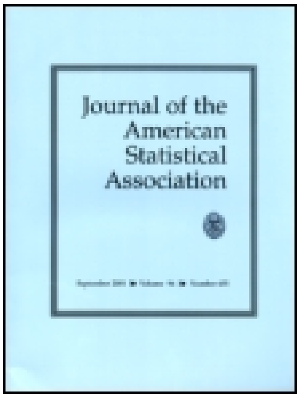

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://amstat.tandfonline.com/loi/uasa20

Partially Linear Additive Hazards Regression With Varying Coefficients

Guosheng Yin[a], Hui Li[a] & Donglin Zeng[a]

[a] Guosheng Yin is Associate Professor, Department of Biostatistics, M. D. Anderson Cancer Center, University of Texas, Houston, TX 77030 . Hui Li is Assistant Professor, School of Mathematical Sciences, Beijing Normal University, Beijing, China. Donglin Zeng is Associate Professor, Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599. The authors thank the editor, associate editor, and three referees for their insightful comments, which led to substantial improvements in the manuscript, and also thank Dr. Banu Arun for providing the breast cancer data. This work was supported in part by funds from the Physician Referral Service at M. D. Anderson Cancer Center and the U.S. Department of Defense grant W81XWH-05-2-0027. This work was conducted when Hui Li was visiting M. D. Anderson Cancer Center.
Published online: 01 Jan 2012.

To cite this article: Guosheng Yin, Hui Li & Donglin Zeng (2008) Partially Linear Additive Hazards Regression With Varying Coefficients, Journal of the American Statistical Association, 103:483, 1200-1213, DOI: 10.1198/016214508000000463

To link to this article: http://dx.doi.org/10.1198/016214508000000463

Partially Linear Additive Hazards Regression With Varying Coefficients

Guosheng YIN, Hui LI, and Donglin ZENG

To explore the nonlinear interactions between some covariates and an exposure variable, we propose the partially linear additive hazards model for survival data. In a semiparametric setting, we construct a local pseudoscore function to estimate the varying and constant coefficients and establish the asymptotic normality of the proposed estimators. Moreover, we develop the weak convergence property for the local estimator of the baseline cumulative hazard function. We conduct simulation studies to empirically examine the finite-sample performance of the proposed methods and use real data from a breast cancer study for illustration.

KEY WORDS: Asymptotic normality; Censored data; Estimating equation; Kernel function; Local polynomial; Semiparametric estimation; Varying-coefficient model.

1. INTRODUCTION

Varying-coefficient models have been extensively investigated in various contexts and are becoming standard statistical tools in many applications (see, e.g., Hoover, Rice, Wu, and Yang 1998; Cai, Fan, and Li 2000; Chiang, Rice, and Wu 2001; Huang, Wu, and Zhou 2002; Zhang 2004; Sun and Wu 2005; Martinussen and Scheike 2006). The proportional hazards model (Cox 1972) can be extended to enhance model flexibility by incorporating time-varying coefficients (see, e.g., Zucker and Karr 1990; Murphy and Sen 1991; Gamerman 1991; Hastie and Tibshirani 1993; Murphy 1993; Nielsen and Linton 1995; Marzec and Marzec 1997; Dabrowska 1997; Nielsen and Tanggaard 2001; Martinussen, Scheike, and Skovgaard 2002; Cai and Sun 2003; Tian, Zucker, and Wei 2005). Although time-varying coefficient models have attracted much attention, in many applications the covariate effects may vary with an exposure variable. This formulation is well suited for exploring the nonlinear interaction effects between risk factors (see Fan, Lin, and Zhou 2006). Fan, Gijbels, and King (1997) studied the nonparametric Cox model, in which the unknown risk function can be estimated by integrating its derivative. Chen and Zhou (2006) proposed to directly estimate the relative risk function by constructing the local partial likelihood around two sets of covariates. By selecting observations in the shrinking neighborhoods of two covariate values, the nonparametric risk function can be easily estimated, and its large-sample theories are rigorously derived.

Alternatively, the additive hazards model produces the risk difference as opposed to the risk ratio (e.g., Aalen 1989; Huffer and McKeague 1991; Lin and Ying 1994; McKeague and Sasieni 1994). Moreover, certain covariate effects may be much more complex than linear effects, which motivates simultaneously modeling the parametric and nonparametric components in the model. For subject i, let T_i be the failure time and C_i be the censoring time; then $X_i = T_i \wedge C_i$ is the observed time, where $a \wedge b$ takes the minimum value of a and b. Let $\Delta_i = I(T_i \le C_i)$ be the failure indicator, where $I(\cdot)$ is the indicator function. The corresponding possibly time-dependent covariates (external as defined by Kalbfleisch and Prentice 2002) are denoted by a p-vector $\mathbf{Z}_i(t)$, a q-vector $\mathbf{V}_i(t)$, and a scalar $W_i(t)$, where $\mathbf{Z}_i(t)$ may interact nonlinearly with the exposure variable $W_i(t)$. Assume that T_i and C_i are conditionally independent given the covariates and that the observed data $\{X_i, \Delta_i, \mathbf{Z}_i(t), \mathbf{V}_i(t), W_i(t), t \in [0,\tau]\}$ are independent and identically distributed (iid) for $i = 1, \ldots, n$, where τ is the end time of a study.

To characterize the varying-covariate effects of $\mathbf{Z}_i(t)$ with respect to $W_i(t)$, we propose the partially linear varying-coefficient additive hazards model

$$\lambda(t|\mathbf{Z}_i, \mathbf{V}_i, W_i) \\ = \lambda_0(t) + \boldsymbol{\beta}^T(W_i(t))\mathbf{Z}_i(t) + \boldsymbol{\gamma}^T\mathbf{V}_i(t) + \alpha(W_i(t)), \quad (1)$$

where $\lambda_0(t)$ is the baseline hazard function, $\boldsymbol{\beta}(W_i(t))$ characterizes the nonlinear interaction between $\mathbf{Z}_i(t)$ and $W_i(t)$, and $\alpha(W_i(t))$ represents the main effect of $W_i(t)$. For model identifiability, we set $\alpha(w_1) = 0$, where w_1 belongs to the interior of the support of $W_i(t)$ denoted by $\mathcal{W}$. Using the local polynomial technique (Fan and Gijbels 1996), we derive a local kernel-weighted estimator, which includes the pseudoscore estimator of Lin and Ying (1994) as a special case. We obtain an analytic solution for the estimator that overcomes the difficulties of numerical convergence and initial value selection (Fan and Chen 1999).

The rest of the article is organized as follows. In Section 2 we propose the local estimating equation under the varying-coefficient additive hazards model. In Section 3 we establish the asymptotic theories for the varying- and constant-coefficient estimators and the local estimator for the baseline cumulative hazard function. In Section 4 we examine the finite-sample properties using simulation studies and illustrate the proposed methods with a recent breast cancer data set. We give concluding remarks in Section 5, and delineate the proofs of our theorems in Appendix A.

Guosheng Yin is Associate Professor, Department of Biostatistics, M. D. Anderson Cancer Center, University of Texas, Houston, TX 77030 (E-mail: *gsyin@mdanderson.org*). Hui Li is Assistant Professor, School of Mathematical Sciences, Beijing Normal University, Beijing, China. Donglin Zeng is Associate Professor, Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599. The authors thank the editor, associate editor, and three referees for their insightful comments, which led to substantial improvements in the manuscript, and also thank Dr. Banu Arun for providing the breast cancer data. This work was supported in part by funds from the Physician Referral Service at M. D. Anderson Cancer Center and the U.S. Department of Defense grant W81XWH-05-2-0027. This work was conducted when Hui Li was visiting M. D. Anderson Cancer Center.

Journal of the American Statistical Association
September 2008, Vol. 103, No. 483, Theory and Methods
DOI 10.1198/016214508000000463

2. ESTIMATION PROCEDURES

Assume that $\boldsymbol{\beta}(\cdot)$ and $\alpha(\cdot)$ are smooth so that their first and second derivatives $\boldsymbol{\beta}'(\cdot)$, $\alpha'(\cdot)$, $\boldsymbol{\beta}''(\cdot)$, and $\alpha''(\cdot)$ exist. By the Taylor series expansion, for each given $w_0 \in \mathcal{W}$, we have that

$$\boldsymbol{\beta}(w) \approx \boldsymbol{\beta}(w_0) + \boldsymbol{\beta}'(w_0)(w - w_0)$$

and

$$\alpha(w) \approx \alpha(w_0) + \alpha'(w_0)(w - w_0).$$

Thus model (1) can be approximated by

$$\lambda(t|\mathbf{Z}_i, \mathbf{V}_i, W_i, w_0) = \lambda_0^*(t, w_0) + \boldsymbol{\xi}^T(w_0)\mathbf{Z}_i^*(t, w_0), \quad (2)$$

where $\lambda_0^*(t, w_0) = \lambda_0(t) + \alpha(w_0)$, $\boldsymbol{\xi}(w_0) = \{\boldsymbol{\beta}^T(w_0), \boldsymbol{\gamma}^T(w_0), (\boldsymbol{\beta}'(w_0))^T, \alpha'(w_0)\}^T$, and $\mathbf{Z}_i^*(t, w_0) = \{\mathbf{Z}_i^T(t), \mathbf{V}_i^T(t), \mathbf{Z}_i^T(t) \times (W_i(t) - w_0), (W_i(t) - w_0)\}^T$. We write $\boldsymbol{\gamma}(w_0)$, even though in our model $\boldsymbol{\gamma}$ is nonvarying, because in the sequel we will consider both local and global estimates of $\boldsymbol{\gamma}$. We write $N_i(t) = I(X_i \leq t, \Delta_i = 1)$ and $Y_i(t) = I(X_i \geq t)$ and define

$$\bar{\mathbf{Z}}(t, w_0) = \frac{\sum_{i=1}^n K_h(W_i(t) - w_0)Y_i(t)\mathbf{Z}_i^*(t, w_0)}{\sum_{i=1}^n K_h(W_i(t) - w_0)Y_i(t)},$$

where $K(\cdot)$ is a kernel density function, h is a bandwidth, and $K_h(\cdot) = K(\cdot/h)/h$. Motivated by the work of Lin and Ying (1994), we propose the local score-type function

$$\mathbf{U}_n(\boldsymbol{\xi}, w_0) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau K_h(W_i(t) - w_0) \times \{\mathbf{Z}_i^*(t, w_0) - \bar{\mathbf{Z}}(t, w_0)\}\, dM_i(t, w_0), \quad (3)$$

where $dM_i(t, w_0) = dN_i(t) - Y_i(t)\{\lambda_0^*(t, w_0) + \boldsymbol{\xi}^T(w_0)\mathbf{Z}_i^*(t, w_0)\}\, dt$.

If we denote the solution to $\mathbf{U}_n(\boldsymbol{\xi}, w_0) = \mathbf{0}$ by $\widehat{\boldsymbol{\xi}}(w_0)$, then we obtain an analytic closed form of

$$\widehat{\boldsymbol{\xi}}(w_0) = \left[\sum_{i=1}^n \int_0^\tau K_h(W_i(t) - w_0)Y_i(t) \times \{\mathbf{Z}_i^*(t, w_0) - \bar{\mathbf{Z}}(t, w_0)\}^{\otimes 2}\, dt\right]^{-1} \times \left[\sum_{i=1}^n \int_0^\tau K_h(W_i(t) - w_0) \times \{\mathbf{Z}_i^*(t, w_0) - \bar{\mathbf{Z}}(t, w_0)\}\, dN_i(t)\right], \quad (4)$$

where $\mathbf{a}^{\otimes k} = 1, \mathbf{a}$, and $\mathbf{a}\mathbf{a}^T$ for $k = 0, 1$, and 2. Note that the coefficient $\alpha(w_0)$ itself cannot be directly estimated, because it is incorporated into the local baseline hazard function $\lambda_0^*(t, w_0)$. But $\alpha(w_0)$ can be estimated by integrating $\widehat{\alpha}'(\cdot)$ over $\mathcal{W}$, based on the trapezoidal rule, and its confidence interval can be obtained by the usual bootstrap method (Efron and Tibshirani 1993). The local baseline cumulative hazard function, $\Lambda_0^*(t, w_0) = \int_0^t \lambda_0^*(u, w_0)\, du$, can be consistently estimated by

$$\widehat{\Lambda}_0^*(t, w_0) = \int_0^t \left(\sum_{i=1}^n K_h(W_i(u) - w_0) \times \left\{dN_i(u) - Y_i(u)\widehat{\boldsymbol{\xi}}^T(w_0)\mathbf{Z}_i^*(u, w_0)\, du\right\}\right) \Bigg/ \left(\sum_{i=1}^n K_h(W_i(u) - w_0)Y_i(u)\right). \quad (5)$$

This is a major generalization of the pseudoscore estimator of Lin and Ying (1994) to nonparametric regression, which is appealing because (4) nicely circumvents the convergence and other numerical challenges.

Because only the local data are used for estimating $\boldsymbol{\gamma}$ in (4), the resulting estimator $\widehat{\boldsymbol{\gamma}}(w_0)$ is not root-n consistent. To improve its convergence rate, we take

$$\widetilde{\boldsymbol{\gamma}} = \int_{\mathcal{W}} \boldsymbol{\Gamma}(w_0)\widehat{\boldsymbol{\gamma}}(w_0)\, dw_0, \quad (6)$$

where the weight matrix $\boldsymbol{\Gamma}(w_0)$ satisfies $\int_{\mathcal{W}} \boldsymbol{\Gamma}(w_0)\, dw_0 = \mathbf{I}_{q\times q}$, an identity matrix. We typically can choose $\boldsymbol{\Gamma}(w_0)$ to be the standardized inverse covariance matrix of $\widehat{\boldsymbol{\gamma}}(w_0)$ (see Tian et al. 2005).

3. ASYMPTOTIC THEORIES

3.1 Notation

Let $\mathbf{H}$ be a $(2p + q + 1)$-diagonal matrix, with the first $p + q$ elements equal to 1 and the remaining $p + 1$ elements equal to h. Let $\mu_j = \int u^j K(u)\, du$, $\nu_j = \int u^j K^2(u)\, du$, $P(t, \mathbf{Z}, \mathbf{V}, W) = \Pr(X \geq t|\mathbf{Z}(t), \mathbf{V}(t), W(t))$, and $\rho(t, \mathbf{Z}, \mathbf{V}, W) = P(t, \mathbf{Z}, \mathbf{V}, W)\lambda(t|\mathbf{Z}, \mathbf{V}, W)$ given the external time-dependent covariates. For $k = 0, 1$, and 2, we define

$$\mathbf{a}_k(t, w_0) = f_W(t, w_0)E\{P(t, \mathbf{Z}, \mathbf{V}, w_0)\mathbf{Z}^{\otimes k}(t)|W(t) = w_0\}$$

and

$$\mathbf{a}_k^*(t, w_0) = f_W(t, w_0)E\{\rho(t, \mathbf{Z}, \mathbf{V}, w_0)\mathbf{Z}^{\otimes k}(t)|W(t) = w_0\},$$

where $f_W(t, w_0)$ is the density function of $W(t)$ evaluated at w_0. Denote $\mathbf{a}_k(w_0) = \int_0^\tau \mathbf{a}_k(t, w_0)\, dt$ and $\mathbf{a}_k^*(w_0) = \int_0^\tau \mathbf{a}_k^*(t, w_0)\, dt$. For $k = 1$ and 2, we define

$$\mathbf{c}_k(t, w_0) = f_W(t, w_0)E\{P(t, \mathbf{Z}, \mathbf{V}, w_0)\mathbf{V}^{\otimes k}(t)|W(t) = w_0\},$$
$$\mathbf{c}_k^*(t, w_0) = f_W(t, w_0)E\{\rho(t, \mathbf{Z}, \mathbf{V}, w_0)\mathbf{V}^{\otimes k}(t)|W(t) = w_0\},$$
$$\mathbf{g}(t, w_0) = f_W(t, w_0) \times E\{P(t, \mathbf{Z}, \mathbf{V}, w_0)\mathbf{Z}(t)\mathbf{V}^T(t)|W(t) = w_0\},$$

and

$$\mathbf{g}^*(t, w_0) = f_W(t, w_0) \times E\{\rho(t, \mathbf{Z}, \mathbf{V}, w_0)\mathbf{Z}(t)\mathbf{V}^T(t)|W(t) = w_0\},$$

where $\mathbf{c}_k(w_0)$, $\mathbf{c}_k^*(w_0)$, $\mathbf{g}(w_0)$, and $\mathbf{g}^*(w_0)$ are defined similarly to $\mathbf{a}_k(w_0)$ and $\mathbf{a}_k^*(w_0)$. Finally, let

$$\begin{aligned}\boldsymbol{\omega}_{11}(w_0) = \nu_0 \int_0^\tau \Bigg\{ &\mathbf{a}_2^*(t,w_0) - \frac{\mathbf{a}_1^*(t,w_0)\mathbf{a}_1^T(t,w_0)}{a_0(t,w_0)} \\ &- \frac{\mathbf{a}_1(t,w_0)(\mathbf{a}_1^*(t,w_0))^T}{a_0(t,w_0)} \\ &+ \frac{\mathbf{a}_1^{\otimes 2}(t,w_0)a_0^*(t,w_0)}{a_0^2(t,w_0)}\Bigg\}\, dt, \\ \boldsymbol{\omega}_{12}(w_0) = \nu_0 \int_0^\tau \Bigg\{ &\mathbf{g}^*(t,w_0) - \frac{\mathbf{a}_1^*(t,w_0)\mathbf{c}_1^T(t,w_0)}{a_0(t,w_0)} \\ &- \frac{\mathbf{a}_1(t,w_0)(\mathbf{c}_1^*(t,w_0))^T}{a_0(t,w_0)} \\ &+ \frac{\mathbf{a}_1(t,w_0)\mathbf{c}_1^T(t,w_0)a_0^*(t,w_0)}{a_0^2(t,w_0)}\Bigg\}\, dt, \\ \boldsymbol{\omega}_{22}(w_0) = \nu_0 \int_0^\tau \Bigg\{ &\mathbf{c}_2^*(t,w_0) - \frac{\mathbf{c}_1^*(t,w_0)\mathbf{c}_1^T(t,w_0)}{a_0(t,w_0)} \\ &- \frac{\mathbf{c}_1(t,w_0)(\mathbf{c}_1^*(t,w_0))^T}{a_0(t,w_0)} \\ &+ \frac{\mathbf{c}_1^{\otimes 2}(t,w_0)a_0^*(t,w_0)}{a_0^2(t,w_0)}\Bigg\}\, dt,\end{aligned}$$

$$\boldsymbol{\Sigma}_{11}(w_0) = \nu_0 \begin{pmatrix} \boldsymbol{\omega}_{11}(w_0) & \boldsymbol{\omega}_{12}(w_0) \\ \boldsymbol{\omega}_{12}^T(w_0) & \boldsymbol{\omega}_{22}(w_0) \end{pmatrix},$$

$$\boldsymbol{\Sigma}_{22}(w_0) = \nu_2 \begin{pmatrix} \mathbf{a}_2^*(w_0) & \mathbf{a}_1^*(w_0) \\ (\mathbf{a}_1^*(w_0))^T & a_0^*(w_0) \end{pmatrix},$$

and

$$\boldsymbol{\Sigma}(w_0) = \operatorname{diag}(\boldsymbol{\Sigma}_{11}(w_0), \boldsymbol{\Sigma}_{22}(w_0)).$$

3.2 Asymptotic Properties

Let $\boldsymbol{\xi}_0(w_0) = \{\boldsymbol{\beta}_0^T(w_0), \boldsymbol{\gamma}_0^T(w_0), (\boldsymbol{\beta}_0'(w_0))^T, \alpha_0'(w_0)\}^T$ be the true parameter vector. The following theorems characterize the asymptotic properties of the proposed local estimator $\widehat{\boldsymbol{\xi}}(w_0)$.

Theorem 1. Under conditions (C.1)–(C.6) in the Appendix, we have that

$$\sqrt{nh}\big\{\mathbf{H}(\widehat{\boldsymbol{\xi}}(w_0) - \boldsymbol{\xi}_0(w_0)) - \tfrac{1}{2}h^2\mu_2\mathbf{D}^{-1}(w_0)\mathbf{b}(w_0)\big\} \xrightarrow{\mathcal{D}} \mathrm{N}\big(\mathbf{0}, \mathbf{D}^{-1}(w_0)\boldsymbol{\Sigma}(w_0)\mathbf{D}^{-1}(w_0)\big),$$

where $\mathbf{b}(w_0) = (\mathbf{b}_1(w_0)^T, \mathbf{b}_2(w_0)^T, \mathbf{0}_{p+1}^T)^T$ with $\mathbf{0}_{p+1}$ denoting a zero column vector of length $p+1$,

$$\mathbf{b}_1(w_0) = \int_0^\tau \left\{\mathbf{a}_2(t,w_0) - \frac{\mathbf{a}_1^{\otimes 2}(t,w_0)}{a_0(t,w_0)}\right\} dt\, \boldsymbol{\beta}_0''(w_0),$$

$$\mathbf{b}_2(w_0) = \int_0^\tau \left\{\mathbf{g}^T(t,w_0) - \frac{\mathbf{c}_1(t,w_0)\mathbf{a}_1^T(t,w_0)}{a_0(t,w_0)}\right\} dt \times \boldsymbol{\beta}_0''(w_0),$$

and

$$\mathbf{D}(w_0) = \operatorname{diag}\left(\begin{pmatrix} \int_0^\tau \big\{\mathbf{a}_2(t,w_0) - \frac{\mathbf{a}_1^{\otimes 2}(t,w_0)}{a_0(t,w_0)}\big\}\, dt & \int_0^\tau \big\{\mathbf{g}(t,w_0) - \frac{\mathbf{a}_1(t,w_0)\mathbf{c}_1^T(t,w_0)}{a_0(t,w_0)}\big\}\, dt \\ \int_0^\tau \big\{\mathbf{g}(t,w_0) - \frac{\mathbf{a}_1(t,w_0)\mathbf{c}_1^T(t,w_0)}{a_0(t,w_0)}\big\}^T dt & \int_0^\tau \big\{\mathbf{c}_2(t,w_0) - \frac{\mathbf{c}_1^{\otimes 2}(t,w_0)}{a_0(t,w_0)}\big\}\, dt \end{pmatrix}, \mu_2 \begin{pmatrix} \mathbf{a}_2(w_0) & \mathbf{a}_1(w_0) \\ \mathbf{a}_1^T(w_0) & a_0(w_0) \end{pmatrix}\right).$$

The proof is outlined in the Appendix. We define $\mathbf{G}_i^*(t,w_0) = \mathbf{H}^{-1}\mathbf{Z}_i^*(t,w_0)$ and

$$\bar{\mathbf{G}}(t,w_0) = \frac{\sum_{i=1}^n K_h(W_i(t)-w_0)Y_i(t)\mathbf{G}_i^*(t,w_0)}{\sum_{i=1}^n K_h(W_i(t)-w_0)Y_i(t)}.$$

To obtain a consistent estimator for the asymptotic variance of $\widehat{\boldsymbol{\xi}}(w_0)$, we replace $\mathbf{D}(w_0)$ and $\boldsymbol{\Sigma}(w_0)$ by their empirical counterparts. The covariance matrix of $\mathbf{H}\{\widehat{\boldsymbol{\xi}}(w_0) - \boldsymbol{\xi}_0(w_0)\}$ can be consistently estimated by $(nh)^{-1}\mathbf{D}_n^{-1}(w_0)\boldsymbol{\Sigma}_n(w_0)\mathbf{D}_n^{-1}(w_0)$, where

$$\mathbf{D}_n(w_0) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau K_h(W_i(t)-w_0)Y_i(t) \times \{\mathbf{G}_i^*(t,w_0) - \bar{\mathbf{G}}(t,w_0)\}^{\otimes 2}\, dt$$

and

$$\boldsymbol{\Sigma}_n(w_0) = \frac{h}{n}\sum_{i=1}^n \int_0^\tau K_h^2(W_i(t)-w_0) \times \{\mathbf{G}_i^*(t,w_0) - \bar{\mathbf{G}}(t,w_0)\}^{\otimes 2}\, dN_i(t).$$

The consistency of such variance estimate naturally follows from the proof of Theorem 1.

We can further obtain the asymptotic distribution of $\widehat{\Lambda}_0^*(t, w_0)$ as defined in (5).

Theorem 2. Under conditions (C.1)–(C.6) in the Appendix, $\sqrt{nh}\{\widehat{\Lambda}_0^*(t,w_0) - \Lambda_0^*(t,w_0) - h^2\mu_2\alpha''(w_0)t/2\}$ converges in distribution to a mean-zero Gaussian process, where the covariance function between time t and s can be consistently estimated by

$$\begin{aligned} &nh\sum_{i=1}^n \int_0^{t\wedge s} \frac{K_h^2(W_i(u)-w_0)\, dN_i(u)}{\{\sum_{i=1}^n K_h(W_i(u)-w_0)Y_i(u)\}^2} \\ &- \int_0^t \bar{\mathbf{G}}^T(u,w_0)\, du\, \mathbf{D}_n^{-1}(w_0)\boldsymbol{\eta}_n(s,w_0) \\ &- \boldsymbol{\eta}_n^T(t,w_0)\mathbf{D}_n^{-1}(w_0)\int_0^s \bar{\mathbf{G}}(u,w_0)\, du \\ &+ \int_0^t \bar{\mathbf{G}}^T(u,w_0)\, du\, \{\mathbf{D}_n^{-1}(w_0)\boldsymbol{\Sigma}_n(w_0)\mathbf{D}_n^{-1}(w_0)\} \\ &\times \int_0^s \bar{\mathbf{G}}(u,w_0)\, du, \end{aligned}$$

with

$$\boldsymbol{\eta}_n(t, w_0) = h \sum_{i=1}^{n} \int_0^t \frac{K_h^2(W_i(u) - w_0)}{\sum_{i=1}^n K_h(W_i(u) - w_0) Y_i(u)} \times \{\mathbf{G}_i^*(u, w_0) - \bar{\mathbf{G}}(u, w_0)\}\, dN_i(u).$$

We establish the asymptotic properties of the proposed estimator $\widetilde{\boldsymbol{\gamma}}$ by arguments similar to those of Tian et al. (2005). In particular, the following result holds.

Theorem 3. Under conditions (C.1)–(C.6) in the Appendix, assume that $\boldsymbol{\Gamma}(w_0)$ is twice-continuously differentiable. Then $\sqrt{n}(\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ converges in distribution to a mean-zero normal distribution.

The proof is given in the Appendix. The asymptotic covariance of $\widetilde{\boldsymbol{\gamma}}$ can be estimated as follows. We let $\boldsymbol{\mathcal{I}} = \{\mathcal{I}_{jk}\}$ be a $q \times (2p + q + 1)$ matrix with elements $\mathcal{I}_{jk} = 1$ for $j = 1, \ldots, q$, $k = p + j$, and $\mathcal{I}_{jk} = 0$ otherwise, and let $\mathbf{Z}_i^\dagger(t) = (\mathbf{Z}_i^T(t), \mathbf{V}_i^T(t), \mathbf{0}_{p+1}^T)^T$ and $\bar{\mathbf{Z}}^\dagger(t) = \sum_{i=1}^n Y_i(t)\mathbf{Z}_i^\dagger(t)/\sum_{i=1}^n Y_i(t)$. Then the limiting variance–covariance matrix can be consistently estimated by

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \boldsymbol{\Gamma}(W_i(t))\boldsymbol{\mathcal{I}}\mathbf{D}_n^{-1}(W_i(t))(\mathbf{Z}_i^\dagger(t) - \bar{\mathbf{Z}}^\dagger(t))^{\otimes 2} \times \mathbf{D}_n^{-1}(W_i(t))\boldsymbol{\mathcal{I}}^T\boldsymbol{\Gamma}^T(W_i(t))\, dN_i(t).$$

3.3 Confidence Bands

For varying-coefficient models, simultaneous confidence bands for the estimated coefficient functions are more desirable than the pointwise confidence intervals. Motivated by the work of Lin, Fleming, and Wei (1994) and Tian et al. (2005), we consider a stochastic perturbation of (3) by replacing $M_i(t, w_0)$ by $N_i(t)\psi_i$, that is,

$$\widetilde{\mathbf{U}}_n(\boldsymbol{\xi}, w_0) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau K_h(W_i(t) - w_0) \times \{\mathbf{Z}_i^*(t, w_0) - \bar{\mathbf{Z}}(t, w_0)\}\, dN_i(t)\psi_i,$$

where $\{\psi_1, \ldots, \psi_n\}$ is an iid sample from the standard normal distribution. Then, within a properly chosen interval $[w_1, w_2]$, we define the standardized distribution

$$\widetilde{\mathcal{S}}_k = \sup_{w_0 \in [w_1, w_2]} |v_k(w_0)\widetilde{\mathcal{U}}_k(\widehat{\boldsymbol{\xi}}, w_0)|, \qquad k = 1, \ldots, (2p+q+1),$$

where $\widetilde{\mathcal{U}}_k(\widehat{\boldsymbol{\xi}}, w_0)$ is the kth component of the vector $\{\mathbf{HD}_n(w_0)\mathbf{H}\}^{-1}\widetilde{\mathbf{U}}_n(\widehat{\boldsymbol{\xi}}, w_0)$ and $v_k(w_0)$ is a positive weight function that converges uniformly to a deterministic function. By repeatedly generating samples of $\{\psi_1, \ldots, \psi_n\}$, the distribution of

$$\mathcal{S}_k = \sup_{w_0 \in [w_1, w_2]} \left| v_k(w_0)\{\widehat{\xi}_k(w_0) - \xi_{0k}(w_0)\}\right|, \qquad k = 1, \ldots, (2p + q + 1),$$

can be approximated by that of $\widetilde{\mathcal{S}}_k$. Let $c_{\alpha k}$ be the $100(1-\alpha)$th percentile of this approximate distribution; then the $(1-\alpha)$ confidence band for $\{\xi_{0k}(w_0), w_0 \in [w_1, w_2]\}$ is given by $\widehat{\xi}_k(w_0) \pm c_{\alpha k} v_k^{-1}(w_0)$. The validity of the approximation of $\mathcal{S}_k$ using the simulation technique can be justified following almost the same arguments as those of Tian et al. (2005); we leave the justification to Appendix B.

3.4 Bandwidth Selection

The bandwidth selection often is a critical part of nonparametric regression. We can select the optimal h by minimizing the asymptotic weighted mean squared error. For the kth component of $\widehat{\boldsymbol{\xi}}(w_0)$, we minimize

$$\int_{\mathcal{W}} \left\{\frac{1}{4}h^4\mu_2^2\phi_k^2(w_0) + \frac{1}{nh}\sigma_{kk}(w_0)\right\}\Psi(w_0)\, dw_0,$$

where $\phi_k(w_0)$ is the kth element of the vector $\mathbf{D}^{-1}(w_0)\mathbf{b}(w_0)$, $\sigma_{kk}(w_0)$ is the kth diagonal element of $\mathbf{D}^{-1}(w_0)\boldsymbol{\Sigma}(w_0)\mathbf{D}^{-1}(w_0)$, and $\Psi(\cdot)$ is a nonnegative and integrable weight function. Therefore, the theoretical optimal bandwidth is given by

$$h_{\mathrm{opt},k} = \left\{\frac{\int \sigma_{kk}(w_0)\Psi(w_0)\, dw_0}{\int \mu_2^2\phi_k^2(w_0)\Psi(w_0)\, dw_0}\right\}^{1/5} n^{-1/5}.$$

Alternatively, we could choose the bandwidth h by the K-fold cross-validation method (see, e.g., Tian et al. 2005; Fan et al. 2006). Here we first divide the data into K equal-sized groups. If we let D_k denote the kth subgroup of data, then the kth prediction error is given by

$$\mathrm{PE}_k(h) = \sum_{i \in D_k}\int_0^\tau \{N_i(t) - \widehat{E}(N_i(t))\}^2\, d\left\{\sum_{j \in D_k} N_j(t)\right\}, \qquad k = 1, \ldots, K,$$

where

$$\widehat{E}(N_i(t)) = \int_0^t Y_i(u)\big[d\widehat{\Lambda}_{0(-k)}(u) + \{\widehat{\boldsymbol{\beta}}_{(-k)}^T(W_i(u))\mathbf{Z}_i(u) + \widetilde{\boldsymbol{\gamma}}_{(-k)}^T\mathbf{V}_i(u) + \widehat{\alpha}_{(-k)}(W_i(u))\}\, du\big],$$

in which $\widehat{\boldsymbol{\beta}}_{(-k)}(\cdot)$, $\widetilde{\boldsymbol{\gamma}}_{(-k)}$, $\widehat{\alpha}_{(-k)}(\cdot)$, and $\widehat{\Lambda}_{0(-k)}(t)$ are estimated using the data from all subgroups other than D_k. The global estimator of $\Lambda_0(t)$ is given by

$$\widehat{\Lambda}_0(t) = \int_0^t \left(\sum_{i=1}^n \big[dN_i(u) - Y_i(u) \times \{\widehat{\boldsymbol{\beta}}^T(W_i(u))\mathbf{Z}_i(u) + \widetilde{\boldsymbol{\gamma}}^T\mathbf{V}_i(u) + \widehat{\alpha}(W_i(u))\}\, du\big]\right) \Big/ \sum_{i=1}^n Y_i(u).$$

The optimal bandwidth can be obtained by minimizing the total prediction error, $\sum_{k=1}^K \mathrm{PE}_k(h)$, with respect to h. We note that the numerical procedure for the bandwidth selection is to balance the trade-offs between the variance and bias, whereas condition (C.3) in the Appendix removes the asymptotic bias.

4. NUMERICAL STUDIES

4.1 Simulations

We carried out two sets of simulation studies to examine the finite-sample properties of the proposed methods. In Simulation I we generated the failure times from the partially linear additive hazards model

$$\lambda(t|Z, V, W) = \lambda_0(t) + \beta(W)Z + \gamma V + \alpha(W), \qquad (7)$$

Table 1. Simulation I: Estimation of coefficients with $h = .12$ and 25% censoring

		$\beta(w) = 1.2 + \sin(2w)$				$\alpha'(w) = .2$				$\gamma = 1$			
n	w_0	Bias	SD	SE	CR (%)	Bias	SD	SE	CR (%)	Bias	SD	SE	CR (%)
200	.5	.190	1.226	1.186	95.4	−.069	1.787	1.740	95.8	.081	.370	.311	91.6
	1.0	.075	1.171	1.205	96.0	.095	1.644	1.565	95.8				
	1.5	.021	1.123	1.078	95.6	.112	1.623	1.559	95.0				
	2.0	.064	.925	.948	96.8	.023	1.561	1.518	95.2				
	2.5	.016	1.048	.994	95.8	.071	1.780	1.804	95.2				
400	.5	−.061	.798	.790	94.6	.065	1.229	1.156	94.2	.025	.223	.218	95.2
	1.0	−.047	.824	.810	94.4	−.102	1.094	1.069	94.6				
	1.5	.069	.768	.734	95.8	.030	.970	1.045	96.0				
	2.0	.104	.643	.638	93.2	−.052	1.037	1.024	94.2				
	2.5	.109	.674	.660	94.4	.098	1.197	1.192	96.4				

where $\beta(w) = 1.2 + \sin(2w)$, $\gamma = 1$, $\alpha(w) = .2w$, and $\lambda_0(t) = .5$. Covariate Z was generated from a uniform distribution, Unif[0, 1], and covariate V was a Bernoulli random variable taking value 0 or 1 with probability .5. The exposure variable W was generated from Unif[0, 3]. The censoring time was taken as the minimum value of τ and a random number independently generated from Unif[$\tau/2, 3\tau/2$]. We took $\tau = .86$ to yield an approximate censoring rate of 25%. We used the Gaussian kernel function and chose 29 even partitions along the range of W, that is, $w_0 = (.1, .2, \ldots, 2.9)$. The bandwidth of $h = .12$ was chosen based on a preliminary investigation in which the cross-validation method was applied to a few simulated data sets. We used sample sizes $n = 200$ and 400. For each configuration, we replicated 500 simulations. Based on each data realization, we computed the estimators for $\beta(w_0)$, $\alpha(w_0)$, γ, and $\Lambda_0^*(t, w_0)$, along with the corresponding standard errors.

We present the regression coefficient estimates of Simulation I in Table 1 and the local baseline cumulative hazard function estimates in Table 2. We report the standard deviations (SDs) characterizing the sample variations over 500 simulations, the average standard errors (SEs) using the asymptotic approximation, and the 95% confidence interval coverage rates (CRs). The SEs are approximately unbiased for the SDs and the 95% confidence interval CRs are centered around the nominal level. The variances of the estimates decrease as the sample size increases. As shown in Table 2, for the selected time points and w_0, the local estimators $\widehat{\Lambda}_0^*(t, w_0)$ are close to the true values, and the SE provides a good approximation for the variation of the point estimates. The CRs of the 95% confidence intervals are close to the nominal value.

Figures 1(a) and (b) show the varying-coefficient estimates and the 95% pointwise confidence intervals averaged over 500 simulations, (c) and (d) correspond to the local estimator and the true function of the baseline cumulative hazard $\Lambda_0^*(t, w_0)$, and (e) shows the global estimator and the true curve of $\Lambda_0(t)$. We can see that the estimated varying-coefficient curves are close to the true curves. For the baseline cumulative hazard function, the local estimator $\widehat{\Lambda}_0^*(t, w_0)$ matches the true pattern of the surface, increasing with respect to both t and w_0 and clearly indicating an interaction between t and w_0. The global estimator $\widehat{\Lambda}_0(t)$ lies slightly above the true line. Overall, our proposed methods behave well with sample sizes of practical use.

Table 2. Simulation I: Estimation of the local baseline cumulative hazard function $\Lambda_0^*(t, w) = .5t + .2wt$ with $h = .12$ and 25% censoring

			$n = 200$				$n = 400$			
t	w_0	True value	$\widehat{\Lambda}_0^*(t, w_0)$	SD	SE	CR (%)	$\widehat{\Lambda}_0^*(t, w_0)$	SD	SE	CR (%)
.25	.5	.150	.121	.170	.166	94.0	.153	.109	.116	95.6
	1.0	.175	.156	.172	.173	95.0	.165	.121	.119	94.8
	1.5	.200	.197	.165	.164	95.0	.189	.116	.112	92.8
	2.0	.225	.228	.152	.150	94.6	.224	.106	.103	93.4
	2.5	.250	.262	.167	.159	95.8	.251	.105	.107	96.0
.5	.5	.300	.243	.300	.297	94.6	.306	.206	.205	94.8
	1.0	.350	.319	.312	.312	94.6	.346	.216	.212	95.4
	1.5	.400	.405	.300	.290	94.6	.391	.214	.201	94.4
	2.0	.450	.470	.276	.276	95.6	.449	.190	.186	95.4
	2.5	.500	.538	.305	.296	95.0	.497	.193	.196	95.0
.7	.5	.450	.388	.428	.429	93.4	.474	.297	.295	94.8
	1.0	.525	.517	.450	.450	93.6	.537	.314	.306	94.8
	1.5	.600	.630	.449	.422	93.8	.591	.310	.292	94.6
	2.0	.675	.728	.417	.403	95.4	.681	.271	.273	95.6
	2.5	.750	.835	.460	.436	94.6	.762	.288	.288	96.4

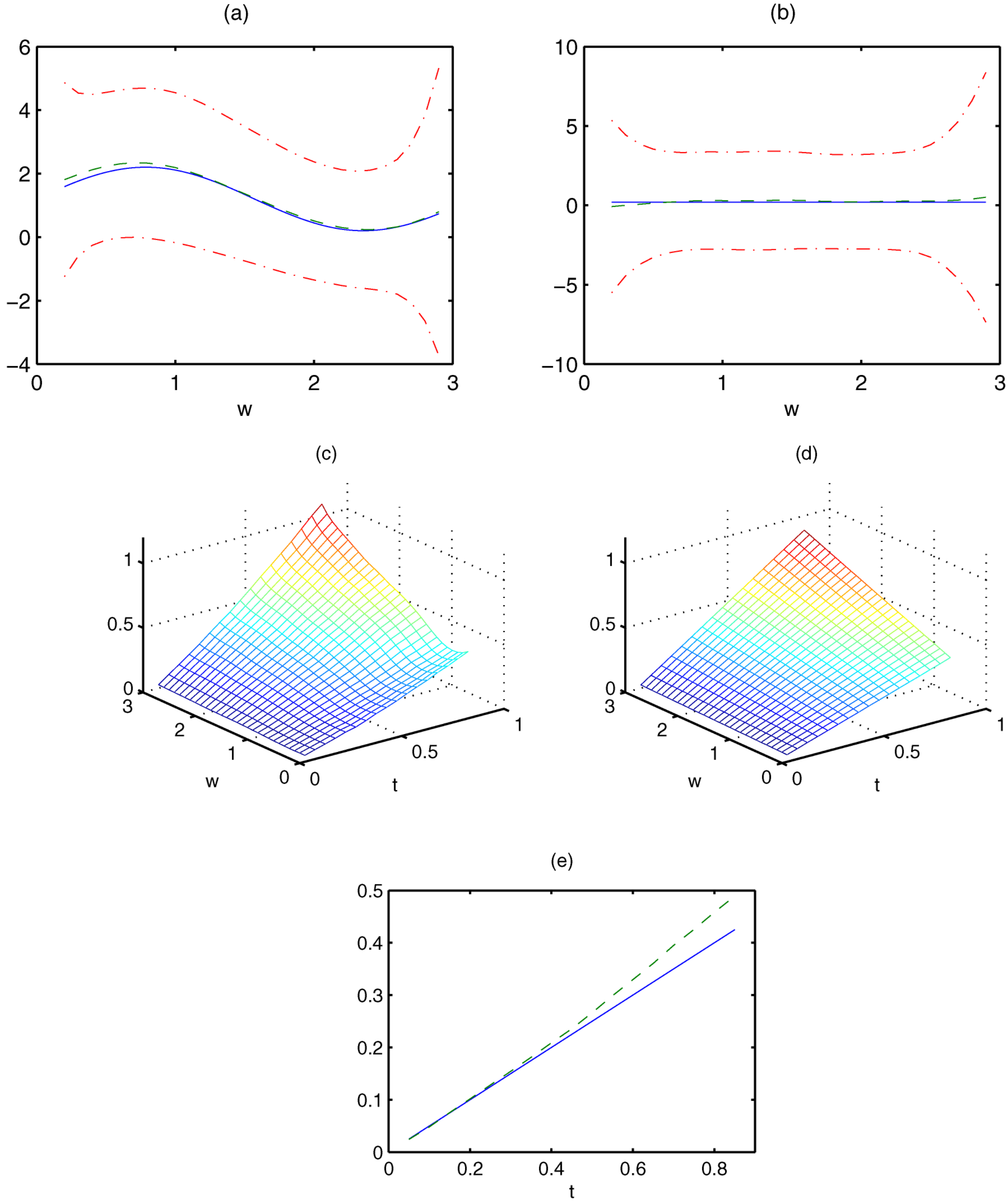

Figure 1. Simulation I with $n = 200$ and $h = .12$ averaging over 500 data replicates. (a) Estimated curves of $\beta(w)$. (b) Estimated curves of $\alpha'(w)$. In (a) and (b) solid lines are the true functions, dashed lines are the estimates of varying coefficients, and dashed–dotted lines are the pointwise 95% confidence intervals. (c), (d) The local estimator $\widehat{\Lambda}_0^*(t, w_0)$ and the true surface of the baseline cumulative hazard. (e) The global estimator $\widehat{\Lambda}_0(t)$ of the baseline cumulative hazard, with the solid line representing the true function and the dashed line representing the estimate.

In Simulation II we examined model (7), where $\beta(w) = 1.2 + \cos(2w)$, $\gamma = 1$, $\alpha(w) = 0$, and $\lambda_0(t) = t$. In this case the exposure variable had no effect on the baseline hazard function as $\alpha(w) = 0$, that is, $\Lambda_0^*(t, w_0) \equiv \Lambda_0^*(t)$. We generated covariates and censoring times in the same way as in Simulation I, and we took $\tau = 1.1$ to yield a censoring rate of 25%. Tables 3 and 4 summarize the simulation results, from which we can see that the biases are quite small, the SEs based on the asymptotic approximation provide good approximation of the variability of the estimators, and the 95% confidence interval CRs are reasonably accurate. Figures 2(a) and (b) show that the averaged varying-coefficient estimates are close to the true curves. Figure 2(c) shows the averaged surface estimate of $\Lambda_0^*(t, w_0)$ over 500 simulations, which matches the pattern of the true surface in (d); that is, it gradually increases with t while staying constant with respect to w_0. Comparing Figures 2(c) and 1(c) demonstrates the role of $\alpha(w_0)$ in the estimation of $\Lambda_0^*(t, w_0)$: If $\alpha(w_0) = 0$, then W has no influence on $\Lambda_0^*(t, w_0)$, such that

Table 3. Simulation II: Estimation of coefficients with $h = .1$ and 25% censoring

		$\beta(w) = 1.2 + \cos(2w)$				$\beta'(w) = -2\sin(2w)$				$\gamma = 1$			
n	w_0	Bias	SD	SE	CR (%)	Bias	SD	SE	CR (%)	Bias	SD	SE	CR (%)
200	.5	.209	1.055	1.003	95.0	−.052	2.114	2.102	96.0	.072	.236	.275	91.0
	1.0	.131	.752	.773	96.8	−.022	1.575	1.529	96.4				
	1.5	.082	.654	.638	96.4	.054	1.190	1.243	96.8				
	2.0	.073	.703	.720	96.8	−.092	1.514	1.423	94.4				
	2.5	.033	1.007	.948	93.0	.028	2.089	1.971	94.4				
400	.5	−.004	.685	.679	95.0	.034	1.421	1.409	95.4	.023	.165	.168	96.0
	1.0	.050	.537	.525	94.8	.112	1.079	1.048	94.2				
	1.5	.128	.456	.443	95.8	.083	.836	.862	95.2				
	2.0	.073	.506	.494	95.4	−.119	.976	.984	94.8				
	2.5	.062	.648	.644	95.0	−.119	1.363	1.348	94.4				

$\widehat{\Lambda}_0^*(t, w_0)$ is parallel to the horizontal axis of W. Figure 2(e) indicates reasonable performance of the global estimator for the baseline cumulative hazard function.

4.2 Breast Cancer Data Analysis

As an illustration, we applied our methods to data from a recent study on 197 patients with high-risk primary or metastatic breast cancer. The study was initiated in April 1992, and patient follow-up continued until November 2005. The primary aim was to determine whether chemotherapy at a dose higher than the standard maximum tolerated dose might induce a better response. Patients were randomized to either systemic chemotherapy with standard doses of 5-FU, doxorubicin, and cyclophosphamide (FAC) or a dose-intense regimen of FAC supported by the granulocyte colony-stimulating factor (G–CSF). G–CSF was shown to reduce the likelihood and severity of neutropenia and its attendant complications. In the neoadjuvant setting, patients in the FAC arm were treated for four cycles with a cycle duration of 21 days, whereas those in the treatment arm of FAC combined with G–CSF received a higher dose of FAC with a shorter cycle duration. But the dose intensity was defined as the amount of drug administered per unit time, which was different for different patients. There were 197 distinct values of dose intensity, which were computed using the method of Ang, Buzdar, Smith, Kau, and Hortobagyi (1989). We were interested in characterizing the relationship between the disease-free survival (DFS) and known risk factors and evaluating how the dose intensity interacted with other covariates, including disease stage, pathological response, number of positive axillary (AX) nodes, tamoxifen use (1 if yes; 0 if no), and menopausal status (1 if premenopausal; 0 otherwise). To explore the nonlinear interactions between dose intensity (W) and other covariates, we started by fitting the fully nonparametric model to the breast cancer data. After observing that the parameters associated with disease stages III and IV and menopausal status appeared to be invariant with respect to dose intensity, we obtained the following model:

$$\begin{aligned}\lambda(t|\mathbf{Z}, \mathbf{V}, W) = \lambda_0(t) &+ \beta_1(W) Z_{\text{Path Resp}} + \beta_2(W) Z_{\text{Tamox}} \\ &+ \beta_3(W) Z_{\text{AX nodes}} + \gamma_1 V_{\text{Stage III}} \\ &+ \gamma_2 V_{\text{Stage IV}} + \gamma_3 V_{\text{Manop}} + \alpha(W).\end{aligned}$$

Table 4. Simulation II: Estimation of the local baseline cumulative hazard function $\Lambda_0^*(t, w) = .5t^2$ with $h = .1$ and 25% censoring

			$n = 200$				$n = 400$			
t	w_0	True value	$\widehat{\Lambda}_0^*(t, w_0)$	SD	SE	CR (%)	$\widehat{\Lambda}_0^*(t, w_0)$	SD	SE	CR (%)
.25	.5	.031	.016	.151	.143	93.8	.027	.098	.116	95.2
	1.0		.022	.123	.116	94.4	.029	.080	.119	95.0
	1.5		.021	.102	.099	94.4	.029	.070	.112	93.4
	2.0		.020	.109	.107	94.4	.023	.076	.103	95.4
	2.5		.028	.138	.133	94.8	.032	.093	.107	95.6
.5	.5	.125	.106	.262	.252	93.8	.122	.167	.173	95.4
	1.0		.118	.228	.211	93.2	.117	.143	.144	95.6
	1.5		.111	.189	.181	94.8	.126	.134	.127	94.2
	2.0		.108	.196	.195	94.0	.114	.138	.138	94.2
	2.5		.119	.252	.240	92.6	.131	.164	.166	96.2
.7	.5	.281	.265	.385	.362	94.2	.278	.243	.248	95.8
	1.0		.271	.331	.307	93.2	.270	.209	.210	94.6
	1.5		.259	.293	.266	92.2	.288	.199	.187	94.6
	2.0		.265	.294	.286	93.4	.275	.208	.201	93.4
	2.5		.278	.379	.346	93.0	.298	.232	.239	96.2

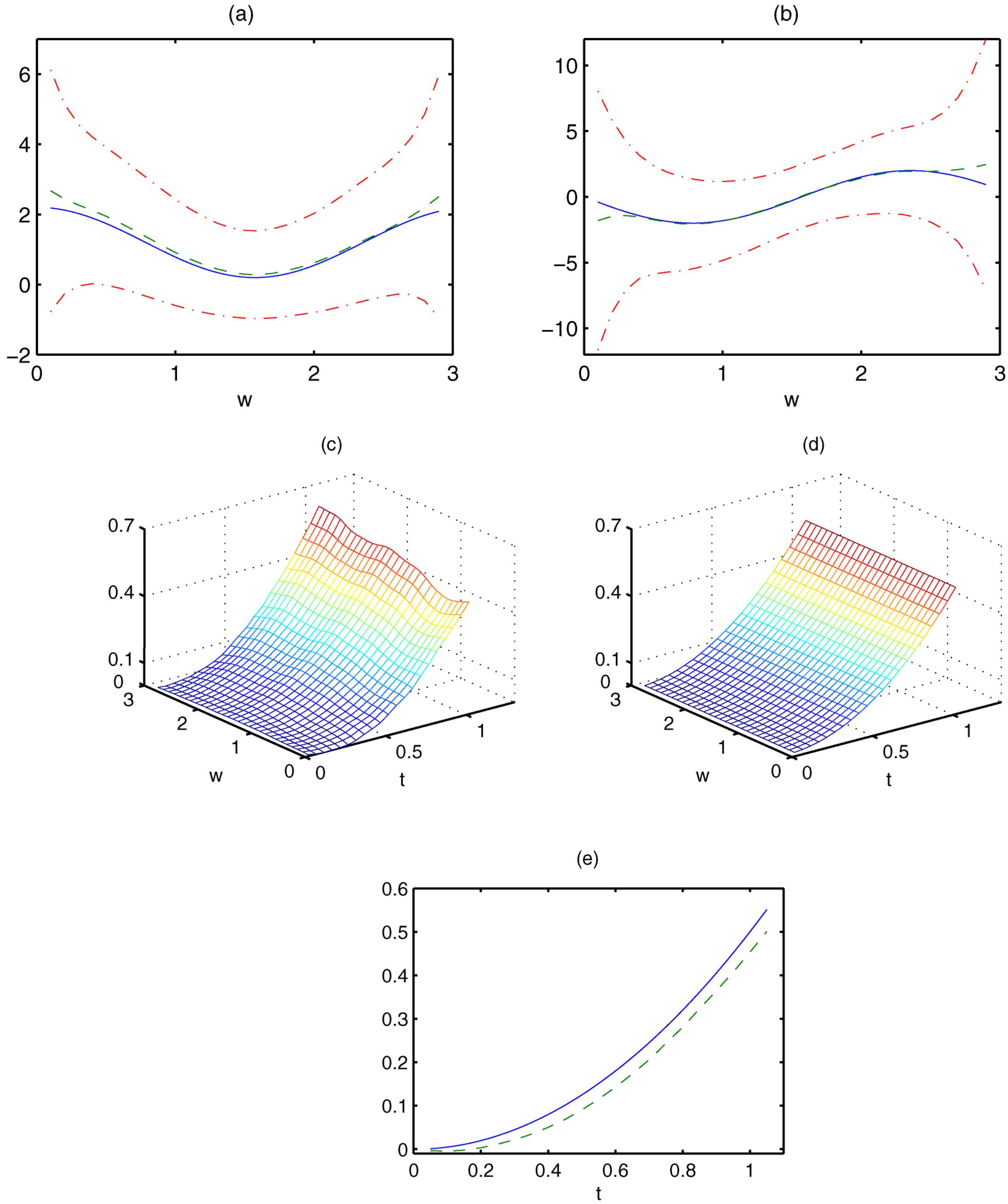

Figure 2. Simulation II with $n = 200$ and $h = .12$ averaging over 500 data replicates. (a) Estimated curves of $\beta(w)$. (b) Estimated curves of $\beta'(w)$. In (a) and (b) the solid lines are the true functions, dashed lines are the estimates of varying coefficients, and dashed–dotted lines are the pointwise 95% confidence intervals. (c), (d) The local estimator $\widehat{\Lambda}_0^*(t, w_0)$ and the true surface of the baseline cumulative hazard. (e) The global estimator $\widehat{\Lambda}_0(t)$ of the baseline cumulative hazard, with the solid line representing the true function and the dashed line representing the estimate.

We used the K-fold cross-validation method to select the optimal bandwidth with $K = 39$. As shown in Figure 3, $h = 2.5$ yielded the smallest prediction error. We used the Gaussian kernel function and partitioned the entire range of W into 70 intervals, with $w_0 = 6.25 + .25(j - 1)$, $j = 1, \ldots, 70$. For the constant coefficient associated with disease stage III, $\widetilde{\gamma}_1 = .035$ (the standard error of .018); for stage IV, $\widetilde{\gamma}_2 = .057(.030)$; and for the menopausal status, $\widetilde{\gamma}_3 = -.026(.016)$. Patients with stage III or IV breast cancer had a higher risk of disease relapse, but the difference was not statistically significant. Menopausal status did not appear to affect the DFS significantly either. The estimated regression curves and their 95% pointwise and simultaneous confidence bands are presented in Figure 4(a)–(d). We see an overall trend of a higher dose intensity associated with a decreased hazard, thus leading to a better DFS. Pathological response decreased the risk significantly at low dose intensities, whereas the risk reduction gradually disappeared with increasing dose intensity. The covariate effects of tamoxifen use and

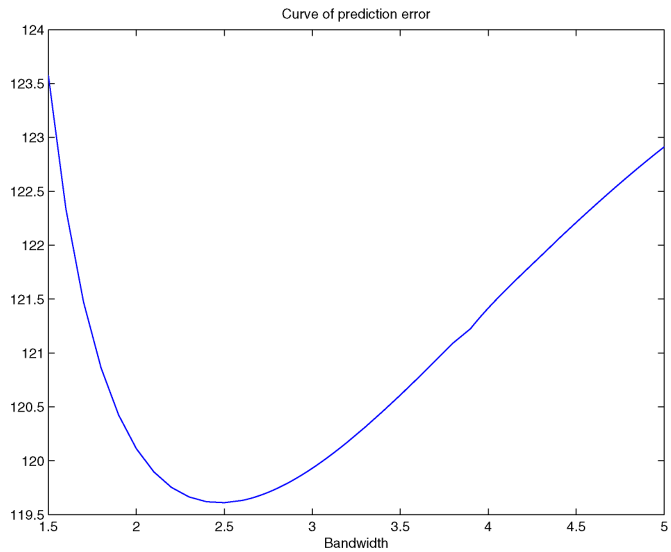

Figure 3. Prediction errors versus bandwidths, indicating the optimal bandwidth $h = 2.5$.

the number of positive AX nodes were not statistically significant. Tamoxifen use resulted in an interesting trend over the dose intensity; for patients treated at low dose intensities, tamoxifen helped prolong the DFS, whereas for patients treated at high dose intensities, its use did not improve survival as much. That is, the effect of tamoxifen use might be offset by increasing the dose intensity. The effect of the number of positive AX nodes showed an increasing pattern as the dose intensity increased; the more positive nodes a patient had, the worse her DFS, particularly at high dose intensities. The dose intensity main effect showed a steady decreasing trend; the higher the dose intensity, the lower the risk of disease relapse. Figures 4(e) and (f) show the local and global estimators of $\Lambda_0^*(t, w_0)$ and $\Lambda_0(t)$. The local estimator $\widehat{\Lambda}_0^*(t, w_0)$ increases with respect to both t and w_0.

5. DISCUSSION

The additive hazards model serves as an important alternative to the proportional hazards model. To enhance modeling flexibility, we have studied the semiparametric partially linear additive hazards model and proposed the estimation and inference procedures. The coefficients may vary and thus exhibit different covariate effects over the level of an exposure variable. Based on a different perspective than the time-varying coefficient model, our model imposes a covariate-varying structure. We applied the local polynomial technique and estimated the coefficient functions nonparametrically. The proposed model and estimation procedure are particularly attractive due to the analytic solution for the estimator, because convergence and initial value selection are the well-known challenges for nonparametric regression. The choice of the smoothing parameter in the kernel function requires more caution in practice. Based on the simulation studies, the estimation procedure appears to be quite robust to the bandwidth. We typically truncate both the left and right sides of $\mathcal{W}$ by the bandwidth h, because the estimates at the boundaries often are not very stable. To gain efficiency, a weight function can be incorporated into the estimating equation; however, this may lessen the ease of computation. The baseline cumulative hazard function is completely unspecified, and we can ensure its monotonicity by forcing the estimator to be nondecreasing over time (Lin and Ying 1994; Peng and Huang 2007).

APPENDIX A: PROOFS OF THEOREMS

We impose the following conditions:

(C.1) The kernel function $K(\cdot) > 0$ is a bounded and symmetric density with a compact bounded support.

(C.2) The functions $\boldsymbol{\beta}(\cdot)$ and $\alpha(\cdot)$ have continuous second derivatives in $\mathcal{W}$ including the boundary and $\int_0^\tau \lambda_0(t)\,dt < \infty$.

(C.3) $h \to 0$, $\log h/\sqrt{nh^2} \to 0$ and nh^4 is bounded.

(C.4) $\inf_{t\in[0,\tau], w_0\in\mathcal{W}} a_0(t, w_0) > 0$, the matrices

$$\mathbf{a}_2(w_0) - \int_0^\tau \frac{\mathbf{a}_1^{\otimes 2}(t, w_0)}{a_0(t, w_0)}\,dt$$

and

$$\begin{pmatrix} \mathbf{a}_2(w_0) & \mathbf{a}_1(w_0) \\ \mathbf{a}_1^T(w_0) & a_0(w_0) \end{pmatrix}$$

are nonsingular, and $\boldsymbol{\Sigma}(w_0)$ is positive definite for all $w_0 \in \mathcal{W}$.

(C.5) The sample path of $(\mathbf{Z}(t), \mathbf{V}(t), W(t))$ has bounded total variation in $[0, \tau]$.

(C.6) The conditional density of $(\mathbf{Z}(t), \mathbf{V}(t))$ given $W(t) = w_0$ is twice continuously differentiable with respect to w_0. The marginal density of $W(t)$ evaluated at w_0, denoted by $f_W(t, w_0)$, is twice continuously differentiable with respect to w_0 and satisfies $\inf_{t\in[0,\tau], w_0\in\mathcal{W}} f_W(t, w_0) > 0$.

All of these conditions are standard conditions in statistical inference using local linear estimation. For (C.3), we can choose the bandwidth $h = n^{-\nu}$ with $\nu \in [1/4, 1/3)$. Condition (C.5) allows discontinuous sample paths of time-dependent covariates.

Let the filtration $\{\mathcal{F}_t : t \in [0, \tau]\}$ be the data history up to time t, that is,

$$\mathcal{F}_t = \sigma\{N_i(s), Y_i(s), 0 \le s \le t; \\ \mathbf{Z}_i(s), \mathbf{V}_i(s), W_i(s), 0 \le s \le \tau, i = 1, \ldots, n\}.$$

Define $M_i(t) = N_i(t) - \int_0^t Y_i(u)\lambda(u|\mathbf{Z}_i, \mathbf{V}_i, W_i)\,du$. Then $M_i(t)$ is a $\mathcal{F}_t$-martingale.

Recall that $\mathbf{G}_i^*(t, w_0) = \mathbf{H}^{-1}\mathbf{Z}_i^*(t, w_0)$. For $t \in [0, \tau]$, $k = 0, 1$, and 2, we define

$$\mathbf{S}_{nk}(t, w_0) = \frac{1}{n}\sum_{i=1}^n K_h(W_i(t) - w_0)Y_i(t)(\mathbf{G}_i^*(t, w_0))^{\otimes k}$$

and

$$\mathbf{S}_{nk}^*(t, w_0) = \frac{1}{n}\sum_{i=1}^n K_h(W_i(t) - w_0) \\ \times Y_i(t)\lambda(t|\mathbf{Z}_i, \mathbf{V}_i, W_i)(\mathbf{G}_i^*(t, w_0))^{\otimes k}.$$

For $w_0 \in \mathcal{W}$, we define

$$s_0(t, w_0) = a_0(t, w_0),$$

$$\mathbf{s}_1(t, w_0) = \left(\mathbf{a}_1^T(t, w_0), \mathbf{c}_1^T(t, w_0), \mathbf{0}_{p+1}^T\right)^T,$$

$$\mathbf{s}_2(t, w_0) = \operatorname{diag}\left(\begin{pmatrix} \mathbf{a}_2(t, w_0) & \mathbf{g}(t, w_0) \\ \mathbf{g}^T(t, w_0) & \mathbf{c}_2(t, w_0) \end{pmatrix},\right. \\ \left.\mu_2\begin{pmatrix} \mathbf{a}_2(t, w_0) & \mathbf{a}_1(t, w_0) \\ \mathbf{a}_1^T(t, w_0) & a_0(t, w_0) \end{pmatrix}\right)$$

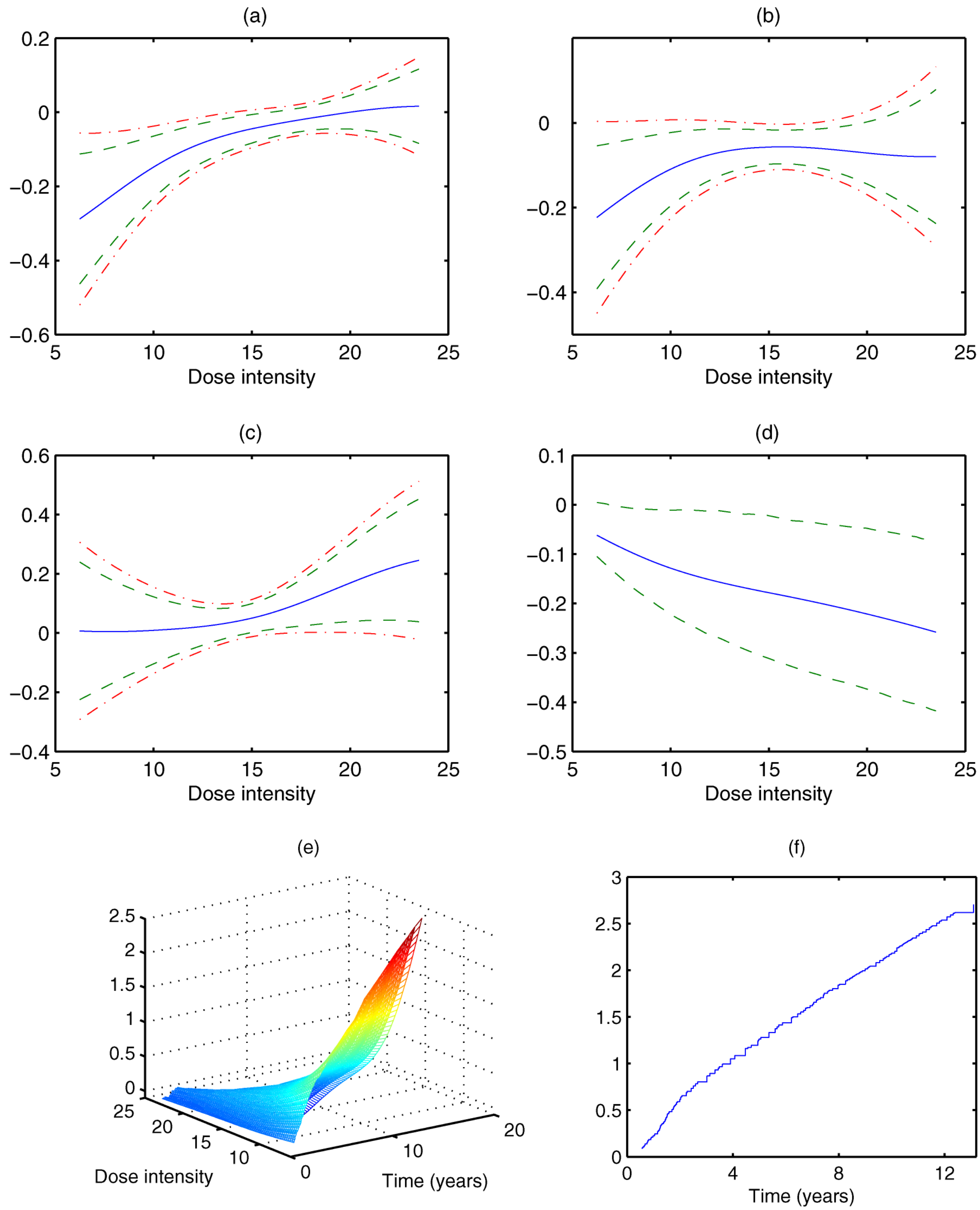

Figure 4. Estimates of regression coefficient curves for the breast cancer data with $h = 2.5$. (a) Pathological response; (b) tamoxifen; (c) number of positive AX nodes; (d) dose intensity main effect. In (a)–(d), the solid lines are the estimates of varying coefficients, dashed lines are the 95% pointwise confidence intervals, and dashed–dotted lines are the 95% simultaneous confidence bands. (e) The local estimator of the baseline cumulative hazard. (f) The global estimator of the baseline cumulative hazard.

and

$$s_0^*(t, w_0) = a_0^*(t, w_0),$$

$$\mathbf{s}_1^*(t, w_0) = \left((\mathbf{a}_1^*(t, w_0))^T, (\mathbf{c}_1^*(t, w_0))^T, \mathbf{0}_{p+1}^T\right)^T,$$

$$\mathbf{s}_2^*(t, w_0) = \mathrm{diag}\left(\begin{pmatrix} \mathbf{a}_2^*(t, w_0) & \mathbf{g}^*(t, w_0) \\ (\mathbf{g}^*(t, w_0))^T & \mathbf{c}_2^*(t, w_0) \end{pmatrix},\right.$$

$$\left.\mu_2 \begin{pmatrix} \mathbf{a}_2^*(t, w_0) & \mathbf{a}_1^*(t, w_0) \\ (\mathbf{a}_1^*(t, w_0))^T & a_0(t, w_0) \end{pmatrix}\right).$$

Lemma A.1. Under Conditions (C.1)–(C.6), we have that, for $k = 0$, 1, and 2,

$$\sup_{t \in [0,\tau], w_0 \in \mathcal{W}} |\mathbf{S}_{nk}(t, w_0) - \mathbf{s}_k(t, w_0)| = O_p\left(\frac{\log h}{\sqrt{nh}}\right) + O(h^2)$$

and

$$\sup_{t\in[0,\tau],w_0\in\mathcal{W}} |\mathbf{S}^*_{nk}(t,w_0)-\mathbf{s}^*_k(t,w_0)| = O_p\left(\frac{\log h}{\sqrt{n}h}\right)+O(h^2).$$

Proof of Lemma A.1. Let $\mathbf{P}_n$ and $\mathbf{G}_n$ denote the empirical measure and the empirical process from n iid observations. We can rewrite $\mathbf{S}_{nk}(t,w_0)$ as

$$\mathbf{S}_{nk}(t,w_0)=\mathbf{P}_n\left[\frac{1}{h}K\left(\frac{W(t)-w_0}{h}\right)Y(t)\begin{pmatrix}\mathbf{Z}(t)\\ \mathbf{V}(t)\\ \mathbf{Z}(t)\frac{W(t)-w_0}{h}\\ \frac{W(t)-w_0}{h}\end{pmatrix}^{\otimes k}\right], \quad k=0,1,2.$$

Note that $W(t)$, $\mathbf{Z}(t)$, $\mathbf{V}(t)$, and $Y(t)$ are stochastic processes with bounded total variation. From lemma 9.10 of Kosorok (2008), they are all VC-subgraph with finite VC-index. Thus, by theorem 2.6.7 of van der Vaart and Wellner (1996), there exist an $m=O(\delta^{-N})$ number of balls covering $\{(W(t),\mathbf{Z}(t),\mathbf{V}(t),Y(t)):t\in[0,\tau]\}$ with $L_r(Q)$ radius $<\delta$, where N is a constant depending only on r and Q is any probability measure. We also partition $\mathcal{W}$ into intervals with the length $<\delta/2$. It is direct to verify that for any pairs of $(W(t),\mathbf{Z}(t),\mathbf{V}(t),Y(t))$ within the same ball and any pairs of w_0 within the same interval, the $L_r(Q)$ distance of the associate function in the class

$$\mathcal{F}=\left\{\frac{1}{h}K\left(\frac{W(t)-w_0}{h}\right)Y(t)\begin{pmatrix}\mathbf{Z}(t)\\ \mathbf{V}(t)\\ \mathbf{Z}(t)\frac{W(t)-w_0}{h}\\ \frac{W(t)-w_0}{h}\end{pmatrix}^{\otimes k} : w_0\in\mathcal{W}, t\in[0,\tau]\right\}$$

cannot exceed $O(h^{-4}\delta)$, that is,

$$N(\delta,\mathcal{F},L_r(Q))\le O\big(h^{-4(N+1)}\delta^{-(N+1)}\big).$$

In addition, $\mathcal{F}$ has a covering function $O(h^{-1})$. According to corollary 19.38 of van der Vaart (1998), we obtain

$$E^*\|\mathbf{G}\|_{\mathcal{F}}\le\int_0^1\sqrt{1+\log O\big(h^{-4(N+1)}\delta^{-(N+1)}\big)}\,d\delta\, h^{-1} = O\left(\frac{\log h}{h}\right).$$

Thus we obtain

$$\sup_{t\in[0,\tau],w_0\in\mathcal{W}}\left|\mathbf{S}_{nk}(t,w_0) - E\left[\frac{1}{h}K\left(\frac{W(t)-w_0}{h}\right)Y(t)\begin{pmatrix}\mathbf{Z}(t)\\ \mathbf{V}(t)\\ \mathbf{Z}(t)\frac{W(t)-w_0}{h}\\ \frac{W(t)-w_0}{h}\end{pmatrix}^{\otimes k}\right]\right| = O_p\left(\frac{\log h}{\sqrt{n}h}\right).$$

Furthermore, because

$$E\left[\frac{1}{h}K\left(\frac{W(t)-w_0}{h}\right)Y(t)\begin{pmatrix}\mathbf{Z}(t)\\ \mathbf{V}(t)\\ \mathbf{Z}(t)\frac{W(t)-w_0}{h}\\ \frac{W(t)-w_0}{h}\end{pmatrix}^{\otimes k}\right] = \int_x K(x)E\left\{Y(t)\begin{pmatrix}\mathbf{Z}(t)\\ \mathbf{V}(t)\\ \mathbf{Z}(t)x\\ x\end{pmatrix}^{\otimes k}\,\middle|\, W(t)=xh+w_0\right\} \times f_W(t,xh+w_0)\,dx,$$

by the Taylor expansion, we have

$$\sup_{t\in[0,\tau],w_0\in\mathcal{W}}\left|E\left[\frac{1}{h}K\left(\frac{W(t)-w_0}{h}\right)Y(t)\begin{pmatrix}\mathbf{Z}(t)\\ \mathbf{V}(t)\\ \mathbf{Z}(t)\frac{W(t)-w_0}{h}\\ \frac{W(t)-w_0}{h}\end{pmatrix}^{\otimes k}\right] - \mathbf{s}_k(t,w_0)\right| = O(h^2).$$

We conclude that

$$\sup_{t\in[0,\tau],w_0\in\mathcal{W}}|\mathbf{S}_{nk}(t,w_0)-\mathbf{s}_k(t,w_0)| = O_p\left(\frac{\log h}{\sqrt{n}h}\right)+O(h^2).$$

The proof of the second half of the lemma follows similar arguments, so we omit it here.

Proof of Theorem 1

First, we note that

$$\mathbf{H}^{-1}\mathbf{U}_n(\boldsymbol{\xi},w_0)=\frac{1}{n}\sum_{i=1}^n\int_0^\tau K_h(W_i(t)-w_0)\{\mathbf{G}^*_i(t,w_0)-\bar{\mathbf{G}}(t,w_0)\} \times\{dN_i(t)-Y_i(t)\boldsymbol{\xi}^T(w_0)\mathbf{Z}^*_i(t,w_0)\,dt\}.$$

Because $\mathbf{U}_n(\widehat{\boldsymbol{\xi}},w_0)=0$, we have

$$\mathbf{H}^{-1}\mathbf{U}_n(\boldsymbol{\xi}_0,w_0)=\mathbf{D}_n(w_0)\mathbf{H}(\widehat{\boldsymbol{\xi}}(w_0)-\boldsymbol{\xi}_0(w_0)), \tag{A.1}$$

where

$$\mathbf{D}_n(w_0)=\frac{1}{n}\sum_{i=1}^n\int_0^\tau K_h(W_i(t)-w_0)Y_i(t) \times\{\mathbf{G}^*_i(t,w_0)-\bar{\mathbf{G}}(t,w_0)\}^{\otimes 2}\,dt.$$

By the definition of the martingale,

$$dM_i(t)=dN_i(t)-Y_i(t)\big\{\boldsymbol{\beta}^T(W_i(t))\mathbf{Z}_i(t) + \boldsymbol{\gamma}^T\mathbf{V}_i(t)+\alpha(W_i(t))\big\}\,dt - Y_i(t)\,d\Lambda_0(t),$$

we have

$$\mathbf{U}_n(\boldsymbol{\xi}_0,w_0)=\frac{1}{n}\sum_{i=1}^n\int_0^\tau K_h(W_i(t)-w_0)\{\mathbf{Z}^*_i(t,w_0)-\bar{\mathbf{Z}}(t,w_0)\} \times\big[Y_i(t)\big\{\boldsymbol{\beta}^T(W_i(t))\mathbf{Z}_i(t)+\boldsymbol{\gamma}^T\mathbf{V}_i(t)+\alpha(W_i(t)) - \alpha(w_0)-\boldsymbol{\xi}_0^T(w_0)\mathbf{Z}^*_i(t,w_0)\big\}\,dt+dM_i(t)\big].$$

Because

$$\begin{aligned}\boldsymbol{\beta}^T(W_i(t))\mathbf{Z}_i(t)+\boldsymbol{\gamma}^T\mathbf{V}_i(t)+\alpha(W_i(t))\\ =\boldsymbol{\xi}_0^T(w_0)\mathbf{Z}_i^*(t,w_0)+\alpha(w_0)\\ +\frac{1}{2}\big\{(\boldsymbol{\beta}''(w_0))^T\mathbf{Z}_i(t)+\alpha''(w_0)\big\}(W_i(t)-w_0)^2\\ +o_p\big((W_i(t)-w_0)^2\big),\end{aligned}$$

it yields $\mathbf{H}^{-1}\mathbf{U}_n(\boldsymbol{\xi}_0,w_0)=\mathbf{A}_n(\tau,w_0)+\mathbf{B}_n(\tau,w_0)+o_p(h^2)$, where

$$\mathbf{A}_n(\tau,w_0)=\frac{1}{n}\sum_{i=1}^n\int_0^\tau K_h(W_i(t)-w_0)\times\{\mathbf{G}_i^*(t,w_0)-\bar{\mathbf{G}}(t,w_0)\}\,dM_i(t)$$

and

$$\begin{aligned}\mathbf{B}_n(\tau,w_0)=\frac{1}{2n}\sum_{i=1}^n\int_0^\tau K_h(W_i(t)-w_0)(W_i(t)-w_0)^2Y_i(t)\\ \times\{\mathbf{G}_i^*(t,w_0)-\bar{\mathbf{G}}(t,w_0)\}\\ \times\big\{(\boldsymbol{\beta}''(w_0))^T\mathbf{Z}_i(t)+\alpha''(w_0)\big\}\,dt.\end{aligned}$$

Note that $\sqrt{nh}\mathbf{A}_n(\tau,w_0)$ is a sum of local square-integrable martingales with the quadratic variation process given by

$$\begin{aligned}&nh\langle\mathbf{A}_n,\mathbf{A}_n\rangle(s,w_0)\\ &=\frac{h}{n}\sum_{i=1}^n\int_0^s K_h^2(W_i(t)-w_0)\left\{\mathbf{G}_i^*(t,w_0)-\frac{\mathbf{S}_{n1}(t,w_0)}{S_{n0}(t,w_0)}\right\}^{\otimes 2}\\ &\quad\times Y_i(t)\lambda(t|\mathbf{Z}_i,\mathbf{V}_i,W_i)\,dt,\end{aligned}$$

for $0\le s\le\tau$. From Lemma A.1, we have

$$\begin{aligned}&\Bigg|nh\langle\mathbf{A}_n,\mathbf{A}_n\rangle(s,w_0)\\ &\quad-\frac{h}{n}\sum_{i=1}^n\int_0^s K_h^2(W_i(t)-w_0)\left\{\mathbf{G}_i^*(t,w_0)-\frac{\mathbf{s}_1(t,w_0)}{s_0(t,w_0)}\right\}^{\otimes 2}\\ &\quad\times Y_i(t)\lambda(t|\mathbf{Z}_i,\mathbf{V}_i,W_i)\,dt\Bigg|\\ &\le O_p\left(\frac{\log h}{\sqrt{nh}}+h^2\right)\frac{h}{n}\sum_{i=1}^n\int_0^s K_h^2(W_i(t)-w_0)Y_i(t)\\ &\quad\times\lambda(t|\mathbf{Z}_i,\mathbf{V}_i,W_i)\,dt.\end{aligned}$$

Using exactly the same argument, we can easily show that the right side is of order $O_p(\frac{\log h}{\sqrt{nh}}+h^2)$ and so converges in probability to 0, and that

$$\frac{h}{n}\sum_{i=1}^n\int_0^\tau K_h^2(W_i(t)-w_0)\left\{\mathbf{G}_i^*(t,w_0)-\frac{\mathbf{s}_1(t,w_0)}{s_0(t,w_0)}\right\}^{\otimes 2}\times Y_i(t)\lambda(t|\mathbf{Z}_i,\mathbf{V}_i,W_i)\}\,dt\xrightarrow{\mathcal{P}}\boldsymbol{\Sigma}(w_0).$$

Moreover, for any $\delta>0$, because $|K_h(W_i(t)-w_0)(\mathbf{G}_i^*(t,w_0)-\bar{\mathbf{G}}(t,w_0))|$ is bounded by $O(h^{-1})$,

$$I\left\{\sqrt{\frac{h}{n}}\big|K_h(W_i(t)-w_0)Y_i(t)\big(\mathbf{G}_i^*(t,w_0)-\bar{\mathbf{G}}(t,w_0)\big)\big|>\delta\right\}=0$$

as n is sufficiently large. Thus

$$\begin{aligned}&\frac{h}{n}\sum_{i=1}^n\int_0^\tau K_h^2(W_i(t)-w_0)g_{ij}^2(t,w_0)Y_i(t)\lambda(t|\mathbf{Z}_i,\mathbf{V}_i,W_i)\\ &\quad\times I\left\{\sqrt{\frac{h}{n}}\big|K_h(W_i(t)-w_0)Y_i(t)\right.\\ &\quad\left.\times\big(\mathbf{G}_i^*(t,w_0)-\bar{\mathbf{G}}(t,w_0)\big)\big|>\delta\right\}dt\\ &\xrightarrow{\mathcal{P}}0,\end{aligned}$$

where $g_{ij}(t,w_0)$ is the jth element of $\mathbf{G}_i^*(t,w_0)-\overline{\mathbf{G}}(t,w_0)$. Thus, from theorem 5.11 of Fleming and Harrington (1991), we conclude that

$$\sqrt{nh}\mathbf{A}_n(\tau,w_0)\xrightarrow{\mathcal{D}}\mathrm{N}(0,\boldsymbol{\Sigma}(w_0)).\qquad\text{(A.2)}$$

On the other hand, from Lemma A.1,

$$\begin{aligned}\frac{1}{h^2}\mathbf{B}_n(\tau,w_0)&=\frac{1}{2n}\sum_{i=1}^n\int_0^\tau K_h(W_i(t)-w_0)\left(\frac{W_i(t)-w_0}{h}\right)^2\\ &\quad\times Y_i(t)\left\{\mathbf{G}_i^*(t,w_0)-\frac{\mathbf{s}_1(t,w_0)}{s_0(t,w_0)}\right\}\\ &\quad\times\big\{(\boldsymbol{\beta}''(w_0))^T\mathbf{Z}_i(t)+\alpha''(w_0)\big\}\,dt\\ &\quad+O_p\left(\frac{\log h}{\sqrt{nh}}+h^2\right),\\ \mathbf{D}_n(w_0)&=\frac{1}{n}\sum_{i=1}^n\int_0^\tau K_h(W_i(t)-w_0)Y_i(t)\\ &\quad\times\left\{\mathbf{G}_i^*(t,w_0)-\frac{\mathbf{s}_1(t,\omega_0)}{s_0(t,\omega_0)}\right\}^{\otimes 2}dt\\ &\quad+O_p\left(\frac{\log h}{\sqrt{nh}}+h^2\right).\end{aligned}$$

Again, following the same arguments as in Lemma A.1, we have that

$$\frac{1}{h^2}\mathbf{B}_n(\tau,w_0)\xrightarrow{\mathcal{P}}\frac{1}{2}\mu_2\mathbf{b}(w_0),\qquad\mathbf{D}_n(w_0)\xrightarrow{\mathcal{P}}\mathbf{D}(w_0).\qquad\text{(A.3)}$$

Combining all the results of (A.1)–(A.3), we have

$$\sqrt{nh}\big\{\mathbf{H}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}_0)-\tfrac{1}{2}h^2\mu_2\mathbf{D}^{-1}(w_0)\mathbf{b}(w_0)\big\}\xrightarrow{\mathcal{D}}\mathrm{N}\big(\mathbf{0},\mathbf{D}^{-1}(w_0)\boldsymbol{\Sigma}(w_0)\mathbf{D}^{-1}(w_0)\big).$$

It is easy to verify that $\mathbf{D}(w_0)^{-1}\mathbf{b}(w_0)=((\boldsymbol{\beta}''(w_0))^T,\mathbf{0}_{p+q+1}^T)^T$.

Proof of Theorem 2

Note that

$$\begin{aligned}&\widehat{\Lambda}_0^*(t,w_0)-\Lambda_0^*(t,w_0)\\ &=\int_0^t\frac{\sum_{i=1}^n K_h(W_i(u)-w_0)\,dM_i(u)}{\sum_{i=1}^n K_h(W_i(u)-w_0)Y_i(u)}\\ &\quad-(\widehat{\boldsymbol{\xi}}(w_0)-\boldsymbol{\xi}_0(w_0))^T\\ &\quad\times\int_0^t\frac{\sum_{i=1}^n K_h(W_i(u)-w_0)Y_i(u)\mathbf{Z}_i^*(u,w_0)\,du}{\sum_{i=1}^n K_h(W_i(u)-w_0)Y_i(u)}\\ &\quad+\int_0^t\left(\sum_{i=1}^n K_h(W_i(u)-w_0)\{\boldsymbol{\beta}^T(W_i(u))\mathbf{Z}_i(u)+\boldsymbol{\gamma}^T\mathbf{V}_i(u)\right.\end{aligned}$$

$$+\alpha(W_i(u))-\alpha(w_0)-\boldsymbol{\xi}_0^T\mathbf{Z}_i^*(u,w_0)\big\}\,du\Bigg)$$

$$\Bigg/\left(\sum_{i=1}^n K_h(W_i(u)-w_0)Y_i(u)\right).$$

By the Taylor expansion, the last term on the right side is equal to

$$h^2\mu_2\left\{\frac{1}{2}\int_0^t\frac{\mathbf{a}_1(u,w_0)}{a_0(u,w_0)}\,du\,\boldsymbol{\beta}''(w_0)+\frac{1}{2}\alpha''(w_0)t\right\}+o(h^2)$$

uniformly in t and w_0. From (A.1) and the proof of Theorem 1, the first two terms can be written as the summation of local square-integrable martingales and a bias term that equals $-h^2\mu_2\int_0^t\frac{\mathbf{a}_1(u,w_0)}{a_0(u,w_0)}\,du \times\boldsymbol{\beta}''(w_0)/2+o(h^2)$. Finally, using the same arguments as in proving Theorem 1 and the uniform central limit theorem for martingale process, we obtain the result.

Proof of Theorem 3

From the proof of Theorem 1, we have that, uniformly in $w_0\in\mathcal{W}$,

$$\begin{aligned}&\sqrt{nh}\mathbf{H}(\widehat{\boldsymbol{\xi}}(w_0)-\boldsymbol{\xi}_0(w_0))\\&\quad=\mathbf{D}_n^{-1}(w_0)\sqrt{\frac{h}{n}}\sum_{i=1}^n\int_0^\tau K_h(W_i(t)-w_0)\\&\qquad\times\{\mathbf{G}_i^*(t,w_0)-\bar{\mathbf{G}}(t,w_0)\}\,dM_i(t)\\&\qquad+O_p\big(\sqrt{n}h^{5/2}\mathbf{D}^{-1}(w_0)\mathbf{b}(w_0)\big)+o_p\big(\sqrt{n}h^{5/2}\big).\end{aligned}$$

Moreover, because $\mathbf{D}^{-1}(w_0)\mathbf{b}(w_0)$ is 0 except for the first p components, we obtain

$$\begin{aligned}&\sqrt{nh}(\widehat{\boldsymbol{\gamma}}(w_0)-\boldsymbol{\gamma}_0(w_0))\\&\quad=\mathcal{I}\mathbf{D}_n^{-1}(w_0)\sqrt{\frac{h}{n}}\sum_{i=1}^n\int_0^\tau K_h(W_i(t)-w_0)\\&\qquad\times\{\mathbf{G}_i^*(t,w_0)-\bar{\mathbf{G}}(t,w_0)\}\,dM_i(t)+o_p\big(\sqrt{n}h^{5/2}\big).\end{aligned}$$

This implies that

$$\begin{aligned}&\sqrt{n}(\tilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)\\&\quad=\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau\int_{\mathcal{W}}\boldsymbol{\Gamma}(w_0)K_h(W_i(t)-w_0)\mathcal{I}\mathbf{D}_n^{-1}(w_0)\\&\qquad\times\{\mathbf{G}_i^*(t,w_0)-\bar{\mathbf{G}}(t,w_0)\}\,dw_0\,dM_i(t)+o_p(\sqrt{n}h^2).\end{aligned}$$

Because

$$\begin{aligned}&\mathbf{D}_n^{-1}(w_0)\{\mathbf{G}_i^*(t,w_0)-\bar{\mathbf{G}}(t,w_0)\}\\&\quad=\mathbf{D}^{-1}(w_0)\left\{\mathbf{G}_i^*(t,w_0)-\frac{\mathbf{s}_1(t,w_0)}{s_0(t,w_0)}\right\}+O_p\left(\frac{\log h}{\sqrt{nh}}+h^2\right)\end{aligned}$$

uniformly in w_0, the quadratic variance of the martingale

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^n\bigg[\int_{\mathcal{W}}\boldsymbol{\Gamma}(w_0)K_h(W_i(t)-w_0)\mathcal{I}\mathbf{D}_n^{-1}(w_0)\\&\qquad\times\{\mathbf{G}_i^*(t,w_0)-\bar{\mathbf{G}}(t,w_0)\}\,dw_0.\\&\qquad-\int_{\mathcal{W}}\boldsymbol{\Gamma}(w_0)K_h(W_i(t)-w_0)\mathcal{I}\mathbf{D}^{-1}(w_0)\\&\qquad\times\left\{\mathbf{G}_i^*(t,w_0)-\frac{\mathbf{s}_1(t,w_0)}{s_0(t,w_0)}\right\}dw_0\bigg]M_i(t)\end{aligned}$$

is $O_p(\log h/\sqrt{n}h+h^2)$. Thus

$$\begin{aligned}&\sqrt{n}(\tilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)\\&\quad=\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau\int_{\mathcal{W}}\boldsymbol{\Gamma}(w_0)K_h(W_i(t)-w_0)\mathcal{I}\mathbf{D}^{-1}(w_0)\\&\qquad\times\left\{\mathbf{G}_i^*(t,w_0)-\frac{\mathbf{s}_1(t,w_0)}{s_0(t,w_0)}\right\}dw_0\,dM_i(t)+o_p(1).\end{aligned}$$

Furthermore, note that

$$\begin{aligned}&\int_{\mathcal{W}}\boldsymbol{\Gamma}(w_0)K_h(W_i(t)-w_0)\mathcal{I}\mathbf{D}^{-1}(w_0)\\&\qquad\times\left\{\mathbf{G}_i^*(t,w_0)-\frac{\mathbf{s}_1(t,w_0)}{s_0(t,w_0)}\right\}dw_0\\&\quad=\boldsymbol{\Gamma}(W_i(t))\mathcal{I}\mathbf{D}^{-1}(W_i(t))\\&\qquad\times\left[\begin{pmatrix}\mathbf{Z}_i(t)\\\mathbf{V}_i(t)\\\mathbf{0}_{p+1}\end{pmatrix}-\frac{E\{Y(t)(\mathbf{Z}^T(t),\mathbf{V}^T(t),\mathbf{0}_{p+1}^T)^T\}}{E\{Y(t)\}}\right]+O_p(h^2).\end{aligned}$$

This gives

$$\begin{aligned}&\sqrt{n}(\tilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)\\&\quad=\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau\boldsymbol{\Gamma}(W_i(t))\mathcal{I}\mathbf{D}^{-1}(W_i(t))\\&\qquad\times\left[\begin{pmatrix}\mathbf{Z}_i(t)\\\mathbf{V}_i(t)\\\mathbf{0}_{p+1}\end{pmatrix}-\frac{E\{Y(t)(\mathbf{Z}^T(t),\mathbf{V}^T(t),\mathbf{0}_{p+1}^T)^T\}}{E\{Y(t)\}}\right]dM_i(t)\\&\qquad+o_p(1).\end{aligned}$$

Applying the martingale central limit theorem to the right side, the proof is complete.

APPENDIX B: APPROXIMATION OF THE CONFIDENCE BANDS

The arguments follow from appendix B of Tian et al. (2005). We assume that $h=n^{-\nu}$ with $\nu\in[1/4,1/3)$. We note that their function $\mathbf{U}$ corresponds to

$$\begin{aligned}&\frac{1}{\sqrt{nh}}\sum_{i=1}^n\int_0^\tau K\left(\frac{W_i(t)-w_0}{h}\right)\\&\qquad\times\left\{\begin{pmatrix}\mathbf{Z}_i(t)\\\mathbf{V}_i(t)\\\mathbf{Z}_i(t)(W_i(t)-w_0)\\(W_i(t)-w_0)\end{pmatrix}-\bar{\mathbf{Z}}(t,w_0)\right\}dN_i(t)\end{aligned}$$

in our case and thus is independent of parameters. With the exact proofs of their propositions B.3 and B.4, the foregoing process can be represented as

$$\begin{aligned}&\frac{1}{\sqrt{h}}\int_{x,\mathbf{z},\mathbf{v},w}K\left(\frac{w-w_0}{h}\right)\left[\begin{pmatrix}\mathbf{z}\\\mathbf{v}\\\mathbf{z}(w-w_0)\\(w-w_0)\end{pmatrix}-\bar{\mathbf{Z}}(x,w_0)\right]d\big[\sqrt{n}\\&\qquad\times\big\{\widehat{F}_n(x,\mathbf{z},\mathbf{v},w)-E(\widehat{F}_n(x,\mathbf{z},\mathbf{v},w))\big\}\big],\qquad\text{(B.1)}\end{aligned}$$

where $\widehat{F}_n$ is the empirical distribution of the observed data $(X^*=X\Delta+\infty(1-\Delta),\mathbf{Z}(X),\mathbf{V}(X),W(X))$. Using the approximation of the empirical process by the Brownian bridge $\mathcal{B}_n$,

$$\sup_{x,\mathbf{z},\mathbf{v},w}\big|\sqrt{n}\big\{\widehat{F}_n(x,\mathbf{z},\mathbf{v},w)-E(\widehat{F}_n(x,\mathbf{z},\mathbf{v},w))\big\}-\mathcal{B}_n(R(x,\mathbf{z},\mathbf{v},w))\big|=O\big(n^{-1/2}(\log n)^2\big),$$

where R is the quantile function of the observed data, we obtain from integration by parts that (B.1) can be replaced by

$$\frac{1}{\sqrt{h}} \int_{x,\mathbf{z},\mathbf{v},\tilde{w}} \{\mathcal{B}_n(R(x,\mathbf{z},\mathbf{v},w_0+\tilde{w}h)) + O(n^{-1/2}(\log n)^2)\} \times K'(\tilde{w})\,d\tilde{w}\,d_{x,\mathbf{z},\mathbf{v}} \left[\begin{pmatrix}\mathbf{z}\\ \mathbf{v}\\ \mathbf{z}\tilde{w}h\\ \tilde{w}h\end{pmatrix} - \bar{\mathbf{Z}}(x,w_0)\right].$$

From Lemma A.1, this is further approximated by

$$\frac{1}{\sqrt{h}} \int_{x,\mathbf{z},\mathbf{v},\tilde{w}} \{\mathcal{B}_n(R(x,\mathbf{z},\mathbf{v},w_0+\tilde{w}h)) + O(n^{-1/2}(\log n)^2)\} \times K'(\tilde{w})\,d\tilde{w}\,d_{x,\mathbf{z},\mathbf{v}} \left[\begin{pmatrix}\mathbf{z}\\ \mathbf{v}\\ \mathbf{z}\tilde{w}h\\ \tilde{w}h\end{pmatrix} - \frac{\mathbf{s}_1(x,w_0)}{s_0(x,w_0)}\right] + O\left(\frac{\log h}{\sqrt{n}h^{3/2}}\right) + O(h^{3/2}).$$

By integration by parts again, we obtain that (B.1) approximates, uniformly in w_0,

$$\frac{1}{\sqrt{h}} \int_{x,\mathbf{z},\mathbf{v},w} K\left(\frac{w-w_0}{h}\right) \times \left[\begin{pmatrix}\mathbf{z}\\ \mathbf{v}\\ \mathbf{z}(w-w_0)\\ (w-w_0)\end{pmatrix} - \frac{\mathbf{s}_1(x,w_0)}{s_0(x,w_0)}\right] d\mathcal{B}_n(R(x,\mathbf{z},\mathbf{v},w)),$$

which is some kernel-smoothed Wiener process. The latter, by the arguments used in proving proposition B.4 of Tian et al. (2005), has the same distribution as the conditional distribution of the simulated process given the observed data.

[Received September 2006. Revised March 2008.]

REFERENCES

Aalen, O. O. (1989), "A Linear Regression Model for the Analysis of Lifetimes," *Statistics in Medicine*, 8, 907–925.

Andersen, P., and Gill, R. (1982), "Cox's Regression Model for Counting Processes: A Large Sample Study," *The Annals of Statistics*, 10, 1100–1120.

Ang, P.-T., Buzdar, A. U., Smith, T. L., Kau, S., and Hortobagyi, G. N. (1989), "Analysis of Dose Intensity in Doxorubicin-Containing Adjuvant Chemotherapy in Stage II and III Breast Carcinoma," *Journal of Clinical Oncology*, 7, 1677–1684.

Bickel, P. J., and Rosenblatt, M. (1973), "On Some Global Measures of the Deviations of Density Function Estimates," *The Annals of Statistics*, 1, 1071–1095.

Cai, Z., and Sun, Y. (2003), "Local Linear Estimation for Time-Dependent Coefficients in Cox's Regression Models," *Scandinavian Journal of Statistics*, 30, 93–111.

Cai, Z., Fan, J., and Li, R. (2000), "Efficient Estimation and Inference for Varying-Coefficient Models," *Journal of the American Statistical Association*, 95, 888–902.

Chen, S., and Zhou, L. (2006), "Local Partial Likelihood Estimation in Proportional Hazards Regression," *The Annals of Statistics*, 35, 888–916.

Chiang, C.-T., Rice, J. A., and Wu, C. O. (2001), "Smoothing Spline Estimation for Varying Coefficient Models With Repeatedly Measured Dependent Variables," *Journal of the American Statistical Association*, 96, 605–619.

Cox, D. R. (1972), "Regression Models and Life Tables" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 34, 187–220.

Dabrowska, D. M. (1997), "Smoothed Cox Regression," *The Annals of Statistics*, 25, 1510–1540.

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, London: Chapman & Hall.

Fan, J., and Chen, J. (1999), "One-Step Local Quasi-Likelihood Estimation," *Journal of Royal Statistical Society*, Ser. B, 61, 927–943.

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman & Hall.

Fan, J., Gijbels, I., and King, M. (1997), "Local Likelihood and Local Partial Likelihood in Hazard Regression," *The Annals of Statistics*, 25, 1661–1690.

Fan, J., Lin, H. Z., and Zhou, Y. (2006), "Local Partial Likelihood Estimation for Lifetime Data," *The Annals of Statistics*, 34, 290–325.

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

Gamerman, D. (1991), "Dynamic Bayesian Methods for Survival Data," *Applied Statistics*, 40, 63–79.

Hastie, T., and Tibshirani, R. (1993), "Varying-Coefficient Models," *Journal of the Royal Statistical Society*, Ser. B, 55, 757–796.

Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L.-P. (1998), "Nonparametric Smoothing Estimates of Time-Varying Coefficient Models With Longitudinal Data," *Biometrika*, 85, 809–822.

Huang, J. Z., Wu, C. O., and Zhou, L. (2002), "Varying-Coefficient Models and Basis Function Approximations for the Analysis of Repeated Measurements," *Biometrika*, 89, 111–128.

Huffer, F. W., and McKeague, I. W. (1991), "Weighted Least Squares Estimation for Aalen's Additive Risk Model," *Journal of the American Statistical Association*, 86, 114–129.

Kalbfleisch, J. D., and Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data* (2nd ed.), New York: Wiley.

Kosorok, M. R. (2008), *Introduction to Empirical Processes and Semiparametric Inference*, New York: Springer.

Lin, D. Y., and Ying, Z. (1994), "Semiparametric Analysis of the Additive Risk Model," *Biometrika*, 81, 61–71.

Lin, D. Y., Fleming, T. R., and Wei, L. J. (1994), "Confidence Bands for Survival Curves Under the Proportional Hazards Model," *Biometrika*, 81, 73–81.

Martinussen, T., and Scheike, T. H. (2006), *Dynamic Regression Models for Survival Data*, New York: Springer-Verlag.

Martinussen, T., Scheike, T. H., and Skovgaard, I. M. (2002), "Efficient Estimation of Fixed and Time-Varying Covariate Effects in Multiplicative Intensity Models," *Scandinavian Journal of Statistics*, 29, 57–74.

Marzec, L., and Marzec, P. (1997), "On Fitting Cox's Regression Model With Time-Dependent Coefficients," *Biometrika*, 84, 901–908.

McKeague, I. W., and Sasieni, P. (1994), "A Partly Parametric Additive Risk Model," *Biometrika*, 81, 501–514.

Murphy, S. (1993), "Testing for a Time-Dependent Coefficient in Cox's Regression Model," *Scandinavian Journal of Statistics*, 20, 35–50.

Murphy, S., and Sen, P. (1991), "Time-Dependent Coefficients in a Cox-Type Regression Model," *Stochastic Processes and Their Applications*, 39, 153–180.

Nielsen, J. P., and Linton, O. (1995), "Kernel Estimation in a Nonparametric Marker Dependent Hazard Estimation," *The Annals of Statistics*, 5, 1735–1748.

Nielsen, J. P., and Tanggaard, C. (2001), "Simple Boundary and Bias-Correction Kernel Hazard Estimation," *Scandinavian Journal of Statistics*, 28, 695–724.

Peng, L., and Huang, Y. (2007), "Survival Analysis With Temporal Covariate Effects," *Biometrika*, 94, 719–733.

Sun, Y., and Wu, H. (2005), "Semiparametric Time-Varying Coefficients Regression Model for Longitudinal Data," *Scandinavian Journal of Statistics*, 32, 21–47.

Tian, L., Zucker, D., and Wei, L. J. (2005), "On the Cox Model With Time-Varying Regression Coefficients," *Journal of the American Statistical Association*, 100, 172–183.

van der Vaart, A. (1998), *Asymptotic Statistics*, Cambridge, U.K.: Cambridge University Press.

van der Vaart, A., and Wellner, J. (1996), *Weak Convergence and Empirical Processes*, New York: Springer-Verlag.

Yandell, B. S. (1983), "Nonparametric Inference for Rates With Censored Survival Data," *The Annals of Statistics*, 11, 1119–1135.

Zhang, D. (2004), "Generalized Linear Mixed Models With Varying Coefficients for Longitudinal Data," *Biometrics*, 60, 8–15.

Zucker, D., and Karr, A. (1990), "Nonparametric Survival Analysis With Time-Dependent Covariate Effects: A Penalized Partial Likelihood Approach," *The Annals of Statistics*, 18, 329–353.