Ann Inst Stat Math (2015) 67:817–841
DOI 10.1007/s10463-014-0484-7

Partially varying coefficient single-index additive hazard models

Xuan Wang · Qihua Wang · Xiao-Hua Andrew Zhou

Received: 27 August 2013 / Revised: 15 February 2014 / Published online: 12 September 2014
© The Institute of Statistical Mathematics, Tokyo 2014

Abstract The partially linear additive hazards model has been proposed to study the interaction between some covariates and an exposure variable. In this paper, we extend it to the partially varying coefficient single-index additive hazard model where the high dimension covariates are collapsed to a single index, due to practical needs. Two sets of estimating equations were proposed to estimate the varying coefficient functions in the linear components: the link function for the single index and the single-index parameter vector separately. It was shown that the proposed local and global estimators are asymptotically normal. Simulation studies were conducted to examine the finite-sample performance of our method to compare the relative performance of our method with existing ones. A real data analysis was used to illustrate the proposed methods.

Keywords Varying coefficient · Partially linear single-index · Two sets of estimating functions · Iteration · Asymptotic normality

1 Introduction

The additive (Aalen 1989; Lin and Ying 1994) and multiplicative risk (Cox 1972) models are the two commonly used models for studying the relationship between

X. Wang · Q. Wang (✉)
Academy of Mathematics and Systems Science, Chinese Academy of Sciences,
Beijing 100190, People's Republic of China
e-mail: qhwang@amss.ac.cn

Q. Wang
Institute of Statistical Science, Shenzhen University,
Shenzhen 518060, People's Republic of China

X.-H. A. Zhou
Department of Biostatistics, University of Washington, Seattle, WA 98198, USA

risk factors and disease occurrence or death. Both the additive and multiplicative models are motivated from applications in biology and have been studied extensively in statistics. They together provide complementary risk analysis of covariate effects. The additive and mutiplicative models can be extended to enhance model flexibility by incorporating time-varying coefficients (see e.g., Cai and Sun 2003; Tian et al. 2005). Furthermore, in many applications, the covariate effects may vary with an exposure variable. The formulation of Fan et al. (2006) characterizes the extent to which the association varies with the level of the exposure variable and is well suited for exploring the nonlinear interaction effects between risk factors. They used local partial likelihood to estimate coefficient functions in multiplicative hazards model, which contains such nonlinear effects.

The additive models are useful alternatives to the multiplicative models when absolute rate differences are of interest. Especially, the risk differences from additive risk models offer additional survival information beyond the risk ratio, which is particularly important in epidemiology and public health study (Breslow and Day 1980, 1987). Various forms of additive models have been studied, including fully nonparametric, partially linear semiparametric models and partially linear varying coefficients models (Huffer and McKeague 1991; McKeague and Sasieni 1994; Yin et al. 2008). To characterize the varying covariate effects in the additive hazard models, Yin et al. (2008) proposed the following partially linear varying coefficient additive model:

$$\lambda(t|\mathbf{Z}, \mathbf{X}, V(t)) = \lambda_0(t) + \beta(V(t))^\top \mathbf{Z}(t) + g(V(t)) + \mathbf{X}(t)^\top \gamma,$$

where $V(t)$ is the exposure variable, both $\mathbf{Z}(t)$ and $\mathbf{X}(t)$ are covariate vectors, of which $\mathbf{Z}(t)$ may interact nonlinearly with the exposure variable $V(t)$, $\lambda_0(t)$ is the baseline hazard function, $\beta(V(t))$ characterizes the nonlinear interaction between $\mathbf{Z}(t)$ and $V(t)$, $g(V(t))$ represents the main effect of $V(t)$ and γ is an unknown parameter vector.

However, in many biomedical studies, the covariate effects may be much more complex than linear effect and new challenges arise in assessing nonlinear effects. This motivates us to consider the extension of the above model to the following partially varying coefficient single-index additive hazards model (PVCSIAHM):

$$\lambda(t|\mathbf{Z}(t), \mathbf{X}(t), V(t)) = \lambda_0(t) + \beta(V(t))^\top \mathbf{Z}(t) + g(V(t)) + \omega(\mathbf{X}(t)^\top \gamma), \quad (1)$$

with $\beta(\cdot)$, $g(\cdot)$ and $\omega(\cdot)$ being unknown functions. For example, in medical studies, $\mathbf{Z}(t)$ may represent the treatment, $V(t)$ represents the biomarker value and $\mathbf{X}(t)$ a vector of explanatory variables such as a vector of age and other variables. Then, the capability of a biomarker to predict the hazard of a patient to one particular treatment over another can be represented by the above model. Besides the estimates of $\beta(\cdot)$ and $g(\cdot)$, we are interested in identifying the effect of explanatory variables, in other words, the estimates of $\omega(\cdot)$ and γ. For ease of presentation, the dependence of covariates on time is dropped as the methods and proofs in this paper are applicable to time-dependent covariates. For the sake of identifiability, we assume that $\|\gamma\| = 1$ and the first component of γ is positive, and $\omega(0) = 0$, $g(0) = 0$, where $(0, 0)$ belongs to the interior of the support of $(\mathbf{X}^\top\gamma, V)$ denoted by $(\mathcal{W}_1, \mathcal{W}_2)$.

Recall that, in regression problems, there are various estimating methods for the partially linear single-index models (PLSIM), such as the backfitting algorithm proposed by Carroll et al. (1997), the penalized spline estimation procedure proposed by Yu and Ruppert (2002) and the minimum average variance estimation method by Xia and Hardle (2006). However, the existing methods for PLSIM can not be applied to model (1) directly since model (1) is a hazard regression model.

To the best of my knowledge, model (1) has not been considered in the literature. For the model of Yin et al. (2008), they used an estimating equation to obtain local estimators of varying coefficient functions and constant parameter vector γ, then use a weighted average method to obtain a global estimator of the global parameter vector γ. Comparing to the model of Yin et al. (2008), new analytic challenges arise in assessing the single-index part and nonparametric effect of it for model (1) because they cannot be estimated simultaneously from just one estimating function. In this paper, we develop an iterative approach based on two sets of estimating equations to estimate the unknown functions $\beta(\cdot)$, $g(\cdot)$, $\omega(\cdot)$ and unknown single-index parameter vector γ, respectively, in which the first estimating function is a generalization of the method of Yin et al. (2008) to more complicated case.

This paper is organized as follows. In Sect. 2, we propose two sets of estimating equations to estimate the unknown functions and unknown parameter vector in PVC-SIAHM. In Sect. 3, we establish the asymptotic theories for the proposed local and global estimators. In Sect. 4, we examine the finite sample property using simulation studies and use a real data set to illustrate the proposed method. In Sect. 5, we give concluding remarks. The detailed proofs are delayed in Appendix A.

2 Estimating procedures

For subject i, let T_i be the failure time and C_i be the censoring time, then $\tilde{T}_i = T_i \bigwedge C_i$ is the observed time. Let $\Delta_i = I(T_i \le C_i)$ be the failure indicator function. The covariates are a p-vector $\mathbf{Z}_i$, a q-vector $\mathbf{X}_i$ and a scaler V_i. Assume that T_i and C_i are conditionally independent given the covariates, we further assume that the observed data $\{\mathbf{X}_i, \mathbf{Z}_i, V_i, \tilde{T}_i, \Delta_i\}$ are independent and identically distributed for $i = 1, \ldots, n$.

We write $N_i(t) = I[\tilde{T}_i \le t, \Delta_i = 1]$ and $Y_i(t) = I[\tilde{T}_i \ge t]$. Let the filtration$\{\mathscr{F}_t : t \in [0, \tau]\}$ be the data history up to time τ that is $\mathscr{F}_t = \sigma\{N_i(s), Y_i(s), 0 \le s \le t, i = 1, 2, \ldots, n\}$. Define $M_i(t) = N_i(t) - \int_0^t Y_i(s)\lambda(s|\mathbf{Z}_i, \mathbf{X}_i, V_i)\,\mathrm{d}s$. Then, $M_i(t)$ is a $\mathscr{F}_t$-martingale.

We will use local linear method to approximate the hazard function (1). Assume $\beta(\cdot)$, $g(\cdot)$, $\omega(\cdot)$ are smooth so that their first and second derivatives exist. By Taylor expansion, we have

$$\begin{aligned}
\beta(\tilde{v}) &\approx \beta(v) + \dot{\beta}(v)(\tilde{v} - v),\\
g(\tilde{v}) &\approx g(v) + \dot{g}(v)(\tilde{v} - v),\\
\omega(\mathbf{X}^\top\gamma) &\approx \omega(u) + \dot{\omega}(u)(\mathbf{X}^\top\gamma - u),
\end{aligned}$$

where (u, v) belongs to $(\mathcal{W}_1, \mathcal{W}_2)$; $\dot{\beta},\ \dot{g},\ \dot{\omega}$ are the derivatives of $\beta,\ g,\ \omega$, respectively. Then, model (1) can be approximated by

$$\lambda(t|\mathbf{Z}, \mathbf{X}, V; u, v) \approx \lambda_0^*(t, u, v) + \xi(u, v)^\top \mathbf{Z}^*(u, v),$$

where

$$\begin{aligned}\lambda_0^*(t, u, v) &= \lambda_0(t) + g(v) + \omega(u),\\ \xi(u, v) &= \Big(\beta(v)^\top, \dot{\beta}(v)^\top, \dot{g}(v), \dot{\omega}(u)\Big)^\top,\\ \mathbf{Z}^*(u, v) &= (\mathbf{Z}^\top, \mathbf{Z}^\top(V - v), V - v, \mathbf{X}^\top\gamma - u)^\top.\end{aligned}$$

To estimate $\xi(u, v)$ and γ, we develop an estimation approach with two sets of estimating equations because $\xi(u, v)$ and γ cannot be estimated simultaneously from just one estimating equation due to the interaction of $w(\cdot)$ and γ and the fact that $\xi(u, v)$ is a function parameter vector and γ is a constant parameter vector. The first estimating function, which will be defined by (2), is a local pseudoscore function, which can only be used to obtain local estimators. The second estimating function in (4) is also a form of pseudoscore function, but global one, and hence can be used to define estimator for the global parameter vector γ. We need to iterate between these two estimating functions to get the final estimates.

Define

$$\bar{M}_i(t) = N_i(t) - \int_0^t Y_i(s)\{\lambda_0^*(s, u, v) + \xi^\top(u, v)\mathbf{Z}_i^*(u, v)\}\,\mathrm{d}s$$

and

$$\bar{\mathbf{Z}}(t, u, v) = \frac{\sum_{i=1}^n K_{1,h_1}(V_i - v)K_{2,h_2}(\mathbf{X}_i{}^\top\gamma - u)Y_i(t)\mathbf{Z}_i^*(u, v)}{\sum_{i=1}^n K_{1,h_1}(V_i - v)K_{2,h_2}(\mathbf{X}_i{}^\top\gamma - u)Y_i(t)},$$

where $K_1(\cdot)$ and $K_2(\cdot)$ are kernel functions, h_1 and h_2 are bandwidths, and $K_{1,h_1}(\cdot) = K_1(\cdot/h_1)/h_1$, $K_{2,h_2}(\cdot) = K_2(\cdot/h_2)/h_2$.

For the estimation of ξ, we propose the following local pseudoscore function

$$\begin{aligned}U_n(\xi, \gamma; u, v) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau & K_{1,h_1}(V_i - v)K_{2,h_2}(\mathbf{X}_i{}^\top\gamma - u)\{\mathbf{Z}_i^*(u, v)\\ & -\bar{\mathbf{Z}}(t, u, v)\}\,\mathrm{d}\bar{M}_i(t).\end{aligned} \tag{2}$$

Given γ, we denote the solution to $U_n(\xi, \gamma; u, v) = 0$ by $\tilde{\xi}(\gamma, u, v) = (\tilde{\beta}(u, v)^\top,$ $\tilde{\dot{\beta}}(u, v)^\top, \tilde{\dot{g}}(u, v), \tilde{\dot{\omega}}(u, v))^\top$. We can obtain an analytic form of $\tilde{\xi}(\gamma, u, v)$ as follows

$$\tilde{\xi}(\gamma, u, v) = \left[\sum_{i=1}^n \int_0^\tau K_{1,h_1}(V_i - v)K_{2,h_2}(\mathbf{X}_i{}^\top\gamma - u)Y_i(t)\{\mathbf{Z}_i^*(u, v) - \bar{\mathbf{Z}}(t, u, v)\}^{\otimes 2}\,\mathrm{d}t\right]^{-1}$$

$$\times \left[\sum_{i=1}^{n} \int_0^{\tau} K_{1,h_1}(V_i - v) K_{2,h_2}(\mathbf{X}_i^\top \gamma - u) \{\mathbf{Z}_i^*(u, v) - \bar{\mathbf{Z}}(t, u, v)\} \, \mathrm{d}N_i(t) \right].$$

The above formula gives estimates $\tilde{\beta}(u, v)^\top$, $\tilde{\dot{\beta}}(u, v)^\top$, $\tilde{\dot{g}}(u, v)$, $\tilde{\dot{\omega}}(u, v)$, which are local estimates depending on both the variables u and v. To get the estimates of the functions $\beta(v)^\top$, $\dot{\beta}(v)^\top$, $\dot{g}(v)$, $\dot{\omega}(u)$, we define

$$\begin{aligned} \tilde{\beta}(V_i) &= \int_{\mathcal{W}_2} \int_{\mathcal{W}_1} I(v \le V_i) \Gamma_1(u) \, \tilde{\dot{\beta}}(u, v) \, \mathrm{d}u \, \mathrm{d}v, \\ \tilde{g}(V_i) &= \int_{\mathcal{W}_2} \int_{\mathcal{W}_1} I(v \le V_i) \Gamma_2(u) \, \tilde{\dot{g}}(u, v) \, \mathrm{d}u \, \mathrm{d}v, \\ \tilde{\omega}(X_i^\top \gamma) &= \int_{\mathcal{W}_2} \int_{\mathcal{W}_1} I(u \le X_i^\top \gamma) \Gamma_3(v) \, \tilde{\dot{\omega}}(u, v) \, \mathrm{d}u \, \mathrm{d}v, \\ \text{and} \quad \tilde{\dot{\omega}}(X_i^\top \gamma) &= \int_{\mathcal{W}_2} \tilde{\dot{\omega}}(X_i^\top \gamma, v) \, \Gamma_4(v) \, \mathrm{d}v, \quad \text{respectively,} \end{aligned} \tag{3}$$

where $\Gamma_1(\cdot)$, $\Gamma_2(\cdot)$, $\Gamma_3(\cdot)$ and $\Gamma_4(\cdot)$ are weight functions satisfying $\int_{\mathcal{W}_1} \Gamma_1(u) \, \mathrm{d}u = I_{p \times p}$, an identity matrix, $\int_{\mathcal{W}_1} \Gamma_2(u) \, \mathrm{d}u = 1$, $\int_{\mathcal{W}_2} \Gamma_3(v) \, \mathrm{d}v = 1$ and $\int_{\mathcal{W}_2} \Gamma_4(v) \, \mathrm{d}v = 1$. And, let

$$\begin{aligned} \mathbf{Z}_i^* &= (\mathbf{Z}_i^\top, 1, 1)^\top, \\ \zeta_{0i}(\gamma) &= (\beta_0(V_i)^\top, g_0(V_i), \omega_0(X_i^\top \gamma))^\top, \\ \tilde{\zeta}_i(\gamma) &= (\tilde{\beta}(V_i)^\top, \tilde{g}(V_i), \tilde{\omega}(X_i^\top \gamma))^\top. \end{aligned}$$

Then,

$$\tilde{\zeta}_i(\gamma) = \int_{\mathcal{W}_1} \int_{\mathcal{W}_2} R_i(\gamma, u, v) \, \tilde{\xi}(\gamma, u, v) \, \mathrm{d}u \, \mathrm{d}v,$$

where $\beta_0(\cdot)$, $g_0(\cdot)$ and $\omega_0(\cdot)$ are the true values of $\beta(\cdot)$, $g(\cdot)$ and $\omega(\cdot)$, respectively, and

$$R_i(\gamma, u, v) = \begin{pmatrix} 0 & I(v \le V_i)\Gamma_1(u) & 0 & 0 \\ 0 & 0 & I(v \le V_i)\Gamma_2(u) & 0 \\ 0 & 0 & 0 & I(u \le X_i^\top \gamma)\Gamma_3(v) \end{pmatrix}_{(p+2)\times(2p+2)}.$$

We assume that $\gamma_1 > 0$ and let $\gamma^{(1)} = (\gamma_2, \gamma_3, \ldots, \gamma_q)^\top$ be a $q-1$ dimensional parameter vector after removing the 1st component γ_1 in γ. Recall that $\|\gamma\| = 1$, then

$$\gamma = \gamma(\gamma^{(1)}) = \left((1 - \|\gamma^{(1)}\|^2)^{1/2}, \gamma_2, \gamma_3, \ldots, \gamma_q \right)^\top.$$

This "remove one component" method for γ is also used in Yu and Ruppert (2002) to avoid taking derivatives to the boundary points of the unit ball.

To obtain the estimator for γ, we consider a Jacobian matrix of γ with respect to $\gamma^{(1)}$,

$$\mathbf{J}_{\gamma^{(1)}} = \frac{\partial \gamma}{\partial \gamma^{(1)}} = \left(-\gamma^{(1)}/(1 - \|\gamma^{(1)}\|^2)^{1/2}, \mathbf{I}_{q-1}\right)^\top.$$

Define

$$\tilde{M}_i(t) = N_i(t) - \int_0^t Y_i(s)\{\lambda_0(s) + \tilde{\beta}(V_i)^\top \mathbf{Z}_i + \tilde{g}(V_i) + \tilde{\omega}(\mathbf{X}_i^\top \gamma)\}\, \mathrm{d}s,$$
$$\text{and} \quad \bar{\mathbf{X}}(t) = \frac{\sum_{i=1}^n \tilde{\dot{\omega}}(\mathbf{X}_i^\top \gamma)\mathbf{X}_i Y_i(t)}{\sum_{i=1}^n Y_i(t)}.$$

Then, γ can be estimated by the solution of the following equation

$$\tilde{U}_n(\gamma^{(1)}) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \mathbf{J}_{\gamma^{(1)}}{}^\top \{\tilde{\dot{\omega}}(\mathbf{X}_i^\top \gamma)\mathbf{X}_i - \bar{\mathbf{X}}(t)\}\, \mathrm{d}\tilde{M}_i(t) = 0. \tag{4}$$

Let the solution be $\tilde{\gamma}^{(1)}$. This defines the estimator of γ, say $\tilde{\gamma}$.

Given γ, we can obtain $\tilde{\xi}(\gamma, u, v)$ by solving the equation $U_n(\xi, \gamma; u, v) = 0$ and hence obtain the estimators of $\beta(\cdot), g(\cdot), \omega(\cdot)$ by (3), which are used to update the estimate of γ by (4). We denote the final estimator of γ by $\hat{\gamma}$, the final estimator of $\xi(u, v)$ by $\hat{\xi}(u, v) = \tilde{\xi}(\hat{\gamma}, u, v)$.

Remark The partially varying coefficient single-index additive hazard model (PVC-SIAHM) stems from real data analysis although it is an extension of the model of YLZ. We developed a two set of estimating equations approach to estimate the unknown quantities of interest. The first equation is an analog of the estimating equation of YLZ but more complicated because of the nonparametric part in the single-index part, while the second equation is constructed especially for additive hazard model with single-index based on a mean zero martingale, which is quite different from the existing methods for single-index regression model in the literature.

3 Asymptotic properties

Let H be a $(2p + 2)$-diagonal matrix, with the first p elements equal to 1, the second p elements equal to h_1 and the last two elements equal to h_1 and h_2, respectively. Let $\mu_j^{(1)} = \int u^j K_1(u)\, \mathrm{d}u$, $\nu_j^{(1)} = \int u^j K_1^2(u)\, \mathrm{d}u$, $\mu_j^{(2)} = \int u^j K_2(u)\, \mathrm{d}u$, $\nu_j^{(2)} = \int u^j K_2^2(u)\, \mathrm{d}u$, $P(t, \mathbf{Z}, \mathbf{X}^\top \gamma, V) = Pr(\tilde{T} \geq t | \mathbf{Z}, \mathbf{X}^\top \gamma, V)$ and $\rho(t, \mathbf{Z}, \mathbf{X}^\top \gamma, V) = P(t, \mathbf{Z}, \mathbf{X}^\top \gamma, V)\lambda(t | \mathbf{Z}, \mathbf{X}, V)$.

For $k = 0, 1, 2$, define

$$a_k(t, u, v) = E[P(t, \mathbf{Z}, u, v)\mathbf{Z}^{\otimes k} | \mathbf{X}^\top \gamma_0 = u, V = v] f_{X^\top \gamma_0, V}(u, v),$$

$$
\begin{aligned}
a_k^*(t,u,v) &= E[\rho(t,\mathbf{Z},u,v)\mathbf{Z}^{\otimes k}|\mathbf{X}^\top\gamma_0=u, V=v]f_{X^\top\gamma_0,V}(u,v),\\
\mathbf{G}_i^*(u,v) &= H^{-1}\mathbf{Z}_i^*(u,v),\\
\bar{\mathbf{G}}(t,u,v) &= \frac{\sum_{i=1}^n K_{1,h_1}(V_i-v)K_{2,h_2}(\mathbf{X}_i^\top\gamma_0-u)Y_i(t)\mathbf{G}_i^*(u,v)}{\sum_{i=1}^n K_{1,h_1}(V_i-v)K_{2,h_2}(\mathbf{X}_i^\top\gamma_0-u)Y_i(t)},
\end{aligned}
$$

where $f_{X^\top\gamma_0,V}(u,v)$ is the joint density function of $(\mathbf{X}^\top\gamma_0, V)$ evaluated at (u,v).

Theorem 1 *Under* (C1)–(C5) *given in Appendix, we have*

$$
\sqrt{n}(\hat{\gamma}-\gamma_0) \longrightarrow_d N\left(0, \mathbf{J}_{\gamma_0^{(1)}}\{D(\gamma_0^{(1)})\}^{-1}\Sigma(\gamma_0^{(1)})\{D(\gamma_0^{(1)})\}^{-1}\mathbf{J}_{\gamma_0^{(1)}}^\top\right),
$$

where γ_0 is the true value of γ under model (1)*, $\gamma_0^{(1)}$ the true value of $\gamma^{(1)}$,*

$$
\begin{aligned}
D(\gamma_0^{(1)}) &= \int_0^\tau E\left[\mathbf{J}_{\gamma_0^{(1)}}^\top\{\tilde{\dot{\omega}}(\mathbf{X}^\top\gamma_0)\mathbf{X}-\bar{\mathbf{X}}(t)\}Y(t)\tilde{\dot{\omega}}(\mathbf{X}^\top\gamma_0)\mathbf{X}\mathbf{J}_{\gamma_0^{(1)}}\right]\mathrm{d}t,\\
\Sigma(\gamma_0^{(1)}) &= \int_0^\tau E\Bigg\{\Big[\mathbf{J}_{\gamma_0^{(1)}}^\top\{\tilde{\dot{\omega}}(\mathbf{X}^\top\gamma_0)\mathbf{X}-\bar{\mathbf{X}}(t)\}\\
&\quad -Q\cdot\{H\cdot D(\mathbf{X}^\top\gamma_0,V)\}^{-1}\{\mathbf{G}^*(\mathbf{X}^\top\gamma_0,V)-\bar{\mathbf{G}}(t,\mathbf{X}^\top\gamma_0,V)\}\Big]^{\otimes 2}\\
&\quad \times Y(t)\lambda(t|\mathbf{Z},\mathbf{X},V)\Bigg\}\mathrm{d}t,
\end{aligned}
$$

with $D(u,v)$ defined in (7) *and*

$$
Q_i = \int_0^\tau E[\mathbf{J}_{\gamma_0^{(1)}}^\top\{\tilde{\dot{\omega}}(\mathbf{X}^\top\gamma)\mathbf{X}-\bar{\mathbf{X}}(s)\}Y(s)(\mathbf{Z}^*)^\top R(\gamma_0,\mathbf{X}_i^\top\gamma_0,V_i)]\,\mathrm{d}s.
$$

Theorem 2 *Under* (C1)–(C5) *given in Appendix, we have*

$$
\sqrt{nh_1h_2}\left\{\hat{\xi}(u,v)-\xi_0(u,v)-\tilde{b}(u,v)\right\} \longrightarrow_d N(0,\Sigma(u,v)),
$$

where $\xi_0(u,v)=\left(\beta_0(v)^\top,\dot{\beta}_0(v)^\top,\dot{g}_0(v),\dot{\omega}_0(u)\right)^\top$ is the true value of $\xi(u,v)$ under model (1)*,*

$$
\begin{aligned}
\Sigma(u,v) &= \int_0^\tau \{H\cdot D(u,v)\}^{-1}\mathrm{diag}\left(a_2^*(t,u,v), \nu_2^{(1)}\begin{pmatrix}a_2^*(t,u,v) & a_1^*(t,u,v)\\ a_1^*(t,u,v) & a_0^*(t,u,v)\end{pmatrix}, \nu_2^{(2)}a_0^*(t,u,v)\right)\\
&\quad \{H\cdot D(u,v)\}^{-1}\mathrm{d}t,\\
\tilde{b}(u,v) &= \{H\cdot D(u,v)\}^{-1}b(u,v),
\end{aligned}
$$

with $b(u,v)=\frac{1}{2}h_1^2\mu_2^{(1)}b_1(u,v)+\frac{1}{2}h_2^2\mu_2^{(2)}b_2(u,v)$ and $b_1(u,v)$, $b_2(u,v)$ defined in (8), (9)*. Note that, $\sqrt{nh_1h_2}\tilde{b}$ goes to zero as $n\to\infty$.*

To obtain consistent estimators for the asymptotic covariance matrices of γ and $\xi(u, v)$, we can replace $D(u, v)$, $Q(t)$, $D(\gamma_0^{(1)})$, $\Sigma(\gamma_0^{(1)})$ and Σ with their empirical counterparts $D_n(u, v)$, $Q_n(t)$, $D_n(\gamma_0^{(1)})$, $\Sigma_n(\gamma_0^{(1)})$ and Σ_n, respectively. The performance of the resulting covariance matrix estimators may be unstable for finite sample size due to the complicated expressions of the asymptotic covariance matrices. Alternatively, resampling method such as that of Jin et al. (2001) can be used to approximate the covariance matrices. The proofs are outlined in the Appendix.

4 Numerical studies

4.1 Simulations

We carried out two simulation studies to assess the finite sample properties of the proposed method. In the first simulation study, we examine the relative performance of the proposed method and the existing method due to Yin et al. (2008) when $\omega(\mathbf{X}^\top\gamma) = \mathbf{X}^\top\gamma$. In the second simulation study, we assess the finite-sample performance when $\omega(\mathbf{X}^\top\gamma)$ is nonlinear in $\mathbf{X}^\top\gamma$. The estimating method of Yin et al. (2008) is invalid in this situation because their model is not suited for this set-up.

The failure times were generated from the partially linear single-index additive hazard model

$$\lambda(t|\mathbf{Z}, \mathbf{X}, V) = \lambda_0(t) + \beta(V)^\top\mathbf{Z} + g(V) + \omega(\mathbf{X}^\top\gamma). \quad (5)$$

In the first simulation, we took $\lambda_0(t) = 0.5$, $\beta(v) = 0.5v(3 - v)$, $g(v) = 0.2\exp(v - 1.5)$ and $\omega(\mathbf{X}^\top\gamma) = \mathbf{X}^\top\gamma$, with $\mathbf{X} = (X_1, X_2, X_3)^\top$, $\gamma = (\gamma_1, \gamma_2, \gamma_3)^\top = (0.5, 0.5, 0.7071)^\top$. Covariate $\mathbf{Z}$ was generated from a uniform distribution Unif[0,1], covariate V was generated from Unif[0,3], X_1 from Unif[0,0.5], X_2 from Unif[0,0.5], X_3 from Unif[0,0.7071]. The censoring time was taken as the minimum value of τ and a random number generated from Unif[$\tau/2, 3\tau/2$]. We took $\tau = 0.86$ to yield an approximate censoring rate of 25 %. The Epanechnikov kernel function was used.

Table 1 Biases and standard deviations (SD) for $\hat{\gamma}$, the proposed estimators (WWZ) and the estimates of Yin et al. (2008) (YLZ)

Estimators	n		$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
WWZ	60	Bias	0.0157	−0.0075	−0.0060
		SD	0.2848	0.3200	0.2962
YLZ		Bias	0.2125	−0.0114	0.2186
		SD	2.3489	2.3197	1.6770
WWZ	120	Bias	0.0270	−0.0165	−0.0081
		SD	0.2547	0.3000	0.2721
YLZ		Bias	0.0010	−0.0941	0.0816
		SD	1.4933	1.4560	0.9042

Four hundreds simulations of sample size $n = 60$, 120 were, respectively, ran to calculate the biases, standard deviations (SD) of the proposed estimators. Yin et al. (2008) developed estimating approach for the special case of $\omega(\mathbf{X}^\top\gamma) = \mathbf{X}^\top\gamma$ as considered here. Hence, we calculated the corresponding biases and SDs of the estimators due to Yin et al. (2008) as a comparison.

The simulation results for the estimators of γ are presented in Table 1. In Figs. 1 and 2, we plot the curves of the estimates of $\beta(\cdot)$, $g(\cdot)$ and $\omega(\cdot)$ and SDs of the estimates of $\beta(\cdot)$, $g(\cdot)$ and $\omega(\cdot)$ at 48 equally spaced points on the support of V and 20 equally spaced points on the support of $\mathbf{X}^\top\gamma$, respectively. In these Tables and Figures, "WWZ" denotes the proposed estimator and "YLZ" the estimator due to Yin

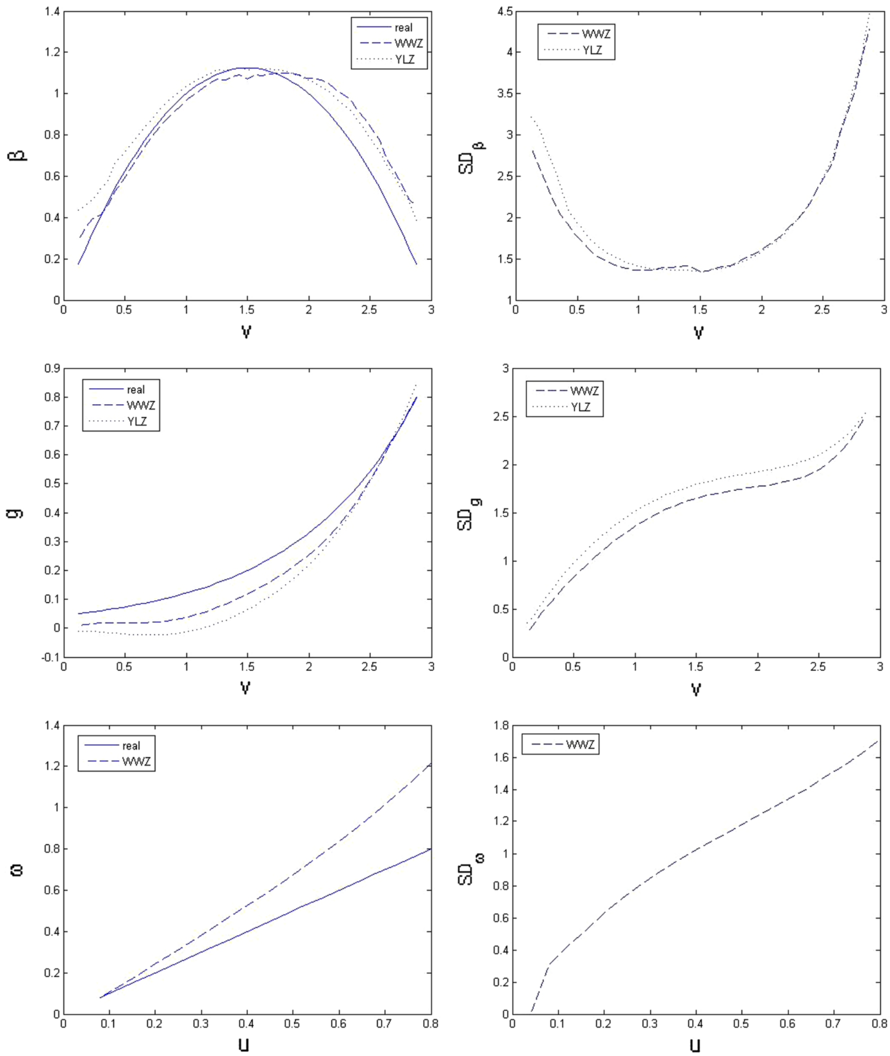


Fig. 1 Curves and standard deviation (SD) curves of $\hat{\beta}(v)$, $\hat{g}(v)$ and $\hat{\omega}(u)$ for $n = 60$

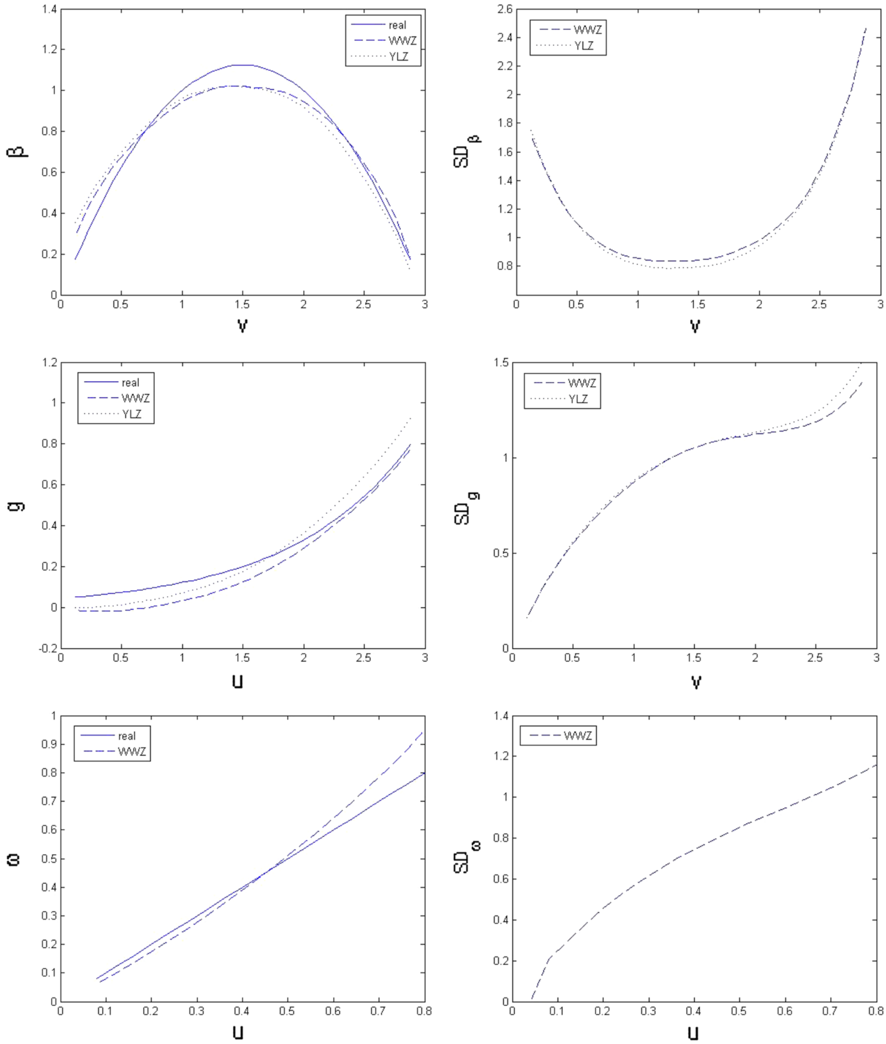


Fig. 2 Curves and standard deviation (SD) curves of $\hat{\beta}(v)$, $\hat{g}(v)$ and $\hat{\omega}(u)$ for $n = 120$

et al. (2008). We also compute the joint integrated mean squared errors (IMSE) (Li et al. 2007) for $\hat{\beta}(\cdot)$ and $\hat{g}(\cdot)$ based on 48 equally spaced points on the support of V in Table 3.

From Table 1, we can see that our estimators of γ are comparable to those of Yin et al. (2008) and perform better in terms of SD. A reason may be that they first obtain a local estimate of γ from an estimating equation and then use a weighted average of the local estimate of γ to obtain a global estimate of γ, while the proposed estimating function for γ is a global one. Figures 1, 2 and Table 3 indicate that the estimators of the two methods perform similarly especially when $n = 120$ as in that case the joint IMSEs for different methods are very close. Combining all

Table 2 Biases and standard deviations (SD) for $\hat{\gamma}$, the proposed estimators (WWZ) and the estimates of Yin et al. (2008) (YLZ)

Estimators	n		$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
WWZ	60	Bias	0.0386	0.0049	−0.0326
		SD	0.2849	0.3051	0.2966
YLZ	60	Bias	−0.0998	−0.3506	−0.2148
		SD	2.5952	2.4991	1.8109
WWZ	120	Bias	−0.0021	−0.0140	0.0112
		SD	0.2587	0.3030	0.2826
YLZ	120	Bias	−0.2952	−0.3909	−0.3026
		SD	1.5865	1.5667	1.0026

Table 3 Joint IMSEs for $\hat{\beta}(\cdot)$ and $\hat{g}(\cdot)$

Estimators	n	Simulation 1	Simulation 2
WWZ	60	17.9722	22.5195
YLZ		20.2125	23.3076
WWZ	120	6.6789	7.0908
YLZ		6.6960	7.5246

the results, we see that both the proposed estimates and those of YLZ are close to the true values and the estimated curves are close to the true curves as sample size increases.

In the second simulation, the failure times were generated from the partially linear single-index additive hazard model (5) with $\lambda_0(t) = 0.5$, $\beta(v) = 1.2 + 0.5v(3 - v)$, $g(v) = 0.2\exp(v - 1.5)$ and $\omega(\mathbf{X}^\top\gamma) = 0.5(\mathbf{X}^\top\gamma)^2$, with $\mathbf{X} = (X_1, X_2, X_3)^\top$, $\gamma = (\gamma_1, \gamma_2, \gamma_3)^\top = (0.5, 0.5, 0.7071)^\top$. Covariate $\mathbf{Z}$ was generated from a uniform distribution Unif[0,1], covariate V was generated from Unif[0,3], X_1 from Unif[0,0.5], X_2 from Unif[0,0.5], X_3 from Unif[0,0.7071]. The censoring time was taken as the minimum value of τ and a random number generated from Unif$[\tau/2, 3\tau/2]$. We took $\tau = 0.86$ to yield an approximate censoring rate of 25 %. Here, we also used the Epanechnikov kernel function. We run four hundreds simulations of sample sizes $n = 60$ and 120 to calculate the biases and SDs of the proposed estimators. In this simulation example, $\omega(\cdot)$ is nonlinear. We also calculated the estimators by the method of Yin et al. (2008) to see its performance.

The simulation results of the estimators of γ are presented in Table 2. Figures 3 and 4 plot the curves of the estimates of $\beta(\cdot)$, $g(\cdot)$ and $\omega(\cdot)$ and the SDs of the estimates of $\beta(\cdot)$, $g(\cdot)$ and $\omega(\cdot)$ at 48 equally spaced points on the support of V and 20 equally spaced points on the support of $\mathbf{X}^\top\gamma$, respectively. Joint IMSEs for $\hat{\beta}(\cdot)$ and $\hat{g}(\cdot)$ at the 48 equally spaced points on the support of V are also calculated in Table 3.

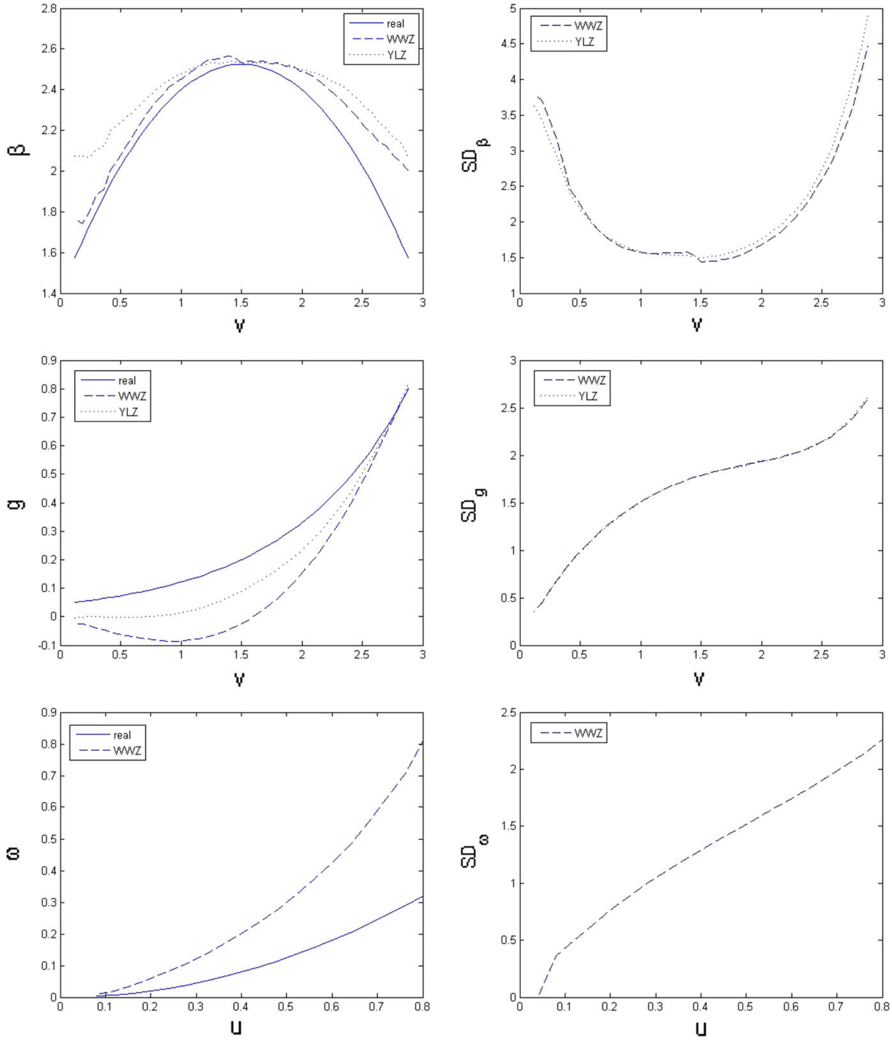


Fig. 3 Curves and standard deviation (SD) curves of $\hat{\beta}(v)$, $\hat{g}(v)$ and $\hat{\omega}(u)$ for $n = 60$

From Tables 2, 3 and Figs. 3, 4, we can see that our estimates of γ are close to the true values and the estimated curves are close to the true ones, and hence the proposed method performs well. However, the YLZ estimates of γ are seriously biased in this set-up. The biases of the YLZ estimates of γ are at least 10 times larger than the proposed ones, and even 100 times larger in some cases. This suggests that YLZ method does not define consistent estimator for γ for the case where $\omega(\cdot)$ is not a linear function. In Table 3, the joint IMSEs of the proposed estimates are smaller than those of YLZ, which indicate that the proposed estimates of $\beta(\cdot)$ and $g(\cdot)$ are closer to the true functions. On the whole, the proposed method outperforms that of YLZ in this set-up.

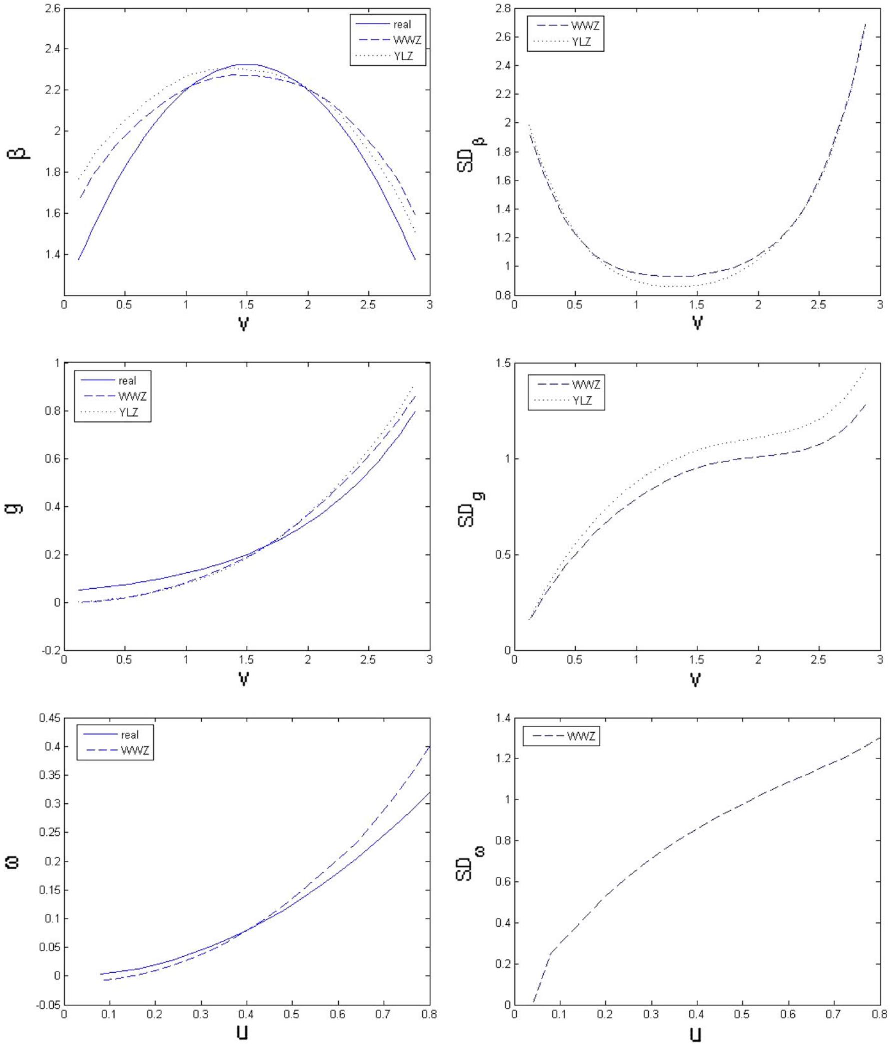


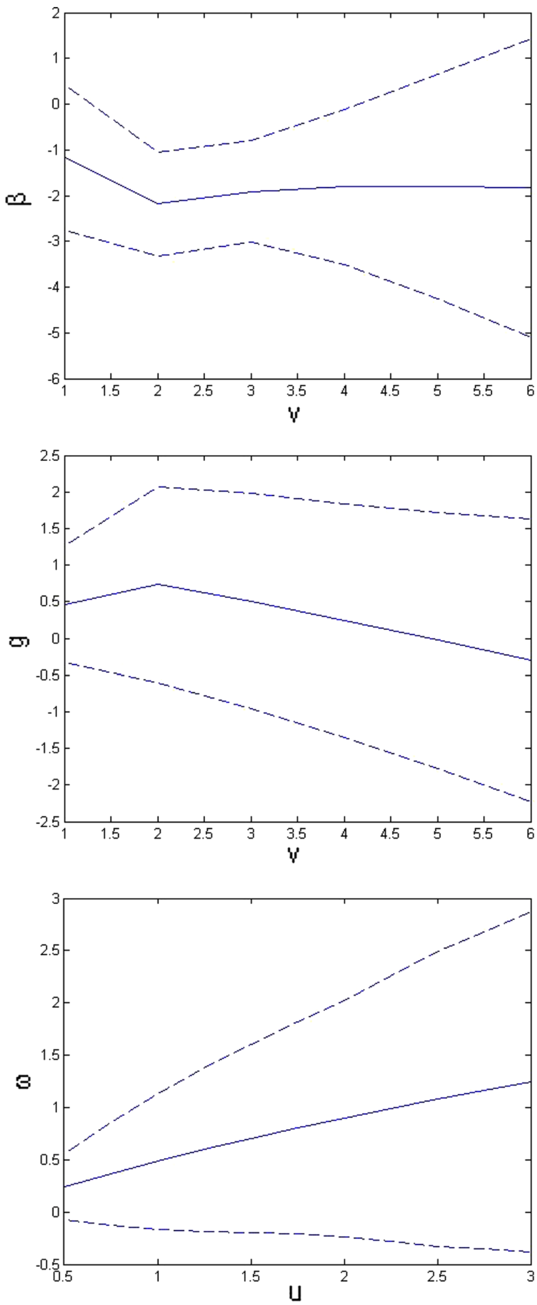
Fig. 4 Curves and standard deviation (SD) curves of $\hat{\beta}(v)$, $\hat{g}(v)$ and $\hat{\omega}(u)$ for $n = 120$

4.2 Data analysis

The proposed approaches are now applied to the The National Alzheimer's Coordinating Center(NACC) and Uniform Data Set(UDS). There are 62,617 observations in all. The study was initiated in May 2005, and patients are followed annually since then.

The interest here is to characterize the relationship between survival and known risk factors and determine how the genotype APOE interacts with other covariates, such as Global CDR(CDRGLOB), Parkinsonism disorder status(PDOTHR), Incontinence-bowel status(INCONTF) and Hypertension status(HYPERTEN). After eliminating

Fig. 5 The estimated coefficient functions for the NACC UDS (real line, estimated functions; *dashed line*, 95 % confidence limit)



observations with FORMVER=1 or PACKET=F (or T) or missing covarite values, we got approximate 100 observations. We first fitted the fully nonparametric model to the data, and found that the estimated parameters for CDRGLOB, PDOTHR, INCONTF, and HYPERTEN did not change much with respect to APOE, so we considered the following model:

$$\lambda(t|Z, \mathbf{X}, V) = \lambda_0(t) + \beta(V)Z + g(V) + \omega(\mathbf{X}^\top\gamma),$$

where V denotes the value of APOE, Z denotes NORMCOG value, $\mathbf{X} = (X_1, X_2, X_3, X_4)$ and X_1, X_2, X_3, X_4 denote the values of CDRGLOB, PDOTHR, INCONTF, and HYPERTEN, respectively. Gaussian kernel functions were used here. The constant coefficients were estimated as $\hat{\gamma}_1 = 0.6530$, $\hat{\gamma}_2 = 0.4474$, $\hat{\gamma}_3 = 0.4444$, and $\hat{\gamma}_4 = 0.4194$. The estimated functions of $\beta(V)$, $g(V)$ and $\omega(\mathbf{X}^\top\gamma)$ are plotted in Fig. 5, and the estimates of the pointwise confidence limits are based on the resampling method of Jin et al. (2001).

From Fig. 5, the negative estimates of β imply a higher NORCOG value associated with a decreased hazard and the positive estimates of ω and γ imply higher CDRGLOB, PDOTHR, INCONTF, and HYPERTEN values associated with an increased hazard, which is reasonable. The estimates of g imply that different APOE values have different influence on the hazard or the survival, some being positive while some being negative.

5 Discussion

We study a class of partially varying coefficient single-index additive hazards model. The coefficients may vary with an exposure variable, thus leading to a covariate varying structure. Also, we incorporate high-dimension explanatory covariates into a single index. The model is very general and includes commonly used additive hazard models as special examples, such as the model of Yin et al. (2008). An inference procedure is proposed by solving a unified local and global estimating equation sets based on local polynomial technique and the martingale representation. It is interesting to consider the further extensions when the coefficients vary with many exposure variables. A single index may be used to combine all the exposure variables through a linear function. A thorough investigation is needed for future research.

6 Appendix

6.1 Proofs of the main results

(C1): The kernel functions $K_1(\cdot) > 0$ and $K_2(\cdot) > 0$ are bounded and symmetric densities with compact bounded supports.
(C2): The covariates $\mathbf{Z}$, $\mathbf{X}$ and V are bounded in $[0, \tau]$.
(C3): The conditional density of $\mathbf{Z}$ given $(\mathbf{X}^\top\gamma = u, V = v)$ is twice continuously differentiable with respect to u and v. The marginal density of $(\mathbf{X}^\top\gamma = u, V = v)$ evaluated at (u, v) is twice continuously differentiable with respect to (u, v) and satisfies $inf_{(u,v)\in(\mathcal{W}_1,\mathcal{W}_2)} f_{X^\top\gamma,V}(u, v) > 0$.

(C4): $inf_{t\in[0,\tau],(u,v)\in(\mathcal{W}_1,\mathcal{W}_2)}a_0(t,u,v)>0$, the matrices $D(u,v)$ and $D(\gamma_0^{(1)})$ are nonsingular, $\Sigma(\gamma_0^{(1)})$ and $\Sigma(u,v)$ are positive definite for all $(u,v)\in(\mathcal{W}_1,\mathcal{W}_2)$.
(C5): $h_1\to 0$, $h_2\to 0$, $logh_1/(n^{1/2}h_1h_2)\to 0$, nh_1^5 and nh_2^5 are bounded.

All of these conditions are standard in local linear estimation. Use the definitions in Sect. 3, as well as for k=0,1,2, define

$$
\begin{aligned}
S_{nk}(t,u,v)&=\frac{1}{n}\sum_{i=1}^{n}K_{1,h_1}(V_i-v)K_{2,h_2}(\mathbf{X}_i^\top\gamma_0-u)Y_i(t)\left(G_i^*(u,v)\right)^{\otimes k},\\
s_0(t,u,v)&=a_0(t,u,v),\\
s_1(t,u,v)&=\left(a_1{}^\top(t,u,v),0_{p+2}^\top\right)^\top,\\
s_2(t,u,v)&=\operatorname{diag}\left(a_2(t,u,v),\mu_2^{(1)}\begin{pmatrix}a_2(t,u,v)&a_1(t,u,v)\\a_1^\top(t,u,v)&a_0(t,u,v)\end{pmatrix},\mu_2^{(2)}a_0(t,u,v)\right).
\end{aligned}
$$

Lemma 1 *Under the conditions given in Appendix, we have, for k=0,1,2,*

$$
\begin{aligned}
&sup_{t\in[0,\tau],(u,v)\in(\mathcal{W}_1,\mathcal{W}_2)}|S_{nk}(t,u,v)-s_k(t,u,v)|\\
&=O_p\left(\frac{\log h_1}{\sqrt{n}h_1h_2}\right)+O(h_1^2)+O(h_2^2).
\end{aligned}
$$

Lemma 1 can be proved using similar arguments to that of Lemma A.1 in Lin et al. (2008).

Proof of Theorem 1 We prove the theorem in the following four steps:
STEP 1. Derivation of the expression for $\tilde{\xi}(\gamma_0,u,v)-\xi_0(u,v)$.
Note that

$$
\begin{aligned}
0-U_n(\xi_0(u,v),\gamma_0;u,v)&=U_n(\tilde{\xi}(\gamma_0,u,v),\gamma_0;u,v)-U_n(\xi_0(u,v),\gamma_0;u,v)\\
&=\left\{\partial U_n(\xi,\gamma_0;u,v)/\partial\xi|_{\xi=\xi_0}+o_p(1)\right\}\\
&\quad\left\{\tilde{\xi}(\gamma_0,u,v)-\xi_0(u,v)\right\}.
\end{aligned}
$$

To obtain the expression of $\tilde{\xi}(\gamma_0,u,v)-\xi_0(u,v)$, we obtain the expressions for $\partial U_n(\xi(u,v),\gamma_0;u,v)/\partial\xi(u,v)|_{\xi(u,v)=\xi_0(u,v)}$ and $U_n(\xi_0(u,v),\gamma_0;u,v)$ separately, and then the asymptotic expression for $\tilde{\xi}(\gamma_0,u,v)-\xi_0(u,v)$.

First, we derive the expressions for $\partial U_n(\xi(u,v),\gamma_0;u,v)/\partial\xi(u,v)|_{\xi(u,v)=\xi_0(u,v)}$. Recalling the definition of $U_n(\xi(u,v),\gamma_0;u,v)$ in (2), we have the following result:

$$
\begin{aligned}
&\partial U_n(\xi(u,v),\gamma_0;u,v)/\partial\xi(u,v)|_{\xi(u,v)=\xi_0(u,v)}\\
&= -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_{1,h_1}(V_i-v)K_{2,h_2}(\mathbf{X}_i^{\top}\gamma_0-u)\{\mathbf{Z}_i^*(u,v)-\bar{\mathbf{Z}}(t,u,v)\}^{\otimes 2}Y_i(t)\,\mathrm{d}t\\
&= -H^{\otimes 2}\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_{1,h_1}(V_i-v)K_{2,h_2}(\mathbf{X}_i^{\top}\gamma_0-u)\{\mathbf{G}_i^*(u,v)-\bar{\mathbf{G}}(t,u,v)\}^{\otimes 2}Y_i(t)\,\mathrm{d}t\\
&:= -H^{\otimes 2}D_n(u,v)\\
&= -H^{\otimes 2}D(u,v)+o_p(1), \qquad (6)
\end{aligned}
$$

where H and $\mathbf{G}_i^*(u,v)$ are defined in Sect. 3,

$$
D_n(u,v) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_{1,h_1}(V_i-v)K_{2,h_2}(\mathbf{X}_i^{\top}\gamma_0-u)\{\mathbf{G}_i^*(u,v)-\bar{\mathbf{G}}(t,u,v)\}^{\otimes 2}Y_i(t)\mathrm{d}t,
$$

$$
D(u,v) = \begin{pmatrix} D_{11} & 0 & 0 & 0\\ 0 & D_{22} & D_{23} & 0\\ 0 & D_{32} & D_{33} & 0\\ 0 & 0 & 0 & D_{44}\end{pmatrix}, \qquad (7)
$$

with

$$
\begin{aligned}
D_{11} &= \int_0^{\tau}\left\{a_2(t,u,v)-\frac{a_1(t,u,v)a_1(t,u,v)^{\top}}{a_0(t,u,v)}\right\}\mathrm{d}t,\\
D_{22} &= \mu_2^{(1)}\int_0^{\tau} a_2(t,u,v)\,\mathrm{d}t,\\
D_{23} &= \mu_2^{(1)}\int_0^{\tau} a_1(t,u,v)\,\mathrm{d}t,\\
D_{32} &= \mu_2^{(1)}\int_0^{\tau} a_1(t,u,v)\,\mathrm{d}t,\\
D_{33} &= \mu_2^{(1)}\int_0^{\tau} a_0(t,u,v)\,\mathrm{d}t,\\
D_{44} &= \mu_2^{(2)}\int_0^{\tau} a_0(t,u,v)\,\mathrm{d}t,
\end{aligned}
$$

and $a_k(t,u,v)$, $k=0,1,2$ defined in Sect. 3.

Second, we derive the expression for $U_n(\xi_0(u,v),\gamma_0;u,v)$.

By definition,

$$
\begin{aligned}
&H^{-1}U_n(\xi_0(u,v),\gamma_0;u,v)\\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_{1,h_1}(V_i-v)K_{2,h_2}(\mathbf{X}_i^{\top}\gamma_0-u)\{\mathbf{G}_i^*(u,v)-\bar{\mathbf{G}}(t,u,v)\}\\
&\quad\times\left[\mathrm{d}N_i(t)-Y_i(t)\{\lambda_0^*(t,u,v)+\xi_0^{\top}(u,v)\mathbf{Z}_i^*(u,v)\}\,\mathrm{d}t\right]
\end{aligned}
$$

$$
\begin{aligned}
= & \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_{1,h_1}(V_i-v)K_{2,h_2}(\mathbf{X}_i^{\top}\gamma_0-u)\{\mathbf{G}_i^*(u,v)-\bar{\mathbf{G}}(t,u,v)\}\,[\mathrm{d}M_i(t) \\
& +Y_i(t)\{\lambda_0(t)+\beta(V_i)^{\top}\mathbf{Z}_i+g(V_i)+\omega(\mathbf{X}_i^{\top}\gamma_0)\}\,\mathrm{d}t \\
& -Y_i(t)\{\lambda_0^*(t,u,v)+\xi_0^{\top}(u,v)\mathbf{Z}_i^*(u,v)\}\,\mathrm{d}t] \\
= & \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_{1,h_1}(V_i-v)K_{2,h_2}(\mathbf{X}_i^{\top}\gamma_0-u)\{\mathbf{G}_i^*(u,v)-\bar{\mathbf{G}}(t,u,v)\}\mathrm{d}M_i(t) \\
& +\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_{1,h_1}(V_i-v)K_{2,h_2}(\mathbf{X}_i^{\top}\gamma_0-u)Y_i(t)\{\mathbf{G}_i^*(u,v)-\bar{\mathbf{G}}(t,u,v)\} \\
& \times\left[\{\frac{1}{2}\ddot{\beta}(v)^{\top}\mathbf{Z}_i+\frac{1}{2}\ddot{g}(v)\}(V_i-v)^2+\frac{1}{2}\ddot{\omega}(u)(\mathbf{X}_i^{\top}\gamma_0-u)^2\right]\mathrm{d}t \\
& +\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_{1,h_1}(V_i-v)K_{2,h_2}(\mathbf{X}_i^{\top}\gamma_0-u)Y_i(t)\{\mathbf{G}_i^*(u,v)-\bar{\mathbf{G}}(t,u,v)\} \\
& \times\sum_{p=3}^{\infty}\left[\{\frac{1}{p!}\beta^{(p)}(v)^{\top}\mathbf{Z}_i+\frac{1}{p!}g^{(p)}(v)\}(V_i-v)^p+\frac{1}{p!}\omega^{(p)}(u)(\mathbf{X}_i^{\top}\gamma_0-u)^p\right]\mathrm{d}t \\
:= & A_n(\tau,u,v)+B_n(\tau,u,v)+C_n(\tau,u,v).
\end{aligned}
$$

For $B_n(\tau, u, v)$, by Lemma 1, we obtain the following result:

$$
\begin{aligned}
B_n(\tau,u,v) = & \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_{1,h_1}(V_i-v)K_{2,h_2}(\mathbf{X}_i^{\top}\gamma_0-u)Y_i(t)\{G_i^*(u,v)-\bar{G}(t,u,v)\} \\
& \times\left[\left\{\frac{1}{2}\ddot{\beta}(v)^{\top}\mathbf{Z}_i+\frac{1}{2}\ddot{g}(v)\right\}(V_i-v)^2+\frac{1}{2}\ddot{\omega}(u)(\mathbf{X}_i^{\top}\gamma_0-u)^2\right]\mathrm{d}t \\
= & h_1^2\Bigg[\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_{1,h_1}(V_i-v)K_{2,h_2}(\mathbf{X}_i^{\top}\gamma_0-u)Y_i(t) \\
& \left\{G_i^*(u,v)-\frac{s_1(t,u,v)}{s_0(t,u,v)}\right\}\times\left\{\frac{1}{2}\ddot{\beta}(v)^{\top}\mathbf{Z}_i+\frac{1}{2}\ddot{g}(v)\right\}\left(\frac{V_i-v}{h_1}\right)^2\mathrm{d}t \\
& +O_p\left(\frac{\log h_1}{\sqrt{n}h_1h_2}\right)+O_p(h_1^2)+O_p(h_2^2)\Bigg]+h_2^2\Bigg[\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_{1,h_1}(V_i-v) \\
& K_{2,h_2}(\mathbf{X}_i^{\top}\gamma_0-u)\left\{G_i^*(u,v)-\frac{s_1(t,u,v)}{s_0(t,u,v)}\right\}\times\frac{1}{2}\ddot{\omega}(u)\left(\frac{\mathbf{X}_i^{\top}\gamma_0-u}{h_2}\right)^2\mathrm{d}t \\
& +O_p\left(\frac{\log h_1}{\sqrt{n}h_1h_2}\right)+O_p(h_1^2)+O_p(h_2^2)\Bigg] \\
= & \frac{1}{2}h_1^2\mu_2^{(1)}b_1(u,v)+\frac{1}{2}h_2^2\mu_2^{(2)}b_2(u,v)+o_p(h_1^2)+o_p(h_2^2) \\
:= & b(u,v)+o_p(h_1^2)+o_p(h_2^2),
\end{aligned}
$$

where $\bar{G}(t,u,v)$, $\mu_2^{(1)}$, $\mu_2^{(2)}$ are defined in Sect. 3,

$$b_1(u,v)=\begin{pmatrix}\int_0^\tau \left\{a_2(t,u,v)-\frac{a_1(t,u,v)a_1(t,u,v)^\top}{a_0(t,u,v)}\right\}\mathrm{d}t\,\ddot{\beta}(v)\\0\\0\\0\end{pmatrix}_{(2p+2)\times 1}+\begin{pmatrix}\int_0^\tau \left\{a_1(t,u,v)-\frac{a_1(t,u,v)}{a_0(t,u,v)}\right\}\mathrm{d}t\,\ddot{g}(v)\\0\\0\\0\end{pmatrix}_{(2p+2)\times 1}, \quad (8)$$

$$b_2(u,v)=\begin{pmatrix}\int_0^\tau \left\{a_1(t,u,v)-\frac{a_1(t,u,v)}{a_0(t,u,v)}\right\}\mathrm{d}t\,\ddot{\omega}(u)\\0\\0\\0\end{pmatrix}_{(2p+2)\times 1}, \quad (9)$$

with $a_k(t,u,v)$ being defined in Sect. 3.

Therefore,

$$\begin{aligned}H^{-1}U_n(\xi_0(u,v),\gamma_0;u,v)&=A_n(\tau,u,v)+B_n(\tau,u,v)+C_n(\tau,u,v)\\&=A_n(\tau,u,v)+b(u,v)+o_p(h_1^2)+o_p(h_2^2).\end{aligned} \quad (10)$$

Finally, we can get the asymptotic expression for $\tilde{\xi}(\gamma_0,u,v)-\xi_0(u,v)$. Recall that

$$\begin{aligned}0-U_n(\xi_0(u,v),\gamma_0;u,v)&=U_n(\tilde{\xi}(\gamma_0,u,v),\gamma_0;u,v)-U_n(\xi_0(u,v),\gamma_0;u,v)\\&=\left\{\partial U_n(\xi,\gamma_0;u,v)/\partial\xi|_{\xi=\xi_0}+o_p(1)\right\}\\&\quad\left\{\tilde{\xi}(\gamma_0,u,v)-\xi_0(u,v)\right\}.\end{aligned}$$

From (6) and (10), we have the following result:

$$\begin{aligned}\tilde{\xi}(\gamma_0,u,v)-\xi_0(u,v)&=\{H\cdot D(u,v)\}^{-1}\{A_n(\tau,u,v)+B_n(\tau,u,v)+C_n(\tau,u,v)\}\\&=\{H\cdot D(u,v)\}^{-1}\\&\quad\times\left[\frac{1}{n}\sum_{i=1}^n\int_0^\tau K_{1,h_1}(V_i-v)K_{2,h_2}(\mathbf{X}_i^\top\gamma_0-u)\{\mathbf{G}_i^*(u,v)\right.\\&\quad\left.-\bar{\mathbf{G}}(t,u,v)\}\mathrm{d}M_i(t)+b(u,v)+o_p(h_1^2)+o_p(h_2^2)\right].\end{aligned} \quad (11)$$

From this formula, it can be seen that

$$\sqrt{nh_1h_2}\left\{\tilde{\xi}(\gamma_0, u, v) - \xi_0(u, v)\right\} = O_p(1). \tag{12}$$

Note that
$0 - \tilde{U}_n(\gamma_0^{(1)}) = \tilde{U}_n(\hat{\gamma}^{(1)}) - \tilde{U}_n(\gamma_0^{(1)}) = \left\{\partial\tilde{U}_n(\gamma^{(1)})/\partial\gamma^{(1)}|_{\gamma^{(1)}=\gamma_0^{(1)}} + o_p(1)\right\}(\hat{\gamma}^{(1)} - \gamma_0^{(1)})$.

To prove the asymptotic normality of $\sqrt{n}(\hat{\gamma^{(1)}} - \gamma_0^{(1)})$, it is sufficient to prove $\partial\tilde{U}_n(\gamma^{(1)})/\partial\gamma^{(1)}|_{\gamma^{(1)}=\gamma_0^{(1)}}$ convergence in probability and $\sqrt{n}\tilde{U}_n(\gamma_0^{(1)})$ convergence to a normal random variable in distribution. We first obtain the asymptotic expressions of $\partial\tilde{U}_n(\gamma^{(1)})/\partial\gamma^{(1)}|_{\gamma^{(1)}=\gamma_0^{(1)}}$ and $\tilde{U}_n(\gamma_0^{(1)})$ in Step 2 and Step 3, so as to obtain the asymptotic property of $\hat{\gamma}^{(1)}$ and hence the asymptotic property of $\hat{\gamma}$ in Step 4.

STEP 2. Derivation of the expression for $\partial\tilde{U}_n(\gamma^{(1)})/\partial\gamma^{(1)}|_{\gamma^{(1)}=\gamma_0^{(1)}}$.

Recalling the definition of $\tilde{U}_n(\gamma^{(1)})$ in (4), we have

$$\begin{aligned}
&\partial\tilde{U}_n(\gamma^{(1)})/\partial\gamma^{(1)}|_{\gamma^{(1)}=\gamma_0^{(1)}}\\
&\quad= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\partial\left(\mathbf{J}_{\gamma_0^{(1)}}^{\top}\{\tilde{\omega}(\mathbf{X}_i^{\top}\gamma_0)\mathbf{X}_i - \bar{\mathbf{X}}(t)\}\right)\Big/\partial\gamma^{(1)}|_{\gamma^{(1)}=\gamma_0^{(1)}}Y_i(t)\\
&\qquad\times\left[\mathrm{d}N_i(t) - Y_i(t)\left\{\lambda_0(t) + \tilde{\beta}(V_i)^{\top}\mathbf{Z}_i + \tilde{g}(V_i) + \tilde{\omega}(\mathbf{X}_i^{\top}\gamma_0)\right\}\mathrm{d}t\right]\\
&\qquad-\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{J}_{\gamma_0^{(1)}}^{\top}\{\tilde{\dot{\omega}}(\mathbf{X}_i^{\top}\gamma_0)\mathbf{X}_i - \bar{\mathbf{X}}(t)\}Y_i(t)\dot{\tilde{\omega}}(\mathbf{X}_i^{\top}\gamma_0)\mathbf{X}_i^{\top}\mathbf{J}_{\gamma_0^{(1)}}\,\mathrm{d}t\\
by(12)&\quad= -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{J}_{\gamma_0^{(1)}}^{\top}\{\tilde{\dot{\omega}}(\mathbf{X}_i^{\top}\gamma_0)\mathbf{X}_i - \bar{\mathbf{X}}(t)\}Y_i(t)\tilde{\dot{\omega}}(\mathbf{X}_i^{\top}\gamma_0)\mathbf{X}_i^{\top}\mathbf{J}_{\gamma_0^{(1)}}\,\mathrm{d}t + o_p(1)\\
&\quad= -\int_0^{\tau}E\left[\mathbf{J}_{\gamma_0^{(1)}}^{\top}\{\tilde{\dot{\omega}}(\mathbf{X}^{\top}\gamma_0)\mathbf{X} - \bar{\mathbf{X}}(t)\}Y(t)\tilde{\dot{\omega}}(\mathbf{X}^{\top}\gamma_0)\mathbf{X}^{\top}\mathbf{J}_{\gamma_0^{(1)}}\right]\mathrm{d}t + o_p(1)\\
&\quad:= -D(\gamma_0^{(1)}) + o_p(1).
\end{aligned} \tag{13}$$

STEP 3. Derivation of the asymptotic expression for $\tilde{U}_n(\gamma_0^{(1)})$.
By (11) and the definitions in Sect. 3, we have

$$\begin{aligned}
\tilde{\zeta}_i(\gamma_0) - \zeta_i(\gamma_0) &= \int_{\mathcal{W}_1}\int_{\mathcal{W}_2}R_i(\gamma_0, u, v)\{\tilde{\xi}(\gamma_0, u, v) - \xi_0(u, v)\}\,\mathrm{d}u\,\mathrm{d}v\\
&= \int_{\mathcal{W}_1}\int_{\mathcal{W}_2}R_i(\gamma_0, u, v)\{H\cdot D(u, v)\}^{-1}\{A_n(\tau, u, v)\\
&\qquad+B_n(\tau, u, v) + C_n(\tau, u, v)\}\,\mathrm{d}u\,\mathrm{d}v\\
&= \int_{\mathcal{W}_1}\int_{\mathcal{W}_2}R_i(\gamma_0, u, v)\{H\cdot D(u, v)\}^{-1}
\end{aligned}$$

$$\times\Bigg[\frac{1}{n}\sum_{j=1}^{n}\int_0^{\tau} K_{1,h_1}(V_j-v)K_{2,h_2}(\mathbf{X}_j^{\top}\gamma_0-u)$$
$$\{\mathbf{G}_j^*(u,v)-\bar{\mathbf{G}}(s,u,v)\}\mathrm{d}M_j(s)$$
$$+B_n(\tau,u,v)+C_n(\tau,u,v)\Bigg]\mathrm{d}u\,\mathrm{d}v$$
$$=\frac{1}{n}\sum_{j=1}^{n}\int_0^{\tau}\int_{\mathcal{W}_1}\int_{\mathcal{W}_2} R_i(\gamma_0,u,v)\{H\cdot D(u,v)\}^{-1}K_{1,h_1}(V_j-v)$$
$$K_{2,h_2}(\mathbf{X}_j^{\top}\gamma_0-u)\times\{\mathbf{G}_j^*(u,v)-\bar{\mathbf{G}}(s,u,v)\}\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}M_j(s)$$
$$=\frac{1}{n}\sum_{j=1}^{n}\int_0^{\tau} R_i(\gamma_0,\mathbf{X}_j^{\top}\gamma_0,V_j)\{H\cdot D(\mathbf{X}_j^{\top}\gamma_0,V_j)\}^{-1}\{\mathbf{G}_j^*(\mathbf{X}_j^{\top}\gamma_0,V_j)$$
$$-\ \bar{\mathbf{G}}(s,\mathbf{X}_j^{\top}\gamma_0,V_j)\}\,\mathrm{d}M_j(s).$$

This gives

$$\tilde{U}_n(\gamma_0^{(1)})=\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{J}_{\gamma_0^{(1)}}^{\top}\{\tilde{\dot{\omega}}(\mathbf{X}_i^{\top}\gamma)\mathbf{X}_i-\bar{\mathbf{X}}(t)\}[\mathrm{d}N_i(t)-Y_i(t)\{\lambda_0(t)+\tilde{\zeta}_i^{\top}(\gamma_0)\mathbf{Z}_i^*\}\mathrm{d}t]$$
$$=\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{J}_{\gamma_0^{(1)}}^{\top}\{\tilde{\dot{\omega}}(\mathbf{X}_i^{\top}\gamma)\mathbf{X}_i-\bar{\mathbf{X}}(t)\}[\mathrm{d}M_i(t)$$
$$+Y_i(t)\{\lambda_0(t)+\beta(V_i)^{\top}\mathbf{Z}_i+g(V_i)+\omega(\mathbf{X}_i^{\top}\gamma_0)\}$$
$$-Y_i(t)\{\lambda_0(t)+\tilde{\zeta}_i^{\top}(\gamma_0)\mathbf{Z}_i^*\}\mathrm{d}t]$$
$$=\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{J}_{\gamma_0^{(1)}}^{\top}\{\tilde{\dot{\omega}}(\mathbf{X}_i^{\top}\gamma)\mathbf{X}_i-\bar{\mathbf{X}}(t)\}[\mathrm{d}M_i(t)$$
$$-Y_i(t)\{\tilde{\zeta}_i^{\top}(\gamma_0)-\zeta_i^{\top}(\gamma_0)\}\mathbf{Z}_i^*\,\mathrm{d}t]$$
$$=\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{J}_{\gamma_0^{(1)}}^{\top}\{\tilde{\dot{\omega}}(\mathbf{X}_i^{\top}\gamma)\mathbf{X}_i-\bar{\mathbf{X}}(t)\}\,\mathrm{d}M_i(t)$$
$$-\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{J}_{\gamma_0^{(1)}}^{\top}\{\tilde{\dot{\omega}}(\mathbf{X}_i^{\top}\gamma)\mathbf{X}_i-\bar{\mathbf{X}}(t)\}Y_i(t)(\mathbf{Z}_i^*)^{\top}$$
$$\times\Bigg[\frac{1}{n}\sum_{j=1}^{n}\int_0^{\tau}R_i(\gamma_0,\mathbf{X}_j^{\top}\gamma_0,V_j)\{H\cdot D(\mathbf{X}_j^{\top}\gamma_0,V_j)\}^{-1}$$
$$\times\{\mathbf{G}_j^*(\mathbf{X}_j^{\top}\gamma_0,V_j)-\bar{\mathbf{G}}(s,\mathbf{X}_j^{\top}\gamma_0,V_j)\}\,\mathrm{d}M_j(s)\Bigg]\mathrm{d}t$$
$$=\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{J}_{\gamma_0^{(1)}}^{\top}\{\tilde{\dot{\omega}}(\mathbf{X}_i^{\top}\gamma)\mathbf{X}_i-\bar{\mathbf{X}}(t)\}\,\mathrm{d}M_i(t)$$
$$-\frac{1}{n}\sum_{j=1}^{n}\int_0^{\tau}\left[\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{J}_{\gamma_0^{(1)}}^{\top}\{\tilde{\dot{\omega}}(\mathbf{X}_i^{\top}\gamma)\mathbf{X}_i-\bar{\mathbf{X}}(t)\}Y_i(t)(\mathbf{Z}_i^*)^{\top}R_i(\gamma_0,\mathbf{X}_j^{\top}\gamma_0,V_j)\,\mathrm{d}t\right]$$

$$
\begin{aligned}
&\times\{H \cdot D(\mathbf{X}_j^\top \gamma_0, V_j)\}^{-1}\{\mathbf{G}_j^*(\mathbf{X}_j^\top \gamma_0, V_j) - \bar{\mathbf{G}}(s, \mathbf{X}_j^\top \gamma_0, V_j)\}\, \mathrm{d}M_j(s) \\
= &\frac{1}{n}\sum_{i=1}^n \int_0^\tau \mathbf{J}_{\gamma_0^{(1)}}^\top \{\tilde{\dot{\omega}}(\mathbf{X}_i^\top \gamma)\mathbf{X}_i - \bar{\mathbf{X}}(t)\}\, \mathrm{d}M_i(t) \\
&-\frac{1}{n}\sum_{i=1}^n \int_0^\tau \left[\frac{1}{n}\sum_{j=1}^n \int_0^\tau \mathbf{J}_{\gamma_0^{(1)}}^\top \{\tilde{\dot{\omega}}(\mathbf{X}_j^\top \gamma)\mathbf{X}_j - \bar{\mathbf{X}}(s)\} Y_j(s)(\mathbf{Z}_j^*)^\top R_j(\gamma_0, \mathbf{X}_i^\top \gamma_0, V_i)\mathrm{d}s\right] \\
&\times\{H \cdot D(\mathbf{X}_i^\top \gamma_0, V_i)\}^{-1}\{\mathbf{G}_i^*(\mathbf{X}_i^\top \gamma_0, V_i) - \bar{\mathbf{G}}(t, \mathbf{X}_i^\top \gamma_0, V_i)\}\, \mathrm{d}M_i(t).
\end{aligned}
$$

Then,

$$
\begin{aligned}
&\langle \sqrt{n}\tilde{U}_n(\gamma_0^{(1)}), \sqrt{n}\tilde{U}_n(\gamma_0^{(1)})\rangle \\
&= \frac{1}{n}\sum_{i=1}^n \int_0^\tau \Bigg\{\mathbf{J}_{\gamma_0^{(1)}}^\top \{\tilde{\dot{\omega}}(\mathbf{X}_i^\top \gamma)\mathbf{X}_i - \bar{\mathbf{X}}(t)\} \\
&\quad -\Bigg[\frac{1}{n}\sum_{j=1}^n \int_0^\tau \mathbf{J}_{\gamma_0^{(1)}}^\top \{\tilde{\dot{\omega}}(\mathbf{X}_j^\top \gamma)\mathbf{X}_j - \bar{\mathbf{X}}(s)\} Y_j(s)(\mathbf{Z}_j^*)^\top R_j(\gamma_0, \mathbf{X}_i^\top \gamma_0, V_i)\, \mathrm{d}s\Bigg] \\
&\quad \times\{H \cdot D(\mathbf{X}_i^\top \gamma_0, V_i)\}^{-1}\{\mathbf{G}_i^*(\mathbf{X}_i^\top \gamma_0, V_i) - \bar{\mathbf{G}}(t, \mathbf{X}_i^\top \gamma_0, V_i)\}\Bigg\}^{\otimes 2} \\
&\quad \times Y_i(t)\lambda(t|\mathbf{Z}_i, \mathbf{X}_i, V_i)\mathrm{d}t \\
&:= \Sigma_n(\gamma_0^{(1)}) = \Sigma(\gamma_0^{(1)}) + o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
\Sigma(\gamma_0^{(1)}) = \int_0^\tau E\Bigg\{ &\Big[\mathbf{J}_{\gamma_0^{(1)}}^\top \{\tilde{\dot{\omega}}(\mathbf{X}^\top \gamma)\mathbf{X} - \bar{\mathbf{X}}(t)\} \\
&-Q \cdot \{H \cdot D(\mathbf{X}^\top \gamma_0, V)\}^{-1}\{\mathbf{G}^*(\mathbf{X}^\top \gamma_0, V) - \bar{\mathbf{G}}(t, \mathbf{X}^\top \gamma_0, V)\}\Big]^{\otimes 2} \\
&\times Y(t)\lambda(t|\mathbf{Z}, \mathbf{X}, V)\Bigg\}\mathrm{d}t,
\end{aligned}
$$

with $Q_i = \int_0^\tau E[\mathbf{J}_{\gamma_0^{(1)}}^\top \{\tilde{\dot{\omega}}(\mathbf{X}^\top \gamma)\mathbf{X} - \bar{\mathbf{X}}(s)\} Y(s)(\mathbf{Z}^*)^\top R(\gamma_0, \mathbf{X}_i^\top \gamma_0, V_i)]\, \mathrm{d}s$.

By verifying, the Lindeberg condition for Martingale Central Limit Theorem holds. We can show that,

$$
\sqrt{n}\tilde{U}_n(\gamma_0^{(1)}) \longrightarrow_d N\left(0, \Sigma(\gamma_0^{(1)})\right). \tag{14}
$$

STEP 4. From steps 2 and 3, we can get the asymptotic normality of $\hat{\gamma}^{(1)}$.

$$0 - \tilde{U}_n(\gamma_0^{(1)}) = \tilde{U}_n(\hat{\gamma}^{(1)}) - \tilde{U}_n(\gamma_0^{(1)})$$
$$= \left\{\partial \tilde{U}_n(\gamma^{(1)})/\partial \gamma^{(1)}|_{\gamma^{(1)}=\gamma_0^{(1)}} + o_p(1)\right\}(\hat{\gamma}^{(1)} - \gamma_0^{(1)}).$$

By (13) and (14), we have

$$\sqrt{n}\left(\hat{\gamma}^{(1)} - \gamma_0^{(1)}\right) \longrightarrow_d N\left(0, \{D(\gamma_0^{(1)})\}^{-1}\Sigma(\gamma_0^{(1)})\{D(\gamma_0^{(1)})\}^{-1}\right).$$

By similar arguments to that in the proof of Theorem 2 in Wang et al. (2010), we have

$$\sqrt{n}(\hat{\gamma} - \gamma_0) \longrightarrow_d N\left(0, \mathbf{J}_{\gamma_0^{(1)}}\{D(\gamma_0^{(1)})\}^{-1}\Sigma(\gamma_0)\{D(\gamma_0^{(1)})\}^{-1}\mathbf{J}_{\gamma_0^{(1)}}^{\top}\right). \tag{15}$$

□

Proof of Theorem 2 Using the results of Step 1–4, we can get the asymptotic property of $\hat{\xi}(u, v)$, or $\tilde{\xi}(\hat{\gamma}, u, v)$.

Define $D_n(\gamma_0, u, v) = \partial\tilde{\xi}(\gamma, u, v)/\partial\gamma^{\top}|_{\gamma=\gamma_0} = D(\gamma_0, u, v) + o_p(1)$.

Note that

$$\sqrt{nh_1h_2}\left\{\hat{\xi}(u, v) - \xi_0(u, v)\right\} = \sqrt{nh_1h_2}\left\{\tilde{\xi}(\hat{\gamma}, u, v) - \tilde{\xi}(\gamma_0, u, v)\right\}$$
$$+ \sqrt{nh_1h_2}\left\{\tilde{\xi}(\gamma_0, u, v) - \xi_0(u, v)\right\}. \tag{16}$$

By (15), we have

$$\sqrt{nh_1h_2}\left\{\tilde{\xi}(\hat{\gamma}, u, v) - \tilde{\xi}(\gamma_0, u, v)\right\}$$
$$= \sqrt{nh_1h_2}\left\{\partial\tilde{\xi}(\gamma, u, v)/\partial\gamma|_{\gamma=\gamma_0}\mathbf{J}_{\gamma_0^{(1)}} + o_p(1)\right\}(\hat{\gamma}^{(1)} - \gamma_0^{(1)})$$
$$= \left\{D_n(\gamma_0, u, v) + o_p(1)\right\}\mathbf{J}_{\gamma_0^{(1)}}\{D_n(\gamma_0^{(1)})\}^{-1}\sqrt{nh_1h_2}\tilde{U}_n(\gamma_0^{(1)})$$
$$= O_p(\sqrt{h_1h_2}), \tag{17}$$

By (11), we have

$$\sqrt{nh_1h_2}\{\tilde{\xi}(\gamma_0, u, v) - \xi_0(u, v)\}$$
$$= \sqrt{nh_1h_2}\{H \cdot D(u, v)\}^{-1}$$
$$\times\left[\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_{1,h_1}(V_i - v)K_{2,h_2}(\mathbf{X}_i^{\top}\gamma_0 - u)\{\mathbf{G}_i^*(u, v) - \bar{\mathbf{G}}(t, u, v)\}\mathrm{d}M_i(t)\right.$$
$$\left. + b(u, v) + o_p(h_1^2) + o_p(h_2^2)\right]. \tag{18}$$

By (16), (17) and (18), we have

$$\begin{aligned}\sqrt{nh_1h_2}\{\hat{\xi}(u,v)-\xi_0(u,v)\} &= \sqrt{nh_1h_2}\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_{1,h_1}(V_i-v)K_{2,h_2}(\mathbf{X}_i^{\top}\gamma_0-u)\\ &\quad \times\{H\cdot D(u,v)\}^{-1}\{\mathbf{G}_i^*(u,v)-\bar{\mathbf{G}}(t,u,v)\}\,\mathrm{d}M_i(t)\\ &\quad +\sqrt{nh_1h_2}\{H\cdot D(u,v)\}^{-1}\{b(u,v)+o_p(h_1^2)\\ &\quad +o_p(h_2^2)\}+O_p(\sqrt{h_1h_2})\\ &:= \tilde{Q}_n+\sqrt{nh_1h_2}\{\tilde{b}(u,v)+o_p(h_1^2)\\ &\quad +o_p(h_2^2)\}+O_p(\sqrt{h_1h_2}),\end{aligned}$$

where $\tilde{b}(u,v)=\{H\cdot D(u,v)\}^{-1}b(u,v)$.

Then,

$$\begin{aligned}\langle\tilde{Q}_n,\tilde{Q}_n\rangle &= \frac{h_1h_2}{n}\sum_{i=1}^{n}\int_0^{\tau} K_{1,h_1}^2(V_i-v)K_{2,h_2}^2(\mathbf{X}_i^{\top}\gamma_0-u)\\ &\quad \times\left[\{H\cdot D(u,v)\}^{-1}\{\mathbf{G}_i^*(u,v)-\bar{\mathbf{G}}(t,u,v)\}\right]^{\otimes 2} Y_i(t)\lambda(t|\mathbf{Z}_i,\mathbf{X}_i,V_i)\mathrm{d}t\\ &:= \Sigma_n(u,v)\\ &= \Sigma(u,v)+o_p(1),\end{aligned}$$

where

$$\Sigma(u,v)=\int_0^{\tau}\{H\cdot D(u,v)\}^{-1}\mathrm{diag}\left(a_2^*(t,u,v),\nu_2^{(1)}\begin{pmatrix}a_2^*(t,u,v) & a_1^*(t,u,v)\\ a_1^*(t,u,v) & a_0^*(t,u,v)\end{pmatrix},\nu_2^{(2)}a_0^*(t,u,v)\right)\{H\cdot D(u,v)\}^{-1}\mathrm{d}t.$$

Thus,

$$\sqrt{nh_1h_2}\left\{\hat{\xi}(u,v)-\xi_0(u,v)-\tilde{b}\right\}\longrightarrow_d N(0,\Sigma(u,v)). \tag{19}$$

□

Acknowledgments Wang's research was supported by the National Science Fund for Distinguished Young Scholars in China (10725106), the National Natural Science Foundation of China (General program 11171331 and Key program 11331011), a grant from the Key Lab of Random Complex Structure and Data Science, CAS and Natural Science Foundation of SZU.

References

Aalen, O.O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine*, *8*, 907–925.

Breslow, N. E., Day, N. E. (1980). Statistical Methods in Cancer Research I, Volume I: The Analysis of Case-Control studies. IARC Scientific Publication 32, International Agency for Research on Cancer, Lyon.

Breslow, N. E., Day, N. E. (1987). Statistical Methods in Cancer Research, Volume II: The Design and Analysis of Cohort studies. IARC Scientific Publication 82, International Agency for Research on Cancer, Lyon.

Cai, Z., Sun, Y. (2003). Local linear estimation for time-dependent coefficients in Cox's regression models. *Scandinavian Journal of Statistics*, *30*, 93–111.

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

Chen, K. N., Tong, X. W. (2010). Varying coefficient transformation models with censored data. *Biometrika*, *97*, 969–976.

Chen, K. N., Jin, Z. Z., Ying, Z. L. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika*, *89*, 659–668.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society Series B*, *34*, 187–220.

Cui, X., Hardle, W., Zhu, L. X. (2011). The EFM approach for single-index models. *The Annals of Statistics*, *39*, 1658–1688.

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

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

Huffer, F. W., McKeague, I.W. (1991). Weighted least squares estimation for Aalen's additive risk model. *Journal of the American Statistical Association*, *86*, 114–129.

Jin, Z., Ying, Z., Wei, L. L. (2001). A simple resampling method by perturbing the minimand. *Biometrka*, *88*, 381–390.

Li, H., Yin, G.S., Zhou, Y. (2007). Local likelihood with time-varying additive hazards model. *The Canadian Journal of Statistics*, *35*, 321–337.

Liang, H., Liu, X., Li, R. Z., Tsai, C. L. (2010). Estimation and testing for partially linear single-index models. *The Annals of Statistics*, *38*, 3811–3836.

Lin, D. Y., Ying, Z. L. (1994). Semiparametric analysis of the additive risk model. *Biometrika*, *81*, 61–71.

Lu, W. B., Zhang, H. H. (2010). On estimation of partially linear transformation models. *Journal of the American Statistical Association*, *105*, 683–691.

McKeague, I. W., Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika*, *81*, 501–514.

Peng, L. M., Huang, Y. J. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association*, *103*, 637–649.

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.

van der vaart, Wellner, J. (1996). *Weak Convergence and Empirical Processes*. New York: Springer.

Xia, Y., Hardle, W. (2006). Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis*, *97*, 1162–1184.

Yin, G. S., Li, H., Zeng, D. L. (2008). Partially linear additive hazards regression with varying coefficients. *Journal of the American Statistical Association*, *103*, 1200–1213.

Yu, Y., Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, *97*, 1042–1054.

Wang, W., Wang, J. L., Wang, Q.H. (2009). Proportional hazards regression with unknown link function. *Institute of Mathematical Statistics*, *57*, 47–66.

Wang, J. L., Xue, L. G., Zhu, L. X., Chong, Y. S. (2010). Estimation for a partial-linear single-index model. *The Annals of Statistics*, *38*, 246–274.