Cox Regression Model with Time-Varying Coefficients in Nested Case-Control Studies

Journal:	*Biostatistics*
Manuscript ID:	Draft
Manuscript Type:	Original Article
Date Submitted by the Author:	
Complete List of Authors:	Liu, Mengling; New York University Lu, Wenbin; North Carolina State University, Statistics Zeleniuch-Jacquotte, Anne; New York University Shore, Roy; Radiation Effects Research Foundation
Keywords:	Kernel estimation, Martingale, Nested case-control study, Proportional hazards model, Risk-set sampling, Time-varying coefficient

Biostatistics (2009), **0**, 0, *pp.* 1–28
doi:10.1093/biostatistics/

Cox Regression Model with Time-Varying Coefficients in Nested Case-Control Studies

MENGLING LIU*
Division of Biostatistics, Department of Environmental Medicine,
New York University School of Medicine, New York, NY 10016, USA
mengling.liu@nyu.edu

WENBIN LU
Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA

ROY E. SHORE
Radiation Effects Research Foundation, Hiroshima 732-0815, Japan

ANNE ZELENIUCH-JACQUOTTE
Division of Epidemiology, Department of Environmental Medicine,
New York University School of Medicine, New York, NY 10016, USA

SUMMARY

The nested case-control (NCC) design is a cost-effective sampling method to study the relationship of disease with risk factors in large prospective epidemiologic studies. NCC data are commonly analyzed using Thomas' partial likelihood approach under Cox's proportional hazards model with constant covariate effects. Here, we are interested in studying the potential time-varying effects of risk factors on disease risk in NCC studies and propose an estimation approach based on kernel-weighted Thomas' partial likelihood. We establish the asymptotic properties of the proposed estimator, propose a numerical approach to construct simultaneous confidence bands of time-varying coefficients, and develop a hypothesis testing

*To whom correspondence should be addressed.

procedure to detect time-varying coefficient using the integrated function of the time-varying estimate. The proposed inference procedure is evaluated in simulations and applied to an NCC study of breast cancer in the NYU Women's Health Study.

Keywords: Kernel estimation; Martingale; Nested case-control study; Proportional hazards model; Risk-set sampling; Time-varying coefficient.

1. INTRODUCTION

Epidemiologic cohort studies of rare diseases are usually expensive because a large number of individuals need to be followed-up for a long time in order to obtain an adequate number of cases. Moreover, the assessment of exposures of interest and confounders for the entire cohort may be prohibitively expensive. The nested case-control (NCC) design (Thomas, 1977) has been widely used as a cost-effective alternative to the full-cohort analysis. For example, a NCC study was conducted in the NYU Women's Health Study (NYUWHS) to investigate the association between breast cancer risk and genetic variations in the nucleotide excision repair (NER) pathway (Shore *and others*, 2008). The NYUWHS is a prospective cohort study that enrolled 14,274 healthy women aged 35-65 between 1985 and 1991 at a breast cancer screening center. Women have been followed-up since enrollment for cancer and other health outcomes. The NER mechanism is important to remove DNA damage, and thus genes in the NER pathway may play a role in the development of cancer and other human diseases involving DNA damage and genetic mutations. It would have been very costly to ascertain genetic information for the entire cohort. Thus, for each of the 612 identified invasive breast cancer cases, one control was selected from the case's risk set. Selected single nucleotide polymorphisms (SNPs) in two genes (XPC and ERCC2) in the NER pathway were genotyped for the cases and their matched controls. Covariate information on demographics, smoking status and pregnancy history was obtained from baseline and follow-up questionnaires. The effects of genetic variants in these two genes, environmental exposures that cause DNA damage and their interactions on breast cancer risk were of primary interest.

Cox's proportional hazards (Cox, 1972) model has been commonly used to analyze NCC data with the successful implementation of the partial likelihood technique (Thomas, 1977; Oaks, 1981). Under the assumption of proportional hazards, i.e. the ratio of hazard functions with different covariate values remains constant over time, the expression of Thomas' partial likelihood function is equivalent to the conditional logistic likelihood for matched case-control studies. The theoretical properties of Thomas' maximum partial likelihood estimator have been formally established using the counting process and martingale theory (Goldstein and Langholz, 1992; Borgan *and others*, 1995; among others).

Due to the nature of long-term observation and complexity of the relationship to be explored in large-scale epidemiologic studies, the proportional hazards assumption may easily be violated. Therefore, researchers have extended Cox's model to improve the modeling flexibility and to study temporal covariate effects by allowing coefficients to vary with time. More specifically, the Cox model with time-varying coefficients (Zucker and Karr, 1990; Hastie and Tibshirani, 1993)) assumes

$$\lambda\{t|Z(t)\} = \lambda_0(t)\exp\{a_0^T(t)Z(t)\}, \tag{1.1}$$

where $\lambda_0(t)$ denotes an unspecified baseline hazard function, $Z(t)$ denotes the p covariate processes, and $a_0(t)$ are p processes characterizing the covariates' temporal effects. The estimation and inference of model (1.1) in cohort studies have been studied by many researchers using various techniques: for example, the penalized partial likelihood approach with smoothing splines (Tibshirani and Hastie, 1987; Zucker and Karr, 1990); the sieve maximum partial likelihood approach with histogram sieves (Murphy and Sen, 1991); the integrated Newton-Raphson equation for the cumulative coefficient functions $\int_0^t a_0(u)du$ (Martinussen *and others*, 2002); the kernel-weighted partial likelihood approach (Cai and Sun, 2003; Tian *and others*, 2005). However, the use of model (1.1) in risk-set sampling designs remains limited.

In this article, we are interested in studying the time-varying effects of risk factors on the disease risk using model (1.1) in NCC studies. We propose to estimate the time-varying coefficient functions using a kernel-weighted partial likelihood approach. A major difference between Thomas' partial likelihood function for NCC data and Cox's partial likelihood function for cohort data is that the size of each at-risk set remains finite rather than going to infinity as sample sizes increases. We show that the kernel-weighted local polynomial fitting technique (Fan and Gijbels, 1996) can be well coupled with Thomas' partial likelihood to study the time-varying effects of covariates in NCC studies. Furthermore, pertinent to the inference of time-varying coefficients, we develop numerical approaches to constructing simultaneous confidence bands and hypothesis testing method to identify time-varying coefficient.

The rest of the article is organized as follows. In Section 2, we first introduce notation and the kernel-weighted maximum partial likelihood estimation approach. We then present the large-sample properties of our proposed estimator and furthermore, consider the inference of an extension to incorporate both time-varying and time-invariant coefficients. In Section 3, we present numerical studies that include simulation studies to evaluate the finite-sample performance of our proposed approaches and the analysis of the NYUWHS breast cancer data. We conclude with some remarks in Section 5. All proofs are relegated in the appendix.

2. METHODS

Consider an NCC study embedded in a cohort of size n. Let $\{T_i^*, C_i, Z_i(\cdot), i = 1, \cdots, n\}$ denote n i.i.d. triplets of failure times, censoring times, and p-dimensional covariate processes. Define $T_i =$

$\min(T_i^*, C_i)$, $\delta_i = I(T_i^* \leqslant C_i)$, $N_i(t) = \delta_i I(T_i \leqslant t)$ and $Y_i(t) = I(T_i \geqslant t)$, where $I(\cdot)$ denotes the indicator function throughout. An NCC study identifies cases as subjects of $\delta_i = 1$ and randomly samples $(m-1)$ controls without replacement from the at-risk set at each failure time excluding the failed subject itself. For a given case i, let R_i^* denote the indices of the $(m-1)$ selected controls and define $R_i = R_i^* \cup \{i\}$. The true covariates are then ascertained for all the cases and selected controls. Therefore, the observed data consist of $\{T_i, Z_i, Z_j : \delta_i = 1, j \in R_i^*, i = 1, \cdots, n\}$.

2.1 *Kernel-weighted local partial likelihood estimation*

Consider model (1.1) and assume time-varying coefficients $a_0(\cdot)$ to be smooth functions with continuous first and second derivatives. Locally around a time point t, we approximate $a_0(\cdot)$ by a linearization using the first-order Taylor expansion,

$$a_0(u) \approx a_0(t) + a_0'(t)(u-t).$$

Let $\boldsymbol{\beta}_0(t) = \{a_{01}(t), \ldots, a_{0p}(t), a_{01}'(t), \ldots, a_{0p}'(t)\}^T$, and $\tilde{\boldsymbol{Z}}_i(u, u-t) = (1, u-t)^T \otimes Z_i(u)$ with $\otimes$ denoting the Kronecker product. To estimate $\boldsymbol{\beta}_0$ locally around t we consider the following kernel-weighted partial likelihood function of $2p-$dimensional parameter $\boldsymbol{\beta}$,

$$l_n(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^{n} \int_0^{\tau_0} K_h(u-t) \left[\boldsymbol{\beta}^T \tilde{\boldsymbol{Z}}_i(u, u-t) - \log\Big\{ \sum_{j \in R_i} e^{\boldsymbol{\beta}^T \tilde{\boldsymbol{Z}}_j(u, u-t)} \Big\} \right] dN_i(u), \qquad (2.1)$$

where $K(\cdot)$ is the kernel function and h denotes the bandwidth parameter; $K_h(\cdot) = h^{-1}K(\cdot/h)$; τ_0 is the upper bound of observation times and satisfies $\tau_0 = \inf\{t : \mathrm{pr}(T > t) = 0\}$. We assume that $K(\cdot)$ is a symmetric density function with a bounded support on $[-1, 1]$ and $h \equiv h_n = O(n^{-v})$ for $0 < v < 1$. This kernel function down-weights the contributions from subjects whose event times are far from time point t and the bandwidth parameter h controls the size of local neighborhood. Thus, the local partial likelihood function (2.1) depends only on the case-control sets with case event times in the close vicinity of t.

The score function of local partial likelihood function (2.1) can be easily derived

$$U_n(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^{n} \int_0^{\tau_0} K_h(u-t) \big[\tilde{\boldsymbol{Z}}_i(u, u-t) - E_{R_i, \tilde{\boldsymbol{z}}}(\boldsymbol{\beta}, u, u-t) \big] dN_i(u), \tag{2.2}$$

where for a set R, $S^{(k)}_{R,\tilde{\boldsymbol{z}}}(\boldsymbol{\beta}, u, u-t) = \sum_{j \in R} e^{\boldsymbol{\beta}^T \tilde{\boldsymbol{Z}}_j(u,u-t)} \tilde{\boldsymbol{Z}}_j^{\otimes k}(u, u-t)$ and $E_{R,\tilde{\boldsymbol{z}}}(\boldsymbol{\beta}, u, u-t) = \frac{S^{(1)}_{R,\tilde{\boldsymbol{z}}}(\boldsymbol{\beta},u,u-t)}{S^{(0)}_{R,\tilde{\boldsymbol{z}}}(\boldsymbol{\beta},u,u-t)}$. For a vector a, let $a^{\otimes k} = 1, a$ and aa^T for $k = 0, 1,$ and 2, respectively. Furthermore, the Hessian matrix of (2.1) equals to

$$l''_n(\boldsymbol{\beta}, t) = -n^{-1} \sum_{i=1}^{n} \int_0^{\tau_0} \frac{K_h(u-t)}{S^{(0)}_{R_i,\tilde{\boldsymbol{z}}}(\boldsymbol{\beta}, u, u-t)} \Big[\sum_{j \in R_i} e^{\boldsymbol{\beta}^T \tilde{\boldsymbol{Z}}_j(u,u-t)} \{ \tilde{\boldsymbol{Z}}_j(u, u-t) - E_{R_i,\tilde{\boldsymbol{z}}}(\boldsymbol{\beta}, u, u-t) \}^{\otimes 2} \Big] dN_i(u), \tag{2.3}$$

It is evident that $l''_n(\boldsymbol{\beta}, t)$ is seimnegative definite and thus the concavity of (2.1) assures a unique maximum. Newton-Raphson method or other gradient-based search algorithms can be easily used to find $\hat{\boldsymbol{\beta}}$ that maximizes (2.1). We denote the kernel-weighted maximum partial likelihood estimate of $a_0(t)$ by $\hat{a}(t)$, which are the first p components of $\hat{\boldsymbol{\beta}}$.

2.2 *Asymptotic properties of* $\hat{a}(t)$

Because of the risk-set sampling mechanism of NCC design, the size of at-risk set R_i for any i is always m rather than increasing to infinity as in Cox's partial likelihood function. Thus, we adopt similar theoretical arguments for Thomas' maximum partial likelihood estimator used in Goldstein and Langholz (1992) and consider its kernel-weighted local version by using the kernel polynomial fitting technique (Fan and

Gijbels, 1996). Let $r = \{1, \ldots, m\}$, $Y_r(t) = \prod_{i \in r} Y_i(t)$, and $p_0(t) = \mathrm{pr}\{Y_1(t) = 1\}$. Denote

$$\Sigma_0(t) = E\Big[m^{-1} \sum_{j \in r} e^{a_0^T(t) Z_j(t)} \Big\{Z_j(t) - \frac{\sum_{j \in r} e^{a_0^T(t) Z_j(t)} Z_j(t)}{\sum_{j \in r} e^{a_0^T(t) Z_j(t)}}\Big\}^{\otimes 2} \Big| Y_r(t) = 1\Big].$$

Indeed, $p_0(t)\lambda_0(t)\Sigma_0(t)$ is the local contribution to the asymptotic information matrix of Thomas' partial likelihood function (Goldstein and Langholz, 1992). Here, we state the main asymptotic results of uniform consistency and asymptotic normality of $\hat{a}(t)$. Technical conditions and proofs are given in the appendix. Let $\mu_j = \int s^j K(s) ds$ and $v_j = \int s^j K^2(s) ds$ for $j = 0, 1, 2$.

PROPOSITION 1 (a) Under Conditions A.1-A.4, $\sup_{t \in (0,\tau_0)} |\hat{a}(t) - a(t)| \to_p 0$ as $n \to \infty$. (b) Under Conditions A.1-A.5, for $t \in (0, \tau_0)$,

$$(nh)^{1/2} \Big\{\hat{a}(t) - a_0(t) - \frac{h^2}{2} \mu_2 a''{}_0(t)\Big\} \to_D N\Big\{0, \frac{v_0}{p_0(t)\lambda_0(t)} \Sigma_0^{-1}(t)\Big\},$$

as $n \to \infty$.

2.3 *Point-wise confidence interval and simultaneous confidence band*

From Proposition 1, it is evident that the optimal bandwidth that minimizes the mean squared error (MSE) or mean integrated squared error (MISE) is $h = O(n^{-1/5})$. This theoretical optimal bandwidth, however, leads to an asymptotically biased estimator, i.e. the bias term is $O(1)$. In the present paper, we prefer to use a slightly faster rate for bandwidth $h = n^{-v}$, with $1/5 < v < 1$, to obtain an unbiased estimator. The main reason is to avoid the complications due to bias on constructing point-wise confidence intervals and simultaneous confidence band. We refer the readers to Härdel and Marron (1991) for more discussion of handling bias in constructing the simultaneous confidence band for general nonparametric curve estimation.

The variance of $\hat{a}(t)$ can be consistently estimated by the upper left $p \times p$ submatrix of

$$\hat{\Gamma}(t) = (nh)^{-1} v_0 \big[-l_n''(\boldsymbol{\beta}, t)|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \big]^{-1},$$

where $l_n''(\boldsymbol{\beta}, t)$ is as specified in (2.3). The $(1-\alpha)\%$ point-wise confidence interval for the jth element of $a_0(\cdot)$ at time t can be constructed as

$$\{\hat{a}_j(t) \pm z_{1-\alpha/2} \hat{\Gamma}_{jj}(t)\}. \tag{2.4}$$

To make the inference for the underlying coefficient function over a specific time region, it is more desirable and informative to consider the simultaneous confidence band than the point-wise confidence interval. Although the Bonferroni method can be applied to point-wise intervals with $(1-\alpha/I)\%$ level at I grid points to obtain a simultaneous error control, it can be of low efficiency due to ignoring the positive correlations across grid points and produce inappropriately wide confidence bands. In practice, simultaneous bands are usually estimated using numerical approaches that mimic the original data structure while assessing variability (Hädle and Marron, 1991; Tian *and others*, 2005).

Here, we construct simultaneous confidence bands by approximating the distribution of

$$S_j = \sup_{[t_1, t_2]} w_n(t) |\hat{a}_j(t) - a_{0j}(t)|, \; j = 1, \ldots, p, \tag{2.5}$$

with the resampling method of Lin *and others* (1994), where the weight function $w_n(t)$ can be a data-related positive function that uniformly converges to a deterministic function. More specifically, in the proof of Proposition 1, we show that, if $h = n^{-v}$ and $1/5 < v < 1$, the kernel-weighted score function (2.2) at the true coefficient values is asymptotically equivalent to

$$U_n^*(\boldsymbol{\beta}_0, t) = n^{-1} \sum_{i=1}^{n} \int_0^{\tau_0} K_h(u-t) \big[\tilde{\boldsymbol{Z}}_i(u, u-t) - E_{R_i, \tilde{\boldsymbol{z}}}(\boldsymbol{\beta}_0, u, u-t) \big] dM_i(u), \tag{2.6}$$

where $dM_i(u) = dN_i(u) - Y_i(u)\exp\{a^T(u)Z_i(u)\}\lambda_0(u)$ and consequently, under standard measurability assumptions $M_i(u)$ is a local martingale. Substituting $M_i(t)$ by $N_i(t)G_i$, where G_i's are independent standard normal random variables, we obtain a randomly perturbed version of $U_n^*(\boldsymbol{\beta}_0, t)$, denoted by $\tilde{U}_n^*(\boldsymbol{\beta}_0, t)$. At each specific t, the conditional limiting distribution of $\tilde{U}_n^*(\boldsymbol{\beta}_0, t)$ given the observed data is the same as the unconditional limiting distribution of $U_n^*(\boldsymbol{\beta}_0, t)$ (Lin *and others*, 1994). Therefore, the distribution of S_j can be numerically approximated by its randomly perturbed counterpart

$$\tilde{S}_j = \sup_{[t_1,t_2]} w_n(t) \Big| \big[\{l_n''(\hat{\boldsymbol{\beta}}, t)\}^{-1}\tilde{U}_n^*(\hat{\boldsymbol{\beta}}, t)\big]_j \Big|.$$

Let $c_{1-\alpha}$ denote the sample $100(1-\alpha)$th percentile of a large number of realizations of $\tilde{S}_j$'s, and thus the simultaneous confidence band for $a_{0j}(t)$ over $[t_1, t_2]$ can be constructed as

$$\{\hat{a}_j(t) \pm c_\alpha w_n^{-1}(t)\}. \tag{2.7}$$

2.4 *Inference of mixed Cox model with time-varying and constant coefficients*

When there is indication of possible time-invariant coefficients for certain covariates, one may want to consider a mixed model with time-varying coefficient functions and constant coefficients. Consider a mixed model with the first q components of $a_0(t)$ being constants, i.e. $a_0^T(t) = \{\gamma_0^T, a_{[-q]}^T(t)\}$ where $a_{[q]}$ denotes the first q elements of a vector a and $a_{0[-q]}$ denotes the remaining elements of the vector a without first q elements. Such a mixed model possesses the advantage of the flexibility of time-varying coefficients and can be more efficient for estimating the constant coefficients. Building on the kernel-weighted partial likelihood estimates $\hat{a}(t)$ proposed in Section 2.1, we estimate the constant coefficients γ by

$$\hat{\gamma} = \int_0^{\tau_0} w_\gamma(t)\hat{a}_{[q]}(t)dt, \tag{2.8}$$

where $w_\gamma(t)$ is a weight function converging to a deterministic function and $\int_0^{\tau_0} w_\gamma(t)dt = I_{q\times q}$ as an identity matrix. As suggested in Tian *and others* (2005), a natural choice for this weight function is the standardized inverse covariance matrix of $\hat{a}_{[q]}(t)$, i.e. $w_\gamma(t) = \{\int_0^{\tau_0} J(t)dt\}^{-1}J(t)$ where $J(t)$ is the inverse of the upper left $q\times q$ submatrix of $\hat{\Gamma}(t)$. To make inference about γ_0, by arguments similar to Tian *and others* (2005), we can show that $n^{1/2}(\hat{\gamma}-\gamma_0)$ converges weakly to a mean-zero normal distribution for $h = n^{-v}$ with $1/4 < v < 1/2$. The limiting covariance matrix can be consistently estimated by $\hat{\Gamma}_\gamma = \{\int_0^{\tau_0} J(t)dt\}^{-1}$.

In practice, after we obtain the estimates $\hat{a}(t)$ and their accompanying simultaneous confidence band, we can examine each plot whether the confidence band encloses a horizontal line to check if a constant assumption for this coefficient is possible. One drawback for this procedure is the slow convergence rate of the local estimates and thus its low sensitivity. As shown in the paragraph above, we can consider the cumulative function of the time-varying coefficient estimates to achieve better a convergence rate (Martinussen *and others*, 2002). To check whether the jth component of $a(t)$ is independent of time, we consider the process $T_j(t) = n^{1/2}\int_0^t\{\hat{a}_j(u)-\hat{\gamma}_j\}du$, where $\hat{a}(t)$ and $\hat{\gamma}$ are the local and integrated estimates, respectively. Based on a similar resampling method described in Section 2.3, we can obtain randomly perturbed realizations of $\tilde{T}_j(t)$ to evaluate this hypothesis graphically and numerically.

More specifically, denote $\tilde{\boldsymbol{\beta}}(t)$ as $\hat{\boldsymbol{\beta}}(t)$ with the jth and the $(p+j)$th elements replaced by $(\hat{\gamma}_j, 0)$. When the hypothesis is true, the standardized cumulative difference process $T_j(t)$ converges weakly to a mean-zero Gaussian process (Tian *and others*, 2005). Its asymptotic distribution can be approximated by the conditional distribution of the randomly perturbed processes of

$$\tilde{T}_j(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau_0}\Big(\int_0^t \Big[\big\{-l_n''(\tilde{\boldsymbol{\beta}}(s),s)\big\}^{-1}K_h(u-s)\big\{\tilde{\boldsymbol{Z}}_i(u,u-s) - E_{R_i,\tilde{\boldsymbol{z}}}(\tilde{\boldsymbol{\beta}}(s),u,u-s)\big\}\Big]ds$$
$$-t\big\{-l_n''(\tilde{\boldsymbol{\beta}}(u),u)\big\}^{-1}\big\{\tilde{\boldsymbol{Z}}_i(u,0) - E_{R_i,\tilde{\boldsymbol{z}}}(\tilde{\boldsymbol{\beta}}(u),u,0)\big\}\Big)_j dN_i(u)G_i. \quad (2.9)$$

A large number of resampled realizations of $\tilde{T}_j(t)$ can be plotted against $T_j(t)$ to provide a visual check about the constant hypothesis. Furthermore, a statistical test can be constructed based on $T_j^* = \max_{[0,\tau_0]}|T_j(t)|$, where the critical value can be obtained using the sample quantile of the resampled counterparts $\tilde{T}_j^* = \max_{[0,\tau_0]}|\tilde{T}_j(t)|$.

2.5 *Bandwidth selection*

In practice, we suggest to use a cross-validation method (Hastie *and others*, 2008) to select the bandwidth parameter h. Specifically, we first randomly split the case-control sets into K subsets. Following Tian *and others* (2005), we may use the minus logarithm of the partial likelihood function as a measure of the prediction error. For each h and the kth data part,

$$PE_k(h) = -\sum_{i\in k\text{th part}}\int_0^{\tau_0}\Big[\{\hat{a}^{(k)}(t)\}^T Z_i(t) - \log\big(\sum_{j\in R_i}\exp[\{\hat{a}^{(k)}(t)\}^T Z_j(t)]\big)\Big]dN_i(t), \quad (2.10)$$

where $\hat{a}^{(k)}(t)$ is estimated using $K-1$ data sets excluding the kth parts with bandwidth h. Then the total predication error with h is $PE(h) = \sum_{k=1}^{K} PE_k(h)$ and we can choose the bandwidth parameter h that minimizes $PE(h)$.

3. NUMERICAL STUDIES

3.1 *Simulations*

Simulation studies were carried out to evaluate the performance of the proposed estimation and inference procedures under finite sample sizes. The failure times were generated from a Cox model with a time-varying coefficient given by $\lambda(t|Z) = \lambda_0(t)e^{a_1(t)Z_1+a_2(t)Z_2}$, where the baseline hazard function $\lambda_0(t) = b_0 + b_1t + b_2t\exp\{b_3(t-2)^2\}$ with (b_0, b_1, b_2, b_3) to be specified later, and covariates Z_1 and Z_2 were generated from the Binomial distribution with success probability of 0.5. We considered two types of time-varying function for $a_1(t)$: polynomial, where $a_1(t) = 1 - 0.25(t-2)^2$ and $(b_0, b_1, b_2, d) = (0.0, 0.005, 0.002, 0.25)$; log-sinusoid where $a_1(t) = \log\{1 - 0.8\sin(t/1.2)\}$ and $(b_0, b_1, b_2, d) = (0.02, 0.01, 0, 0)$. We assumed $a_2(t) = \gamma$ to be a constant coefficient of -0.5. In addition, the censoring time was generated as $C = \min(5, C^*)$ where C^* was from a uniform distribution on $(3, 6)$. As nested case-control studies are commonly implemented when the disease incidence rate is low, our simulated incidence rates were all about $10\% - 13\%$. We simulated the nested case-control data from full cohorts sized 1000 and 2000, and selected two controls for each case. We used the Epanechnikov kernel function and performed 1000-run simulations for each setting. The proposed simultaneous confidence band and testing procedure were carried out with 5000 resampling runs.

Table 1 presents the results regarding the simultaneous confidence bands for the time-varying coefficient $a_1(t)$. The bandwidth parameter h was set to change from 0.6 to 1.4 by an increment of 0.2 and the simultaneous bands were constructed over $[1, 4]$. We found that the performance of simultaneous confidence band depended on bandwidth parameter h. Small h led to small biases but large variance and thus higher coverage probabilities; while large h led to reverse results. When the balance was reached between the bias and the variance, for example when $h = 1.2$ for $n = 1000$ and $h = 1$ for $n = 2000$, the simultaneous confidence bands yielded satisfactory coverage probability. The overall performance improved with increased sample size. For example, when $n = 2000$ and $h = 1.0$, the resampling-based simultaneous coverage probabilities matched the nominal level reasonably well and the empirical quantiles of observed $\hat{S}$ based on 1000 simulated data sets were very close to the resampling-based threshold. Furthermore, when $a_1(t)$ was the quadratic polynomial function, the curvature of this polynomial function, i.e. $a_2''(t)$ was -0.5 for all t; while $a_1(t)$ being the log-sinusoid function, its curvature ranged from -0.5 to 2.8. Therefore, by comparing the results between these two different types of time-varying coefficients, we found that the performance of simultaneous confidence band for the polynomial coefficient was less sensitive to the selection of h than those with the log-sinusoid coefficient function. This observation is not surprising because it is usually more difficult to characterize a more variable function and the bias of local linear fitting also depends on the magnitude of $a''(t)$.

Figures 1 and 2 showed the estimated time-varying coefficient curves with sample size of 2000, the 95% point-wise confidence intervals averaged over 1000 simulations, and 95% confidence envelope that was constructed using the point-wise 2.5% and 97.5% quantiles of the estimated curves over 1000 simulations. When h was small, the estimated curves were very close to the true curves but the point-wise confidence intervals were wide. As h increased, the estimated curves showed biases at ”the valley” but the point-wise confidence intervals were narrower.

We summarized the estimation results for the constant coefficient $a_2(t) \equiv \gamma$ in Table 2. We report the estimation bias, the sample standard deviations (SD) of the estimates over 1000 simulations, the average standard errors (SE) using the asymptotic approximation, and the 95% coverage probability (CP) of Wald-type confidence intervals. The biases were all small and the SDs and SEs decreased as the sample size increased. The performance of the integrated estimator $\hat{\gamma}$ for the constant coefficient was stable with respect to h. Such an observation reconfirmed the theoretical result that the integrated estimator has the rate of $n^{-1/2}$ and is independent of h (given the rate of h in certain range). Overall, the SEs and SDs matched well and the 95% CPs were close to the nominal level.

Finally, we assessed the performance of our proposed testing procedure to identify time-varying coefficients. In Table 3, we reported 5% error rates for the tests for $a_1(t)$ and $a_2(t)$, respectively, with sample size of 2000. The empirical thresholds that were estimated by the 95th quantile of the sample test statistics from 1000 runs of simulations, and the average resampling-based threshold that was defined by the mean of 1000 thresholds found by the Monte Carlo resampling approach. For the constant coefficient $a_2(t)$, the proposed testing procedure showed good error rate control; for time-varying coefficient, it yielded reasonable power. As we discussed for Table 1, the coefficient functions that are more variable are easier to be identified, and thus the power was higher for the log-sinusoid coefficient.

3.2 *Breast cancer study in the NYUWHS*

We apply our proposed approaches to the NCC study of breast cancer in the NYUWHS as introduced in Section 1. The details of this NCC study have been reported in Shore *and others* (2008). As an illustration, we estimated the gene-environmental interaction effects of gene XPC in the NER pathway and smoking exposure which leads to DNA damage. Based on the results of Shore *and others* (2008), we assumed a recessive model for gene XPC (1 for XPC-PAT allele +/+; 0 otherwise) and coded Smoking (1 for ever-smoking; 0 never). We considered model (1.1) with four covariates: `Ethnicity` (1 for Caucasian; 0 otherwise); `XPC-Smoking-10` (1 for XPC=1 and Smoking=1; 0 otherwise); `XPC-Smoking-01` (1 for XPC=0 and Smoking=1; 0 otherwise); `XPC-Smoking-11` (1 for XPC=1 and Smoking=1; 0 otherwise). Because there were very few incidences before age 45, we focused our analysis on the NCC data with age of breast cancer diagnosis between 45 and 75.

We fitted the kernel-weighted partial likelihood approach using the Epanechnikov kernel function with bandwidth parameter $h = 10$ selected using 10-fold cross-validation. The estimated coefficient curves, the point-wise confidence intervals and the simultaneous confidence bands were presented in the top panel of Figure 3. We found that the Caucasian group had lower risk in early age than the non-Caucasian group but then the risk increased and peaked around 60 to 65. Both the 90% point-wise confidence interval and the simultaneous confidence bands excluded zero. There was no significantly increased risk in the XPC-PAT +/+ nonsmokers group (`XPC-Smoking-10`) comparing to the reference group of XPC wild-type and nonsmokers as the point-wise and simultaneous confidence intervals all included 0. In the group of XPC wild-type smokers (`XPC-Smoking-01`), the risk seemed to be elevated at early ages then diminished, but the overall effect was not significant according to the simultaneous confidence band. Finally, the risk of the group of XPC-PAT+/+ smokers was uniformly elevated across all ages, and the effect was borderline significant, which agreed with the findings in Shore *and others* (2008)

We next applied the proposed resampling testing procedure to this NCC data to formally check the constant coefficient assumption for each covariate. We plotted $T_j(t)$ and ten resampling realizations in the lower panel of Figure 3, and the p-values based on 5000 resampling runs were also presented in the plots. The constant coefficient assumption for `Ethnicity` variable was rejected at 0.05 level (p-value=0.0497), but this assumption seemed to be reasonable for all other covariates. Therefore, we further estimated the constant parameters using the proposed integrated estimator (2.8). Using XPC wild-type nonsmokers as the reference group, there were no significantly increased risk in `XPC-Smoking-10` group (OR=1.01, 95% CI: 0.57-1.76, p=0.99) or in `XPC-Smoking-01` group (OR=0.94, 95% CI: 0.68-1.30, p=0.70), but an significant increase in `XPC-Smoking-11` group (OR=1.93, 95% CI: 1.15-3.25, p=0.01). Again, these observations confirmed the results of Shore *and others* (2008).

4. DISCUSSION

As a way of enhancing modeling flexibility and providing an alternative and diagnostic tool for Cox's proportional hazards model in NCC studies, we developed inference procedures for the Cox model with time-varying coefficients for NCC data. We investigated the asymptotic properties of the proposed maximum kernel-weighted partial likelihood estimator and provided numerical tools to construct simultaneous confidence bands and to identify time-invariant coefficients. The proposed inference procedures were evaluated through simulations and illustrated with the NCC study of breast cancer in the NYUWHS. Our data analysis further confirmed the findings of interaction between gene XPC in the nucleotide excision repair pathway and the DNA damaging agent of smoking. In addition, the findings that the Caucasian group has lower risk at earlier age and then elevated risk at older age compared to the non-Caucasian group was consistent with the well-described black-to-white ethnic crossover in breast cancer incidence rate (Gray *and others*, 1980; Joslyn *and others*, 2005; Anderson *and others*, 2008) and demonstrated the issue of race disparity in the disease risk. In summary, the Cox model with time-varying coefficients can elucidate the effect of risk factors on the disease and provide intuitive tool to visualize such effects.

ACKNOWLEDGMENTS

Conflict of Interest: None declared.

REFERENCES

ANDERSON, W. F., ROSENBERG, P. S., MENASHE, I., MITANI, A., AND PFEIFFER, R. M. (2008). Age-Related Crossover in Breast Cancer Incidence Rates between Black and White Ethnic Groups. *Journal of National Cancer*

Institute **100**, 1804-1814.

BORGAN, O., GOLDSTEIN, L., AND LANGHOLZ, B. (1995). Methods for the Analysis of Sampled Cohort Data in the Cox Proportional Hazards Model. *Annals of Statistics* **23**, 1749–1778.

CAI, Z. W., AND SUN, Y. Q. (2003). Local Linear Estimation for Time-Dependent Coefficients in Cox's Regression Models. *Scandinavian Journal of Statistics* **30**, 93–111.

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

FAN, J. AND GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall: London.

GOLDSTEIN, L., AND LANGHOLZ, B. (1992). Asymptotic Theory for Nested Case-Control Sampling in the Cox Regression-Model, *Annals of Statistics* **20**, 1903–1928.

GRAY, G. E., HENDERSON, B. E., AND PIKE, M. C. (1980). Changing Ratio of Breast Cancer Incidence Rates with Age of Black Females Compared with White Females in the United States. *Journal of National Cancer Institute* **64**, 461-463.

HÄRDLE, W. AND MARRON, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Annals of Statistics* **19**, 778–796.

HASTIE, T., TIBSHIRANI, R. AND FRIEDMAN, J. (2008). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer-Verlag, New York.

JOSLYN, S. A., FOOTE, M. L., NASSERI, K., COUGHLIN, S. S., AND HOWE, H. L. (2005). Racial and Ethnic Disparities in Breast Cancer Rates by Age: Naaccr Breast Cancer Project. *Breast Cancer Research and Treatment* **92**, 97-105.

LEHR, S., AND SCHEMPER, M. (2007). Parsimonious Analysis of Time-Dependent Effects in the Cox Model. *Statistics in Medicine* **26**, 2686–2698.

LIN, D. Y., FLEMING, T. R., AND WEI, J. L. (1994). Confidence Bands for Survival Curves under the Proportional Hazards Model. *Biometrika* **81**, 73-81.

MARTINUSSEN, T., SCHEIKE, T. H., AND SKOVGAARD, I. M. (2002). Efficient Estimation of Fixed and Time-

Varying Covariate Effects in Multiplicative Intensity Models. *Scandinavian Journal of Statistics* **29**, 57–74.

MURPHY, S. A., AND SEN, P. K. (1991). Time-Dependent Coefficients in a Cox-Type Regression-Model. *Stochastic Processes and Their Applications* **39**, 153–180.

OAKES, D. (1981). Survival times - aspects of partial likelihood. *International Statistical Review* **49**, 235–252.

SHORE, R. E., ZELENIUCH-JACQUOTTE, A., CURRIE, D., MOHRENWEISER, H., AFANASYEVA, Y.,; KOENIG, K. L., ARSLAN, A. A., TONIOLO, P., WIRGIN, I. (2008). Polymorphisms in Xpc and Ercc2 Genes, Smoking and Breast Cancer Risk. *International Journal of Cancer* **122**, 2101–2105.

THOMAS, D. C. (1977). Addendum to "Methods of Cohort Analysis - Appraisal by Application to Asbestos Mining" by Liddell, F. D. K., McDonald, J. C., and Thomas, D. C., J. R. *Journal of the Royal Statistical Society Series A* **140**, 469–491.

TIAN, L., ZUCKER, D., AND WEI, L. J. (2005). On the Cox Model with Time-Varying Regression Coefficients. *Journal of the American Statistical Association* **100**, 172–183.

TIBSHIRANI, R. AND HASTIE, T. (1987). Local likelihood estimation. *Journal of the American Statistical Association* **82**, 559–568.

ZUCKER, D. M., AND KARR, A. F. (1990). Nonparametric Survival Analysis with Time-Dependent Covariate Effects - a Penalized Partial Likelihood Approach. *Annals of Statistics* **18**, 329–353.

APPENDIX

We first introduce additional notation to simplify the exposition. Define $\boldsymbol{\alpha} = H(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ and $\tilde{\boldsymbol{U}}_i(u, u - t) = H^{-1}\tilde{\boldsymbol{Z}}_i(u, u-t)$, where $H = diag(1, \ldots, 1, h, \ldots, h)$. Therefore, proving $\hat{\boldsymbol{\alpha}} \to_p 0$ suffices the proof of consistency of $\hat{\boldsymbol{\beta}}$. Furthermore, for $k = 0, 1, 2$ and an arbitray set R, denote

$$S^{*(k)}_{R,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha}, u, u-t) = m^{-1}\sum_{j\in R} e^{\boldsymbol{\beta}_0^T \tilde{\boldsymbol{Z}}_j(u,u-t) + \boldsymbol{\alpha}^T \tilde{\boldsymbol{U}}_j(u,u-t)} \tilde{\boldsymbol{U}}_j^{\otimes k}(u, u-t),$$
$$S^{*(k)}_{R,\tilde{\boldsymbol{z}}}(\boldsymbol{\alpha}, u, s) = m^{-1}\sum_{j\in R} e^{a^T(u)Z_j(u) + \boldsymbol{\alpha}^T \tilde{\boldsymbol{Z}}_j(u,s)} \tilde{\boldsymbol{Z}}_j^{\otimes k}(u, s),$$

$E^*_{R,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha}, u, u-t) = \frac{S^{*(1)}_{R,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha},u,u-t)}{S^{*(0)}_{R,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha},u,u-t)}$, and $E^*_{R,\tilde{\boldsymbol{z}}}(\boldsymbol{\alpha}, u, s) = \frac{S^{*(1)}_{R,\tilde{\boldsymbol{z}}}(\boldsymbol{\alpha},u,s)}{S^{*(0)}_{R,\tilde{\boldsymbol{z}}}(\boldsymbol{\alpha},u,s)}$. We also assume the following technical conditions to hold:

A.1 The time-varying coefficient function $a(t)$ have continuous and bounded second derivative for $t \in (0, \tau_0)$.

A.2 The sequence $h \to 0$ and $nh \to \infty$ as $n \to \infty$, and nh^5 is bounded.

A.3 The hazard function $\lambda\{t|Z(t)\} < \infty$ for $t \in (0, \tau_0)$ and the at-risk probability $p_0(t)$ is continuous with $p_0(\tau_0) > 0$.

A.4 There exists a random variable V of $p\times 1$ dimension, such that $EV^{\otimes 2} < \infty$, $E[Y(t)e^{|a(t)|'V}V^{\otimes 2}] < \infty$, and $\sup_u |Z(u)| \leqslant V$.

A.5 $\Sigma_0(t)$ is nonsingular $t \in (0, \tau_0)$.

We first state a lemma to be used throughout. Its proof is a straightforward based on Lemma 1 in Goldstein and Langholz (1992) and thus omitted.

LEMMA 4.1 Assume that Conditions 1-4 hold. For $k = 1, 2$ and $l = 0, 1$, let

$$G_n(u, u-t) = n^{-1} \sum_{i=1}^{n} \frac{S^{*(k)}_{R_i,\tilde{\boldsymbol{u}}}(\alpha, u, u-t)}{S^{*(0)}_{R_i,\tilde{\boldsymbol{u}}}(\alpha, u, u-t)} Y_i(u) e^{a^T(u) Z_i(u)} \tilde{\boldsymbol{U}}_i^{\otimes l}(u, u-t),$$

$$G(u, u-t) = p_0(u) E\left\{ \frac{S^{*(k)}_{r,\tilde{\boldsymbol{u}}}(\alpha, t, u-t)}{S^{*(0)}_{r,\tilde{\boldsymbol{u}}}(\alpha, t, u-t)} S^{*(l)}_{r,\tilde{\boldsymbol{u}}}(0, t, u-t) | Y_r(u) = 1 \right\}.$$

Then,

$$\sup_{t \in [0,\tau_0]} |G_n(t) - G(t)| = O_p(n^{-1/2}).$$

Proof of Proposition 1(a).

Let $l_n(\boldsymbol{\alpha}, t)(\tau)$ denote the process of

$$n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)\big[\boldsymbol{\beta}_0^T\tilde{\boldsymbol{Z}}_i(u, u-t) + \boldsymbol{\alpha}^T\tilde{\boldsymbol{U}}_i(u, u-t) - \log\{mS^{(0)}_{R_i,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha}, u, u-t)\}\big]dN_i(u),$$

with respect to τ and note that $l_n(\boldsymbol{\alpha}, t) \equiv l_n(\boldsymbol{\alpha}, t)(\tau_0)$. Then

$$l_n(\boldsymbol{\alpha}, t)(\tau) - l_n(0, t)(\tau) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)\left[\boldsymbol{\alpha}^T\tilde{\boldsymbol{U}}_i(u, u-t) - \log\left\{\frac{S^{(0)}_{R_i,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha}, u, u-t)}{S^{(0)}_{R_i,\tilde{\boldsymbol{u}}}(0, u, u-t)}\right\}\right]dN_i(u),$$

can be written into $A_n(\boldsymbol{\alpha}, t)(\tau) + B_n(\boldsymbol{\alpha}, t)(\tau)$, where

$$A_n(\boldsymbol{\alpha}, t)(\tau) = n^{-1}\int_0^{\tau} K_h(u-t)\left[\boldsymbol{\alpha}^T\tilde{\boldsymbol{U}}_i(u, u-t) - \log\left\{\frac{S^{(0)}_{R_i,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha}, u, u-t)}{S^{(0)}_{R_i,\tilde{\boldsymbol{u}}}(0, u, u-t)}\right\}\right]Y_i(u)e^{a^T(u)Z_i(u)}\lambda_0(u)du,$$

$$B_n(\boldsymbol{\alpha}, t)(\tau) = n^{-1}\int_0^{\tau} K_h(u-t)\left[\boldsymbol{\alpha}^T\tilde{\boldsymbol{U}}_i(u, u-t) - \log\left\{\frac{S^{(0)}_{R_i,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha}, u, u-t)}{S^{(0)}_{R_i,\tilde{\boldsymbol{u}}}(0, u, u-t)}\right\}\right]dM_i(u).$$

Thus, $B_n(\boldsymbol{\alpha}, t)(\tau)$ is a square integrable martingale with quadratic variation at time τ of

$$n^{-2}\sum_{i=1}^{n}\int_0^{\tau} K_h^2(u-t)\left[\boldsymbol{\alpha}^T\tilde{\boldsymbol{U}}_i(u, u-t) - \log\left\{\frac{S^{(0)}_{R_i,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha}, u, u-t)}{S^{(0)}_{R_i,\tilde{\boldsymbol{u}}}(0, u, u-t)}\right\}\right]^{\otimes 2} Y_i(u)e^{a^T(u)Z_i(u)}\lambda_0(u)du. \quad (4.1)$$

By Lemma 1 and Taylor expansion, it is easy to see that (4.1) is is bounded by $O_p((nh)^{-1})$, which is $o_p(1)$, uniformly in $\tau \in (0, \tau_0)$. Next,

$$\frac{\partial A_n(\boldsymbol{\alpha}, t)}{\partial \boldsymbol{\alpha}} = n^{-1}\sum_{i=1}^{n}\int_0^{\tau_0} K_h(u-t)\big\{\tilde{\boldsymbol{U}}_i(u, u-t) - E^*_{R_i,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha}, u, u-t)\big\}Y_i(u)e^{a^T(u)Z_i(u)}\lambda_0(u)du.$$

As shown in Tian *and others* (2005), for a generic function $\omega(\cdot)$ with bounded second derivative over the region of (a, b),

$$\sup_{t\in(a,b)} \left|\int K_h(u-t)\{\omega(u) - \omega(t)\}du\right| \leqslant ch^2. \quad (4.2)$$

The simple calculations and using Lemma 1 and (4.2),

$$\sup_{t\in[9,\tau_0]}\left|\frac{\partial A_n(\boldsymbol{\alpha}, t)(s)}{\partial \boldsymbol{\alpha}} - p_0(t)\lambda_0(t)\int K(s)[Q(0, t, s) - Q(\boldsymbol{\alpha}, t, s)]ds\right| = o_p(1), \quad (4.3)$$

where

$$Q(\boldsymbol{\alpha}, t, s) = E\left\{\frac{S^{*(1)}_{r,\tilde{\boldsymbol{z}}}(\boldsymbol{\alpha}, t, s)}{S^{*(0)}_{r,\tilde{\boldsymbol{z}}}(\boldsymbol{\alpha}, t, s)} S^{*(0)}_{r,\tilde{\boldsymbol{z}}}(0, t, s) \Big| Y_r(t) = 1\right\}.$$

Thus, $A_n(\boldsymbol{\alpha}, t)$ uniformly converges to a function with the first derivative 0 at $\boldsymbol{\alpha} = 0$. By the Lenglart inequality, it implies that

$$\sup_{t\in[0,\tau_0]} \left| l'_n(\boldsymbol{\alpha}, t) - l'_n(0, t) - p_0(t)\lambda_0(t) \int K(s)[Q(0, t, s) - Q(\boldsymbol{\alpha}, t, s)]ds \right| \to_p 0.$$

Following the similar derivation that leads to (4.3), it can be shown that the limit of the second derivative of $A_n(\boldsymbol{\alpha}, t)$ equals to

$$\Gamma(\boldsymbol{\alpha}, t) = -p_0(t)\lambda_0(t) \int K(s) E\Bigg[\frac{S^{*(0)}_{r,\tilde{\boldsymbol{z}}}(\boldsymbol{\alpha}, t, s)}{S^{*(0)}_{r,\tilde{\boldsymbol{z}}}(0, t, s)} \sum_{j\in r} e^{a^T(t)Z_j(t)+\boldsymbol{\alpha}^T\tilde{\boldsymbol{Z}}_j(t,s)} \{\tilde{\boldsymbol{Z}}_j(t, s) - E^*_{r,\tilde{\boldsymbol{z}}}(\boldsymbol{\alpha}, t, s)\}^{\otimes 2} \Big| Y_r(t) = 1\Bigg] ds,$$

which is non-positive definite. It follows from the Arzela-Ascoli theorem and a subsequence argument that

$$\sup_{t\in(0,\tau_0)} |\hat{\boldsymbol{\alpha}}(t)| \to_p 0.$$

$\square$

Proof of Proposition 1(b). We first show that $(nh)^{1/2}\{l'_n(0, t) - b_n(t)\} \to_D N\{0, V(t)\}$ where the bias term $b_n(t)$ and covariance matrix $V(t)$ will be specified below. By martingale representation, $l'_n(0, t)(\tau)$ can be expressed by

$$n^{-1} \sum_{i=1}^{n} \int_0^\tau K_h(u - t)\big\{\tilde{\boldsymbol{U}}_i(u, u - t) - E^*_{R_i,\tilde{\boldsymbol{u}}}(0, u, u - t)\big\} dM_i(u),$$
$$+ n^{-1} \sum_{i=1}^{n} \int_0^\tau K_h(u - t)\big\{\tilde{\boldsymbol{U}}_i(u, u - t) - E^*_{R_i,\tilde{\boldsymbol{u}}}(0, u, u - t)\big\} Y_i(u) \left[e^{a^T(u)Z_i(u)} - e^{\boldsymbol{\beta}_0^T\tilde{\boldsymbol{Z}}_i(u,u-t)}\right] \lambda_0(u)du,$$
$$+ n^{-1} \sum_{i=1}^{n} \int_0^\tau K_h(u - t)\big\{\tilde{\boldsymbol{U}}_i(u, u - t) - E^*_{R_i,\tilde{\boldsymbol{u}}}(0, u, u - t)\big\} Y_i(u) e^{\boldsymbol{\beta}_0^T\tilde{\boldsymbol{Z}}_i(u,u-t)} \lambda_0(u)du \quad (4.4)$$

The first term of (4.4) is a stochastic integral of a predictable process against a martingale and thus itself is a martingale, denoted by $W_n(0,t)(\tau)$. We can use the martingale central limit theory (Andersen and Gill, 1982) to prove the asymptotic normality. The quadratic variation process of $(nh)^{1/2}W_n(0,t)(\tau)$ is

$$V_n(t)(\tau) = n^{-1}h\sum_{i=1}^{n}\int_0^{\tau} K_h^2(u-t)\left\{\tilde{\boldsymbol{U}}_i(u,u-t) - \frac{S^{(1)}_{R_i,\tilde{\boldsymbol{u}}}(0,u,u-t)}{S^{(0)}_{R_i,\tilde{\boldsymbol{u}}}(0,u,u-t)}\right\}^{\otimes 2} Y_i(u)e^{a^T(u)Z_i(u)}\lambda_0(u)du.$$

By Lemma 1 and Taylor expansion, we obtain

$$V_n(t) \to_p V(t) = p_0(t)\lambda_0(t)\{\int \tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}^T K^2(s)ds\} \otimes \Gamma(t),$$

where $\tilde{\boldsymbol{s}} = (1,s)'$. Next, consider the kth component of $(nh)^{1/2}W_n(0,t)(\tau)$ and denote its integrand as

$$D_{ij}(u,u-t) = K_h(u-t)\{\tilde{\boldsymbol{U}}_{ij}(u,u-t) - E_{R_i,\tilde{\boldsymbol{u}},j}(0,u,u-t)\}.$$

We need to check the Lindeberg condition, and by the boundedness of the covariates and the kernel function, we have

$$n^{-1}\sum_{i=1}^{n}\int_0^{\tau} D_{ij}^2(u,u-t)I\{|D_{ij}(u,u-t)| > \sqrt{n/h}\varepsilon\}Y_i(u)e^{a^T(u)Z_i(u)}\lambda_0(u)du \to_p 0,$$

for all $\varepsilon > 0$ because the indicator function becomes 0 as n is large. This implies that

$$(nh)^{1/2}W_n(0,t) \to_D N\{0, V(t)\}. \tag{4.5}$$

Furthermore, by Taylor's expansion,

$$e^{a^T(u)Z_i(u)} - e^{\boldsymbol{\beta}_0^T\tilde{\boldsymbol{Z}}_i(u,u-t)} = \frac{(u-t)^2}{2}e^{a^T(t)Z_i(u)}\{a''(t)\}^T Z_i(u) + o_p((u-t)^2). \tag{4.6}$$

Therefore, we have that the second term of (4.4) is asymptotically equivalent to

$$b_n(t) = \frac{h^2}{2}p_0(t)\lambda_0(t)\{\int \tilde{\boldsymbol{s}}s^2K(s)ds\} \otimes \Gamma(t)a''(t) + o_p(h^2).$$

As shown in Goldstein and Langholz (1992), for every u,

$$n^{-1}\sum_{i=1}^{n}\Big\{\tilde{\boldsymbol{U}}_i(u,u-t)-E^{*}_{R_i,\tilde{\boldsymbol{u}}}(0,u,u-t)\Big\}Y_i(u)e^{\boldsymbol{\beta}_0^T\tilde{\boldsymbol{Z}}_i(u,u-t)}=o_p(n^{-1/2}).$$

Moreover, using Condition A.4, it is easy to strength the results to uniform for $u\in[0,\tau_0]$. Then by (4.2), the third term in (4.4) is of the order of $o_p((nh)^{-1/2})$.

Next, we show that for any $\boldsymbol{\alpha}^*\to 0$, $l_n''(\boldsymbol{\alpha},t)\to_p\Gamma(t)$. Note that

$$l_n''(\boldsymbol{\alpha},t)(\tau)=-n^{-1}\sum_{i=1}^{n}\int_0^{\tau}K_h(u-t)\left[\frac{S^{(2)}_{R_i,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha},u,u-t)}{S^{(0)}_{R_i,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha},u,u-t)}-\left\{\frac{S^{(1)}_{R_i,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha},u,u-t)}{S^{(0)}_{R_i,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha},u,u-t)}\right\}^{\otimes 2}\right]dN_i(u).$$

By Condition A.1 and Taylor expansion, it is easy to see that $||l_n''(\boldsymbol{\alpha}^*,t)-l_n''(0,t)||$ is bounded by $O_p(||\boldsymbol{\alpha}^*||)$. Thus, for any consistent estimator $\boldsymbol{\alpha}^*$, it suffice to show that $l_n''(0,t)\to_p\Gamma(t)$. The Lenglart's inequality implies that

$$l_n''(0,t)\approx-n^{-1}\sum_{i=1}^{n}\int_0^{\tau}K_h(u-t)\left[\frac{S^{(2)}_{R_i,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha},u,u-t)}{S^{(0)}_{R_i,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha},u,u-t)}-\left\{\frac{S^{(1)}_{R_i,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha},u,u-t)}{S^{(0)}_{R_i,\tilde{\boldsymbol{u}}}(\boldsymbol{\alpha},u,u-t)}\right\}^{\otimes 2}\right]Y_i(u)e^{a(u)'Z_i(u)}\lambda_0(u),$$

and by Lemma 1 and Cai and Sun (2003), we have

$$-\,l_n''(0,t)\to_p\Gamma(t)\equiv p_0(t)\lambda_0(t)\Big\{\int\tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}^TK(s)ds\Big\}\otimes\Gamma(t). \tag{4.7}$$

Now, by the definition of $\hat{\boldsymbol{\alpha}}$, we have $l_n'(\hat{\boldsymbol{\alpha}},t)=0$. Following Taylor expansion around 0,

$$0=l_n'(\hat{\boldsymbol{\alpha}},t)=l_n'(0,t)+l_n''(\boldsymbol{\alpha}^*,t)\hat{\boldsymbol{\alpha}},$$

where $\boldsymbol{\alpha}^*$ lies between $\hat{\boldsymbol{\alpha}}$ and 0. By the consistency property of $\hat{\boldsymbol{\alpha}}$, $\boldsymbol{\alpha}^*\to_p 0$. Therefore, by (4.5), (4.7) and Slutsky's theorem, we have

$$(nh)^{1/2}\{\hat{\boldsymbol{\alpha}}-\Gamma^{-1}(t)b_n(t)\}\to_D N\{0,\Gamma^{-1}(t)V(t)\Gamma^{-1}(t)\}.$$

The first p components, $\hat{a}(t)$ thus converges to $N\Big\{\frac{h^2}{2}\mu_2a''(t),\,\frac{v_0}{p_0(t)\lambda_0(t)}\Gamma^{-1}(t)\Big\}$.

Table 1. *Simulation results: simultaneous confidence band for* $\hat{a}(t)$ *over* $[1, 4]$

		$N = 1000$				$N = 2000$			
$a(t)$	h	S.CP†	E.S‡	R.S♯	SD(R.S)§	S.CP†	E.S‡	R.S♯	SD(R.S)§
Polynomial	0.6	98.4	2.524	2.826	0.146	96.4	2.689	2.810	0.089
	0.8	97.9	2.526	2.755	0.126	95.5	2.712	2.740	0.082
	1.0	97.2	2.511	2.698	0.116	94.8	2.709	2.686	0.078
	1.2	96.0	2.529	2.661	0.113	93.8	2.724	2.653	0.078
	1.4	95.3	2.609	2.640	0.112	92.6	2.806	2.635	0.078
Sinusoid	0.6	97.8	2.494	3.078	9.185	96.0	2.728	2.803	0.127
	0.8	97.1	2.514	2.733	0.166	95.2	2.695	2.732	0.107
	1.0	96.5	2.568	2.682	0.150	94.5	2.689	2.673	0.099
	1.2	95.2	2.643	2.645	0.140	92.3	2.866	2.632	0.093
	1.4	92.7	2.739	2.622	0.132	87.3	3.083	2.607	0.089

†S. CP: Simultaneous 95% coverage probability over $[1, 4]$; ‡E.S: Empirical 95% quantile of $\hat{S}$ in 1000 simulations; ♯R.S: average resampling-based thresholds of $\hat{S}$ from 1000 simulations using 5000 resampling runs; §SD(R.S): standard deviation of resampling-based thresholds.

Table 2. *Simulation results: estimation of constant coefficient* γ

		$N = 1000$				$N = 2000$			
$a(t)$	h	Bias	SD†	SE‡	CP♯	Bias	SD†	SE‡	CP♯
Polynomial	0.6	0.002	0.247	0.252	0.964	-0.007	0.168	0.167	0.948
	0.8	-0.002	0.244	0.248	0.967	-0.006	0.167	0.167	0.949
	1.0	-0.003	0.242	0.248	0.967	-0.005	0.166	0.169	0.953
	1.2	-0.002	0.241	0.251	0.966	-0.004	0.165	0.172	0.960
	1.4	-0.002	0.240	0.255	0.971	-0.003	0.164	0.175	0.965
Sinusoid	0.6	0.005	0.262	0.276	0.969	0.001	0.183	0.180	0.951
	0.8	-0.002	0.257	0.270	0.969	-0.001	0.179	0.180	0.955
	1.0	-0.004	0.253	0.270	0.969	-0.002	0.177	0.182	0.959
	1.2	-0.005	0.250	0.272	0.973	-0.002	0.177	0.185	0.959
	1.4	-0.005	0.249	0.276	0.976	-0.002	0.177	0.188	0.963

†SD: Sample standard deviation of the proposed estimates in 1000 simulations; ‡SE: average standard error estimates in 1000 simulations; ♯CP: coverage probability of 95% Wald-type confidence interval.

□

Table 3. *Simulation results: identifying time-varying coefficients with* $N = 2000$

		constant coef.				time-varying coef.			
$a(t)$	h	5% rate	E.T[‡]	R.T[♯]	SD(R.T)[§]	5% rate	E.T[‡]	R.T[♯]	SD(R.T)[§]
Polynomial	0.6	0.065	1.220	1.111	0.160	0.412	2.053	1.123	0.160
	0.8	0.058	1.099	1.038	0.140	0.401	1.797	1.050	0.148
	1.0	0.050	1.084	1.008	0.137	0.363	1.649	1.020	0.147
	1.2	0.042	1.043	0.996	0.136	0.314	1.605	1.008	0.146
	1.4	0.039	1.024	0.993	0.134	0.277	1.562	1.004	0.143
Sinusoid	0.6	0.097	1.614	1.337	0.182	0.897	4.558	1.358	0.182
	0.8	0.068	1.369	1.237	0.157	0.882	3.820	1.263	0.155
	1.0	0.061	1.259	1.171	0.146	0.850	3.403	1.208	0.149
	1.2	0.055	1.171	1.124	0.141	0.798	2.997	1.171	0.149
	1.4	0.048	1.086	1.093	0.139	0.732	2.653	1.146	0.151

[‡]E.T: Empirical 95% quantile of the test statistics in 1000 simulations; [♯]R.T: average resampling-based thresholds of the test statistics from 1000 simulations using 5000 resampling runs; [§]SD(R.T): standard deviation of resampling-based thresholds.

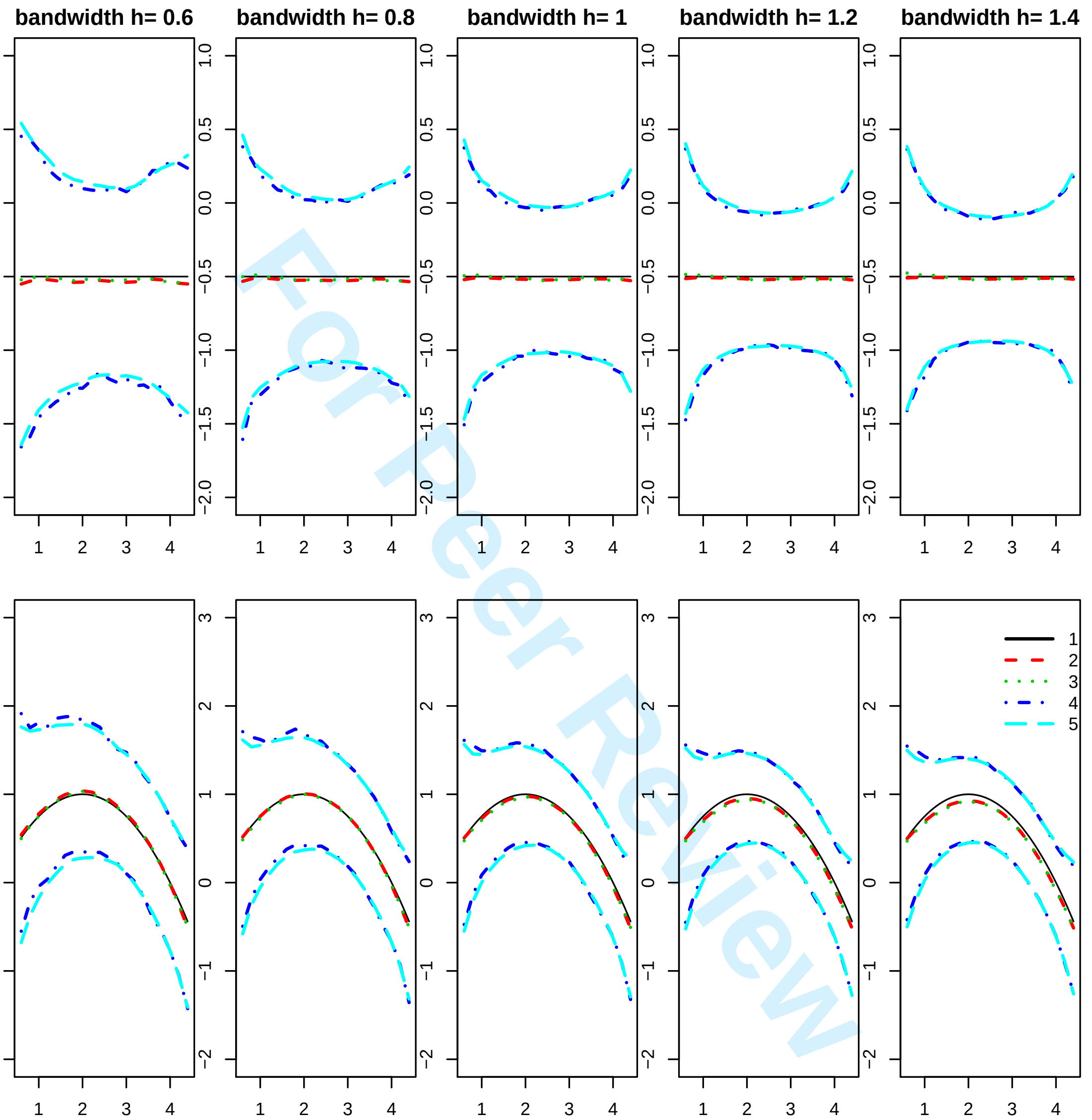

Fig. 1. Estimated coefficient curves with different bandwidth parameters with the polynomial curve. 1: the true underlying coefficient; 2: the point-wise average of 1000 estimated curves; 3: the median of 1000 estimated curves; 4: the average of 95% point-wise confidence intervals; 5: confidence envelop of 2.5% and 97.5% quantiles of 1000 estimated curves.

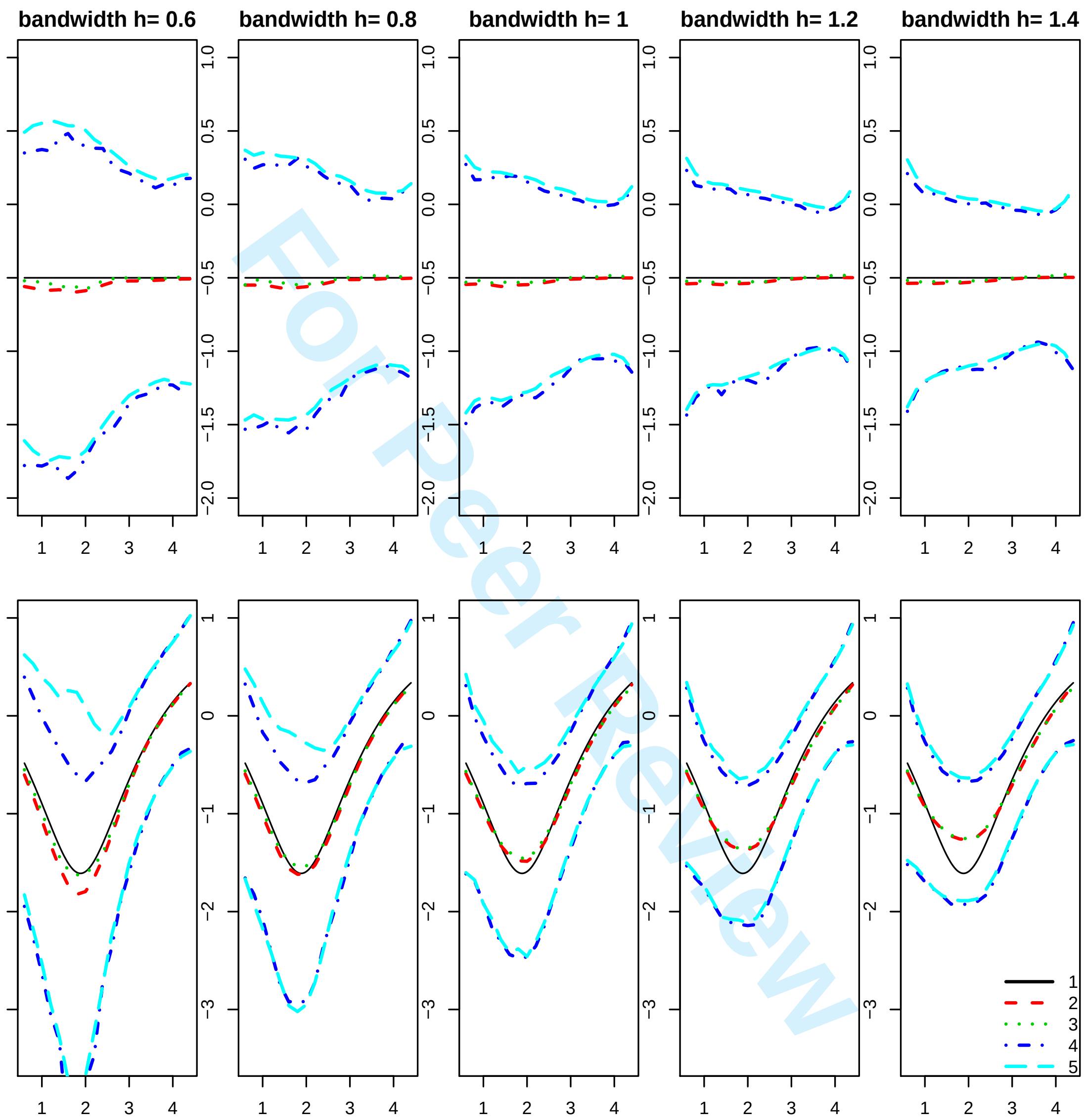

Fig. 2. Estimated coefficient curves with different bandwidth parameters with log-sinusoid curve.1: the true underlying coefficient; 2: the point-wise average of 1000 estimated curves; 3: the median of 1000 estimated curves; 4: the average of 95% point-wise confidence intervals; 5: confidence envelop of 2.5% and 97.5% quantiles of 1000 estimated curves.

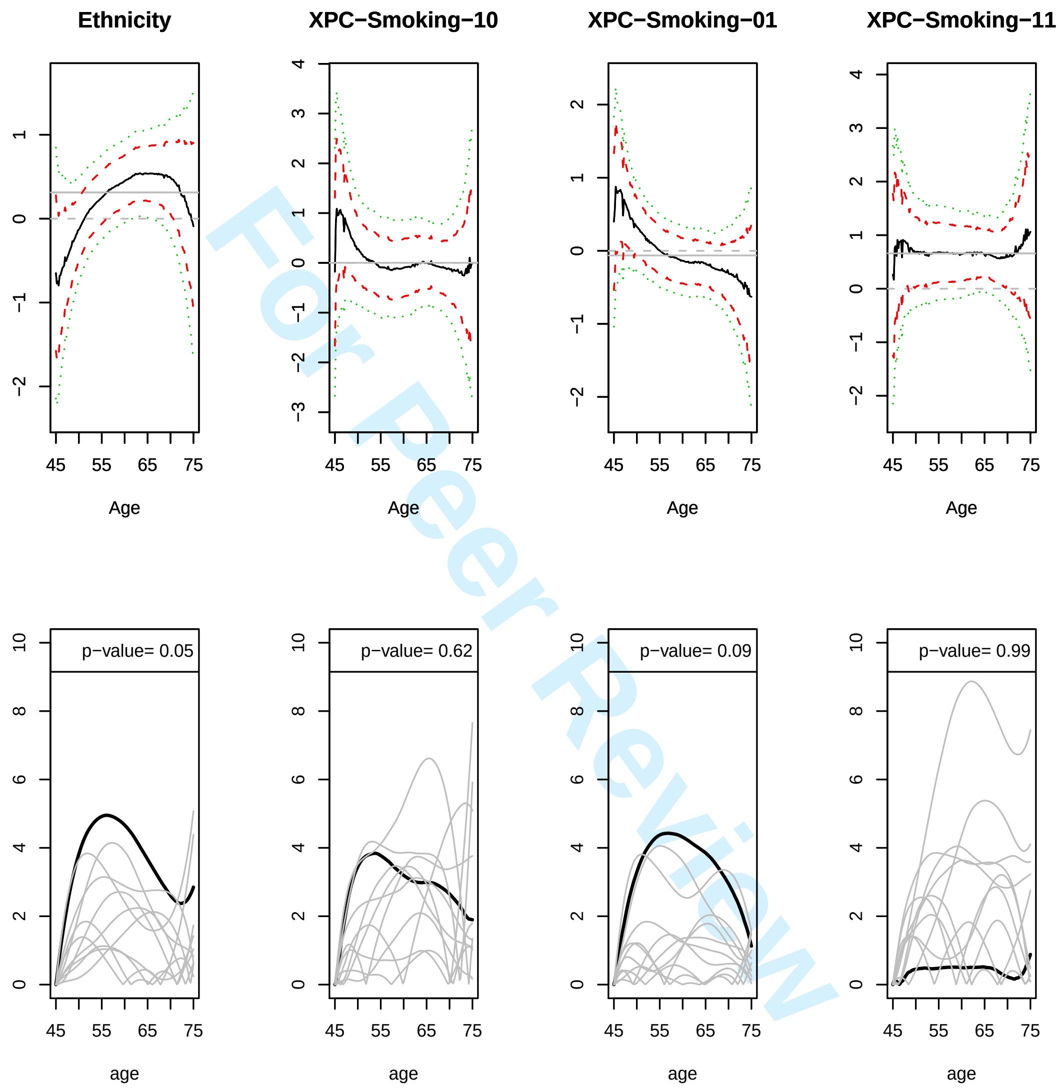

Fig. 3. Time-varying coefficient curves in the NCC study of the NYUWHS. The solid curves are the estimated coefficients; the dash lines are 90% point-wise confidence intervals; the dotted lines are 90% simultaneous confidence bands.