Biometrika (2011), **xx**, x, *pp*. 1–14

Printed in Great Britain

Efficient estimation for the Cox model with varying coefficients

BY KANI CHEN
Department of Mathematics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
makchen@ust.hk

HUAZHEN LIN
School of Statistics, Southwestern University of Finance and Economics, Wenjiang, Chengdu, Sichuan, China, 611130
linhz@swufe.edu.cn

AND YONG ZHOU
Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing, China, 100190
yzhou@amss.ac.cn

SUMMARY

A proportional hazards model with varying coefficients allows one to examine the extent to which covariates interact nonlinearly with an exposure variable. A global partial likelihood method, in contrast with the local partial likelihood method of Fan et al. (2006), is proposed for estimation of varying coefficient functions. The proposed estimators are proved to be consistent and asymptotically normal. Semiparametric efficiency of the estimators is demonstrated in terms of their linear functionals. Evidence in support of the superiority of the method is presented in numerical studies and real examples.

Some key words: Global partial likelihood, proportional hazards model, semiparametric efficiency, varying coefficient.

1. INTRODUCTION

The Cox proportional hazards model is one of the most celebrated models for analyzing lifetime data. It explicitly postulates the covariate effects on the hazard risk to be $\lambda(t) = \lambda_0(t)\exp\{f(Z)\}$, where $\lambda_0(\cdot)$ is the baseline hazard function and $f(\cdot)$ represents the covariate effects and is commonly assumed to be a linear function. In practice, however, some covariates may interact nonlinearly with other exposure variables. For example, the effects of a medical treatment Z may vary with age W. The variation often cannot be calibrated properly in a parametric form, so the following varying coefficient Cox model is considered by Fan et al. (2006):

$$\lambda(t) = \lambda_0(t)\exp\{\beta(W)'Z + g(W)\}, \tag{1}$$

where $\beta(\cdot)$ and $g(\cdot)$ are unknown functions of interest, characterizing the extent to which the association varies with the scalar exposure variable W. When the variable W represents time, model (1) becomes a Cox model with time-dependent coefficients, which has been extensively studied; see Zucker & Karr (1990), Murphy & Sen (1991), Gamerman (1991), Murphy (1993), Marzec & Marzec (1997), Martinussen et al. (2000), Cai & Sun (2003), Tian et al. (2005), among others.

With parametric $g(\cdot)$ and $\beta(\cdot)$ in model (1), Cox's partial likelihood approach is the standard tool for statistical analysis; see Cox (1972), Kalbfleisch & Prentice (1980), Cox & Oakes (1984) and Andersen et al. (1993), among many others. For model (1) with unknown $g(\cdot)$ and $\beta(\cdot)$, Fan et al. (2006) propose a local partial likelihood approach for estimating $g(\cdot)$ and $\beta(\cdot)$. A local partial likelihood is based on observations with W_i in a small neighborhood of a given w. The method is properly motivated and rather simple to implement and analyze. However, the localization suffers a loss of efficiency since the observations outside the neighborhood which carry information about $g(w)$ and $\beta(w)$ are not used. Moreover, $g(w)$ cannot be directly estimated, as it cancels out of the local partial likelihood. As a result, only $\beta(\cdot)$ and $\dot{g}(\cdot)$, the derivative of $g(\cdot)$, can be directly estimated. Although $g(\cdot)$ can be estimated by subsequently integrating the estimate of $\dot{g}(\cdot)$, the large sample properties of this estimate are not formally established and statistical inference is not immediately available. Also, the local partial likelihood approach cannot handle a discrete exposure variable.

In this paper, we propose a global partial likelihood method for the estimation of β and g. The main idea is to derive estimating functions directly from Cox's partial likelihood rather than from local partial likelihood. All observations are used for the estimation of $\beta(w)$ and $g(w)$. The superiority of this method is reflected in its semiparametric efficiency in terms of linear functionals. The coefficient functions β and g and their derivatives are estimated directly and simultaneously. This approach reduces to Cox's partial likelihood approach when W is discrete. We show that the proposed estimates are uniformly consistent and asymptotically normal. The optimal bandwidth is found to be at the order of $n^{-1/5}$, the same as in classical nonparametric regression.

2. PARTIAL LIKELIHOOD ESTIMATION

2·1. *Notation and related estimation*

Let T_i and C_i denote the failure time and censoring time of subject i. Let $\mathcal{T}_i = \min(T_i, C_i)$ be the observed event time and Δ_i be the failure/censoring index. Assume that T_i and C_i are conditionally independent of each other given covariates Z_i and W_i. The observations are $(\mathcal{T}_i, \Delta_i, Z_i, W_i)$ for $i = 1, \ldots, n$, where $Z_i = (Z_{i1}, \ldots, Z_{ip})'$ and W_i are two types of covariates. Consider the varying coefficient model (1). Following Fan et al. (2006), for ease of presentation, we drop the dependence of covariates on time t, with the understanding that the methods and proofs in this paper are applicable to time-dependent covariates. The partial likelihood for model (1) is

$$L\{\beta(\cdot), g(\cdot)\} = \prod_{i=1}^{n} \left[\frac{\exp\{\beta(W_i)'Z_i + g(W_i)\}}{\sum_{j \in \mathcal{R}(\mathcal{T}_i)} \exp\{\beta(W_j)'Z_j + g(W_j)\}} \right]^{\Delta_i}, \tag{2}$$

where $R(t) = \{i : \mathcal{T}_i \geq t\}$ is the risk set at time t.

Suppose all components of $\beta(u)$ and $g(u)$ are unspecified and smooth. For every given w and u close to w, Taylor expansion implies

$$\begin{aligned} \beta(u) &\approx \beta(w) + \dot{\beta}(w)(u - w) \equiv \delta + \eta(u - w), \\ g(u) &\approx g(w) + \dot{g}(w)(u - w) \equiv \alpha + \gamma(u - w), \end{aligned} \tag{3}$$

where $\dot{g}(u) = dg(u)/du$. Let $N_i(t) = I(T_i \le t, \Delta_i = 1)$, $Y_i(t) = I(\mathcal{T}_i \ge t)$,

$$\zeta = (\delta', h\eta', h\gamma')' \text{ and } V_i = \{Z_i', Z_i'(W_i - w)/h, (W_i - w)/h\}'.$$

Substituting (3) into (2), Fan et al. (2006) consider nonparametric estimation based on the local partial likelihood score equation:

$$\sum_{i=1}^n \int_t K_i(w)\Big\{V_i - \frac{S_1(t,\zeta)}{S_0(t,\zeta)}\Big\} dN_i(t) = 0, \tag{4}$$

where $S_m(t,\zeta) = \sum_{j=1}^n Y_j(t) V_j^{\otimes m} \exp(\zeta' V_j) K_j(w)$, $V_j^{\otimes 0} = 1$, $V_j^{\otimes 1} = V_j$, $K_i(w) = K_h(W_i - w)$, $K_h(\cdot) = K(\cdot/h)/h$, K is a kernel function, and h represents the size of the local neighborhood. Note, however, that $g(w)$ is not directly estimable because (4) does not involve the intercept $\alpha = g(w)$, which cancels out. Furthermore, if W is a discrete random variable taking a finite number of values, a small window around w would only contain itself. Then (4) reduces to an equation which does not involve $g(\cdot)$. Thus the local likelihood estimator is not applicable when W is a discrete random variable.

2·2. *Global partial likelihood estimation*

The motivation of global partial likelihood is quite straightforward. For every fixed w in the range of W, suppose that β and g are known outside a neighborhood $B_n(w)$ of w. Set

$$\xi = (\delta', \alpha, h\eta', h\gamma')' \text{ and } X_i = \{Z_i', 1, Z_i'(W_i - w)/h, (W_i - w)/h\}'.$$

Then one can consider maximizing

$$\prod_{i=1}^n \left[\frac{I_i \exp(\xi' X_i) + (1 - I_i)\exp(\psi_i)}{\sum_{j\in\mathcal{R}(\mathcal{T}_i)} \{I_j \exp(\xi' X_j) + (1 - I_j)\exp(\psi_j)\}} \right]^{\Delta_i}, \tag{5}$$

where $\psi_j = \beta(W_j)' Z_j + g(W_j)$ and I_i equals 1 if $W_i \in B_n(w)$ and 0 otherwise. Replacing I_i with a kernel function K_i that decreases smoothly to zero, we obtain

$$\prod_{i=1}^n \left(\frac{K_i(w)\exp(\xi' X_i) + \{1 - K_i(w)\}\exp(\psi_i)}{\sum_{j\in\mathcal{R}(\mathcal{T}_i)} [K_j(w)\exp(\xi' X_j) + \{1 - K_j(w)\}\exp(\psi_j)]} \right)^{\Delta_i}. \tag{6}$$

Since the true $\beta(\cdot)$ and $g(\cdot)$ are unknown, the objective function (6) is not directly useful. The key idea of our method is to use an iterative algorithm and replace the unknown $\beta(\cdot)$ and $g(\cdot)$ with values from the previous step of the iteration. That is, we may estimate ξ in step m by maximizing

$$\prod_{i=1}^n \left(\frac{K_i(w)\exp(\xi' X_i) + \{1 - K_i(w)\}\exp(\psi_i^{(m-1)})}{\sum_{j\in\mathcal{R}(\mathcal{T}_i)} \left[K_j(w)\exp(\xi' X_j) + \{1 - K_j(w)\}\exp(\psi_j^{(m-1)})\right]} \right)^{\Delta_i}, \tag{7}$$

where $\psi_j^{(m-1)} = \beta^{(m-1)}(W_j)' Z_j + g^{(m-1)}(W_j)$. We differentiate the logarithm of (7) with respect to ξ and solve the following equations for ξ:

$$\sum_{i=1}^n \Delta_i \left(K_i(w) X_i - \frac{\sum_{j\in\mathcal{R}(\mathcal{T}_i)} K_j(w)\exp(\xi' X_j) X_j}{\sum_{j\in\mathcal{R}(\mathcal{T}_i)} \left[K_j(w)\exp(\xi' X_j) + \{1 - K_j(w)\}\exp(\psi_j^{(m-1)})\right]} \right) = 0. \tag{8}$$

It follows from (5) that

$$\exp(\psi_j) = \{K_j(w) + 1 - K_j(w)\}\exp(\psi_j) \approx K_j(w)\exp(\xi' X_j) + \{1 - K_j(w)\}\exp(\psi_j).$$

We then consider the following set of equations for ξ:

$$\sum_{i=1}^{n}\Delta_i\left[K_i(w)X_i-\frac{\sum_{j\in\mathcal{R}(\mathcal{T}_i)}K_j(w)\exp(\xi'X_j)X_j}{\sum_{j\in\mathcal{R}(\mathcal{T}_i)}\exp(\psi_j^{(m-1)})}\right]=0, \tag{9}$$

which is the following global partial likelihood score, using the counting process notation,

$$\sum_{i=1}^{n}\int_0^{\tau}\left[X_iK_i(w)-\frac{S_{n1}(t;\xi)}{S_{n0}\left(t;\beta^{(m-1)},g^{(m-1)}\right)}\right]dN_i(t)=0, \tag{10}$$

with

$$S_{n0}(t;\beta,g)=\sum_{j=1}^{n}Y_j(t)\exp\{\beta(W_j)'Z_j+g(W_j)\},\ \ S_{n1}(t;\xi)=\sum_{j=1}^{n}Y_j(t)X_jK_j(w)\exp(\xi'X_j)$$

and $\tau=\infty$. For technical development, τ is often assumed to be finite to avoid the tail problem. As $g(\cdot)$ is identifiable up to a location shift, we set $g^{(m)}(W_n)=0$ for all $m\geq 0$ for notational and computational convenience. The iterative algorithm is formally presented in the following.

- Step 0. Choose initial values of $\beta^{(0)}(w)$ and $g^{(0)}(w)$ for $w=W_1,\ldots,W_n$. For example, one could choose the local partial likelihood estimates of $\beta(\cdot)$ and $g(\cdot)$ (Fan et al., 2006).
- Step m. For every given $w=W_1,\ldots,W_n$, obtain ξ by solving the equations (10), which involves $\beta^{(m-1)}(W_i)$ and $g^{(m-1)}(W_i)$, $1\leq i\leq n$. Let $\widehat{\xi}(w)=\left\{\widehat{\delta}(w)',\widehat{\alpha}(w),h\widehat{\eta}(w)',h\widehat{\gamma}(w)\right\}'$ be the solution of ξ. Then $\beta^{(m)}(W_i)=\widehat{\delta}(W_i)$ for $i=1,\ldots,n$ and $g^{(m)}(W_i)=\widehat{\alpha}(W_i)$ for $i=1,\ldots,n-1$.
- Repeat the above steps for $m=1,2,\ldots$ till $\beta^{(m)}(W_i)$ and $g^{(m)}(W_i)$, $1\leq i\leq n$, converge.
- For every w in the range of W, the estimates of $\beta(w)$, $\dot{\beta}(w)$, $g(w)$ and $\dot{g}(w)$, denoted by $\widehat{\beta}(w)$, $\widehat{\dot{\beta}}(w)$, $\widehat{g}(w)$ and $\widehat{\dot{g}}(w)$, respectively, are obtained by solving the equations (10) for δ,η,α and γ. This can be achieved by replacing $\beta^{(m-1)}(W_i)$ and $g^{(m-1)}(W_i)$ with their estimators,which are defined as the limits of the above iteration.

Suppose the exposure variable W takes finite values, $b_1,\ldots,b_K$. Assume $g(b_K)=0$ for identifiability. Then, as long as the bandwidth h is smaller than $\min\left(|b_k-b_l|:l\neq k\right)$, we have $W_i=W_j$ for $|W_i-W_j|\leq h$. As a result, (10) reduces to

$$\sum_{i=1}^{n}\Delta_i\left[Z_i^*I(W_i=b_k)-\frac{\sum_{j\in\mathcal{R}(\mathcal{T}_i)}Z_j^*I(W_j=b_k)\exp\{\beta(b_k)'Z_j+g(b_k)\}}{\sum_{j\in\mathcal{R}(\mathcal{T}_i)}\sum_{l=1}^{K}I(W_j=b_l)\exp\{\beta(b_l)'Z_j+g(b_l)\}}\right]=0,$$

where $Z_i^*=(Z_i',1)'$. These are the same as the Cox partial likelihood estimating equation. The solutions of $\beta(b_k)$ and $g(b_k)$ are the Cox partial likelihood estimates of $\beta(b_k)$ and $g(b_k)$. Hence, the proposed estimate reduces to the Cox partial likelihood estimate when W is discrete.

Comparing the proposed global partial likelihood score (10) with its local counterpart (4), we can see that the local partial likelihood score is based on the estimating equation for V_i using only local data, while the global partial likelihood score is based on the estimating equation for $X_iK_i(w)$ using all of the data. As expected, the global partial likelihood estimator is more efficient. Its semiparametric efficiency is presented in Theorem 3. In addition, the local partial likelihood estimation uses the kernel functions twice and cannot be extended to the Cox model with time-varying coefficients, in which the variable W is time. However, in this case, the global partial likelihood estimator based on (10) reduces to the Cox partial likelihood estimate, as considered by Cai & Sun (2003) and Tian et al. (2005) for the Cox model with time-varying coefficients. In this respect, the proposed method is more flexible than the local partial likelihood method.

When the variable W is time, Zucker & Karr (1990) use a penalized partial likelihood approach and Murphy & Sen (1991) use the sieve approach for estimation. Their methods are also global. However, the Zucker–Karr and Murphy–Sen estimators involve maximization of functions over a parameter space whose dimension increases with the sample size. This optimization problem can be rather complex. In addition, no guidelines are available for choosing the segments of the sieve method and the asymptotic distribution of the Zucker–Karr estimator is fully established only for the case with a single covariate (Tian et al., 2005). Further, the theoretical efficiency of these estimators has not been formally established. Martinussen et al. (2002) propose a one-step estimation for the cumulative parameter function $\int_0^t \beta(u)du$ of the semiparametric hazard model with time-varying regression coefficients. Using the fixed point theorem, the estimator is shown to be $n^{1/2}$-consistent and semiparametric efficient. However, the $n^{1/2}$-consistency is obtained only for the cumulative function, not for the covariate effects function itself.

3. Large sample properties

We now establish the uniform consistency and asymptotic normality of the global partial likelihood estimator. Without loss of generality, we fix $\widehat{g}(0) = g(0) = 0$ and assume that the support of W is $[0, 1]$. Additional regularity conditions are stated in the Appendix. The uniform consistency of $\widehat{\beta}(\cdot)$ and $\widehat{g}(\cdot)$ is presented in Theorem 1. The proofs of Theorems 1–3 are given in Part A of the Supplementary Material.

Theorem 1. *Under Conditions (C1)–(C7) stated in Appendix A, as $n \to \infty$, we have*

$$\sup_{0<w<1} \|\widehat{\beta}(w) - \beta(w)\| \to 0 \ \textit{and} \ \sup_{0<w<1} |\widehat{g}(w) - g(w)| \to 0$$

in probability.

Let $\theta(w) = \{\beta(w)', g(w)\}'$, $\widehat{\theta}(w) = \{\widehat{\beta}(w)', \widehat{g}(w)\}'$ and $Z^* = (Z', 1)'$. To express explicitly the asymptotic expression of the estimator $\widehat{\theta}(w) - \theta(w)$, we introduce some notation. Let

$$\mu_i = \int x^i K(x)dx, \ \ \nu_i = \int x^i K^2(x)dx, \ \ \mathrm{pr}(t \mid z, w) = \mathrm{pr}(\mathcal{T} \geq t \mid Z = z, W = w),$$

and $f(w \mid z)$ be the conditional density function of W given $Z = z$. Denote

$$\Gamma(z, w) = \int_0^\tau \mathrm{pr}(t \mid z, w)\lambda_0(t)dt, \ \ D(w) = E\left[Z^* Z^{*\prime}\Gamma(Z, w)\exp\{\theta(w)'Z^*\}f(w \mid Z)\right],$$

$$s_0(t; \delta) = E\left[\mathrm{pr}(t \mid Z, W)\exp\{\delta(W)'Z^*\}\right],$$

$$s_1(t; \delta, w) = E\left[Z^*\mathrm{pr}(t \mid Z, w)\exp\{\delta(w)'Z^*\}f(w \mid Z)\right]$$

and $\Psi(u; w) = \int_0^\tau s_1(t; \theta, w)s_1(t; \theta, u)'\lambda_0(t)/s_0(t; \theta)dt$.

THEOREM 2. *Under Conditions (C1)–(C7) stated in Appendix A, for* $0 < w < 1$, $\widehat{\theta}(w) - \theta(w)$ *satisfies the Fredholm integral equation*

$$\widehat{\theta}(w) - \theta(w) = D(w)^{-1} \int_0^1 \Psi(u; w)\{\widehat{\theta}(u) - \theta(u)\}du + (nh)^{-1/2}\Sigma_0(w)\varphi + \frac{1}{2}h^2\mu_2\ddot{\theta}(w) + o_p\{h^2 + (nh)^{-1/2}\}, \qquad (11)$$

where $\Sigma_0(w)\Sigma_0(w)' = \nu_0 D(w)^{-1}$, φ *is a standard normal random vector and* $\ddot{g}(w) = d^2g(w)/dw^2$.

Let $\mathcal{A}$ be the linear operator satisfying $\mathcal{A}(\phi)(w) = D(w)^{-1}\int_0^1 \Psi(u; w)\phi(u)du$ for any function ϕ. Let I be the identity operator. Then (11) can be written as

$$(I - \mathcal{A})(\widehat{\theta} - \theta)(w) = (nh)^{-1/2}\Sigma_0(w)\varphi + \frac{1}{2}h^2\mu_2\ddot{\theta}(w) + o_p\{h^2 + (nh)^{-1/2}\}. \qquad (12)$$

We prove $(I - \mathcal{A})^{-1}$ exists and is bounded in Part A of the Supplementary Material. Then, (12) yields

$$\widehat{\theta}(w) - \theta(w) = (nh)^{-1/2}(I - \mathcal{A})^{-1}(\Sigma_0)(w)\varphi + \frac{1}{2}h^2\mu_2(I - \mathcal{A})^{-1}(\ddot{\theta})(w) + o_p\{h^2 + (nh)^{-1/2}\}.$$

The order of the asymptotic bias of $\widehat{\theta}(w) - \theta(w)$ is h^2 and the order of the asymptotic variance is $(nh)^{-1}$. As a consequence, the theoretical optimal bandwidth is of order $n^{-1/5}$. Corollary 1 below follows from (12) in a straightforward fashion.

COROLLARY 1. *Under Conditions (C1)–(C7) stated in Appendix A, if* $nh^7 \to 0$,

$$(nh)^{1/2}\{\widehat{\theta}(w) - \theta(w) - \frac{1}{2}h^2\mu_2(I - \mathcal{A})^{-1}(\ddot{\theta})(w)\} \to N(0, \Sigma(w)), \qquad (13)$$

for $w \in (0, 1)$, *where* $\Sigma(w) = \{(I - \mathcal{A})^{-1}(\Sigma_0)(w)\}\{(I - \mathcal{A})^{-1}(\Sigma_0)(w)\}'$.

When $Z = 0$, our model reduces to the nonparametric Cox model considered by Chen et al. (2010) and the asymptotic expression (11) reduces to that in Theorem 2 of Chen et al. (2010). The difference between the varying coefficient Cox model and the nonparametric Cox model is the same as that between the varying coefficient model and the simple nonparametric regression model. The varying coefficient Cox model, compared with the nonparametric Cox model, is more practical and more complex. It allows covariates to interact nonlinearly with another exposure variable. However, due to the presence of Z, the closed forms used in Chen et al. (2010) are no longer available. Based on our current study, the global method can be extended to most of the cases where local linear smoothing is valid and the global method performs generally better. In this sense, the global method can be regarded as a general methodology, although theoretical justification is still made on a case-by-case basis.

For any function $\phi(\cdot)$ defined on $(0, 1)$ with a continuous second derivative, let $\int_0^1 \phi(w)'\widehat{\theta}(w)dw$ be an estimator of $\int_0^1 \phi(w)'\theta(w)dw$. Optimality of this estimator is shown in Theorem 3.

THEOREM 3. *Assume Conditions (C1)–(C6) stated in Appendix A hold. If* $nh^4 \to 0$ *and* $nh^2/(\log n)^2 \to \infty$, *then* $\int_0^1 \phi(w)'\widehat{\theta}(w)dw$ *is an efficient estimator of* $\int_0^1 \phi(w)'\theta(w)dw$.

Estimation of $\int_0^1 \phi(w)'\theta(w)dw$ at the rate $n^{-1/2}$ requires undersmoothing with $h = o(n^{-1/4})$ to ensure that the bias is of order $o(n^{-1/2})$. The necessity of undersmoothing to obtain root-

n-consistent estimation is standard in nonparametric regression; see, for example, Carroll et al. (1997) and Hastie & Tibshirani (1990).

With estimators of β and g, we can use kernel smoothing (Fan et al., 2006) to estimate the baseline hazard function by $\widehat{\lambda}_0(t) = \int K_b(t-u)d\widehat{\Lambda}_0(u)$, where b is a given bandwidth and

$$\widehat{\Lambda}_0(t) = \frac{1}{n}\sum_{i=1}^n \int_0^t \frac{dN_i(u)}{n^{-1}\sum_{j=1}^n Y_j(u)\exp\{\widehat{\beta}(W_j)'Z_j + \widehat{g}(W_j)\}}. \quad (14)$$

Using Corollary 1 and the proof of Fan, et al. (2006), we can show that $\widehat{\lambda}_0(t)$ and $\widehat{\Lambda}_0(t)$ are uniformly consistent on $(0,\tau)$.

4. Estimation of the fixed and varying regression coefficients

In practice, effects of some covariables may be fixed and a mixed model with both fixed and varying coefficients is desirable. Martinussen et al. (2002) and Tian et al. (2005) examine a mixed model in the setting of the traditional time-dependent coefficient Cox model where $\beta(\cdot)$ is a function of the time t, and McKeague & Sasieni (1994) study a similar mixed model in the setting of an additive hazard model. Consider a mixed model

$$\lambda(t) = \lambda_0(t)\exp\{\alpha' Z_1(t) + \beta(W)'Z_2(t) + g(W)\}. \quad (15)$$

Again, for the purpose of identifiability, we set $g^{(m)}(W_n) = 0$ for all $m \geq 0$. Based on the idea of the global partial likelihood, we estimate α, $\beta(\cdot)$ and $g(\cdot)$ using the following iterative algorithm.

- Step 0. Choose initial values of functions $\alpha^{(0)}$, $\beta^{(0)}(w)$ and $g^{(0)}(w)$ for $w = W_1, \ldots, W_n$.
- Step m, Part A. For every given $w = W_1, \ldots, W_n$, solve the following equations for $\xi = \{\beta(w)', g(w), h\dot{\beta}(w)', h\dot{g}(w)\}'$:

$$\sum_{i=1}^n \Delta_i \left[K_i(w)U_i - \frac{\sum_{j\in\mathcal{R}(\mathcal{T}_i)} K_j(w)\exp\{\alpha^{(m-1)\prime}Z_{1j} + \xi' U_j\}U_j}{\sum_{j\in\mathcal{R}(\mathcal{T}_i)}\exp\{\alpha^{(m-1)\prime}Z_{1j} + \beta^{(m-1)}(W_j)'Z_{2j} + g^{(m-1)}(W_j)\}}\right] = 0, \quad (16)$$

  where $U_i = \{Z_{2i}', 1, Z_{2i}'(W_i - w)/h, (W_i - w)/h\}'$. Let $\widehat{\xi}(w) = \left\{\widehat{\beta}(w)', \widehat{g}(w), h\widehat{\dot{\beta}}(w)', h\widehat{\dot{g}}(w)\right\}'$ be the solutions of ξ. Then $\beta^{(m)}(W_i) = \widehat{\beta}(W_i)$ for $i = 1, \ldots, n$ and $g^{(m)}(W_i) = \widehat{g}(W_i)$ for $i = 1, \ldots, n-1$.
- Step m, Part B. Solve the following equations for α:

$$\sum_{i=1}^n \Delta_i \left[Z_{1i} - \frac{\sum_{j\in\mathcal{R}(\mathcal{T}_i)} Z_{1j}\exp\{\alpha' Z_{1j} + \beta^{(m)}(W_j)'Z_{2j} + g^{(m)}(W_j)\}}{\sum_{j\in\mathcal{R}(\mathcal{T}_i)}\exp\{\alpha' Z_{1j} + \beta^{(m)}(W_j)'Z_{2j} + g^{(m)}(W_j)\}}\right] = 0. \quad (17)$$

- Repeat the above steps for $m = 1, 2, \ldots$ till $\alpha^{(m)}$, $\beta^{(m)}(W_i)$ and $g^{(m)}(W_i)$, $1 \leq i \leq n$, converge.
- For every w in the range of W, the estimates of $\beta(w)$, $\dot{\beta}(w)$, $g(w)$ and $\dot{g}(w)$, denoted by $\widehat{\beta}(w)$, $\widehat{\dot{\beta}}(w)$, $\widehat{g}(w)$ and $\widehat{\dot{g}}(w)$, respectively, are obtained by solving the equations (16) for ξ by replacing $\alpha^{(m-1)}$, $\beta^{(m-1)}(W_i)$ and $g^{(m-1)}(W_i)$ with their estimators defined as the limits of the above iteration.

Following similar arguments to those for Theorems 1–2, we have the following Theorems 4–6. The proof of Theorem 4 is analogous to that of Theorem 1. The proofs of Theorems 5 and 6 are given in Part B of the Supplementary Material.

THEOREM 4. *Under Conditions (C1)–(C7) stated in Appendix A, with* $Z = (Z_1', Z_2')'$, *we have*

$$\|\widehat{\alpha} - \alpha\| \to 0, \quad \sup_{0<w<1} \|\widehat{\beta}(w) - \beta(w)\| \to 0 \;\; \textit{and} \;\; \sup_{0<w<1} |\widehat{g}(w) - g(w)| \to 0$$

in probability.

THEOREM 5. *Under the conditions of Theorem 4, if* $nh^4 = o(1)$, *then as* $n \to \infty$

$$n^{1/2} (\widehat{\alpha} - \alpha) \to N\{0, A^{-1}B(A^{-1})'\},$$

where A *and* B *are defined in Appendix B.*

THEOREM 6. *Under the conditions of Theorem 4, for* $0 < w < 1$, *we have the Fredholm integral equation,*

$$D(w)\left\{\widehat{\theta}(w) - \theta(w)\right\} = \int_0^1 \Psi(v;w)\{\widehat{\theta}(v) - \theta(v)\}dv$$
$$+(nh)^{-1/2}\Sigma(w)\varphi + \frac{\mu_2 h^2}{2} D(w)\ddot{\theta}(w) + o_p\{h^2 + (nh)^{-1/2}\},$$

where $\theta(w) = \{\beta(w), g(w)\}'$, $\Sigma(w)\Sigma(w)' = \nu_0 D(w)$, $D(w)$ *and* $\Psi(v;w)$ *are defined in Appendix B, and* φ *is a standard normal random vector.*

Theorem 5 shows that $\widehat{\alpha}$ is a $n^{1/2}$-consistent and asymptotically normal estimator of α. For any function $\phi(w) = \{\phi_1', \phi_2(w)'\}'$, which has a continuous second derivative on $[0, \tau]$, let $\phi_1'\widehat{\alpha} + \int_0^\tau \phi_2'(w)\widehat{\theta}(w)dw$ be the proposed estimator of $\phi_1'\alpha + \int_0^\tau \phi_2'(w)\theta(w)dw$. Theorem 7 below shows that $\widehat{\alpha}$ is also an efficient estimator of α.

THEOREM 7. *Under the conditions of Theorem 4, if* $nh^4 = o(1)$, *then, as* $n \to \infty$, $\phi_1'\widehat{\alpha} + \int_0^\tau \phi_2'(w)\widehat{\theta}(w)dw$ *is an efficient estimator of* $\phi_1'\alpha + \int_0^\tau \phi_2'(w)\theta(w)dw$.

Theorem 7 implies that, by taking $\phi_2(w) \equiv 0$, $\widehat{\alpha}$ is an efficient estimator of α and that, by taking $\phi_1 = 0$, $\int_0^\tau \phi_2'(w)\widehat{\theta}(w)dw$ is an efficient estimator of $\int_0^\tau \phi_2'(w)\theta(w)dw$.

In practice, we need to investigate which components have varying and which have fixed effects. For this purpose, we consider testing a hypothesis that $\beta_j(\cdot)$ is a constant against the alternative that it is not. Using an idea of Martinussen et al. (2002), we propose a test statistic

$$Q = n^{1/2} \sup_{w\in(0,1)} \left| \int_0^w u(s)\widehat{\beta}_j(s)ds - \beta_j^* \int_0^w u(s)ds \right|, \tag{18}$$

where $u(\cdot)$ is a given weight function, $\widehat{\beta}_j(\cdot)$ is the estimator of $\beta_j(\cdot)$ under the alternative hypothesis, while β_j^* is computed under the null hypothesis. The limit distribution of Q can be simulated by the bootstrap method.

5. NUMERICAL EXAMPLES

5·1. *Simulations*

In this section we investigate the performance of the proposed global partial likelihood estimator. The performance of estimator $\widehat{\beta}(\cdot)$ is assessed via the weighted mean squared errors, WMSE $= n_g^{-1} \sum_{j=1}^{p} \sum_{k=1}^{n_g} a_j \{\widehat{\beta}_j(w_k) - \beta_j(w_k)\}^2$, where $w_k, k = 1, \ldots, n_g$ are the grid points at which the functions $\beta(\cdot)$ are estimated. In the following examples, the Epanechnikov kernel will be used, $n_g = 200$ and a_j is reciprocal of the sample variance of $\{\widehat{\beta}_j(w_k)\}$. We adopt the same setting as Fan et al. (2006) have and consider a varying-coefficient model, $\lambda(t) = 4t^3 \exp[b\{Z_1(t), Z_2, W\}]$, with

$$b\{Z_1(t), Z_2, W\} = 0.5W(1.5 - W)Z_1(t) + \sin(2W)Z_2 + 0.5\{\exp(W - 1.5) - \exp(-1.5)\},$$

where W is a random variable uniformly distributed on $[0, 3]$, the covariate $Z_1(t)$ is time-dependent and is defined as $Z_1(t) = Z_1 I(t \leq 1)/4 + Z_1 I(t > 1)$, and Z_1 and Z_2 are jointly normal with correlation 0.5, each with mean 0 and standard deviation 5. The censoring variable C given (Z_1, Z_2, W) is distributed uniformly on $[0, a(Z_1, Z_2, W)]$, where

$$a(Z_1, Z_2, W) = c_1 I(b(Z_1, Z_2, W) > b_0) + c_2 I(b(Z_1, Z_2, W) \leq b_0)$$

with b_0 being the mean function of $b(Z_1, Z_2, W)$. The constants $c_1 = 0.8$ and $c_2 = 20$ are chosen so that about 30%-40% of the data is censored in each region of the function $a(\cdot)$.

We conducted 200 simulations with a sample size of 300. Figure 1 depicts the distribution of the weighted mean squared errors over the 200 replications, using the proposed global partial likelihood estimator with bandwidth $h = 0.2, 0.3, 0.4, 0.5, 0.6, 1$ and the local partial likelihood estimator with bandwidth $h = 0.3, 0.4, 0.5, 0.6, 0.7, 1$. It is evident that the minimum weighted mean squared error of the global partial likelihood estimator is smaller than that of the local partial likelihood estimator, suggesting that the global partial likelihood estimator is better. The optimal bandwidth for the global partial likelihood estimator is smaller than that for the local partial likelihood estimator. This is likely due to the following. Given h, we can see from a comparison of (4) and (10) that the amount of data used by the global partial likelihood estimator is more than that by the local partial likelihood estimator. The local partial likelihood estimator needs to compensate for its lower data usage by enlarging the included range of data. In support of this conclusion, Table 1 presents the empirical standard deviations of 200 estimated values of $\widehat{g}(w)$, $\widehat{\beta}_1(w)$ and $\widehat{\beta}_2(w)$ using the global partial likelihood estimator with $h = 0.3$ and the local partial likelihood estimator with $h = 0.6$. Here $h = 0.3$ and $h = 0.6$ are the optimal bandwidths for the global and local partial likelihood estimators, respectively. We take $w = 0.3, 0.75, 1.5, 2.25$ and 2.7, which correspond to the 10th, 25th, 50th, 75th and 90th percentiles of the distribution of W. From Table 1, we can see that the variance of the global partial likelihood estimator is smaller than that of the local partial likelihood estimator in most of the cases. These results also suggest that the global partial likelihood estimator is more efficient than the local partial likelihood estimator.

The calculation of the standard errors is carried out using bootstrap resampling, in which each subject is treated as a resampling unit in order to preserve the inherent feature of the data. In our simulation studies with bootstrap sample size over 150, the bootstrap standard errors are stable and generally close to the empirical standard errors, suggesting that the bootstrap as a resampling method for estimating the standard errors is reasonably adequate.

5·2. *Real example*

The proposed approach is applied to the nursing home data described and analyzed by Morris et al. (1994). The data were obtained from an experiment sponsored by the National Center for Health Services Research between 1980 and 1982, involving 36 for-profit nursing homes in San Diego, California. The sample size is 1601. The study was designed to evaluate the effects of different financial incentives on the durations of stay, among other things. This motivated Morris

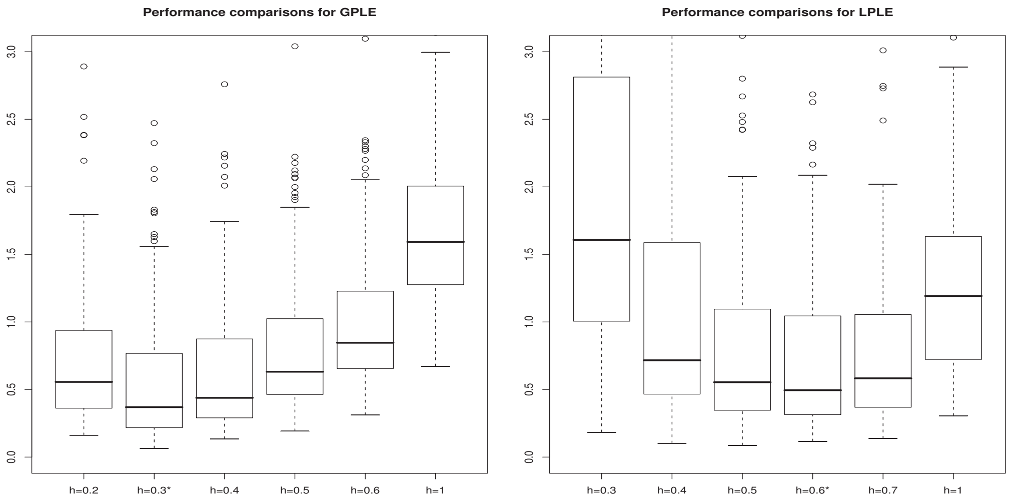

Figure 1: Boxplots of the distribution of the weighted mean squared errors over 200 replications using (a) the global partial likelihood estimator (GPLE) with bandwidth $h = 0.2, 0.3, 0.4, 0.5, 0.6, 1$ and (b) the local partial likelihood estimator (LPLE) with bandwidth $h = 0.3, 0.4, 0.5, 0.6, 0.7, 1$. * stands for the minimum of weighted mean squared errors.

Table 1. *Standard deviations of the global partial likelihood estimator (GPLE) and the local partial likelihood estimator (LPLE) with bandwidths 0.3 and 0.6.*

		$h = 0.3$				
Function	Method	w =0.3	0.75	1.5	2.25	2.70
$g(w)$	GPLE	0.300	0.373	0.361	0.371	0.394
	LPLE	0.633	0.754	0.841	0.930	1.005
$\beta_1(w)$	GPLE	0.061	0.076	0.079	0.110	0.157
	LPLE	0.090	0.128	0.145	0.226	0.468
$\beta_2(w)$	GPLE	0.050	0.059	0.045	0.061	0.064
	LPLE	0.117	0.244	0.063	0.199	0.190
		$h = 0.6$				
Function	Method	w =0.3	0.75	1.5	2.25	2.70
$g(w)$	GPLE	0.195	0.262	0.257	0.267	0.301
	LPLE	0.297	0.448	0.445	0.465	0.487
$\beta_1(w)$	GPLE	0.042	0.035	0.038	0.062	0.108
	LPLE	0.065	0.054	0.063	0.120	0.270
$\beta_2(w)$	GPLE	0.033	0.037	0.023	0.036	0.048
	LPLE	0.077	0.090	0.040	0.100	0.106

et al. (1994) to take the duration of stay in days T as the response variable and use the Cox model with treatment indicator x_1, gender variable x_2, marital status indicator x_3, three binary health status indicators x_4, x_5, x_6, and age x_7 as the covariates. The treatment $x_1 = 1$ if treated at a nursing home and 0 otherwise. The gender $x_2 = 1$ for male and 0 for female. The marital status $x_3 = 1$ if married and 0 otherwise. The health status x_4, x_5, x_6 correspond to very healthy,

healthy and not healthy. The age x_7 ranges from 65 to 104. To explore possible interactions, Fan & Li (2001) add interaction terms such as $x_7x_1, x_7x_2, \ldots$. Fan et al. (2006) consider the more general model,

$$\lambda(t, x) = \lambda_0(t) \exp\left\{\sum_{j=1}^{6} \beta_j(x_7)x_j + g(x_7)\right\}.$$

This permits us to examine how age interacts with covariates such as treatment, gender, marital status, among others. Fan et al. (2006) use the penalized likelihood method with smoothly clipped absolute deviation to show that treatment and marital status are not significant. We fit the data to the model

$$\lambda(t, y) = \lambda_0(t) \exp\left\{\sum_{j=1}^{4} \beta_j(z_5)z_j + g(z_5)\right\},$$

where z_1 is a gender variable, the three binary health status indicators z_2, z_3, z_4 correspond to very healthy, healthy and not healthy, and z_5 is age.

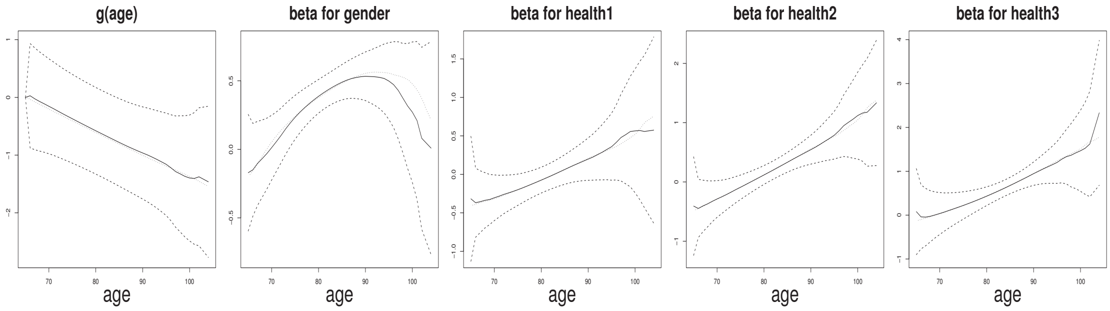

Figure 2: The estimated coefficient functions (solid) by the global partial likelihood method, their 95% confidence limits (dashed), and the estimated coefficient functions (dotted) by the local partial likelihood method for the nursing home data.

The global partial likelihood method and local partial likelihood method were applied to the dataset with bandwidths $h = 20$ and $h = 25$, respectively. The bandwidths $h = 20$ and $h = 25$ were chosen by *K*-fold cross-validation (Cai et al., 2000; Hoover et al., 1998; Fan et al., 2006) to minimize the prediction error

$$\int_0^{\tau} [\{N_i(t) - \widehat{E}\{N_i(t)\}]^2 d\{\sum_{k=1}^{n} N_k(t)\},$$

where $\widehat{E}\{N_i(t)\} = \int_0^t Y_i(u) \exp\{\widehat{\beta}(W_i)' Z_i(u) + \widehat{g}(W_i)\} d\widehat{\Lambda}_0(u)$ is the estimate of the expected number of failures up to time t. We chose $K = 20$ and obtained estimated coefficient functions and their 95% confidence bands as shown in Figure 2. The results for the global partial likelihood estimator and the local partial likelihood estimator are similar. They demonstrate the extent to which the gender effect and the health effect vary with age. In particular, the gender effect is not significant for the patients aged below 73 or above 98, but among the patients aged between 73 and 98, the male ones are more likely to stay at nursing homes. This is because the patients aged below 73 are able to take care of themselves but the patients aged above 98 are not, regardless of whether they are male or female. However, among the patients aged between 73 and 98, the female patients may be able to take better care of themselves than the male ones. Figure 2 shows that the likelihood of staying in nursing homes increases as health declines.

ACKNOWLEDGEMENT

Chen's research was supported by grants from the Research Grant Council of Hong Kong. Lin's research was supported partially by the National Natural Science Foundation of China. Zhou's work was supported by the National Natural Science Foundation of China. A supplementary material of this paper is posted at http://huazhenlin.ys168.com/.

APPENDIX A: CONDITIONS FOR THEOREMS

Let $\vartheta = (\delta', \alpha)'$ and $\mathcal{C}_0 = \{\vartheta(w) : w \in [0,1], \alpha(0) = 0, \delta(w)$ and $\alpha(w)$ are continuous on $[0,1]\}$.

Condition:

(C1). The kernel function $K(\cdot)$ is a symmetric density function with compact support $[-1, 1]$ and bounded derivative.

(C2). Variables W and Z are bounded with compact support $[0,1]$ and $[0,1]^p$, respectively, and $\mathrm{pr}(C = 0 \mid W = w, Z = z) < 1$ for any $w \in [0,1]$ and $z \in [0,1]^p$.

(C3). The conditional density function $f(\cdot \mid z)$ and the density function $f(\cdot)$ are positive and have continuous second derivatives on $[0,1]$ for each $z \in [0,1]^p$.

(C4). The functions β and g have a continuous second derivative on the corresponding support.

(C5). The conditional probability $\mathrm{pr}(t \mid z, \cdot)$ is positive, and has a continuous second derivative on $[0,1]$ for each $t \in [0,\tau]$ and $z \in [0,1]^p$.

(C6). For any $\vartheta \in \mathcal{C}_0$, if $\vartheta(W)'Z^*$ is a constant almost surely, then $\vartheta(\cdot) \equiv 0$.

(C7). As $n \to \infty$, $h^2 \log(n) \to 0$ and $nh/\{\log(n)\}^2 \to \infty$.

These conditions are used for deriving the convergence properties, asymptotic representation and efficiency of $\widehat{\beta}(\cdot)$ and $\widehat{g}(\cdot)$. Conditions (C1)–(C5) are similar to those in Fan et al. (1997) and Fan et al. (2006). Condition (C6) is for the purpose of identifiability.

APPENDIX B: NOTATION FOR THEOREMS 5 AND 6

To express explicitly the asymptotical expression of $\widehat{\alpha} - \alpha$, $\widehat{\theta}(w) - \theta(w)$ and $\widehat{g}(w) - g(w)$, we introduce some necessary notation. Denote $\mathrm{pr}(u \mid Z_{1i}, Z_{2i}, W_i) = \mathrm{pr}(\mathcal{T}_i \geq u \mid Z_{1i}, Z_{2i}, W_i)$, $Z_i = (Z'_{1i}, Z'_{2i})'$, $Z^*_{2i} = (Z'_{2i}, 1)'$, $\theta(w) = \{\beta'(w), g(w)\}'$,

$$\Gamma(z,w) = \int_0^\tau \mathrm{pr}(u \mid z, w)\lambda_0(u)du, \;\; \Gamma_i(u,\gamma,\vartheta) = \mathrm{pr}(u \mid Z_{1i}, Z_{2i}, W_i)\exp\left\{Z'_{1i}\gamma + Z^{*\prime}_{2i}\vartheta(W_i)\right\},$$

$$D(w) = E\left[Z^*_{2i}Z^{*\prime}_{2i}\Gamma(Z_i,w)\exp\{Z'_{1i}\alpha + \theta(w)'Z^*_{2i}\}f(w \mid Z_i)\right],$$

$$s_{r0}(u,\gamma,\vartheta) = E\left\{\Gamma_i(u,\gamma,\vartheta)Z_{1i}^{\otimes r}\right\} \quad \text{for} \quad r = 0, 1 \text{ and } 2,$$

$$s_{0r}(u,\gamma,\vartheta,w) = E\left\{\Gamma_i(u,\gamma,\vartheta)Z_{2i}^{*\otimes r} \mid W_i = w\right\}f(w) \quad \text{for} \quad r = 1 \text{ and } 2,$$

$$s_{11}(u,\gamma,\vartheta,w) = E\left\{\Gamma_i(u,\gamma,\vartheta)Z^*_{2i}Z'_{1i} \mid W_i = w\right\}f(w),$$

$$\Psi(v;w) = \int_0^\tau \frac{s_{01}(u,\alpha,\theta,w)s'_{01}(u,\alpha,\theta,v)}{s_{00}(u,\alpha,\theta)}\lambda_0(u)du,$$

$$\Xi_0 = \int_0^\tau \left\{ \frac{s_{10}(u,\alpha,\theta)s'_{10}(u,\alpha,\theta)}{s_{00}(u,\alpha,\theta)} - s_{20}(u,\alpha,\theta) \right\} \lambda_0(u)du,$$

$$\Xi_1(w) = \int_0^\tau \left\{ \frac{s_{10}(u,\alpha,\theta)s'_{01}(u,\alpha,\theta,w)}{s_{00}(u,\alpha,\theta)} - s'_{11}(u,\alpha,\theta,w) \right\} \lambda_0(u)du.$$

Let $G(w)$ satisfy the following integral equation:

$$\Xi_1(w) = -G(w)D(w) + \int_0^1 G(v)\Psi(v;w)dv \quad (0 < w < 1),$$

$s_{G(01)}(u) = E\left\{\Gamma_i(u,\alpha,\theta)G(W_i)Z_{2i}^*\right\}$, $\;A = \Xi_0 - \int_0^1 G(w)\Xi_1'(w)dw$ and

$$B = \int_0^\tau E\left[\left\{ Z_{2i}^* G(W_i) - \frac{s_{G(01)}(u)}{s_{00}(u)} \right\} - \left\{ Z_{1i} - \frac{s_{10}(u)}{s_{00}(u)} \right\} \right]^2 \Gamma_i(u,\alpha,\theta)\lambda_0(u)du.$$

REFERENCES

ANDERSEN, P. K., BORGAN, O., GILL, R. D. & KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. & WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* New York: Springer-Verlag.

CAI, J., FAN, J., ZHOU, H. & ZHOU, Y. (2007). Marginal hazard models with varying-coefficients for multivariate failure time data. *Ann. Statist.* **35**, 324–354.

CAI, Z. & SUN, Y. (2003). Local linear estimation for time-dependent coefficients in Cox'S regression models. *Scand. J. Statist.* **30**, 93–111.

CARROLL, R. J., FAN, J., GIJBELS, I. & WAND, M. P. (1997). Generalized partially linear single-index models. *J. Am. Statist. Assoc.* **92**, 477–489.

CHEN, K., GUO, S., SUN, L. & WANG, J. (2010). Global partial likelihood for nonparametric proportional hazards models. *J. Am. Statist. Assoc.* **105**, 750–760.

COX, D. R. (1972). Regression models and life tables (with discussion). *J. R. Statist. Soc.* B. **4**, 187–220.

COX, D. R., & OAKES, D. (1984). *Analysis of Survival Data*. London: Chapman and Hill.

FAN, J., GIJBELS, I. & KING, M. (1997). Local likelihood and local partial likelihood in hazard regression. *Ann. Statist.* **25**, 1661–1690.

FAN, J., LIN, H. & ZHOU, Y. (2006). Local partial-likelihood estimation for lifetime data. *Ann. Statist.* **34**, 290–325.

FLEMING, T. R. & HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

GAMERMAN, D. (1991). Markov chain monte carlo for dynamic generalized linear models. *Biometrika* **85**, 215–227.

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

HASTIE, T. J. & TIBSHIRANI, R. J. (1990). *Generalized Additive Models.* London: Chapman and Hall.

Hoover, D. R., Rice, J. A., Wu, C. O. & Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, **85**, 809–822.

KALBFLEISCH, J. D. & PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

MARTINUSSEN, T., SCHEIKE, T. H. & SKOVGAARD, I. M. (2000). Efficient estimation of fixed and time-varying covariates effects in multiplicative intensity models. *Scand. J. Statist.* **29**, 57–74.

MARZEC, L. & MARZEC, P.(1997). On fitting Cox's regression model with time-dependent coefficients. *Biometrika* **84**, 901–908.

MORRIS, C. N., NORTON, E. C. & ZHOU, X. H. (1994). Parametric duration analysis of nursing home usage. In *Case Studies in Biometry*, Ed. N. Lange, L. Ryan, L. Billard, D. Brillinger, L. Conquests and J. Greenhouse, pp. 231-248. New York: Wiley.

MURPHY, S. A. (1993). Testing for a time dependent coefficient in Cox's regression model. *Scand. J. Statist.* **20**, 35–50.

MURPHY, S. A. & SEN, P. K.(1991). Time-dependent coefficients in a Cox-type regression model. *Stochastic Process. Appl.* **39**, 153–180.

TIAN, L., ZUCKER, D. & WEI, L. J. (2005). On the Cox model with time-varying regression coefficients. *J. Am. Statist. Assoc.* **100**, 172–183.

ZUCKER, D. M. & KARR, A. F. (1990). Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *Ann. Statist.* **18**, 329–353.