concerning the estimators of α were also noted in Cai et al. (2008), where the model of interest is (4). Like these previous studies, our proposed solution is based on profile-likelihood, but there are additional challenges since the combined model (5) includes both $g(\cdot)$ and α, which makes the task of deriving asymptotic properties of estimators immensely more difficult. Undoubtedly, the inability to express asymptotically the score function of α as an integral of a predictable process with respect to a martingale due to the utilization of all observed information for estimating $g(\cdot)$ is a major obstacle in the derivation of asymptotic results. However, as we shall see, these derivations are made possible using the asymptotic results developed in the Appendix.

2.1 Profile local partial-likelihood estimation (PLPLE)

To describe the PLPLE approach, define $\varphi = (\delta^{\mathrm{T}}, \gamma^{\mathrm{T}}, b)^{\mathrm{T}}$, $\tilde{X}_{ij} = (Z_{ij}^{\mathrm{T}}, Z_{ij}^{\mathrm{T}}(W_{ij} - w_0), (W_{ij} - w_0))^{\mathrm{T}}$, $N_{ij}(t) = I(T_{ij} \le t, \Delta_{ij} = 1)$ and $Y_{ij}(t) = I(X_{ij} \ge t)$. As in Cai et al. (2007b), we restrict our attention to the time interval $[0, \tau]$ for convenience purposes, where τ is a constant denoting the time of the end of the study. For a given α, let $\tilde{\varphi}(w_0, \alpha)$ be the maximum of the estimator $\tilde{\varphi}(\cdot, \alpha)$ of $\varphi(\cdot)$, obtained by maximizing

$$
\begin{aligned}
\ell_n^*(\varphi, \alpha, \tau) &= n^{-1} \sum_{j=1}^{m} \sum_{i=1}^{n} \int_0^{\tau} K_{h_1}(W_{ij} - w_0)(\varphi^{\mathrm{T}} \tilde{X}_{ij} + \alpha^{\mathrm{T}} V_{ij}) \mathrm{d}N_{ij}(u) \\
&\quad - n^{-1} \sum_{j=1}^{m} \sum_{i=1}^{n} \int_0^{\tau} K_{h_1}(W_i - w_0) \\
&\quad \times \log \left\{ \sum_{l=1}^{n} Y_{lj}(u) \exp(\varphi^{\mathrm{T}} \tilde{X}_{lj} + \alpha^{\mathrm{T}} V_{lj}) K_{h_1}(W_{lj} - w_0) \right\} \mathrm{d}N_{ij}(u) \qquad (9)
\end{aligned}
$$

with respect to φ, and $\tilde{\beta}(w_0, \alpha)$ and $\tilde{g}(w_0, \alpha)$ be the respective maximums of the corresponding estimators $\tilde{\beta}(\cdot, \alpha)$ and $\tilde{g}(\cdot, \alpha)$ of β and $g(\cdot)$, where h_1 is the bandwidth for estimating $\varphi(\cdot)$. Upon substituting these estimators into (8), we obtain the following logarithm of the profile local partial-likelihood function:

$$
\begin{aligned}
\ell_p(\alpha) &= \frac{1}{n} \sum_{j=1}^{m} \sum_{i=1}^{n} \Delta_{ij} \times \Bigg\{ \tilde{\beta}^{\mathrm{T}}(W_{ij}, \alpha) Z_{ij} + \alpha^{\mathrm{T}} V_{ij} + \tilde{g}(W_{ij}, \alpha) \\
&\quad - \log \left(\sum_{l \in \mathcal{R}_j(X_{ij})} \exp\{\tilde{\beta}^{\mathrm{T}}(W_{lj}, \alpha) Z_{lj} + \alpha^{\mathrm{T}} V_{lj} + \tilde{g}(W_{ilj}, \alpha)\} \right) \Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{j=1}^{m}\sum_{i=1}^{n}\int_0^{\tau}\{\tilde{\beta}^{\mathrm{T}}(W_{ij},\alpha)Z_{ij}(u)+\alpha^{\mathrm{T}}V_{ij}(u)+\tilde{g}(W_{ij},\alpha)\}\mathrm{d}N_{ij}(u)\\
&\quad -\frac{1}{n}\sum_{j=1}^{m}\sum_{i=1}^{n}\int_0^{\tau}\log\left(\sum_{l=1}^{n}Y_{lj}(u)\exp\{\tilde{\beta}^{\mathrm{T}}(W_{lj},\alpha)Z_{lj}(u)\right.\\
&\quad \left. +\alpha^{\mathrm{T}}V_{lj}(u)+\tilde{g}(W_{lj},\alpha)\}\right)\mathrm{d}N_{ij}(u). \qquad (10)
\end{aligned}
$$

Following previous studies that emphasize narrow versions of (5), we assume, for the purpose of facilitating exposition, that Z's are independent of time, with the understanding that the method discussed here is applicable to external time-dependent covariates. Now, suppose that (10) is maximized at $\alpha = \hat{\alpha}_p$ and let $\hat{\varphi}_p = \tilde{\varphi}(w_0, \hat{\alpha}_p)$. The estimators of α, $\beta(\cdot)$ and $g(\cdot)$ are $\hat{\alpha}_p$, $\hat{\beta}_p(w_0) = \tilde{\beta}(w_0, \hat{\alpha}_p)$, and $\hat{g}_p(w_0) = \tilde{g}(w_0, \hat{\alpha}_p)$, respectively, where $\hat{g}_p(w_0)$ may be obtained by integrating $\hat{g'}_p(w_0) = \tilde{g'}(w_0, \hat{\alpha}_p)$. A backfitting algorithm similar to those developed by Cai et al. (2007a, 2008) may be used to compute these estimators based on PLPLE. One strong feature of this algorithm is that it takes care of the fact that $g(\cdot, \alpha)$ is implicitly defined. Let w_k $(k = 1, 2, \ldots, N)$ be a grid of points on the range of the exposure variable W. The algorithm may be succinctly summarized in the following steps:

Step 1. *Initialization* Set $\tilde{\alpha}$, the initial estimator of α, to $\bar{\alpha} = N^{-1}\sum_{k=1}^{N}\hat{\alpha}(w_k)$, the average of the local partial-likelihood estimator of α.

Step 2. *Estimation of the non-parametric component* Maximize $\ell_n^*(\varphi, \tilde{\alpha}, \tau)$ at each grid point w_k and obtain the non-parametric estimators $\tilde{\beta}(\cdot, \tilde{\alpha})$ and $\tilde{g}(\cdot, \tilde{\alpha})$ at all the grid points. Then obtain the non-parametric estimator at point W_{ij} by linear interpolation using the non-parametric estimators at the nearest grids to W_{ij}. Let these grid points be w_{k1} and w_{k2} such that $W_{ij} \in (w_{k1}, w_{k2})$. These non-parametric estimators may be written as

$$
\tilde{\beta}(W_{ij},\tilde{\alpha}) = \tilde{\beta}(w_{k1},\tilde{\alpha}) + (\tilde{\beta}(w_{k2},\tilde{\alpha}) - \tilde{\beta}(w_{k1},\tilde{\alpha}))\frac{W_{ij}-w_{k1}}{w_{k2}-w_{k1}},
$$
$$
\text{and } \tilde{g}(W_{ij},\tilde{\alpha}) = \tilde{g}(w_{k1},\tilde{\alpha}) + (\tilde{g}(w_{k2},\tilde{\alpha}) - \tilde{g}(w_{k1},\tilde{g}))\frac{W_{ij}-w_{k1}}{w_{k2}-w_{k1}}.
$$

Step 3. *Estimation of parametric component* With the estimates of α, $\beta(\cdot)$ and $g(\cdot)$ obtained in the previous steps, maximize the profile partial-likelihood $\ell_p(\alpha)$ with $\tilde{\beta}(\cdot, \alpha) = \tilde{\beta}(\cdot, \tilde{\alpha})$ and $\tilde{g}(\cdot, \alpha) = \tilde{g}(\cdot, \tilde{\alpha})$ using the Newton–Raphson algorithm.

Step 4. *Iteration* Iterate between Steps 2 and 3 until convergence. The final estimator of α is $\hat{\alpha}_p$.

Step 5. *Re-estimation of the nonparametric component* Set α to its final estimated value $\hat{\alpha}_p$ from Step 4. The final estimators $\hat{\beta}_p(\cdot)$ and $\hat{g}_p(\cdot)$ are then $\tilde{\beta}(\cdot, \hat{\alpha}_p)$ and $\tilde{p}(\cdot, \hat{\alpha}_p)$.

Only a small number of iterations are expected of Step 4 as the initial estimator of α is consistent. As for the number of iterations required for the Newton–Raphson algorithm in Step 3, Robinson (1988) showed that 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 $O_p(n^{-ak})$. Now, the initial estimators of $\beta(\cdot)$ and $g(\cdot)$ in Step 3 have at least the rate $O_p(n^{-\frac{1}{5}})$, and with $k = 3$, the order of error is $o(n^{-1/2})$. Thus, three iterations in the Newton Raphson algorithm are sufficient to yield estimators of $\beta(\cdot)$ and $g(\cdot)$ that are sufficiently close to the maximum likelihood estimators. Although Robinson (1988) theoretical results and the rule-of-thumb are obtained for parametric models, results in other contexts (e.g., Cai et al. 2007a, 2008) showed that they provide a good approximation to non-parametric applications.

The estimators $\hat{\beta}_p(\cdot)$, $\hat{\alpha}_p$ and $\hat{g}_p(\cdot)$ thus obtained also enable the derivation of the following consistent estimator of the cumulative baseline hazard function $\Lambda_{0j} = \int_0^t \lambda_{0j}(u)\mathrm{d}u$:

$$\hat{\Lambda}_{0jp}(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\mathrm{d}N_{ij}(u)}{n^{-1}\sum_{l=1}^{n} Y_{lj}(u)\exp(\hat{\beta}_p^{\mathrm{T}}(W_{lj})Z_{lj}(u) + \hat{\alpha}_p^{\mathrm{T}} V_{lj}(u) + \hat{g}_p(W_{lj}))}.$$

2.2 Asymptotic properties

Notwithstanding the close analogy between the estimation steps described in the last section and those for the narrow models discussed in the literature, obtaining asymptotic properties of the unknowns in the present context poses major difficulties. New technical challenges arise because of the co-existence of α, $g(\cdot)$ and $\beta(\cdot)$ in (5). As shown previously, $\hat{\beta}_p(\cdot)$, the estimator of $\beta(\cdot)$, is an implicit estimator, while $\hat{g}_p(\cdot)$, the estimator of $g(\cdot)$, utilizes all information from the data. These make neither the common martingale methods nor the techniques developed in Cai et al. (2007a, 2007b, 2008), applicable in analyzing the estimators' asymptotic properties. In spite of these difficulties, we derive and present in this section the asymptotic properties of these estimators and establish their asymptotic normality. Our analysis is made possible through the asymptotic theory developed in Lemmas 1–3 given in the Appendix, yielding an asymptotic representation of the local linear kernel estimator of the varying-coefficient function.

Our theoretical analysis relies on certain technical assumptions which are relegated to Sect. 5.1 of the Appendix for ease of exposition. To introduce notations, let $\mu_i = \int x^i K(x)\mathrm{d}x$, and $\nu_i = \int x^i K^2(x)\mathrm{d}x$. Denote $A(u) = (Z^{\mathrm{T}}(u), V^{\mathrm{T}}(u))^{\mathrm{T}}$,

$$P(u, z, v, w_0) = P(X \geq u | Z = z, V = v, W = w_0), \text{ and}$$
$$\rho(u, z, v, w_0) = P(u, z, v, w_0)\exp\{\beta^{\mathrm{T}}(w_0)z + \alpha^{\mathrm{T}} v + g(w_0)\}.$$

Also, define, for $k = 0, 1, 2$, $j = 1, \ldots m$,

$$b_{jk}(u) = b_{jk}(u, w_0) = f_j(w_0)E\{\rho(u, Z_j, v, w_0)Z_j^{\otimes k} | w = w_0\},$$
$$t_{jk}(u) = t_{jk}(u, w_0) = t_{jk}(u, \alpha_0, w_0) = f_j(w_0)E[\rho(u, Z_j(u), V_j(u), w_0) \cdot Z_j^{\otimes k}(u)$$

$$\otimes V(u)|w = w_0], \text{ and}$$
$$t_j^*(u) = t_j^*(u, w_0) = t_{jk}(u, \alpha_0, w_0) = f_j(w_0)E[\rho(u, Z_j(u), V_j(u), w_0) \cdot Z_j^{\mathrm{T}}(u) \otimes V(u)|w = w_0],$$

where $f_j(\cdot)$ is the density of W_j, and $Z^{\otimes k} = 1, Z$ and ZZ^{T} for $k = 0, 1$ and 2, respectively. Additionally, set

$$b_{jk} = b_{jk}(w_0) = \int_0^\tau b_{jk}(u, w_0)\mathrm{d}\Lambda_{0j}(u), \text{ and}$$
$$b_k = \sum_{j=1}^m b_{jk},$$

and wherever there is no ambiguity, we will drop the dependence of $b_k(u, w_0)$ and $b_k(w_0)$ on w_0. Furthermore, let

$$\Gamma^* = \Gamma^*(w_0) = \left\{ \sum_{j=1}^m \left[b_2 - \int_0^\tau \frac{b_{j1}^{\otimes 2}(u, w_0)}{b_{j0}(u, w_0)} \right] \mathrm{d}\Lambda_{0j}(u) \right\}^{-1}$$

and

$$\mathbb{A}^* = \begin{pmatrix} \Gamma^{*-1} & 0_{p\times p} & 0_p \\ 0_{p\times p} & \mu_2 b_2 & \mu_2 b_1 \\ 0_p^{\mathrm{T}} & \mu_2 b_1^{\mathrm{T}} & \mu_2 b_0 \end{pmatrix}.$$

Note that b_0 is a scalar.

We will now show that the estimator of α resulting from PLPLE is consistent and asymptotically normal.

Theorem 1 *Assume that conditions* (C.1)–(C.8) *in Appendix* 5.1 *are satisfied. Then, $\hat{\alpha}_p \to \alpha_0$ with probability tending to one, where $\hat{\alpha}_p$ is the estimator of α that maximizes the profile partial-likelihood $\ell_p(\alpha)$ and α_0 is the true value of α.*

Proof See Appendix 5.2. □

Theorem 2 *Assume that conditions* (C.1)–(C.8) *in Appendix* 5.1 *are satisfied. If $nh_1^2 \to \infty$ and $nh_1^4 \to 0$, then the sequence of estimators in Theorem* 1 *satisfies the following convergence property:*

$$\sqrt{n}(\hat{\alpha}_p - \alpha_0) \xrightarrow{\mathcal{L}} N(0, \Omega),$$

where $\Omega = I^{-1}\Sigma^* I^{-1}$,

$$I = I(\alpha_0) = \sum_{j=1}^m \int_0^\tau \left[\frac{r_{j2}(u, \alpha_0)}{r_{j0}(u, \alpha_0)} - \frac{r_{j1}^{\otimes 2}(u, \alpha_0)}{r_{j0}^2(u, \alpha_0)} \right] r_{j0}(u, \alpha_0)\mathrm{d}\Lambda_{0j}(u),$$

$I(\alpha_0)$ and $\Sigma^* = \Sigma^*(\alpha_0)$ are positive definite, and

$$\Sigma^*(\alpha_0) = E\left[\sum_{j=1}^{m}\int_0^{\tau}\left\{\left[V_{\cdot j}(u) + \chi_2(W_{\cdot j})Z_j + \chi_1(W_{\cdot j}) - \frac{r_{j1}(u,\alpha_0)}{r_{j0}(u,\alpha_0)}\right]\right.\right.$$
$$\left.\left. - Q(u, W_{\cdot j})s(W_{\cdot j})\right\} \mathrm{d}M_{\cdot j}(u)\right]^2.$$

The definitions of the terms embedded in $I(\alpha_0)$ and $\Sigma^*(\alpha_0)$ can be found in the Appendix.

Proof See Appendix 5.2.

Remark 1 The unknown I and Σ^* in Ω may be estimated by an empirical plug-in method. Let $\hat{I}$ and $\hat{\Sigma}^*$ be the estimates. Hence we have $\hat{\Omega} = \hat{I}^{-1}\hat{\Sigma}^*\hat{I}^{-1}$ as the estimated asymptotic covariance of α_p. Also, $\hat{\Omega}$ is consistent since both $\hat{I}$ and $\hat{\Sigma}^*$ are consistent estimators of their respective unknowns.

Now, define the following Wald test statistic for the testing problem $H_0 : \alpha = \alpha_0 \; rmvs \; H_1 : \alpha \neq \alpha_0$:

$$W_n = n(\hat{\alpha}_p - \alpha_0)^{\mathrm{T}}\hat{\Omega}^{-1}(\hat{\alpha}_p - \alpha_0). \tag{11}$$

The following theorem relates to the asymptotic null distribution of W_n.

Theorem 3 *Let the conditions of Theorem 2 be satisfied. The asymptotic null distribution of W_n is $\chi^2(p)$, where p is the dimension of α.*

Proof The proof is omitted as it is a straightforward consequence of Theorem 2.

Results for testing a subset of the coefficient of α may be developed analogously to the above. The next theorem concerns the properties of estimators in the non-parametric component of the model.

Theorem 4 *Assume that conditions* (C.1)–(C.8) *in Appendix 5.1 are satisfied. If $\hat{\alpha}_p$ is root n-consistent and nh_1^5 is bounded, then*

$$\sqrt{nh_1}[\mathbf{H}_1(\hat{\varphi}_p(w_0, \hat{\alpha}_p) - \varphi_0) - \mathbf{b}_n(w_0)] \xrightarrow{\mathcal{L}} N(0, \mathbf{V}(w_0)),$$

where h_1 is a bandwidth, $\mathbf{H}_1$ is a diagonal matrix with the first p diagonal elements equalling 1 and the remaining $p+1$ diagonal elements equalling h_1, $\mathbf{b}_n(w_0) = \mathbb{A}^{*-1}\mathbb{B}_n^*(\tau, w_0)$, and $\mathbf{V}(w_0) = \mathbb{A}^{*-1}\Pi^*(\tau, w_0)\mathbb{A}^{*-1}\sigma(w_0)$.

Proof See Appendix 5.2.

Remark 2 The biases and variances, $\mathbf{b}_n(w_0)$ and $\mathbf{V}(w_0)$ of the profile partial-likelihood estimators $\mathbf{H}_1(\hat{\varphi}_p(w_0, \hat{\alpha}_p) - \varphi_0)$ can be estimated by $\hat{\mathbf{b}}_n(w_0) = \hat{\mathbb{A}_n^{*-1}}(\tau, w_0)\hat{\mathbb{B}^*}_n(\tau, w_0)$ and $\hat{\mathbf{V}}(w_0) = (nh_1)\hat{\mathbb{A}_n^{*-1}}(\tau, w_0)\hat{\Pi^*}_n(\tau, w_0)\ \hat{\mathbb{A}_n^{*-1}}(\tau, w_0)$, respectively, where

$$\hat{\mathbb{A}_n^*}(\tau, w_0) = \frac{1}{n}\sum_{j=1}^{m}\sum_{i=1}^{n}\int_0^\tau K_{h_1}(W_{ij} - w_0)\frac{\hat{\Phi}_{nj2}(u, w_0)\hat{\Phi}_{nj0}(u, w_0) - \hat{\Phi}_{nj1}^{\otimes 2}(u, w_0)}{\hat{\Phi}_{nj0}^2(u, w_0)}\mathrm{d}N_{ij}(u),$$

$$\hat{\mathbb{B}^*}_n(\tau, w_0) = \frac{1}{n}\sum_{j=1}^{m}\sum_{i=1}^{n}\int_0^\tau K_{h_1}(W_{ij} - w_0)\left[\tilde{U}_{ij}(w_0) - \frac{\hat{\Phi}_{nj1}(u, w_0)}{\hat{\Phi}_{nj0}(u, w_0)}\right]Y_{ij}(u)\hat{\lambda}_{ijp}(u)\mathrm{d}u,$$

and

$$\hat{\Pi^*}_n(\tau, w_0) = \frac{h_1}{n}\left[\sum_{j=1}^{m}\sum_{i=1}^{n}\int_0^\tau K_{h_1}(W_{ij} - w_0)\tilde{U}_{ij}(w_0) - \frac{\hat{\Phi}_{nj1}(u, w_0)}{\hat{\Phi}_{nj0}(u, w_0)}Y_{ij}(u)\hat{\lambda}_{ijp}(u)\mathrm{d}u\right]^{\otimes 2},$$

with

$$\hat{\lambda}_{ijp}(u) = \exp\{\hat{\beta}_p^{\mathrm{T}}(W_{ij})Z_{ij}(u) + \hat{\alpha}_p^{\mathrm{T}}V_{ij}(u) + \hat{g}_p(W_{ij})\}\hat{\lambda}_{0jp}(u),$$
$$\hat{\Phi}_{njk}(u, w_0) = \Phi_{njk}(u, \hat{\alpha}_p, \hat{\varphi}_p(w_0, \hat{\alpha}_p)),\quad k = 0, 1, 2,\ \ j = 1, \ldots, m.$$

and $\tilde{U}_{ij}(\cdot)$ and $\Phi_{njk}(\cdot, \cdot, \cdot)$ being defined in Appendix 5.1. Moreover, the estimator of the covariance matrix Ω of $\hat{\alpha}_p$ is also available. We present the details in Remark 4 in Appendix 5.1.

From Theorem 4, we have the following result on the asymptotic normality of the estimators of $\beta(\cdot)$ and $g(\cdot)$:

Corollary 1

$$\sqrt{nh_1}\left[\hat{\beta}_p(w_0, \hat{\alpha}_p) - \beta_0(w_0) - \frac{h_1^2\mu_2}{2}\sigma^{-1}(w_0)\beta_0''(w_0)\right] \xrightarrow{\mathcal{L}} N(0, v_1^2(w_0))$$
$$\text{and } \sqrt{nh_1^3}[\hat{g'}_p(w_0, \hat{\alpha}_p) - g_0'(w_0)] \xrightarrow{\mathcal{L}} N(0, v_2^2(w_0)),$$

where $v_1^2(w_0) = \Gamma^*\nu_0\sigma(w_0)$ *and* $v_2^2(w_0) = \nu_2\mu_2^{-2}\sigma(w_0)\begin{pmatrix} E[ZZ^{\mathrm{T}}] & EZ \\ EZ^{\mathrm{T}} & 1 \end{pmatrix}^{-1}$.

Corollary 1 enables the derivation of the bandwidth that is optimal for estimating β_k, in the sense of minimizing the asymptotic weighted mean integrated square error

$$\int\left\{\left[\frac{h_1^2\mu_2}{2}\sigma^{-1}(w)\beta_k''(w)\right]^2 + \frac{1}{nh_1}\Gamma_k^*(w)\nu_0\sigma(w)\right\}W(w)\mathrm{d}w,$$

where $W(.)$ is a weighted function. This optimal bandwidth is given by

$$h_{1k,opt} = \left[\frac{\int \Gamma_k^*(w)\nu_0\sigma(w)W(w)\mathrm{d}w}{\int \mu_2^2\beta_k''^2(w)\sigma^{-2}(w)W(w)\mathrm{d}w} \right]^{\frac{1}{5}} n^{-\frac{1}{5}}, \tag{12}$$

where Γ_k^* and β_k are the kth components of Γ^* and β_0, respectively.

Theorem 5 *Suppose conditions* (C1)–(C8) *in Appendix* 5.1 *are satisfied. Then, we have, for each* $j = 1, \ldots, m$,

$$\hat{\Lambda}_{0jp}(t) \longrightarrow \Lambda_{0j}(t) \quad \textit{and} \quad \hat{\lambda}_{0jp}(t) \longrightarrow \lambda_{0j}(t)$$

uniformly on $(0, \tau]$ *in probability.*

Proof Available from the online supplementary document at http://personal.cb.cityu.edu.hk/msawan/research.htm.

2.3 Bandwidth selection via the K-fold cross-validation method

The optimal bandwidth $h_{1k,opt}$ in (12) contains unknown parameters. In practice, data driven methods may be utilized for the selection of bandwidth. Here, we adopt the K-fold cross-validation method based on the prediction error (Tian et al. 2005; Fan et al. 2006) for bandwidth selection.

To implement the method, we divide each fold data into K equal-sized subgroups denoted by D_{jk} ($j = 1, \ldots, J, k = 1, \ldots, K$). The kth prediction error is given by

$$PE_k(h) = \sum_{j=1}^{J} \sum_{i \in D_{jk}} \int_0^\tau \{N_{ij}(t) - \hat{E}N_{ij}(t)\}^2 \mathrm{d}\left\{ \sum_{l \in D_{jk}} N_{lj}(t) \right\},$$

where

$$\hat{E}N_{ij}(t) = \int_0^t I(Y_{ij} \geq t)\Delta_{ij} \exp\{\hat{\alpha}_{(-k,p)}^{\mathrm{T}} V_{ij} + \hat{\beta}_{(-k,p)}^{\mathrm{T}}(W_{ij})Z_{ij} + \hat{g}_{(-k,p)}(W_{ij})\}\mathrm{d}\hat{\Lambda}_{0j(-k,p)}(u).$$

The quantities $\hat{\alpha}_{(-k,p)}$, $\hat{\beta}_{(-k,p)}(W_{ij})$, $\hat{g}_{(-k,p)}(W_{ij})$ and $\hat{\Lambda}_{0j(-k,p)}(u)$ are estimates obtained using data from all subgroups other than the subgroup D_{jk}. The optimal bandwidth is then obtained by minimizing the total prediction error $PE(h) = \sum_{k=1}^{K} PE_k(h)$ with respect to h.

3 A simulation study and a real data application

3.1 Simulation study

In this section, we compare the LPLE and PLPLE approaches in finite samples by simulations. The data are generated from a multivariate extension of the model of Clayton and Cuzick (1985), where the joint survival function of $(T_1, \ldots, T_J)$ conditional on $(Z_1, \ldots, Z_J)$, $(V_1, \ldots, V_J)$ and $(W_1, \ldots, W_J)$ is given by

$$\begin{aligned} &S(t_1, \ldots, t_J; Z_1, \ldots, Z_J, V_1, \ldots, V_J, W_1, \ldots, W_J) \\ &\quad = \left\{ \sum_{j=1}^{J} S_j(t_j)^{-1/\theta} - (J-1) \right\}^{-\theta}, \end{aligned} \tag{13}$$

where $S_j(t)$ is the marginal survival probability of the jth failure type at time t, and θ is a parameter representing the degree of dependence within a subject. Note that $\varsigma = \theta/(2+\theta)$ is Kendall's rank correlation coefficient. In our simulations, we set J to 3, and θ to 10.0 and 0.1, representing, respectively, low and high dependence within a subject. We further assume that T_{ij} has an exponential marginal distribution with the failure rate described by (5) and baseline hazard function depending on time through $\lambda_{0j}(t) = 4t^3\lambda^*_{0j}$; we set λ^*_{0j} to 0.2, 1.0 and 1.5, for $j = 1, 2$ and 3, respectively. Additionally, we set $\alpha = 0.8$, $\beta(W) = (0.5W(1.5 - W), \sin 2W)^{\mathrm{T}}$, and $g(W) = 2\sin(2W)$, where W is uniform over $[0, 3]$. Moreover, for each $j = 1, 2, 3$, we let $V_{\cdot j} \sim N(0, 5)$ and $Z_{\cdot j} = (Z_{\cdot j1}, Z_{\cdot j2})^{\mathrm{T}} \sim N(0, \Sigma)$ such that $Z_{\cdot j1}$ and $Z_{\cdot j2}$ each have a variance of 5 and the covariance between $Z_{\cdot j1}$ and $Z_{\cdot j2}$ is $\sqrt{5}$. Along the lines of Cai and Shen (2000), the failure times (t_{i1}, t_{i2}, t_{i3}) are generated by

$$\begin{aligned} t_{i1} &= [-\log(1-u_{i1})\Pi(Z_{i1}, V_{i1}, W_{i1}, \lambda^*_{01})]^{1/4}, \\ t_{i2} &= [\theta \log(1 - a_{i1} + a_{i1}(1-u_{i2})^{(\theta^{-1}+1)^{-1}}\Pi(Z_{i2}, V_{i2}, W_{i2}, \lambda^*_{02})]^{1/4}, \text{ and} \\ t_{i3} &= [\theta \log(1-(a_{i1}+a_{i2})+(a_{i1}+a_{i2})-1)(1-u_{i3})^{(\theta^{-1}+2)^{-1}}\Pi(Z_{i3}, V_{i3}, W_{i3}, \lambda^*_{03})]^{1/4}, \end{aligned}$$

where $a_{il} = (1-u_{il})^{-\theta}$ for $l = 1, 2, i = 1, \ldots, n$, $\Pi(Z, V, W, \lambda) = \exp\{\beta^{\mathrm{T}}(W)Z + \alpha^{\mathrm{T}}V + g(W)\}/\lambda^*$ and (u_{i1}, u_{i2}, u_{i3}) are independent Uniform random variables. Moreover, we generate the censoring time C from a Uniform distribution over $(0, c)$, where c is a constant controlling the censoring rate; here, c is set to 5 and 15, corresponding to censoring rate of approximately 30 and 10 %, respectively. We use a Gaussian kernel function in both LPLE and PLPLE, and the K-folder cross-validation method to select the optimal bandwidth with $K = 5$. We set the sample size to $n = 200$ and the reported sampling properties are based on 200 replications.

Some representative results are displayed in Fig. 1a, h, where the fitted curves of $\beta(w)$ and $g(w)$ based on LPLE and PLPLE under censoring rate of 10 % and $\theta = 0.1, 10$ are plotted alongside the curves of the actual functions. In the figures, the solid curves represent the true functions, while the dotted and dashed curves are the estimated functions based on LPLE and PLPLE, respectively. These curves show

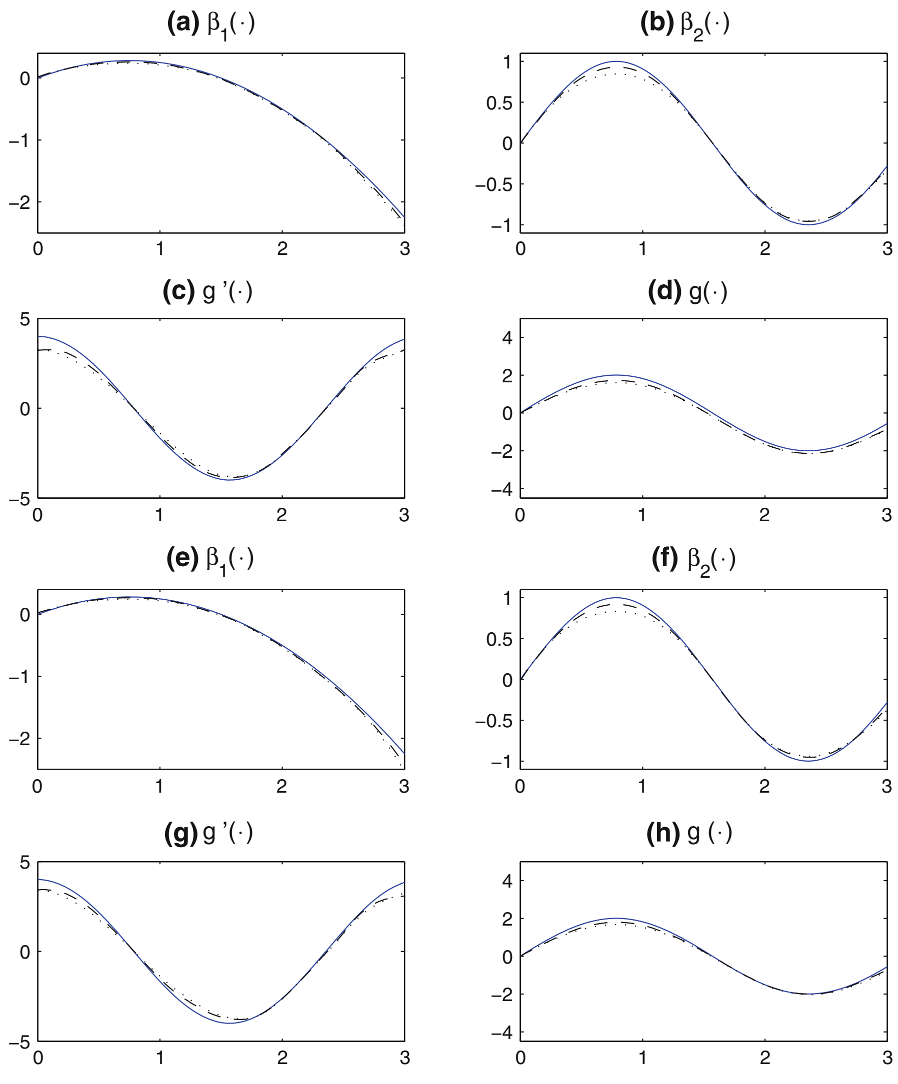

Fig. 1 These figures provide the estimated curves of $\beta_1(w)$, $\beta_2(w)$, $g(w)$ and $g'(w)$, with $\theta = 10$ for **a–d**, and $\theta = 0.1$ for **e–f**. The censoring rate is set to 10 % in all cases. The *solid*, *dotted* and *dashed curves* represent, respectively, the true function and estimated functions based on the LPLE and PLPLE approaches

that both LPLE and PLPLE result in estimated functions of $\beta(w)$ and $g(w)$ that are very close to the corresponding true coefficient functions, indicating that the proposed methods are not far from matching the true structure. Moreover, the observed results seem to be invariant with respect to the degree of dependence within a subject, as the shapes of the estimated curves do not change very much when θ alternates between 0.1 and 10. Other choices of censoring rates and θ have also been considered and the results are largely similar. These are omitted for presentation here to conserve space, but details are available on request from the authors.

Table 1 Sampling performance of the LPLE and PLPLE when estimating α

θ	c.r. (%)	bias($\hat{\alpha}$)	bias($\hat{\alpha}_p$)	SE($\hat{\alpha}$)	SE($\hat{\alpha}_p$)	SD($\hat{\alpha}$)	SD($\hat{\alpha}_p$)	CP($\hat{\alpha}$)	CP($\hat{\alpha}_p$)
10	10	0.0211	0.0077	0.0940	0.0292	0.1011	0.0325	0.9375	0.9435
	30	0.096	0.063	0.1065	0.0326	0.1198	0.0484	0.9368	0.9451
0.1	10	−0.0014	−0.0006	0.0983	0.0296	0.1069	0.0450	0.9306	0.9375
	30	0.0215	0.0201	0.1106	0.0333	0.1237	0.0543	0.9294	0.9404

The preceding analysis shows that the LPLE and PLPLE approaches are both effective in estimating the non-parametric components of the model and the two approaches do not yield substantially different estimates of $\beta(\cdot)$ and $g(\cdot)$. Next, we present simulation results on the properties of the estimators of α based on the two approaches. The inference comparisons are in terms of magnitude of bias (bias), standard deviations (SD), standard errors of the estimators (SE) and the proximity of actual confidence interval coverage to the nominal target coverage of 95 % (CP). The results for samples of size 200, $\theta = 0.1$, 10 and censoring rate of 10 and 30 % are shown in Table 1. These reported sampling properties are based on 200 replications.

Table 1 shows that for the cases considered, the PLPLE is less biased than the LPLE, although the magnitudes of bias produced by the two estimators are not large relative to the magnitude of the parameter being estimated (recall that α is set to 0.8). In terms of SE and SD, the PLPLE is again superior to LPLE, and usually by a wide margin. Other things being equal, an increase in the censoring rate (c.r.) generally leads to an increase in standard deviations and standard errors of the estimates. When comparing with respect to confidence interval widths, in all cases, the PLPLE exhibits widths that are closer to the target 95 % level, irrespective of the levels of θ and the censoring rate.

3.2 A real example

The following example, based on data from the Busselton Population Health Surveys, concerns the effects of different factors on the risk of death due to CHD. The Busselton surveys are a series of cross-sectional health surveys conducted in the town of Busselton in Western Australia every 3 years from 1966 to 1981 by means of questionnaires and clinical visits. See Knuiman et al. (1994) for a detailed description of the surveys. Cai et al. (2007b) used data from the same surveys to illustrate their proposed method for estimating the unknown parameters in model (2). Our dataset comprises 2202 observations from participants of 619 families. Here, the main exposure variable of interest is AGE (age in years), the covariates that possibly interact with AGE are GENDER (1 for female and 0 for male), BMI (body mass index, in kg/m^2), $rmBMI^2$, CHOL (serum cholesterol level, in mmol/L), SMOKE1 (=1 for ex-smoker and 0 otherwise) and SMOKE2 (=1 for current smoker and 0 otherwise). The covariate that exerts linear effect on the logarithm of hazards rate is RXHYPER (=1 if participant is diagnosed with hypertension). Thus, the model being considered is

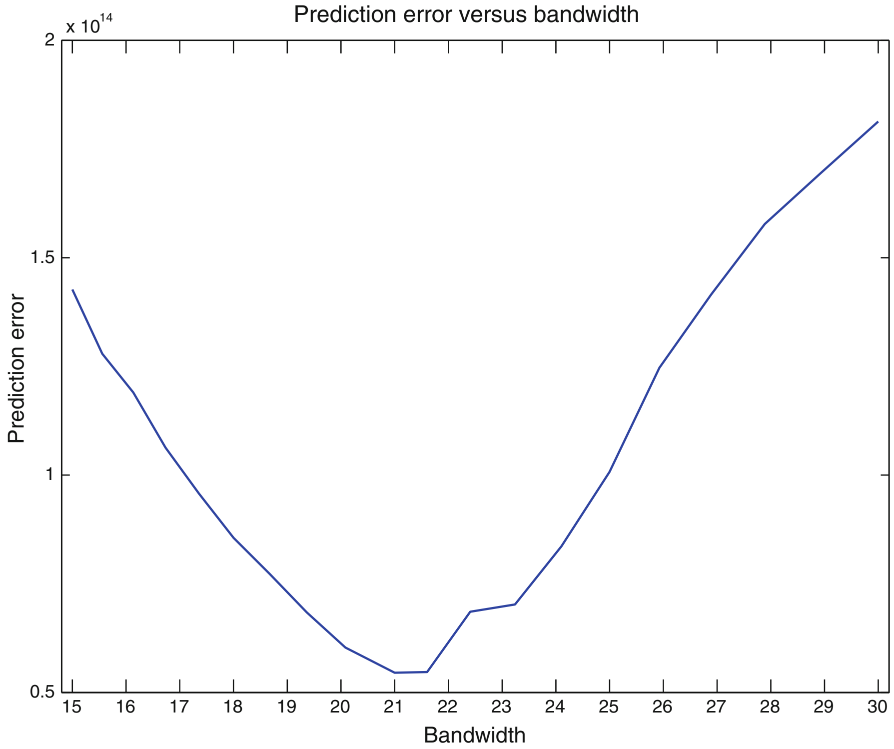

Fig. 2 Plot of prediction errors against different bandwidths for the Busselton Population Health Survey data

$$\begin{aligned}\lambda_i(t|\mathcal{F}_i) = \lambda_0(t|\mathcal{F}_i)\exp(&\beta_1(\mathrm{AGE}_i) * \mathrm{GENDER}_i + \beta_2(\mathrm{AGE}_i) * \mathrm{BMI}_i \\ &+ \beta_3(\mathrm{AGE}_i) * \mathrm{BMI}_i^2 + \beta_4(\mathrm{AGE}_i) * \mathrm{CHOL}_i + \beta_5(\mathrm{AGE}_i) * \mathrm{SMOKE1}_i \\ &+ \beta_6(\mathrm{AGE}_i) * \mathrm{SMOKE2}_i + \alpha * \mathrm{RXHYPER} + g(\mathrm{AGE}_i)).\end{aligned}$$

In the dataset, the mean values of AGE, BMI and CHOL are 41.7, 24.8 and 5.65, respectively, and the percentages of males, ex-smokers and current smokers are 49, 17 and 34 %, respectively. The main exposure variable, AGE (in years), ranges from 16.3 to 89. We divide [16.3, 89] equally into 100 intervals, and use all 101 boundary points as grids. In our estimation, we select the bandwidth by the K-fold cross-validation method with $K = 10$. As shown in Fig. 2, we found that the bandwidth of $h_1 = 21$ (in years) yields the smallest prediction error.

Applying the proposed PLPLE technique, we obtain an α estimate of 0.1478 with a standard error of 0.0348, and estimated coefficient functions as depicted in Fig. 3a, h. The appropriateness of model (14), where linear, non-linear and interaction effects of covariates are all included, is substantiated by the significance of α, the dependence of the time to CHD on AGE (Fig. 3h), the non-liner form of the estimated $g(\cdot)$ (Fig. 3h), and the clear indication that GENDER, BMI, BMI^2, CHOL, SMOKE1 and SMOKE2 vary with AGE in producing significant effects on the hazards rate (Fig. 3a–f).

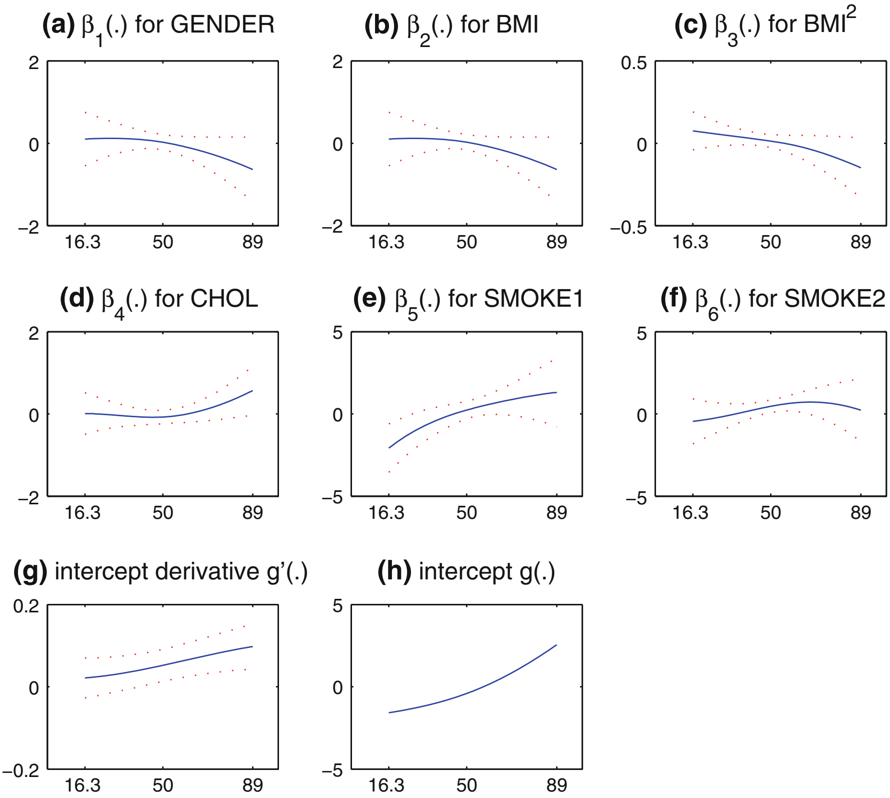

Fig. 3 The estimated coefficient functions (*solid lines*) and their 95 % confidence limits (*dashed lines*)

4 Conclusions

In the context of a marginal proportional hazards model that incorporates linear, nonlinear and interaction effects between covariates, we have demonstrated a profile local-partial likelihood approach that produces estimators with asymptotically valid results. The resultant estimators also have attractive precision properties in finite samples. In particular, inference performance of the estimators is comparable in bias magnitude and standard errors to estimators resulting from the local-partial likelihood approach while exhibiting the potential for smaller standard deviation and closer proximity of actual confidence interval coverage to nominal target coverage. In addition, the profile local-partial likelihood estimators exhibit robustness with respect to correlations among survival times. Overall, the model presented in this paper, which amalgamates several existing marginal proportional hazards models, represents a credible alternative that deserves further attention from both applied and theoretical statisticians. There are ways that the current lines of research can be extended that may result in an even more effective procedure. For example, variable selection could be introduced into the process and, when valid, could contribute to greater estimation efficiency. Work in progress considers this along the lines of Fan and Li (2001, 2002, 2008).

5 Appendix

This is a three-part Appendix organized as follows. Appendix 5.1 defines the notations and gives the technical assumptions used in the lemmas and the proofs of theorems. Appendix 5.2 provides the proofs of Theorems 1, 2 and 4, and Appendix 5.3 states the supporting lemmas that underline these proofs. The proofs of these lemmas and Theorem 5 are not included here as they are lengthy and tedious. We provide them in an online supplementary document available at http://personal.cb.cityu.edu.hk/msawan/research.htm.

5.1 Notations and assumptions

For ease of reference, we first define the notations and conditions to be used. Let $(\Upsilon, \mathcal{F}, \mathcal{P}_{(\alpha,\beta,g)})$ be a family of complete probability spaces with a history $\mathcal{F} = \{\mathcal{F}_t\}$ for an increasing right-continuous filtration $\mathcal{F}_t \subset \mathcal{F}$. We assume that W_{ij} is $\mathcal{F}_{t,ij}-$measurable, and $N_{ij}(u)$, $V_{ij}(u)$ and $Z_{ij}(u)$ are $\mathcal{F}-$adapted. Write $\mathcal{F}_{t,ij} = \sigma\{X_{ij} \le u, Z_{ij}(u), V_{ij}(u), W_{ij}, Y_{ij}(u), 0 \le u \le t\}$, $\mathcal{F}_t = \sigma\{X_{ij} \le u, Z_{ij}(u), V_{ij}(u), W_{ij}, Y_{ij}(u), i = 1, 2, \ldots, n, j = 1, \ldots, m, 0 \le u \le t\}$, and $M_{ij}(t) = N_{ij}(t) - \int_0^{\mathrm{T}} \lambda_{ij}(u)\mathrm{d}u, i = 1, 2, \ldots, n,\ \ j = 1, \ldots, m$. Obviously, $M_{ij}(t)$ is a $\mathcal{F}_t$ martingale, $j = 1, 2, \ldots, m$.

Let $\|\cdot\|$ denote the L_2-norm, and $\|\cdot\|_J$ be the sup-norm of a function or a process on a set J. The support of the random variable W is denoted by $\mathcal{W}$. For a compact subset J_W of $\mathcal{W}$, we define the neighborhood set of J_W as

$$J_{W,\varepsilon} = \{w : \inf_{w_0 \in J_W} |w - w_0| \le \varepsilon\}$$

for some $\varepsilon > 0$.

Here, we assume that Z is time-independent. The time-dependent covariate Z model can be similarly developed. To facilitate technical arguments, we reparametrize the profile local partial-likelihood (9) by the transformation $\varphi^* = \mathbf{H}_1\varphi$ and $\tilde{U}_{ij} = \mathbf{H}_1^{-1}\tilde{X}_{ij}$. Then, we have

$$\begin{aligned}\tilde{\ell}_n^*(\varphi^*, \alpha, \tau) = n^{-1} \sum_{j=1}^{m}\sum_{i=1}^{n} \int_0^{\tau} K_{h_1}(W_{ij} - w_0)(\varphi^{*T}\tilde{U}_{ij}(w_0) + \alpha^{\mathrm{T}} V_{ij}(u))\mathrm{d}N_{ij}(u) \\ -n^{-1} \sum_{j=1}^{m}\sum_{i=1}^{n} \int_0^{\tau} K_{h_1}(W_{ij} - w_0) \times \log\left\{\sum_{l=1}^{n} Y_{lj}(u) \exp(\varphi^{*T}\tilde{U}_{lj}(w_0)\right. \\ \left. + \alpha^{\mathrm{T}} V_{lj}(u)) K_{h_1}(W_{lj} - w_0)\right\} \mathrm{d}N_{ij}(u),\end{aligned} \tag{14}$$

and

$$\Phi_{njk}(u, \alpha, \varphi^*) = \frac{1}{n}\sum_{i=1}^{n} \tilde{S}_{ij}(u, \alpha, \varphi^*)(\tilde{U}_{ij}(w_0))^{\otimes k} K_{h_1}(W_{ij} - w_0), \tag{15}$$

for $k = 0, 1, 2$, and $j = 1, \ldots, m$, with $\tilde{S}_{ij}(u, \alpha, \varphi^*) = Y_{ij}(u) \exp\{\varphi^{*^{\mathrm{T}}} \tilde{U}_{ij}(w_0) + \alpha^{\mathrm{T}} V_{ij}(u)\}$.

Given the identifiability condition $\hat{g}_p(0, \alpha) = 0$, we have $\hat{g}_p(w_0, \alpha) = \int_0^{w_0} \hat{g}'_p(w, \alpha)\mathrm{d}w$. Let $\chi_{n1}(w_0)$ and $\chi_{n2}(w_0)$ be the first derivatives of $\hat{g}_p(w_0, \alpha)$ and $\hat{\beta}_p^{\mathrm{T}}(w_0, \alpha)$ with respect to α_0, respectively. Also, let $\kappa_{n1}(w_0)$ and $\kappa_{n2}(w_0)$ be the second derivatives of $\hat{g}_p(w_0, \alpha)$ and $\hat{\beta}_p^{\mathrm{T}}(w_0, \alpha)$ with respect to α_0, respectively.

Then for any α in a neighborhood of α_0, by Taylor series expansion, we have

$$\hat{g}_p(w, \alpha) \approx \hat{g}_p(w, \alpha_0) + \chi_{n1}^{\mathrm{T}}(w_0)(\alpha - \alpha_0) + \frac{1}{2}(\alpha - \alpha_0)^{\mathrm{T}} \kappa_{n1}(w_0)(\alpha - \alpha_0), \text{ and}$$
$$\hat{\beta}_p(w, \alpha) \approx \hat{\beta}_p(w, \alpha_0) + \chi_{n2}^{\mathrm{T}}(w_0)(\alpha - \alpha_0) + \frac{1}{2}(\alpha - \alpha_0)^{\mathrm{T}} \kappa_{n2}(w_0)(\alpha - \alpha_0).$$

Furthermore, Let $C_{j2}(u, w_0)$ be a $(2p + 1) \times d$-dimensional matrix with the first $p \times d$-dimensional sub-matrix defined as $b_{j0}^{-2}(u)[b_{j0}(u)t_{j1}(u) - b_{j1}(u) \otimes t_{j0}(u)]$ and other elements 0. Also, let $C_{j3}(u, w_0)$ be a $(2p + 1) \times d$-dimensional matrix with the last $(p + 1) \times d$-dimensional sub-matrix defined as $\mu_2 b_{j0}^{-2}(u)\big([b_{j0}(u)\mathcal{D}_w[t_{j1}(u)] - \mathcal{D}_w[b_{j1}(u)] \otimes t_{j0}(u)]^{\mathrm{T}}, [b_{j0}(u)\mathcal{D}_w[t_{j0}(u)] - \mathcal{D}_w[b_{j0}(u)] \otimes t_{j0}(u)]^{\mathrm{T}}\big)^{\mathrm{T}}$ and other elements 0. Denote

$$C_2(w_0) = \sum_{j=1}^{m} \int_0^{\tau} C_{j2}(u, w_0) b_{j0}(u, w_0) \mathrm{d}\Lambda_{0j}(u),$$
$$C_3(w_0) = \sum_{j=1}^{m} \int_0^{\tau} C_{j3}(u, w_0) b_{j0}(u, w_0) \mathrm{d}\Lambda_{0j}(u),$$
$$\chi_1(W_{ij}) = h_1^{-1} \int_0^{W_{ij}} \mathrm{e}_1^{\mathrm{T}} \mathbb{A}^{*-1}(W) C_2(W) \mathrm{d}W + \int_0^{W_{ij}} \mathrm{e}_1^{\mathrm{T}} \mathbb{A}^{*-1}(W) C_3(W) \mathrm{d}W, \text{ and}$$
$$\chi_2(W_{ij}) = \mathbf{e}_p^{\mathrm{T}} \mathbb{A}^{*-1}(W_i) C_2(W_{ij}) + \mathbf{e}_p^{\mathrm{T}} \mathbb{A}^{*-1}(W_{ij}) h_1 C_3(W_{ij}),$$

where $\mathbf{e}_p$ is a $p \times 2p + 1$ matrix, with the first p diagonal elements equalling 1 and remaining elements 0, and e_1 is a $2p + 1$-order vector with the last element equalling 1 and remaining elements 0. Also, let

$$r_{jk}(u, \alpha_0) = E\{\rho(u, V_{\cdot j}(u), Z_{\cdot j}, W_{\cdot j})[V_{\cdot j}(u) + \chi_2^{\mathrm{T}}(W_{\cdot j}) Z_j + \chi_1^{\mathrm{T}}(W_{\cdot j})]^{\otimes k}\}$$

for $k = 0, 1, 2$.

Our proofs require the following technical assumptions:

C.1. The kernel function $K \geq 0$ is a bound, symmetric density function with compact support.

C.2. $nh_1 \to \infty$ and $h_1 \to 0$, as $n \to \infty$.

C.3. The density $f_j(\cdot)$ of W_{1j} is of compact support and has a bounded second derivative.

C.4. $\int_0^{\tau} \lambda_0(u)\mathrm{d}u < \infty$, for every $j = 1, \ldots, m$. The function $\beta_0(\cdot)$ and $g_0(\cdot)$ have continuous second-derivatives with $g_0(0) = 0$.

C.5. The conditional expectations

$$E[\tilde{S}_{ij}(u,\alpha,\varphi^*)\tilde{U}_{ij}^{\otimes k}(w)|w_0]$$

are equi-continuous in $w_0 \in \cup_{j=1}^m supp(f_j)$, for $k = 0, 1$ and $j = 1, \ldots, m$. For $k = 0$, this conditional expectation has a continuous second derivative with respect to w_0.

C.6. $b_{j1}(u, w_0) = E(Z|w_0)b_{j0}(u, w_0) + O_p(h_1)$ and $t_{j1}(u, w_0) = E(Z|w_0) \otimes t_{j0}(u, w_0) + O_p(h_1)$ for every $w_0 \in \cup_{j=1}^m supp(f_j)$,

C.7. There exists a neighborhood $\mathcal{A}$ of α_0 such that for $k = 0, 1, 2, 3$, and $j = 1, \ldots, m$,

$$E\left[\sup_{(\alpha,u)\in\mathcal{A}\times[0,\tau]} Y_{\cdot j}(u)||\tilde{X}_{1j}(u)||^k \exp\{\alpha^{\mathrm{T}} V_{\cdot j}(u)+\beta^{\mathrm{T}}(W_{\cdot j})Z_{1j}+g(W_{\cdot j})\}\right]<\infty.$$

C.8. The functions $r_{j0}(u,\cdot)$, $r_{j1}(u,\cdot)$ and $r_{j2}(u,\cdot)$ are continuous in $\alpha \in \mathcal{A}$, uniformly in $u \in [0,\tau]$; r_0 is bounded away from zero on $\mathcal{A} \times [0,\tau]$, and r_1 and r_2 are bounded on $\mathcal{A} \times [0,\tau]$. The matrices

$$I(\alpha_0) = \sum_{j=1}^m \int_0^\tau \left[\frac{r_{j2}(u,\alpha_0)}{r_{j0}(u,\alpha_0)} - \frac{r_{j1}^{\otimes 2}(u,\alpha_0)}{r_{j0}^2(u,\alpha_0)}\right] r_{j0}(u,\alpha_0)\mathrm{d}\Lambda_{0j}(u),$$

and

$$\Sigma^*(\alpha_0) = E\left[\sum_{j=1}^m \int_0^\tau \left\{\left[V_{\cdot j}(u) + \chi_2(W_{\cdot j})Z_j + \chi_1(W_{\cdot j}) - \frac{r_{j1}(u,\alpha_0)}{r_{j0}(u,\alpha_0)}\right]\right.\right.$$
$$\left.\left. -Q(u, W_{\cdot j})s(W_{\cdot j})\right\} \mathrm{d}M_{\cdot j}(u)\right]^2$$

are positive definite, where the definition of $Q(.,.)$ can be found in Lemma 2 in Appendix 5.3.

The above conditions are similar to those in Andersen and Gill (1982) and Fan et al. (1997). Conditions C.1–C.5 are standard conditions for local partial-likelihood nonparametric estimation. Conditions C.7–C.8 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.

Remark 3 It is easy to show that condition C.6 is satisfied as long as one of Z and V is a non-random matrix. Furthermore, let $Z_i \equiv 1,\ i = 1, \ldots p$, then by assuming that $E[\beta(w)] = 0_p$ for model identifiability, model (5) reduces to the semi-parametric additive Cox hazards model,

$$\lambda_{ij}(t)=\lambda_{0j}(t)\exp\left\{\sum_{k=1}^{p}\beta_k(W_{ij}(t))+\alpha^{\mathrm{T}}V_{ij}(t)+g(W_{ij}(t))\right\}.$$

Remark 4 To obtain a consistent estimator of the covariance matrix Ω of $\hat{\alpha}_p$, we first have to estimate $I(\alpha_0)$ and $\Sigma^*(\alpha_0)$ consistently. Let $\hat{F}_j(w)$ be the empirical distribution function of w based on the observed $\{W_{ij}\}_{i=1}^{n}$. Write

$$\begin{aligned}
\hat{b}_{jk}(u,W_{ij}) &= \hat{E}[Y_{\cdot j}(u)\exp\{\hat{\beta_p}^{\mathrm{T}}(W_{ij})Z_{\cdot j}+\hat{\alpha_p}^{\mathrm{T}}V_{\cdot j}+\hat{g_p}(W_{ij})\}Z_{\cdot j}^{\otimes k}|w=W_{ij}],\\
\hat{t}_{jk}(u,W_{ij}) &= \hat{E}[Y_{\cdot j}(u)\exp\{\hat{\beta_p}^{\mathrm{T}}(W_{ij})Z_{\cdot j}+\hat{\alpha_p}^{\mathrm{T}}V_{\cdot j}+\hat{g_p}(W_{ij})\}Z_{\cdot j}^{\otimes k}\otimes V_{\cdot j}(u)|w=W_{ij}],\\
\hat{C}_{j2}(u,w_0) &= \begin{pmatrix} \frac{\hat{b}_{j0}(u)\hat{t}_{j1}(u)-\hat{b}_{j1}(u)\otimes\hat{t}_{j0}(u)}{\hat{b}_{j0}^2(u)}\\ 0_{p\times d}\\ 0_d^{\mathrm{T}}\end{pmatrix},\\
\hat{C}_2(w_0) &= \sum_{j=1}^{m}\int_0^{\tau}\hat{C}_{j2}(u,w_0)\hat{b}_{j0}(u,w_0)d\hat{\Lambda}_{0jp}(u),\\
\hat{C}_{j3}(u,w_0) &= \mu_2\begin{pmatrix} 0_{p\times d}\\ \frac{\hat{b}_{j0}(u)\mathcal{D}_w[\hat{t}_{j1}(u)]-\mathcal{D}_w[\hat{b}_{j1}(u)]\otimes\hat{t}_{j0}(u)}{\hat{b}_{j0}^2(u)}\\ \frac{\hat{b}_{j0}(u)\mathcal{D}_w[\hat{t}_{j0}(u)]-\mathcal{D}_w[\hat{b}_{j0}(u)]\otimes\hat{t}_{j0}(u)}{\hat{b}_{j0}^2(u)}\end{pmatrix},\\
\hat{C}_3(w_0) &= \sum_{j=1}^{m}\int_0^{\tau}\hat{C}_{j3}(u,w_0)\hat{b}_{j0}(u,w_0)d\hat{\Lambda}_{0jp}(u),
\end{aligned}$$

where $\hat{E}(\cdot|\cdot)$ is a consistent estimator of $E(\cdot|\cdot)$ (e.g., the Nadaraya–Watson estimator, or the local linear estimator). Then, the plug-in estimators of $\chi_1(W_{ij})$ and $\chi_2(W_{ij})$ are

$$\begin{aligned}
&\hat{\chi}_1(W_{ij})=h_1^{-1}\int_0^{W_{ij}}\mathrm{e}_1^{\mathrm{T}}\hat{\mathbb{A}}^{*-1}(W)\hat{C}_2(W)\mathrm{d}W+\int_0^{W_{ij}}\mathrm{e}_1^{\mathrm{T}}\hat{\mathbb{A}}^{*-1}(W)\hat{C}_3(W)\mathrm{d}W, \textit{and}\\
&\hat{\chi}_2(W_{ij})=\mathbf{e}_p^{\mathrm{T}}\hat{\mathbb{A}}^{*-1}(W_i)\hat{C}_2(W_{ij})+\mathbf{e}_p^{\mathrm{T}}\hat{\mathbb{A}}^{*-1}(W_{ij})h_1\hat{C}_3(W_{ij}),
\end{aligned}$$

respectively. Let the empirical estimator of $r_k(u,\alpha_0)$ be

$$\begin{aligned}
\hat{r}_{jk}(u,\alpha_0)=\frac{1}{n}\sum_{i=1}^{n}&Y_{ij}(u)\exp\{\hat{\beta}_p^{\mathrm{T}}(W_{ij})Z_{ij}+\hat{\alpha}_p^{\mathrm{T}}V_{ij}+\hat{g}_p(W_{ij})\}[V_{ij}+\hat{\chi}_2^{\mathrm{T}}(W_{ij})Z_{ij}\\
&+\hat{\chi}_1^{\mathrm{T}}(W_{ij})]^{\otimes k}.
\end{aligned}$$

Then, the empirical estimator of $I(\alpha_0)$ is

$$\hat{I}(\alpha_0)=\frac{1}{n}\sum_{j=1}^{m}\sum_{i=1}^{n}\Delta_{ij}\left[\frac{\hat{r}_{j2}(X_{ij})}{\hat{r}_{j0}(X_{ij})}-\frac{\hat{r}_{j1}^{\otimes 2}(X_{ij})}{\hat{r_0}^2(X_{ij})}\right].$$

The matrix $\hat{\Sigma}^*(\alpha_0)$ is defined as follows. Let the plug-in estimator of $s(w)$ be $\hat{s}(w) = \sum_{j=1}^m \int_{-\infty}^w \hat{a}(w^*)d\hat{F}_j(w^*)$, where

$$\begin{aligned}\hat{a}(w) &= \int_0^\tau \Big[\hat{E}[Y_{ij}(u)\exp\{\hat{\beta}_p^{\mathrm{T}}(W_{ij})Z_{ij} + \hat{\alpha}_p^{\mathrm{T}}V_{ij} + \hat{g}_p(W_{ij})\}|w \\ &= W_{ij}]\left\{V_{ij} + \hat{\chi}_2^{\mathrm{T}}(W_{ij})Z_{ij} + \hat{\chi}_1^{\mathrm{T}}(W_{ij}) - \frac{\hat{r}_{j1}(u,\alpha_0)}{\hat{r}_{j0}(u,\alpha_0)}\right\}\Big] \cdot \hat{r}_{j0}^{-1}(u,\alpha_0)d\bar{N}_{\cdot j}(u)\end{aligned}$$

is the plug-in estimator of $a(w)$. Set the empirical plug-in estimator to

$$\begin{aligned}\hat{\mathbf{M}}_{ij}(X_{ij}) = \Delta_{ij}\hat{G}_{ij}(X_{ij}) - \frac{1}{n}\sum_{k=1}^n \Delta_{kj}Y_{ij}(X_{Kj})\exp\{\hat{\beta}_p^{\mathrm{T}}(W_{kj})Z_{kj} + \hat{\alpha}_p^{\mathrm{T}}V_{kj} \\ +\hat{g}_p(W_{kj})\}\hat{r_{j0}}^{-1}(X_{kj},\alpha_0)\hat{G}_{ij}(X_{kj}),\end{aligned}$$

where $\hat{G}_{ij}(u) = V_{ij} + \hat{\chi}_2^{\mathrm{T}}(W_{ij})Z_{ij} + \hat{\chi}_1^{\mathrm{T}}(W_{ij}) - \frac{\hat{r}_{j1}(u,\alpha_0)}{\hat{r}_{j0}(u,\alpha_0)} - \hat{s}(W_{ij})\hat{Q}(u, W_{ij})$, with $\hat{Q}(u, W_{ij})$, the plug-in estimator, defined as

$$\hat{Q}(u, W_{ij}) = \mathrm{e}_1^{\mathrm{T}}\mu_2\left\{\hat{\mathbb{A}}^{*-1}(W_{ij})\begin{pmatrix} 0_p \\ \frac{\mathcal{D}_w[\hat{t}_{j1}(u,W_{ij})]}{\hat{t}_{j0}(u,W_{ij})} \\ \mathcal{D}_w[\log(\hat{t}_{j0}(u, W_{ij}))]\end{pmatrix} + [\hat{\mathbb{A}}^{*-1}(W_{ij})]'\begin{pmatrix} 0_p \\ Z_{ij} \\ 1\end{pmatrix}\right\}.$$

Then, the empirical estimator of Σ^* is $\hat{\Sigma}^* = \frac{1}{n}\sum_{i=1}^n[\sum_{j=1}^m \hat{\mathbf{M}}_{ij}(X_{ij})]^{\otimes 2}$.

Given the identifiability condition $\tilde{g}(0,\alpha) = 0$, we have $\tilde{g}(w_0,\alpha) = \int_0^{w_0}\tilde{g}'(w,\alpha)\mathrm{d}w$. Recall that the global profile partial-likelihood is (10). By Taylor series expansion around α_0, we have

$$\begin{aligned}\ell_p(\alpha) = \ell_p(\alpha_0) + (\alpha - \alpha_0)^{\mathrm{T}}\frac{\partial \ell_p(\alpha)}{\partial\alpha}\Big|_{\alpha=\alpha_0} + \frac{1}{2}(\alpha-\alpha_0)^{\mathrm{T}}\frac{\partial^2\ell_p(\alpha)}{\partial\alpha\partial\alpha^{\mathrm{T}}}\Big|_{\alpha=\alpha_0} \\ \times(\alpha-\alpha_0) + \mathcal{R}_n(\alpha^*), \qquad (16)\end{aligned}$$

where α^* lies between α and α_0, and

$$\mathcal{R}_n(\alpha^*) = \frac{1}{6}\sum_{j,k,l}(\alpha_j - \alpha_{0j})(\alpha_k - \alpha_{0k})(\alpha_l - \alpha_{0l})\left[\frac{\partial^3\ell_p(\alpha)}{\partial\alpha_j\partial\alpha_k\partial\alpha_l}|\alpha = \alpha^*\right], \quad (17)$$

with α_j and α_{0j} being the jth elements of α and α_0, respectively. It can be shown from Lemma 1 in Appendix 5.3 that $\frac{\partial^3\ell_p(\alpha)}{\partial\alpha_j\partial\alpha_k\partial\alpha_l}$ is bounded in probability, and hence $\mathcal{R}_n(\alpha) = O_p(\|\alpha-\alpha_0\|^3)$ for $\alpha \in \mathcal{A}$.

Note that $\varphi^{*T}\tilde{U}_{ij} = \beta^{\mathrm{T}}(W_{ij})Z_{ij} + g(W_{ij}) - g(w_0) + O_p(h_1^2)$. Simple algebra shows that

$$
\begin{aligned}
E[\Phi_{nj0}(u,\alpha_0,\varphi^*)] &= \mathrm{e}^{-g(w_0)} b_{j0}(u,w_0) + O(h_1^2),\\
E[\Phi_{nj1}(u,\alpha_0,\varphi^*)] &= \mathrm{e}^{-g(w_0)} \begin{pmatrix} b_{j1}(u,w_0)\\ 0_p \\ 0 \end{pmatrix} + h_1 \mathrm{e}^{-g(w_0)}\mu_2 \begin{pmatrix} 0_p \\ \mathcal{D}_w[b_{j1}(u,w_0)] \\ \mathcal{D}_w[b_{j0}(u,w_0)] \end{pmatrix}\\
&\quad + O(h_1^2),\\
E[\Phi_{nj2}(u,\alpha_0,\varphi^*)] &= \mathrm{e}^{-g(w_0)} \begin{pmatrix} b_{j2}(u,w_0) & 0_{p\times p} & 0_p \\ 0_{p\times p} & \mu_2 b_{j2}(u,w_0) & \mu_2 b_{j1}(u,w_0) \\ 0_p^{\mathrm{T}} & \mu_2 b_{j1}^{\mathrm{T}}(u,w_0) & \mu_2 b_{j0}(u,w_0) \end{pmatrix}\\
&\quad + h_1 \mathrm{e}^{-g(w_0)}\mu_2 \begin{pmatrix} 0_{p\times p} & \mathcal{D}_w[b_{j2}(u,w_0)] & \mathcal{D}_w[b_{j1}(u,w_0)] \\ \mathcal{D}_w[b_{j2}(u,w_0)] & 0_{p\times p} & 0_p \\ \mathcal{D}_w[b_{j1}^{\mathrm{T}}(u,w_0)] & 0_p^{\mathrm{T}} & 0 \end{pmatrix}\\
&\quad + O(h_1^2) \qquad (18)
\end{aligned}
$$

and $Var[\Phi_{nj0}(u,\alpha_0,\varphi^*)] = O(\frac{1}{nh_1})$, uniformly for $u \in [0,\tau]$. Then, using the same argument as for Lemma 1 of Fan et al. (1997), we obtain

$$
\sup_{0\le u\le\tau} \left\| \frac{\Phi_{nj1}(u,\alpha_0,\varphi^*)}{\Phi_{nj0}(u,\alpha_0,\varphi^*)} - \mathbf{C}_{j1}(u,w_0) \right\| \xrightarrow{P} 0, \qquad (19)
$$

and

$$
\frac{\Phi_{nj0}(u,\alpha_0,\varphi^*)\Phi_{nj2}(u,\alpha_0,\varphi^*) - \Phi_{nj1}^{\otimes 2}(u,\alpha_0,\varphi^*)}{\Phi_{nj0}^2(u,\alpha_0,\varphi^*)} = \mathbb{A}_j^*(u,w_0) + o_p(1) \qquad (20)
$$

uniformly for $u \in [0,\tau]$, where $\mathbf{C}_{j1}(u,w_0) = \left(\frac{b_{j1}^{\mathrm{T}}(u,w_0)}{b_{j0}(u,w_0)}, 0_p^{\mathrm{T}}, 0\right)^{\mathrm{T}}$, and

$$
\mathbb{A}_j^*(u,w_0) = \begin{pmatrix} \frac{b_{j2}(u,w_0)b_{j0}(u,w_0) - b_{j1}^{\otimes 2}(u,w_0)}{b_{j0}^2(u,w_0)} & 0_{p\times p} & 0_p \\ 0_{p\times p} & \mu_2 \frac{b_{j2}(u,w_0)}{b_{j0}(u,w_0)} & \mu_2 \frac{b_{j1}(u,w_0)}{b_{j0}(u,w_0)} \\ 0_p^{\mathrm{T}} & \mu_2 \frac{b_{j1}^{\mathrm{T}}(u,w_0)}{b_{j0}(u,w_0)} & \mu_2 \end{pmatrix}.
$$

5.2 Proofs of Theorems 1, 2 and 4

Proof of Theorem 1 By Lemma 2 in Appendix 5.3, $\frac{\partial \ell_p(\alpha)}{\partial \alpha}\ |_{\alpha=\alpha_0} \xrightarrow{P} 0$. Thus, with probability tending to one, for any small ε with a positive value, if $\alpha \in S_\varepsilon = \{\alpha : ||\alpha - \alpha_0|| < \varepsilon\}$, then

$$
(\alpha - \alpha_0)^{\mathrm{T}} \left[\left. \frac{\partial \ell_p(\alpha)}{\partial \alpha} \right|_{\alpha=\alpha_0} \right] \le \varepsilon. \qquad (21)
$$

Let a be the minimum eigenvalue of positive definitive matrix $I(\alpha_0)$. By Lemma 3 in Appendix 5.3, we have, for all $\alpha \in S_\varepsilon$,

$$(\alpha - \alpha_0)^{\mathrm{T}} \left[\frac{\partial^2 \ell_p(\alpha)}{\partial \alpha \partial \alpha^{\mathrm{T}}} \bigg|_{\alpha=\alpha_0} \right] (\alpha - \alpha_0) \leq -a\varepsilon^2 \tag{22}$$

with probability that tends to one. By the argument stated right after (17), with probability tending to one, there is a constant $C > 0$ such that

$$|\mathcal{R}_n(\alpha)| \leq C \cdot \varepsilon^3. \tag{23}$$

Upon substituting (21)–(23) in (16), we have, when ε is sufficiently small, with probability tending to one,

$$\ell_p(\alpha) - \ell_p(\alpha_0) \leq 0. \tag{24}$$

Therefore, $\ell_p(\alpha)$ has a local maximum in the interior of S_ε, and with probability tending to one, there exists a consistent estimator sequence $\hat{\alpha}$ of α_0 which maximizes the local profile partial-likelihood $\ell_p(\alpha)$. Then, the desired results holds.

Proof of Theorem 2 The result of Lemma 3 in Appendix 5.3 implies that $\frac{\partial^2 \ell_p(\alpha)}{\partial \alpha \partial \alpha^{\mathrm{T}}} \,|_{\alpha=\alpha_0} \xrightarrow{P} -I(\alpha_0)$. Note that $\hat{\alpha}_p$ is consistent. Using the above expression in (16), we have

$$\begin{aligned} \ell_p(\hat{\alpha}_p) = \ell_p(\alpha_0) + (\hat{\alpha}_p - \alpha_0)^{\mathrm{T}} \frac{\partial \ell_p(\alpha)}{\partial \alpha} \,|_{\alpha=\alpha_0} - \frac{1}{2}(\hat{\alpha}_p - \alpha_0)^{\mathrm{T}} I(\alpha_0)(\hat{\alpha}_p - \alpha_0) \\ + o_p\{(||\hat{\alpha}_p - \alpha_0|| + \frac{1}{\sqrt{n}})^2\}. \end{aligned} \tag{25}$$

Using Corollary 1 of Murphy and van der Vaart (2000) and Lemma 2 in Appendix 5.3, we obtain

$$\begin{aligned} &\sqrt{n}(\hat{\alpha}_p - \alpha_0) \\ &= I^{-1}(\alpha_0) \cdot \frac{1}{\sqrt{n}} \sum_{j=1}^{m} \sum_{i=1}^{n} \int_0^\tau \left\{ \left[V_{ij}(u) + \chi_2^{\mathrm{T}}(W_{ij}) Z_{ij} + \chi_1^{\mathrm{T}}(W_{ij}) - \frac{r_{j1}(u, \alpha_0)}{r_{j0}(u, \alpha_0)} \right] \right. \\ &\quad \left. - Q(u, W_{ij}) s(W_{ij}) \right\} \mathrm{d}M_{ij}(u) + o_p(1 + \sqrt{n}||\hat{\alpha}_p - \alpha_0||). \end{aligned}$$

Then by the martingale Central Limit Theorem and Slutsky Theorem, we have

$$\sqrt{n}(\hat{\alpha}_p - \alpha_0) \xrightarrow{\mathcal{D}} N(0, I^{-1}(\alpha_0) \Sigma^*(\alpha_0) I^{-1}(\alpha_0)).$$

The desired results thus hold.

Proof of Theorem 4 By (14), $\hat{\varphi}_p^* \equiv \hat{\varphi}_p^*(w_0, \hat{\alpha}_p)$ satisfies

$$\frac{\partial \tilde{\ell}_n^*(\hat{\varphi}_p^*, w_0, \tau)}{\partial \varphi^*} = n^{-1} \sum_{j=1}^{m} \sum_{i=1}^{n} \int_0^\tau K_{h_1}(W_{ij} - w_0) \times \left[\tilde{U}_{ij}(w_0) - \frac{\Phi_{nj1}(u, \hat{\alpha}_p, \hat{\varphi}_p^*)}{\Phi_{nj0}(u, \hat{\alpha}_p, \hat{\varphi}_p^*)} \right] \mathrm{d}N_{ij}(u) = 0. \tag{26}$$

It can be shown from the assumption $\hat{\alpha}_p - \alpha_0 = O_p(n^{-1/2})$ that

$$\sup_{u \in [0,\tau]} \left\| \frac{\Phi_{nj1}(u, \hat{\alpha}_p, \hat{\varphi}_p^*)}{\Phi_{nj0}(u, \hat{\alpha}_p, \hat{\varphi}_p^*)} - \frac{\Phi_{nj1}(u, \alpha_0, \hat{\varphi}_p^*)}{\Phi_{nj0}(u, \alpha_0, \hat{\varphi}_p^*)} \right\| = O_p(n^{-1/2}).$$

Thus, $\hat{\varphi}_p^*$ satisfies

$$n^{-1} \sum_{j=1}^{m} \sum_{i=1}^{n} \int_0^\tau K_{h_1}(W_{ij} - w_0) \left[\tilde{U}_{ij}(w_0) - \frac{\Phi_{nj1}(u, \alpha_0, \hat{\varphi}_p^*)}{\Phi_{nj0}(u, \alpha_0, \hat{\varphi}_p^*)} \right] \mathrm{d}N_{ij}(u) = O_p(n^{-1/2}).$$

We denote by $\hat{U}(\hat{\varphi}_p^*, w_0)$ the left-hand side of the above equation. Then, $\hat{U}(\hat{\varphi}_p^*, w_0) = o_p(1/\sqrt{nh_1})$. By Taylor series expansion, we obtain

$$\hat{U}(\varphi_0^*, w_0) + \frac{\partial \hat{U}(\tilde{\varphi}^*, w_0)}{\partial \varphi^*} (\hat{\varphi}_p^* - \varphi_0^*) = o_p(1/\sqrt{nh_1}), \tag{27}$$

where $\tilde{\varphi}^*$ lies between $\hat{\varphi}_p^*$ and φ_0^*, and hence $\tilde{\varphi}^* \to \varphi_0^*$ in probability. Simple algebra shows that

$$-\frac{\partial \hat{U}(\varphi_0^*, w_0)}{\partial \varphi^*} = \frac{1}{n} \sum_{j=1}^{m} \sum_{i=1}^{n} \int_0^\tau K_{h_1}(W_{ij} - w_0) \times \frac{\Phi_{nj0}(u, \alpha_0, \varphi_0^*) \Phi_{nj2}(u, \alpha_0, \varphi_0^*) - \Phi_{nj1}^{\otimes 2}(u, \alpha_0, \varphi_0^*)}{\Phi_{nj0}^2(u, \alpha_0, \varphi_0^*)} \mathrm{d}N_{ij}(u).$$

It follows from (20) that

$$
\begin{aligned}
-\frac{\partial \hat{U}(\varphi_0^*, w_0)}{\partial \varphi^*} &= \sum_{j=1}^{m} \int_0^{\tau} \mathbb{A}_j^*(u, w_0) dF_{w,j}(u) + o_p(1) \\
&= \begin{pmatrix} \sum_{j=1}^{m} [b_2 - \int_0^{\tau} \frac{b_{j1}^{\otimes 2}(u,w_0)}{b_{j0}(u,w_0)}] \mathrm{d}\Lambda_{0j}(u) & 0_{p\times p} & 0_p \\ 0_{p\times p} & \mu_2 b_2 & \mu_2 b_1 \\ 0_p^{\mathrm{T}} & \mu_2 b_1^{\mathrm{T}} & \mu_2 b_0 \end{pmatrix} \\
&= \begin{pmatrix} \Gamma^{*-1} & 0_{p\times p} & 0_p \\ 0_{p\times p} & \mu_2 b_2 & \mu_2 b_1 \\ 0_p^{\mathrm{T}} & \mu_2 b_1^{\mathrm{T}} & \mu_2 b_0 \end{pmatrix} = \mathbb{A}^*. \qquad (28)
\end{aligned}
$$

By the Doob–Meyer decomposition, we have

$$
\begin{aligned}
\hat{U}(\varphi_0^*, w_0) = & n^{-1} \sum_{j=1}^{m} \sum_{i=1}^{n} \int_0^{\tau} K_{h_1}(W_{ij} - w_0) \left[\tilde{U}_{ij}(w_0) - \frac{\Phi_{nj1}(u, \alpha_0, \varphi_0^*)}{\Phi_{nj0}(u, \alpha_0, \varphi_0^*)} \right] \mathrm{d}M_{ij}(u) \\
& + n^{-1} \sum_{j=1}^{m} \sum_{i=1}^{n} \int_0^{\tau} K_{h_1}(W_{ij} - w_0) \left[\tilde{U}_{ij}(w_0) - \frac{\Phi_{nj1}(u, \alpha_0, \varphi_0^*)}{\Phi_{nj0}(u, \alpha_0, \varphi_0^*)} \right] \\
& \times Y_{ij}(u) \lambda_0(u) \exp\{\beta_0^{\mathrm{T}}(W_{ij}) Z_{ij}(u) + \alpha_0^{\mathrm{T}} V_{ij}(u) + g_0(W_{ij})\} \mathrm{d}u \\
\triangleq & \sum_{j=1}^{m} \mathbf{d}_{nj}(\tau) + \mathbf{q}_n(\tau). \qquad (29)
\end{aligned}
$$

Note that

$$
\begin{aligned}
& \exp\{\alpha^{\mathrm{T}} V_{ij}(u) + \beta^{\mathrm{T}}(W_{ij}) Z_{ij}(u) + g(W_{ij})\} - \exp\{\alpha^{\mathrm{T}} V_{ij}(u) + \varphi^{*T} \tilde{U}_{ij}(w_0) + g(w_0)\} \\
& = \exp\{\alpha^{\mathrm{T}} V_{ij}(u) + \varphi^{*T} \tilde{U}_{ij}(w_0) + g(w_0)\} \frac{1}{2} \{[\beta''^{\mathrm{T}}(w_0) Z_{ij}(u) + g''(w_0)] \\
& \quad \times (W_{ij} - w_0)^2 + o_p(h_1^2)\}(1 + O_p(h_1^4)).
\end{aligned}
$$

Then,

$$
\begin{aligned}
\mathbf{q}_n(\tau) = & \frac{1}{2n} \sum_{j=1}^{m} \sum_{i=1}^{n} \int_0^{\tau} K_{h_1}(W_{ij} - w_0) \left[\tilde{U}_{ij}(w_0) - \frac{\Phi_{nj1}(u, \alpha_0, \varphi_0^*)}{\Phi_{nj0}(u, \alpha_0, \varphi_0^*)} \right] \\
& \times \exp\{\alpha_0^{\mathrm{T}} V_{ij}(u) + \varphi_0^{*T} \tilde{U}_{ij}(w_0) + g_0(w_0)\} \\
& \times Y_{ij}(u) \lambda_{0j}(u) \{[\beta_0''^{\mathrm{T}}(w_0) Z_{ij}(u) + g_0''(w_0)](W_{ij} - w_0)^2\} \mathrm{d}u + o_p(h_1^2).
\end{aligned}
$$

Hence we have

$$\begin{aligned}\mathbf{q}_n(\tau) &= \frac{h_1^2}{2}\mu_2 \begin{pmatrix} \Gamma^{*-1} & 0_{p\times p} & 0_p \\ 0_{p\times p} & 0_{p\times p} & 0_p \\ 0_p^{\mathrm{T}} & 0_p^{\mathrm{T}} & 0 \end{pmatrix} \varphi_0''(w_0) + o_p(h_1^2) \\ &\triangleq \mathbb{B}_n^*(\tau, w_0) + o_p(h_1^2), \end{aligned} \tag{30}$$

where $\varphi_0''(w_0) = (\beta_0''(w_0), 0_p, g_0''(w_0))^{\mathrm{T}}$. Let $\mathbf{d}_{nj}^*(\tau) = \sqrt{nh_1}\mathbf{d}_{nj}(\tau)$. Then, combined with (19), we have

$$\mathrm{Var}\left(\sum_{j=1}^m \mathbf{d}_{nj}^*(\tau)\right) = \Pi^*(\tau, w_0) = \Pi_1^*(\tau, w_0) + \Pi_2^*(\tau, w_0),$$

where

$$\begin{aligned}\Pi_1^*(\tau, w_0) &= \sum_{j=1}^m \Pi_{j1}^*(\tau, w_0) = \sum_{j=1}^m \lim_{n\to\infty} E\langle \mathbf{d}_{nj}^*, \mathbf{d}_{nj}^*\rangle(\tau) \\ &= \begin{pmatrix} \nu_0\Gamma^{*-1} & 0_{p\times p} & 0_p \\ 0_{p\times p} & \nu_2 b_2 & \nu_2 b_1 \\ 0_p^{\mathrm{T}} & \nu_2 b_1^{\mathrm{T}} & \nu_2 b_0 \end{pmatrix},\end{aligned}$$

and

$$\begin{aligned}\Pi_2^*(\tau, w_0) = \lim_{n\to\infty} E\Bigg\{ h_1 \sum_{k\neq l} \int_0^\tau K_{h_1}(W_{1k}-w_0)[\tilde{U}_{1k} - \frac{\Phi_{nk1}(u, \alpha_0, \varphi_0^*)}{\Phi_{nk0}(u, \alpha_0, \varphi_0^*)}]\mathrm{d}M_{1k}(u) \\ \times \int_0^\tau K_{h_1}(W_{1l}-w_0)[\tilde{U}_{1l}^{\mathrm{T}} - \frac{\Phi_{nl1}^{\mathrm{T}}(u, \alpha_0, \varphi_0^*)}{\Phi_{nl0}(u, \alpha_0, \varphi_0^*)}]\mathrm{d}M_{1l}(u)\Bigg\}.\end{aligned}$$

By conditions C.1–C.8, and results from Andersen and Gill (1982), it is easy to show that the Lindeberg condition of $\mathbf{d}_n^*(\tau)$ holds. Then, we have

$$\sqrt{nh_1}(\hat{U}(\varphi_0^*, w_0) - \mathbb{B}_n^*(\tau, w_0)) \xrightarrow{L} N(0, \Pi^*(\tau, w_0)).$$

Together with (27) and (28), this leads to

$$\sqrt{nh_1}(\mathbf{H}_1(\hat{\varphi}_p - \varphi_0) - \frac{h_1^2\mu_2}{2}(\beta_0''^{T}(w_0), 0_p^{\mathrm{T}}, 0)^{\mathrm{T}}) \xrightarrow{L} N(0, \mathbf{V}(w_0)),$$

where

$$\mathbf{V}(w_0) = \begin{pmatrix} \nu_0\Gamma^* & 0_{p\times p+1} \\ 0_{p+1\times p} & Q^* \end{pmatrix} + \mathbb{A}^{*-1}\Pi_2^*(\tau, w_0)\mathbb{A}^{-1*T}$$

with

$$Q^* = \nu_2 \mu_2^{-2} \begin{pmatrix} b_2 & b_1 \\ b_1^{\mathrm{T}} & b_0 \end{pmatrix}^{-1}.$$

The proof of Theorem 4 is completed.

5.3 Lemmas for the proofs of main results

Let

$$C_{nj}(t) = n^{-1} \sum_{i=1}^{n} Y_{ij}(t) g(W_{ij}, (W_{ij} - w_0)/h, Z_{ij}(t), V_{ij}(t)) K_{h_1}(W_{ij} - w_0)$$

for a function $g(\cdot,\cdot,\cdot,\cdot)$. The following Lemmas 1 and 3 are required for proving Theorems 1, 2 and 4, while Lemma 2 is required for proving Theorems 1 and 2.

Lemma 1 *Assume that conditions* C.1 *and* C.4 *hold. Suppose that $g(\cdot,\cdot,\cdot,\cdot)$ is continuous in all of its four arguments, and $E(g(W, u, Z(t), V(t))|W = w_0)$ is continuous at the point w_0. If $h_1 \to 0$ so that $nh_1/\log n \to \infty$, then*

$$\sup_{0\le t\le\tau} |C_{nj}(t) - C_j(t)| \xrightarrow{P} 0,$$

where $C_j(t) = f_j(w_0) \int E[Y(t) g(w_0, u, Z_j(t), V_j(t))|W_j = w_0] K(u)\mathrm{d}u$, $j = 1, \ldots, m$.

Proof The proof is similar to the proof of Lemma 1 in Fan et al. (2006), and is omitted here for brevity.

Lemma 2 *Assume the conditions* C.1–C.8 *hold. If $nh_1^2 \to \infty$ and $nh_1^4 \to 0$, then*

$$\begin{aligned}
&\sqrt{n}\frac{\partial \ell_p(\alpha)}{\partial \alpha}|_{\alpha=\alpha_0} \\
&= \frac{1}{\sqrt{n}} \sum_{j=1}^{m}\sum_{i=1}^{n} \int_0^{\tau} \left\{ \left[V_{ij}(u) + \chi_2^{\mathrm{T}}(W_{ij}) Z_{ij} + \chi_1^{\mathrm{T}}(W_{ij}) - \frac{r_{j1}(u,\alpha_0)}{r_{j0}(u,\alpha_0)} \right] \right. \\
&\quad \left. - Q(u, W_{ij}) s(W_{ij}) \right\} \mathrm{d}M_{ij}(u) + o_p(1),
\end{aligned}$$

where $Q(u, W_{ij}) = \mathrm{e}_1^{\mathrm{T}} \mu_2 \{\mathbb{A}^{*-1}(W_{ij}) \begin{pmatrix} 0_p \\ \frac{\mathcal{D}_w[t_{s1}(u, W_{ij})]}{t_{s0}(u, W_{ij})} \\ \mathcal{D}_w[\log(t_{j0}(u, W_{rs}))] \end{pmatrix} + [\mathbb{A}^{*-1}(W_{ij})]' \begin{pmatrix} 0_p \\ Z_{ij} \\ 1 \end{pmatrix}\}$ *and* $s(W_{ij}) = \sum_{l=1}^{m} \int_{-\infty}^{W_{ij}} a_l(W_{ij}) f_l(W_{ij}) \mathrm{d}w$.

Proof See the online supplementary document at http://personal.cb.cityu.edu.hk/msawan/research.htm.

Lemma 3 *Suppose that conditions* C.1–C.8 *hold. Then,*

$$\left.\frac{\partial^2 \ell_p(\alpha)}{\partial\alpha\,\partial\alpha^{\mathrm{T}}}\right|_{\alpha=\alpha_0} \xrightarrow{P} -I(\alpha_0).$$

Proof See the online supplementary document at http://personal.cb.cityu.edu.hk/msawan/research.htm.

Acknowledgments This work was partially supported by the National Natural Science Fund for Distinguished Young Scholar (No. 70825004), the National Natural Science Foundation of China (NSFC) (Nos. 10731010, 11301424 and 10628104), the National Basic Research Program (No. 2007CB814902), Creative Research Groups of China (No. 10721101), and Leading Academic Discipline Program, 211 Project for Shanghai University of Finance and Economics (the 3rd phase) (No. B803). The authors thank the editor, an associate editor and two referees for helpful comments and suggestions. The usual disclaimer applies.

References

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

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

Cai, J., Fan, J., Jiang, J., Zhou, H. (2007a). Partially linear hazard regression for multivariate survival data. *Journal of the American Statistical Association*, *102*, 538–551.

Cai, J., Fan, J., Zhou, H., Zhou, Y. (2007b). Hazard models with varying coefficients for multivariate failure time data. *Annals of Statistics*, *35*, 324–354.

Cai, J., Fan, J., Jiang, J., Zhou, H. (2008). Partially linear hazard regression with varying coefficients for multivariate survival data. *Journal of the Royal Statistical Society (Series B)*, *70*, 141–158.

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

Fan, J., Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*, 1348–1360.

Fan, J., Li, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. *Annals of Statistics*, *30*, 74–99.

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

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

Hastie, T., Tibshirani, R. (1993). Varying-coefficient models (with discussion). *Journal of the Royal Statistical Society (Series B)*, *55*, 757–796.

Knuiman, M.W., Cullen, K.J., Bulsara, M.K., Welborn, T.A., Hobbs, M.S.T. (1994). Mortality trends 1965 to 1989 in Busselton, the site of repeated health surveys and interventions. *Australian Journal of Public Health*, *18*, 129–135.

Li, R., Liang, H. (2008). Variable selection in semiparametric regression modeling. *Annals of Statistics*, *36*, 261–286.

Liang, K.Y., Self, S.G., Chang, Y.C. (1993). Modelling marginal hazards in multivariate failure time data. *Journal of the Royal Statistical Society (Series B)*, *55*, 441–453.

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

Prentice, R.L., 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 econometrics and statistics. *Econometrica*, *56*, 531–547.

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

Tian, L., Zucker, D., Wei, L.J. (2005). On the Cox model with time-varying regression coefficients. *Journal of the American Statistical Association*, *100*, 172–183.

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