Global Partial Likelihood for Nonparametric Proportional Hazards Models

Kani CHEN, Shaojun GUO, Liuquan SUN, and Jane-Ling WANG

As an alternative to the local partial likelihood method of Tibshirani and Hastie and Fan, Gijbels, and King, a global partial likelihood method is proposed to estimate the covariate effect in a nonparametric proportional hazards model, $\lambda(t|x) = \exp\{\psi(x)\}\lambda_0(t)$. The estimator, $\hat{\psi}(x)$, reduces to the Cox partial likelihood estimator if the covariate is discrete. The estimator is shown to be consistent and semiparametrically efficient for linear functionals of $\psi(x)$. Moreover, Breslow-type estimation of the cumulative baseline hazard function, using the proposed estimator $\hat{\psi}(x)$, is proved to be efficient. The asymptotic bias and variance are derived under regularity conditions. Computation of the estimator involves an iterative but simple algorithm. Extensive simulation studies provide evidence supporting the theory. The method is illustrated with the Stanford heart transplant data set. The proposed global approach is also extended to a partially linear proportional hazards model and found to provide efficient estimation of the slope parameter. This article has the supplementary materials online.

KEY WORDS: Cox model; Local linear smoothing; Local partial likelihood; Semiparametric efficiency.

1. INTRODUCTION

The Cox proportional hazards model (Cox 1972) is widely used in the analysis of time-to-failure data in biomedical, economic, and social studies. The covariate effect in the Cox model is usually assumed to be log-linear; that is, the logarithm of the hazard function is a linear function of the covariate. The regression parameter retains interpretability and can be easily estimated through the partial likelihood method. The assumption of log-linearity may not hold in practice, however.

A nonparametric proportional hazards model, in which the form of the covariate effect is unspecified, provides a useful variant. Specifically, let T be the survival time and let X be a one-dimensional covariate. The nonparametric proportional hazards model assumes that the conditional hazard of T given $X = x$ takes the form

$$\lambda(t|x) = \exp\{\psi(x)\}\lambda_0(t), \tag{1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function and $\psi(x)$ is an unknown smooth function. Several statistical methods involving smoothing techniques, including nearest-neighbor, spline, and local polynomial smoothing methods, have been developed for this model (see, e.g., Tibshirani and Hastie 1987; O'Sullivan 1988, 1993; Hastie and Tibshirani 1990; Sleeper and Harrington 1990; Gentleman and Crowley 1991; Gray 1992; Kooperberg, Stone, and Truong 1995; Fan, Gijbels, and King 1997; LeBlanc and Crowley 1999; Wang 2004; Huang and Liu 2006; Chen and Zhou 2007). In particular, Tibshirani and Hastie (1987) and Fan, Gijbels, and King (1997) applied nearest-neighbor and local polynomial smoothing methods, respectively, and developed a local partial likelihood approach. The main idea of that approach is quite insightful and nontrivial. The local partial likelihood is used to estimate the derivative of the link function $\psi(x)$, and an estimate of $\psi(x)$ is obtained by integrating the estimated derivative. The reason for estimating the derivative first is that the link function can be identified only up to an additive constant, so only the derivative can be identified. An extra condition, such as $\psi(0) = 0$, is needed to identify the link function. In essence, the local partial likelihood approach corresponds to a particular smoothing method, but it resembles and inherits the major advantages of the partial likelihood approach. Fan, Gijbels, and King (1997) developed applications of local polynomial smoothing, along with asymptotic theory. This approach enjoys the advantages of the local polynomial smoothing method (Fan and Gijbels 1996), such as numerical simplicity and design adaptivity; however, it is not efficient, as we demonstrate in this article. Recently, Chen and Zhou (2007) developed an elegant approach by considering local partial likelihood at two points, x and x_0, where one obtains an estimator for the difference, $\psi(x) - \psi(x_0)$. Aiming at the difference has the advantage that the target is identifiable, and this approach may gain efficiency over the local partial likelihood method. Nevertheless, this approach is still local and similar overall to earlier local partial likelihood methods.

Motivated by the deficiency of all local partial likelihood methods, we propose a global version of partial likelihood. Specifically, this means that all observations are used to estimate ψ at x, in contrast to previous approaches, which use only observations with covariates in the neighborhood of x and/or another point, x_0. This global approach leads to an efficient estimator, $\hat{\psi}$, of the link function ψ in the sense that $\int \phi(x)\hat{\psi}(x)\,dx$ is a semiparametric efficient estimator of $\int \phi(x)\psi(x)\,dx$ for any function ϕ with $\int \phi(x)\,dx = 0$. In addition, we prove that the

Kani Chen is Professor, Department of Mathematics, HKUST, Kowloon, Hong Kong, China (E-mail: *makchen@ust.hk*). Shaojun Guo is Assistant Professor, Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, P.R. China (E-mail: *guoshaoj@amss.ac.cn*). Liuquan Sun is Professor, Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, P.R. China (E-mail: *slq@amt.ac.cn*). Jane-Ling Wang is Professor, Department of Statistics, University of California Davis, Davis, CA 95616 (E-mail: *wang@wald.ucdavis.edu*). The research of K. Chen was supported by the Hong Kong RGC grants (600706 and 600307). The research of S. Guo was supported by Key Laboratory of RCSDS, CAS (No. 2008DP173182). The research of L. Sun was supported in part by the National Natural Science Foundation of China grants (No. 10731010, 10971015 and 10721101), the National Basic Research Program of China (973 Program) (No. 2007CB814902) and Key Laboratory of RCSDS, CAS (No. 2008DP173182). The research of J.-L. Wang was supported in part by the U.S. National Science Foundation (DMS04-06430 and DMS09-06813) and by the National Institute of Health (1R01AG025218-01). The authors gratefully acknowledge the constructive suggestions and comments from the editor, associate editor, and two referees that greatly improved this article.

Journal of the American Statistical Association
June 2010, Vol. 105, No. 490, Theory and Methods
DOI: 10.1198/jasa.2010.tm08636

Breslow-type estimator of the cumulative baseline hazard function, using the proposed estimator $\hat{\psi}$, is semiparametric efficient. The efficiency gain initially seems unusual, because as local smoothing yields optimal procedures for nonparametric regression functions. But the situation is quite different for a hazard-based model, such as the proportional hazards model. To see this, consider first the conventional Cox partial likelihood method for a parametric regression with a known link function but unknown regression parameter β. The partial likelihood function for the regression parameter contributed by each observation involves all subjects who are at risk. In our setting with unknown link function and a single covariate, the counterpart for the regression parameter is the link function ψ. Therefore, one would expect an efficient estimate for ψ to utilize global information rather than be locally constrained.

Although several global procedures using spline smoothing methods have been reported (O'Sullivan 1988; Sleeper and Harrington 1990; Gray 1992; Huang 1999; LeBlanc and Crowley 1999), the aforementioned efficiency of a global smoother has not been pointed out or demonstrated theoretically. One advantage of the local polynomial approach is its design-adaptive feature. For example, when the covariates are concentrated near a few points, our approach, being design-adaptive, reduces naturally to Cox partial likelihood. In contrast, the spline method has difficulty doing this properly, because the observations are too sparse between the few points.

The article is organized as follows. In Section 2 we introduce local and global partial likelihood and present an iterative algorithm for computing the proposed estimates. The main idea is to derive estimating functions directly from the partial likelihood, rather than from the local partial likelihood. In contrast to the local partial likelihood method, this approach reduces to the partial likelihood approach when the covariate has a discrete distribution. Asymptotic properties and semiparametric efficiency of the estimates are presented in Section 3. The optimal bandwidth is found to be of order $n^{-1/5}$, the same as for standard local linear smoothing in nonparametric regression. Section 4 presents simulation results, evaluating the finite-sample properties of the new method and comparing the global procedure with two local procedures by Fan, Gijbels, and King (1997) and Chen and Zhou (2007). An analysis of the Stanford heart transplant data set with the global procedure is reported. In Section 5, the methodology developed in Section 2 for a single covariate is extended to a partially linear proportional hazards model. A brief discussion is given in Section 6. All proofs are relegated to the Appendix.

2. LOCAL AND GLOBAL PARTIAL LIKELIHOOD

In the presence of censoring, let C be the censoring variable, and assume that T and C are conditionally independent given X, but that the distribution of C may depend on X. Let $\tilde{T} = \min(T, C)$ be the event time and $\delta = I(T \le C)$ be the failure/censoring indicator, where $I(A)$ is an indicator function. Let $(T_i, C_i, \tilde{T}_i, \delta_i, X_i)$, for $i = 1, \ldots, n$, be iid copies of $(T, C, \tilde{T}, \delta, X)$. The observations are $\{(\tilde{T}_i, \delta_i, X_i), i = 1, \ldots, n\}$. Define $N_i(t) = \delta_i I(\tilde{T}_i \le t)$ and $Y_i(t) = I(\tilde{T}_i \ge t)$. The (global) partial likelihood, due to Cox (1972), is

$$\prod_{i=1}^{n} \left[\frac{\exp\{\psi(X_i)\}}{\sum_{j=1}^{n} Y_j(\tilde{T}_i) \exp\{\psi(X_j)\}} \right]^{\delta_i}. \tag{2}$$

Using local linear approximation of ψ, Tibshirani and Hastie (1987) suggested a local partial likelihood near a value $x \in \mathcal{R}$, given by

$$\prod_{\substack{i=1,\ldots,n \\ X_i \in B_n(x)}} \left[\frac{\exp\{\alpha + \beta(X_i - x)\}}{\sum_{\substack{j=1,\ldots,n \\ X_j \in B_n(x)}} Y_j(\tilde{T}_i) \exp\{\alpha + \beta(X_j - x)\}} \right]^{\delta_i}, \tag{3}$$

where $\alpha = \psi(x), \beta = \psi'(x)$, and $B_n(x)$ is a neighborhood around x. This is essentially the partial likelihood (2), restricted to observations for which the covariates are near x. Note that the local partial likelihood is actually free of α, because α cancels out in the ratio in (3). This is in accordance with the identifiability of ψ up to a shift. As a result, $\psi(x)$ is not directly estimable from the local partial likelihood, but the derivative $\psi'(x)$ can be estimated by maximizing (3). Tibshirani and Hastie (1987) suggested a trapezoidal rule as an ad hoc version of integration to obtain an estimate of $\psi(x)$, but no theory is available for this estimate.

A much more refined version of local partial likelihood estimation through local polynomial smoothing was proposed and studied by Fan, Gijbels, and King (1997). Let h be a bandwidth, $K(x)$ be a kernel function with support on $[-1, 1]$, and $K_i(x) = h^{-1}K\{(X_i - x)/h\}$. With a local polynomial smoother of order p, the logarithm of the local partial likelihood of Fan, Gijbels, and King (1997) is given by

$$\sum_{i=1}^{n} \delta_i K_i(x) \Bigg\{ \alpha + \beta_1(X_i - x) + \cdots + \beta_p(X_i - x)^p - \log\Bigg[\sum_{j=1}^{n} Y_j(\tilde{T}_i) \times \exp\{\alpha + \beta_1(X_j - x) + \cdots + \beta_p(X_j - x)^p\} K_j(x) \Bigg] \Bigg\}, \tag{4}$$

where $\beta_k = \psi^{(k)}(x)/k!$ and $\psi^{(k)}(x)$ is the kth derivative of $\psi(x)$. Note that (3) is a special case of (4) for $p = 1$ and the uniform kernel if $B_n(x)$ corresponds to the interval $[x-h, x+h]$. Similar to (3), the logarithm of the local partial likelihood in (4) is also free of α, so maximizing (4) leads to estimates of the derivatives of ψ. Fan, Gijbels, and King (1997) suggested estimating $\psi(b) - \psi(a)$ by integrating the estimate of $\psi'(x)$ from a to b, and the final estimate of $\psi(x)$ is the same as the estimate for $\psi(x) - \psi(0)$. They also provided a comprehensive theoretical justification of this method and closed-form expressions for asymptotic bias and variance of the derivative estimates. Chen and Zhou (2007) took an alternative approach, aiming directly at the differences $\psi(x) - \psi(x_0)$, using local partial likelihood in the neighborhood of both x and x_0. They proposed a two-step algorithm, requiring the choice of a bandwidth for the estimates of $\psi'(x)$ and $\psi'(x_0)$ in a first step and then of another bandwidth for estimating $\psi(x) - \psi(x_0)$ in a second step. Asymptotic properties, including asymptotic bias and variance, were derived. Because this procedure still utilizes a local partial likelihood, improvement is possible by using a global partial likelihood method instead, as we demonstrate herein.

The global approach is motivated as follows. Given an x, if we know ψ except in a neighborhood of x, we can estimate $\psi(x)$ by maximizing

$$\prod_{i=1}^{n}\Bigg\{\exp\big[\{\alpha+\beta(X_i-x)\}I\{X_i\in B_n(x)\}+\psi(X_i)I\{X_i\notin B_n(x)\}\big] \Big/\Bigg(\sum_{\substack{j=1,\ldots,n\\X_j\in B_n(x)}} Y_j(\tilde{T}_i)\exp\{\alpha+\beta(X_j-x)\} + \sum_{\substack{j=1,\ldots,n\\X_j\notin B_n(x)}} Y_j(\tilde{T}_i)\exp\{\psi(X_j)\}\Bigg)\Bigg\}^{\delta_i}. \quad (5)$$

This is a crude nearest-neighbor (global) partial likelihood, analogous to (3) but it also uses the information available at data outside the neighborhood region $B_n(x)$. It follows from (5) that the score functions of α and β are

$$\sum_{i=1}^{n}\int_0^\infty\left[I\{X_i\in B_n(x)\} - \frac{\sum_{\substack{j=1,\ldots,n\\X_j\in B_n(x)}} Y_j(t)\exp\{\alpha+\beta(X_j-x)\}}{S_n(\alpha,\beta,t;\psi,x)}\right]dN_i(t)$$

and

$$\sum_{i=1}^{n}\int_0^\infty\left[(X_i-x)I\{X_i\in B_n(x)\} - \frac{\sum_{\substack{j=1,\ldots,n\\X_j\in B_n(x)}} (X_j-x)Y_j(t)\exp\{\alpha+\beta(X_j-x)\}}{S_n(\alpha,\beta,t;\psi,x)}\right]dN_i(t),$$

where $S_n(\alpha,\beta,t;\psi,x)=\sum_{\substack{j=1,\ldots,n\\X_j\in B_n(x)}} Y_j(t)\exp\{\alpha+\beta(X_j-x)\}+\sum_{\substack{j=1,\ldots,n\\X_j\notin B_n(x)}} Y_j(t)\exp\{\psi(X_j)\}$. Note that the first term is asymptotically negligible relative to the second term in $S_n(\alpha,\beta,t;\psi,x)$. Replacing the nearest-neighbor smoother by a local linear smoother for computational convenience, which amounts to replacing $I\{X_i\in B_n(x)\}$ with $K_i(x)$, we arrive at the following asymptotic equations:

$$\sum_{i=1}^{n}\int_0^\infty\left[K_i(x) - \frac{\sum_{j=1}^{n} K_j(x)\exp\{\alpha+\beta(X_j-x)\}Y_j(t)}{\sum_{j=1}^{n}\exp\{\psi(X_j)\}Y_j(t)}\right]dN_i(t)=0 \quad (6)$$

and

$$\sum_{i=1}^{n}\int_0^\infty\left[(X_i-x)K_i(x) - \frac{\sum_{j=1}^{n}(X_j-x)K_j(x)\exp\{\alpha+\beta(X_j-x)\}Y_j(t)}{\sum_{j=1}^{n}\exp\{\psi(X_j)\}Y_j(t)}\right]dN_i(t) = 0. \quad (7)$$

Because the true ψ is unknown, an iterative procedure is needed to solve (6) and (7).

Algorithm.

Let $\psi_{(m)}$ denote the mth iteration and fix $\psi_{(m)}(X_n)=0$ for all $m\geq 0$.

Initialization ($m=0$). Choose initial values for $\psi_{(0)}(x)$ at $x=X_1,\ldots,X_{n-1}$.

Iteration step from $m-1$ *to* m. For every given $x=X_1,\ldots,X_{n-1}$. Let $\hat{\alpha}(x)$ and $\hat{\beta}(x)$ be the solutions of α and β in eqs. (6) and (7), where ψ is replaced by $\psi_{(m-1)}$. Then $\psi_{(m)}(X_i)=\hat{\alpha}(X_i)$ for $i=1,\ldots,n-1$.

Final estimates. This iteration is continued until convergence, and for every x, the final estimates of $\psi(x)$ and $\psi'(x)$, denoted as $\hat{\psi}(x)$ and $\hat{\psi}'(x)$, are obtained as the solutions of α and β for eqs. (6) and (7) at convergence.

Remark 1. We fix $\psi_{(m)}(X_n)=0$, because ψ is not identifiable and can be identified only up to a constant shift. Therefore, only differences of the function ψ at various arguments are estimable from the data. Another possibility would be to require $\psi(0)=0$; then the final estimate of $\psi(x)$ would be $\hat{\psi}(x)-\hat{\psi}(0)$.

Remark 2. The proposed estimate reduces to the partial likelihood estimate when the covariate assumes only finitely many (distinct) values, say $a_1,\ldots,a_K$. To see this, assume that $\psi(a_K)=0$ for the purpose of identifiability. Then, as long as the bandwidth h is smaller than $\min(|a_k-a_l|, l\neq k)$, $X_i-X_j=0$ for $|X_i-X_j|\leq h$. As a result, (7) is always satisfied, and the limit of (6) reduces to

$$\sum_{i=1}^{n}\int_0^\infty\left[I(X_i=a_k) - \frac{\sum_{j=1}^{n} I(X_j=a_k)\exp(\alpha_k)Y_j(t)}{\sum_{l=1}^{K}\sum_{j=1}^{n} I(X_j=a_l)\exp(\alpha_l)Y_j(t)}\right]dN_i(t)=0 \quad (8)$$

for $k=1,\ldots,K-1$ and $\alpha_K=0$, where $\alpha_k=\psi(a_k)$. This is exactly the same as the Cox partial likelihood estimating equation, and thus the solution of α_k is the partial likelihood estimator of $\psi(a_k)$. This observation lends support to the optimality of the proposed global method of estimation. We note that the estimates of Tibshirani and Hastie (1987), Fan, Gijbels, and King (1997), and Chen and Zhou (2007) do not reduce to the Cox partial likelihood estimates in this case, and neither do any spline methods.

3. ASYMPTOTIC THEORY

In this section we assume that the random variable X is bounded with compact support. Without loss of generality, let the support be [0, 1]. Additional regularity conditions are stated in the Appendix. For identifiability and without loss of generality, we fix $\hat{\psi}(0)=\psi(0)=0$. We start with the uniform consistency of $\hat{\psi}(x)$.

Theorem 1. Suppose that the regularity conditions (C1)–(C7) stated in the Appendix hold. Then $\sup_{x\in[0,1]}|\hat{\psi}(x)-\psi(x)|\to 0$ in probability, as $n\to\infty$.

We now introduce some notation to show the asymptotic expression of $\hat{\psi}(x)-\psi(x)$ at each fixed point $x\in(0,1)$. Let $f(x)$ denote the density of the random variable X, and let $\psi''(x)$ be the second derivative of $\psi(x)$. Set $\mu_2=$

$\int_{-1}^{1} u^2K(u)\,du$, $\nu = \int_{-1}^{1} K^2(u)\,du$, $P(t|x) = P(\tilde{T} \ge t|X = x)$, $\Gamma(x) = \int_0^\tau P(t|x)\lambda_0(t)\,dt$, $s_0(t) = E[P(t|X)\exp\{\psi(X)\}]$, and

$$\Phi(u|x) = \{\Gamma(x)\}^{-1} f(u)\exp\{\psi(u)\} \times \int_0^\tau P(t|u)P(t|x)\{s_0(t)\}^{-1}\lambda_0(t)\,dt.$$

Theorem 2. Under the regularity conditions (C1)–(C7) in the Appendix, for any fixed point $x \in (0, 1)$, $\hat{\psi}(x) - \psi(x)$ satisfies the Fredholm integral equation

$$\hat{\psi}(x) - \psi(x) = \int_0^1 \Phi(u|x)\{\hat{\psi}(u) - \psi(u)\}\,du + \frac{1}{2}h^2\mu_2\psi''(x) + (nh)^{-1/2}\xi_n(x) + o_p\big(h^2 + (nh)^{-1/2}\big), \quad (9)$$

where the o_p term depends on the point x and $\xi_n(x)$ converges to the normal distribution with mean 0 and variance

$$\sigma^2(x) = \nu\big[f(x)\Gamma(x)\exp\{\psi(x)\}\big]^{-1}.$$

Remark 3. Let $\mathcal{A}$ be the linear operator satisfying $\mathcal{A}(H)(x) = \int_0^1 \Phi(u|x)H(u)\,du$ for any function $H(\cdot)$, and let $\mathcal{I}$ be the identity operator. Then (9) can be written as

$$(\mathcal{I} - \mathcal{A})(\hat{\psi} - \psi)(x) = \tfrac{1}{2}h^2\mu_2\psi''(x) + (nh)^{-1/2}\xi_n(x) + o_p\big(h^2 + (nh)^{-1/2}\big), \quad (10)$$

which means that the order of the asymptotic bias of $(\mathcal{I} - A)(\hat{\psi} - \psi)(x)$ is $O(h^2)$. If nh^5 is bounded, which is a regular condition for bandwidths in standard nonparametric function estimation, then it follows from (10) that the optimal bandwidth for $(\mathcal{I} - A)(\hat{\psi} - \psi)(x)$ is of order $n^{-1/5}$ and $\hat{\psi}$ attains the optimal rate of convergence.

To evaluate the optimality of the global partial likelihood approach, in the following theorem we provide a justification via semiparametric efficiency. Define $\mathcal{D}_0 = \{\phi : \phi(x)$ has a continuous second derivative on $[0, 1]$ and $\int_0^1 \phi(x)\,dx = 0\}$. For any fixed function $\phi \in \mathcal{D}_0$, $\int_0^1 \hat{\psi}(x)\phi(x)\,dx$ is naturally an estimator of $\int_0^1 \psi(x)\phi(x)\,dx$.

Theorem 3. Suppose that the regularity conditions (C1)–(C6) and (C7′) in the Appendix are satisfied. Then for any function $\phi \in \mathcal{D}_0$, we have

$$\sqrt{n}\left\{\int_0^1 \hat{\psi}(x)\phi(x)\,dx - \int_0^1 \psi(x)\phi(x)\,dx\right\} \xrightarrow{\mathcal{L}} \mathrm{N}(0, \tilde{\sigma}^2),$$

where $\tilde{\sigma}^2$ is as defined in (A.17). Furthermore, $\int_0^1 \hat{\psi}(x)\phi(x)\,dx$ is a semiparametric efficient estimator of $\int_0^1 \psi(x)\phi(x)\,dx$.

Remark 4. Note that $\int_0^1 \phi(x)\,dx = 0$ is an identifiability condition for estimating $\int_0^1 \psi(x)\phi(x)\,dx$, because ψ is identifiable only up to an additive constant. The semiparametric efficiency in Theorem 3 is meaningful in the sense that any functional $\int_0^1 \psi(x)\phi(x)\,dx$ of the link function $\psi(x)$ can be estimated with semiparametric efficiency by the corresponding functional of the estimate $\hat{\psi}(x)$. (See Klaassen, Lee, and Ruymgaart 2005 for a related type of efficiency based on functionals.) In addition, it is noteworthy that Theorem 3 implies the semiparametric efficiency of the cumulative hazard function as presented in Theorem 4.

Let $\Lambda_0(t) = \int_0^t \lambda_0(s)\,ds$ be the cumulative baseline hazard function. Using the global likelihood estimate $\hat{\psi}$, we can estimate $\Lambda_0(t)$ by the so-called Breslow estimator,

$$\hat{\Lambda}_0(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n \exp\{\hat{\psi}(X_j)\}Y_j(s)}.$$

The following theorem gives the efficiency for $\hat{\Lambda}_0(t)$, which means that using our estimator $\hat{\psi}$ should produce efficient estimation for $\Lambda_0(t)$, whereas using others, such as those of Fan, Gijbels, and King (1997) or Chen and Zhou (2007), would not.

Theorem 4. Suppose that the regularity conditions (C1)–(C6) and (C7′) in the Appendix are satisfied. Then for any bounded measurable function $b(t)$, $\int_0^\tau b(t)\,d\hat{\Lambda}_0(t)$ is a semiparametric efficient estimator of $\int_0^\tau b(t)\,d\Lambda_0(t)$. In particular, $\hat{\Lambda}_0(t)$ is a semiparametric efficient estimator of $\Lambda_0(t)$ for any $t \in [0, \tau]$.

4. NUMERICAL STUDIES

4.1 Simulations

In this section we report simulation studies regarding the finite-sample performance of the global partial likelihood method (designated GPL hereinafter). We compare the performance of our method with that of the local partial likelihood methods of Fan, Gijbels, and King (1997) (FGK hereinafter) and Chen and Zhou (2007) (CZ hereinafter).

In the numerical examples that follow, Weibull baseline hazard functions of the form $\lambda_0(t) = 3\lambda t^2$ are used, and the censoring time C is distributed uniformly in $[0, a(x)]$ given the covariate $X = x$, where $a(x) = \exp\{c_1 I(\psi(x) > b)/3 + c_2 I(\psi(x) \le b)/3\}$, with the constants λ, c_1, c_2, and b chosen such that the total censoring rate is about 30%–40%. The Epanechnikov kernel $K(u) = 0.75(1 - u^2)I(|u| \le 1)$ was used in all simulations. The results presented herein are based on 500 replications and sample size $n = 200$. We consider the following six settings, in which ρ is a Bernoulli random variable with $P(\rho = 1) = P(\rho = 0) = 0.5$:

Model 1. $X \sim \mathrm{Uniform}[-1, 1]$ and $\psi(x) = x$.

Model 2. $X \sim \rho\mathrm{N}(-0.6, 0.3^2) + (1 - \rho)\mathrm{N}(0.6, 0.3^2)$ and further truncated outside $[-1, 1]$, and $\psi(x) = x$.

Model 3. $X \sim \mathrm{Uniform}[-1, 1]$ and $\psi(x) = x^3$.

Model 4. $X \sim \rho\mathrm{N}(-0.6, 0.3^2) + (1 - \rho)\mathrm{N}(0.6, 0.3^2)$ and further truncated outside $[-1, 1]$, and $\psi(x) = x^3$.

Model 5. $X \sim \mathrm{Uniform}[-2, 2]$ and $\psi(x) = 4\sin(2x)$.

Model 6. $X \sim \rho\mathrm{N}(-1, 0.5^2) + (1 - \rho)N(1, 0.5^2)$ and further truncated outside $[-2, 2]$, and $\psi(x) = 4\sin(2x)$.

We chose $c_1 = 11, c_2 = 5, b = 0$, for all of the models, and $\lambda = e^{-4.5}$ in Models 1 and 2, $\lambda = e^{-4}$ in Models 3 and 4, and $\lambda = e^{-5}$ in Models 5 and 6. Models 1 and 2 correspond to the

Table 1. Comparison between the global and local approaches based on the WMISE in Models 1–2

	Model 1			Model 2		
Bandwidth h	GPL	FGK	CZ $h_0 = 0.7$ ($1.25h$)	GPL	FGK	CZ $h_0 = 0.7$ ($1.25h$)
0.30	0.095	0.196	0.087 (0.238)	0.127	0.333	0.126 (0.148)
0.40	0.068	0.133	0.068 (0.142)	0.077	0.159	0.076 (0.090)
0.50	0.053	0.101	0.059 (0.115)	0.052	0.104	0.054 (0.065)
0.60	0.044	0.083	0.059* (0.088)	0.038	0.073	0.046 (0.054)
0.70	0.039	0.072	0.064 (0.081)	0.032	0.057	0.044* (0.050)
0.80	0.034	0.066	0.076 (0.076)	0.027	0.050	0.047 (0.047)
0.90	0.030	0.062	0.091 (0.069)	0.024	0.048	0.053 (0.043)
1.00	0.026*	0.058*	0.111 (0.068*)	0.022*	0.046*	0.066 (0.039*)

NOTE: (a) $*$ stands for the minimum WMISE. (b) Two initial bandwidths h_0 are specified for CZ. (c) Average standard errors of the WMISE for GPL, FGK, and CZ in Model 1 are 0.0250, 0.049, and 0.050, respectively and in Model 2, 0.020, 0.041, and 0.038, respectively.

conventional Cox model with identity ψ-function, and Models 3 and 4 correspond to situations with more rapidly increasing risks associated with the covariate values. We did not include decreasing link functions, because the sign of X can be flipped in Models 1–4 to make the link function decreasing. Whereas it is unlikely that a link function will have both a local maximum and a local minimum, as in the settings of Models 5 and 6, we included these to allow comparison of all three methods under a nonmonotone link function.

We next describe how we computed the estimators based on various methods with the same constraint $\hat{\psi}(0) = 0$. For GPL, we first calculated the curve estimates $\hat{\psi}(x)$ based on (6) and (7) with constraints $\sum_{i=1}^{n} \hat{\psi}(X_i) = 0$, and then replaced $\hat{\psi}(x)$ by $\hat{\psi}(x) - \hat{\psi}(0)$. For FGK, we obtained the estimates through $\hat{\psi}(x) = \int_0^x \hat{\psi}'(t)\,dt$. For CZ, as suggested by Chen and Zhou (2007), we used a two-step procedure to first calculate estimates for $\psi'(x)$ as in FGK using a bandwidth, h_0. We then calculated the estimates of $\psi(x) - \psi(x_0)$ at the second step using another bandwidth h, where x_0 is a prespecified point, and obtained the final estimates by $\hat{\psi}(x) - \hat{\psi}(0)$. We chose $x_0 = 0$ for CZ in Models 1, 3, 5, and 6 and $x_0 = -0.6$ in Models 2 and 4. The choices of the initial bandwidth are prespecified at $h_0 = 0.7, 0.88, 1.0$, respectively, in Tables 1–3, because these values provide good estimates for $\hat{\psi}'$. In the second step, we adopted a different bandwidth h, as specified in the first columns of Tables 1–3, to obtain the final estimate of $\psi(x) - \psi(x_0)$. Chen and Zhou (2007) proposed using $h_0 = 1.25h$, and we included their suggestion for comparison purposes.

The performance of the various estimators $\hat{\psi}(x)$ was assessed via the weighted mean integrated squared error (WMISE),

$$WMISE = E \int_{-a}^{a} \{\hat{\psi}(x) - \psi(x)\}^2 w(x)\,dx,$$

where $w(\cdot)$ is taken as the density function of X. We chose $a = 1.0$ in Models 1–4 and $a = 2.0$ in Models 5 and 6.

Tables 1–3 summarize the results for the WMISE of Models 1–6 under various bandwidth choices, with WMISE reported as an average over the 500 replications. As can be seen from the tables, the minimum WMISE of GPL is smaller than that of FGK and CZ, except for Model 6, where the minimum WMISE of CZ is slightly smaller than that of GPL. This confirms the benefit of the GPL approach compared with local approaches. We also see that GPL performs better when the covariate x has a uniform distribution than when it has a normal mixture distribution. When $\psi(x) = x$, as in Table 1, a large bandwidth should be chosen to estimate the link function and GPL and FGK both attain their minimum WMISEs at the largest bandwidth, $h = 1.0$. The optimal bandwidth h for CZ depends on the choice of the prespecified bandwidth h_0, however. Table 3 also shows that the optimal bandwidth for GPL is smaller than those for FGK and CZ. This is likely because the local methods need to enlarge the included range of data to compensate for the use of less data information than the global procedure. For Models 5 and 6, we also tried $x_0 = -1$ and 1 for CZ, but the results were not as good as those shown in Table 3 and thus are not presented here. The performance

Table 2. Comparison between the global and local approaches based on the WMISE in Models 3–4

	Model 3			Model 4		
Bandwidth h	GPL	FGK	CZ $h_0 = 0.88$ ($1.25h$)	GPL	FGK	CZ $h_0 = 0.88$ ($1.25h$)
0.30	0.081	0.169	0.079 (0.223)	0.106	0.234	0.102 (0.128)
0.40	0.061	0.120	0.067* (0.118)	0.071	0.126	0.070 (0.077)
0.50	0.050	0.097	0.069 (0.100)	0.052	0.088	0.055 (0.061)
0.60	0.044	0.085	0.083 (0.093)	0.042	0.070	0.052* (0.054*)
0.70	0.041	0.079	0.109 (0.092*)	0.038	0.064*	0.058 (0.054)
0.80	0.039	0.077	0.144 (0.093)	0.037	0.067	0.069 (0.057)
0.90	0.037	0.076	0.213 (0.094)	0.037	0.072	0.081 (0.058)
1.00	0.035*	0.075*	0.257 (0.094)	0.036*	0.078	0.099 (0.060)

NOTE: (a) $*$ stands for the minimum WMISE. (b) Two initial bandwidths h_0 are specified for CZ. (c) Average standard errors of the WMISE for GPL, FGK, and CZ in Model 3 are 0.029, 0.062, and 0.046, respectively and in Model 4, 0.024, 0.052, and 0.042, respectively.

Table 3. Comparison between the global and local approaches based on the WMISE in Models 5–6

	Model 5			Model 6		
Bandwidth h	GPL	FGK	CZ $h_0 = 1.0$ (1.25h)	GPL	FGK	CZ $h_0 = 1.0$ (1.25h)
0.20	0.449	13.53	7.529 (20.99)	–	–	–
0.25	0.256*	6.807	3.441 (5.106)	10.07	34.51	13.91 (49.97)
0.30	0.260	4.218	1.238 (2.161)	0.764	23.46	8.107 (29.58)
0.35	0.318	2.716	0.421 (1.463)	0.605*	12.49	3.106 (18.53)
0.40	0.420	1.276	0.299* (0.627)	0.765	6.405	1.276 (2.250)
0.45	0.556	1.015	0.308 (0.371)	1.125	3.355	0.606 (1.219)
0.50	0.719	0.851	0.336 (0.346*)	1.177	2.022	0.590* (0.938)
0.55	0.902	0.747	0.378 (0.3762)	1.402	1.399	0.671 (0.677*)
0.60	1.100	0.683	0.434 (0.449)	1.630	1.096	0.740 (0.703)
0.65	1.310	0.646	0.510 (0.557)	1.855	0.912	0.872 (0.812)
0.70	1.527	0.634*	0.592 (0.698)	2.080	0.816	1.114 (1.003)
0.75	1.750	0.643	0.700 (0.879)	2.302	0.778*	1.309 (1.263)
0.80	1.977	0.674	0.807 (1.098)	2.522	0.794	1.500 (1.622)

NOTE: (a) $*$ stands for the minimum WMISE. (b) Two initial bandwidths h_0 are specified for CZ. (c) Average standard errors of the WMISE for GPL, FGK, and CZ in Model 5 are 0.253, 0.588, and 0.301, respectively and in Model 6, 0.293, 0.535, and 0.438, respectively.

of CZ depends heavily on the bandwidth choice; the values given in parentheses in Table 3 reflect this variation. The results in Tables 1–3 show that our bandwidth choices generally lead to better performance for CZ than the previously suggested $h_0 = 1.25h$ (reported in parentheses).

Figure 1 shows the biases of the different estimates based on the optimal bandwidths of Tables 1–3. It can be seen that the estimated curves based on all three methods are close to the true curves, reflecting the effectiveness of the global and local methods. To appreciate the sample variability of the estimated nonparametric functions at each point, we also calculated the pointwise standard errors at some grid points based on 500 replications. The results showed that the proposed GPL method had better accuracy than FGK and CZ, which were very similar

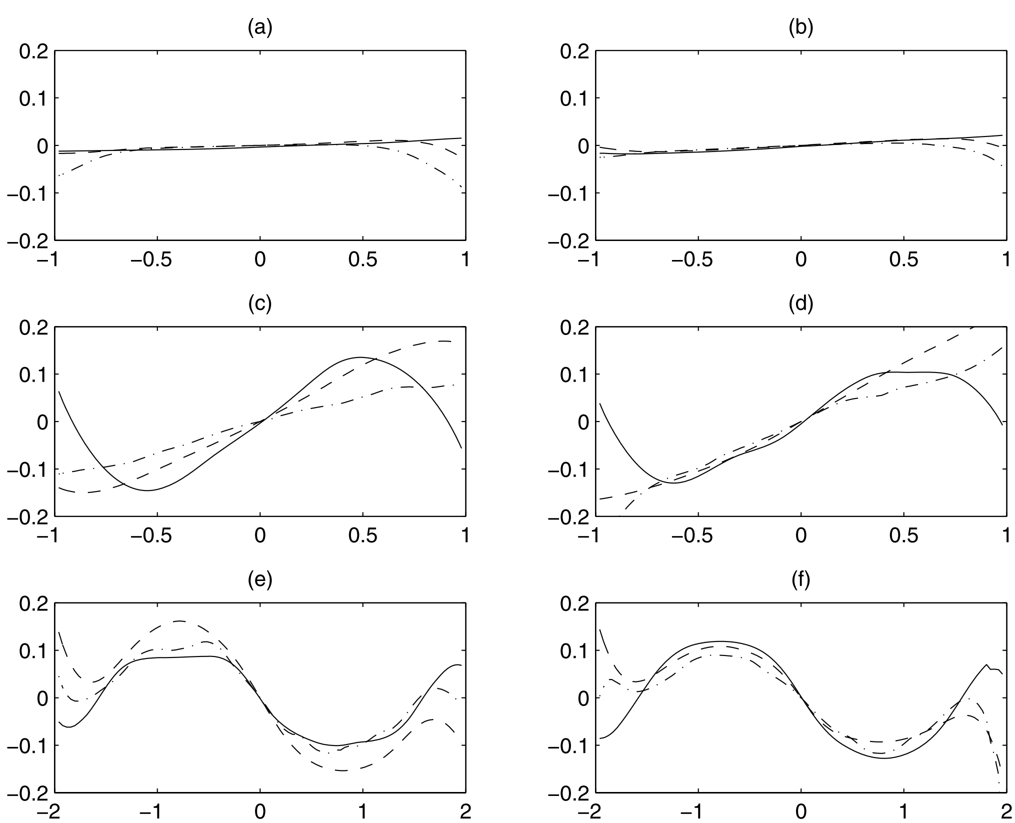

Figure 1. Biases of the estimators based on 500 replications in Models 1–6. In (a)–(f), the solid curve is based on GPL, the dashed curve is based on FGK, and the dash-dotted curve is based on CZ. (a) and (b) We take $h_{GPL} = h_{FGK} = h_{CZ} = 1.0$ ($h_0 = 1.25$). (c) We take $h_{GPL} = h_{FGK} = 0.70$, $h_{CZ} = 0.40$ ($h_0 = 0.80$). (d) We take $h_{GPL} = h_{FGK} = 0.70$, $h_{CZ} = 0.60$ ($h_0 = 0.80$). (e) We take $h_{GPL} = 0.25$, $h_{FGK} = 0.70$, and $h_{CZ} = 0.40$ ($h_0 = 0.70$). (f) We take $h_{GPL} = 0.35$, $h_{FGK} = 0.75$, and $h_{CZ} = 0.50$ ($h_0 = 0.70$).

to the summary results of the WMISE reported in Tables 1–3. To save space, we do not report them here.

In summary, FGK is usually less efficient than CZ, which requires the choice of prespecified h_0 and x_0; different choices may yield very different final estimates. In contrast, GPL does not require such prespecified choices and is generally more stable and efficient than CZ. Therefore, the simulation results clearly demonstrate the advantages of the GPL method and also support the theoretical findings of its efficiency. In another simulation not reported here, we compared the GPL method with the regression spline smoother of Leblanc and Crowley (1999). The two procedures are comparable with GPL slightly better than the regression spline method in some cases.

4.2 Applications

We now illustrate the proposed method with the well-known Stanford Heart Transplant data set. In this data, 184 of the 249 patients received heart transplants from October 1967 to February 1980. Patients alive beyond February 1980 were considered censored. More details about this data and some related work in the literature can (Crowley and Hu 1977; Miller and Halpern 1982). Previous analyses have included quadratic functions of age (in years) at transplantation. Instead of speculating which order of polynomials or other parametric functions would work for the data, a nonarametric link function on age, such as the one proposed in this article, is a good way to explore the data structure. Fan and Gijbels (1994) and Cai (2003) also reanalyzed the data using nonparametric regression.

Following the common literature, we limit our analysis to the 157 patients who had completed tissue typing. Among the 152 patients with complete records and survival time exceeding 10 days, 55 were censored, constituting a censoring proportion of 36%. The estimates of $\psi(x)$ with the GPL method and bandwidths $h = 7$ and 10 are presented in Figure 2. These estimates suggest that the risk decreases with age at transplantation for younger patients (age < 20), remain constant (nearly 0) in the middle age range (20–40), and then increases with age for older patients (age > 40). Whereas further tests are needed to confirm these effects, a quadratic function of age is not suitable to model the age effects. Instead, a piecewise linear function with breakpoints at age 20 and 40 (the breakpoints could be estimated as well, but we have not done so) and zero risks between the breakpoints might be a better alternative parametric model. The solid curve in Figure 2, based on such a piecewise linear fit for the ψ-function, supports this speculation, as well as that the risk associated with the age at transplantation is relevant only for younger and older patients.

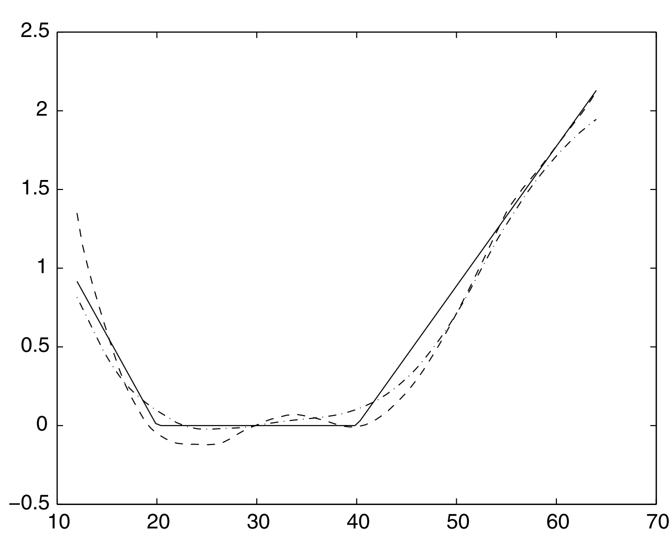

Figure 2. Estimates of $\psi(x)$ for the Stanford Heart Transplant Data (the dashed curve is based on GPL with bandwidth $h = 7$, the dashed-dotted curve is based on GPL with bandwidth $h = 10$, and the solid curve is based on partial likelihood with a piecewise linear function joined at age 20 and 40 and set to be zero in between 20 and 40).

5. GENERALIZATION

In many applications, there will be more than one covariate. The global approach developed in previous sections can be extended to many important cases with high-dimensional covariates, for instance, partially linear proportional hazards models, single index proportional hazards models and partially linear additive models and so on. In this regard, the global approach can be viewed as a building block for general semiparametric or nonparametric approaches. As an illustration, in this section we show how to extend the methodology to provide efficient estimation of the slope parameter in a partially linear proportional hazards model.

In addition to X, let $\mathbf{Z}$ be another vector covariate, whose effects on the risk follows the parametric link function in conventional Cox model. This leads to the partially linear proportional hazards model with the following conditional hazard of T given $(\mathbf{Z}, X) = (\mathbf{z}, x)$:

$$\lambda(t|\mathbf{z}, x) = \exp\{\boldsymbol{\theta}^\top \mathbf{z} + \psi(x)\}\lambda_0(t), \tag{11}$$

where $\boldsymbol{\theta}$ is a p-vector of unknown regression parameters and $\psi(x)$ is an unknown smooth function like the one in model (1). Let $(T_i, C_i, \tilde{T}_i, \delta_i, \mathbf{Z}_i, X_i)$, for $i = 1, \ldots, n$, be iid copies of $(T, C, \tilde{T}, \delta, \mathbf{Z}, X)$. The observations are $\{(\tilde{T}_i, \delta_i, \mathbf{Z}_i, X_i),\ i = 1, \ldots, n\}$.

For any fixed $\boldsymbol{\theta}$, model (11) is a nonparametric proportional hazards model. Therefore, we can use the global method to estimate $\psi(x)$, denoted by $\hat{\psi}(x; \boldsymbol{\theta})$, by solving the following equations, similar to (6) and (7):

$$\sum_{i=1}^n \int_0^\infty \Bigg[K_i(x) - \sum_{j=1}^n K_j(x) \exp\{\boldsymbol{\theta}^\top \mathbf{Z}_j + \alpha(x) + \beta(x)(X_j - x)\} Y_j(t) \Bigg/ \Bigg(\sum_{j=1}^n \exp\{\boldsymbol{\theta}^\top \mathbf{Z}_j + \alpha(X_j)\} Y_j(t) \Bigg) \Bigg] dN_i(t) = 0, \tag{12}$$

$$\sum_{i=1}^n \int_0^\infty \Bigg[(X_i - x) K_i(x) - \sum_{j=1}^n (X_j - x) K_j(x) \times \exp\{\boldsymbol{\theta}^\top \mathbf{Z}_j + \alpha(x) + \beta(x)(X_j - x)\} Y_j(t) \Bigg/ \Bigg(\sum_{j=1}^n \exp\{\boldsymbol{\theta}^\top \mathbf{Z}_j + \alpha(X_j)\} Y_j(t) \Bigg) \Bigg] dN_i(t) = 0. \tag{13}$$

To estimate $\boldsymbol{\theta}$, we can use the partial likelihood due to Cox (1972) and replace $\psi(x)$ with its estimator $\hat{\psi}(x;\boldsymbol{\theta})$ obtained earlier. This leads to the following equation:

$$\mathbf{U}_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\int_0^\infty \Bigg[\mathbf{Z}_i - \frac{\sum_{j=1}^{n}\mathbf{Z}_j \exp\{\boldsymbol{\theta}^\top \mathbf{Z}_j + \hat{\psi}(X_j;\boldsymbol{\theta})\}Y_j(t)}{\sum_{j=1}^{n}\exp\{\boldsymbol{\theta}^\top \mathbf{Z}_j + \hat{\psi}(X_j;\boldsymbol{\theta})\}Y_j(t)}\Bigg]\,dN_i(t), \quad (14)$$

whose solution is $\hat{\boldsymbol{\theta}}_n$. This process can be iterated a few times until the algorithm converges.

Step 0. Choose $\boldsymbol{\theta}^{(0)}$.

Step m. For $m \geq 1$, compute $\boldsymbol{\theta}^{(m)}$ as follows. First, replace $\boldsymbol{\theta}$ with $\boldsymbol{\theta}^{(m-1)}$ in (12) and (13), and apply the iteration therein to obtain $\hat{\psi}(X_j;\boldsymbol{\theta}^{(m-1)})$ $(j = 1,\dots,n-1)$. Then substitute $\hat{\psi}(X_j;\boldsymbol{\theta}^{(m-1)})$ into (14), and solve $\mathbf{U}_n(\boldsymbol{\theta}) = 0$ to obtain $\boldsymbol{\theta}^{(m)}$.

Repeat this iteration procedure until convergence, and the final $\boldsymbol{\theta}^{(m)}$ is the estimate $\hat{\boldsymbol{\theta}}_n$, whose asymptotic efficiency is outlined in the next paragraph.

Let $\boldsymbol{\theta}_0$ be the true value of $\boldsymbol{\theta}$ and $\tilde{M}(t) = \delta I(\tilde{T} \leq t) - \int_0^t Y(u)\exp\{\boldsymbol{\theta}_0^\top \mathbf{Z} + \psi(X)\}\lambda_0(u)\,du$, where $Y(t) = I(\tilde{T} \geq t)$. Define $\xi_{\mathbf{Z}} = \int_0^\tau [\mathbf{Z} - E(\mathbf{Z}|\tilde{T} = t, \delta = 1)]\,d\tilde{M}(t)$ and $\xi_g = \int_0^\tau [g(X) - E(g(X)|\tilde{T} = t, \delta = 1)]\,d\tilde{M}(t)$ for any smooth function g, and $\tau = \inf\{t : P(\tilde{T} > t) = 0\}$.

Theorem 5. Under the regularity conditions (C1)–(C5) and (C6′)–(C8′) stated in the Appendix, we have that $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ converges in distribution to a mean-0 normal random vector with covariance matrix $\boldsymbol{\Sigma}_0^{-1}$, where $\boldsymbol{\Sigma}_0 = \mathrm{var}(\xi_{\mathbf{Z}} - \xi_{\mathbf{g}^*})$, and $\xi_{\mathbf{g}^*} = \Pi(\xi_{\mathbf{Z}}|\xi_g$ for all smooth g) is the projection of $\xi_{\mathbf{Z}}$ onto the linear space spanned by ξ_g for all smooth functions g. Furthermore, $\hat{\boldsymbol{\theta}}_n$ is semiparametric efficient in the sense that its asymptotic variance attains the semiparametric information bound for $\hat{\boldsymbol{\theta}}_n$.

It is noteworthy that the global feature of the partial likelihood approach turns out to be the key to obtaining a semiparametric efficient estimator of $\boldsymbol{\theta}$, whereas a local partial likelihood method, such as that of Fan, Gijbels, and King (1997) and Chen and Zhou (2007), cannot. In addition, the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_n$ can be consistently estimated by $\boldsymbol{\Sigma}_n^{-1} = -\frac{1}{n}[\frac{\partial \mathbf{U}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n}]^{-1}$.

To evaluate the finite-sample performance of the proposed GPL method relative to the local partial likelihood methods (e.g., Fan, Gijbels, and King 1997; Chen and Zhou 2007), we conducted a small simulation study. We chose $\lambda_0(t) = 3e^{-5}t^2$ and $\psi(x) = 2\sin(2x)$, and chose the covariate Z to be a Bernoulli random variable with success probability 0.5 and X to be a uniform random variable distributed on [0, 4]. The distribution of the censoring time, C, was similar to that reported in Section 4, except that $c_1 = 10$, $c_2 = 6$, and $b = 0$. The Epanechnikov kernel $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$ was used in all simulations, and the results reported here are based on 500 replications. The sample size was 200, and various bandwidths were used to reflect the different levels of smoothness. Table 4 presents the simulation results on estimation of $\theta_0 = 0.5$. The table includes the Monte Carlo biases (bias) and standard deviations (SD) of $\hat{\theta}_n$, as well as relative efficiencies (REs) between the GPL and FGK/CZ methods. The RE was calculated as the ratio of the sampling variance estimator of $\hat{\theta}_n$ based on FGK or CZ to that of GPL. It can be seen from Table 4 that the proposed procedure performed very well for the situations considered here. Compared with the results of local partial likelihood methods such as FGK and CZ, our proposed estimators are more stable and efficient. In addition, Table 4 also shows that our proposed method is insensitive to the choice of bandwidth.

6. CONCLUDING REMARKS

We have proposed a GPL method for estimating the covariate effect in a nonparametric proportional hazards model. The estimation procedure uses all of the data, and the proposed estimates are consistent and semiparametrically efficient in the sense described in Theorem 3. The proposed algorithm involves iteration but is easy to implement, and it converges toward a solution of the score equations. The estimation procedure reduces to the Cox partial likelihood approach when the covariate is discrete, and the simulation results demonstrate that the proposed methods work well for the situations considered herein. For the Stanford Heart Transplant data set, our approach leads to interesting findings that shed new light on these data. We also have shown how our methodology can be extended to partially linear proportional hazards models and obtain efficient estimation of the slope parameter. A small simulation study demonstrated that the new GPL method performs very well and is more stable and efficient than local partial likelihood methods.

One important issue that we have not conclusively addressed in this article is the choice of the bandwidth for the nonparametric estimate of $\psi(x)$. This is a difficult question for hazard-based regression models, and one that has eluded many researchers who have studied nonparametric Cox models. Some model selection procedures have been suggested by Tibshrani and Hastie (1987), and another possibility is to minimize the asymptotic mean integrated square error. The former is an ad hoc approach, and the latter requires explicit expressions for the asymptotic bias and variance of the estimator and might not work well in practice because of the need to truncate certain

Table 4. Comparison between the global and local procedures with $\theta_0 = 0.5$ in the partially linear proportional hazards model

	GPL	FGK	CZ	RE
(n, h)	Bias (SD)	Bias (SD)	Bias (SD)	(FGK, CZ)
(200, 0.25)	0.018 (0.202)	0.077 (0.361)	−0.029 (0.422)	(3.208, 4.386)
(200, 0.35)	−0.002 (0.189)	0.055 (0.264)	−0.023 (0.268)	(1.955, 2.007)
(200, 0.45)	−0.019 (0.181)	0.013 (0.220)	−0.040 (0.187)	(1.480, 1.072)

NOTE: Bias is the mean of the estimates of θ_0 minus θ_0 and SD is the standard deviation of the estimates of θ_0.

boundary regions. One promising direction is the development of a cross-validation approach based on the GPL, similar to the that proposed by O'Sullivan (1988). This is a separate project warranting further investigation. For the time being, we rely on a subjective method to visually determine the proper amount of smoothness, as illustrated in the real data example.

APPENDIX: PROOFS

Here we outline the proofs of Theorems 1–5. We provide more detail in the supplementary technical report.

A.1 Regularity Conditions and Notations

We first state the following regularity conditions:

- (C1) The kernel function $K(x)$ is a symmetric density function with compact support $[-1, 1]$ and continuous derivative.
- (C2) Let $\tau = \inf\{t: P(\tilde{T} > t) = 0\}$. τ is finite, $P(T > \tau) > 0$, and $P(C = \tau) > 0$.
- (C3) X is bounded with compact support $[0, 1]$, and $P(C = 0|X = x) < 1$.
- (C4) The density function $f(x)$ of X is bounded away from 0 and has a continuous second-order derivative on $[0, 1]$.
- (C5) The function $\psi(x)$ is twice continuously differentiable on $[0, 1]$.
- (C6) The conditional probability $P(t|x)$ is continuous in $(t, x) \in [0, \tau] \times [0, 1]$ and has a continuous second derivative of x on $[0, 1]$ for any $t \in [0, \tau]$, where $P(t|x)$ is as defined in Section 3.
- (C6′) $\boldsymbol{\rho}_k(t|x, \boldsymbol{\theta})$ $(k = 0, 1, 2)$ are continuous at $(t, x, \boldsymbol{\theta})$, where $(t, x, \boldsymbol{\theta}) \in [0, \tau] \times [0, 1] \times \boldsymbol{\Theta}$, and $\boldsymbol{\Theta}$ is a neighborhood of $\boldsymbol{\theta}_0$, $\boldsymbol{\rho}_k(t|x, \boldsymbol{\theta}) = E\{\mathbf{Z}^{\otimes k} \exp\{\boldsymbol{\theta}^\top \mathbf{Z}\} Y(t)|X = x\}$ for $k = 0, 1, 2$, and for a vector $\mathbf{v}$, $\mathbf{v}^{\otimes 0} = 1$, $\mathbf{v}^{\otimes 1} = v$, and $\mathbf{v}^{\otimes 2} = \mathbf{v}\mathbf{v}^\top$.
- (C7) $h^2 \log n \to 0$ and $nh/(\log n)^2 \to \infty$, as $n \to \infty$.
- (C7′) $nh^4 \to 0$ and $nh^2/(\log n)^2 \to \infty$, as $n \to \infty$.
- (C8′) Let $\mathbf{s}_k(t; \boldsymbol{\theta}) = E[\{\mathbf{Z} - \mathbf{g}^*(X)\}^{\otimes k} \exp\{\boldsymbol{\theta}^\top \mathbf{Z} + \psi(X)\} Y(t)]$ for $k = 0, 1, 2$, where $\mathbf{g}^*(x)$ is as defined in Theorem 5. The matrix $\boldsymbol{\Sigma}_0$ is positive definite, where

$$\boldsymbol{\Sigma}_0 = \int_0^\tau \left[\frac{\mathbf{s}_2(t; \boldsymbol{\theta}_0)}{\mathbf{s}_0(t; \boldsymbol{\theta}_0)} - \left\{\frac{\mathbf{s}_1(t; \boldsymbol{\theta}_0)}{\mathbf{s}_0(t; \boldsymbol{\theta}_0)}\right\}^{\otimes 2}\right] \mathbf{s}_0(t; \boldsymbol{\theta}_0) \lambda_0(t)\, dt.$$

Let the superscript $a^\top$ denote the transpose of a vector a. For $i = 1, \ldots, n$, define

$$\begin{aligned} \boldsymbol{\omega}_i(x) &= (\omega_{i1}(x), \omega_{i2}(x))^\top, \\ \tilde{\omega}_{nk}(x) &= \frac{1}{n} \sum_{i=1}^n K_i(x) \{(X_i - x)/h\}^k, \qquad k = 0, 1, 2, \\ \omega_{i1}(x) &= K_i(x) \{\tilde{\omega}_{n2}(x) - \tilde{\omega}_{n1}(x)(X_i - x)/h\} \\ &\quad /\{\tilde{\omega}_{n2}(x)\tilde{\omega}_{n0}(x) - \tilde{\omega}_{n1}^2(x)\}, \qquad \text{and} \\ \omega_{i2}(x) &= K_i(x) \{\tilde{\omega}_{n0}(x)(X_i - x)/h - \tilde{\omega}_{n1}(x)\} \\ &\quad /\{\tilde{\omega}_{n2}(x)\tilde{\omega}_{n0}(x) - \tilde{\omega}_{n1}^2(x)\}. \end{aligned}$$

Define $\mu_k(x) = \int_{u \in \mathcal{Z}(x,h)} u^k K(u)\, du$, $k = 0, 1, 2, 3$, where $\mathcal{Z}(x, h) = \{u: x - hu \in [0, 1]\} \cap [-1, 1]$. If $\mathcal{Z}(x, h) = [-1, 1]$, then write $\mu_k = \mu_k(x)$. It is seen that $\mu_0 = 1, \mu_1 = \mu_3 = 0$. For any functions η and γ defined on $[0, 1]$, set $S_{n0}(t; \eta) = \frac{1}{n} \sum_{i=1}^n Y_i(t) \exp\{\eta(X_i)\}$, and $\mathbf{S}_{n1}(t; \eta, \gamma, x) = \frac{1}{n} \sum_{i=1}^n \boldsymbol{\omega}_i(x) Y_i(t) \exp\{\eta(x) + \gamma(x)(X_i - x)/h\}$. Then $(\hat{\psi}, h\hat{\psi}')$ is the solution to the equation $\mathbf{U}(\eta, \gamma; x) = 0$, where

$$\mathbf{U}(\eta, \gamma; x) = n^{-1} \sum_{i=1}^n \int_0^\tau \left[\boldsymbol{\omega}_i(x) - \frac{\mathbf{S}_{n1}(t; \eta, \gamma, x)}{S_{n0}(t; \eta)}\right] dN_i(t). \qquad (A.1)$$

Let $s_0(t; \eta) = E[P(t|X) \exp\{\eta(X)\}]$, $\mathbf{s}_1(t; \eta, \gamma, x) = P(t|x) \exp\{\eta(x)\} \times (\xi_1(x, \gamma), \xi_2(x, \gamma))^\top / \{\mu_2(x)\mu_0(x) - \mu_1^2(x)\}$, $\xi_1(x, \alpha) = \mu_2(x) \times \int_{\mathcal{Z}(x,h)} \exp(\alpha u) K(u)\, du - \mu_1(x) \int_{\mathcal{Z}(x,h)} \exp(\alpha u) u K(u)\, du$, and $\xi_2(x, \alpha) = \mu_0(x) \int_{\mathcal{Z}(x,h)} \exp(\alpha u) u K(u)\, du - \mu_1(x) \int_{\mathcal{Z}(x,h)} \exp(\alpha u) K(u)\, du$. Define

$$\begin{aligned} \mathbf{u}(\eta, \gamma; x) &= \Gamma(x) \exp\{\psi(x)\}(1, 0)^\top \\ &\quad - \int_0^\tau \frac{\mathbf{s}_1(t; \eta, \gamma, x)}{s_0(t; \eta)} s_0(t; \psi) \lambda_0(t)\, dt, \\ \mathcal{B}_n &= \big\{\eta: \|\eta\| \le D, |\eta(x) - \eta(y)| \le d[|x - y| + b_n], \\ &\qquad x, y \in [0, 1]\big\}, \quad \text{and} \\ \mathcal{C}_0 &= \{\eta: x \in [0, 1], \\ &\qquad \eta(0) = 0, \eta(x) \text{ is continuous on } [0, 1]\}, \end{aligned}$$

where $b_n = h + (nh)^{-1/2} (\log n)^{1/2}$, $D > 0$, $d > 0$. Let $U_1(\eta, \gamma; x)$ and $\mathbf{u}_1(\eta, \gamma; x)$ denote the first entries of $\mathbf{U}(\eta, \gamma; x)$ and $\mathbf{u}(\eta, \gamma; x)$, respectively.

A.2 Proof of Theorem 1

Set $\hat{\gamma}(x) = h\hat{\psi}'(x)$, and $\bar{N}(t) = n^{-1} \sum_{i=1}^n N_i(t)$. First, we show that

$$\mathbf{u}(\eta, \gamma; x) \equiv 0 \quad \text{with } \eta \in \mathcal{C}_0 \quad \Longleftrightarrow \quad \eta \equiv \psi, \qquad \gamma \equiv 0. \qquad (A.2)$$

Next, we show that

$$\sup_{x \in [0,1], \eta, \gamma \in \mathcal{B}_n} \|\mathbf{U}(\eta, \gamma; x) - \mathbf{u}(\eta, \gamma; x)\| = o_p(1). \qquad (A.3)$$

It follows from the uniform strong law of large numbers (Pollard 1990, p. 41) that for fixed continuous functions η and γ,

$$\sup_{x \in [0,1]} \|\mathbf{U}(\eta, \gamma; x) - \mathbf{u}(\eta, \gamma; x)\| = o_p(1). \qquad (A.4)$$

By construction of the ε-net of $\mathcal{B}_n$, it is seen that

$$\log N(\varepsilon, \mathcal{B}_n, \|\cdot\|_\infty) \le O\left(\frac{1}{\varepsilon} \log\left(\frac{1}{\varepsilon}\right)\right) = o(n), \qquad (A.5)$$

where $N(\varepsilon, \mathcal{B}_n, \|\cdot\|_\infty)$ is the covering number of the class $\mathcal{B}_n$ with the norm $\|\cdot\|_\infty$. Thus (A.3) follows by (A.4), (A.5), and theorem 8.2 of Pollard (1990). Applying theorem 2.37 of Pollard (1984) and then using the mean value theorem, we have

$$P\{\hat{\psi} \in \mathcal{B}_n, \hat{\gamma} \in \mathcal{B}_n\} \to 1. \qquad (A.6)$$

Using the Arzela–Ascoli theorem and (A.6), we show that for any subsequence of $\{(\hat{\psi}, \hat{\gamma})\}$, there exists a further convergent subsequence $(\hat{\psi}_{nk}, \hat{\gamma}_{nk})$ such that, uniformly over $x \in [0, 1]$, $(\hat{\psi}_{nk}, \hat{\gamma}_{nk}) \to (\psi^*, \gamma^*)$ in probability, with $\psi^* \in \mathcal{C}_0$. Note that

$$\begin{aligned} \mathbf{u}(\psi^*, \gamma^*; x) &= \big\{\mathbf{u}(\psi^*, \gamma^*; x) - \mathbf{u}\big(\hat{\psi}_{n_k}, \hat{\gamma}_{n_k}; x\big)\big\} \\ &\quad + \big\{\mathbf{u}\big(\hat{\psi}_{n_k}, \hat{\gamma}_{n_k}; x\big) - U\big(\hat{\psi}_{n_k}, \hat{\gamma}_{n_k}; x\big)\big\}. \end{aligned}$$

It follows from (A.3) and $\mathbf{U}(\hat{\psi}, \hat{\gamma}; x) \equiv 0$ that the left side converges to 0 uniformly over $x \in [0, 1]$. As a result, $\mathbf{u}(\psi^*, \gamma^*; x) = 0$. Because $\mathbf{u}(\eta, \gamma; x) = 0$ has a unique solution $(\psi, 0)$ from (A.2), we have $\psi^* \equiv \psi$ on $[0, 1]$, which ensures the uniform consistency of $\hat{\psi}(x)$. The proof of Theorem 1 is complete.

A.3 Proof of Theorem 2

Observe that $\mathbf{U}(\hat{\psi}, h\hat{\psi}'; x) - \mathbf{U}(\psi, h\psi'; x) \equiv -\mathbf{U}(\psi, h\psi'; x)$. By Taylor expansion, the consistency of $(\hat{\psi}, h\hat{\psi}')$, and the empirical approximation as theorem 2.37 of Pollard (1984), it can be shown that, uniformly over $x \in [0,1]$,

$$\hat\psi(x) - \psi(x) - \frac{1}{n}\sum_{j=1}^{n}\exp\{\psi(X_j)\}[\hat\psi(X_j) - \psi(X_j)] \times \left[\int_0^\tau Y_j(t)P(t|x)\{\Gamma(x)s_0(t)\}^{-1}\lambda_0(t)\,dt\right] = \left[\Gamma(x)\exp\{\psi(x)\}\right]^{-1}U_1(\psi, h\psi'; x) + o_p\big(h^2 + (nh)^{-1/2}\big). \tag{A.7}$$

Furthermore, (A.7) implies that

$$\sup_{x\in[0,1]}\left|\frac{1}{n}\sum_{i=1}^{n}\exp\{\psi(X_i)\}[\hat\psi(X_i) - \psi(X_i)] \times \left[\int_0^\tau Y_i(t)P(t|x)\{\Gamma(x)s_0(t)\}^{-1}\lambda_0(t)\,dt\right] - \int_0^\tau \Phi(u|x)\{\hat\psi(u) - \psi(u)\}\,du\right| = o_p\big(h^2 + (nh)^{-1/2}\big). \tag{A.8}$$

Therefore, it is seen from (A.7) and (A.8) that

$$\hat\psi(x) - \psi(x) - \int_0^\tau \Phi(u|x)\{\hat\psi(u) - \psi(u)\}\,du = [\Gamma(x)\exp\{\psi(x)\}]^{-1}U_1(\psi, h\psi'; x) + o_p\big(h^2 + (nh)^{-1/2}\big), \tag{A.9}$$

uniformly over $x \in [0,1]$. To prove (9), it remains to show the asymptotic expansion of $U_1(\psi, h\psi'; x)$. Toward this end, let $S_{n1}(t;x)$ and $s_1(t;x)$ be the first entries of $\mathbf{S}_{n1}(t;\psi, h\psi', x)$ and $\mathbf{s}_1(t;\psi, 0, x)$, respectively. Note that $M_i(t) = N_i(t) - \int_0^t Y_i(s)\exp\{\psi(X_i)\}\lambda_0(s)\,ds$. Then $U_1(\psi, h\psi'; x)$ can be expressed as

$$U_1(\psi, h\psi'; x) = V_n(x) + B_n(x), \tag{A.10}$$

where $V_n(x) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau [\omega_{i1}(x) - \frac{S_{n1}(t;x)}{S_{n0}(t;\psi)}]\,dM_i(t)$ and

$$B_n(x) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau\left[\omega_{i1}(x) - \frac{S_{n1}(t;x)}{S_{n0}(t;\psi)}\right]Y_i(t)\exp\{\psi(X_i)\}\lambda_0(t)\,dt.$$

The martingale central limit theorem (Fleming and Harrington 1991, thm. 5.3.5) implies that

$$\sqrt{nh}V_n(x) = \sqrt{nh}\cdot\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau\left[\omega_{i1}(x) - \frac{s_1(t;x)}{s_0(t;\psi)}\right]dM_i(t) + o_p(1), \tag{A.11}$$

is asymptotically normal with mean 0 and variance $\sigma_0^2(x) = \Gamma(x) \times \exp\{\psi(x)\}\nu/f(x)$. On the other hand, it follows from the Taylor expansion that

$$B_n(x) = \tfrac{1}{2}h^2\mu\Gamma(x)\exp\{\psi(x)\}\psi''(x) + o_p(h^2). \tag{A.12}$$

Therefore, (9) follows from (A.9)–(A.12). The proof of Theorem 2 is complete.

A.4 Proof of Theorem 3

For a fixed function $\phi \in \mathcal{D}_0$, let $g \in \mathcal{C}_0$ satisfy the integral equation

$$[g(x)\Gamma(x) - \Lambda_g(x)]f(x)\exp\{\psi(x)\} = \phi(x), \tag{A.13}$$

where $\mathcal{D}_0$ is as given in Section 3, $\mathcal{C}_0$ is as defined in (A.2), $\Lambda_g(x) = \int_0^\tau \bar g(t)P(t|x)\lambda_0(t)\,dt$, and $\bar g(t) = E[P(t|X)\exp\{\psi(X)\}g(X)]/s_0(t)$. It can be verified that there exists a unique solution $g \in \mathcal{C}_0$ with a continuous second derivative to eq. (A.13).

First, we derive the asymptotic distribution of $\int_0^1 \hat\psi(x)\phi(x)\,dx$. Under the conditions $nh^4 \to 0$ and $nh^2/(\log n)^2 \to \infty$, a similar argument for (A.9) leads to

$$\hat\psi(x) - \psi(x) - \int_0^\tau \Phi(u|x)\{\hat\psi(u) - \psi(u)\}\,du = \left[\Gamma(x)\exp\{\psi(x)\}\right]^{-1}U_1(\psi, h\psi'; x) + o_p\big(n^{-1/2}\big), \tag{A.14}$$

uniformly over $x \in [0,1]$. It follows from (A.13) and (A.14) that

$$\int_0^1\{\hat\psi(x) - \psi(x)\}\phi(x)\,dx = \int_0^1 U_1(\psi, h\psi'; x)f(x)g(x)\,dx + o_p\big(n^{-1/2}\big). \tag{A.15}$$

Let $S_{ng}(t) = n^{-1}\sum_{i=1}^n Y_i(t)\exp\{\psi(X_i)\}g(X_i)$. Note that $nh^4 \to 0$. As a result,

$$\int_0^1 U_1(\psi, h\psi'; x)f(x)g(x)\,dx = n^{-1}\sum_{i=1}^{n}\int_0^\tau\left[g(X_i) - \frac{S_{ng}(t)}{S_{n0}(t)}\right]dM_i(t) + o_p\big(n^{-1/2}\big).$$

Applying the martingale central limit theorem (Fleming and Harrington 1991, thm. 5.3.5), we have

$$\sqrt{n}\int_0^1\{\hat\psi(x) - \psi(x)\}\phi(x)\,dx \xrightarrow{\mathcal{L}} \mathrm{N}(0, \tilde\sigma^2), \tag{A.16}$$

where $\bar g(t) = E\{P(t|X)\exp\{\psi(X)\}g(X)\}/s_0(t)$ and

$$\tilde\sigma^2 = E\left[\int_0^\tau\{g(X) - \bar g(t)\}^2 P(t|X)\exp\{\psi(X)\}\lambda_0(t)\,dt\right]. \tag{A.17}$$

To show the asymptotic efficiency of $\int_0^1 \hat\psi(x)\phi(x)\,dx$, we consider a parametric submodel with parameter θ, such that $\psi(x;\theta) = \psi(x) + \theta g(x)$, and $\lambda(t;\theta) = \lambda_0(t) - \theta\bar g(t)\lambda_0(t)$, where θ is an unknown parameter and $g(x)$ is the solution to the integral equation (A.13) for a given $\phi \in \mathcal{D}_0$. Let $\theta_0 = 0$ be the true value of θ. The score for this parametric submodel at θ_0 is

$$\int_0^\tau\{g(X) - \bar g(t)\}\,dM(t)$$

with variance $\tilde\sigma^2$, where $M(t) = N(t) - \int_0^t I(\tilde T \ge s)\exp\{\psi(X)\}\lambda_0(s)\,ds$. Thus the maximum likelihood estimator of θ, denoted by $\tilde\theta$, satisfies

$$\sqrt{n}(\tilde\theta - \theta_0) \to \mathrm{N}(0, \tilde\sigma^{-2}).$$

Observe that $\int_0^1 g(x)\phi(x)\,dx = E[\int_0^\tau g(X)\{g(X) - \bar g(t)\}\,dN(t)] = \tilde\sigma^2$. It follows that

$$\sqrt{n}\int_0^1\{\psi(x;\tilde\theta) - \psi(x;\theta_0)\}\phi(x)\,dx \xrightarrow{\mathcal{L}} \mathrm{N}(0, \tilde\sigma^2). \tag{A.18}$$

This, together with (A.16) and (A.18), shows that the asymptotic variance of $\int_0^1 \hat\psi(x)\phi(x)\,dx$ is the same as that of $\int_0^1 \psi(x;\tilde\theta)\phi(x)\,dx$. As explained by Bickel et al. (1993, p. 46), $\int_0^1 \hat\psi(x)\phi(x)\,dx$ is asymptotically efficient for the estimation of $\int_0^1 \psi(x)\phi(x)\,dx$. This completes the proof of Theorem 3.

A.5 Proof of Theorem 4

Recall the definitions of $s_0(t) = E[\exp\{\psi(X)\}Y(t)]$ and $P(t|x) = E\{Y(t)|X = x\}$. Set $\phi(x) = \int_0^\tau b(t)\exp\{\psi(x)\}P(t|x)\lambda_0(t)/s_0(t)\,dt\,f(x)$. Let g be the solution of (A.21), which can be reexpressed as

$$\int_0^\tau\{g(x) - \bar g(t) - b(t)/s_0(t)\}\exp\{\psi(x)\}P(t|x)\lambda_0(t)\,dt = 0.$$

Consequently,

$$\operatorname{cov}\left(\int_0^\tau \{g(X) - \bar{g}(t) - b(t)/s_0(t)\}\, dM(t), \int_0^\tau g(X)\, dM(t)\right) = 0. \tag{A.19}$$

To show the asymptotic normality of $\int_0^\tau b(t)\, d\hat{\Lambda}_0(t)$, using a Taylor expansion and the uniform strong law of large numbers, we get

$$\begin{aligned}
&\int_0^\tau b(t)\, d\hat{\Lambda}_0(t) - \int_0^\tau b(t)\, d\Lambda_0(t) \\
&= \frac{1}{n}\int_0^\tau \frac{b(t)}{s_0(t)} \sum_{i=1}^n dM_i(t) - \int_0^1 \{\hat{\psi}(x) - \psi(x)\} \\
&\quad \times \int_0^\tau \frac{b(t)\exp\{\psi(x)\}P(t|x)\lambda_0(t)}{s_0(t)}\, dt\, f(x)\, dx + o_p(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{\frac{b(t)}{s_0(t)} - g(X_i) + \bar{g}(t)\right\} dM_i(t) + o_p(n^{-1/2}),
\end{aligned}$$

where the last equality is due to (A.14)–(A.15). As a result,

$$\sqrt{n}\left(\int_0^\tau b(t)\, d\hat{\Lambda}_0(t) - \int_0^\tau b(t)\, d\Lambda_0(t)\right) \to \mathrm{N}(0, \sigma^2),$$

where, by (A.19), $\sigma^2 = \int_0^\tau b(t)\{\frac{b(t)}{s_0(t)} + \bar{g}(t)\}\lambda_0(t)\, dt$. Using the same arguments as in the proof of Theorem 3, we also obtain that the estimator is semiparametric efficient. The proof of Theorem 4 is complete.

A.6 Proof of Theorem 5

First, we show the existence and consistency of $\hat{\boldsymbol{\theta}}_n$. It follows from the uniform law of large numbers (Pollard 1990) that $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}) = -\partial \mathbf{U}_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ converges to some nonrandom matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ in probability uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The uniform convergence of $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})$, the continuity of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, and the nonsingularity of $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ imply that there exists a small neighborhood of $\boldsymbol{\theta}_0$ inside which the eigenvalues of $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})$ are bounded away from 0 for all large n and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Note that by the uniform law of large numbers, $U_n(\boldsymbol{\theta}_0) \to 0$ in probability. Then, by the inverse function theorem, we have that inside a small neighborhood of $\boldsymbol{\theta}_0$ there exists a unique solution, $\hat{\boldsymbol{\theta}}_n$, to $\mathbf{U}_n(\boldsymbol{\theta}) = 0$ for all sufficiently large n.

By Taylor expansion of $\mathbf{U}_n(\hat{\boldsymbol{\theta}}_n)$ at $\boldsymbol{\theta}_0$, $\mathbf{0} = \mathbf{U}_n(\boldsymbol{\theta}_0) - \hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$, where $\boldsymbol{\theta}^*$ is between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$. From the uniform convergence of $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})$ and consistency of $\hat{\boldsymbol{\theta}}_n$, it follows that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \{\boldsymbol{\Sigma}_0^{-1} + o_p(1)\}\sqrt{n}\mathbf{U}_n(\boldsymbol{\theta}_0). \tag{A.20}$$

Let $\mathbf{W}_i^* = \mathbf{Z}_i - \mathbf{g}^*(X_i), i = 1, \ldots, n$, where $\mathbf{g}^*(x)$ is as defined in Theorem 5. Then, by empirical approximation, $\sqrt{n}\mathbf{U}_n(\boldsymbol{\theta}_0)$ can be rewritten as $\sqrt{n}\mathbf{U}_n(\boldsymbol{\theta}_0) = \mathbf{I}_n + o_p(1)$, where

$$\mathbf{I}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau \left[\mathbf{W}_i^* - \frac{\sum_{j=1}^n \mathbf{W}_j^* \exp\{\boldsymbol{\theta}_0^\top \mathbf{Z}_j + \psi(X_j)\}Y_j(t)}{\sum_{j=1}^n \exp\{\boldsymbol{\theta}_0^\top \mathbf{Z}_j + \psi(X_j)\}Y_j(t)}\right] d\tilde{M}_i(t).$$

Therefore, it follows from the martingale central limit theorem that

$$\sqrt{n}\mathbf{U}_n(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} \mathrm{N}(0, \boldsymbol{\Sigma}_0),$$

where $\boldsymbol{\Sigma}_0$ is as defined in (C8′). Using the asymptotic distribution of $\sqrt{n}\mathbf{U}_n(\boldsymbol{\theta}_0)$, together with (A.20), we get that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} \mathrm{N}(0, \boldsymbol{\Sigma}_0^{-1}). \tag{A.21}$$

Following Bickel et al. (1993), we obtain that the information bound for $\boldsymbol{\theta}_0$ is

$$\begin{aligned}
\operatorname{var}(\xi_{\mathbf{Z}} - \xi_{\mathbf{g}^*}) &= E\big[\delta\{\mathbf{Z} - \mathbf{g}^*(X) - E(\mathbf{Z} - \mathbf{g}^*(X)|T = t, \delta = 1)\}\big]^{\otimes 2} \\
&= \boldsymbol{\Sigma}_0.
\end{aligned}$$

Therefore, it follows that $\hat{\boldsymbol{\theta}}_n$ is semiparametric efficient. The proof of Theorem 5 is complete.

SUPPLEMENTAL MATERIALS

Supplement: The supplemental materials address three issues: (1) A detailed discussion is given in Section 6. (2) All detailed proofs are provided in the Appendix. (3) A proof of the convergence of the algorithm is provided for a special case in Appendix B. (supplementalTechnicalReport.pdf)

[Received November 2008. Revised January 2010.]

REFERENCES

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993), *Efficient and Adaptive Estimation for Semiparametric Models*, New York: Springer-Verlag. [759,760]

Cai, Z. (2003), "Weighted Local Linear Approach to Censored Nonparametric Regression," in *Recent Advances and Trends in Nonparametric Statistics*, eds. M. G. Akritas and D. M. Politis, Amsterdam: Elsevier, pp. 217–231. [756]

Chen, S., and Zhou, L. (2007), "Local Partial Likelihood Estimation in Proportional Hazards Regression," *The Annals of Statistics*, 35, 888–916. [750-754,757]

Cox, D. R. (1972), "Regression Models and Life Tables" (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 4, 187–220. [750,751,757]

Crowley, J., and Hu, M. (1977), "Covariance Analysis of Heart Transplant Survival Data," *Journal of the American Statistical Association*, 72, 27–36. [756]

Fan, J., and Gijbels, I. (1994), "Censored Regression: Local Linear Approximations and Their Applications," *Journal of the American Statistical Association*, 89, 560–570. [756]

——— (1996), *Local Polynomial Modeling and Its Applications*, London: Chapman & Hall. [750]

Fan, J., Gijbels, I., and King, M. (1997), "Local Likelihood and Local Partial Likelihood in Hazard Regression," *The Annals of Statistics*, 25, 1661–1690. [750-753,757]

Fleming, T. R., and Harrington, D. P. (1991), *Counting Processes and Survival Analysis*, New York: Wiley. [759]

Gentleman, R., and Crowley, J. (1991), "Local Full Likelihood Estimation for the Proportional Hazards Model," *Biometrics*, 47, 1283–1296. [750]

Gray, R. (1992), "Flexible Methods for Analyzing Survival Data Using Splines, With Applications to Breast Cancer Prognosis," *Journal of the American Statistical Association*, 87, 942–951. [750,751]

Hastie, T., and Tibshirani, R. (1990), "Exploring the Nature of Covariate Effects in the Proportional Hazards Model," *Biometrics*, 46, 1005–1016. [750]

Huang, J. (1999), "Efficient Estimation of the Partly Linear Additive Cox Model," *The Annals of Statistics*, 27, 1536–1563. [751]

Huang, J., and Liu, L. X. (2006), "Polynomial Spline Estimation and Inference of Proportional Hazards Regression Models With Flexible Relative Risk Form," *Biometrics*, 62, 793–802. [750]

Klaassen, C. A., Lee, E., and Ruymgaart, F. H. (2005), "Asymptotically Efficient Estimation of Linear Functional in Inverse Regression Models," *Journal of Nonparametric Statistics*, 17, 819–831. [753]

Kooperberg, C., Stone, C. J., and Truong, Y. (1995), "Hazard Regression," *Journal of the American Statistical Association*, 90, 78–94. [750]

LeBlanc, M., and Crowley, J. (1999), "Adaptive Regression Splines in the Cox Model," *Biometrics*, 55, 204–213. [750,751,756]

Miller, R., and Halpern, J. (1982), "Regression With Censored Data," *Biometrika*, 69, 521–531. [756]

O'Sullivan, F. (1988), "Nonparametric Estimation of Relative Risk Using Splines and Cross-Validation," *SIAM Journal on Scientific and Statistical Computing*, 9, 531–542. [750,751,758]

——— (1993), "Nonparametric Estimation in the Cox Model," *The Annals of Statistics*, 21, 124–145. [750]

Pollard, D. (1984), *Convergence of Stochastic Processes*, New York: Springer. [758]

——— (1990), *Empirical Processes: Theory and Applications*, Hayward, CA: Institute of Mathematical Statistics. [758,760]

Sleeper, L. A., and Harrington, D. P. (1990), "Regression Splines in the Cox Model With Applications to Covariate Effects in Liver Disease," *Journal of the American Statistical Association*, 85, 941–949. [750,751]

Tibshirani, R., and Hastie, T. (1987), "Local Likelihood Estimation," *Journal of the American Statistical Association*, 82, 559–567. [750-752,757]

Wang, W. (2004), "Proportional Hazards Regression Models With Unknown Link Function and Time-Dependent Covariates," *Statistica Sinica*, 14, 885–905. [750]