Hazard Models with Varying-coefficients for Multivariate Failure Time Data *

JIANWEN CAI1, JIANGQING FAN2,3, HAIBO ZHOU1 AND YONG ZHOU3

1 Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420
2 Department of Operations Research and Financial Engineering, Princeton University, NJ 08540
3 Institute of Applied Mathematics and Center of Statistics, Academy of Mathematics and System Sciences, Chinese Academy of Science

Abstract

Statistical estimation and inference for the marginal hazard models with varying-coefficients for multivariate failure time data are important subjects in survival analysis. A local pseudo-partial likelihood procedure is proposed for estimating the unknown coefficient functions. A weighted average estimator is also proposed in an attempt to improve the efficiency of the estimator. Consistency and the asymptotic normality of the proposed estimators are established and standard error formulas for the estimated coefficients are derived and empirically tested. To reduce the computational burden of the maximum local pseudo-partial likelihood estimator, a simple and useful one-step estimator is proposed. Statistical properties of the one-step estimator are established and simulation studies are conducted to compare the performance of the one-step estimator to the maximum local pseudo-partial likelihood estimator. The results show that the one-step estimator can save computational cost without deteriorating its performance both asymptotically and empirically and that the optimal weighted average estimator is more efficient than the maximum local pseudo-partial likelihood estimator. A data set from the Busselton Population Health Surveys is analyzed to illustrate our proposed methodology.

Keywords: Local pseudo-partial likelihood, Marginal hazard model, Martingale, Multivariate failure time, One-step estimator, Varying coefficients.

This research was partially supported by NIH grant R01 HL69720. Fan’s research was also partially supported by NSF grants DMS-0355179 and DMS-0354223, and a RGC grant CUHK4262/01P of HKSAR. Y. Zhou’s research was also partially supported by the Fund of National Natural Science (No.10471140) of China. The authors thank Dr. Matthew Knuiman and the Busselton Population Medical Research Foundation in Western Australia for providing the data used in the illustration and are grateful to the AE and referees for helpful comments.
1 Introduction

Multivariate failure time data are encountered in many biomedical studies when related subjects are at risk of a common event or study subject is at risk of different types of events or recurrence of the same event. Some examples are: in epidemiological cohort studies in which the ages of disease occurrence are recorded for members of families; in animal experiments where treatments are applied to samples of litter mates; in clinical trials in which individual study subjects are followed for the occurrence of multiple events; or, in intervention trials involving group randomization. A common feature of the data in these examples is that the failure times might be correlated. For example, in clinical trials where the patients are followed for repeated recurrent events, the times between recurrences for a given patient may be correlated.

When there is at most one event for each subject and these subjects are mutually independent, the Cox (1972) proportional hazards model has commonly been used to assess the effects of covariate on failure times. For multivariate failure time data, research efforts have been concentrated on marginal hazards models and frailty models. Related literature includes, but not limited to, Wei, Lin and Weissfeld (1989), Lin (1994), Cai and Prentice (1995, 1997), Spiekerman and Lin (1998), for the marginal models and Clayton and Cuzick (1985), Anderson and Louis (1995), Oakes and Jeong (1998), and Fan and Li (2002), for the frailty models. The statistical methods developed for dealing with the failure time data typically assume that the covariate effects on the logarithm of the hazard function are linear and the regression coefficients are constants. These assumptions, however, are mainly chosen for their mathematical convenience. True associations in practical studies are usually more complex than a simple linear model.

An important extension of the standard regression model with constant coefficient is the varying coefficient model. The varying coefficient model addresses an issue frequently encountered by investigators in practical studies. For example, the effect of an exposure variable on the hazard function may change with the level of a confounding covariate. This is traditionally modelled by including an interaction term in the model. Such an approach is a simplification of the true underlying association, since a cross product of the exposure and the confounding variable means
that the effect of the exposure on the hazard function changes linearly with the confounding variable. In many studies, however, investigators believe that the rate of change is not linear and they would like to examine how each level of the exposure interacts with confounding variables. For example, in the Cancer Risk in Uranium Miners Study (Sandler et al. 1999), the radon exposure for uranium miners is measured by the working level months for 16,434 underground miners from 1945-1976 in the Czech Republic. The mining industry’s working place safety measurements which affect the true inhalation of the radon gas, such as ventilation condition, is changing over the last 50 years. Therefore the effect of a fixed amount of exposure in the 50’s should not be treated the same as in the 70’s. How to handle this issue is of current active research interest in epidemiology. This leads to a general varying-coefficient model where the coefficient for the radon exposure is a function of the calendar year, and this function can be nonlinear over time. Parametric models for the varying-coefficient functions can be most efficient if the underlying functions are correctly specified. However, misspecification may cause serious bias and the model constraints may distort the trend in local areas. Non-parametric modelling is appealing in these situations.

Varying-coefficient models have been studied in many non-failure time data settings such as multi-dimensional nonparametric regression, generalized linear models, analysis of longitudinal data and nonlinear time series. It is particularly appealing for models in longitudinal studies because it allows one to explore the extent to which covariates effect changing over time. Related literature includes Hastie and Tibshirani (1993), Carroll, Ruppert, and Welsh (1998), and Cai, Fan and Li (2000). For the univariate survival time, the time-varying effect has been carefully studied by Murphy (1993), Cai and Sun (1999) and Tian, Zuker and Wei (2002). Applications of the varying-coefficients models to survival analysis, particularly in the context of multivariate failure time, remain to be studied. New technical challenges arise in dealing with within-cluster dependence and the varying effects of an exposure variable. The local pseudo-partial likelihood in our setting is more sophisticated than that based on the time-varying model. In fact, the latter is no longer a proportional hazards model.

In this paper, we study the marginal hazards model with varying coefficient for multivariate failure time data. The rest of this paper is organized as follows. In Section 2, we formulate the varying coefficient model and propose local pseudo-partial likelihood procedures for coefficient functions. We also present the asymptotic prop-
erties and propose a variance estimator, and we consider a computationally efficient one-step procedure and show that it is asymptotically equivalent to the local pseudo-partial likelihood estimator. In Section 3, we propose a weighted average approach to estimate the coefficient functions. We evaluate the proposed procedures through simulation studies and illustrate the proposed approach via an application to the Busselton Population Health Surveys data set in Section 4. Final remarks are given in Section 5. Proofs of theoretical results are given in Section 6.

2 Marginal hazards model with varying coefficients

Suppose that there is a random sample of $n$ clusters from an underlying population and that there are $J$ members in each cluster. Let $i$ indicate cluster, $(i, j)$ denote the $j$th member in the $i$th cluster, and let $X_{ij}$ denote the observed time, for the $(i, j)$ member. Let $T_{ij}$ and $C_{ij}$ denote the failure time and censoring time, respectively, and $X_{ij} = \min(T_{ij}, C_{ij})$ the observed time, for the $(i, j)$ member $(i = 1, \ldots, n, j = 1, \ldots, J)$. Let $\Delta_{ij}$ be an indicator which equals 1 if $X_{ij}$ is a failure time and 0 otherwise. Varying cluster size can be accommodated by defining $T_{ij} = C_{ij} = 0$. Let $F_{ij}(t)$ represent the failure, censoring and covariate information up to time $t$ for the $(i, j)$ member as well as the covariate information of the other members in the $i$th cluster up to time $t$. The marginal hazard function is defined as

$$
\lambda_{ij}(t; F_{ij}(t)) = \lim_{h \rightarrow 0} \frac{1}{h} P[T_{ij} \leq t + h | T_{ij} > t, F_{ij}(t)].
$$

The observed data structure is given by $\{X_{ij}, \Delta_{ij}, Z_{ij}(t), V_{ij}(t)\}$ for $i = 1, \ldots, n$, where $Z_{ij}(t) = (Z_{ij1}(t), \ldots, Z_{ijp}(t))^T$ and $V_{ij}(t)$ are two types of covariates, with $V$ being an exposure variable of interest. We assume that the censoring times are independent of the failure times conditional on the covariates and the observation period is $[0, \tau]$, where $\tau$ is a constant denoting the study ending time.

To explore the extent to which the hazard regression function interacts with different level of a covariate variable $V$, we consider the following varying-coefficient model

$$
\lambda_{ij}(t; F_{ij}) = \lambda_{0j}(t) \exp\{\beta(V_{ij}(t))^T Z_{ij}(t) + g(V_{ij}(t))\},
$$

(1)

where $\lambda_{0j}(\cdot)$ is an unspecified baseline hazard function pertaining to the $j$th member of each response vector, $\beta(\cdot)$ is the regression coefficient vector that may be a function of the covariate $V_{ij}$, $g(\cdot)$ is a nonlinear effect of $V_{ij}$, and both $\beta(\cdot)$ and $g(\cdot)$ are unspecified continuously differentiable functions. Let $N_{ij}(t) = I(X_{ij} \leq t, \Delta_{ij} = 1)$ denote the counting process corresponding to $T_{ij}$ and $Y_{ij}(t) = I(X_{ij} \geq t)$ denote the at risk indicator process. Set $M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(s) \lambda_{ij}(s) ds$. 

4
Note that $M_{ij}(t)$ is a martingale with respect to the marginal filtration $\mathcal{F}_{t,ij} = \sigma\{N_{ij}(s^-), Y_{ij}(s), Z_{ij}(s), 0 \leq s \leq t\}$ and the union $\sigma\{-\} \cup_{i=1}^n \mathcal{F}_{t,ij} = \sigma\{N_{ij}(s^-), Y_{ij}(s), Z_{ij}(s), 0 \leq s \leq t, i = 1, 2, \cdots, n\}$, respectively. However, $M_{ij}(t)$ ($i = 1, 2, \cdots, n, j = 1, 2, \cdots, J$) is no longer a martingale with respect to the entire union $\sigma\{-\} \cup_{i=1}^n \cup_{j=1}^J \mathcal{F}_{t,ij} = \sigma\{N_{ij}(s^-), Y_{ij}(s), Z_{ij}(s), 0 \leq s \leq t, i = 1, 2, \cdots, n, j = 1, 2, \cdots, J\}$ because the observations within a cluster might be dependent.

For ease of presentation, we drop the dependence of covariates on the time $X_{ij}$, with the understanding that the methods and proofs in this paper are applicable to external time-dependent covariates (Kalbfleisch and Prentice, 2002). If all the observations are independent, the partial likelihood for model (1) is

$$L(\beta(\cdot), g(\cdot)) = \prod_{j=1}^J \prod_{i=1}^n \left\{ \frac{\exp\{\beta(V_{ij})^T Z_{ij} + g(V_{ij})\}}{\sum_{l \in \mathcal{R}_j(X_{ij})} \exp\{\beta(V_{lj})^T Z_{lj} + g(V_{lj})\}} \right\}^{\Delta_{ij}}, \quad (2)$$

where $\mathcal{R}_j(t) = \{i : X_{ij} \geq t\}$ denote the set of the individuals at risk just prior to time $t$. Since the observations within a cluster are not independent, we refer to (2) as the pseudo-partial likelihood. Wei, Lin, and Weissfeld (1989) considered the parametric counter part for (2).

### 2.1 Local pseudo-partial likelihood estimation

If the unknown functions $\beta(\cdot)$ and $g(\cdot)$ are parameterized, the parameters can be estimated by maximizing (2). For our nonparametric estimation, the forms of the unknown functions are not available. Directly solving the pseudo-partial likelihood (2) for the unknown functions $\beta(\cdot)$ and $g(\cdot)$ is hardly possible because of the infinite dimension of the unknown parameters. We choose to use the local polynomial method for our nonlinear modelling for the following reasons. Firstly, it is relatively easy to program because existing software on parametric fitting can be modified, via introducing a weighting scheme, to deal with local parametric problems. Secondly, the sampling properties of the local polynomial fitting can be derived and efficient semiparametric estimators can be constructed.

Assume that each function in the component of $\beta(\cdot)$ and $g(\cdot)$ are smooth so that they admit Taylor’s expansions: for each given $v$ and $u$, where $u$ is close to $v$,

$$\beta(u) \approx \beta(v) + \beta'(v)(u-v) \equiv \delta + \eta(u-v),$$
$$g(u) \approx g(v) + g'(v)(u-v) \equiv \alpha + \gamma(u-v), \quad (3)$$

where $\beta'(u) = d\beta(u)/du$. Substituting these local models into (2), we obtain the
following logarithm of the local pseudo-partial likelihood:

\[
\ell(\gamma, \delta, \eta) = \sum_{j=1}^{J} \sum_{i=1}^{n} K_h(V_{ij} - v) \Delta_{ij} \left\{ \delta^T Z_{ij} + \eta^T Z_{ij} (V_{ij} - v) + \gamma (V_{ij} - v) \right\} - \log \left( \sum_{l \in R_j(X_{ij})} \exp \left( \delta^T Z_{lj} + \eta^T Z_{lj} (V_{lj} - v) + \gamma (V_{lj} - v) \right) K_h(V_{lj} - v) \right), \tag{4}
\]

where \( K_h(\cdot) = K(\cdot/h)/h \), \( K(\cdot) \) is a probability density called a kernel function, and \( h \) represents the size of the local neighborhood called a bandwidth. The kernel weight is introduced to confine the fact that the local model (3) is only applied to the data around \( v \).

Using counting process notation and letting \( X_{ij}^* = (Z_{ij}^T, Z_{ij}^T (V_{ij} - v), V_{ij} - v)^T \), and \( \xi = (\delta^T, \eta^T, \gamma)^T \), the local pseudo-partial likelihood function (4) can be expressed as \( n \cdot \ell_n(\xi, \infty) \), where

\[
\ell_n(\xi, \tau) = n^{-1} \sum_{j=1}^{J} \sum_{i=1}^{n} \int_{0}^{\tau} K_h(V_{ij} - v) \left[ \xi^T X_{ij}^* \right] \left\{ \sum_{l=1}^{n} Y_{ij}(w) \exp(\xi^T X_{ij}^*) K_h(V_{ij} - v) \right\} dN_{ij}(w). \tag{5}
\]

Maximizing \( \ell(\gamma, \delta, \eta) \) in (4) is equivalent to maximizing \( \ell_n(\xi, \tau) \) in (5). For a technical reason, following the work in the literature, we maximize (5) for a given finite \( \tau \).

Let \( \hat{\xi}(v) = (\hat{\gamma}(v)^T, \hat{\delta}(v)^T, \hat{\eta}(v)^T)^T \) be the maximizer of (5). Then, \( \hat{\beta}(v) = \hat{\delta}(v) \) is a local linear estimator for the coefficient function \( \beta(\cdot) \) at the point \( v \). Similarly, an estimator of \( \gamma'(\cdot) \) at the point \( v \) is simply the local slope \( \hat{\gamma}(v) \), namely \( \hat{\gamma}(v) = \gamma(v) \). The curve \( \hat{\gamma}(\cdot) \) can be estimated by integration on the function \( \hat{\gamma}'(v) \). Following Hastie and Tibshirani (1990), the integration can be approximated by using the trapezoidal rule. The local pseudo-partial likelihood estimator in (4) is particularly easy to compute. It can be implemented by using existing software such as SAS or Splus with the Cox regression procedure. The only difference is that one needs to incorporate the kernel weights in the Cox regression and repeatedly applies the procedure at a grid of points in the range of the variable \( V \).

### 2.2 Assumptions and Notation

To express explicitly asymptotic bias and asymptotic variance of the estimator, we introduce some necessary assumptions and notation. Let \( \mu_i = \int x' K(x) dx \)
and \( \nu_i = \int x^i K^2(x) dx \) for \( i = 0, 1, 2 \). Denote \( P(w, z, v) = P(X \geq w | Z = z, V = v) \) and \( \rho(w, z, v) = P(w, z, v) \exp \{ \beta_0(v)^T z + g_0(v) \} \). For \( k = 0, 1, 2 \), define \( a_{jk}(w, v) = f_j(v) E \{ \rho(w, Z_j, v)Z_j^{\otimes k} | V_j = v \} \), where \( f_j(\cdot) \) is the density of \( V_j \) and \( Z^{\otimes k} = 1, Z \) and \( ZZ^T \) for \( k = 0, 1, \) and \( 2 \), respectively. Let \( a_{jk}(v) = \int_0^\tau a_{jk}(w, v)dw \) and \( a_k(v) = \sum_{j=1}^J a_{jk}(v) \). We will drop the dependence of \( a_k(w, v), a_{jk}(w, v), a_{jk}(v) \) and \( a_k(v) \) on \( v \) when there is no ambiguity. Finally, let

\[
\Gamma = \Gamma(v) = \left\{ \sum_{j=1}^J \left( a_{j2} - \int_0^\tau a_{j1}(w) a_{j0}(w)^{-1} \lambda_{0j}(w) dw \right) \right\}^{-1},
\]

and

\[
Q = \begin{pmatrix}
Q_1^{-1} & -\left( a_0^{-1}\right)^T Q_1^{-1} a_1 \\
-a_1^T Q_1^{-1} a_0^{-1} & \left( a_0 - a_1^T a_2 a_1^{-1} \right)^{-1}
\end{pmatrix},
\]

where \( Q_1 = a_2 - a_1 a_1^T a_0^{-1} \).

Let \( || \cdot || \) denote the \( L_2 \)-norm and \( || \cdot ||_\Phi \) be the sup-norm of a function or a process on a set \( \Phi \). The support of the random variable \( V \) is denoted by \( V \). For a compact subset \( \Phi_\nu \) of \( V \), we define the neighborhood set of \( \Phi_\nu \) as \( \Phi_{\nu, \varepsilon} = \{ u : \inf_{v \in \Phi_\nu} | u - v | \leq \varepsilon \} \) for some \( \varepsilon > 0 \). For \( k = 0, 1, 2 \), let

\[
s_{jk}(w, \zeta, v) = f_j(v) \int \left[ P(w, Z_j, \nu) \Delta(y, Z_j, \nu) R_j^{\otimes k}(y, y) | V_j = v \right] K(y) dy,
\]

where \( R_j(y, w) = (Z_j^T (w), Z_j^T (w) y, y)^T \) and

\[
\Delta(y, Z_j, \nu) = \exp \{ \zeta^2 R_j(y, y) + \xi_0 R_j(0, w) \}.
\]

The following conditions are needed in the proof of the main results.

(i) The kernel function \( K(\cdot) \geq 0 \) is a bounded, symmetric function with a compact support.

(ii) The functions \( \beta(\cdot) \) and \( g(\cdot) \) have a continuous third derivative around the point \( v \).

(iii) \( f_j(\cdot) \) is continuous at the point \( v \).

(iv) The conditional probability \( P(w, Z_j(w), \cdot) \) is equi-continuous at \( v \) and \( Z_j(w) \) is continuous about \( w \) for each \( j \).

(v) \( nh/\log n \to \infty \) and \( nh^5 \) is bounded and \( \int_0^\tau \lambda_{0j}(t)dt < \infty \) for each \( j \in \{1, 2, \cdots, J \} \).
(vi) $s_{jr}(t, \theta, v)$ ($j = 1, 2, \ldots, J, r = 0, 1, 2$) is bounded away from 0 on the product space $[0, \tau] \times C \times \Phi_{V, \varepsilon}$, that is

$$\inf_{t \in [0, \tau]} \inf_{\theta \in C} \inf_{v \in \Phi_{V, \varepsilon}} s_{jr}(t, \theta, v) > 0,$$

where $\theta = (\beta^T, g)$ and

$$\sup_{t \in [0, \tau]} \sup_{\theta \in C} \| Z_j(t) \|^2 \exp(\beta^T Z_j(t) + g) < \infty$$

for each $j \in \{1, 2, \ldots, J\}$. Meanwhile $s_{jr}(t, \theta, v)$ ($j = 1, 2, \ldots, J$) are continuous function for $(t, \theta, v) \in [0, \tau] \times C \times \Phi_{V, \varepsilon}$ uniformly in $t \in [0, \tau]$ and

$$s_{j1}(t, \theta, v) = \frac{\partial}{\partial \theta} s_{jr}(t, \theta, v),$$

and

$$s_{j2}(t, \theta, v) = \frac{\partial^2}{\partial \theta^2} s_{jr}(t, \theta, v).$$

(vii) (Asymptotic Variance) The matrix

$$\sum_{j=1}^J \left( a_{j2} - \int_0^\tau \frac{a_{j1}(w)a_{j1}(w)^T}{a_{j0}(w)} d\Lambda_{0j}(w) \right)$$

is positive definite for any $v \in \Phi_{V, \varepsilon}$, and the matrix

$$Q_2 = \begin{pmatrix} a_2 & a_1 \\ a_1^T & a_0 \end{pmatrix}$$

is nonsingular at $v \in \Phi_{V, \varepsilon}$.

(viii) The conditional probability $P(u, Z_j(u), w)$ is equi-continuous in the arguments $(u, w)$ on $[0, \tau] \times \Phi_{W, \varepsilon}$.

(ix) The compact set $\Phi_V \subset W$ has the following property: $\inf_{u \in \Phi_{W, \varepsilon}} f_j(u) > 0$ for each $j$ and some $\varepsilon > 0$, and $\| f_j \|_{\Phi_V} < \infty$.

(x) The covariate process $Z_j(t)$ has continuous sample path in a subset $Z$ of the continuous function space, and $|Z_{ijk}(0)| + \int_0^\tau |dZ_{ijk}(t)| \leq B_z$ a.s. for all $i, j, k$ and some constants $B_z < \infty$. 8
The above conditions will be used for deriving the pointwise convergence properties of \( \hat{\xi} \) and its asymptotic normality. Conditions (i)-(v) are similar to those in Fan et al. (1997) and Conditions (vii)-(viii) are similar to Conditions C and D of Andersen and Gill (1982). In order to derive the uniform consistent result, Conditions (ix)-(x) are also necessary. From the proof of the theorems, continuity of \( Z_j(t) \) in assumption (x) can be weakened to \( Z_j(t) \) being left-continuous with right-hand limits and \( \mathbb{E}[\exp\{\beta(W)^T Z(t)\}|Z(t)\otimes k|W = w] \) and \( \mathbb{E}(Z(t)\otimes k|W = w) \) are continuous functions of \( t \), for \( k = 0, 1, 2 \).

2.3 Asymptotic properties

We now establish the asymptotic properties of the local pseudo-partial likelihood estimator. We summarize the results here and provide the outline of the proofs in Section 6. As shown in Section 6, the local pseudo-partial likelihood function \( \ell_n(\xi, \tau) \) is concave in \( \xi \) and its maximizer exists with probability tending to one.

Let \( H \) be a \((2p + 1) \times (2p + 1)\) diagonal matrix, with the first \( p \) diagonal elements being 1 and the rest being \( h \), where \( p \) is the number of elements in \( Z \).

**Theorem 1** Under Conditions (i)-(viii), we have
\[
H(\hat{\xi}(v) - \xi_0(v)) \overset{P}{\longrightarrow} 0,
\]
where \( \xi_0(v) = (\beta_0^T(v), \beta_0'(v)^T, p_0(v))^T \) is the vector of the true parameter functions.
If, in addition, Conditions (ix)-(x) are satisfied, then we have the following uniform consistency:
\[
\sup_{u \in \Phi_V} |H(\hat{\xi}(u) - \xi_0(u))| \overset{P}{\longrightarrow} 0,
\]
where \( \Phi_V \) is any compact subset of the support of the random variable \( V \).

**Theorem 2** Assume that Conditions (i)-(viii) are satisfied. Then the random vector \( (nh)^{-1/2} \{ \ell'_n(\xi_0(u), \tau) - \frac{1}{2} h^2 \nu_2 \left( (\Gamma^{-1} u(\xi_0))^T, 0^T, 0^T \right) \} \) converges in distribution to a \((2p + 1)\)-variate normal vector with mean zero and covariance matrix \( \Pi \), where \( \ell'_n(\xi, \tau) = \partial \ell_n(\xi, \tau) / \partial \xi \), \( 0 \) is a \( p \)-variate column vector with all entries 0, and \( \Pi = \Pi_0 + D \), in which \( D = \text{blockdiag}(\Gamma^{-1} \nu_0, Q_2 \nu_2) \) and
\[
\Pi_0 = \sum_{l=1}^{J} \sum_{j=1, j \neq l}^{J} \lim_{n \to \infty} \mathbb{E} h B_{n1j}(\tau) B_{n1l}(\tau)^T,
\]
where the definitions of \( B_{n1j}(\tau) \) and \( B_{n1l}(\tau) \) can be found in the proof of Theorem 2.
Theorem 3 (Asymptotic normality) Assume that Conditions (i)-(viii) are satisfied. Then
\[ \sqrt{n h} \{ \hat{H}(\xi(v) - \xi_0(v)) - \frac{1}{2} h^2 \mathbf{e}_p \xi_0''(v) \nu_2 \} \xrightarrow{L} N(0, \Sigma(\tau, v)), \]
where \( \mathbf{e}_p \) is a \((2p+1) \times (2p+1)\) matrix, with the first \( p \times p \) elements being 1 and the rest of the elements being 0, \( \xi_0(v) = (\beta^T_0(v), \beta_0(v)^T, g_0(v))^T, \Sigma = A^{-1} \Pi (A^{-1})^T. \)

From the expressions of the asymptotic bias and variance matrix \( \Sigma \) in Section 6, it can be shown that they can be consistently estimated by
\[ \hat{A}^{-1}_n(\tau, v) \hat{B}_n(\tau, v) \quad \text{and} \quad (nh)^{-1} \hat{A}^{-1}_n(\tau, v) \hat{P}_n(\tau, v) \hat{A}^{-1}_n(\tau, v), \]
where
\[ \hat{A}_n(\tau, v) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^J \int_0^\tau K_h(V_{ij} - v) \left( \frac{\hat{S}_{njk}(w, v)}{\hat{S}_{njo}(w, v)} - \hat{E}_{ij}(w, v) \right) dN_{ij}(w), \]
\[ \hat{B}_n(\tau, v) = \frac{1}{nh} \sum_{i=1}^n \sum_{j=1}^J \int_0^\tau K_h(V_{ij} - v) \left( \mathbf{u}^*_i(w) - \hat{E}_{ij}(w, v) \right) dN_{ij}(w), \]
\[ \hat{P}_n(\tau, v) = \frac{1}{nh} \sum_{i=1}^n \left\{ \sum_{j=1}^J \int_0^\tau K_h(V_{ij} - v) \left( \mathbf{u}^*_i(w) - \hat{E}_{ij}(w, v) \right) d\hat{M}_{ij}(w) \right\} \otimes^2, \]
with \( \hat{S}_{njk}(w, v) = \frac{1}{n} \sum_{i=1}^n K_h(V_{ij} - v) Y_{ij}(w) \exp(\xi_0^T(v) \mathbf{X}_{ij}(w))(','(\mathbf{u}^*_i(w))^{\otimes k}, \text{for } k = 0, 1, 2, \mathbf{u}^*_i = \mathbf{H}^{-1} \mathbf{X}^*_i, \hat{E}_{ij}(w, v) = \frac{\hat{S}_{nj1}(w, v)}{\hat{S}_{njo}(w, v)}, \text{and } \hat{M}_{ij}(t) = N_{ij}(t) - \int_0^t \hat{\lambda}_{ij}(s) ds, \text{in which } \hat{\lambda}_{ij}(s) = \hat{\lambda}_{0j}(s) \exp\{\beta(V_{ij}(s))Z_{ij}(s) + \hat{g}(V_{ij}(s))\} \text{ and } \hat{\lambda}_{0j}(s) \text{ is given in the following section.} \]

2.4 Estimation of the baseline hazard function

With estimators of \( \beta(\cdot) \) and \( g(\cdot) \), we can estimate the baseline hazard function by using a kernel smoothing:
\[ \hat{\lambda}_{0j}(t) = \int W_b(t - x) d\hat{\Lambda}_{0j}(x), \]
where \( W_b \) is a given kernel function and \( b \) is a given bandwidth. The cumulative hazard function \( \Lambda_{0j}(\cdot) \) can be estimated by
\[ \hat{\Lambda}_{0j}(t) = \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{dN_{ij}(w)}{n^{-2} \sum_{i=1}^n Y_{ij}(w) \exp(\beta(V_{ij})^T \mathbf{z}_{ij}(w) + \hat{g}(V_{ij}))}. \]
The properties of \( \hat{\Lambda}_{0j}(\cdot) \) and \( \hat{\lambda}_{0j}(\cdot) \) are summarized in the following theorem and the outline of the proof is provided in Section 6.
Theorem 4 Under Conditions (i)-(x), we have
\[ \hat{\Lambda}_0(t) \rightarrow \Lambda_0(t) \quad \text{and} \quad \hat{\lambda}_0(t) \rightarrow \lambda_0(t) \]
uniformly on \((0, \tau]\) in probability.

To investigate the asymptotic properties of the estimated cumulative hazard function, we assume for simplicity that \(g(V) = 0\). The function \(g(\cdot)\) needs to be estimated by integrating its derivative estimator from the partial likelihood, and hence its asymptotic properties are challenging to obtain. When \(g(\cdot) = 0\) or is known, our task is simplified somewhat. The generalized Breslow estimator for \(\Lambda_0(t)\) is given by
\[ \hat{\Lambda}_0(t) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{dN_{ij}(w)}{n-1} \sum_{l=1}^{n} Y_{lj}(w) \exp\{\hat{\beta}(V_{lj})^T Z_{lj}(w)\}. \]

Write \(W_j(t) = n^{1/2} \{\hat{\Lambda}_0(t) - \Lambda_0(t)\}\). This is a stochastic process defined in the metric space \(\Omega = C[0, \tau]\) with the norm \(\rho(f, g) = \max_{1 \leq j \leq J} \sup_{0 \leq t \leq \tau} |f_j(t) - g_j(t)|\).

Theorem 5 Assume that conditions (i)-(x) are satisfied, and let \(nh^4 \rightarrow 0\). Then the random process vector \(W(t) = (W_1(t), \ldots, W_J(t))\) converges weakly to a zero-mean Gaussian random field \(G(t)\).

Remark 1 The covariance structure of the Gaussian field \(G(t)\) is very complex. It is very difficult to calculate this covariance by asymptotic methods directly. The wild bootstrap provides a useful method for computing the covariance or approximating the distribution of \(G(t)\) (see Spiekerman and Lin 1998).

Remark 2 Theorem 5 shows that the estimator \(\hat{\Lambda}_0(t)\) is root-n consistent, if the nonparametric estimators are under-smoothed. This means that in practical applications, one uses the right amount of smoothing for estimating coefficient functions and then chooses a smaller amount of smoothing for estimating the cumulative hazard functions. The situation here is very different from the one-step likelihood estimation of Carroll et al (1997), but similar to their one-step procedure.

2.5 One-step local pseudo-partial likelihood estimator
To estimate the functions \(\beta(\cdot)\) and \(g(\cdot)\) over an interval of interest, we usually need to maximize the local pseudo-partial likelihood (5) at hundreds of points. This can be very computationally intensive. In addition, for certain given \(v\), the local pseudo-partial likelihood estimator might not exist due to limited amount of data around
These drawbacks make computing the local pseudo-partial likelihood estimator over an interval less appealing. We consider the following one-step estimator as a feasible alternative.

To facilitate notation, we drop the dependence of \( \ell_n(\xi, \tau) \) on \( \tau \). The local pseudo-partial likelihood estimator \( \hat{\xi} \) satisfies \( \ell_n'(\xi) = 0 \). For a given initial estimator \( \hat{\xi}_0 \), by Taylor's expansion, we have

\[
\ell_n'(\hat{\xi}_0) + \ell_n''(\hat{\xi}_0)(\hat{\xi} - \hat{\xi}_0) \approx 0.
\]

Thus, the one-step estimator \( \hat{\xi}_{os} \) is defined as

\[
\hat{\xi}_{os} = \hat{\xi}_0 - \{\ell_n''(\hat{\xi}_0)\}^{-1}\ell_n'(\hat{\xi}_0).
\]

In the Newton-Raphson algorithm, the above equation is iterated until convergence. As shown in Section 6, the function \( \ell_n(\xi) \) is concave. Hence, its maximizer exists and is unique when \( \ell_n(\xi) \) is strictly concave. In practice, we do not have to iterate (7) until convergence — once or a few times suffice. Robinson (1988) gives the results on the distance between the estimators based on a few steps of iteration and the maximum likelihood estimator. A natural question arises as to how good the initial estimator \( \hat{\xi}_0 \) has to be in order for the one-step estimator to have the same performance as the maximum local pseudo-partial likelihood estimator. It is not hard to show that a sufficient condition is

\[
H(\hat{\xi}_o - \xi_0) = O_P(h^2 + (nh)^{-1/2}).
\]

See Fan and Chen (1999) for a derivation in the local likelihood context. When condition (8) is not satisfied, multiple-step estimator is needed. By repeatedly applying the one-step result \( k \) times as in Robinson (1988), condition (8) can be relaxed to

\[
H(\hat{\xi}_o - \xi_0) = O_P\{(h^2 + (nh)^{-1/2})^{1/k}\}.
\]

Cai et al. (2000) provide a useful strategy on the choice of initial estimators in the context of generalized linear models and their idea can be adapted to the current setting. Their idea is to exploit the smoothness of nonparametric functions. Compute the local pseudo-partial likelihood estimates at a few fixed points. Use these estimates as the initial values of their nearest grid points and obtain the one-step estimates at these grid points. Use the newly computed one-step estimates as the initial values of their nearest grid points to compute the one-step estimates and propagate, until the one-step estimates at all grid points are computed. For
example, in our simulation studies, we shall evaluate the functions at \( n_{\text{grid}} = 200 \) grid points and are willing to compute the maximum local pseudo-partial likelihood at five distinct points. A sensible placement of these points is \( w_{20}, w_{60}, w_{100}, w_{140} \) and \( w_{180} \). We shall use, for instance, \( \hat{\beta}(w_{60}) \) as an initial value for calculating the one-step estimator for \( \hat{\beta}(w_{59}) \) and \( \hat{\beta}(w_{61}) \), and then proceed to use the resulting estimates as the initial value for calculating the one-step estimator for \( \hat{\beta}(w_{58}) \) and \( \hat{\beta}(w_{62}) \), respectively. We continue this process until all the one-step estimates at \( w_i \) for \( i = 40, \ldots, 79 \) are calculated.

3 Weighted average estimator

An alternative approach is to fit a varying coefficient model for each failure type. That is, fitting the following model for event type \( j \):

\[
\lambda_{ij}(t; F_{ij}) = \lambda_{0j}(t) \exp\{\beta_j(V_{ij}(t))^T Z_{ij}(t) + g_j(V_{ij}(t))\}, \quad \text{for } i = 1, \ldots, n,
\]

resulting in \( \hat{\xi}_j(v) \) for estimating \( \xi_j(v) = (\beta_j^T(v), (\beta_j'(v))^T, g_j'(v)) \). Under model (1), we have \( \xi_1 = \xi_2 = \cdots = \xi_J = \xi \). Thus, we can estimate \( \xi(v) \) by a linear combination \( c_1 \hat{\xi}_1(v) + \cdots + c_J \hat{\xi}_J(v) \) with \( \sum_{j=1}^J c_j = 1 \). Weights \( c_j \) can be chosen to optimize the performance. Note that the weights \( c_j \) can be generalized to a matrix \( C_j \) to allow for different linear combination for different component of \( \xi(v) \), i.e., the linear combination can be generalized to \( C_1 \hat{\xi}_1 + \cdots + C_J \hat{\xi}_J \) with \( C_1 + \cdots + C_J = \text{diag}(1, \ldots, 1) \) being the identity matrix.

In order to establish the asymptotic distribution of the weighted average estimator, we need to derive the asymptotic distribution of \( \hat{\Psi}(v) = (\hat{\xi}_1^T, \ldots, \hat{\xi}_J^T)^T \). We define \( \Psi(v) \) and \( \Psi''(v) \) similarly to \( \hat{\Psi}(v) \) except that \( \hat{\xi}_j \) are replaced by \( \xi_j \) and \( \xi_j'' \), respectively, for \( j = 1, 2, \ldots, J \). Using similar arguments to those for Theorems 2 and 3, it can be shown that

**Theorem 6** Under conditions of Theorem 2, we have that

\[
\sqrt{nh} \left\{ \mathbf{I}_p \otimes H[\hat{\Psi}(v) - \Psi(v)] - \frac{h^2}{2} \mathbf{R} \Psi''(v) \right\}
\]

is asymptotically normal with mean 0 and a covariance matrix \( \Sigma^* = (G_{kl}(\xi_k, \xi_l)) \) for \( k, l = 1, \ldots, J \), where \( \mathbf{R} = \text{diag}(\mathbf{R}_1, \cdots, \mathbf{R}_J) \) and \( \mathbf{R}_j \) is a \((2p + 1) \times (2p + 1)\) matrix with the first \( p \times p \) elements being \( \mathbf{I}_p \), which is an identity matrix, and the rest of the elements being 0, and \( G_{kl}(\xi_k, \xi_l) \) is defined at the end of this section.

The asymptotic normality of the weighted average estimator follows easily from Theorem 6. For example, suppose that we are interested in estimation of the \( k \)th
Obviously, where \( A \) is given by

\[
\frac{(nh)^{1/2}}{\mathcal{L}} \left\{ \hat{\beta}_w(v) - \beta_0(v) - \frac{h^2}{2} \mathcal{C}_\mathbf{R} \Psi''(v) \right\} \to N(0, \Sigma_w),
\]

where \( \beta_w = (\beta_{1k}, \ldots, \beta_{kj})^T \) and \( \Sigma_w = \mathcal{C}^T \Sigma_w \mathcal{C} \), in which \( \beta_{kj} \) is the \( k \)th entry of \( \beta_j \). The optimal weight which minimizes \( c^T \Sigma_w c \) with \( \sum_{j=1}^J c_j = 1 \) is

\[
c_k = (e^T \Sigma_w^{-1} e)^{-1} \Sigma_w^{-1} e.
\]

Since failure times \( T_{ij} \) are usually dependent for different failure type, \( \hat{\xi}_j (j = 1, \ldots, J) \) are likely to be dependent, hence, the variance \( \Sigma_w \) is not necessarily diagonal. This implies that the optimal weight is unlikely to be uniform. In other words, the weighted average estimator with the optimal weight is, in general, more efficient than the estimator with working independence weight. This is supported by the simulation results reported in Table 2.

We now give the expressions of the asymptotic variance-covariance matrix in Theorem 6 and its estimate. From Theorem 3, it is easy to show that the asymptotic covariance matrix between \((nh)^{1/2} \mathbf{H}(\hat{\xi}_k(v) - \xi_{k0}(v))\) and \((nh)^{1/2} \mathbf{H}(\hat{\xi}_l(v) - \xi_{l0}(v))\) is given by

\[
G_{kl}(\xi_k, \xi_l) = \mathbf{A}^{-1}_k(\xi_k) \lim_{n \to \infty} E \{ \Pi_{1k}(\xi_k) \Pi_{1l}(\xi_l) \} \mathbf{A}^{-1}_l(\xi_l),
\]

where \( \Pi_{jk}(\xi_k) = \int_0^\tau K_h(V_{jk} - v) \left[ U_{jk}^* s_{k1}(w, \zeta, v)/s_{k0}(w, \zeta, v) \right] dM_j(w), \xi = \mathbf{H}(\xi - \xi_0) \), and \( s_{kd}(w, \zeta, v) \) (d=0.1) are defined as in Section 6. From the definition of \( \Pi_{jk}(\xi_k) \), it is natural to estimate \( \lim_{n \to \infty} E \{ \Pi_{1k}(\xi_k) \Pi_{1l}(\xi_l) \} \) by

\[
\hat{D}_{kl}(\hat{\xi}_k, \hat{\xi}_l) = n^{-1} \sum_{j=1}^n W_{jk}(\hat{\xi}_k) W_{jl}^T(\hat{\xi}_l),
\]

where

\[
W_{jk}(\xi_k) = \Delta_{jk} \left\{ U_{jk}^* (X_{jk}) - \frac{\hat{S}_{nj1}(X_{jk}, v)}{\hat{S}_{nj0}(X_{jk}, v)} \right\} K_h(V_{jk} - v) - \sum_{m=1}^n \frac{\Delta_{mk} Y_{jk}(X_{mk}) \exp \{ \hat{\beta}_{k}^T (V_{jk}(X_{mk})) Z_{jk}(X_{mk}) + \hat{g}_k(V_{mk}(X_{mk})) \} \} \times \left\{ U_{jk}^* (X_{mk}) - \frac{\hat{S}_{nj1}(X_{mk}, v)}{\hat{S}_{nj0}(X_{mk}, v)} \right\} K_h(V_{jk} - v).
\]

Obviously, \( \mathbf{A}_j(\xi) \) can be estimated by

\[
\hat{A}_j(\xi) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(V_{ij} - v) \frac{\hat{S}_{nj2}(w, v) \hat{S}_{nj0}(w, v) - \hat{S}_{nj2}^2(w, v)}{(\hat{S}_{nj0}(w, v))^2} dN_{ij}(w).
\]
Write
\[ \hat{G}_{kl}(\xi_k, \xi_l) = \hat{A}_k^{-1}(\xi_k)\hat{D}_{kl}(\hat{\xi}_k, \hat{\xi}_l)\hat{A}_l^{-1}(\xi_l). \]

By some tedious proofs, we can show that \( \hat{G}_{kl}(\xi_k, \xi_l) \) is a consistent estimator of \( G_{kl}(\xi_k, \xi_l) \). Hence the covariance matrix of \( (\hat{\xi}_1, \cdots, \hat{\xi}_J) \) can be consistently estimated by \( \hat{\Sigma}^* = \left( n h \right)^{-1}(\hat{G}_{ij}(\hat{\xi}_i, \hat{\xi}_j))_{i,j=1}^J \). These results provide a basis for simultaneous inferences about the \( \xi_j, j = 1, 2, \cdots, J \) as well as for the weighted average estimator \( \sum_{j=1}^J c_j\hat{\xi}_j \) for \( \xi \).

4 Numerical Examples

4.1 Simulations

We perform a series of simulation studies to evaluate the performance of the proposed estimation method. Multivariate failure times are generated from a multivariate extension of the model of Clayton and Cuzick (1985) in which the joint survival function of \( (T_1, \cdots, T_J) \) given \( (Z_1, \cdots, Z_J) \) and \( (V_1, \cdots, V_J) \) is:

\[ F(t_1, \cdots, t_J; Z_1, \cdots, Z_J, V_1, \cdots, V_J) = \left\{ \sum_{j=1}^J S_j(t_j)^{-\theta} - (J - 1) \right\}^{-1/\theta}, \quad (11) \]

where \( J \) takes integer values and \( S_j(t) \) is the marginal survival probability for the \( j \)th member, depending on covariates \( Z_j \) and \( V_j \). Note that \( \theta \) is a parameter which represents the degree of dependence of \( T_i \) and \( T_j \) \( (i, j = 1, 2, \cdots, J) \). The relationship between Kendall’s \( \tau \) and \( \theta \) is \( \tau = \theta/(2+\theta) \). Specifically, \( \theta = 0.25 \) and \( \theta = 4 \) represent weak and strong positive dependence, respectively, with \( \theta \to 0 \) giving independence and \( \theta \to \infty \) giving maximal positive dependence. In our simulation, \( \theta \) was chosen to be 0.25, 1.5 and 4.0 which correspond to low, moderate or high positive dependence, respectively. The Gaussian kernel function is used for the estimates.

In our first set of simulations, we examine the performance of the local pseudo-partial likelihood estimators. We consider the marginal distribution of \( T_{ij} \) to be exponential with the failure rate

\[ \lambda_{ij}(t) = \lambda_{0j}(t) \exp\{\beta(V_{ij})Z_{ij} + g(V_{ij})\}. \quad (12) \]

We choose the baseline hazard function to depend on time through \( \lambda_{0j}(t) = 4t^3\lambda_{0j}^* \), where \( j = 1, 2, 3 \). We take \( \lambda_{0j}^* \) to be 0.2, 1.0 and 1.5 for \( j = 1, 2, 3 \), respectively. Failure times \( (t_{i1}, t_{i2}, t_{i3}) \) \( (i = 1, 2, \cdots, n) \) are generated from the distribution function (11) with marginal distribution (12). We generate the covariate
vector $\mathbf{Z}_{ij} = (Z_{ij1}, Z_{ij2}, \ldots, Z_{ijp})^T$ from a multivariate normal distribution with marginal mean of 0, standard deviation of 5, and the correlation between $Z_{ijl}$ and $Z_{ijk}$ of $\rho^{ij}$ and $\rho = 1/\sqrt{5}$. We consider $p = 2$ and the varying-coefficients

$$\beta_1(V) = 0.5V(1.5 - V), \quad \beta_2(V) = \sin(2V), \quad \text{and} \quad g(V) = 0.5\{e^{V - 1.5} - e^{-1.5}\},$$

g where $V$ is generated from a uniform distribution over $[0, 3]$. In our simulation, using similar derivation as in Cai and Shen (2000), with given covariates $(\mathbf{Z}_{ij}, V_{ij})$, $(j = 1, 2, 3)$, the failure times $(t_{i1}, t_{i2}, t_{i3})$ are generated through independent uniform random variables $(w_{i1}, w_{i2}, w_{i3})$ via

$$t_{i1} = [- \log(1 - w_{i1})\Upsilon(V_{i1}, \mathbf{Z}_{i1}, \lambda_0^*)]^{1/4},$$
$$t_{i2} = [\theta \log(1 - a_{i1} + a_{i1}(1 - w_{i2}))^{(\theta - 1+1)^{-1}}\Upsilon(V_{i2}, \mathbf{Z}_{i2}, \lambda_0^*)]^{1/4},$$
$$t_{i3} = [\theta \log(1 - (a_{i1} + a_{i2}) + (a_{i1} + a_{i2} - 1)(1 - w_{i3}))^{(\theta - 1+2)^{-1}}\Upsilon(V_{i3}, \mathbf{Z}_{i3}, \lambda_0^*)]^{1/4},$$

where $a_{il} = (1-w_{il})^{-\theta}$ for $l = 1, 2$ and $i = 1, 2, \cdots, n$ and $\Upsilon(V, \mathbf{Z}, \lambda^*) = \exp\{\beta(V)\mathbf{Z} + g(V)\}/\lambda^*$. Censoring times $C_{ij}$ are generated from uniform distribution over $(0, c)$, where $c$ is a constant which is set to control censoring rate. There are about 10% censoring when $c = 5$ and about 30% censoring when $c = 2$. For each of the configuration studied, 500 simulations were carried out.

**Table 1 and Figure 1 are about here.**

Table 1 summarizes the simulation results for the local pseudo-partial likelihood estimator of $\beta(\cdot)$ and $g'(\cdot)$ with the number of the clusters of 200, $\theta=0.25$, and $c=2$. We present the estimates of the functions evaluated at $v = 0.5$, 1.0, 1.5, 2.0 and 2.5. The bandwidths we considered are $h=0.075$, 0.1, 0.15, 0.2 and 0.4. The averages of the 500 estimates for $\beta_1(v)$, $\beta_2(v)$ and $g'(v)$ subtracted by their true values are given in the “bias” columns, and the standard deviations of the 500 estimates are given in the corresponding SD columns. The SE columns give the average of the estimated standard errors. Figure 1 provides the average estimates for $\beta_1(\cdot)$, $\beta_2(\cdot)$ and $g'(\cdot)$ based on different bandwidths. It gives us an idea how large the biases are for different bandwidths. From Table 1, we can also see that as the bandwidth increases, the variance decreases. As expected, with a large bandwidth $h=0.4$, the bias is large and variance is small. Note that the absolute biases exhibit a U-shape in Table 1. This is unusual, but can happen. The bias depends on function values on a local neighborhood and is continuous in $h$. If the bias associated with $h = 0.075$ is negative and with $h = 0.4$ is positive as in this example, then the bias has to cross
zero and the U-shape absolute biases emerge. Table 1 also shows that the average SE is close to SD when the bandwidth is 0.15 and above which indicates good performance of the variance estimator. We have also examined the situation with moderate and high dependence of the failure times with $\theta=1.5$ and 4, respectively, as well as with lighter censoring with $c=5$. The conclusions are similar.

**Table 2 is about here.**

We also examined the performance of the weighted average estimator. The results are presented in Table 2. The rows indicated by “W” are those based on the weighted average estimates for $\beta_p$ with optimal weight $c_p = (\mathbf{e}^T \hat{\Sigma}_p^{-1} \mathbf{e})^{-1} \hat{\Sigma}_p^{-1} \mathbf{e}$, where $\hat{\Sigma}_p$ is the estimator of the asymptotic variance-covariance matrix of $(\hat{\beta}_{p1}, \hat{\beta}_{p2}, \hat{\beta}_{p3})^T$. The maximum local pseudo-partial likelihood estimates are indicated by “P”. The performance of an estimator $\hat{\beta}(\cdot)$ is assessed via the square-root of average square errors (RASE),

$$\text{RASE} = \left( \frac{1}{n_{\text{grid}}} \sum_{k=1}^{n_{\text{grid}}} [\hat{\beta}(w_k) - \beta(w_k)]^2 \right)^{1/2},$$

(13)

where \(\{w_k, k = 1, \cdots, n_{\text{grid}}\}\) are the grid points at which the functions $\beta(\cdot)$ are estimated. In our simulations, we take $n_{\text{grid}} = 200$.

For the local pseudo-partial likelihood estimates, we used the bandwidth of $h=0.15$. For the weighted average estimates, the bandwidth of $h=0.225$ was used, since the data used for estimation of the covariate effects for each event type was significantly less. The censoring parameter $c$ was set to be 5. The weighted average estimate cannot always be calculated since the data for each event type could be too sparse to permit an estimate for each type. We only report those estimates which exist. From Table 2, we can see that the weighted average estimator has smaller RASE in most of the cases when they can be calculated.

In the second set of simulations, we compare the performance of the one-step estimator (OS) to the maximum local pseudo-partial likelihood estimator (P). We take model (11) with somewhat different configurations. In particular, $V$ is now generated from the standard uniform distribution over $[0, 1]$, $Z$ is independently generated from a standard normal distribution, and $\theta = 0.25$ and 4 and $(\lambda_{01}^*, \lambda_{02}^*, \lambda_{03}^*) = (0.2, 1.0, 1.5)$. Censoring times $C_{ij}$ are generated from uniform distribution over $(0, c)$ with $c=2$ and 5. We take $g(u) = 8u(1-u)$ and $\beta(u) = \exp(2u - 1)$.

**Table 3 and Figure 2 are about here.**
Table 3 presents the summary of the average square errors (ASE = RASE²) for the one-step estimator (OS) and the maximum local pseudo-partial likelihood estimator (P) under various realizations. From the table we can see that the performance of the one-step estimator is very close to that for the maximum local pseudo-partial likelihood estimator. Figure 2, which presents the box-plot and the estimated curves for the two methods, also confirms it. The curves based on the one-step estimator and the maximum local pseudo-partial likelihood estimator are hardly distinguishable in Figure 2. We have also conducted simulations using the parameters considered in the first set of simulations. The results are similar.

4.2 Application to Busselton population health surveys

We illustrate the proposed method by analyzing a data set from the Busselton Population Health Surveys. The Busselton Population Health 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, general health information for adult participants were collected by means of questionnaire and clinical visit. Details of the study are described in Cullen (1972) and Knuiman et al. (1994). Data for several cardiovascular risk factors are available for 2202 persons who make up 619 families. In this analysis we investigate the effect of cardiovascular risk factors on the risk of death due to cardiovascular disease (CVD) based on these family data. Since the death times of the family members might be correlated due to genetic factors and cohabitation, we are dealing with multivariate failure time data.

The risk factors we considered here includes age (in years), body mass index (bmi, in kg/m²), serum cholesterol level (chol), and smoking status. Serum cholesterol (in mmol/L) was determined from a blood sample. Participant’s smoking status was classified into three categories: never smoker, ex-smoker and current smoker. Two indicator variables are created to indicate the three levels of smoking status: smoke1 is coded as 1 for ex-smoker and 0 otherwise; and smoke2 is coded as 1 for current smoker and 0 otherwise.

If a person took part in more than one of the Busselton surveys, only one record from the survey at which that person’s age was closest to 45 years is included. Forty-eight percent of the participants are males (gender=0 for male and 1 for female). The average age in the data analyzed is 41.7 years, ranging from 16.3 to 89.0 years old. The average cholesterol reading was 5.65 mmol/L. The average body mass index was 24.8 kg/m². The prevalence of the never-smokers, ex-smokers, and current smokers
are 49%, 17%, and 34%, respectively. Of the 619 families, there are 154 families
with one event, 28 families with two events, and 3 families with more than two
events. There are 219 observed events in all.

For this analysis, we are interested in investigating how the effect of the risk
factors changing with age. We consider the following model:

$$\lambda_{ij}(t|F_{ij}) = \lambda_{0j}(t) \exp(\beta_1(\text{age}_{ij}) \ast \text{gender}_{ij} + \beta_2(\text{age}_{ij}) \ast \text{bmi}_{ij} + \beta_3(\text{age}_{ij}) \ast \text{chol}_{ij} + \beta_4(\text{age}_{ij}) \ast \text{smoke1}_{ij} + \beta_5(\text{age}_{ij}) \ast \text{smoke2}_{ij} + g(\text{age}_{ij})),$$

where $j = 1$ and 2 denote the parents and the children of the family, respectively, and
smoke1 and smoke2 are indicators for ex-smoker and current smoker, respectively.
We take the bandwidth to be $h = 0.15 \ast (\max(\text{age}) - \min(\text{age})) = 10.905$.

**Figure 3 is about here.**

Figure 3 presents the estimates for the varying-coefficients as functions of age.
From Figure 3(a) we can see that men have higher risk of dying from CVD than
women with the hazard ratio being 1.96 with 95% confidence interval (CI) of (1.30,
3.03) at age 50. The effect does not seem to change much with age for those older
than 35. From Figure 3(b), BMI has little effect for the risk because the coefficient
is close to zero over the span of age. From Figure 3(c), higher cholesterol level is
associated with higher risk of dying from CVD and the effect of cholesterol increased
with age. The hazard ratio for 1 mmol/L change in cholesterol is 1.01 (95% CI:
[0.80,1.28]) at age 40 and 1.30 (95% CI: [1.12,1.53]) at age 65. From Figure 3(d)
and (e), ex-smokers have similar risk as the never smokers, while current smokers
have higher risk of dying from CVD compared to the never smokers. The effect of
current smoking is higher for younger people with the hazard ratio being 5.60 (95%
CI: [2.19,14.34]) at age 40 and 1.07 (95% CI: [0.72,1.60]) at age 65.

5 Concluding Remarks

The local pseudo-partial likelihood is a powerful and a straightforward approach to
analyzing multivariate failure time data. The estimator follows asymptotically a
normal distribution. Simulation results show that the asymptotic approximation is
applicable to finite samples with moderate number of clusters.

The weighted average estimator, when it can be calculated, can be a more ef-
ficient alternative to the maximum local pseudo-partial likelihood estimator. A
disadvantage of the weighted average estimator is that it cannot always be calculated, since it involves estimating the covariate coefficient for each failure type and the data for each failure type could be too sparse to permit a reliable estimate.

The one-step estimator is an effective way to reduce the computational burden of estimator involving iterations. We showed theoretically and empirically that the one-step estimator is an excellent approximation of the fully iterated maximum local pseudo-partial likelihood estimator.

Our proposed methods are sensitive to the choice of bandwidths in constructing local smooth estimation. The one-step estimator and maximum local pseudo-partial likelihood estimator have the same asymptotic distribution and share the same asymptotic bandwidth. We can use the sophisticated bandwidth selection rule proposed in Fan et. al. (1998) for these estimators.

The methods proposed in this paper can be easily extended to the more general form of the multivariate failure time data. More specifically, suppose that there are \( n \) clusters and in each cluster there are \( K \) correlated individuals and for each individual there are \( J \) possible distinct failure types. The marginal hazard function for the \( j \)th type of failure of the \( k \)th individual in the \( i \)th cluster is related to the corresponding covariate vector \( Z_{ijk}(t) \) by

\[
\lambda_{ijk}(t, Z_{ijk}) = \lambda_{0j}(t) \exp\{\beta^T(W_{ijk})Z_{ijk}(t) + g(W_{ijk})\},
\]

where \( \lambda_{0j}(t) \) (\( j = 1, 2, \cdots, J \)) are unspecified positive functions and \( \beta(\cdot) \) and \( g(\cdot) \) are defined as in (1). The maximum local pseudo-partial likelihood for the more general model can be derived similarly and the asymptotic properties can be established with similar but more tedious approach.

### 6 Proofs

Let \((\Omega, \mathcal{F}, P(\beta, g, \lambda))\) be a family of complete probability spaces provided with a history \( \mathcal{F} = \{\mathcal{F}_t\} \) for an increasing right-continuous filtration \( \mathcal{F}_t \subset \mathcal{F} \). Let \( Y_{ij}(t) = I(X_{ij} \geq t) \). We assume that \( V_{ij} \) is \( \mathcal{F}_{t_{ij}} \)-measurable, and \( N_{ij}(w) \) and \( Z_{ij}(w) \) are \( \mathcal{F} \)-adapted. Let \( \mathcal{F}_{t_{ij}} = \sigma\{N_{ij}(w^-), Z_{ij}(w), V_{ij}, Y_{ij}(w), 0 \leq w \leq t\}, i = 1, 2, \cdots, n, j = 1, 2, \cdots, J \) and \( M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(w)\lambda_{ij}(w)dw, i = 1, 2, \cdots, n \). Obviously, \( M_{ij}(t) \) is a \( \cup_{i=1}^n \mathcal{F}_{t_{ij}} \) martingale.

To facilitate technical arguments, we will re-parameterize the local pseudo-partial likelihood (5) via the transform: \( \zeta = H(\xi - \xi_0) \). Hence, the logarithm
of local pseudo-partial likelihood function has the form $\tilde{\ell}_n(\mathbf{\zeta}, t) = \ell_n(\mathbf{H}^{-1} \mathbf{\xi} + \mathbf{\xi}_0, t)$. By simplification, we have

$$\tilde{\ell}_n(\mathbf{\zeta}, t) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{J} \int_0^t K_h(V_{ij} - v) \left[ \mathbf{\zeta}^T \mathbf{U}_{ij}^*(w) + \mathbf{\xi}_0^T \mathbf{X}_{ij}^*(w) - \log S_{njk}(w, \mathbf{\zeta}, v) \right] dN_{ij}(w)$$

where $\mathbf{U}_{ij}^*(w) = \mathbf{H}^{-1} \mathbf{X}_{ij}^*(w)$ and

$$S_{njk}(w, \mathbf{\zeta}, v) = \frac{1}{n} \sum_{i=1}^{n} K_h(V_{ij} - v) Y_{ij}(w) \exp\{ \mathbf{\zeta}^T \mathbf{U}_{ij}^*(w) + \mathbf{\xi}_0^T \mathbf{X}_{ij}^*(w) \} \{ \mathbf{U}_{ij}^*(w) \}^{\otimes k}.$$}

Furthermore, for each $w \in [0, \tau]$ and $k = 0, 1, 2$, we rewrite $\tilde{\ell}_n(\mathbf{\zeta}) = \tilde{\ell}(\mathbf{\zeta}, \tau)$ and define

$$S_{njk}^*(w, \mathbf{\theta}, v) = \frac{1}{n} \sum_{i=1}^{n} K_h(V_{ij} - v) Y_{ij}(w) \exp(\mathbf{\beta}^T (V_{ij}) \mathbf{Z}_{ij}(w) + g(V_{ij})) \{ \mathbf{U}_{ij}^*(w) \}^{\otimes k},$$

where $\mathbf{\xi}(\cdot) = (\mathbf{\beta}^T(\cdot), \mathbf{\beta}'(\cdot)^T, g(\cdot))^T, \mathbf{\theta}(\cdot) = (\mathbf{\beta}^T(\cdot), g(\cdot))^T, v \in \Phi_V$.

Recall the notation $\rho(w, \mathbf{z}, v)$ introduced in §2.3. Define, for $v \in \Phi_{V, \varepsilon}$,

$$s_{j0}^*(w, \mathbf{\theta}, v) = f_j(v) \mathbb{E} \left[ \rho(w, \mathbf{Z}_{j}(w), v) \big| \mathbf{V}_j = v \right],$$

$$s_{j1}^*(w, \mathbf{\theta}, v) = f_j(v) \mathbb{E} \left[ \rho(w, \mathbf{Z}_{j}(w), v) (\mathbf{Z}_{j}(w), 0, 0)^T \big| \mathbf{V}_j = v \right],$$

$$s_{j2}^*(w, \mathbf{\theta}, v) = f_j(v) \mathbb{E} \left[ \rho(w, \mathbf{Z}_{j}(w), v) \mathbf{\Xi}(\mathbf{Z}_{j}, w) \big| \mathbf{V}_j = v \right].$$

where

$$\mathbf{\Xi}(\mathbf{Z}_{j}, w) = \begin{pmatrix} \mathbf{Z}_{j}(w) \mathbf{Z}_{j}^T(w) & 0 & 0 \\ 0 & \mathbf{Z}_{j}(w) \mathbf{Z}_{j}^T(w) \mathbf{\mu}_2 & \mathbf{Z}_{j}(w) \mathbf{\mu}_2 \\ 0 & 0 & \mathbf{\mu}_2 \end{pmatrix}.$$}

To facilitate notation, the true functions $\mathbf{\theta}_0(u) = (\mathbf{\beta}_0^T(u), g_0(u))^T, \mathbf{\xi}_0(u), \mathbf{\zeta}_0 = 0$ and $v$ shall be omitted in $S_{njk}^*(t, \mathbf{\theta}, v), S_{njk}(t, \mathbf{\zeta}, v)$ and $s_{j0}^*(t, \mathbf{\theta}, v), s_{j1}(t, \mathbf{\zeta}, v)$ whenever there is no ambiguity, e.g.,

$$S_{njk}^*(t) = S_{njk}^*(t, v) = S_{njk}^*(t, \mathbf{\theta}_0, v), \quad s_{j0}^*(t) = s_{j0}^*(t, v) = s_{j0}^*(t, \mathbf{\theta}_0, v),$$

$$S_{njk}(t) = S_{njk}(t, v) = S_{njk}(t, \mathbf{0}, v), \quad s_{j1}(t) = s_{j1}(t, v) = s_{j1}(t, \mathbf{0}, v),$$

$$S_{njk}(t, \mathbf{\zeta}) = S_{njk}(t, \mathbf{\zeta}, v), \quad s_{j0}(t, \mathbf{\zeta}) = s_{j0}(t, \mathbf{\zeta}, v).$$

21
We will need the following two lemmas in the proof of the theorems. Let

\[ C_{nj}(t) = n^{-1} \sum_{i=1}^{n} Y_{ij}(t) \psi(V_{ij}, (V_{ij} - v)/h, Z_{ij}(t)) K_h(V_{ij} - v), \]

for a function \( \psi(\cdot, \cdot, \cdot) \).

**Lemma 1** Assume that Conditions (i) and (iv) hold. Assume that \( \psi(\cdot, \cdot, \cdot) \) is continuous for its three arguments and \( E(\psi(V_j, w, Z_j(t))|V_j = v) \) is continuous at the point \( v \) for each \( j \) and \( w \). If \( h \to 0 \) in such a way that \( nh/\log n \to \infty \), then

\[
\sup_{0 \leq t \leq \tau} \sum_{j=1}^{J} |C_{nj}(t) - C_j(t)| \xrightarrow{P} 0,
\]

where \( C_j(t) = f_j(u) \int E(Y(t)\psi(v, w, Z_j(t))|V_j = v) K(w) dw \), where \( f_j(u) \) is the density function of \( V \). Under Conditions (viii)-(x), we have

\[
\sup_{0 \leq t \leq \tau \in B} \sup_{v \in B} \sum_{j=1}^{J} |C_{nj}(t, v) - C_j(t, v)| \xrightarrow{P} 0,
\]

where \( B \) is a compact set satisfying \( \inf_{u \in B} f_j(u) > 0 \).

**Proof.** By the assumption of \( h \), it is easy to show that for every \( t \in [0, \tau], \)

\[
|C_{nj}(t) - C_j(t)| \xrightarrow{P} 0. \tag{14}
\]

Now we divide \([0, \tau]\) into \( M \) sub-intervals \([t_{i-1}, t_i]\) with a given length not exceeding \( \delta \). Note that \( \delta \) does not depend on \( n \). Then

\[
\max_{1 \leq i \leq M} |C_{nj}(t_i) - C_j(t_i)| \xrightarrow{P} 0. \tag{15}
\]

Thus, we need only to deal with the term

\[
\max_{1 \leq i \leq M} \sup_{|t-t_{i-1}| < \delta} |C_{nj}(t) - C_j(t) - \{C_{nj}(t_{i-1}) - C_j(t_{i-1})\}|. \tag{16}
\]

By decomposing \( \psi \) into the positive part and negative part, we can decompose \( C_{nj}(t) \) into \( C_{nj}^{+}(t) \) and \( C_{nj}^{-}(t) \). Hence, we need only to show that as \( n \to \infty \) and \( \delta \to 0 \)

\[
\max_{1 \leq i \leq M} \sup_{|t-t_{i-1}| < \delta} |C_{nj}^{+}(t) - C_{nj}^{+}(t_{i-1})| + \max_{1 \leq i \leq M} \sup_{|t-t_{i-1}| < \delta} |C_{j}^{+}(t) - C_{j}^{+}(t_{i-1})| \xrightarrow{P} 0. \tag{17}
\]
and a similar result for \( C_{nj}^{-} \) and \( C_{j}^{-} \).

We now focus on the first term of (17), which is bounded by \( J_1 + J_2 \), where

\[
J_1 = \max_{1 \leq i \leq M} \sup_{|t-t_i-1| \leq \delta} \left| n^{-1} \sum_{i=1}^{n} Y_{ij}(t)K_h(V_{ij} - v)\{\psi^+(V_{ij}, (V_{ij} - v)/h, Z_{ij}(t)) - \psi^+(V_{ij}, (V_{ij} - v)/h, Z_{ij}(t_{i-1}))\} \right|
\]

and

\[
J_2 = \max_{1 \leq i \leq M} \sup_{|t-t_i-1| \leq \delta} \left| n^{-1} \sum_{i=1}^{n} \{Y_{ij}(t) - Y_{ij}(t_{i-1})\}\psi^+(V_{ij}, (V_{ij} - v)/h, Z_{ij}(t_{i-1}))K_h(V_{ij} - v) \right|
\]

Note that \( Z_{ij}(t) \) \((l = 1, 2, \ldots, n)\) is continuous on \([0, \tau]\). Thus, \( J_1 \) is bounded by

\[
\max_{1 \leq l \leq n} \max_{1 \leq i \leq M} \sup_{|t-t_i-1| \leq \delta} \left| \psi^+(V_{ij}, (V_{ij} - v)/h, Z_{ij}(t)) - \psi^+(V_{ij}, (V_{ij} - v)/h, Z_{ij}(t_{i-1})) \right| n^{-1} \sum_{i=1}^{n} K_h(V_{ij} - v),
\]

which tends to zero in probability. Since \( Y_{ij}(t) \) is a decreasing function of \( t \), we have, for any \( \varepsilon > 0 \), the probability \( P(J_2 > \varepsilon) \) is bounded by

\[
MP \left( n^{-1} \sum_{l=1}^{n} I(t_{i-1} < X_{ij} < t_i)\psi^+(V_{ij}, (V_{ij} - v)/h, Z_{ij}(t_{i-1}))K_h(V_{ij} - v) > \varepsilon \right),
\]

It is easy to show that

\[
n^{-1} \sum_{l=1}^{n} I(t_{i-1} < X_{ij} < t_i)\psi^+(V_{ij}, (V_{ij} - v)/h, Z_{ij}(t_{i-1}))K_h(V_{ij} - v)
= \frac{P}{P} \int E\{I(t_{i-1} < X_{ij} < t_i)\psi^+(v, w, Z_{ij}(t_{i-1}))|V_j = v\}K(w)dw.
\]

Note that by the Cauchy-Schwartz inequality

\[
E(I(t_{i-1} < X_j < t_i)\psi^+(v, w, Z_{ij}(t_{i-1}))|V_j = v)
= |P(t_{i-1}|V_j = v) - P(t_i|V_j = v)|^{1/2}E^{1/2}(\psi^2(v, w, Z_{ij}(t_{i-1}))|V = v) < M^*\delta^{1/2},
\]

since \( |t_i - t_{i-1}| \leq \delta \), where \( M^* \) is some constant. Hence, \( J_2 \to 0 \) as \( n \to \infty \) first and then \( \delta \to 0 \).
The second term of (17) is bounded by

\[
\max_{1 \leq i \leq M} \sup_{|t - t_{i-1}| \leq \delta} f_j(v) \left| \int E \{ Y_j(t)(\psi^+(v, w, Z_j(t)) - \psi^+(v, w, Z_j(t_{i-1})) | V_j = v \} K(w) dw \right|
\]

which tends to zero as \( \delta \to 0 \). This implies that (17) holds. Hence, it completes the proof of Lemma 1.

**Lemma 2** Under Conditions (i)-(vi), we have for \( k = 0, 1, 2 \)

\[
S_{njk}^*(w) = s_{jk}^*(w) + o_p(1),
\]

uniformly for \( w \in (0, \tau] \), where \( s_{jk}^*(w) = s_{jk}(w, \theta_0, v) \) and in addition, under Conditions (viii)-(x), we have

\[
\sup_{w \in (0, \tau], v \in \Phi} ||S_{nj}(w, v) - s_{jk}(w, v)|| = o_p(1).
\]

Furthermore, we have

\[
\sup_{w \in (0, \tau]} ||S_{nj}(w, \zeta) - s_{jk}(w, \zeta)|| = o_p(1).
\]

Lemma 2 can be easily proved by Lemma 1.

**Proof of Theorem 1.** We first show that \( \hat{\zeta} \to 0 \) in probability, where \( \hat{\zeta} = H(\hat{\xi} - \xi_0) \), in which \( \hat{\xi} \) is the maximum local pseudo-partial likelihood estimator of \( \xi_0 \). Let

\[
X_{nj}(t, \zeta) = \frac{1}{n} \sum_{i=1}^{n} \int_{t-1}^{t} K_h(V_{ij} - v) \left[ \zeta^T U_{ij}^*(w) - \log \frac{S_{nj0}(w, \zeta)}{S_{nj0}(w, 0)} \right] dM_{ij}(w).
\]

Then, it is easy to show that

\[
\tilde{\ell}_n(t, \zeta) - \tilde{\ell}_n(t, 0) = \sum_{j=1}^{J} X_{nj}(t, \zeta) + Y_n(t, \zeta),
\]

where

\[
Y_n(t, \zeta) = \sum_{j=1}^{J} \int_{0}^{t} \left[ (S_{nj1}(w))^T \zeta - \log \frac{S_{nj0}(w, \zeta)}{S_{nj0}(w, 0)} \right] \lambda_{0j}(w) dw.
\]
By Lemma 2, we obtain that
\[
Y_n(t, \zeta) = \sum_{j=1}^{J} \int_0^t \left[ (s_{j1}(w))^T \zeta - \log \frac{s_{j0}(w, \zeta)}{s_{j0}(w, 0)} \right] \lambda_0(w) dw + o_P(1)
\equiv Y(t, \zeta) + o_P(1).
\]

By similar argument as in Andersen and Gill (1982), it can be shown that each term in the sum of the asymptotic representation of \(Y_n(t, \zeta)\) is a strictly concave function in \(\zeta\), and it has the maximum value at \(\zeta = 0\). The first term in (19) is a sum of \(J\) local square integrable martingale with the square variation process being
\[
< X_{nj}, X_{nj} > (t) = \frac{1}{n^2} \sum_{i=1}^{n} \int_0^t K_h^2(V_{ij} - v) \left[ \zeta^T U_{ij}(w) - \log \left( \frac{S_{nj0}(w, \zeta)}{S_{nj0}(w, 0)} \right) \right]^2 \times Y_{ij}(w) \exp(\beta_0(V_{ij})^T Z_{ij}(w) + g_0(V_{ij})) \lambda_0(w) dw.
\]

It follows from Lemma 1 that
\[
EX_{nj}^2(t, \zeta) = E < X_{nj}, X_{nj} > (t) = O((nh)^{-1}) \to 0, \ 0 < t \leq \tau.
\]
This implies that \(X_{nj}(t, \zeta) \to 0\) in probability for \(1 \leq j \leq J\). Hence we obtain that
\[
\hat{\ell}_n(t, \zeta) - \hat{\ell}_n(t, 0) = Y(t, \zeta) + o_P(1).
\]

We can easily show that \(\hat{\zeta}\) maximizes the strictly concave function \(\hat{\ell}_n(t, \zeta) - \hat{\ell}_n(t, 0)\). By Lemma A.1 of Carroll, Fan, Gijbels and Wand (1997), it follows that \(\hat{\zeta} \to 0\) in probability.

Now we prove the second result of Theorem 1. By the same argument as above, we can prove from Lemma 1 that
\[
\sup_{t \in [0, \tau]} \sup_{\zeta \in \mathbb{C}^*} \sup_{v \in \Phi_V} |\hat{\ell}_n(t, \zeta) - \hat{\ell}_n(t, 0) - Y(t, \zeta)| \longrightarrow 0
\]
in probability, here \(\zeta = H(\xi - \xi_0)\) and \(\mathbb{C}^*\) is a convex and compact set of \(R^{2p+1}\). Therefore, it follows from Lemma A.1 of Carroll, Fan, Gijbels and Wand (1997) that \(\sup_{v \in \Phi_V} |\hat{\zeta}| \longrightarrow 0\) in probability. Hence we complete the proof of Theorem 1.
Proof of Theorems 2 and 3. Note that we have proved in Theorem 1 that $\mathbf{H}(\tilde{\xi}(v) - \xi_0(v)) \to 0$ in probability. This result is very useful for proving Theorem 2. We divide the proofs into the following three steps.

(a) The asymptotic normality of $\tilde{\ell}_n'(0)$. The logarithm of the local pseudo-partial likelihood function can be written as

$$
\tilde{\ell}_n'(0) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{J} \int_0^\tau K_h(V_{ij} - v) \left[ U_{ij}'(w) - \frac{S_{nj1}(w,v)}{S_{nj0}(w,v)} \right] dM_{ij}(w)
$$

$$
+ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{J} \int_0^\tau K_h(V_{ij} - v) \left[ U_{ij}'(w) - \frac{S_{nj1}(w,v)}{S_{nj0}(w,v)} \right]
$$

$$
\times \exp \left( \beta_0(V_{ij})^T \mathbf{Z}_{ij}(w) + g_0(V_{ij}) \right) Y_{ij}(w) \lambda_{0j}(w) dw,
$$

$$
\equiv I_1(\tau,0) + I_2(\tau,0).
$$

We first deal with $I_2(\tau,0)$. Noting that

$$
I_2(\tau,0) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{J} \int_0^\tau \left( U_{ij}'(w) - \frac{S_{nj1}(w,v)}{S_{nj0}(w,v)} \right) \left[ \exp \{ \beta_0(V_{ij})^T \mathbf{Z}_{ij}(w) + g_0(V_{ij}) \} \right]
$$

$$
- \exp \left( \xi_0^T \mathbf{X}_{ij}^* + g_0(v) \right) K_h(V_{ij} - v) Y_{ij}(w) \lambda_{0j}(w) dw,
$$

it follows from Taylor’s expansion and Lemma 1 that

$$
I_2(\tau,0) = \frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{J} \int_0^\tau \left( U_{ij}'(w) - \frac{s_{j1}(w)}{s_{j0}(w)} \right) Y_{ij}(w) \exp \left( \xi_0^T \mathbf{X}_{ij}^* + g_0(v) \right)
$$

$$
\times \left[ \mathbf{B}_0(v)^T \mathbf{Z}_{ij}(w) + g_0''(v) \right] (V_{ij} - v)^2 K_h(V_{ij} - v) \lambda_{0j}(w) dw (1 + O_P(h))
$$

$$
= \frac{1}{2} h^2 \sum_{j=1}^{J} f_j(v) \int_0^\tau \mathbf{E} \left[ \left( \mathbf{Z}(w) \mu_2 \right) - \frac{s_{j1}(w)\mu_2}{s_{j0}(w)} \right] \rho(w, \mathbf{Z}(w), v)
$$

$$
\times \left[ \beta_0'(v)^T \mathbf{Z}(w) + g_0''(v) \right] V_j = v \lambda_{0j}(w) dw (1 + O_P(h)),
$$

where $s_{jk}(w) = s_{jk}(w, \theta_0, v)$ for $k = 0, 1, 2$. Since $K(\cdot)$ is a symmetric function, by simple calculation, we have

$$
I_2(\tau,0) = \mathbf{B}_n(\tau, v) = \frac{1}{2} h^2 \nu_2 (\Gamma^{-1} \beta_0''(v))^T, 0^T, 0]^T (1 + O_P(h)) (20).
$$
We now consider $I_1(\tau, 0)$. Let 

$$B_{nij}(\tau) = \int_0^\tau K_h(V_{ij} - v) \left[ \frac{U^*_ij(w) - s_{1j}(w, \xi, v)}{s_{0j}(w, \xi, v)} \right] dM_{ij}(w).$$

By Conditions (vi)-(x), Lemma 2 and Lemma A.1 of Spiekerman and Lin (1998), we can prove that by some tedious and routine calculation,

$$\frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(V_{ij} - v) \left[ \frac{S_{nj1}(w, v) - s_{1j}(w, v)}{S_{nj0}(w, v)} \right] dM_{ij}(w) = O_P((nh)^{-1/2}).$$

Hence, it follows that

$$I_1(\tau, 0) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^J B_{nij}(\tau) + o_P(1).$$

Note that $\sqrt{n}hI_1(\tau, 0)$ is a sum of iid random vectors $\sum_{j=1}^J B_{nij}(\tau)$ with zero mean and finite variance, the desired asymptotic normality follows from the multivariate central limit theorem by using the Liapounov condition. It can be shown that the asymptotic variance is

$$\Pi = \lim_{n \to \infty} Eh \left( \sum_{i=1}^J B_{nij}(\tau) \right)^2 = \sum_{j=1}^J \lim_{n \to \infty} Eh B_{nj1}(\tau) B_{nj1}(\tau)^T.$$ (21)

Note that $\sum_{i=1}^n B_{nij}(t)$ is a local square-integrable martingale with respect to the filtration $\bigcup_{i=1}^n F_{t,ij} = \sigma\{N_{ij}(w^-), Z_{ij}(w), V_{ij}, Y_{ij}(w), 0 \leq w \leq t, i = 1, 2, \cdots, n\}$. Hence, it can be shown that the first term of (21) converges to $D$. By Cauchy-Schwartz inequality, we easily know that $lim_{n \to \infty} Eh B_{nj1}(\tau) B_{nj1}(\tau)^T$ exists. Write $\Pi_{jl}(\tau, v) = \lim_{n \to \infty} Eh B_{nj1}(\tau) B_{nj1}(\tau)^T$. Hence, we can prove that the second term of (21) converges to $\Pi_{0}(\tau, v) = \sum_{l\neq j} \Pi_{lj}(\tau, v)$ for the limit matrix $\Pi_{lj}(\tau, v)$. The proof of Theorem 2 is completed by using the asymptotic results of $I_1(\tau, 0)$ and $I_2(\tau, 0)$.

(b) *Convergence of the Hessian matrix.* We shall show that the second derivative of the logarithm of the local pseudo-partial likelihood function converges to a finite constant matrix. We have shown in Theorem 1 that $\hat{\zeta} \to 0$ in probability. Hence by the mean-value theorem, we have

$$\bar{\ell}''(\hat{\zeta}) = \bar{\ell}''(0) + o_P(1).$$ (22)
Since \( s_{jk}^*(w) = s_{jk}(w) \exp(g_0(v)) \), \( k = 0, 1, 2 \), from Lemma 2, we can obtain

\[
\tilde{p}_n^\prime(0) = \frac{1}{n} \int_0^\tau n \sum_{i=1}^n \sum_{j=1}^J K_h(V_{ij} - v) \frac{s_{j2}^*(w) s_{j0}^*(w) - s_{j1}^*(w) (s_{j1}^*(w))^T}{(s_{j0}^*(w))^2} dN_{ij}(w) + o_P(1).
\]

Write \( F_u(w) = P(X \le w, \Delta = 1|V_j = w) \) and its corresponding conditional empirical distribution \( \tilde{F}_{nj}(w) = \frac{1}{n} \sum_{i=1}^n K_h(V_{ij} - v)I(X_{ij} \le w, \Delta_{ij} = 1) \). By the conventional argument in the kernel smoothing, together with the empirical process theory, it can be shown that

\[
\tilde{p}_n^\prime(0) = \frac{1}{n} \int_0^\tau \sum_{i=1}^n \sum_{j=1}^J K_h(V_{ij} - v) \frac{s_{j2}^*(w) s_{j0}^*(w) - s_{j1}^*(w) (s_{j1}^*(w))^T}{(s_{j0}^*(w))^2} d\tilde{F}_{nj}(w) = -A(\tau, v) + o_P(1), \tag{23}
\]

where \( A(\tau, v) = \sum_{j=1}^J \int_0^\tau \frac{s_{j2}^*(w) s_{j0}^*(w) - s_{j1}^*(w) (s_{j1}^*(w))^T}{(s_{j0}^*(w))^2} dF_u(w) \). It is easy to show that \( A(\tau, v) \) is positive definite by Condition (vii).

(c) Asymptotic normality of \( \tilde{\xi}(v) \). Since \( \tilde{\xi} \) maximizes \( \tilde{\ell}_n(\xi) \), by Taylor’s expansion around 0, we have

\[
-\tilde{\ell}_n(0) = \tilde{\ell}_n(\tilde{\xi}) - \tilde{\ell}_n(0) = (\tilde{\ell}_n(\tilde{\xi}^*)^T \tilde{\xi},
\]

where \( \tilde{\xi}^* \) lies between 0 and \( \tilde{\xi} \) (strictly speaking, the immediate point can depend on the element of \( \tilde{p}_n^\prime \), but this does not alter the proof). Hence \( \tilde{\xi}^* \to 0 \) in probability.

It follows from (23) that

\[
\tilde{\xi} - A(\tau, v)^{-1} B_n(\tau, v) = -\left(\tilde{\ell}_n(\tilde{\xi}^*)\right)^{-1} \left(\tilde{\ell}_n(0) - B_n(\tau, v)\right) + o_P(1).
\]

By Theorem 1, (23), and Slutsky’s theorem, we obtain that

\[
\sqrt{nh} \left(\tilde{\xi} - A(\tau, v)^{-1} B_n(\tau, v)\right) \to N(0, A^{-1}(\tau, v) \Pi(\tau, v) A^{-1}(\tau, v)).
\]

Now we simplify the matrix \( A(\tau, v) \). By some simple calculation, we have

\[
s_{j2}^*(w) = \begin{pmatrix} a_{j2}(w) & 0 & 0 \\ 0 & a_{j2}(w) \mu_2 & a_{j1}(w) \mu_2 \\ 0 & a_{j1}(w) \mu_2 & a_{j0}(w) \mu_2 \end{pmatrix}.
\tag{24}
\]

Similarly, we obtain that \( (s_{j1}^*(w))^{\otimes 2} = diag(a_{j1}(w) a_{j1}^T(w), 0) \). Note that \( s_{j0}^*(w) = a_{j0}(w) \). By some tedious basic calculation, we have \( A(\tau, v) = diag(\Gamma^{-1}, Q_{2\mu_2}) \).
Hence, the asymptotic bias of estimator $\hat{\xi}(v)$ is $b(\tau, v) = A^{-1}(\tau, v)B_n(\tau, v) = h^2e_2\xi''_0(v)\mu_2/2$, and the asymptotic covariance is

\[
\Sigma(\tau, v) = A^{-1}(\tau, v)\Pi(\tau, v)(A^{-1}(\tau, v))^T = \text{diag}(\Gamma, Q\mu_2^2\nu_2) + A^{-1}\Pi_0(A^{-1})^T.
\]

Therefore, we have finished the proof of the asymptotic normality of the maximum local pseudo-partial likelihood function estimator.

From the proof of Theorems 2 and 3, we have the following result. Under $nh^4 \to 0$, then

\[
\sup_{\nu \in \Phi_{V,\epsilon}} |(nh)^{1/2}H(\hat{\xi}(u) - \xi_0(u)) - A^{-1}(\tau, \xi_0(u))\ell'_{n}(\xi_0(u), \tau)| = o_P(h^{-1/2}). \tag{25}
\]

**Proof of Theorem 4.** By similar arguments as those used for proving Lemma 1, it can be shown that

\[
\sup_{t \in [0, \tau]} \sup_{||\theta - \theta_0|| \leq ||\hat{\theta} - \theta_0||} n^{-1}||\psi_{n}\jmath(t, \theta) - \psi_{n}\jmath(t, \theta_0)|| \to 0 \tag{26}
\]

in probability, where

\[
\psi_{n}\jmath(t, \theta) = \sum_{i=1}^{n} I(V_{ij} \in J_{ij})Y_{ij}(t)\exp\{\beta^T(V_{ij})Z_{ij}(t) + g(V_{ij})\},
\]

where $\theta = (\beta^T(\cdot), g(\cdot))^T$. By the definition of $\hat{\Lambda}_{0j}(t)$, we can show that $\hat{\Lambda}_{0j}(t) - \Lambda_{0j}(t)$ can be represented by a summation of three terms that are functionals of $\psi_{n}\jmath(t, \theta) - \psi_{n}\jmath(t, \theta_0)$, and it follows that these terms are negligible in the sense of probability. Hence, $\hat{\Lambda}_{0j}(t) \to \Lambda_{0j}(t)$, uniformly on $(0, \tau]$. Therefore, we can prove by the standard argument of kernel estimation that $\hat{\lambda}_{0j}(t) \to \lambda_{0j}(t)$, uniformly on $(0, \tau]$.

**Proof of Theorem 5.** By (25), Theorem 2 and the similar argument of Theorem 3 in Spiekerman and Lin (1998), we can prove Theorem 5.

**REFERENCES**


Table 1: Summary of simulation results based on local pseudo-partial likelihood procedures

<table>
<thead>
<tr>
<th>$v$</th>
<th>$h$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
<th>$\hat{g}'$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>bias SE SD</td>
<td>bias SE SD</td>
<td>bias SE SD</td>
</tr>
<tr>
<td>0.5</td>
<td>0.075</td>
<td>-0.071 0.194 0.262</td>
<td>-0.179 0.266 0.370</td>
<td>-0.059 1.931 1.501</td>
</tr>
<tr>
<td>0.100</td>
<td>-0.036 0.158 0.177</td>
<td>-0.095 0.215 0.266</td>
<td>-0.011 1.118 0.956</td>
<td></td>
</tr>
<tr>
<td>0.150</td>
<td>-0.007 0.121 0.133</td>
<td>-0.004 0.160 0.175</td>
<td>0.003 0.538 0.493</td>
<td></td>
</tr>
<tr>
<td>0.200</td>
<td>0.028 0.096 0.101</td>
<td>0.077 0.121 0.125</td>
<td>0.044 0.328 0.295</td>
<td></td>
</tr>
<tr>
<td>0.400</td>
<td>0.077 0.076 0.086</td>
<td>0.252 0.087 0.108</td>
<td>0.247 0.167 0.123</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.075</td>
<td>-0.041 0.187 0.252</td>
<td>-0.173 0.279 0.354</td>
<td>0.085 1.817 1.495</td>
</tr>
<tr>
<td>0.100</td>
<td>-0.009 0.153 0.188</td>
<td>-0.094 0.218 0.259</td>
<td>0.079 1.140 0.933</td>
<td></td>
</tr>
<tr>
<td>0.150</td>
<td>0.004 0.115 0.118</td>
<td>0.006 0.156 0.164</td>
<td>-0.007 0.456 0.451</td>
<td></td>
</tr>
<tr>
<td>0.200</td>
<td>0.027 0.092 0.093</td>
<td>0.101 0.113 0.124</td>
<td>-0.066 0.305 0.257</td>
<td></td>
</tr>
<tr>
<td>0.400</td>
<td>0.108 0.063 0.064</td>
<td>0.382 0.066 0.082</td>
<td>-0.055 0.132 0.093</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>0.075</td>
<td>-0.004 0.174 0.231</td>
<td>-0.031 0.181 0.236</td>
<td>0.066 1.843 1.527</td>
</tr>
<tr>
<td>0.100</td>
<td>0.016 0.142 0.164</td>
<td>-0.015 0.145 0.157</td>
<td>0.032 1.040 0.949</td>
<td></td>
</tr>
<tr>
<td>0.150</td>
<td>0.019 0.110 0.114</td>
<td>-0.007 0.116 0.115</td>
<td>0.035 0.533 0.496</td>
<td></td>
</tr>
<tr>
<td>0.200</td>
<td>0.023 0.092 0.094</td>
<td>0.010 0.096 0.097</td>
<td>0.035 0.329 0.304</td>
<td></td>
</tr>
<tr>
<td>0.400</td>
<td>0.075 0.061 0.060</td>
<td>0.063 0.065 0.068</td>
<td>-0.181 0.131 0.095</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>0.075</td>
<td>0.097 0.196 0.277</td>
<td>0.145 0.235 0.326</td>
<td>0.140 1.886 1.503</td>
</tr>
<tr>
<td>0.100</td>
<td>0.076 0.167 0.195</td>
<td>0.070 0.192 0.224</td>
<td>0.104 1.075 0.956</td>
<td></td>
</tr>
<tr>
<td>0.150</td>
<td>0.047 0.129 0.142</td>
<td>0.004 0.143 0.139</td>
<td>0.078 0.566 0.521</td>
<td></td>
</tr>
<tr>
<td>0.200</td>
<td>0.037 0.107 0.112</td>
<td>-0.044 0.114 0.117</td>
<td>0.091 0.355 0.339</td>
<td></td>
</tr>
<tr>
<td>0.400</td>
<td>0.049 0.072 0.076</td>
<td>-0.254 0.068 0.076</td>
<td>-0.098 0.138 0.116</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>0.075</td>
<td>0.206 0.315 0.413</td>
<td>0.141 0.261 0.329</td>
<td>0.470 1.982 1.635</td>
</tr>
<tr>
<td>0.100</td>
<td>0.135 0.260 0.305</td>
<td>0.057 0.206 0.251</td>
<td>0.210 1.160 1.058</td>
<td></td>
</tr>
<tr>
<td>0.150</td>
<td>0.074 0.200 0.216</td>
<td>-0.004 0.151 0.166</td>
<td>0.095 0.633 0.570</td>
<td></td>
</tr>
<tr>
<td>0.200</td>
<td>0.039 0.166 0.179</td>
<td>-0.074 0.119 0.128</td>
<td>0.005 0.385 0.384</td>
<td></td>
</tr>
<tr>
<td>0.400</td>
<td>-0.014 0.125 0.130</td>
<td>-0.222 0.087 0.098</td>
<td>-0.261 0.189 0.181</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Comparison of the local pseudo-partial likelihood estimator (P) and the weighted average estimator (W)

<table>
<thead>
<tr>
<th>EST</th>
<th>$\tilde{\beta}_1(\cdot)$</th>
<th>$\tilde{\beta}_2(\cdot)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abias</td>
<td>SD</td>
</tr>
<tr>
<td>$\theta = 0.25$</td>
<td>P</td>
<td>0.0693</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>0.0767</td>
</tr>
<tr>
<td>$\theta = 4.00$</td>
<td>P</td>
<td>0.0653</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>0.0750</td>
</tr>
</tbody>
</table>

**Note:** Abias is the average absolute bias of estimator $\hat{\beta}_j$ for $j = 1, 2$, and RASE denotes the square root of average square errors of estimator $\hat{\beta}_j$. 
Table 3: Comparison of the average square errors of the local pseudo-partial likelihood estimator (P) with the one-step estimator (OS)

<table>
<thead>
<tr>
<th>h</th>
<th>estimator</th>
<th>θ = 0.25, c = 2</th>
<th>mean</th>
<th>median</th>
<th>std</th>
<th>θ = 4.0, c = 2</th>
<th>mean</th>
<th>median</th>
<th>std</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>P</td>
<td></td>
<td>0.0879</td>
<td>0.0738</td>
<td>0.0602</td>
<td></td>
<td>0.0710</td>
<td>0.0569</td>
<td>0.0550</td>
</tr>
<tr>
<td></td>
<td>OS</td>
<td></td>
<td>0.0879</td>
<td>0.0740</td>
<td>0.0603</td>
<td></td>
<td>0.0716</td>
<td>0.0573</td>
<td>0.0551</td>
</tr>
<tr>
<td>0.2</td>
<td>P</td>
<td></td>
<td>0.0276</td>
<td>0.0199</td>
<td>0.0239</td>
<td></td>
<td>0.0287</td>
<td>0.0220</td>
<td>0.0273</td>
</tr>
<tr>
<td></td>
<td>OS</td>
<td></td>
<td>0.0276</td>
<td>0.0198</td>
<td>0.0239</td>
<td></td>
<td>0.0287</td>
<td>0.0220</td>
<td>0.0272</td>
</tr>
<tr>
<td>0.4</td>
<td>P</td>
<td></td>
<td>0.0107</td>
<td>0.0084</td>
<td>0.0020</td>
<td></td>
<td>0.0216</td>
<td>0.0140</td>
<td>0.0214</td>
</tr>
<tr>
<td></td>
<td>OS</td>
<td></td>
<td>0.0107</td>
<td>0.0084</td>
<td>0.0020</td>
<td></td>
<td>0.0216</td>
<td>0.0140</td>
<td>0.0214</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>h</th>
<th>estimator</th>
<th>θ = 0.25, c = 5</th>
<th>mean</th>
<th>median</th>
<th>std</th>
<th>θ = 4.0, c = 5</th>
<th>mean</th>
<th>median</th>
<th>std</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>P</td>
<td></td>
<td>0.0350</td>
<td>0.0279</td>
<td>0.0256</td>
<td></td>
<td>0.0584</td>
<td>0.0499</td>
<td>0.0506</td>
</tr>
<tr>
<td></td>
<td>OS</td>
<td></td>
<td>0.0350</td>
<td>0.0278</td>
<td>0.0255</td>
<td></td>
<td>0.0586</td>
<td>0.0500</td>
<td>0.0506</td>
</tr>
<tr>
<td>0.2</td>
<td>P</td>
<td></td>
<td>0.0200</td>
<td>0.0137</td>
<td>0.0182</td>
<td></td>
<td>0.0205</td>
<td>0.0148</td>
<td>0.0199</td>
</tr>
<tr>
<td></td>
<td>OS</td>
<td></td>
<td>0.0200</td>
<td>0.0137</td>
<td>0.0182</td>
<td></td>
<td>0.0204</td>
<td>0.0148</td>
<td>0.0199</td>
</tr>
<tr>
<td>0.4</td>
<td>P</td>
<td></td>
<td>0.0162</td>
<td>0.0113</td>
<td>0.0148</td>
<td></td>
<td>0.0162</td>
<td>0.0117</td>
<td>0.0146</td>
</tr>
<tr>
<td></td>
<td>OS</td>
<td></td>
<td>0.0162</td>
<td>0.0113</td>
<td>0.0148</td>
<td></td>
<td>0.0162</td>
<td>0.0117</td>
<td>0.0146</td>
</tr>
</tbody>
</table>

**Note:** “mean”, “median” and “std” denote the average, median and sample standard derivation of the average square errors, respectively, based on 300 simulations.
Figure 1: Simulation results for the local pseudo-partial likelihood estimator with 200 clusters and the dependence parameter of failure time $\theta = 0.25$. (a), (b) and (c) provide the average estimates of $\beta_1(\cdot)$, $\beta_2(\cdot)$ and $g(\cdot)$ for heavily censored data with $c = 2$, respectively. Solid curves are true functions. Three bandwidths are used: dash-dotted curves for bandwidth $h = 0.1$, dashed curves for $h = 0.2$, and dotted curves for $h = 0.4$.

Figure 2: Simulation results for the comparison of the maximum local pseudo-partial likelihood estimator (P) with the one-step estimator (OS). (a) The boxplots for the distribution of the ASE over the 300 replications, using three bandwidths $h = 0.1, 0.2, 0.4$ (from left to right). Column numbers 1, 3 and 5 denote maximum local pseudo-partial likelihood estimator, and column numbers 2, 4 and 6 plot one-step estimator (OS) for heavily censored data with $c = 2$; (b) the same as (a) for moderately censored data $c = 5$. 
Figure 3: Data analysis for Busselton Population Health Surveys study. The marginal hazard rate model is
\[ \lambda_{ij}(t) = \lambda_0(t) \exp(\sum_{k=1}^{5} \beta_k(V_{ij}(t))Z_{ij}(t) + g(V_{ij}(t))) \],
where \( V(t) \) = age and \( Z^T(t) = (\text{Gender, BMI, CHOL, Smoke1, Smoke2}) \), corresponding to the plots (a)-(e), respectively. (f) is the plot of \( \hat{g}'(\cdot) \), and (g) is the plot of \( \hat{g}(\cdot) \), where smoke1 is coded as 1 for ex-smoker and 0 otherwise; and smoke2 is coded as 1 for current smoker and 0 otherwise. The dotted curve is the confidence curve on nominal level \( \alpha = 0.05 \). In this setting, the chosen bandwidth is \( h = 0.15(\max(\text{age}) - \min(\text{age})) = 10.905 \). The x-axis is for age.