The Canadian Journal of Statistics
Vol. 47, No. 2, 2019, Pages 204–221
La revue canadienne de statistique

Evaluating functional covariate-environment interactions in the Cox regression model

Ling ZHOU[1,3], Haoqi LI[2], Huazhen LIN[3], and Peter X.-K. SONG[1*]

[1]*Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.*
[2]*School of Economics and Statistics, Guangzhou University, Guangzhou, China*
[3]*Center of Statistical Research, Southwestern University of Finance and Economics, Chengdu, China*

Key words and phrases: Generalized likelihood ratio test; global partial likelihood; index coefficient; mixture of toxicants; semiparametric efficiency.

MSC 2010: Primary 62G08; secondary 62P12

Abstract: Children exposed to mixtures of endocrine disrupting compounds such as phthalates are at high risk of experiencing significant friction in their growth and sexual maturation. This article is primarily motivated by a study that aims to assess the toxicants-modified effects of risk factors related to the hazards of early or delayed onset of puberty among children living in Mexico City. To address the hypothesis of potential nonlinear modification of covariate effects, we propose a new Cox regression model with multiple functional covariate-environment interactions, which allows covariate effects to be altered nonlinearly by mixtures of exposed toxicants. This new class of models is rather flexible and includes many existing semiparametric Cox models as special cases. To achieve efficient estimation, we develop the global partial likelihood method of inference, in which we establish key large-sample results, including estimation consistency, asymptotic normality, semiparametric efficiency and the generalized likelihood ratio test for both parameters and nonparametric functions. The proposed methodology is examined via simulation studies and applied to the analysis of the motivating data, where maternal exposures to phthalates during the third trimester of pregnancy are found to be important risk modifiers for the age of attaining the first stage of puberty. *The Canadian Journal of Statistics* 47: 204–221; 2019 © 2019 Statistical Society of Canada

Résumé: Les enfants exposés à des perturbateurs endocriniens comme les phtalates courrent un risque élevé de problèmes relatifs à leur croissance et leur maturation sexuelle. Les auteurs s'intéressent à une étude visant à évaluer l'effet des toxines sur les facteurs de risque liés à une puberté précoce ou retardée chez les enfants vivant à Mexico. Afin d'accommoder l'hypothèse que certaines modifications des effets pourraient s'avérer non linéaires, ils proposent un modèle de régression de Cox avec de nombreuses interactions fonctionnelles entre les covariables et l'environnement, ce qui permet une altération non linéaire de l'effet des covariables suite à une exposition à un mélange de toxines. Cette nouvelle classe de modèles présente une grande flexibilité, au point où plusieurs modèles de Cox semi-paramétriques en sont des cas particuliers. Pour estimer le modèle, les auteurs développent la méthode au maximum de vraisemblance partielle globale dont ils établissent les propriétés clés, notamment la convergence, la normalité asymptotique, l'efficacité semi-paramétrique, et la distribution du test au rapport de vraisemblance généralisé pour les paramètres et pour les fonctions non paramétriques. Les auteurs examinent la méthodologie proposée au moyen d'études de simulation et l'appliquent aux données ayant motivé son développement. Ils constatent que l'exposition aux phtalates lors du troisième trimestre de grossesse modifie substantiellement l'effet des facteurs contribuant à l'âge d'atteinte de la puberté. *La revue canadienne de statistique* 47: 204–221; 2019 © 2019 Société statistique du Canada

---

Additional Supporting Information may be found in the online version of this article at the publisher's website.
* *Author to whom correspondence may be addressed.*
E-mail: pxsong@umich.edu

1. INTRODUCTION

The fundamental hypothesis of ''developmental origins'' in environmental health sciences postulates that environmental exposures during fetal and early postnatal life influence developmental plasticity, thereby altering susceptibility to chronic diseases later on (Perera & Herbstman, 2011; Vaiserman, 2014; Gluckman et al., 2008). A key endeavour has concerned the assessment of the potential developmental and reproductive effects associated with the near ubiquitous environmental exposure to endocrine disrupting compounds (EDCs), such as heavy metals and phthalates, experienced by women and children during sensitive developmental periods (Meeker, 2012; Ma & Song, 2015). Analyzing simultaneous exposures to mixtures of toxic agents is notoriously difficult in the environmental health sciences, and so far only a few statistical methods have been developed that are well-suited to this purpose. This article develops a new Cox regression model that can assess whether or not, and if so, to what extent and in which fashion, mixtures of toxic agents may modify the effects of risk factors related to the timing of pubertal development.

We consider an example concerning growth that involves the covariate concurrent height, which is known to be a strong predictor of the age of pubertal development. We are interested in assessing how the effect of a child's concurrent height on the age of attaining the first stage of puberty may be modified by level of exposure to phthalates. To this end, a statistical analysis needs to address three important questions: (i) Whether or not phthalates modify the effect of a child's concurrent height on the time to reach pubertal first stage? (ii) If so, which phthalates, for example, monobutyl phthalate (MBP), monoethyl phthalate (MEP) or mono-3-carboxypropyl phthalate (MCPP), are responsible for the modification? And (iii) In what form (linear or nonlinear) does the toxicant mixture, that is, a combination of important phthalates, modify the effect of concurrent height on the timing of pubertal development?

To address these questions, we consider a flexible form of functional covariate modification to the usual Cox regression model represented by

$$\lambda(t) = \lambda_0(t) \exp\left\{ \sum_{k=1}^{d} \beta_k(\mathbf{X}'\boldsymbol{\alpha}_k) Z_k \right\}, \tag{1}$$

where $\lambda(t)$ is the hazard at time t, $\lambda_0(t)$ is the baseline hazard and represents the hazard when all of the covariates, $(\mathbf{X}, Z_k, k = 1, \ldots, d)$, are equal to zero. Here $\mathbf{X} = (X_1, \ldots, X_q)'$ is a vector of exposed toxicants, $\boldsymbol{\alpha}_k$ is an unknown q-dimensional vector of parameters, which hereafter we call the loading coefficients, some elements of which may be zero, resulting in different types of mixtures for different covariates Z_k. The covariates $\mathbf{X}$ and $Z_k, k = 1, \ldots, d$ may be correlated. The parameter $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \ldots, \beta_d(\cdot))'$ is a vector of d unknown functions that characterize the forms and extents to which modification of a covariate effect alters with respect to the level of exposure to a combination of multiple toxicants encoded in $\mathbf{X}$.

When a function $\beta_k(\cdot)$ varies in different forms, such as zero, constant, linear or nonlinear, the model specified in Equation (1) allows us to answer if and how groups of phthalates may modify the effects that a child's height and maternal parity exhibit on the age of pubertal development. Recently, Lin, Tan & Li (2016) studied a single-index varying coefficient model with homogeneous loadings across the covariates, namely, $\boldsymbol{\alpha}_k \equiv \boldsymbol{\alpha}, k = 1, \ldots, d$. Compared to the model we have specified in Equation (1), their model aims merely at dimension reduction and does not characterize interaction effects at all; see Ma & Song (2015) for the explanation. Technically, their model may be regarded as a special case of the model we identified in Equation (1); thus, via hypothesis testing, our model may be used to justify their assumption of homogeneous loading coefficients. Furthermore, by using different specifications of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_k$, the model that we identified in Equation (1) covers many other existing semiparametric models. Relevant details can be found in the associated Supplementary Material.

In this article, inference based on the model we have proposed is developed using the idea of efficient global partial likelihood (GPL); see Chen et al. (2010), Chen, Lin & Zhou (2012) and Lin, Tan & Li (2016). GPL has been considered previously for a simpler case with a common argument variable, say m, in functions $\beta_k(m)$, which is the case studied by Lin, Tan & Li (2016). In contrast, the model we have specified in Equation (1) pertains to a varying-coefficient model with different $m_k = \mathbf{X}'\boldsymbol{\alpha}_k$ in $\beta_k(\cdot)$, and represents a much more difficult problem from the technical perspective. That is, the problem studied by Fan, Lin & Zhou (2006) allows for simultaneous estimation of β_k, which avoids the problem of the curse of dimensionality. However, in the model we have proposed different index variables, $m_k = \mathbf{X}'\boldsymbol{\alpha}_k$, lead to a high-dimensional setting, in which extending GPL to the additive Cox model, $\lambda(t) = \lambda_0(t)\exp\left\{\sum_{k=1}^{d}\beta_k(m_k)Z_k\right\}$, is not a trivial challenge. The reason we use GPL is its very attractive efficiency property for nonparametric estimation. In the current literature, such efficiency has not been widely investigated for the additive Cox model, despite many methods that have been proposed for nonparametric function estimation, including spline smoothing (Hastie & Tibshirani, 1990a; 1990b; Sleeper & Harrington, 1990; Huang, 1999), the marginal integration method (Linton, Nielsen, & Van de Geer, 2003; Honda, 2005) and the backfitting method (Mammen, Linton & Nielsen, 1999; Honda, 2005).

To address questions (i)–(iii), we also develop a generalized likelihood ratio (GLR) test for the model we have proposed. In Cox models, methods of hypothesis testing for nonparametric functions have not been well studied. The GLR statistic originally proposed by Fan, Zhang & Zhang (2001) is based on a local linear estimator which, regrettably, is not directly applicable to the GPL setting for testing a function $\beta_k(\cdot)$. In effect, the proposed GLR test represents a useful extension of the classical GLR test investigated in Fan, Zhang & Zhang (2001).

This article is organized as follows. Section 2 introduces GPL estimation; we establish its uniform consistency, asymptotic normality and semiparametric efficiency. Section 3 outlines the theory of inference for the loading parameters and nonparametric functions. Simulation experiments and data examples are described in Sections 4 and 5, respectively. Section 6 consists of some concluding remarks. Some notation and relevant conditions are listed in the Appendix. All technical proofs may be found in the associated Supplementary Material.

2. METHOD OF ESTIMATION

To address the identifiability issue for the model specified in Equation (1), we follow the conditions of the single-index model (Carroll et al., 1997; Wang et al., 2010), which are: (i) $\|\boldsymbol{\alpha}_k\|_2^2 = 1, k = 1, \ldots, d$ and the sign of the first component in each $\boldsymbol{\alpha}_k$ is positive; (ii) covariate $\mathbf{X}$ contains no intercept term; and (iii) for a constant Z_{ik}, say $Z_{i1} \equiv 1$, the corresponding parameter $\beta_1(\cdot)$ is centered and has mean 0. In the following subsections we provide a brief review of local partial likelihood (LPL) and then describe our GPL method.

2.1. Local Partial Likelihood

The independent data replicates are $\{\mathcal{T}_i, \Delta_i, \mathbf{Z}_i, \mathbf{X}_i\}$ from subject $i = 1, \ldots, n$, where $\mathcal{T}_i = \min(T_i, C_i)$, T_i and C_i are failure and censoring times, respectively; Δ_i is an indicator that equals 1 when $\mathcal{T}_i$ is an observed failure time ($\mathcal{T}_i = T_i$) and 0 otherwise ($\mathcal{T}_i = C_i$). The variable $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{id})' \in R^d$ is a set of d covariates of interest, which may be time-dependent. Finally, $\mathbf{X}_i = (X_{i1}, \ldots, X_{iq})' \in R^q$ is a vector of toxicant exposures. We assume that T_i and C_i are conditionally independent, given the covariates $(\mathbf{X}_i, \mathbf{Z}_i)$.

Based on the assumption that T_i follows the model specified in Equation (1), the partial likelihood of the unknown parameters $\boldsymbol{\alpha} = \left(\boldsymbol{\alpha}'_1, \ldots, \boldsymbol{\alpha}'_d\right)'$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)'$ equals

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\frac{\exp\left\{ \sum_{k=1}^{d} \beta_k(m_{ik}) Z_{ik} \right\}}{\sum_{j\in\mathcal{R}(\mathcal{T}_i)} \exp\left\{ \sum_{k=1}^{d} \beta_k(m_{jk}) Z_{jk} \right\}} \right]^{\Delta_i}, \tag{2}$$

where $m_{ik} = \mathbf{X}_i'\boldsymbol{\alpha}_k$ and $\mathcal{R}(t) = \{i : \mathcal{T}_i \geq t\}$ is the set of individuals at risk immediately prior to time t. With $\boldsymbol{\alpha}_k$ being temporarily fixed, we estimate the functions $\beta_k(v_k), k = 1, \ldots, d$ at a point v_k in the range of $\{\mathbf{X}_i'\boldsymbol{\alpha}_k\}_{i=1}^n$ under the assumption that each β_k is continuously first-order differentiable. Thus, for each given value v_k, a Taylor series expansion leads to

$$\beta_k(m_{ik}) \approx \beta_k(v_k) + \dot{\beta}_k(v_k)(m_{ik} - v_k) \stackrel{\text{def}}{=} \zeta_k + \gamma_k(m_{ik} - v_k), \tag{3}$$

where m_{ik} is a certain value in the neighbourhood of v_k. In this article, $\dot{a}(\cdot)$ denotes the first-order derivative of the function $a(\cdot)$. Replacing $\boldsymbol{\beta}(\cdot)$ in Equation (2) by the linear approximation defined in Equation (3), componentwise, gives rise to a set of parameters $\boldsymbol{\eta}_k = (\zeta_k, \gamma_k)'$, $k = 1, \ldots, d$, which are then estimated by maximizing the local partial log-likelihood function

$$\sum_{i=1}^{n} \Delta_i \left(\prod_{k=1}^{d} \mathcal{K}_{ik}(v_k) \right) \left\{ \sum_{k=1}^{d} \overline{\beta}_i(\boldsymbol{\eta}_k, \boldsymbol{\alpha}_k, v_k) Z_{ik} \right.$$
$$\left. - \log \left[\sum_{j\in\mathcal{R}(\mathcal{T}_i)} \left(\prod_{k=1}^{d} \mathcal{K}_{jk}(v_k) \right) \exp\left(\sum_{k=1}^{d} \overline{\beta}_j(\boldsymbol{\eta}_k, \boldsymbol{\alpha}_k, v_k) Z_{jk} \right) \right] \right\}, \tag{4}$$

where $\overline{\beta}_i(\boldsymbol{\eta}_k, \boldsymbol{\alpha}_k, v_k) = M_i(\boldsymbol{\alpha}_k, v_k)'\boldsymbol{\eta}_k$, $M_i(\boldsymbol{\alpha}_k, v_k) = (1, \mathbf{X}_i'\boldsymbol{\alpha}_k - v_k)'$, and the local kernel weighting is allocated by $\mathcal{K}_{ik}(v_k) = \mathcal{K}_{h_k}(m_{ik} - v_k)$, $\mathcal{K}_{h_k}(x) = \mathcal{K}(x/h_k)/h_k$, with $\mathcal{K}$ being a one-dimensional kernel density function and h_k representing the bandwidth.

It is known that the above local method suffers from the curse of dimensionality when $d \geq 2$. The backfitting iterative algorithm (Hastie & Tibshirani, 1990b) is a popular remedy for overcoming this difficulty but establishing its large-sample theory is notoriously challenging (Yu, Park & Mammen, 2008). On the other hand, the existing local scoring backfitting is still based on an LPL approach (Fan, Gijbels & King, 1997), which uses data in a neighbourhood of each fixed value of v_k to estimate $\beta_k(v_k)$. The localization suffers from a potential loss of efficiency because data outside the neighbourhood which may provide information about $\beta_k(\cdot)$ are not used. To overcome this deficiency, here we adopt the method of GPL, rather than LPL, to estimate $\beta_k(\cdot)$.

2.2. Global Partial Likelihood

The GPL approach was first studied by Chen et al. (2010) and Chen, Lin & Zhou (2012) for the simple setting of a single nonparametric function, where there is no curse of dimensionality. In this article we consider a more general GPL method in order to estimate multiple nonparametric functions $\beta_k(\cdot)$, $k = 1, \ldots, d$ with different arguments. Denote a neighbourhood of a target value v_k by $B_n(v_k)$. Let I_{ik} be an indicator that equals 1 if $m_{ik} \in B_n(v_k)$, and 0 otherwise. We consider a first-order expansion of the function $\beta_k(\cdot)$, namely

$$\begin{aligned} \beta_k(m_{ik}) &\approx \{\zeta_k + \gamma_k(\mathbf{X}_i'\boldsymbol{\alpha}_k - v_k)\} I_{ik} + \beta_k(m_{ik})(1 - I_{ik}) \\ &= \overline{\beta}_i(\boldsymbol{\eta}_k, \boldsymbol{\alpha}_k, v_k) I_{ik} + \beta_k(m_{ik})(1 - I_{ik}), \end{aligned}$$

where the second term $\beta_k(m_{ik})$ remains with no approximation if m_{ik} falls outside of $B_n(v_k)$. Moreover, as suggested by Chen et al. (2010), we replace the step function I_{ik} by a smooth function $h_k\mathcal{K}_{h_k}(m_{ik} - v_k)$, thereby obtaining

$$\beta_k(m_{ik}) \approx \overline{\beta}_i(\boldsymbol{\eta}_k, \boldsymbol{\alpha}_k, v_k)h_k\mathcal{K}_{h_k}(m_{ik} - v_k) + \beta_k(m_{ik})\left\{1 - h_k\mathcal{K}_{h_k}(m_{ik} - v_k)\right\}.$$

The right-hand side of this expression represents a global linear approximation, which we will denote by $\beta_i^g(\boldsymbol{\eta}_k, \boldsymbol{\alpha}_k, v_k)$. Substituting $\beta_i^g(\boldsymbol{\eta}_k, \boldsymbol{\alpha}_k, v_k)$ into Equation (2), with fixed $\boldsymbol{\alpha}$, we estimate the parameters $\boldsymbol{\eta}_k, k = 1, \ldots, d$, by maximizing the objective function

$$l_{g,\boldsymbol{\eta}}(\boldsymbol{\eta}; \boldsymbol{\alpha}) = \sum_{i=1}^{n} \Delta_i \left(\sum_{k=1}^{d} \beta_i^g(\boldsymbol{\eta}_k, \boldsymbol{\alpha}_k, v_k)Z_{ik} - \log\left[\sum_{j\in\mathcal{R}(\mathcal{T}_i)} \exp\left\{ \sum_{k=1}^{d} \beta_j^g(\boldsymbol{\eta}_k, \boldsymbol{\alpha}_k, v_k)Z_{jk} \right\}\right]\right). \quad (5)$$

It is worth mentioning that Equation (5) provides the standard full partial likelihood estimator, instead of the LPL estimator that results from maximimizing Equation (4). As a result, the proposed method of estimation based on the use of Equation (5) has some attractive optimality properties due to suitable choices of where and how local approximation is implemented. Estimates that rely on the use of Equation (5) apply the local approximation in the function, whereas the LPL associated with reliance on Equation (4) imposes the local approximation directly on the likelihood.

To estimate $\boldsymbol{\alpha}_k$, denote $\boldsymbol{\eta}_k$ by $\boldsymbol{\eta}_{ik}$ evaluated at $\mathbf{X}_i'\boldsymbol{\alpha}_k$. With $\boldsymbol{\eta}_{ik}$ fixed, we propose to maximize the objective function

$$l_{g,\boldsymbol{\alpha}}(\boldsymbol{\alpha}; \boldsymbol{\eta}) = \sum_{l=1}^{n} \left\{ \hat{f}_{\boldsymbol{\alpha}_k}(m_{lk}) \right\}^{-1} \sum_{i=1}^{n} \Delta_i \left(\sum_{k=1}^{d} \beta_i^g(\boldsymbol{\eta}_{lk}, \boldsymbol{\alpha}_k, \mathbf{X}_l'\boldsymbol{\alpha}_k)Z_{ik} - \log\left[\sum_{j\in\mathcal{R}(\mathcal{T}_i)} \exp\left\{ \sum_{k=1}^{d} \beta_j^g(\boldsymbol{\eta}_{lk}, \boldsymbol{\alpha}_k, \mathbf{X}_l'\boldsymbol{\alpha}_k)Z_{jk} \right\}\right]\right), \quad (6)$$

which corresponds to a summation of Equation (5) over $\mathbf{X}_l'\boldsymbol{\alpha}_k$, $l = 1, \ldots, n$. Here $\hat{f}_{\boldsymbol{\alpha}_k}(v) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{h_k}\left(\mathbf{X}_i'\boldsymbol{\alpha}_k - v\right)$ corresponds to a kernel estimate of the density function $f_{\boldsymbol{\alpha}_k}(v)$ for $\mathbf{X}_i'\boldsymbol{\alpha}_k$.

2.3. Implementation

The objective functions identified in Equations (5) and (6) are not explicitly solvable since the true $\beta_k(\cdot), k = 1, \ldots, d$ are unknown. Instead, this maximization problem may be solved by alternately updating $\boldsymbol{\eta}_{ik} = (\zeta_{ik}, \gamma_{ik})' = \left(\beta_k(m_{ik}), \dot{\beta}_k(m_{ik})\right)'$ and $\boldsymbol{\alpha}_k, k = 1, \ldots, d$. The following constitutes the steps in this algorithm.

Step 0. Choose suitable initial values $\boldsymbol{\eta}_{ik}^{(0)}$ and $\boldsymbol{\alpha}_k^{(0)}$ such that each $\boldsymbol{\alpha}_k^{(0)}$ satisfies $\|\boldsymbol{\alpha}_k^{(0)}\| = 1$ and its first element is positive, $k = 1, \ldots, d$, $i = 1, \ldots, n$. For example, the loading coefficients of the principal components may be used as initial values for $\boldsymbol{\alpha}_k$. Then, with these fixed values for $\boldsymbol{\alpha}_k$, a spline smoothing technique may be used to obtain initial values for the functions $\beta_k(\cdot)$. Our numerical results show that such choices for the initial values work well; see Section 4 for details.

Part 1 of Step s. Find solutions for $\boldsymbol{\eta}_{lk}$ to the score equations

$$\sum_{i=1}^{n} \Delta_i \left[\varpi_{il,k}^{(s-1)} \mathcal{K}_{il,k}^{(s-1)} Z_{ik} - \left\{ \sum_{j \in \mathcal{R}(\mathcal{T}_i)} \exp\left(\sum_{r=1}^{d} \zeta_{jr}^{(s-1)} Z_{jr} \right) \right\}^{-1} \sum_{j \in \mathcal{R}(\mathcal{T}_i)} \varpi_{jl,k}^{(s-1)} \times \mathcal{K}_{jl,k}^{(s-1)} Z_{jk} \exp\left(\beta_j^g(\boldsymbol{\eta}_{lk}, \boldsymbol{\alpha}_k^{(s-1)}, m_{lk}^{(s-1)}) Z_{jk} + \sum_{r \neq k} \zeta_{jr}^{(s-1)} Z_{jr} \right) \right] = \mathbf{0},$$

where $\varpi_{il,k}^{(s-1)} = \left(1, m_{ik}^{(s-1)} - m_{lk}^{(s-1)}\right)'$, $\mathcal{K}_{il,k}^{(s-1)} = \mathcal{K}_{h_k}(m_{ik}^{(s-1)} - m_{lk}^{(s-1)})$, $m_{ik}^{(s-1)} = \mathbf{X}_i' \boldsymbol{\alpha}_k^{(s-1)}$ and $\beta_j^g\left(\boldsymbol{\eta}_{lk}, \boldsymbol{\alpha}_k^{(s-1)}, m_{lk}^{(s-1)}\right) = \boldsymbol{\eta}_{lk}' \varpi_{jl,k}^{(s-1)} h_k \mathcal{K}_{jl,k}^{(s-1)} + \zeta_{jk}^{(s-1)}\left(1 - h_k \mathcal{K}_{jl,k}^{(s-1)}\right)$. Denote these solutions by $\hat{\boldsymbol{\eta}}_{lk}^{(s)} = \left(\hat{\zeta}_{lk}^{(s)}, \hat{\gamma}_{lk}^{(s)}\right)'$ for $l = 1, \ldots, n$ and $k = 1, \ldots, d$.

Part 2 of Step s. Update $\boldsymbol{\alpha}_k$ by solving the partial score equation

$$\sum_{l=1}^{n} \left\{ \hat{f}_{\boldsymbol{\alpha}_k}(m_{lk}^{(s-1)}) \right\}^{-1} \sum_{i=1}^{n} \Delta_i \Bigg(\gamma_{lk}^{(s)} \tilde{\mathbf{X}}_{il} \mathcal{K}_{il,k}^{(s-1)} Z_{ik} - \left\{ \sum_{j \in \mathcal{R}(\mathcal{T}_i)} \exp\left(\sum_{r=1}^{d} \zeta_{jr}^{(s-1)} Z_{jr} \right) \right\}^{-1} \left[\sum_{j \in \mathcal{R}(\mathcal{T}_i)} \gamma_{lk}^{(s)} \tilde{\mathbf{X}}_{jl} \mathcal{K}_{jl,k}^{(s-1)} Z_{jk} \times \exp\left\{ \beta_j^g(\boldsymbol{\eta}_{lk}^{(s)}, \boldsymbol{\alpha}_k, \mathbf{X}_l' \boldsymbol{\alpha}_k) Z_{jk} + \sum_{r \neq k} \zeta_{jr}^{(s)} Z_{jr} \right\} \right] \Bigg) = \mathbf{0},$$

where

$$\beta_j^g(\boldsymbol{\eta}_{lk}^{(s)}, \boldsymbol{\alpha}_k, \mathbf{X}_l' \boldsymbol{\alpha}_k) = \boldsymbol{\eta}_{lk}^{(s)\prime} \varpi_{jl}(\boldsymbol{\alpha}_k) h_k \mathcal{K}_{jl,k}^{(s-1)} + \zeta_{jk}^{(s)} \left(1 - h_k \mathcal{K}_{jl,k}^{(s-1)}\right),$$

$\varpi_{jl}(\boldsymbol{\alpha}_k) = \left(1, \tilde{\mathbf{X}}_{jl}' \boldsymbol{\alpha}_k\right)'$, and $\tilde{\mathbf{X}}_{ij} = \mathbf{X}_i - \mathbf{X}_j$. Denote the solution by $\boldsymbol{\alpha}_k^{(s)}$. Then set $\boldsymbol{\alpha}_k^{(s)} = \boldsymbol{\alpha}_k^{(s)} / \|\boldsymbol{\alpha}_k^{(s)}\|$ with the first element of $\boldsymbol{\alpha}_k^{(s)}$ positive for $k = 1, \ldots, d$.

Part 3. Repeat Step s until convergence is achieved and then collect the output.

As part of this implementation, the bandwidth $h_k, k = 1, 2, \ldots, d$ must be selected. We adopt the K-fold cross-validation procedure (Efron & Tibshirani, 1993; Tian, Zucker & Wei, 2005). In particular, we use the adaptive bandwidth selection method of Brockmann, Gasser & Herrmann (1993), where the bandwidth, $h_Q(v)$, for each target point, v say, is defined in such a way that Q percent of all the data points is used in the analysis. Following Cai, Fan & Li (2000), we choose a value of Q that minimizes the prediction error

$$PE(Q) = \int_0^{\tau} \left[N_i(t) - \hat{E}_Q\{N_i(t)\} \right]^2 d \left\{ \sum_{k=1}^{n} N_k(t) \right\}, \tag{7}$$

where $\hat{E}_Q\{N_i(t)\} = \int_0^t Y_i(u) \exp\{\sum_{k=1}^d \hat{\beta}_{k,Q}(\hat{\boldsymbol{\alpha}}_{k,Q}' \mathbf{X}_i) Z_{ik}\} d\hat{\Lambda}_{0,Q}(u)$ is an estimate of the expected number of events up to time t for a fixed Q value, and $N_i(t) = I(T_i \le t, \Delta_i = 1)$ is the counting process for the total number of observed responses.

2.4. Large-Sample Properties

We now establish uniform consistency and asymptotic normality of the GPL estimators derived by maximizing the objective functions identified in Equations (5) and (6); we denote these by $\hat{\boldsymbol{\alpha}} = (\hat{\boldsymbol{\alpha}}_1', \ldots, \hat{\boldsymbol{\alpha}}_d')'$ and $\hat{\boldsymbol{\beta}}(\cdot) = (\hat{\beta}_1(\cdot), \ldots, \hat{\beta}_d(\cdot))'$. Without loss of generality let the support of $\mathbf{X}'\boldsymbol{\alpha}_k$ be $[0,1]$ and assume $h_k \sim h, k=1,\ldots,d$, namely, all bandwidths have the same asymptotic order. Denote by $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0(\mathbf{v}) = (\beta_{10}(v_1),\ldots,\beta_{d0}(v_d))'$ the true values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}(\cdot)$, respectively. Proofs of the following theorems may be found in the associated Supplementary Material.

Theorem 1. *Under the regularity conditions (C1)–(C7) listed in the Appendix, as $n \to \infty$ we have*

(i) $\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\| \xrightarrow{p} 0$ *and* $\sup_{\mathbf{v}\in\mathcal{V}} \|\hat{\boldsymbol{\beta}}(\mathbf{v}) - \boldsymbol{\beta}_0(\mathbf{v})\| \xrightarrow{p} 0$, *where* $\mathcal{V} = \{(v_1,\ldots,v_d)' : v_k \in [0,1], k = 1,\ldots,d\}$.

(ii) If $nh^4 \to 0$, *then* $\sqrt{n}\left(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right) \xrightarrow{d} N(0,\mathbb{V})$, *where the asymptotic covariance* $\mathbb{V}$ *is specified in Equation (A.1) in the Appendix.*

(iii) $(nh)^{1/2}\left\{\hat{\boldsymbol{\beta}}(\mathbf{v}) - \boldsymbol{\beta}_0(\mathbf{v}) - \frac{1}{2}\nu_2 h^2 \mathcal{A}^{-1}(\boldsymbol{\Omega})(\mathbf{v})\right\} \xrightarrow{d} N(0,\boldsymbol{\Pi}(\mathbf{v})), \mathbf{v}\in\mathcal{V}$, *where the pointwise asymptotic covariance* $\boldsymbol{\Pi}(\mathbf{v}) = \nu_0\left\{\mathcal{A}^{-1}(\boldsymbol{\Sigma}^{1/2})(\mathbf{v})\right\} \times \left\{\mathcal{A}^{-1}(\boldsymbol{\Sigma}^{1/2})(\mathbf{v})\right\}'$, *and* $\nu_r = \int x^r\mathcal{K}^2(x)dx, r = 0,1,2$. *Here the linear operator* $\mathcal{A}$ *and its inverse are the quantities defined in Equation (A.2) of the Appendix; the functions* $\boldsymbol{\Sigma}(\mathbf{v})$ *and* $\boldsymbol{\Omega}(\mathbf{v})$ *are also defined in the Appendix.*

When $d = 1$, Theorem 1 reduces to Theorems 1–3 of Lin, Tan & Li (2016). To achieve the parametric convergence rate $n^{-1/2}$ for the estimator $\hat{\boldsymbol{\alpha}}$, it is commonly required to undersmooth the nonparametric estimation using a kernel technique (Carroll et al., 1997; Hastie & Tibshirani, 1990b). Part (ii) of Theorem 1 in fact requires the bandwidth $h = o(n^{-1/4})$, leading to a scenario of undersmoothing. As usual, the asymptotic normality given in part (iii) of Theorem 1 indicates that the order of the asymptotic bias is $O(h^2)$ and the order of the asymptotic covariance is $(nh)^{-1}$. Consequently, the theoretical optimal bandwidth $O(n^{-1/5})$ may be in principle applied for the nonparametric estimation.

To establish semiparametric efficiency in the sense of Bickel et al. (1993) for both $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}(\cdot)$, we consider a function $\boldsymbol{\phi}(\mathbf{v}) = (\boldsymbol{\phi}_1', \boldsymbol{\phi}_2(\mathbf{v})')'$, which has continuous second-order derivatives on $\mathcal{V}$. Let $\boldsymbol{\phi}_1'\hat{\boldsymbol{\alpha}} + \int_{\mathcal{V}} \boldsymbol{\phi}_2'(\mathbf{v})\hat{\boldsymbol{\beta}}(\mathbf{v})d\mathbf{v}$ be an estimator of $\boldsymbol{\phi}_1'\boldsymbol{\alpha}_0 + \int_{\mathcal{V}} \boldsymbol{\phi}_2'(\mathbf{v})\boldsymbol{\beta}_0(\mathbf{v})d\mathbf{v}$, where $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}(\cdot)$ are the proposed GPL estimators.

Theorem 2. *Under the regularity conditions (C1)–(C7) listed in the Appendix, if* $nh^4 \to 0$ *and* $nh^2 \to \infty$, $\boldsymbol{\phi}_1'\hat{\boldsymbol{\alpha}} + \int_{\mathcal{V}} \boldsymbol{\phi}_2'(\mathbf{v})\hat{\boldsymbol{\beta}}(\mathbf{v})d\mathbf{v}$ *is an efficient estimator of* $\boldsymbol{\phi}_1'\boldsymbol{\alpha}_0 + \int_{\mathcal{V}} \boldsymbol{\phi}_2'(\mathbf{v})\boldsymbol{\beta}_0(\mathbf{v})d\mathbf{v}$.

It follows that with a choice of $\boldsymbol{\phi}_2(\mathbf{v}) = 0$, $\hat{\boldsymbol{\alpha}}$ is an efficient estimator of $\boldsymbol{\alpha}_0$; likewise, with a choice of $\boldsymbol{\phi}_1 = 0$, $\int_{\mathcal{V}} \boldsymbol{\phi}_2'(\mathbf{v})\hat{\boldsymbol{\beta}}(\mathbf{v})d\mathbf{v}$ is an efficient estimator of $\int_{\mathcal{V}} \boldsymbol{\phi}_2'(\mathbf{v})\boldsymbol{\beta}_0(\mathbf{v})d\mathbf{v}$.

Given the GPL estimates $\hat{\boldsymbol{\beta}}(\cdot)$ and $\hat{\boldsymbol{\alpha}}$, we adopt the method of kernel smoothing described in Fan, Lin & Zhou (2006) to estimate the baseline hazard function by $\hat{\lambda}_0(t) = \int \mathcal{K}_b(t-u)d\hat{\Lambda}_0(u)$, where b is a bandwidth and the estimated cumulative baseline hazard function $\hat{\Lambda}_0(t)$ is

$$\hat{\Lambda}_0(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{dN_i(u)}{n^{-1}\sum_{j=1}^{n} Y_j(u)\exp\left\{\sum_{k=1}^{d}\hat{\beta}_k(\mathbf{X}_j'\hat{\boldsymbol{\alpha}}_k)Z_{jk}\right\}}.$$

Given the results of Theorem 1, we follow the proof in Fan, Lin & Zhou (2006) to show that both $\hat{\lambda}_0(t)$ and $\hat{\Lambda}_0(t)$ are uniformly consistent estimators on $(0,\tau)$, where τ is defined via condition (C2) in the Appendix.

3. INFERENCE

3.1. Inference for the Loading Coefficients

Utilizing the asymptotic normality described in part (ii) of Theorem 1, we now construct a Wald statistic to test a null hypothesis $H_0 : \alpha_{k_1 l} = \cdots = \alpha_{k_r l} = 0$, which pertains to a subset, say $\boldsymbol{\alpha}_{l(r)} = (\alpha_{k_1 l}, \ldots, \alpha_{k_r l})'$, of the lth vector of loading coefficients, where $(k_1, \ldots, k_r)$ is a subset of the indices in $\{2, \ldots, q\}$. Clearly, the Wald test statistic is $\chi^2_W = (\hat{\boldsymbol{\alpha}}_{l(r)} - 0)' \left\{\hat{\mathbb{V}}_{l(r)}\right\}^{-1} (\hat{\boldsymbol{\alpha}}_{l(r)} - 0)$, where $\left\{\hat{\mathbb{V}}_{l(r)}\right\}^{-1}$ is the inverse of the estimated asymptotic covariance matrix corresponding to subvector $\hat{\boldsymbol{\alpha}}_{l(r)}$. Under the null hypothesis H_0, the statistic χ^2_W has an asymptotic chi-squared distribution with r degrees of freedom.

3.2. Inference for the Nonparametric Functions

The estimated value of $\boldsymbol{\beta}(\cdot)$ helps us understand in which form the effect of covariate Z_k is modified by an exposure mixture $m_k = \mathbf{X}'\boldsymbol{\alpha}_k$. Such an analysis simplifies the model specification, say, to a linear interaction model. We now propose a goodness-of-fit test using a GLR statistic for a constant function or a linear function.

In the case of a linear function, we set the null and alternative hypotheses as follows: $H_0 : \beta_l(\cdot)$ is a linear function versus $H_1 : \beta_l(\cdot)$ is not a linear function. Under the alternative H_1, we obtain the GPL estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}(\cdot)$ described above. Under the null H_0, the function $\beta_l(\cdot)$ is estimated by $\tilde{\beta}_l(v_l) = \hat{\theta}_{l0} + \hat{\theta}_{l1} v_l$, with $\hat{\theta}_{l0}$ and $\hat{\theta}_{l1}$ being the conventional partial likelihood estimates, given that all other parameters $\boldsymbol{\alpha}$ and functions $\boldsymbol{\beta}_{-l}(\cdot) = (\beta_k(\cdot), k \neq l)'$ have been estimated under the alternative H_1. Then a GLR statistic is constructed as the difference

$$\lambda_{n,l} = \log \mathcal{L}_n \left(\hat{\boldsymbol{\alpha}}, \hat{\beta}_l, \hat{\boldsymbol{\beta}}_{-l}\right) - \log \mathcal{L}_n \left(\hat{\boldsymbol{\alpha}}, \tilde{\beta}_l, \hat{\boldsymbol{\beta}}_{-l}\right),$$

where

$$\mathcal{L}_n \left(\boldsymbol{\alpha}, \beta_l, \boldsymbol{\beta}_{-l}\right) = \exp \left\{ \sum_{i=1}^{n} \Delta_i \left(W_i(\boldsymbol{\alpha}, \beta_l, \boldsymbol{\beta}_{-l}) - \log \left[\sum_{j \in \mathcal{R}(\mathcal{T}_i)} \exp \left\{ W_j \left(\boldsymbol{\alpha}, \beta_l, \boldsymbol{\beta}_{-l}\right) \right\} \right] \right) \right\},$$

and $W_i(\boldsymbol{\alpha}, \beta_l, \boldsymbol{\beta}_{-l}) = \beta_l(\mathbf{X}_i'\boldsymbol{\alpha}_l) Z_{il} + \sum_{k \neq l} \beta_k(\mathbf{X}_i'\boldsymbol{\alpha}_k) Z_{ik}$.

Theorem 3. *Suppose the regularity conditions (C1)–(C7) listed in the Appendix hold.*

(i) Under H_0: $\beta_l(v_l)$ follows a linear function form given by $\theta_{l0} + \theta_{l1} v_l$, we have

$$\gamma_k \lambda_{n,l} \xrightarrow{d} \chi^2_{\gamma_k \mu_{nl}} \quad \text{as} \;\; n \to \infty, \tag{8}$$

where $\gamma_k = \left\{ \mathcal{K}(0) - \frac{1}{2} \int_{-1}^{1} \mathcal{K}^2(t) dt \right\} / \int_{-1}^{1} \left\{ \mathcal{K}(t) - \frac{1}{2} \mathcal{K} * \mathcal{K}(t) \right\}^2 dt$, $\mu_{nl} = |\mathbb{D}_l| h^{-1} \left\{ \mathcal{K}(0) - \frac{1}{2} \times \int_{-1}^{1} \mathcal{K}^2(t) dt \right\}$, $\mathbb{D}_l = \{v_l : v_l = \boldsymbol{x}'\boldsymbol{\alpha}_l, f_{\boldsymbol{\alpha}_l}(v_l) > 0, \boldsymbol{\alpha}_l, \boldsymbol{x} \in R^q\}$ and $|\mathbb{D}_l|$ is the length of interval $\mathbb{D}_l$. Here "$*$" denotes the operation of convolution.

(ii) Consider H_0: β_l is a constant versus H_1: β_l is not a constant. Under this version of H_0, the result specified in Equation (8) continues to hold.

Theorem 3 shows that the asymptotic null distribution of the proposed GLR statistic is nearly χ^2 with a degree-of-freedom parameter that does not depend on the nuisance parameters θ_{l0} and θ_{l1}. This aspect is known as the Wilks phenomenon; see Fan, Zhang & Zhang (2001). With this property, the advantages of the classical likelihood ratio tests are fully inherited. For additional discussion, consult Fan, Zhang & Zhang (2001).

The procedure just described has been implemented as part of an R package called CoxGPLE that is presented in the associated Supplementary Material.

4. SIMULATION STUDIES

We conducted simulation studies to examine the performance of our proposed GPL method. We evaluated the performance of GPL for the nonparametric estimator $\hat{\boldsymbol{\beta}}(\cdot)$ by the weighted mean squared error (WMSE), WMSE $= n_g^{-1} \sum_{k=1}^{d} \sum_{j=1}^{n_g} w_k \{\hat{\beta}_k(v_j) - \beta_k(v_j)\}^2$, where w_j is the reciprocal sample variance of $\beta_j(v_k)$ over a set of grid points $\{v_k, k = 1, \ldots, n_g\}$. We assessed the performance of GPL for the parametric estimator $\hat{\boldsymbol{\alpha}}$ via its bias, empirical standard deviation (ESE) and the root mean squared error (RMSE). In all the cases we considered, we used the Epanechnikov kernel, set $n_g = 100$, and calculated summary statistics based on 200 simulations with sample size $n = 500$.

We specified a Cox model with three covariates, $(1, Z_1(t), Z_2)$, and three exposure variables, (X_1, X_2, X_3), that were defined as follows: $\lambda(t|\mathbf{X}, \mathbf{Z}) = \lambda_0(t) \exp\{\eta(\mathbf{X}, \mathbf{Z})\}$, where the baseline hazard function was $\lambda_0(t) = 4t^3$ and the linear predictor was $\eta(\mathbf{X}, \mathbf{Z}) = \beta_1(\mathbf{X}'\boldsymbol{\alpha}_1) + \beta_2(\mathbf{X}'\boldsymbol{\alpha}_2)Z_1(t) + \beta_3(\mathbf{X}'\boldsymbol{\alpha}_3)Z_2$. The three coefficient functions were specified as $\beta_1(v) = 0.4\{\exp(2v - 0.5) - \exp(-0.5)\}$, $\beta_2(v) = 1.3v(0.25 - v)$ and $\beta_3(v) = \sin(2v)$. In addition, to generate $\mathbf{Z} = (Z_1(t), Z_2)'$, with $Z_1(t)$ as a time-dependent covariate and Z_2 a time-independent covariate, we first obtained independent observations from a bivariate normal distribution, $(\tilde{Z}_1, Z_2)' \sim N\left\{0, 52\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right\}$, and then set $Z_1(t) = \tilde{Z}_1 I(t \leq 1)/4 + \tilde{Z}_1 I(t > 1)$. The three exposure variables, $\mathbf{X} = (X_1, X_2, X_3)'$, were sampled independently, where the first two variables were binary from Bernoulli(0.5) and $X_3 \sim$ Unif(0, 1). The true parameter values were set at $\boldsymbol{\alpha}_{10} = (2, 2, 2)'/5$, $\boldsymbol{\alpha}_{20} = (2, 2, -2)'/3$, and $\boldsymbol{\alpha}_{30} = (2, 2, -2)'/3$, which all have norm 1. The censoring variable C, given $(\mathbf{Z}, \mathbf{X})$, was simulated uniformly on the interval $(0, u(\mathbf{Z}, \mathbf{X}))$, where the upper limit was specified as $u(\mathbf{Z}, \mathbf{X}) = c_1 I(\eta(\mathbf{Z}, \mathbf{X}) > \eta_0) + c_2 I(\eta(\mathbf{Z}, \mathbf{X}) \leq \eta_0)$; the cutoff η_0 was set at 0.52, the mean function of $\eta(\mathbf{Z}, \mathbf{X})$, $c_1 = 2$, and $c_2 = 15$. The censoring rate was approximately 20%.

To evaluate the proposed GPL method, we focused on its efficiency loss relative to two comparison models that were close to the true model. In these two comparison models, the coefficient functions are either specified as the true functions or take the same functional forms as those of the true functions. They were $\lambda_1(t) = \lambda_0(t) \exp\{\eta_1(\mathbf{Z}, \mathbf{X}; \boldsymbol{\theta}_1)\}$ for Simulation Model 1 (simM1), and $\lambda_2(t) = \lambda_0(t) \exp\{\eta_2(\mathbf{Z}, \mathbf{X}; \boldsymbol{\theta}_2)\}$ for Simulation Model 2 (SimM 2), where, with $v_k = \mathbf{X}'\boldsymbol{\alpha}_k$, $k = 1, 2, 3$, $\eta_1(\mathbf{Z}, \mathbf{X}; \boldsymbol{\theta}_1) = \theta_{11}\beta_1(v_1) + (\theta_{12}v_2 + \theta_{13}v_2^2)Z_1(t) + \theta_{14}\beta_3(v_3)Z_2$ and $\eta_2(\mathbf{Z}, \mathbf{X}; \boldsymbol{\theta}_2) = \left\{\theta_{21}\beta_1(v_1) + \theta_{22}v_1\right\} + (\theta_{23} + \theta_{24}v_2 + \theta_{25}v_2^2)Z_1(t) + \left\{\theta_{26} + \theta_{27}v_3 + \theta_{28}\beta_3(v_3)\right\} Z_2$. The vectors of parameters in the above parametric linear predictors were $\boldsymbol{\theta}_1 = (\theta_{11}, \ldots, \theta_{14})'$, $\boldsymbol{\theta}_2 = (\theta_{21}, \ldots, \theta_{28})'$, and clearly these true coefficient functions were parametrically nested in $\eta_1(\cdot)$ and $\eta_2(\cdot)$, respectively. When both models are correctly specified, SimM1 corresponds to a smaller parameter space while SimM2 pertains to a larger parameter space. For models M1 and M2, the loading coefficients $\boldsymbol{\alpha}_k$ in $v_k = \mathbf{X}'\boldsymbol{\alpha}_k$, $k = 1, 2, 3$, together with $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ were estimated using conventional partial likelihood.

Panels (a), (b) and (c) of Figure 1 in the Supplementary Material display the estimated coefficient functions at bandwidth $h = 0.3$, together with their empirical pointwise 95% confidence bands based on 200 simulations. It is easy to see that all estimated curves (denoted by solid lines) are close to the true curves, which are indicated by dashed lines. Some numerical results concerning the GPL method are summarized in Table 1, including average estimation bias and empirical standard error for functional values at $v = -0.60, -0.26, 0.32, 0.90$ and 1.24 over 200 replicates. These values correspond, approximately, to the 10th, 25th, 50th, 75th and 90th percentiles of the distribution of index $v_k = \mathbf{X}'\boldsymbol{\alpha}_k$ for $k = 1, 2, 3$. Both Figure 1 in the Supplementary Material and the results summarized in Table 1 indicate that the GPL method performed well in this simulation setting.

TABLE 1: Simulation study results for nonparametric estimation of the functions $\beta_k(\cdot)$ using the proposed GPL with bandwith $h = 0.3$ over 200 replicates.

Function	Summary statistic	Value of index v				
		−0.60	−0.26	0.32	0.90	1.24
	Bias	0.274	0.202	0.095	0.003	−0.654
$\beta_1(\cdot)$	ESE	0.279	0.279	0.142	0.170	0.242
	RMSE	0.391	0.344	0.171	0.170	0.697
	Bias	0.066	−0.023	−0.008	0.062	0.156
$\beta_2(\cdot)$	ESE	0.133	0.069	0.036	0.129	0.267
	RMSE	0.148	0.073	0.037	0.143	0.309
	Bias	0.108	0.084	−0.083	−0.147	−0.063
$\beta_3(\cdot)$	ESE	0.075	0.043	0.036	0.040	0.058
	RMSE	0.131	0.094	0.090	0.152	0.086

Table 2 reports the summary results from our simulation study concerning estimation of the loading coefficients $\boldsymbol{\alpha}_k, k = 1, 2, 3$ in the proposed Cox regression model and the two comparison parametric models SimM1 and SimM2. These results include average estimation bias, ESE and RMSE of the GPL method with bandwidth $h = 0.2, 0.3, 0.4$ compared to the classical partial likelihood estimation (PLE) method used for the comparison models M1 and M2. Since Model M1 has the coefficient functions in the linear predictor $\eta_1(\cdot)$ specified as special parametric forms of the true functions, it is not surprising that the PLE method with fewer parameters (Model M1) performed the best among the various cases that we considered. The performance of the GPL method fell between performance of the PLE methods for Models M1 and M2, suggesting that there exists a parametric model with which the proposed nonparametric method would exhibit similar performance. In effect, this observation suggests that the GPL method for estimating the loading coefficients is parametrically efficient.

Using the same simulations, we also tried to compare the GPL method of estimation to the local partial likelihood estimation (LPLE) method that we described in Section 2.1. However, we encountered numerous instances of failure to converge numerically. For example, at bandwidth $h = 2.5$ (a case of excessive oversmoothing), out of 200 simulations 67.5% failed to achieve the required convergence criterion for LPLE. Such a numerical challenge for the LPLE method is largely attributable to the curse of dimensionality, since the LPLE method needs to estimate three nonparametric functions in addition to the loading coefficients $\boldsymbol{\alpha}_k, k = 1, 2, 3$. In comparison, only 6% of the simulations failed to converge at $h = 0.4$ with the proposed GPL method. Because of this numerical instability for the LPLE method, we have omitted any details concerning comparison of the GPL method with the LPLE method.

5. APPLICATION

The height of an adolescent has been reported as an important predictor for the age of pubertal development (Salsberry, Regan & Pager, 2009; Karapanou & Papadimitriou 2010). Maternal parity (the number of previous pregnancies) has also been identified as an important predictor of pubertal onset (Ong et al., 2002). It is known that phthalates may affect the tempo of physical growth during sensitive periods of development in childhood, which in itself is related to chronic disease risk as well as the timing and tempo of pubertal development (Salazar et al., 2004). Several

TABLE 2: Simulation study results of the estimation for the loading coefficients $\boldsymbol{\alpha}_k, k = 1, 2, 3$ using either the proposed GPL method with bandwidth h or classical partial likelihood estimation (PLE) with parametric models SimM1 or SimM2 over 200 replicates.

			Loading coefficients					
Method	Bandwidth	Statistic	α_{12}	α_{13}	α_{22}	α_{23}	α_{32}	α_{33}
		Bias	0.005	0.037	0.013	−0.022	0.002	−0.006
	0.2	ESE	0.106	0.141	0.076	0.088	0.020	0.043
		RMSE	0.107	0.146	0.077	0.091	0.020	0.044
		Bias	0.009	0.028	0.012	−0.012	0.002	−0.007
GPL	0.3	ESE	0.089	0.123	0.078	0.106	0.020	0.039
		RMSE	0.089	0.126	0.079	0.106	0.020	0.040
		Bias	0.016	0.028	0.029	−0.030	0.007	−0.009
	0.4	ESE	0.098	0.121	0.080	0.087	0.020	0.038
		RMSE	0.099	0.125	0.085	0.092	0.021	0.039
PLE.SimM1		Bias	0.009	0.005	0.000	−0.002	−0.002	0.001
	–	ESE	0.102	0.109	0.039	0.063	0.015	0.021
		RMSE	0.102	0.109	0.039	0.063	0.015	0.022
PLE.SimM2		Bias	0.033	0.040	−0.006	0.001	0.017	−0.020
	–	ESE	0.182	0.266	0.043	0.071	0.074	0.084
		RMSE	0.186	0.269	0.044	0.071	0.076	0.086

studies have shown that exposure to mixtures of reproductive toxicants may disrupt complex signalling pathways and result in cumulative effects on a child's growth (Rider et al., 2010).

This section presents our analysis of the pubertal development data that were introduced in Section 1. Working through multiple steps of data cleaning and validation under the guidance of our collaborators, we obtained a sample of 549 children aged 9.8–18.1 years for the analysis. The age at which the first stage of pubic hair developed is treated clinically as the age of attainment for study participants, or is right-censored at the age of a subject's last completed assessment. The time to this event was observed during an average follow-up time of 14.3 years, and the rate of right censoring in the study data was 14.6%. Exposure variables of interest included prenatal exposure to MEP (X_1), MBP (X_2) and MCPP (X_3), measured during the third trimester of pregnancy. Other covariates of interest were maternal parity (Z_1) and each child's concurrent height (Z_2). We normalized both exposure variables and covariates in our analysis.

The primary objective of this analysis was to evaluate the functional exposure-covariate interactions, which would allow us to answer whether or not the early life exposure in utero to phthalates may modify the effects of maternal parity and child's height on the age of pubertal development. We began with a preliminary analysis that was based on two traditional Cox models:

M1: Model 1 with only main effects

$$\lambda(t \mid \mathbf{X}, \mathbf{Z}) = \lambda_0(t) \exp(\alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \beta_1 Z_1 + \beta_2 Z_2),$$

and

TABLE 3: Analysis results for the pubertal development data. Regression parameter estimates from two proportional hazards models: Model 1 involves only main covariate effects whereas Model 2 incorporates linear interactions. Also, estimated loading coefficients for phthalate mixtures in our proposed Cox model (Cox). The standard errors cited for Models 1 and 2 were obtained from the R package survival, whereas those for the proposed Cox model were derived using 500 bootstrap samples.

Model	Parameter	Estimate	Standard error	P-value
	α_1	−0.163	0.047	0.001
	α_2	0.285	0.056	0.000
Model 1	α_3	−0.215	0.057	0.000
	β_1	−0.122	0.047	0.001
	β_2	−0.449	0.059	0.000
	α_{11}	−0.163	0.050	0.001
	α_{12}	0.308	0.059	0.000
	α_{13}	−0.228	0.060	0.000
	α_{20}	−0.117	0.047	0.013
	α_{21}	−0.036	0.048	0.453
Model 2	α_{22}	−0.083	0.060	0.167
	α_{23}	0.006	0.054	0.920
	α_{30}	−0.487	0.060	0.000
	α_{31}	0.016	0.060	0.784
	α_{32}	−0.172	0.066	0.009
	α_{33}	0.146	0.070	0.037
	α_{11}	0.386	0.138	0.005
	α_{12}	−0.771	0.060	0.000
	α_{13}	0.507	0.118	0.000
	α_{21}	0.420	0.177	0.018
Cox	α_{22}	−0.765	0.153	0.000
	α_{23}	−0.488	0.156	0.002
	α_{31}	0.359	0.095	0.000
	α_{32}	−0.696	0.060	0.000
	α_{33}	0.623	0.077	0.000

M2: Model 2 with linear interactions

$$\lambda(t \mid \mathbf{X}, \mathbf{Z}) = \lambda_0(t) \exp\left\{\alpha_{11}X_1 + \alpha_{12}X_2 + \alpha_{13}X_3 + \left(\alpha_{20} + \alpha_{21}X_1 + \alpha_{22}X_2 + \alpha_{23}X_3\right) Z_1 + \left(\alpha_{30} + \alpha_{31}X_1 + \alpha_{32}X_2 + \alpha_{33}X_3\right) Z_2\right\}.$$

Table 3 reports the results from our analysis of models M1 and M2 using the R package *survival*. This preliminary analysis with M1 indicated that the phthalates MEP, MBP and MCPP

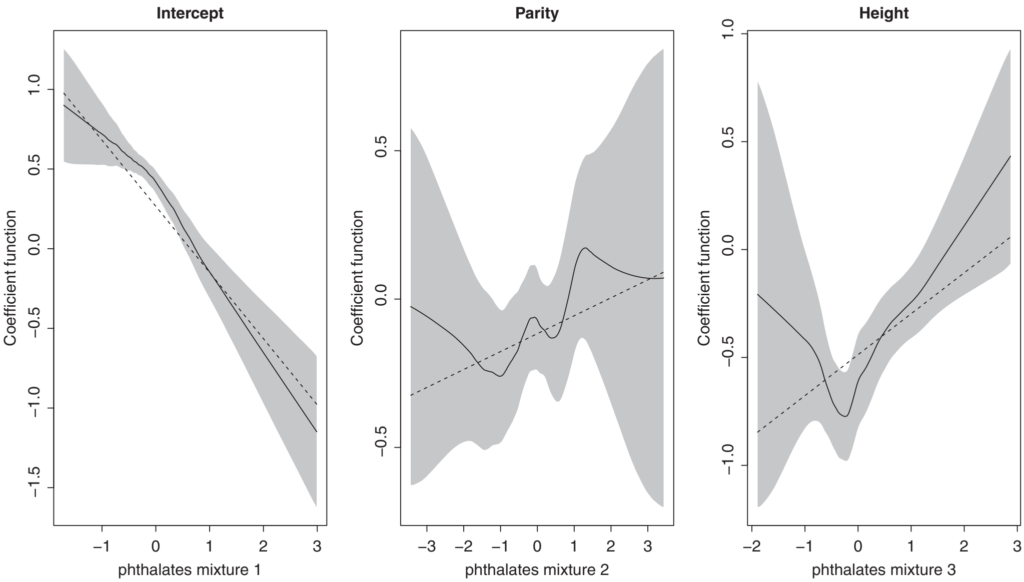

FIGURE 1: Analyzing the pubertal development study. Estimated coefficient functions (solid curves) and their corresponding 95% confidence intervals (shaded regions) for $\beta_k(\cdot)$, $k = 1, 2, 3$, together with the estimated linear coefficient functions (dashed lines) for an intercept (X_1), maternal parity (X_2) and a child's concurrent height (X_3).

were significantly associated with the age of attaining the first stage of puberty. The results from M2 suggested that the effect of height on response was modified by a combination of MBP and MCPP (or a mixture of phthalates), whereas the effect of maternal parity was not modified linearly by MEP, MBP or MCPP.

It is of great interest to investigate whether these phthalate mixtures may have modified the other covariate effects in a nonlinear fashion. To proceed, we used our proposed Cox model with functional covariate-exposure interactions, that is, $\lambda(t \mid \mathbf{X}, Z_1, Z_2) = \lambda_0(t) \exp\left\{\beta_1(\mathbf{X}'\boldsymbol{\alpha}_1) + \beta_2(\mathbf{X}'\boldsymbol{\alpha}_2)Z_1 + \beta_3(\mathbf{X}'\boldsymbol{\alpha}_3)Z_2\right\}$, where the $\beta_k(\cdot)$ represent unknown, smooth functions and the $\boldsymbol{\alpha}_k = (\alpha_{k1}, \alpha_{k2}, \alpha_{k3})'$ for $k = 1, 2, 3$ denote loading parameters that need to be estimated. The initial values of both functions and loading parameters were chosen by fitting the model using regression splines with four knots.

Based on the cross-validation criterion that we specified in Equation (7), we found $Q = 0.5$, which was used for the bandwidth selection. The estimates of the loading coefficients are reported in the lower portion of Table 3, and the estimated functions are plotted in Figure 1. To calculate standard errors, we used the method of bootstrap resampling with 500 bootstrap samples, in which each subject is treated as a resampling unit in order to preserve the inherent features of the data from individual subjects. The choice of 500 bootstrap samples was determined by monitoring the stability of the resulting standard error estimates.

To determine whether or not the covariate-exposure interactions were linear, we considered the hypothesis $H_0 : \beta_k(\cdot)$ is linear, that is, the existence of linear interactions, versus $H_1 : \beta_k(\cdot)$ is not linear, $k = 1, 2, 3$. The P values of the associated GLR tests were obtained from the χ^2 null distribution identified in Theorem 3. They were 0.396 for $\beta_1(\cdot)$, 0.015 for $\beta_2(\cdot)$ and 0.020 for $\beta_3(\cdot)$, respectively. These results suggest that exposure to a mixture of these phthalates alters the effects of maternal parity and concurrent height on the response of interest. In addition, we considered the hypothesis $H_0 : \beta_1(\cdot)$ is constant, that is, the absence of any main effect, versus

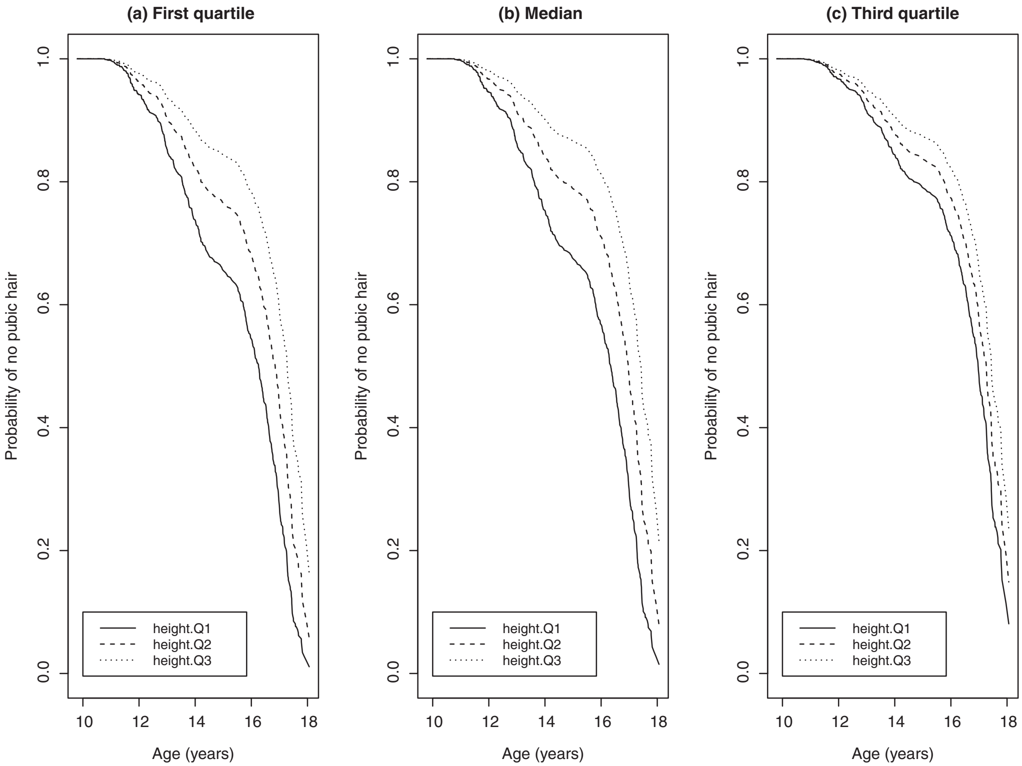

FIGURE 2: Analyzing the pubertal development study. Estimated survival functions for the three height groups with a fixed exposure to mixtures of phthalates m_k, $k = 1, 2, 3$ at the first quartile, median, and third quartiles of mixture $m = \alpha_{31}X_1 + \alpha_{32}X_2 + \alpha_{33}X_3$. The height groups correspond to girls with a height equal to the first quartile (solid line), median (dashed line) or third quartile (dotted line), and an average value for maternal parity.

$H_1 : \beta_1(\cdot)$ is not constant. The corresponding P value of the GLR test in this case was 0.000. Combining the results from these two tests concerning β_1, we concluded that a linear main effect for exposure adequately captures the underlying functional form. This finding agrees with the graphical evidence of linear functions that is displayed in Figure 1.

From the fitted model labelled ''Cox'' summarized in Table 3, as well as the estimated coefficient functions plotted in Figure 1, it appears that greater exposure to a combination of phthalates is associated with a delay in attaining the first stage of puberty. In general, although a taller adolescent is more likely to reach the first stage of puberty at an older age, the level of exposure to a mixture of maternal phthalates appears to alter, nonlinearly, the relationship between concurrent height and the age at which the first stage of puberty is attained. Likewise, with respect to maternal parity, its effect on response is also modified in a nonlinear manner by exposure to a mixture of maternal phthalates.

Panels (a), (b) and (c) in Figure 2 display estimated survival functions for the three height groups with the quartiles of exposure to phthalates mixtures $m = \alpha_{31}X_1 + \alpha_{32}X_2 + \alpha_{33}X_3$, when maternal parity is fixed at its average value in the study data. Readers can see that at a fixed quartile level of exposure to the phthalates mixture and at an average value of maternal parity, a taller adolescent tends to experience a slower progression to sexual maturation, as measured by observing the Tanner stage for pubic hair. Notice, also, that more severe exposure to the phthalates mixture is associated with a longer delay in attaining the first stage of puberty. Unfortunately, the two parametric models, namely Model 1 and Model 2, failed to capture these nonlinear interaction effects. Such estimated patterns and interpretations with respect to the effect of prenatal exposure to the phthalates mixture and other risk factors, such as maternal parity and concurrent height, have been detected and estimated via the modelling of nonlinear effects. In this instance, the insights gained represented meaningful scientific knowledge for our collaborators.

6. CONCLUDING REMARKS

This article has focused on developing a new Cox regression model to address methodological needs in the evaluation of nonlinear interaction effects arising during studies in the environmental health sciences, where existing methods have shown that they are unable to provide satisfactory solutions. One advantage of the proposed method pertains to the estimating efficiency of the proposed GPL method of estimation that uses all the data in both nonparametric and parametric parameter estimation, compared to existing methods such as the local scoring backfitting method introduced by Buja, Hastie & Tibshirani (1989) that uses only local data. We established both estimation consistency and asymptotic normality as part of our investigation of this new methodology. We also proposed a GLR test that enabled us to test for a particular hypothesized form of functional interaction effects, such as constants and linear functions. We showed that this test satisfies the Wilks phenomenon that makes implementing the test straightforward.

In addition, we developed the R package coxphGPLE to implement our proposed methodology. The estimates given in our R package are assumed to satisfy regularity condition (C6), which is listed in the Appendix to the article. Initial values are critically important in the search for reliable solutions. Inspecting bootstrap estimates is useful to check local convexity of the objective functions empirically, which is essential for consistent parametric and nonparametric estimation.

Although the methodology that we have proposed is primarily motivated by modelling functional covariate-environment interactions, the proposed methods, as well as the corresponding theoretical results, are quite general and should be applicable to problems from other fields of study. As usual, bandwidth selection and use of the bootstrap to calculate standard errors are computationally demanding. To estimate the unknowns in a model with functional interactions any researcher would certainly need a reasonably large sample size. In addition, we also assumed that any right censoring of the response was independent, which may not hold in some applications. An important future project would involve extending the current model to permit a high-dimensional vector of toxicants in the formation of mixtures. In this case, the number of loading coefficients would also be high-dimensional and hard to estimate. We believe that developing a regularized method of estimation that can accommodate such high-dimensional situations is well worth exploring.

ACKNOWLEDGEMENTS

The authors thank the editor, the associate editor and two referees for constructive comments and suggestions that led to notable improvements in this article. This research was funded by grants from the National Institutes of Health (Zhou and Song, P01ES02284401, R01ES024732), as well as the National Science Foundation (DMS1513595) and the National Nature Science Foundation of China (Lin, 11528102).

BIBLIOGRAPHY

Bickel, P. J., Klaassen, C. A., Wellner, J. A., & Ritov, Y. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore.

Brockmann, M., Gasser, T., & Herrmann, E. (1993). Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association*, 88, 1302–1309.

Buja, A., Hastie, T., & Tibshirani, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, 17, 453–510.

Cai, Z., Fan, J., & Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95, 888–902.

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

Chen, K., Guo, S., Sun, L., & Wang, J. -L. (2010). Global partial likelihood for nonparametric proportional hazards models. *Journal of the American Statistical Association*, 105, 750–760.

Chen, K., Lin, H., & Zhou, Y. (2012). Efficient estimation for the Cox model with varying coefficients. *Biometrika*, 99, 379–392.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRC, Boca Raton, FL.

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

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

Fan, J., Zhang, C., & Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *The Annals of Statistics*, 29, 153–193.

Gluckman, P. D., Hanson, M. A., Cooper, C., & Thornburg, K. L. (2008). Effect of in utero and early-life conditions on adult health and disease. *New England Journal of Medicine*, 359, 61–73.

Hastie, T. & Tibshirani, R. (1990a). Exploring the nature of covariate effects in the proportional hazards model. *Biometrics*, 46, 1005–1016.

Hastie, T. & Tibshirani, R. (1990b). *Generalized Additive Models*. Chapman & Hall/CRC, Boca Raton, FL.

Honda, T. (2005). Estimation in additive Cox models by marginal integration. *Annals of the Institute of Statistical Mathematics*, 57, 403–423.

Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *The Annals of Statistics*, 27, 1536–1563.

Karapanou, O. & Papadimitriou, A. (2010). Determinants of menarche. *Reproductive Biology and Endocrinology*, 8, 2092–2096.

Lin, H., Tan, M. T., & Li, Y. (2016). A semiparametrically efficient estimator of single-index varying coefficient Cox proportional hazards models. *Statistica Sinica*, 26, 779–807.

Linton, O. B., Nielsen, J. P., & Van de Geer, S. (2003). Estimating multiplicative and additive hazard functions kernel methods. *The Annals of Statistics*, 31, 464–492.

Ma, S. & Song, P. X. -K. (2015). Varying index coefficient models. *Journal of the American Statistical Association*, 110, 341–356.

Mammen, E., Linton, O., & Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *The Annals of Statistics*, 27, 1443–1490.

Meeker, J. D. (2012). Exposure to environmental endocrine disruptors and child development. *Archives of Pediatrics & Adolescent Medicine*, 166, 952–958.

Ong, K. K., Preece, M. A., Emmett, P. M., Ahmed, M. L., & Dunger, D. B. (2002). Size at birth and early childhood growth in relation to maternal smoking, parity and infant breast-feeding: Longitudinal birth cohort study and analysis. *Pediatric Research*, 52, 863–867.

Perera, F. & Herbstman, J. (2011). Prenatal environmental exposures, epigenetics, and disease. *Reproductive Toxicology*, 31, 363–373.

Rider, C. V., Furr, J. R., Wilson, V. S., & Gray Jr, L. E. (2010). Cumulative effects of in utero administration of mixtures of reproductive toxicants that disrupt common target tissues via diverse mechanisms of toxicity. *International Journal of Andrology*, 33, 443–462.

Salazar, V., Castillo, C., Ariznavarreta, C., Campón, R., & Tresguerres, J. A. (2004). Effect of oral intake of dibutyl phthalate on reproductive parameters of Long Evans rats and pre-pubertal development of their offspring. *Toxicology*, 205, 131–137.

Salsberry, P. J., Reagan, P. B., & Pajer, K. (2009). Growth differences by age of menarche in African-American and white girls. *Nursing Research*, 58, 382–390.

Sleeper, L. A. & Harrington, D. P. (1990). Regression splines in the Cox model with application to covariate effects in liver disease. *Journal of the American Statistical Association*, 85, 941–949.
Tian, L., Zucker, D., & Wei, L. (2005). On the Cox model with time-varying regression coefficients. *Journal of the American Statistical Association*, 100, 172–183.
Vaiserman, A. (2014). Early-life exposure to endocrine disrupting chemicals and later-life health outcomes: An epigenetic bridge?. *Aging and Disease*, 5, 419–429.
Wang, J. L., Xue, L., Zhu, L., & Chong, Y. S. (2010). Estimation for a partial-linear single-index model. *The Annals of Statistics*, 38, 246–274.
Yu, K., Park, B. U., & Mammen, E. (2008). Smooth backfitting in generalized additive models. *The Annals of Statistics*, 36, 228–260.

Appendix

Notation

Let $\mathcal{C}_0 = \{\boldsymbol{\delta}(\mathbf{v}) = (\delta_1(v_1), \dots, \delta_d(v_d))' : \mathbf{v} \in \mathcal{V}, \boldsymbol{\delta}(\mathbf{v})$ is continuous on $\mathcal{V}\}$. Let $\boldsymbol{\Theta}$ denote the support of $\boldsymbol{\alpha}$, f be the density function of the p-dimensional vector $\mathbf{X}$, $f_{\boldsymbol{\alpha}_k}(\cdot)$ be the density function of $m_{jk} = \mathbf{X}_j'\boldsymbol{\alpha}_k$, $f_{\boldsymbol{\alpha}_k,\boldsymbol{\alpha}_r}(v_k, v_r)$ be the density function of (m_{jk}, m_{jr}) and $g_{\boldsymbol{\alpha}_k}(v) = E(\mathbf{X}|\mathbf{X}'\boldsymbol{\alpha}_k = v)$. To simplify the notation, we use $f_k(v) = f_{\boldsymbol{\alpha}_k}(v)$, $f_{kr}(v_k, v_r) = f_{\boldsymbol{\alpha}_k,\boldsymbol{\alpha}_r}(v_k, v_r)$ and $g_k(v) = g_{\boldsymbol{\alpha}_{k0}}(v)$. Denote the survival function by $P(t \mid \mathbf{z}, \mathbf{x}) = \Pr(\mathcal{T} > t \mid \mathbf{Z} = \mathbf{z}, \mathbf{X} = \mathbf{x})$, and let $[a_i]_1^d = (a_1, \dots, a_d)'$. Denote $W(\boldsymbol{\alpha}, \boldsymbol{\delta}) = \boldsymbol{\delta}(\boldsymbol{\alpha}\circ\mathbf{X})'\mathbf{Z}$ with $\boldsymbol{\alpha}\circ\mathbf{X} = [\boldsymbol{\alpha}_l'\mathbf{X}]_{l=1}^d$. Let $W = W(\boldsymbol{\alpha}_0, \boldsymbol{\delta}_0)$, $W_i = W_i(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) = \boldsymbol{\beta}_0(\boldsymbol{\alpha}_0\circ\mathbf{X}_i)'\mathbf{Z}_i$ and $m_{jk,0} = \mathbf{X}_j'\boldsymbol{\alpha}_{k0}$.

The asymptotic covariance $\mathbb{V}$ mentioned in part (ii) of Theorem 1 is

$$\mathbb{V} = \Pi_2^{-1} \left[\int_0^\tau E\left\{\xi_i(t)^{\otimes 2} P(t|\mathbf{Z}_i, \mathbf{X}_i) \exp\left(W_i\right)\right\} \lambda_0(t)dt \right] (\Pi_2^{-1})', \tag{A.1}$$

where $\xi_i(t)$ and Π_2 can be found in the corresponding Supplementary Material.

Functions $\boldsymbol{\Sigma}(\mathbf{v})$ and $\boldsymbol{\Omega}(\mathbf{v})$ that appear in the asymptotic expression given in part (iii) of Theorem 1 are equal to $\boldsymbol{\Omega}(\mathbf{x}) = \boldsymbol{\Sigma}(\mathbf{x})\ddot{\boldsymbol{\beta}}_0(\mathbf{x})$, $\boldsymbol{\Sigma}(\mathbf{x}) = \mathrm{diag}(\Xi_{1,20}(x_1), \dots, \Xi_{d,20}(x_d))$, $\ddot{\boldsymbol{\beta}}_0(\mathbf{x}) = \left[\ddot{\beta}_{k0}(x_k)\right]_{k=1}^d$, $\Xi_{k,ij}(v) = \Xi_{k,ij}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, v)$, $\Xi_{k,ij}(\boldsymbol{\alpha}, \boldsymbol{\delta}_1, v) = \int_0^\tau s_{k,ij}(t; \boldsymbol{\alpha}, \boldsymbol{\delta}_1, v)\lambda_0(t)dt$. The linear operator $\mathcal{A}$ in part (iii) of Theorem 1 is equal to

$$\mathcal{A}(\boldsymbol{\phi})(\mathbf{x}) = \int_0^1 \left\{(\mathcal{H}(\mathbf{x}) - \mathbb{E}(\mathbf{x}))\, \Pi_2^{-1} \mathcal{S}_{10}(v) + (\mathcal{D}(\mathbf{x}, v) - \mathcal{J}(\mathbf{x}, v))\right\} \boldsymbol{\phi}(v)dv, \tag{A.2}$$

for any vector function $\boldsymbol{\phi}$. Let $\boldsymbol{\psi}(\mathbf{x}) = \mathcal{A}^{-1}(\boldsymbol{\Omega})(\mathbf{x})$, which means $\boldsymbol{\psi}(\mathbf{x})$ is the solution that satisfies $\mathcal{A}(\boldsymbol{\psi})(\mathbf{x}) = \boldsymbol{\Omega}(\mathbf{x})$. Additional details concerning notation can be found in the associated Supplementary Material.

Regularity Conditions

(C1) The kernel function $\mathcal{K}(x)$ is a symmetric density function with compact support $[-1, 1]$ and continuous derivatives.
(C2) The quantity τ is a finite positive value such that $\Pr(T > \tau) > 0$ and $\Pr(C = \tau) > 0$.
(C3) $(\mathbf{Z}, \mathbf{X})$ are bounded with compact support, and $P(C = 0 \mid \mathbf{Z} = \mathbf{z}, \mathbf{X} = \mathbf{x}) < 1$.
(C4) $\boldsymbol{\alpha} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is a bounded compact set.
(C5) Let $g_{\boldsymbol{\alpha}_k}(v) = E(\mathbf{X}_j|\mathbf{X}_j'\boldsymbol{\alpha}_k = v)$. The density function $f_{\boldsymbol{\alpha}_k}(v)$ of $\mathbf{X}'\boldsymbol{\alpha}_k$ is bounded away from zero; $g_{\boldsymbol{\alpha}_k}(v)$ and $f_{\boldsymbol{\alpha}_k}(v)$ have continuous second-order derivatives with respect to v for any $\boldsymbol{\alpha}_k$. The function $\boldsymbol{\beta}(v)$ and $s_{k,ij}(t; \boldsymbol{\alpha}_k, \boldsymbol{\theta}, \boldsymbol{\delta}_1, v)$ are twice continuously differentiable with respect to $v \in [0, 1]$ for any $t \in [0, \tau]$, $\boldsymbol{\alpha} \in \boldsymbol{\Theta}$, $\boldsymbol{\delta}_1 \in \mathcal{C}_0$.

(C6) For $k = 1, \ldots, d$, there exists a unique root $(\boldsymbol{\alpha}, \boldsymbol{\delta}_1)$ of the following equations:

$$\int_0^\tau \left\{ r_k(t; \boldsymbol{\alpha}_k, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\delta}_2) - r_k(t; \boldsymbol{\alpha}_k, \boldsymbol{\alpha}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2) \frac{s_{00}(t)}{s_{00}(t; \boldsymbol{\alpha}, \boldsymbol{\delta}_1)} \right\} \lambda_0(t) dt = 0,$$
$$\int_0^\tau \left\{ s_{k,10}(t; \boldsymbol{\alpha}_k, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, m_{ik}) - s_{k,10}(t; \boldsymbol{\alpha}_k, \boldsymbol{\alpha}, \boldsymbol{\delta}_1, m_{ik}) \frac{s_{00}(t)}{s_{00}(t; \boldsymbol{\alpha}, \boldsymbol{\delta}_1)} \right\} \lambda_0(t) dt = 0,$$

in $\boldsymbol{\delta}_1 \in C_0$ and $\boldsymbol{\alpha} \in \boldsymbol{\Theta}$ for any bounded $\boldsymbol{\delta}_2$ and $m_{ik} = \mathbf{X}_i' \boldsymbol{\alpha}_k$.

(C7) $h^2 \log(n) \to 0$ and $nh^3 \to \infty$.

Received 21 January 2018
Accepted 20 October 2018