Partially Linear Single Index Cox Regression Model in Nested Case-Control Studies

Shulian Shang[a,*], **Mengling Liu**[a], **Anne Zeleniuch-Jacquotte**[a], **Tess V. Clendenen**[a], **Vittorio Krogh**[b], **Goran Hallmans**[c], and **Wenbin Lu**[d]

[a]Departments of Population Health and Environmental Medicine, New York University, School of Medicine, New York, USA

[b]Nutritional Epidemiology Unit, National Cancer Institute, Milan, Italy

[c]Public Health and Clinical Medicine/Nutritional Research, Umeå University, Umeå, Sweden

[d]Department of Statistics, North Carolina State University, Raleigh, USA

Abstract

The nested case-control (NCC) design is widely used in epidemiologic studies as a cost-effective subcohort sampling method to study the association between a disease and its potential risk factors. NCC data are commonly analyzed using Thomas' partial likelihood approach under the Cox proportional hazards model assumption. However, the linear modeling form in the Cox model may be insufficient for practical applications, especially when there are a large number of risk factors under investigation. In this paper, we consider a partially linear single index proportional hazard model, which includes a linear component for covariates of interest to yield easily interpretable results and a nonparametric single index component to adjust for multiple confounders effectively. We propose to approximate the nonparametric single index function by polynomial splines and estimate the parameters of interest using an iterative algorithm based on the partial likelihood. Asymptotic properties of the resulting estimators are established. The proposed methods are evaluated using simulations and applied to an NCC study of ovarian cancer.

Keywords

nested case-control study; risk-set sampling; nonparametric regression; nonlinear effect; single index model

1. Introduction

Large cohort studies are precious resources to the study of disease etiology. However, it is costly to measure all the risk factors for the entire cohort, especially when disease is rare. As an alternative to the full-cohort design, the nested case-control (NCC) design (Thomas,

*Corresponding author. Tel: +1 212 263 0371. Fax: +1 212 263 8570: shulian.shang@gmail.com.

1979) has been widely used as a cost-effective subcohort sampling method. In this design, cases are ascertained within a large cohort. At the failure time of each case, a small number of controls are sampled among subjects who are still at risk, possibly matched to the case by some known confounders. Then covariates of interest are only measured on the cases and selected controls. The NCC design maintains the attractive feature of the full-cohort design to analyze biological specimens collected before the disease onset, providing an appropriate time sequence for a cause-effect relationship. In addition, both absolute risk and relative risk can be estimated under the NCC design (Langholz & Borgan, 1997).

NCC data are commonly analyzed using Thomas' partial likelihood approach under the Cox proportional hazards (PH) model (Thomas, 1979; Oakes, 1981), for which the hazard function is specified as $\lambda(t/x) = \lambda_0(t) \exp\{x^T\beta\}$, where $\lambda_0(t)$ is the unknown baseline hazard function and x is a p-dimensional covariate vector. A major assumption of the Cox PH model is that covariates have log-linear effects on the disease hazard. In epidemiologic studies where a large number of covariates are considered, covariates often exhibit more complex effects than log-linear format and there may exist interactions between them. Flexible models which could handle potential nonlinear effects of covariates with high dimensionality are greatly desired.

The single index model (Stoker, 1986; Hardle & Stoker, 1989; Ichimura, 1993) is a semiparametric model which achieves dimension reduction and avoids the "curse of dimensionality". In the linear regression setting, the single index model is an extension of the generalized linear model, with link function unspecified (Yu & Ruppert, 2002). In the survival analysis context, the single index model has been incorporated into the multiplicative hazard model (Wang, 2004; Huang & Liu, 2006):

$$\lambda(t|x) = \lambda_0(t)\exp\{\psi(x^T\beta)\}, \quad (1)$$

where $\psi(\cdot)$ is an unknown univariate smooth function. The multi-dimensional covariates are reduced to a linear combination as $x^T\beta$, namely a single index, and the coefficient β characterizes the relative importance of x. Note that if $\psi(\cdot)$ is monotone, β has a similar interpretation as the coefficient in the Cox PH model. Researchers have proposed various methods for fitting the single index model, such as kernel smoothing technique (Ichimura, 1993; Hardle et al, 1993; Wang, 2004), average derivatives method (Stoker, 1986; Hardle & Stoker, 1989) and polynomial spline approximation (Yu & Ruppert, 2002; Huang & Liu, 2006).

In model (1), all components of x are treated equally in the sense that no distinction is made between covariates of primary interest and nuisance ones. In epidemiologic studies, there are usually major risk factors of interest and multiple confounders such as demographics, anthropometric measures and socioeconomic status. Covariates that are most interesting to investigators would be modeled parametrically to render easy interpretation on their effects. Therefore, a partially linear single index (plSI) model has been proposed to extend model (1),

$$\lambda(t|v,x) = \lambda_0(t)\exp\{v^T\alpha + \psi(x^T\beta)\}, \quad (2)$$

where $v \in R^q$, $x \in R^p$ and $\psi(\cdot)$ is the unknown link function as above. In the linear regression setting, researchers have proposed to use local linear method (Carroll et al, 1997), kernel smoothing method (Xia et al, 1999) and penalized spline method (Yu & Ruppert, 2002) to fit the partially linear single index model. In survival analysis, Lu et al. (2006) considered model (2) with a parametric baseline hazard function; Sun et al. (2008) studied this model with the polynomial spline technique; Li & Zhang (2011) extended model (2) to time-varying coefficients.

To the best of our knowledge, the inference of model (2) has not been studied for the NCC design. In this paper, we develop methods for the statistical inference of model (2) for NCC data and establish asymptotic properties of the resulting estimators. We are motivated by an NCC study investigating the association of inflammation-related cytokins and their modulators with the risk of ovarian cancer (Clendenen et al., 2011). This case-control study was nested within three prospective cohorts and for each case, two controls were selected at random from cohort members who fulfilled the risk set criteria. In total, we observed 230 cases and 432 matched controls. The levels of cytokines and cytokine modulators were measured from stored blood samples collected at enrollment. Potential confounders included body mass index and medical history. Our main interest is to estimate the effect of biomarkers on the risk of ovarian cancer while adjusting for confounders. We thus study the partially linear single index proportional hazards model, which allows flexible and parsimonious modeling of nonlinear effects of confounders and easy interpretation on the parameters for covariates of interest.

This paper is organized as follows. In Section 2, we present methods for estimation, inference and implementation of the proposed model. Section 3 includes simulation studies evaluating the finite sample performance of our proposed estimator and the analysis of the NCC data on ovarian cancer as an illustration. We conclude in Section 4 with discussions and provide all the technical details in the appendix.

2. Methods

2.1. Notation and model

Suppose that we have a cohort with size of n. For the ith subject, $i = 1,\ldots,n$, let $z_i = \min(t_i,c_i)$ be the observed survival time subject to censoring, where t_i denotes the survival time and c_i denotes the censoring time. Define $\delta_i = I(t_i \leq c_i)$ as the censoring indicator. At a specific time t, let $\tilde{R}(t) = \{i:z_i \geq t\}$ denote the risk set. By the NCC design, subjects with observed event, i.e. $\delta_i = 1$, are identified as cases. At the failure time of each case, $(M-1)$ controls are randomly sampled without replacement from the risk set, excluding the case itself. For case i, let R_i^* denote the indices of the $(M-1)$ selected controls and define the case-control set $R_i = R_i^* \cup \{i\}$. Then covariate information is assembled for the cases and selected controls, consisting of two components: the q-dimensional vector v denotes the primary risk factors to be modeled parametrically, and the p-dimensional vector x denotes confounders to be included in the nonparametric single index component.

For the purpose of identifiability of the partially linear single index model (2), we impose the constraints that $\psi(0) = 0$, $\| \beta \| = (\beta^T \beta)^{1/2} = 1$ and the first nonzero component of β is positive (Wang, 2004). Following Huang & Liu (2006) and Sun et al. (2008), we use a polynomial spline function to first approximate the derivative of the unknown function $\psi(\cdot)$ by

$$\psi'(x^T\beta)=\sum_{j=1}^{k}\gamma_j B_j(x^T\beta)=\gamma^T \mathbf{B}(\mathbf{x^T}\beta), \quad (3)$$

where $B_j(u), j = 1,\ldots, k$, are the B-spline basis functions (De Boor, 1978), k equals the sum of the number of interior knots and the order of B-spline, $\mathbf{B}(u) = \{B_1(u),\ldots,B_k(u)\}^T$ and $\gamma = (\gamma_1, \ldots, \gamma_k)^T$. This approximation technique facilitates the incorporation of the constraint $\psi(0) = 0$ as described below. The B-spline is chosen here for numerical stability, and other basis such as truncated power function basis can also be used. From the constraint $\psi(0) = 0$ and (3), we obtain

$$\psi(x^T\beta) = \int_0^{x^T\beta}\sum_{j=1}^{k}\gamma_j B_j(t)\,dt=\sum_{j=1}^{k}\gamma_j \tilde{B}(x^T\beta) = \gamma^T \tilde{\mathbf{B}}(x^T\beta),$$

where $\tilde{B}_j(u) = \int_0^u B_j(s)\,ds, j = 1,\ldots,k$, are the integrals of the B-spline basis functions and $\tilde{\mathbf{B}}(u) = \{\tilde{B}_1(u),\ldots, \tilde{B}_k(u)\}^T$. In our numerical studies, quadratic B-splines are used in the basis expansion of $\psi'(\cdot)$ and thus $\psi(\cdot)$ is a cubic spline.

Let $\tau_1 < \ldots < \tau_m$ be m distinctive ordered event times and (v_i, x_i) be the covariates associated with the subject that fails at τ_i. Then the log-partial likelihood function for NCC data under model (2) is

$$pl(\beta,\gamma,\alpha)=\sum_{i=1}^{m}\left(v_i^T\alpha+\gamma^T\tilde{\mathbf{B}}(x_i^T\beta) - \log\left[\sum_{l\in R_i}\exp\{v_l^T\alpha+\gamma^T\tilde{\mathbf{B}}(x_l^T\beta)\}\right]\right). \quad (4)$$

2.2. Parameter estimation

To maximize the log-partial likelihood (4), we first examine its score functions and Hessian matrix. Specifically, the score functions are:

$$S_\beta(\beta,\gamma,\alpha)=\sum_{i=1}^{m}\left\{\gamma^T\mathbf{B}(\mathbf{x_i^T}\beta)\mathbf{x_i} - \sum_{l\in\mathbf{R_i}}\omega_{li}\gamma^T\mathbf{B}(\mathbf{x_l^T}\beta)\mathbf{x_l}\right\},$$

$$S_\gamma(\beta, \gamma, \alpha) = \sum_{i=1}^{m} \left\{ \tilde{\mathbf{B}}(x_i^T \beta) - \sum_{l \in R_i} \omega_{li} \tilde{\mathbf{B}}(x_l^T \beta) \right\},$$

$$S_\alpha(\beta, \gamma, \alpha) = \sum_{i=1}^{m} \left(v_i - \sum_{l \in R_i} \omega_{li} v_l \right),$$

where $\omega_{li} = \exp\{v_l^T \alpha + \gamma^T \tilde{\mathbf{B}}(x_l^T \beta)\} / \sum_{j \in R_i} \exp\{v_j^T \alpha + \gamma^T \tilde{\mathbf{B}}(x_j^T \beta)\}$, and the Hessian matrix is given in the appendix. It is easily seen that the log partial likelihood is a concave function of (γ, α) for fixed β, which leads us to consider an iterative alternating optimization procedure to calculate the maximum partial likelihood estimator. Specifically we propose the following iterative optimization algorithm:

- Step 1. Start with initial values $\beta^{(\hat{0})}, \gamma^{(\hat{0})}$ and $\alpha^{(\hat{0})}$.

- Step 2. Given the current values $\beta^{(\hat{d})}, \gamma^{(\hat{d})}$ and $\alpha^{(\hat{d})}$, update the estimate of β using one step of the Newton-Raphson method:

$$\hat{\beta}^{(d+1)} = \hat{\beta}^{(d)} - \left\{ H_{\beta,\beta}(\hat{\beta}^{(d)}, \hat{\gamma}^{(d)}, \hat{\alpha}^{(d)}) \right\}^{-1} S_\beta(\hat{\beta}^{(d)}, \hat{\gamma}^{(d)}, \hat{\alpha}^{(d)}),$$

  where S_β and $H_{\beta,\beta}$ are the marginal score function and hessian matrix with respect to β. Standardize $\beta^{(\hat{d}+1)}$ such that $\|\beta^{(\hat{d}+1)}\| = 1$ and its first component is positive.

- Step 3. Given the current values $\beta^{(\hat{d}+1)}, \gamma^{(\hat{d})}$ and $\alpha^{(\hat{d})}$, update the estimates of γ and α using one step of the Newton-Raphson method with step-halving as follows:

$$(\hat{\gamma}^{(d+1)}, \hat{\alpha}^{(d+1)}) = (\hat{\gamma}^{(d)}, \hat{\alpha}^{(d)}) - 2^{-k} \left\{ H_{\gamma,\alpha}(\hat{\beta}^{(d+1)}, \hat{\gamma}^{(d)}, \hat{\alpha}^{(d)}) \right\}^{-1} S_{\gamma,\alpha}(\hat{\beta}^{(d+1)}, \hat{\gamma}^{(d)}, \hat{\alpha}^{(d)}),$$

  where k is the smallest nonnegative integer such that $l(\beta^{(\hat{d}+1)}, \gamma^{(\hat{d}+1)}, \alpha^{(\hat{d}+1)})$ $l(\beta^{(\hat{d}+1)}, \gamma^{(\hat{d})}, \alpha^{(\hat{d})})$, and $S_{\gamma,\alpha}$ and $H_{\gamma,\alpha}$ are the joint score function and hessian matrix with respect to (γ, α).

- Step 4. Repeat steps 2 and 3 until the parameter convergence criterion of 10^{-4} is met.

As the log-partial likelihood in (4) is concave in (γ, α) but not in β, it is not guaranteed that the algorithm converges to the global maximum. We use a variety of randomly generated initial values and choose the final estimate to be the one giving the largest log-partial likelihood. As pointed out by Huang & Liu (2006) and Sun et al. (2008), although the log-partial likelihood could be maximized simultaneously with respect to (β, γ, α), the iterative alternating procedure is numerically more stable. In our simulation studies, the algorithm performs quite well. It converges within a few iterations about 96% of the time. The program is terminated if it does not converge in 60 iterations. To find initial values in a real

application, one could first fit some simple models such as a completely parametric (e.g. a linear Cox proportional hazards model) or a nonparametric model (e.g. a single index model). One could examine the estimated link function $\psi(\cdot)$ from the single index model, and if it is close to some known functions such as the trigonometric functions, one could fit a model with a specific link function to obtain initial values.

In our numerical studies, we use 3 to 10 knots equally spaced in the range of the estimated index values $\hat{\beta^T}x$, and choose the final number of knots by minimizing the Akaike (AIC) or Bayesian information criterion (BIC). When the single index values are skewed or unevenly distributed, we suggest placing knots at sample quantiles of the single index values, which avoids placing a large number of knots in regions where data are sparse.

2.3. Inference

To estimate the variance-covariance matrix of the parameters, we first reparametrize $\beta = \beta(\sigma) = \{(1-\|\sigma\|^2)^{1/2}, \sigma_1, \ldots, \sigma_{p-1}\}^T$ with $\sigma = (\sigma_1, \ldots, \sigma_{p-1})^T$ such that the constraints $\|\beta\| = 1$ and $\beta_1 \geq 0$ are automatically satisfied. Because of the risk-set sampling mechanism of the NCC design, the size of each case-control set is fixed to be M and the asymptotics is driven by the increasing number of case-control sets (Langholz, 2005). We adopt the formulation for Thomas' maximum partial likelihood estimator used in Goldstein & Langholz (1992) and consider that the nonparametric function $\psi(\cdot)$ is a spline function with pre-specified knots. We show that the maximum partial likelihood estimators $(\hat{\sigma}, \hat{\gamma}, \hat{a})$ are consistent and asymptotically normal. We estimate the asymptotic variance-covariance matrix by $\{-H_{(\sigma,\gamma,a)}\}^{-1}$, where $H_{(\sigma,\gamma,a)}$ is the Hessian matrix of (4). Details of regularity conditions and proofs are given in the appendix. By the delta method, $(\hat{\beta}, \hat{\gamma}, \hat{a})$ are also asymptotically normal and the estimated asymptotic variance-covariance matrix $\mathrm{var}(\hat{\beta}, \hat{\gamma}, \hat{a})$ is

$$\begin{pmatrix} \frac{\hat{\beta}_2}{\hat{\beta}_1} & \cdots & \frac{\hat{\beta}_p}{\hat{\beta}_1} & 0_{1\times(k+q)} \\ & I_{p-1+k+q} & \end{pmatrix} \{-H_{(\hat{\sigma},\hat{\gamma},\hat{a})}\}^{-1} \begin{pmatrix} \frac{\hat{\beta}_2}{\hat{\beta}_1} & \cdots & \frac{\hat{\beta}_p}{\hat{\beta}_1} & 0_{1\times(k+q)} \\ & I_{p-1+k+q} & \end{pmatrix}^T, \quad (5)$$

where $I_{p-1+k+q}$ is an identity matrix of size $(p - 1 + k + q)$ and $\mathbf{0}_{1 \times (k+q)}$ is a zero vector with dimension $1 \times (k + q)$.

From the diagonal elements of the matrix $\mathrm{var}(\hat{\beta}, \hat{\gamma}, \hat{a})$ we can get the variance estimate for each of the estimated parameters. For a fixed u, the variance of the function $\psi(\cdot)$ evaluated at u can be estimated as $\mathrm{var}\{\hat{\psi}(u)\} = \mathbf{\tilde{B}}(u)^T\mathrm{var}(\hat{\gamma})\mathbf{\tilde{B}}(u)$. An approximate 95% point-wise confidence interval for $\psi(u)$ is given by $\hat{\psi(u)} \pm 1.96\{\mathrm{var}(\hat{\psi(u)})\}^{1/2}$.

3. Numerical studies

3.1. Efficacy simulations

To evaluate the finite sample performance of our proposed methods, we conducted extensive simulations under various settings. We generated the survival time T from the following models

- Log quadratic: $\lambda_0(t) \exp[v^T a + \log\{1 + (x^T\beta)^2\}]$;

- Sine curve: $\lambda_0(t) \exp\{ v^T \alpha + 5\sin(x^T\beta/2)\}$;

- Linear: $\lambda_0(t) \exp\{ v^T \alpha + x^T\beta \}$,

where $V = (V_1, V_2) \sim U(-3, 3)$ independently, and $X = (X_1, X_2) \sim U(-4,4)$, $X_3 \sim N(0, 2)$ independently with X_3 truncated to be within $[-4, 4]$. The true parameters were $\alpha_0 = (-1, 2)^T$ and $\beta_0 = (1, -1, 1)^T/\sqrt{3}$. The baseline hazard function was $\lambda_0(t) = 1$. As NCC studies are usually used when the disease incidence rate is low, the censoring time C was generated independently by a Cox PH model with the same relative risk function but different baseline hazard functions to yield the incidence rates about 10% or 20%. The size of full cohort was 1000 or 2000, and 2 controls were selected for each case. For each setting, 500 runs of simulations were conducted.

We used both the AIC and BIC to choose the number of knots for spline approximation as in Huang & Liu (2006), with knots equally spaced in the range of $\hat{\beta}^T x$. For comparison, the true model with a known link function was fitted as a benchmark and the standard Cox PH model was also assessed. To evaluate the estimated coefficient β for the single index component, we used the angle between the true parameter vector β_0 and its estimate $\hat{\beta}$, defined as

$$\omega(\beta_0, \hat{\beta}) = \arccos\left(\frac{\langle \beta_0, \hat{\beta} \rangle}{\|\beta_0\| \cdot \|\hat{\beta}\|} \right),$$

where $\langle a, b \rangle$ denotes the inner product of two vectors a and b.

Table 1 shows the results for the log quadratic model, and indicates that the proposed method estimates the parameters reasonably well. The empirical coverage probabilities of the 95% confidence intervals for α are close to the nominal level, indicating that the standard error estimates are accurate. For a fixed censoring rate, when sample size increases, both the biases and standard errors of the estimates of α decrease; the same can be seen for the mean and standard deviation of the angle between β_0 and $\hat{\beta}$. The Cox PH model gives biased estimates for α, and the angles between β_0 and $\hat{\beta}$ are very large. Figure 1 shows the median of estimated function $\psi(\cdot)$ and 95% pointwise Monte Carlo intervals which are constructed using the 2.5% and 97.5% sample quantiles of the estimated link function from 500 simulations. The estimated function approximates the true function very well. Table 2 presents the results when the link function is a sine function, and shows similar results.

To evaluate the efficiency loss when the log relative risk function is indeed linear, we conducted simulation with the linear link function. Table 3 shows the parameter estimates from the Cox PH model and the proposed model. The estimates from the partially linear single index model are close to the Cox model and the relative efficiency for estimating α is about 0.90 and 0.96 for cohort size 1000 and 2000, respectively. The angles between β_0 and $\hat{\beta}$ from the two models are very close. Thus, the proposed model maintains good efficiency when the true log-hazard function is a linear function.

When the number of knots is 5, the number of matched controls is fixed and the number of cases is 200, the computation time of one simulation run is about 6 seconds with a randomly generated initial value, and about 3 seconds with a good initial value close to the true parameter, on a 2.66 GHz processor with 4 GB of memory.

3.2. Sensitivity analysis

We performed sensitivity analysis to evaluate the proposed approaches when the model is misspecified. We first considered the scenario where the true model is the single index model, where the hazard function for the survival time T was specified as

$$\lambda(t|v,x) = \lambda_0(t) \exp\{\psi(v^T\alpha + x^T\beta)\}.$$

We generated independent covariates $V \sim U(-4,4)$, $X_1, X_2, X_3 \sim U(-4,4)$, $X_4 \sim N(0,2)$ with X_4 truncated to be within $[-4, 4]$. The true parameters were $\alpha_0 = 1/\sqrt{5}$ and $\beta_0 = (1, -1, 1, -1)^T/\sqrt{5}$. The baseline hazard function was $\lambda_0(t) = 1$. The link function $\psi(\cdot)$ was the log-quadratic link or the sine link as specified previously The size of full cohort was 1000 or 2000, the disease incidence rate was about 20%, and 2 controls were selected for each case. The number of simulation runs was 200.

Since the true model is the single index model, only the direction of the parameter is identifiable. The angles between $\hat{\beta}$ and β_0 are given in Table 4a for the single index (SI) model, partially linear single index (plSI) model and Cox PH model. When fitting the plSI model, the covariate V was included in the parametric part of model (2). The results show that the plSI model performs reasonably well compared to the correctly specified single index model, and better than the Cox PH model.

We also considered the scenarios where the true model is the partially linear single index model but some covariates components are misspecified. The hazard function for T has the log-quadratic form specified previously. Four scenarios were considered.

(S1). A redundant variable was included in the X part. All other settings were the same as in Section 3.1, except that one extra variable X_4 which followed $N(0,1)$ independently of the other covariates was included in the nonlinear part when fitting the proposed model.

(S2). A covariate was omitted from the fitting of single index part. We considered that the single index component X included four covariates, with $X_1, X_2 \sim U(-4,4)$ independently; X_3, X_4 following a bivariate Normal distribution with mean 0, standard deviation 1 and correlation 0.8, and truncated to be within $[-3, 3]$. The true parameters were $\alpha_0 = (-1, 2)$ and $\beta_0 = (1/\sqrt{3}, -1/\sqrt{3}, 1/\sqrt{3}, 0.1)^T$. Covariate X_4 was omitted when fitting models.

(S3). The membership of V and X was misspecified with a linear covariate modeled nonlinear effect. The linear component V included three covariates, with $V_1, V_2 \sim U(-3,3)$, $V_3 \sim N(0,2)$ independently and V_3 truncated to be within $[-4, 4]$; the nonlinear part X included three covariates, with $X_1, X_2, X_3 \sim U(-4,4)$ independently. The true

parameters were $\alpha_0 = (-1,2,0.1)^T$ and $\beta_0 = (0.574,-0.574,0.574)^T$. When fitting the proposed model, V_3 was included in the single index part.

(S4). The membership of V and X was misspecified with a covariate with nonlinear effect modeled in the linear part. X included four covariates, with $X_1, X_2, X_3 \sim U(-4,4)$, $X_4 \sim N(0,2)$ independently. The true parameters were $\alpha_0 = (-1,2)^T$ and $\beta_0 = (0.574, -0.574, 0.574, 0.1)^T$. When fitting the proposed model, X_4 was assigned to the linear part.

The cohort size was 1000 and censoring rate was about 80%. We selected 2 controls for each case and ran 200 simulations for each setting. The results are shown in Table 4b. We observe that including a redundant covariate (S1) does not affect the proposed method much and the results are similar to Table 1, with the angle between β_0 and $\hat{\beta}$ slightly greater. For the other settings (S2)-(S4), the empirical coverage probability of 95% confidence interval for α deviates from the nominal level, and the angle between β_0 and $\hat{\beta}$ enlarges. In all of the scenarios the proposed model outperforms the Cox PH model and shows better flexibility of accommodating various model misspecifications. When other link functions are used, similar results are obtained (results not shown).

3.3. Analysis of the NCC study on ovarian cancer

The NCC study of ovarian cancer (Clendenen et al., 2011) assessed the association between circulating inflammatory cytokines and the risk of epithelial ovarian cancer. As an illustration, we studied the cytokine IL-6 adjusting for the confounders body mass index (BMI), age at menarche, ever been pregnant, ever use of oral contraceptive (OC) and ever use of hormone replacement therapy (HRT). The cytokine IL-6 was first cutted to categories using cohort-specific quartiles, and the first quartile was treated as the baseline. For the partially linear single index model, the indicators of IL-6 quartiles were assigned to the linear component, and the confounders were assigned to the nonlinear component. B-splines with equally spaced knots were used and 5 knots were selected by the AIC criterion. The standard Cox PH model with all the covariates assumed linear effects was also fitted for comparison. Since the Cox model is nested in the proposed model, a likelihood ratio test can be used to examine whether the Cox PH model is appropriate and has an approximate χ^2 distribution with $k-1$ degrees of freedom. The null hypothesis was rejected ($p = 0.006$), indicating that the Cox PH model is insufficient for this dataset. When the relative risk scores of each model were used to classify cases and controls, the areas under the Receiver Operating Characteristic (ROC) curves are 0.59 and 0.63 for the Cox model and proposed model, respectively ($p = 0.060$, DeLong's test).

Figure 2 shows the estimated link function $\psi(\cdot)$, which is nonlinear and non-monotone. Table 5 presents the estimated parameters, standard errors and p-values from the two models. For ease of comparison, coefficients of the covariates in the nonlinear component of the proposed model were rescaled to have the same norm as those for the Cox model. In the Cox model, the fourth quartile of IL-6 has a significantly higher risk comparing to the first quartile (OR = 1.61, $p = 0.045$). Using the proposed model, the estimated coefficients and standard errors for IL-6 quartiles are similar to those of the Cox model. When the other covariates are fixed, the odds ratio of the fourth quartile vs. the first quartile of IL-6 is 1.57.

We also modeled IL-6 linearly in its continuous level and obtained similar results. Regarding the confounders, ever use of HRT is a significant risk factor of ovarian cancer and has the largest effect size in both models. Age at menarche, ever been pregnant and ever use of OC are significant in the proposed model, but not in the Cox model. The angle between the estimated parameter vector $\hat{\beta}$ from the two models is 21.84°.

4. Discussion

The partially linear single index model is a natural extension of partially linear model and single index model. The high-dimensional nuisance covariates with possible nonlinear effects are first combined as a single index, providing a flexible and parsimonious way of modeling. We have shown that the proposed model performs better than the standard Cox PH model for various link functions. Moreover, coefficients of covariates in the linear component have easy interpretation as the log hazard ratio.

In this paper we use polynomial splines to approximate the nonparametric single index function. Several other approaches have been proposed to fit the partially linear single index model in the full-cohort setting, such as kernel smoothing method and penalized spline method. However, the kernel weighted smoothing technique may not always be applicable to estimation of the plSI proportional hazards model with NCC data because the risk set for each case only consists of the case itself and its controls. If covariate X is an important risk factor with significant different distribution between cases and controls, the index value ($\beta^T X$) of the control is rarely in the neighborhood of the case. This will cause the optimization of the kernel weighted partial likelihood run into difficulty.

We choose to use the polynomial regression spline instead of penalized spline due to computational and theoretical reasons. Penalized spline can be viewed as a compromise between regression spline and smoothing spline. While the fitting is more stable and less dependent on the location of knots for penalized spline, the computation and inference are more complicated. In particular, selecting a suitable value of smoothing parameter is crucial. As for survival data an iterative algorithm has to be utilized for the optimization of the penalized partial likelihood, the search of optimal smoothing parameter becomes computationally expensive.

With large and complex data, the single index assumption may be further relaxed to multiple index modeling. One could consider a partially linear multiple index model: $\lambda(t|\upsilon,x) = \lambda_0(t) \exp\{\upsilon^T\alpha + \psi(x^T\beta_1,\ldots,x^T\beta_s)\}$, where $\psi(\cdot)$ is an unknown s-variate function and s is a pre-specified integer less than p. The multiple index model has been studied by many researchers (Cook & Bing, 2002; Xia et al, 2002; Yin & Cook, 2002; Chen et al., 2011). When the number of indices s is large, the s-variate unknown function may be replaced by s univariate unknown functions, leading to the additive-index model (Chiou & Muller, 2004).This model is closely related to the projection pursuit regression (Friedman & Stuetzle, 1981).

In this paper, we have assumed that the unknown function is a spline function with fixed number of knots in establishing the asymptotic properties. The bias caused by spline approximation is small compared to the variance of the estimated function, as shown by our

simulation studies. Alternatively without assuming the unknown function is a spline function, the number of knots needs to increase as sample size increases. Developing asymptotic results in that context is interesting but beyond the scope of this paper.

Appendix A: Formulas of the Hessian matrices of (β, γ, α)

Let

$$\omega_{li} = \exp\{v_l^T \alpha + \gamma^T \tilde{\mathbf{B}}(x_l^T \beta)\} / \sum_{j \in Ri} \exp\{v_j^T \alpha + \gamma^T \tilde{\mathbf{B}}(x_j^T \beta)\},$$

$$\mathbf{B}'(u) = (B_1'(u), \dots, B_k'(u))^T,$$

$$H_1 = \sum_{i=1}^{m} \gamma^T B'(x_i^T \beta) x_i x_i^T,$$

$$H_2 = \sum_{i=1}^{m} \sum_{l \in R_i} \omega_{li} \left\{ \gamma^T \mathbf{B}'(x_l^T \beta) + \{\gamma^T \mathbf{B}(\mathbf{x_l^T} \beta)\}^2 \right\} x_l x_l^T - \sum_{i=1}^{m} \left\{ \sum_{l \in R_i} \omega_{li} \gamma^T \mathbf{B}(\mathbf{x_l^T} \beta) \mathbf{x_l} \sum_{l \in \mathbf{R_i}} \omega_{li} \gamma^{\mathbf{T}} \mathbf{B}(\mathbf{x_l^T} \beta) \mathbf{x_l^T} \right\}.$$

We have

$$H_{\beta,\beta} = \frac{\partial^2 l}{\partial \beta \partial \beta^T} = H_1 - H_2,$$

$$H_{\gamma,\gamma} = \frac{\partial^2 l}{\partial \gamma \partial \gamma^T} = -\sum_{i=1}^{m} \left\{ \sum_{l \in R_i} \omega_{li} \tilde{B}(x_l^T \beta) \tilde{B}^T(x_l^T \beta) - \sum_{l \in R_i} \omega_{li} \tilde{B}(x_l^T \beta) - \sum_{l \in R_i} \omega_{li} \tilde{B}^T(x_l^T \beta) \right\},$$

$$H_{\alpha,\alpha} = \frac{\partial^2 l}{\partial \alpha \partial \alpha^T} = -\sum_{i=1}^{m} \left\{ \sum_{l \in R_i} \omega_{li} v_l v_l^T - \sum_{l \in R_i} \omega_{li} v_l \sum_{l \in R_i} \omega_{li} v_l^T \right\},$$

$$H_{\beta,\gamma} = \frac{\partial^2 l}{\partial \beta \partial \gamma} = \sum_{i=1}^{m} \left[\mathbf{B}(\mathbf{x_i^T} \beta) \mathbf{x_i} - \sum_{l \in \mathbf{R_i}} \omega_{li} \left\{ \tilde{\mathbf{B}}(\mathbf{x_l^T} \beta) \gamma^{\mathbf{T}} \mathbf{B}(\mathbf{x_l^T} \beta) + \mathbf{B}(\mathbf{x_l^T} \beta) \right\} \mathbf{x_l^T} \right] + \sum_{i=1}^{m} \left\{ \sum_{l \in R_i} \omega_{li} \tilde{\mathbf{B}}(x_l^T \beta) \sum_{l \in R_i} \omega_{li} \gamma^T \mathbf{B}(\mathbf{x_l^T} \beta) \mathbf{x_l^T} \right\},$$

$$H_{\beta,\alpha}=\frac{\partial^2 l}{\partial\beta\partial\alpha}=-\sum_{i=1}^{m}\left\{\sum_{l\in R_i}\omega_{li}\gamma^T\mathbf{B}(\mathbf{x_l^T}\beta)\upsilon_{\mathbf{l}}\mathbf{x_l^T}-\sum_{l\in\mathbf{R_i}}\omega_{li}\upsilon_{\mathbf{l}}\sum_{l\in\mathbf{R_i}}\omega_{li}\gamma^{\mathbf{T}}\mathbf{B}(\mathbf{x_l^T}\beta)\mathbf{x_l^T}\right\},$$

$$H_{\gamma,\alpha}=\frac{\partial^2 l}{\partial\gamma\partial\alpha}=-\sum_{i=1}^{m}\left\{\sum_{l\in R_i}\omega_{li}\tilde{\mathbf{B}}(x_l^T\beta)\upsilon_l^T-\sum_{l\in R_i}\omega_{li}\tilde{\mathbf{B}}(x_l^T\beta)\sum_{l\in R_i}\omega_{li}\upsilon_l^T\right\}.$$

Appendix B: Formulas for H(σ,γ,α)

The Hessian matrix $H_{(\sigma,\gamma,\alpha)}$ is

$$H_{(\sigma,\gamma,\alpha)}=\left(\begin{array}{ccc}H_{\sigma,\sigma}&H_{\sigma,\gamma}&H_{\sigma,\alpha}\\H_{\sigma,\gamma}^T&H_{\gamma,\gamma}&H_{\gamma,\alpha}\\H_{\sigma,\alpha}^T&H_{\gamma,\alpha}^T&H_{\alpha,\alpha}\end{array}\right),$$

where $H_{\gamma,\gamma}$, $H_{\gamma,\alpha}$ and $H_{\alpha,\alpha}$ are the same as those given in Appendix A. The rest of the components of $H_{(\sigma,\gamma,\alpha)}$ are given as follows.

Let the vector $x_i\tilde{}=(x_{i2},\ldots,x_{ip})^T$, $\xi_i=-x_{i1}\sigma/(1-\|\sigma\|^2)^{\frac{1}{2}}+\tilde{x}_i^T$, $i=1,\ldots,n$. Let $A=(a_{ij})$ be a $(p-1)\times(p-1)$ matrix with entries $a_{ii}=1+\sigma_i^2/(1-\|\sigma\|^2)$ and $a_{ij}=\sigma_i\sigma_j/(1-\|\sigma\|^2)$, i $\ j$, i, $j=1,\ldots,p-1$. In other words, $A=I_{p-1}+\sigma\sigma^T/(1-\|\sigma\|^2)$.

$$\begin{aligned}H_{\sigma,\sigma}=&\sum_{i=1}^{m}\left\{\gamma^T\mathbf{B}(x_i^T\beta)\left(-\frac{x_{i1}}{\sqrt{1-\|\sigma\|^2}}\right)A+\gamma^T\mathbf{B}'(x_i^T\beta)\xi_i\xi_i^T\right\}\\&-\sum_{i=1}^{m}\sum_{l\in R_i}\omega_{li}\left\{\gamma^T\mathbf{B}(\mathbf{x_l^T}\beta)\left(-\frac{\mathbf{x_{l1}}}{\sqrt{1-\|\sigma\|^2}}\right)\mathbf{A}\right\}\\&+\sum_{i=1}^{m}\sum_{l\in R_i}\omega_{li}[\gamma^T\mathbf{B}'(x_l^T\beta)\xi_l\xi_l^T+\{\gamma^T\mathbf{B}(\mathbf{x_l^T}\beta)\}^2\xi_l\xi_l^T]\\&+\sum_{i=1}^{m}\left\{\sum_{l\in R_i}\omega_{li}\gamma^T\mathbf{B}(\mathbf{x_l^T}\beta)\xi_{\mathbf{l}}\sum_{l\in\mathbf{R_i}}\omega_{li}\gamma^{\mathbf{T}}\mathbf{B}(\mathbf{x_l^T}\beta)\xi_{\mathbf{l}}^{\mathbf{T}}\right\},\end{aligned}$$

$$\begin{aligned}H_{\sigma,\gamma}=&\sum_{i=1}^{m}\left[\xi_i\mathbf{B}^T(x_i^T\beta)-\sum_{l\in R_i}\omega_{li}\left\{\xi_l\mathbf{B}^T(x_l^T\beta)+\gamma^T\mathbf{B}(\mathbf{x_l^T}\beta)\xi_{\mathbf{l}}\tilde{\mathbf{B}}^{\mathbf{T}}(\mathbf{x_l^T}\beta)\right\}\right]\\&+\sum_{i=1}^{m}\left\{\sum_{l\in R_i}\omega_{li}\gamma^T\mathbf{B}(\mathbf{x_l^T}\beta)\xi_{\mathbf{l}}\sum_{l\in\mathbf{R_i}}\omega_{li}\tilde{\mathbf{B}}^{\mathbf{T}}(\mathbf{x_l^T}\beta)\right\},\end{aligned}$$

$$H_{\sigma,\alpha}=\sum_{i=1}^{m}\left\{-\sum_{l\in R_i}\omega_{li}\gamma^T\mathbf{B}(\mathbf{x_l^T}\beta)\xi_1 v_\mathbf{l}^\mathbf{T}+\sum_{l\in\mathbf{R_i}}\omega_{li}\gamma^\mathbf{T}\mathbf{B}(\mathbf{x_l^T}\beta)\xi_1\sum_{l\in\mathbf{R_i}}\omega_{li}v_\mathbf{l}^\mathbf{T}\right\}.$$

Appendix C: Consistency and asymptotic normality of the maximum partial likelihood estimator (β, γ̂, α̂)

There are two types of asymptotics, one with increasing number of knots and one with fixed number of knots. We study the case of fixed number of knots because it is simpler and gives a practically useful result. It is assumed that the nonparametric function $\psi(\cdot)$ is a spline function with fixed knots.

Define $N_i(t) = I(Z_i \quad t, \delta_i = 1)$ and $Y_i(t) = I(Z_i \quad t)$. Then $N_i(t)$ can be uniquely decomposed into the sum of its cumulative intensity process $\Lambda_i(t)$ and a local square integrable martingale M_i:

$$N_i(t)=\Lambda_i(t)+M_i(t)$$

for $i = 1,\ldots, n$ and $t \in [0,1]$, where $\Lambda_i(t)=\int_0^t\lambda_i(s)\,ds$. Note that we consider the interval $[0, 1]$ for simplicity. The argument can be easily extended to the interval $[0, \infty]$. Let $W = (W_1, \ldots, W_n)^T$ denote the covariate processes such that $W_i(t)=(V_i^T(t), X_i^T(t))^T$ denote the counting process histories up to time t. The intensity process, in the manner of Cox (1972), can be written as

$$\lambda_i(t)=Y_i(t)\exp\{\psi(W_i(t);\theta_0)\}\lambda_0(t),$$

for $i = 1,\ldots,n$, where $\psi(\cdot; \theta_0)$ is a function with known form and θ_0 is a vector of unknown parameters. In our context,

$$\psi(W_i(t);\theta)=V_i(t)^T\alpha+\sum_{j=1}^{k}\gamma_j\tilde{B}_j(X_i(t)^T\beta),$$

where $V_i(t) = (V_{i1}, \ldots, V_{iq})$, $X_i(t) = (X_{i1}, \ldots, X_{ip})$ and $\theta = (\beta, \gamma, \alpha)$.

As in Goldstein & Langholz (1992), define $\tilde{R(t)}$, the risk set at time $t+$, by

$$\tilde{R}(t)=\{j:Y_j(t+)=1\}.$$

Let $T_1, T_2, \ldots$ be the ordered collection of event times of the Y_i and N_i processes. Let $\tilde{R}_k = \tilde{R}(T_k)$. If $i \in \tilde{R}_{k-1}$, let $P_{m,i}(\tilde{R}_{k-1})$ be the set of all subsets of $\tilde{R}_{k-1}$ of size M that include i. Let $R_{k,i}$ be independently and uniformly chosen from $P_{m,i}(\tilde{R}_{k-1})$. If $i \notin \tilde{R}_{k-1}$, we let $R_{k,i}$ be the empty set. Set $\eta_{ij}(0) = 0$. We note that the preceding construction makes

$$\eta_{ij}(t) = \sum_{k \geq 1} I(j \in \overline{R}_{k,i}) I(T_{k-1} < t \leq T_k)$$

predictable (Goldstein & Langholz, 1992).

The log partial likelihood can be written as

$$pl(\theta, t) = \sum_{i=1}^{n} \int_0^t \left(\psi(W_i(s); \theta) - \log \left[\sum_{j=1}^{n} \eta_{ij}(s) \exp\{\psi(W_j(s); \theta)\} \right] \right) dN_i(s).$$

For a column vector a, denote $|a| = (a^T a)^{1/2}$, $\| a \| = \sup_i a_i$ and $a^{\otimes 2}$ for the matrix aa^T. For a matrix C, denote $\| C \| = \sup_{i,j} C_{i,j}$. For a function $\psi(x; \theta)$, let $\dot{\psi}(x; \theta)$ and $\ddot{\psi}(x; \theta)$ denote the gradient and Hessian of $\psi(\cdot)$ with respect to θ. Define

$$S^{(0)}(\theta, t) = \frac{1}{n} \sum_{j=1}^{n} \eta_{ij}(t) \exp\{\psi(W_j(t); \theta)\},$$

$$S^{(1)}(\theta, t) = \frac{1}{n} \sum_{j=1}^{n} \dot{\psi}(W_j(t); \theta) \eta_{ij}(t) \exp\{\psi(W_j(t); \theta)\},$$

$$S^{(2)}(\theta, t) = \frac{1}{n} \sum_{j=1}^{n} \{\ddot{\psi}(W_j(t); \theta) + \dot{\psi}(W_j(t); \theta) \dot{\psi}^T(W_j(t); \theta)\} \eta_{ij}(t) \exp\{\psi(W_j(t); \theta)\}.$$

Assume the following conditions:

C.1 $\int_0^1 \lambda_0(t) dt < \infty$. The at risk probability $b(t) = P(Y(t) = 1) > 0$ for every $t \in [0,1]$.

C.2 The functions $\psi(W_i,(t); \theta_0)$, $\dot{\psi}(W_i(t); \theta_0)$ and $\ddot{\psi}(W_i(t); \theta_0)$, for $i = 1, \ldots, n, t \in [0,1]$ are locally bounded.

C.3 (Lindeberg condition)

$$\int_0^1 \frac{1}{n} \sum_{i=1}^{n} \left\{ \dot{\psi}(W_i(s); \theta_0)_j - \left(\frac{S^{(1)}(\theta_0, s)}{S^{(0)}(\theta_0, s)} \right)_j \right\}^2 Y_i(s) \exp\{\psi(W_i(s); \theta_0)\}$$

$$I\left\{|\dot{\psi}(W_i(s);\theta_0)_j - \left(\frac{S^{(1)}(\theta_0,s)}{S^{(0)}(\theta_0,s)}\right)_j| > \sqrt{n}\varepsilon\right\}\lambda_0(s)ds \to_p 0$$

for any $\varepsilon > 0$ and $j = 1, 2,\ldots,q + p$.

C.4 Let $p_j = \dfrac{\exp\{\psi(W_j(s);\theta)\}}{\sum_{i \in U}\exp\{\psi(W_i(s);\theta)\}}$, where $U = \{1, \ldots, m\}$. Define

$$\Gamma(\theta,t) = E\left[\frac{1}{m}\int_0^t b(s)\sum_{j \in U}\exp\{\psi(W_j(s);\theta)\}\left\{\sum_{j \in U}\dot{\psi}(W_j(s);\theta)\dot{\psi}^T(W_j(s);\theta)p_j - \left\{\sum_{j \in U}\dot{\psi}(W_j(s);\theta)p_j\right\}^{\otimes 2}\right\}\lambda_0(s)ds \Big| Y_U(t) = 1\right]$$

where $Y_U(t) = \prod_{i \in U} Y_i(t)$. The matrix $\Gamma = \Gamma(\theta_0,1)$ is positive definite.

We first state a lemma. The proof is straightforward based on Lemma 1 in Goldstein & Langholz (1992) and thus omitted. For simplicity, write $Y = Y(s)$, $W = W(s)$ and $b = b(s)$ for $s \in [0,1]$.

LEMMA 1. Let $\rho \in \{1, 2\}$, (Y_i, W_i), $i \in \{1,\ldots, n\}$ be independent copies of (Y, W) with $W \in \mathbf{R}^{q+p}$, $Y \in \{0,1\}$ and $b = P(Y = 1) > 0$. Let $R = \{j: Y_j = 1\}$, $P = \{T \subset R, |T| = M\}$ and $P_i = \{T \in P: i \in T\}$. With $T \in P$, let

$$\omega(T) = \left[\frac{\sum_{j \in T}\dot{\psi}(W_j;\theta)\exp\{\psi(W_j;\theta)\}}{\sum_{j \in T}\exp\{\psi(W_j;\theta)\}}\right]^{\otimes \rho},$$

$A_i = \exp\{\psi(W_i;\theta_0)\}$ and $S_n = \dfrac{1}{n}\sum_{i=1}^n \omega(\tilde{R}_i)Y_i A_i$. Assume that conditions C.1 - C.4 hold. Then

$$S_n \to_p r,$$

where $r = bE\left[\omega(U)\dfrac{1}{m}\sum_{j \in U} A_j \Big| Y_U = 1\right]$.

Proposition 1

If the nonparametric function $\psi(\cdot)$ with fixed knots satisfy conditions C.1 - C.4, there exists a sequence of roots $\hat{\theta_n}$ of the partial likelihood equation such that $\hat{\theta_n} \to_p \theta_0$.

Proof. Let

$$Z_n(\theta, t) = \frac{1}{n}[\log L(\theta, t) - \log L(\theta_0, t)] = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \left[\psi(W_i(s);\theta) - \psi(W_i(s);\theta_0) - \log\frac{S^{(0)}(\theta, s)}{S^{(0)}(\theta_0, s)}\right] dN_i(s)$$

and

$$A_n(\theta, t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \left[\psi(W_i(s);\theta) - \psi(W_i(s);\theta_0) - \log\frac{S^{(0)}(\theta, s)}{S^{(0)}(\theta_0, s)}\right] \lambda_i(s)\,ds.$$

Then the process

$$Z_n(\theta, t) - A_n(\theta, t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \left[\psi(W_i(s);\theta) - \psi(W_i(s);\theta_0) - \log\frac{S^{(0)}(\theta, s)}{S^{(0)}(\theta_0, s)}\right] dM_i(s)$$

is a locally square integrable martingale for each θ, with predictable variation process at t given by

$$\langle Z_n(\theta, \cdot) - A_n(\theta, \cdot), Z_n(\theta, \cdot) - A_n(\theta, \cdot)\rangle(t) = \frac{1}{n^2}\int_0^t\sum_{i=1}^{n}\left\{\psi(W_i(s);\theta) - \psi(W_i(s);\theta_0) - \log\frac{S^{(0)}(\theta, s)}{S^{(0)}(\theta_0, s)}\right\}^2 \lambda_i(s)\,ds.$$

By conditions C.1, C.2 and Cauchy-Schwarz inequality, it is easy to show that $\langle Z_n(\theta, \cdot) - A_n(\theta, \cdot), Z_n(\theta, \cdot) - A_n(\theta, \cdot)\rangle \to_p 0$. By the Lenglart inequality

$$\lim_{n\to\infty}\{Z_n(\theta, t) - A_n(\theta, t)\} = 0$$

in probability for all $\theta \in \Theta$. Since Θ is a compact set, we have that $Z_n(\theta, t)$ converges to $A_n(\theta, t)$ in probability uniformly for $\theta \in \Theta$. Next,

$$\frac{\partial A_n(\theta, 1)}{\partial\theta} = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \left[\dot\psi(W_j(s);\theta) - \frac{S^{(1)}(\theta, s)}{S^{(0)}(\theta, s)}\right] \times Y_i(s)\exp\{\psi(W_i(s);\theta)\}\lambda_0(s)\,ds.$$

By Lemma 1,

$$\lim_{n\to\infty}\frac{\partial A_n(\theta, 1)}{\partial\theta} = \int_0^t [r(\theta_0, s) - r(\theta, s)]\lambda_0(s)\,ds$$

in probability, where

$$r(\theta, s) = b(s) E\left[\frac{\sum_{j \in U} \dot{\psi}(W_j; \theta) \exp\{\psi(W_j; \theta)\}}{\sum_{j \in U} \exp\{\psi(W_j; \theta)\}} \frac{1}{m} \sum_{j \in U} \exp\{\psi(W_j(s); \theta_0)\} \Big| Y_U(s) = 1\right].$$

So $A_n(\theta, 1)$ converges to a function with first derivative 0 at $\theta = \theta_0$.

With Γ in condition C.4, the second derivative of the limit of $A_n(\theta, 1)$ equals to minus a nonnegative definite matrix for every θ and at θ_0 equals to $-\Gamma$. Thus, θ_0 is a local maximizer of $A_n(\theta, 1)$. Therefore, the maximizer $\hat{\theta}$ of $Z_n(\theta, 1)$ converges to θ_0 in probability.

Proposition 2

(*Asymptotic normality of $\hat{\theta}$*) If the nonparametric function $\psi(\cdot)$ with fixed knots satisfy conditions C.1 - C.4, there exists a sequence of roots $\hat{\theta}_n$ of the partial likelihood equation such that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_D N(0, \Gamma^{-1}).$$

Proof. Consider the score process

$$U(\theta, t) = \sum_{i=1}^{n} \int_0^t \left[\dot{\psi}(W_i(s); \theta) - \frac{S^{(1)}(\theta, s)}{S^{(0)}(\theta, s)}\right] dN_i(s)$$

and the information process

$$H(\theta, t) = \int_0^t \sum_{i=1}^{n} \left[-\ddot{\psi}(W_i(s); \theta) + \frac{S^{(2)}(\theta, s)}{S^{(0)}(\theta, s)} - \left\{\frac{S^{(1)}(\theta, s)}{S^{(0)}(\theta, s)}\right\}^{\otimes 2}\right] dN_i(s).$$

Let

$$F(\theta, s) = \frac{S^{(2)}(\theta, s)}{S^{(0)}(\theta, s)} - \left\{\frac{S^{(1)}(\theta, s)}{S^{(0)}(\theta, s)}\right\}^{\otimes 2}.$$

By the Taylor expansion,

$$U(\theta, 1) - U(\theta_0, 1) = -H(\theta^*, 1)(\theta - \theta_0),$$

where θ^* lies between θ and θ_0. Substitute $\hat{\theta}_n$ for θ,

$$n^{-1}H(\theta^*, 1)\sqrt{n}(\hat{\theta}_n - \theta_0) = n^{-\frac{1}{2}}U(\theta_0, 1).$$

We will show that

$$n^{-1}H(\theta^*, 1) \to_p \Gamma$$

and

$$n^{-\frac{1}{2}}U(\theta_0, 1) \to_D N(0, \Gamma).$$

Define

$$C(\theta, t) = \int_0^t \frac{1}{n}\sum_{i=1}^n \{F(\theta, s) - \ddot{\psi}(W_i(s); \theta)\}\lambda_i(s)\,ds.$$

Then $\dfrac{1}{n}H(\theta, t) - C(\theta, t)$ is a local square integrable martingale with predictable variation process at t given by

$$\left\langle \frac{1}{n}H(\theta, \cdot) - C(\theta, \cdot), \frac{1}{n}H(\theta, \cdot) - C(\theta, \cdot)\right\rangle(t) = \frac{1}{n^2}\int_0^t \sum_{i=1}^n \left\{F(\theta, s) - \ddot{\psi}(W_i(s); \theta)\right\}^2 \lambda_i(s)\,ds$$

By conditions C.1 and C.2, $\left\langle \dfrac{1}{n}H(\theta, \cdot) - C(\theta, \cdot), \dfrac{1}{n}H(\theta, \cdot) - C(\theta, \cdot)\right\rangle(t) \to_p 0.$

By Lenglart's inequality

$$\lim_{n\to\infty}\left\{\frac{1}{n}H(\theta, 1) - C(\theta, 1)\right\} \to_p 0.$$

Using Lemma 1, $C(\theta, 1) \to_p \Gamma$. Moreover, consistency of $\hat{\theta}$ implies that $\theta^* \to_p \theta_0$. By conditions C.1 and C.2, we have $C(\theta^*, 1) \to_p \Gamma$. Hence,

$$\frac{1}{n}H(\theta^*, 1) \to_p \Gamma.$$

Next we show that $n^{-\frac{1}{2}}U(\theta_0, 1) \to_D N(0, \Gamma)$. Let $G_i(s) = \dfrac{s^{(1)}(\theta, s)}{s^{(0)}(\theta, s)}$ and

$$G_{(s)} = \frac{\sum_{j=1}^{n} Y_j(s) \dot{\psi}(W_j(s);\theta) \exp\{\psi(W_j(s);\theta)\}}{\sum_{j=1}^{n} Y_j(s) \exp\{\psi(W_j(s);\theta)\}}.$$

Then

$$U(\theta_0, t) = \sum_{i=1}^{n} \int_0^t [\dot{\psi}(W_j(s);\theta) - G_i(s)] dN_i(s)$$

$$= \sum_{i=1}^{n} \left[\int_0^t \{\dot{\psi}(W_j(s);\theta) - G(s)\} dN_i(s) + \int_0^t \{G(s) - G_i(s)\} dN_i(s) \right]$$

$$= \sum_{i=1}^{n} \left[\int_0^t \{\dot{\psi}(W_j(s);\theta) - G(s)\} dM_i(s) + \int_0^t \{G(s) - G_i(s)\} d\Lambda_i(s) \right].$$

The term $D_t = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t \{\dot{\psi}(W_j(s);\theta) - G(s)\} dM_i(s)$ is a stochastic integral of a predictable process against a martingale and is thus a martingale. By Lemma 1,

$$\langle D, D \rangle(t) = \int_0^t \frac{1}{n} \sum_{i=1}^{n} \{\dot{\psi}(W_j(s);\theta) - G(s)\}^{\otimes 2} \lambda_i(s) ds \to_p \Gamma(\theta_0, t)$$

as $n \to \infty$. By condition C.3 and the martingale central limit theorem (Andersen & Gill, 1982),

$$D_1 \to_D N(0, \Gamma).$$

Moreover, as in Goldstein & Langholz (1992),

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t \{G(s) - G_i(s)\} d\Lambda_i(s) \to_p 0.$$

Therefore, $n^{-\frac{1}{2}} U(\theta_0, 1) \to_D N(0, \Gamma)$. The proof is complete.

References

Andersen PK, Gill RD. Cox regression-model for countingprocesses - a large sample study. Annals of Statistics. 1982; 10:1100–1120.

Carroll RJ, Fan JQ, Gijbels I, Wand MP. Generalized partially linear single-index models. Journal of the American Statistical Association. 1997; 92:477–489.

Chen D, Hall P, Mueller HG. Single and multiple index functional regression models with nonparametric link. Annals of Statistics. 2011; 39:1720–1747.

Chiou JM, Muller HG. Quasi-likelihood regression with multiple indices and smooth link and variance functions. Scandinavian Journal of Statistics. 2004; 31:367–386.

Clendenen TV, Lundin E, Zeleniuch-Jacquotte A, Koenig KL, Berrino F, Lukanova A, Lokshin AE, Idahl A, Ohlson N, Hallmans G, Krogh V, Sieri S, Muti P, Marrangoni A, Nolen BM, Liu M, Shore RE, Arslan AA. Circulating inflammation markers and risk of epithelial ovarian cancer. Cancer Epidemiology Biomarkers and Prevention. 2011; 20:799–810.

Cook RD, Bing L. Dimension reduction for conditional mean in regression. Annals of Statistics. 2002; 30:455–474.

Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society Series B-Statistical Methodology. 1972; 34:187–220.

De Boor, C. A practical guide to splines. Springer; New York: 1978.

Friedman JH, Stuetzle W. Projection pursuit regression. Journal of the American Statistical Association. 1981; 76:817–823.

Goldstein L, Langholz B. Asymptotic theory for nested case-control sampling in the cox regression-model. Annals of Statistics. 1992; 20:1903–1928.

Hardle W, Hall P, Ichimura H. Optimal smoothing in single-index models. Annals of Statistics. 1993; 21:157–178.

Hardle W, Stoker TM. Investigating smooth multiple-regression by the method of average derivatives. Journal of the American Statistical Association. 1989; 84:986–995.

Huang JHZ, Liu L. Polynomial spline estimation and inference of proportional hazards regression models with flexible relative risk form. Biometrics. 2006; 62:793–802. [PubMed: 16984322]

Ichimura H. Semiparametric least-squares (sls) and weighted sls estimation of single-index models. Journal of Econometrics. 1993; 58:71–120.

Langholz, B. Encyclopedia of Bio statistics. Vol. 1. Wiley; New York: 2005. Case control study, nested; p. 646-655.

Langholz B, Borgan O. Estimation of absolute risk from nested case-control data. Biometrics. 1997; 53:767–774. [PubMed: 9192463]

Li J, Zhang R. Partially varying coefficient single index proportional hazards regression models. Computational Statistics and Data Analysis. 2011; 55:389–400.

Lu XW, Chen GM, Singh RS, Song PXK. A class of partially linear single-index survival models. Canadian Journal of Statistics-Revue Canadienne De Statistique. 2006; 34:97–112.

Oakes D. Survival times - aspects of partial likelihood. International Statistical Review. 1981; 49:235–252.

Stoker TM. Consistent estimation of scaled coefficients. Econometrica. 1986; 54:1461–1481.

Sun J, Kopciuk KA, Lu X. Polynomial spline estimation of partially linear single-index proportional hazards regression models. Computational Statistics and Data Analysis. 2008; 53:176–188.

Thomas, DC. Addendum to "methods of cohort analysis - appraisal by application to asbestos mining". In: Liddell, FDK.; McDonald, JC.; Thomas, DC., editors. Journal of the Royal Statistical Society A. Vol. 140. 1979. p. 469-491.

Wang W. Proportional hazards regression models with unknown link function and time-dependent covariates. Statistica Sinica. 2004; 14:885–905.

Xia YC, Tong H, Li WK. On extended partially linear single-index models. Biometrika. 1999; 86:831–842.

Xia YC, Tong H, Li WK, Zhu LX. An adaptive estimation of dimension reduction space. Journal of the Royal Statistical Society Series B-Statistical Methodology. 2002; 64:363–388.

Yin XR, Cook RD. Dimension reduction for the conditional kth moment in regression. Journal of the Royal Statistical Society Series B-Statistical Methodology. 2002; 64:159–175.

Yu Y, Ruppert D. Penalized spline estimation for partially linear single-index models. Journal of the American Statistical Association. 2002; 97:1042–1054.
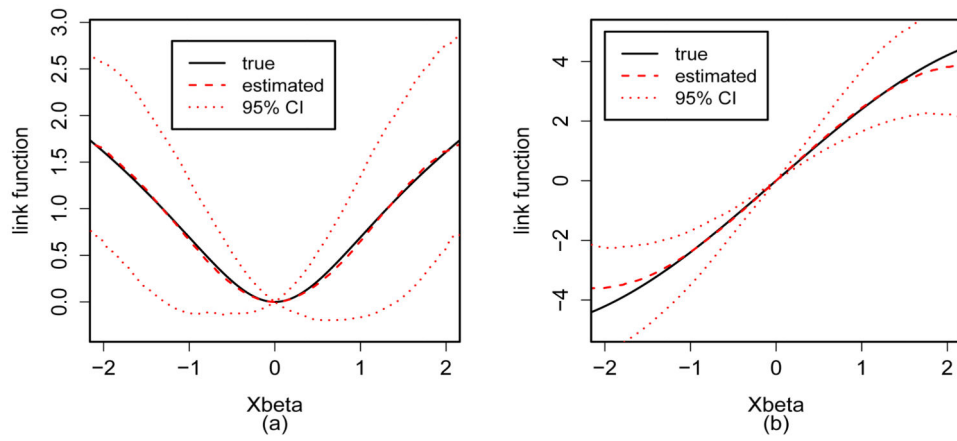
Figure 1.
The median of estimated link function and 95% pointwise Monte Carlo intervals for (a): log quadratic model; (b): sine model. Cohort size = 1000, censoring rate = 80%.
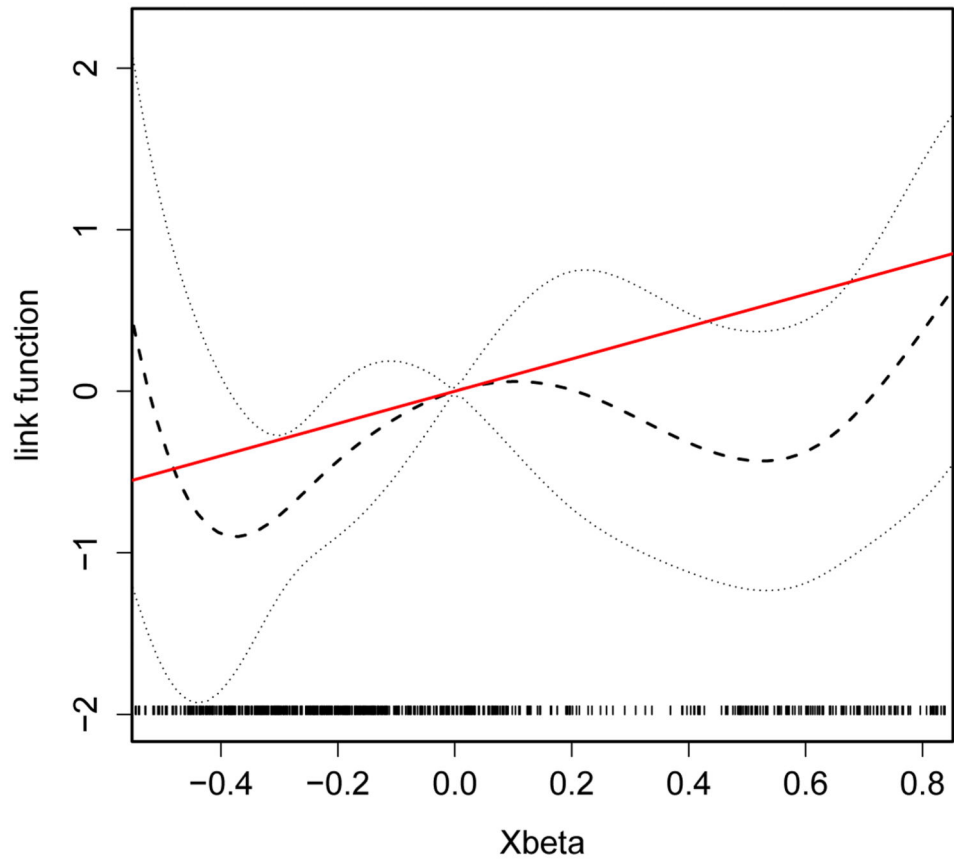
Figure 2.
The estimated link function (dashed) and 95% pointwise confidence interval (dotted) for the ovarian cancer data. The solid curve is the identity link function (*y* = *x*). The distribution of the covariates in the *X* part multiplied by estimated *β* was indicated in the bottom.

Table 1

Log quadratic model: results of parameter estimates using the true model, proposed model and Cox PH model

|	α_1				α_2				angle $(\beta, \hat{\beta})$	
	bias	SD	SEE	CP	bias	SD	SEE	CP	mean	SD
N = 1000, censoring rate = 80%, 200 cases										
true model	0.047	0.153	0.143	0.950	-0.091	0.271	0.251	0.944	6.074	3.517
pISI	0.077	0.163	0.153	0.940	-0.154	0.296	0.272	0.944	7.300	5.430
Cox PH	-0.216	0.114	0.112	0.490	0.425	0.188	0.187	0.374	65.206	35.733
N = 2000, censoring rate = 80%, 400 cases										
true model	0.015	0.097	0.097	0.950	-0.034	0.172	0.170	0.952	3.934	2.154
pISI	0.025	0.100	0.101	0.952	-0.055	0.179	0.179	0.950	4.350	2.426
Cox PH	-0.232	0.076	0.077	0.160	0.460	0.122	0.128	0.078	66.075	35.201
N = 1000, censoring rate = 90%, 100 cases										
true model	0.082	0.237	0.211	0.936	-0.181	0.432	0.374	0.924	9.417	6.723
pISI	0.153	0.282	0.243	0.952	-0.327	0.502	0.438	0.928	18.288	22.626
Cox PH	-0.196	0.161	0.165	0.722	0.369	0.285	0.279	0.654	64.235	34.882
N = 2000, censoring rate = 90%, 200 cases										
true model	0.038	0.154	0.141	0.948	-0.075	0.267	0.246	0.958	6.203	3.750
pISI	0.074	0.163	0.153	0.944	-0.146	0.274	0.271	0.946	7.653	7.039
Cox PH	-0.217	0.111	0.112	0.496	0.435	0.180	0.186	0.354	64.183	34.367

bias: empirical bias; SD: sample standard deviation; SEE: standard error estimate; CP: empirical coverage probability of 95% confidence interval; N: cohort size; pISI: partially linear single index model.

Table 2

Sine curve model: results of parameter estimates using the true model, proposed model and Cox PH model

	α₁				α₂				angle ($\beta_0'\hat{\beta}_0$)	
	bias	SD	SEE	CP	bias	SD	SEE	CP	mean	SD
N = 1000, censoring rate = 80%, 200 cases										
true model	0.005	0.104	0.097	0.934	-0.001	0.125	0.123	0.942	2.563	1.418
pISI	0.088	0.191	0.168	0.938	-0.177	0.326	0.297	0.940	2.748	1.465
Cox PH	-0.304	0.132	0.108	0.252	0.617	0.192	0.172	0.132	4.356	2.633
N = 2000, censoring rate = 80%, 400 cases										
true model	0.002	0.068	0.068	0.948	-0.006	0.088	0.086	0.952	1.755	0.915
pISI	0.039	0.116	0.110	0.940	-0.080	0.201	0.194	0.954	1.848	0.957
Cox PH	-0.322	0.076	0.074	0.024	0.644	0.124	0.118	0.006	3.043	1.658
N = 1000, censoring rate = 90%, 100 cases										
true model	-0.003	0.146	0.141	0.938	-0.007	0.178	0.180	0.960	3.961	2.206
pISI	0.214	0.309	0.286	0.932	-0.446	0.553	0.518	0.944	4.114	2.376
Cox PH	-0.278	0.180	0.161	0.502	0.552	0.302	0.260	0.394	6.035	3.520
N = 2000, censoring rate = 90%, 200 cases										
true model	0.002	0.103	0.097	0.930	-0.002	0.121	0.124	0.962	2.627	1.368
pISI	0.084	0.182	0.167	0.944	-0.168	0.324	0.295	0.938	2.702	1.415
Cox PH	-0.310	0.120	0.108	0.232	0.618	0.184	0.173	0.116	4.266	2.405

bias: empirical bias; SD: sample standard deviation; SEE: standard error estimate; CP: empirical coverage probability of 95% confidence interval; N: cohort size; pISI: partially linear single index model.

Table 3

Linear model: results of parameter estimates using Cox PH model and the proposed model

	α_1					α_2					angle $(\beta, \hat{\beta}_0)$	
	bias	SD	SEE	CP	RE	bias	SD	SEE	CP	RE	mean	SD
	N = 1000, censoring rate = 80%, 200 cases											
Cox PH	0.034	0.154	0.150	0.962	NA	-0.072	0.274	0.263	0.952	NA	4.780	2.650
pISI	0.069	0.168	0.157	0.956	0.907	-0.144	0.298	0.278	0.954	0.898	5.150	2.820
	N = 2000, censoring rate = 80%, 400 cases											
Cox PH	0.019	0.101	0.103	0.958	NA	-0.041	0.179	0.181	0.956	NA	3.213	1.665
pISI	0.034	0.105	0.105	0.952	0.958	-0.072	0.186	0.185	0.956	0.955	3.320	1.700

bias: empirical bias; SD: sample standard deviation; SEE: standard error estimate; CP: empirical coverage probability of 95% confidence interval; RE: relative efficiency; N: cohort size; pISI: partially linear single index model.

Table 4

Sensitivity analysis: results of parameter estimates under model misspecification

Analysis a: angle between β and $\hat{\beta}$ when the true model is the single index model

	log quardratic link				sine link			
	N = 1000		N = 2000		N = 1000		N = 2000	
	mean	SD	mean	SD	mean	SD	mean	SD
SI	7.371	4.174	4.548	1.909	3.441	1.543	2.238	1.071
pISI	14.827	7.493	8.469	3.945	4.559	1.787	3.059	1.305
Cox PH	71.073	32.413	71.473	32.310	5.354	2.466	3.967	1.693

Analysis b: the true model is the partially linear single index model, scenarios S1–S4

	α_1				α_2				angle $(\beta, \hat{\beta})$	
	bias	SD	SEE	CP	bias	SD	SEE	CP	mean	SD
S1: include a noise variable in X										
pISI	0.087	0.150	0.155	0.965	-0.187	0.274	0.277	0.955	8.110	5.174
Cox PH	-0.209	0.114	0.113	0.495	0.411	0.181	0.190	0.435	81.930	44.423
S2: omit a covariate in X										
pISI	0.079	0.185	0.152	0.905	-0.155	0.315	0.271	0.925	11.257	8.135
Cox PH	-0.192	0.125	0.113	0.530	0.393	0.194	0.190	0.415	86.079	38.615
S3: misspecify membership of covariates, linear to nonlinear										
pISI	0.049	0.160	0.150	0.930	-0.116	0.312	0.267	0.935	7.816	6.031
Cox PH	-0.243	0.108	0.109	0.385	0.465	0.179	0.182	0.260	90.100	44.880
S4: misspecify membership of covariates, nonlinear to linear										
pISI	0.070	0.154	0.152	0.970	-0.140	0.272	0.270	0.950	9.668	5.006
Cox PH	-0.233	0.104	0.110	0.425	0.464	0.159	0.182	0.270	89.795	37.962

N: cohort size; bias: empirical bias; SD: sample standard deviation; SEE: standard error estimate; CP: empirical coverage probability of 95% confidence interval; SI: single index model; pISI: partially linear single index model.

Table 5

Results of the ovarian cancer NCC study

	Cox PH		plSI	
	Est. (se)	*p*	Est. (se)	*p*
IL-6_q2	0.175 (0.243)	0.471	0.192 (0.249)	0.441
IL-6.q3	0.272 (0.241)	0.258	0.215 (0.246)	0.382
IL-6.q4	0.478 (0.238)	0.045	0.448 (0.244)	0.066
BMI	-0.026 (0.022)	0.241	-0.006 (0.003)	0.081
Age at menarche	0.035 (0.061)	0.564	0.037 (0.017)	0.032
Ever been pregnant	-0.248 (0.237)	0.294	-0.149 (0.069)	0.032
Ever use of OC	-0.277 (0.212)	0.192	-0.102 (0.042)	0.015
Ever use of HRT	0.468 (0.220)	0.033	0.571 (0.023)	< 0.001

IL-6_q2: the second quartile of IL-6; IL-6_q3: the third quartile of IL-6; IL-6_q4: the fourth quartile of IL-6.