Global Partial Likelihood for Nonparametric Proportional Hazards Models

(Supplemental Technical Report)

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

A global partial likelihood method, in contrast to the local partial likelihood method of Tibshirani and Hastie (1987) and Fan, Gijbels and King (1997), is proposed to estimate the covariate effect in a nonparametric proportional hazards model, $\lambda(t|x) = \exp\{\psi(x)\}\lambda_0(t)$. The estimator $\hat{\psi}(\cdot)$ 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 $\psi(\cdot)$. Moreover, the Breslow-type estimation of the cumulative baseline hazard function, using the proposed estimator $\hat{\psi}(\cdot)$, is proved to be efficient. Under regularity conditions, the asymptotic bias and variance are derived. The computation of the estimator involves an iterative but simple algorithm. Extensive simulation studies show supportive evidence of the theory. The method is illustrated with the Stanford heart transplant data and the PBC data.

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

[1]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 at Davis, CA, USA (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 L. Sun was supported in part by the National Natural Science Foundation of China (10571169 and 10731010) and the National Basic Research Program of China (2007CB814902). The research of J.-L. Wang was supported in part by the U.S. National Science Foundation (DMS04-06430).

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, i.e., 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. However, the assumption of log-linearity may not hold in practice.

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 X 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 $\lambda_0(\cdot)$ is an unspecified baseline hazard function and $\psi(\cdot)$ is an unknown smooth function. Several statistical methods involving smoothing techniques, such as nearest neighbor, spline and local polynomial smoothing methods have been developed for this model; see Tibshirani and Hastie (1987), O'Sullivan (1988), Hastie and Tibshirani (1990), Gentleman and Crowley (1991), Kooperberg, Stone and Truong (1995), Fan, Gijbels and King (1997), Chen and Zhou (2007), among others. 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 this approach is quite insightful and nontrivial. The local partial likelihood is used to estimate the derivative of the link function $\psi(\cdot)$, and an estimate of $\psi(\cdot)$ is obtained by integrating the estimated derivative. The reason for estimating the derivative first is that the link function can only be identified 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. Applications of local polynomial smoothing are developed in Fan, Gijbels and King (1997), 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 paper. Recently, Chen and Zhou (2007) developed an elegant approach by considering local partial likelihood at two points, x and x_0, whence 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 overall similar to previous 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 in order to estimate ψ at x, all observations are used, in contrast to previous approaches, which only use observations with covariate 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 Breslow-type estimator of the cumulative baseline hazard function, using the proposed estimator $\hat{\psi}$, is semiparametric efficient. The efficiency gain seems unusual at first as local smoothing yields optimal procedures for nonparametric regression functions. The situation is however 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 that 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 to be locally constrained.

The paper is organized as follows. In section 2, we introduce local and global partial likelihood and present an iterative algorithm to compute 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. With this new approach, the targeted link function ψ and its derivatives are estimated directly and simultaneously, similar to a standard local linear smoothing approach for nonparametric regression. The computation of the estimates is rather simple, with a built-in practical and feasible iterative algorithm. A proof of the convergence of the algorithm is provided for a special case in Appendix B. Asymptotic properties and semiparametric efficiency of the estimates are presented in Section 3. The optimal bandwidth is found to be of the order $n^{-1/5}$, which is the same as for standard local linear smoothing in nonparametric regression. Section 4 contains 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 and of the PBC data with the global procedure is described in Section 5. 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, however the distribution of C may depend on X. Let $\tilde{T} = \min(T, C)$ be the event time and $\delta = I(T \leq C)$ be the failure/censoring indicator. Let $(T_i, C_i, \tilde{T}_i, \delta_i, X_i)$, for $i = 1, ..., n$, be i.i.d copies of $(T, C, \tilde{T}, \delta, X)$. The observations are $(\tilde{T}_i, \delta_i, X_i)$, $i = 1, ..., n$. Define $N_i(t) = \delta_i I(\tilde{T}_i \leq t)$ and $Y_i(t) = I(\tilde{T}_i \geq t)$. The (global) partial likelihood, due to Cox (1972), is

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

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

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

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, since α cancels out in the ratio in (3). This is in accordance with the identifiability of ψ up to a shift. As a result, $\psi(\cdot)$ is not directly estimable from the local partial likelihood, but the derivative $\psi'(\cdot)$ 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 $\psi(\cdot)$, but no theory is available for this estimate.

A much refined version of local partial likelihood estimation through local polynomial smoothing was proposed and studied in Fan, Gijbels and King (1997). Let h be a bandwidth, $K(\cdot)$ 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 in Fan, Gijbels and King (1997) is given by

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

where $\beta_k = \psi^{(k)}(x)/k!$ and $\psi^{(k)}(x)$ is the k-th 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 $\psi'(\cdot)$ from a to b, and the final estimate of $\psi(x)$ is the same as the estimate for $\psi(x) - \psi(0)$. A comprehensive theoretical justification of this method and closed form expressions for asymptotic bias and variance of the derivative estimates were also provided by Fan, Gijbels and King (1997). Chen and Zhou (2007) took an alternative approach, aiming directly at the differences $\psi(x) - \psi(x_0)$, by employing a local partial likelihood in the neighborhood of both x and x_0. A two-step algorithm was proposed in Chen and Zhou (2007), 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 were derived including asymptotic bias and variance. As this procedure still utilizes a local partial likelihood, improvement is possible by employing a global partial likelihood method instead, as we demonstrate in the following.

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

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

This is a crude nearest neighbor (global) partial likelihood, analogous to (3). Since the true ψ is unknown, (5) is not directly useable, but it motivates an iterative approach. With a slight variant of local linear smoothing for computational convenience, we propose the following iterative algorithm.

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

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

Iteration Step from $m-1$ to m. For every given $x = X_1, ..., X_{n-1}$, solve the following equations for α and β:

$$\sum_{i=1}^{n}\int_0^\infty\Big[K_i(x)-\frac{\sum_{j=1}^n K_j(x)\exp\{\alpha+\beta(X_j-x)\}Y_j(t)}{\sum_{j=1}^n \exp\{\psi_{(m-1)}(X_j)\}Y_j(t)}\Big]dN_i(t)=0, \tag{6}$$

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

Let $\hat{\alpha}(x)$ and $\hat{\beta}(x)$ be the solutions of α and β. Then $\psi_{(m)}(X_i)=\hat{\alpha}(X_i)$ for $i=1,...,n-1$. This iteration is continued until convergence.

Finally, for every x, the 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 equations (6) and (7) at convergence.

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

Remark 2. The proposed estimate reduces to the partial likelihood estimate when the covariate assumes only finitely many (distinct) values, say $a_1,...,a_K$. To see this, assume $\psi(a_K)=0$ 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

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

for $k=1,...,K-1$, and $\alpha_K=0$, where $\alpha_k=\psi(a_k)$. This is exactly the same as the Cox partial likelihood estimating equation, and therefore 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.

Remark 3. The convergence of the algorithm is shown in Appendix B for the case of local constant fitting, i.e. when $p=0$ in (4) or equivalently when $\beta=0$ in the estimating equation (6). Although we do not have a general proof for arbitrary p, the algorithm always converged when we used a local linear polynomial, which corresponds to $p = 1$ in (4), in the simulation studies reported in Section 4.

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. In particular, for $\tau = \inf \{t : P(\tilde{T} > t) = 0\}$, we assume that $P(T > \tau) > 0$ and $P(C = \tau) > 0$. These conditions are needed for technical convenience; it is possible that the results remain true without such restrictions, as suggested by the simulation results presented in section 4. These conditions are imposed in, for example, Chen, Jin and Ying (2002) and frequently appear in the literature. They are satisfied in medical follow-up studies when a study ends at a fixed time τ before all failures are observed.

For identifiability and without loss of generality, we fix $\hat{\psi}(0) = \psi(0) = 0$. We start with the uniform consistency of $\hat{\psi}(\cdot)$.

Theorem 1. *Suppose 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* $n \to \infty$.

Some notations are introduced to show the asymptotic expression of $\hat{\psi}(x) - \psi(x)$ at each fixed point $x \in (0, 1)$. Let $f(\cdot)$ denote the density of the random variable X and $\psi''(\cdot)$ be the second derivative of $\psi(\cdot)$. Set $\mu = \int_{-1}^{1} u^2 K(u)du$, $\nu = \int_{-1}^{1} K^2(u)du$, $P(t|x) = P(\tilde{T} \geq t|X = x)$, $\Gamma(x) = \int_0^\tau P(t|x)\lambda_0(t)dt$, $s_0(t) = E[P(t|X) \exp\{\psi(X)\}]$ and

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

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

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

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

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

Remark 4. 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 $\mathcal{I}$ be the identity operator. Then (9) can be written as

$$(\mathcal{I}-\mathcal{A})(\hat{\psi}-\psi)(x)=\frac{1}{2}h^2\mu\psi''(x)+(nh)^{-1/2}\xi_n(x)+o_p(h^2+(nh)^{-1/2}), \tag{10}$$

which means that the order of the asymptotic bias of $(\mathcal{I}-\mathcal{A})(\hat{\psi}-\psi)(x)$ 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 $(\mathcal{I}-\mathcal{A})(\hat{\psi}-\psi)(x)$ is of the order of $n^{-1/5}$.

Next, we show the asymptotic expression of $\hat{\psi}(x)-\psi(x)$ for a boundary point x. Such an investigation is necessary since it is not obvious that the estimator $\hat{\psi}$ has the same asymptotic behavior at the boundary as in the interior of the support. In the setting of nonparametric regression, Fan (1992) showed that a boundary modification is not required for a local linear smoother. Proposition 1 below confirms that the same holds true for our global partial likelihood estimate when a local linear smoother is employed.

Proposition 1. *Under the regularity conditions (C1)-(C7) stated in the Appendix and for $x_n = ch$ with $c \in [0,1)$, $\hat{\psi}(x_n)-\psi(x_n)$ satisfies the following Fredholm integral* *equation,*

$$\hat{\psi}(x_n) - \psi(x_n) = \int_0^1 \Phi(u|0)\{\hat{\psi}(u) - \psi(u)\}du + \frac{1}{2}h^2\tilde{\mu}_c\psi''(0) + (nh)^{-1/2}\xi_n$$
$$+o_p(h^2 + (nh)^{-1/2}), \qquad (11)$$

where ξ_n *converges to the normal distribution with mean zero and variance*

$$\sigma_c^2 = \tilde{\nu}_c[f(0)\Gamma(0)]^{-1},$$

and $\tilde{\mu}_c$ *and* $\tilde{\nu}_c$ *are defined in (A.52) in the Appendix.*

Similar results as (11) can be established for a boundary point $x_n = 1-ch,\ c \in [0,1)$, at the other end of the interval. Details are omitted.

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

Theorem 3. *Suppose the regularity conditions (C1)-(C6) stated in the Appendix are satisfied. Assume that* $nh^4 \to 0$ *and* $nh^2/(\log n)^2 \to \infty$ *as* $n \to \infty$. *Then, for any function* $\phi(\cdot) \in \mathcal{D}_0$, $\int_0^1 \hat{\psi}(x)\phi(x)dx$ *is a semiparametric efficient estimator of* $\int_0^1 \psi(x)\phi(x)dx$.

Remark 5. Note that $\int_0^1 \phi(x)dx = 0$ is an identifiability condition for estimating $\int_0^1 \psi(x)\phi(x)dx$, since $\psi(\cdot)$ 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 $\psi(\cdot)$ can be estimated with semiparametric efficiency by the corresponding functional of the estimate $\hat{\psi}(\cdot)$. See Klaassen, Lee and Ruymgaart (2005) for a related type of efficiency based on functionals.

Let $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ be the cumulative baseline hazard function. Using the global likelihood estimate $\hat{\psi}(\cdot)$, 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 $\psi(\cdot)$ should produce efficient estimation for $\Lambda_0(t)$, while using others like Fan, Gijbles and King (1997) or Chen and Zhou (2007) will not.

Theorem 4. *Suppose the regularity conditions (C1)-(C6) stated in the Appendix are satisfied. Assume that $nh^4 \to 0$ and $nh^2/(\log n)^2 \to \infty$ as $n \to \infty$. Then, for any bounded measurable function $b(\cdot)$, $\int_0^\tau b(t)d\hat{\Lambda}_0(t)$ 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

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

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 $a(x) = \exp[c_1 I(\psi(x) > b)/3 + c_2 I(\psi(x) \leq b)/3]$, and the constants λ, c_1, c_2 and b are chosen such that the total censoring rate is about 30%-40%. The Epanechnikov kernel $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$ was used in all simulations. The results presented below are based on 500 replications and sample size $n = 200$. We consider six settings; where in the following, ρ is a Bernoulli random variable with $P(\rho = 1) = P(\rho = 0) = 0.5$.

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

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

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

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

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

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

We choose $c_1 = 11,\ c_2 = 5, b = 0$ for all 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 conventional Cox model with identity ψ-function, and Models 3 and 4 correspond to situations with faster increasing risks associated with the covariate values. We do not include decreasing link functions, as the sign of X can be flipped in Models 1-4 to make the link function decreasing. While 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, these are included to compare all three methods under a non-monotone link function.

Next, we describe how to compute the estimators based on various methods with the same constraint $\hat{\psi}(0) = 0$. For GPL, we first calculate the curve estimates $\hat{\psi}(\cdot)$ based on (6) and (7) with constraints $\sum_{i=1}^{n} \hat{\psi}(X_i) = 0$, and then replace $\hat{\psi}(\cdot)$ by $\hat{\psi}(\cdot) - \hat{\psi}(0)$. For FGK, we obtain the estimates through $\hat{\psi}(\cdot) = \int_0^{\cdot} \hat{\psi}'(t)dt$. For CZ, as suggested in Chen and Zhou (2007), we employed a two-step procedure to first calculate estimates for $\psi'(\cdot)$ as in FGK using a bandwidth h_0. We then calculated the estimates of $\psi(\cdot) - \psi(x_0)$ at the second step using another bandwidth h, where x_0 is a prespecified point, and obtained the final estimates by $\hat{\psi}(\cdot) - \hat{\psi}(0)$. 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 Table 1-3, as these choices provide good estimates for $\hat{\psi}'$. In the second step, we adopted a different bandwidth h, as specified in the first columns of Table 1-3, to obtain the final estimate of $\psi(\cdot) - \psi(x_0)$. A different suggestion to use $h_0 = 1.25h$ was proposed in Chen and Zhou (2007), and we include their suggestion for comparison purposes.

The performance of the various estimators $\hat{\psi}(\cdot)$ is 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. To avoid boundary effects, we choose $a = 1.0$ in Models 1-4 and $a = 2.0$ in Models 5-6.

Tables 1-3 summarize the results for the WMISE of Models 1-6 under various bandwidth choices, where WMISE is 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 global partial likelihood approach in comparison with local approaches. We also observe 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, we should choose a big bandwidth to estimate the link function and GPL and FGK both attained their minimum WMISEs at the largest bandwidth $h = 1.0$. However, the optimal bandwidth h for CZ depends on the choice of the prespecified bandwidth h_0. We also see from Table 3 that the optimal bandwidth for GPL is smaller than those for FGK and CZ. This is likely due to the fact that the local methods need to enlarge the included range of data to compensate for the fact of using less data information than the global procedure. For Models 5-6, we also tried $x_0 = -1$ and 1 for CZ, but the results were not as good as those in Table 3 and are not included. The performance of CZ depends heavily on the bandwidth choice; the values given in parenthesis 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 previous suggested choice of $h_0 = 1.25h$ (reported in parenthesis).

Figures 1-6 show the biases of the different estimates based on the optimal bandwidths of Tables 1-6. It can be seen from these figures that all methods are asymptotically unbiased and comparable when optimal bandwidths were used.

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.

5. APPLICATIONS

We now illustrate the proposed method with two historical data sets.

Example 1: *Stanford Heart Transplant data.*

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 be found in Crowley and Hu (1977) and 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 nonparametric link function on age, such as the one proposed in this paper, 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 $\psi(\cdot)$ with the GPL method and bandwidths $h = 7$ and 10 are presented in Figure 7. These estimates suggest that the risk decreases with age at transplantation for younger patients (age less than 20), remain constant (nearly zero) in the middle age range (20-40), and then increases with age for older patients (older than 40). While 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 7, based on such a piecewise linear fit for the ψ-function, supports this speculation and that the risk associated with the age at transplantation is relevant only for younger and older patients.

Example 2: *Primary Biliary Cirrhosis (PBC) data.*

PBC is a rare, chronic and fatal liver disease. This data was collected by the Mayo Clinic between January 1974 and May 1984. There were 312 patients in the randomized trial, of which 125 died with only 11 deaths not attributable to PBC. The data was analyzed in Fleming and Harrington (1991), Kooperberg, Stone and Truong, (1995) and Fan and Gijbels (1994), among others.

There were many covariates in the original PBC data. In order to check which covariate may have a nonlinear effect on log hazards, we select just one covariate at a time. This is suitable at initial data analysis for exploratory purpose. We select the most significant covariate, Bilirubin, which is a waste product that accumulates when liver function is poor and is responsible for jaundice when its level is high. Following common practice, we take the logarithm of bilirubin as the covariate X. The estimate of $\psi(\cdot)$ is presented in Figure 8, choosing bandwidth $h = 0.3$. It can be seen from the result that bilirubin has a nonlinear effect on survival time. Specifically, the effect of log(bilirubin) decreases linearly when it is less than -0.5, and then increases thereafter.

Our results are in agreement with those presented in Fan and Gijbels (1994) and Cai (2003). This is not surprising, as all three procedures assume nonparametric link functions. Since their procedures are local, we expect our procedure to be more efficient as demonstrated in both theory and simulations.

6. CONCLUDING REMARKS

We propose a global partial likelihood method to estimate the covariate effect in the 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 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 show that the proposed methods work well for the situations considered. For the Stanford Heart Transplant data, our approach leads to interesting findings that shed new light on these data.

So far, we have focused on estimating the link function $\psi(\cdot)$. However, if it is of interest to estimate the derivative of $\psi(\cdot)$, we recommend to use a local polynomial of higher order rather than local linear fitting. The method developed here for local linear fitting is applicable to general local polynomial fitting, although some modifications are needed (e.g., Fan and Gijbels, 1996).

In most applications, there will be more than one covariate. Although the procedure and theory developed in this paper can be readily extended to the case of a nonparametric multivariate ψ function, such a completely nonparametric approach is subject to the curse of dimensionality. A more practical way to accommodate high dimensional covariates is to include dimension reduction features in the model; partial linear or additive models are common dimension reduction approaches. Both can be embedded into a larger model called "partial linear additive models". For the proportional survival model this corresponds to:

$$\lambda(t|Z = z, X = x) = \exp\{\beta' z + \psi_1(x_1) + \cdots + \psi_k(x_k)\}\lambda_0(t),$$

where Z is a p-dimensional covariate with unknown regression parameters β, $x = (x_1, \cdots, x_k)$ is a k-dimensional covariate with nonparametric link functions $\psi_i(x_i)$ for each x_i, and $\lambda_0(\cdot)$ is an unspecified baseline hazard function. A backfitting algorithm can be added to our procedure to iteratively estimate each $\psi_i(\cdot)$, and the regression parameter can be estimated through a profile approach similar to the approach of partial linear models studied in Speckman (1988). Thus, it is fairly straightforward to extend our algorithm to accommodate multiple covariates under the partial linear additive proportional hazards model. The challenge is on the theoretical front and beyond the scope of this paper. This is certainly a worthy direction to pursue. We conjecture a similar efficiency result for the nonparametric estimators of $\psi_i(\cdot)$ and semiparametric efficiency for the estimator of β for this case. Related semiparametric efficiency results for the estimation of β were obtained in Huang (1999), who studied the maximum partial likelihood estimator using polynomial splines for the nonparametric components but did not pursue the efficiency of the nonparametric components. The proposed method is also potentially useful in order to deal with other semiparametric and nonparametric models, such as the transformation model with varying covariate effects.

One important issue that we have not conclusively addressed in this paper is the choice of the bandwidth for the nonparametric estimate of $\psi(\cdot)$. 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 are suggested in Tibshrani and Hastie (1987), and another possibility is to minimize 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 may not work well in practice due to the need to truncate certain boundary regions. One promising direction is to develop a cross-validation approach based on the global partial likelihood, but this would be a separate project for future investigation. For the moment, we rely on a subjective method to visually inspect the right amount of smoothness as illustrated in the two data examples.

APPENDIX: PROOFS

A.1. Regularity conditions and Lemmas

We first state the following regularity conditions.

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

(C2) 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 $f(\cdot)$ of X is bounded away from zero and has a continuous second-order derivative on $[0, 1]$.

(C5) The function $\psi(\cdot)$ 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 $t \in [0, \tau]$.

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

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

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

Define $\mu_k(x) = \int_{u \in \mathcal{Z}(x,h)} u^k K(u) du, \quad k = 0, 1, 2, 3,$ where $\mathcal{Z}(x, h) = \{u : x - hu \in [0, 1]\} \cap [-1, 1]$. If $\mathcal{Z}(x, h) = [-1, 1]$, we write $\mu_k = \mu_k(x)$. It is seen that $\mu_0 = 1,\ \mu_1 = \mu_3 = 0$. For any functions $\eta(\cdot)$ and $\gamma(\cdot)$ defined on [0, 1], set

$$\begin{aligned}
S_{n0}(t; \eta) &= \frac{1}{n}\sum_{i=1}^{n} Y_i(t) \exp\{\eta(X_i)\} \qquad \text{and} \\
S_{n1}(t; \eta, \gamma, x) &= \frac{1}{n}\sum_{i=1}^{n} \omega_i(x) Y_i(t) \exp\{\eta(x) + \gamma(x)(X_i - x)/h\}.
\end{aligned}$$

Then, $(\hat{\psi}, h\hat{\psi}')$ is the solution to the equation $U(\eta, \gamma; x) = 0$, where

$$U(\eta, \gamma; x) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \Big[\omega_i(x) - \frac{S_{n1}(t;\eta,\gamma,x)}{S_{n0}(t;\eta)}\Big]dN_i(t). \tag{A.1}$$

Let

$$\begin{aligned}
&s_0(t;\eta) = E[P(t|X)\exp\{\eta(X)\}],\\
&s_1(t;\eta,\gamma,x) = P(t|x)\exp\{\eta(x)\}(\xi_1(x,\gamma),\xi_2(x,\gamma))^\top/[\mu_2(x)\mu_0(x) - \mu_1^2(x)],\\
&\xi_1(x,\alpha) = \mu_2(x)\int_{\mathcal{Z}(x,h)}\exp(\alpha u)K(u)du - \mu_1(x)\int_{\mathcal{Z}(x,h)}\exp(\alpha u)uK(u)du \quad \text{and}\\
&\xi_2(x,\alpha) = \mu_0(x)\int_{\mathcal{Z}(x,h)}\exp(\alpha u)uK(u)du - \mu_1(x)\int_{\mathcal{Z}(x,h)}\exp(\alpha u)K(u)du. \qquad (A.2)
\end{aligned}$$

Define

$$\begin{aligned}
&\boldsymbol{u}(\eta,\gamma;x) = \Gamma(x)\exp\{\psi(x)\}(1,0)^\top - \int_0^\tau \frac{s_1(t;\eta,\gamma,x)}{s_0(t;\eta)}s_0(t;\psi)\lambda_0(t)dt, \qquad (A.3)\\
&\mathcal{B}_n = \Big\{\eta : \|\eta\| \le D,\ |\eta(x) - \eta(y)| \le d[|x-y| + b_n],\ x,\ y \in [0,1]\Big\} \quad \text{and}\\
&\mathcal{C}_0 = \Big\{\eta(x) : x \in [0,1],\ \eta(0) = 0,\ \eta(x) \text{ is continuous on } [0{,}1]\Big\} \qquad (A.4)
\end{aligned}$$

where $b_n = h + (nh)^{-1/2}(\log n)^{1/2}$, $D > 0$, $d > 0$. Denote $U_1(\eta,\gamma;x)$ and $\boldsymbol{u}_1(\eta,\gamma;x)$ the first entries of $U(\eta,\gamma;x)$ and $\boldsymbol{u}(\eta,\gamma;x)$, respectively, and $U_2(\eta,\gamma;x)$ and $\boldsymbol{u}_2(\eta,\gamma;x)$ are the corresponding second entries.

For the proofs of Theorems 1 and 2, we need the following lemmas.

Lemma 1. Suppose condition (C1) holds. Then

(i). $\xi_2(x,\alpha) = 0$ if and only if $\alpha = 0$ for $x \in [0,1]$.

(ii). There exist positive constants σ_1 and σ_2 which depend only on $K(\cdot)$ such that for each $x \in [0,1]$ and $\sigma_0 > 0$,

$$\inf_{|\alpha|\ge\sigma_0}\Big|\frac{\xi_2(x,\alpha)}{\xi_1(x,\alpha)}\Big| \ge \sigma_1\sigma_2[1 - \exp(-2\sigma_0\sigma_1)]. \tag{A.5}$$

Proof. (i). Observe that $\xi_2(x,0) \equiv 0$ for $x \in [0,1]$. Furthermore, we have

$$\xi_2(x,\alpha) = \frac{1}{2}\int\int_{u,v\in\mathcal{Z}(x,h)}[u-v][\exp(\alpha u) - \exp(\alpha v)]K(u)K(v)dudv.$$

Then, $\xi_2(x,\alpha)\alpha > 0$ for any $\alpha \neq 0$ and $x \in [0,1]$, which means that $\xi_2(x,\alpha) = 0$ if and only if $\alpha = 0$ for any $x \in [0,1]$.

(ii). Consider $x \in [h, 1-h]$. It is seen that $\mathcal{Z}(x,h) = [-1,1]$, implying

$$\xi_1(x,\alpha) = \mu_2 \int_{-1}^{1} \exp(\alpha u) K(u) du \quad \text{and} \quad \xi_2(x,\alpha) = \mu_0 \int_{-1}^{1} \exp(\alpha u) u K(u) du.$$

Choose $0 < \sigma_1 < 1$ such that $\int_0^{\sigma_1} K(u)du \leq \int_{\sigma_1}^1 K(u)du$ and $\sigma_2 = \mu_0/(4\mu_2)$. Observe that $\int_0^{\sigma_1} \exp(\alpha u)K(u)du \leq \int_{\sigma_1}^1 \exp(\alpha u)K(u)du$ and that $\exp(\alpha u) - \exp(-\alpha u) \geq \exp(\alpha u)[1 - \exp(-2\sigma_0\sigma_1)]$ for $u \in [\sigma_1, 1]$ and $\alpha \geq \sigma_0 > 0$. Then, for $\alpha \geq \sigma_0 > 0$,

$$\begin{aligned}
\frac{\xi_2(x,\alpha)}{\xi_1(x,\alpha)} &= \frac{\mu_0}{\mu_2} \frac{\int_{-1}^1 \exp(\alpha u) u K(u) du}{\int_{-1}^1 \exp(\alpha u) K(u) du} \\
&\geq \frac{\mu_0}{2\mu_2} \frac{\int_0^1 [\exp(\alpha u) - \exp(-\alpha u)] u K(u) du}{\int_0^1 \exp(\alpha u) K(u) du} \\
&\geq \frac{\mu_0}{4\mu_2} \frac{\int_{\sigma_1}^1 [\exp(\alpha u) - \exp(-\alpha u)] u K(u) du}{\int_{\sigma_1}^1 \exp(\alpha u) K(u) du} \\
&\geq \frac{\mu_0 \sigma_1}{4\mu_2} \frac{\int_{\sigma_1}^1 [\exp(\alpha u) - \exp(-\alpha u)] K(u) du}{\int_{\sigma_1}^1 \exp(\alpha u) K(u) du} \\
&\geq \sigma_1\sigma_2[1 - \exp(-2\sigma_0\sigma_1)].
\end{aligned}$$

In a similar fashion, $-\xi_2(x,\alpha)/\xi_1(x,\alpha) \geq \sigma_1\sigma_2[1 - \exp(-2\sigma_0\sigma_1)]$, for $\alpha \leq -\sigma_0 < 0$ and $x \in [h, 1-h]$. Therefore, (A.5) is established for $x \in [h, 1-h]$. Similar proofs apply for $x \in [0,1]\backslash[h, 1-h]$. The details are omitted.

Lemma 2. Suppose conditions (C2)-(C6) hold. Let

$$\mathcal{K}(\eta)(x) = \psi(x) - \log\Big\{\int_0^\tau \frac{s_0(t;\psi)}{s_0(t;\eta)} P(t|x)\lambda_0(t)dt\Big\} + \log\{\Gamma(x)\}, \ x \in [0,1].$$

Then, $\mathcal{K}(\eta)(\cdot) = \eta(\cdot)$ has exactly one solution $\eta(\cdot) = \psi(\cdot)$ in $\mathcal{C}_0$.

Proof. Recall that $\Gamma(x) = \int_0^\tau P(t|x)\lambda_0(t)dt$. It is easy to check that $\mathcal{K}(\psi)(\cdot) = \psi(\cdot)$. It suffices to prove that $\psi(\cdot)$ is the unique solution. To this end, let $\mathcal{K}'_\eta(h)(\cdot)$ be the Gateaux derivative of $\mathcal{K}(\eta)(\cdot)$ at the point $\eta(\cdot)$. A straightforward calculation gives

$$\mathcal{K}'_\eta(h)(x) = \int_0^1 h(u)\Big[\int_0^\tau a(u|t;\eta)b(t|x;\eta)dt\Big]du,$$

where $h(\cdot) \in \mathcal{C}_0$, $a(u|t;\eta) = f(u)\exp\{\eta(u)\}P(t|u)/s_0(t;\eta)$ and

$$b(t|x;\eta) = \frac{P(t|x)s_0(t;\psi)\lambda_0(t)}{s_0(t;\eta)}\Big[\int_0^\tau \frac{P(t|x)s_0(t;\psi)\lambda_0(t)}{s_0(t;\eta)}dt\Big]^{-1}.$$

Suppose there exist two functions $\eta(\cdot)$ and $\eta(\cdot) + h(\cdot)$ in $\mathcal{C}_0$ such that $\mathcal{K}(\eta)(\cdot) = \eta(\cdot)$ and $\mathcal{K}(\eta + h)(\cdot) = (\eta + h)(\cdot)$. Then,

$$\begin{aligned} h(x) &= \mathcal{K}(\eta + h)(x) - \mathcal{K}\eta(x) = \int_0^1 \mathcal{K}'_{\eta+\xi h}(h)(x)d\xi \\ &= \int_0^1 \int_0^1 h(u)\Big[\int_0^\tau a(u|t;\eta + \xi h)b(t|x;\eta + \xi h)dt\Big]dud\xi \\ &= \int_0^1 h(u)\kappa(u, x; \eta, h)du, \end{aligned} \tag{A.6}$$

where $\kappa(u, x; \eta, h) = \int_0^1 \Big[\int_0^\tau a(u|t;\eta + \xi h)b(t|x;\eta + \xi h)dt\Big]d\xi$. Note that $\kappa(u, x; \eta, h)$ is a transitional density from x to u and positive on $[0, 1]^2$. Let $x^* = \text{argmax}\{h(x) : x \in [0, 1]\}$. Then, $0 = \int_0^1 [h(x^*) - h(u)]\kappa(u, x^*; \eta, h)du \geq 0$, which implies that $h(\cdot)$ is constant on $[0, 1]$. The definition of $\mathcal{C}_0$ ensures that $h(\cdot) \equiv 0$. The proof is complete.

Lemma 3. Suppose conditions (C2)-(C6) hold. Then, for any fixed function $\phi(\cdot) \in \mathcal{D}_0$, there exists a unique function $g(\cdot) \in \mathcal{C}_0$ with a continuous second derivative such that

$$\Big[g(x)\Gamma(x) - \Lambda_g(x)\Big]f(x)\exp\{\psi(x)\} = \phi(x), \tag{A.7}$$

where $\mathcal{D}_0$ is given in Section 3, $\mathcal{C}_0$ is defined in (A.4), $\Lambda_g(x) = \int_0^\tau \bar{g}(t)P(t|x)\lambda_0(t)dt$, and $\bar{g}(t) = E[P(t|X)\exp\{\psi(X)\}g(X)]/s_0(t)$. Moreover, if $\mathcal{F}_0 \subset \mathcal{D}_0$ is uniformly bounded, then the corresponding $\tilde{\mathcal{F}}_0$ is also uniformly bounded, where $\tilde{\mathcal{F}}_0 = \{g(\cdot) : \ g(\cdot) \in \mathcal{C}_0,\ g(\cdot) \text{ satisfies (A.7) and } \phi(\cdot) \in \mathcal{F}_0\}$.

Proof. Rewrite (A.7) as

$$g(x) = \int_0^1 \Phi(u|x)g(u)du + [\Gamma(x)f(x)\exp\{\psi(x)\}]^{-1}\phi(x), \tag{A.8}$$

which is a Fredholm integral equation. Since $\Phi(u|x)$ is a transitional density from x to u and positive on $[0, 1]^2$, by the same arguments as in Lemma 2, there exists a unique solution $g(\cdot) \in \mathcal{C}_0$ to equation (A.7). Note that $f(\cdot)$, $\psi(\cdot)$ and $\phi(\cdot)$ have continuous second derivatives on $[0, 1]$. A straightforward calculation yields that $g(\cdot)$ has a continuous second derivative on [0,1].

It remains to show that $\tilde{\mathcal{F}}_0$ is uniformly bounded. It is checked from conditions (C2)-(C6) that $\delta_0 = \int_0^1 \inf_{x\in[0,1]} \Phi(u|x)du > 0$. Let $x^{**} = argmax\{g(x) : x \in [0,1]\}$, $x^* = argmin\{g(x) : x \in [0,1]\}$. Then,

$$\begin{aligned}
&[\Gamma(x^{**})f(x^{**})\exp\{\psi(x^{**})\}]^{-1}\phi(x^{**}) \\
&= \int_0^1 \Phi(u|x^{**})(g(x^{**}) - g(u))du \geq \int_0^1 \inf_{x\in[0,1]} \Phi(u|x)(g(x^{**}) - g(u))du, \qquad \text{(A.9)}
\end{aligned}$$

and

$$\begin{aligned}
&-[\Gamma(x^{*})f(x^{*})\exp\{\psi(x^{*})\}]^{-1}\phi(x^{*}) \\
&= \int_0^1 \Phi(u|x^{*})(g(u) - g(x^{*}))du \geq \int_0^1 \inf_{x\in[0,1]} \Phi(u|x)(g(u) - g(x^{*}))du. \qquad \text{(A.10)}
\end{aligned}$$

Since $\mathcal{F}_0$ is uniformly bounded, there exists a positive constant M such that

$$\sup_{x\in[0,1]} |[\Gamma(x)f(x)\exp\{\psi(x)\}]^{-1}\phi(x)| \leq M$$

uniformly over $\phi(\cdot) \in \mathcal{F}_0$. Therefore, it follows from (A.9) and (A.10) that

$$[g(x^{**}) - g(x^*)] \leq 2M/\delta_0.$$

Note that $g(0) = 0$. This shows that $\tilde{\mathcal{F}}_0$ is uniformly bounded.

Lemma 4. Suppose conditions (C1)-(C7) hold and $g(x, y, z)$ is any continuous function. Then,

$$\sup_{x\in[0,1],\ t\in[0,\tau]} \Big|c_n(x,t) - Ec_n(x,t)\Big| = O_p((nh)^{-1/2}(\log n)^{1/2}),$$

where $c_n(x,t) = n^{-1}\sum_{i=1}^n Y_i(t)g(X_i, (X_i - x)/h, x)K_h(X_i - x)$.

Proof. The proof follows directly from Theorem 2.37 and Example 38 in Chapter 2 of Pollard (1984). The details are omitted.

Lemma 5. Suppose conditions (C1)-(C7) hold. Then,

$$(i) \qquad \sup_{y\in[0,1]} \Big|\frac{1}{n}\sum_{i=1}^{n} g(X_i,y)U_1(\psi,h\psi';X_i) - \int_0^1 g(u,y)U_1(\psi,h\psi';u)f(u)du\Big| = o_p(n^{-1/2}),$$

$$\text{and} \quad (ii) \qquad \sup_{y\in[0,1]} \Big|\frac{1}{n}\sum_{i=1}^{n}[\int_0^\tau (Y_i(t)-P(t|X_i))P(t|y)(s_0(t))^{-1}\lambda_0(t)dt] [\Gamma(X_i)]^{-1}U_1(\psi,h\psi';X_i)\Big| = o_p(n^{-1/2}),$$

where $g(x,y) = (\Gamma(x))^{-1}\int_0^\tau P(t|x)P(t|y)(s_0(t))^{-1}\lambda_0(t)dt,\ x,y\in[0,1]$.

Proof. Denote $\tilde{S}_{n1}(t;x) = n^{-1}\sum_{i=1}^{n}\omega_{i1}(x)Y_i(t)\exp\{\psi(x)+\psi'(x)(X_i-x)\}$. Recall that $M_i(t) = N_i(t) - \int_0^t \exp\{\psi(X_i)\}Y_i(s)\lambda_0(s)ds$, and

$$U_1(\psi,h\psi';x) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau [\omega_{i1}(x) - \frac{\tilde{S}_{n1}(t;x)}{S_{n0}(t)}]dN_i(t).$$

To prove (i), it suffices to show that

$$(a) \qquad \sup_{y\in[0,1]}\left|\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \Big[\frac{1}{n}\sum_{j=1}^{n} g(X_j,y)\omega_{i1}(X_j) - \int_0^1 g(u,y)\omega_{i1}(u)f(u)du\Big]dM_i(t)\right| = o_p(n^{-1/2}),$$

$$(b) \qquad \sup_{y\in[0,1]}\left|\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau [S_{n0}(t)]^{-1}\Big[\frac{1}{n}\sum_{j=1}^{n} g(X_j,y)\tilde{S}_{n1}(t;X_j) - \int_0^1 g(u,y)\tilde{S}_{n1}(t;u)f(u)du\Big]dM_i(t)\right| = o_p(n^{-1/2}),$$

$$\text{and} \qquad (c) \qquad \sup_{y\in[0,1]}\left|\int_0^\tau \Big[\frac{1}{n}\sum_{i=1}^{n} g(X_i,y)h_n(X_i,t) - \int_0^1 g(u,y)h_n(u,t)f(u)du\Big]\lambda_0(t)dt\right| = o_p(n^{-1/2}),$$

where $h_n(x,t) = n^{-1}\sum_{i=1}^{n}\omega_{i1}(x)Y_i(t)[\exp\{\psi(X_i)\} - \exp\{\psi(x)+\psi'(x)(X_i-x)\}]$. Denote $\bar{h}(x,t) = E[h_n(x,t)]$. Then, Lemma 4 implies

$$\sup_{t\in[0,\tau],\ x\in[0,1]}\Big|h_n(x,t)-\bar{h}(x,t)\Big| = h^2\cdot O_p((nh)^{-1/2}\log^{1/2}(n)) = o_p(n^{-1/2}).$$

Thus, (c) is equivalent to show that

$$(d) \qquad \sup_{y\in[0,1]}\Big|\int_0^\tau \Big[\frac{1}{n}\sum_{i=1}^n g(X_i,y)\bar{h}(X_i,t) - \int_0^1 g(u,y)\bar{h}(u,t)f(u)du\Big]\lambda_0(t)dt\Big| = o_p(n^{-1/2}).$$

We give a proof of part (a). Parts (b) and (d) follow from similar arguments.

Denote $\mathcal{F}=\{g(\cdot,y), y\in[0,1]\}$. Note that there exists one positive constant r such that $\sup_{x\in[0,1]}\Big|g(x,y_1)-g(x,y_2)\Big|\le r|y_1-y_2|$ for any $y_1,y_2\in[0,1]$. Therefore,

$$N(\varepsilon,\mathcal{F},||\cdot||_\infty)=O(\frac{1}{\varepsilon}), \tag{A.11}$$

where $N(\varepsilon,\mathcal{F},||\cdot||_\infty)$ is the covering number with the norm $||\cdot||_\infty$ of the class $\mathcal{F}$, i.e., the minimal number of balls of $||\cdot||_\infty$-radius ε needed to cover $\mathcal{F}$. Write

$$\zeta(y)=\frac{1}{n}\sum_{i=1}^n\int_0^\tau\Big[\frac{1}{n}\sum_{j=1}^n g(X_j,y)\omega_{i1}(X_j)-\int_0^1 g(u,y)\omega_{i1}(u)f(u)du\Big]dM_i(t). \tag{A.12}$$

Define $||\zeta(y)||_2=\sqrt{E[\zeta(y)]^2}$. Notice that $d\langle M_i,M_i\rangle(t)=\exp\{\psi(X_i)\}Y_i(t)\lambda_0(t)dt$. It then follows that $E[\zeta(y_1)-\zeta(y_2)]^2\le c(\log n)/(n^2h)\cdot|y_1-y_2|^2$, which implies

$$||\zeta(y_1)-\zeta(y_2)||_2\le c\cdot\frac{1}{\sqrt{n^2h\log n}}|y_1-y_2|. \tag{A.13}$$

By Theorem 2.2.4 of van der Vaart and Wellner (1996) and (A.11)-(A.13), we obtain

$$\|\sup_{y\in[0,1]}\zeta(y)\|_2\le O(1)\cdot\frac{1}{\sqrt{n^2h\log n}}\int_0^1\sqrt{\frac{1}{\varepsilon}}d\varepsilon=o(n^{-1/2}),$$

which implies (a).

The proof of (ii) is analogous and is omitted.

A.2. Proof of Theorem 1

Set $\hat{\gamma}(x)=h\hat{\psi}'(x)$. The proof is presented in four steps.

Step (i). Show that $(\psi,0)$ is the unique root to the equation $\boldsymbol{u}(\eta,\gamma;\cdot)=0$, where $\eta(\cdot)\in\mathcal{C}_0$.

Lemma 1(i) shows that $\gamma(\cdot) \equiv 0$ is sufficient and necessary for $\boldsymbol{u}_2(\eta, \gamma; \cdot) = 0$. Substituting $\gamma(\cdot) \equiv 0$ into $\boldsymbol{u}_1(\eta, \gamma; x)$, we obtain

$$\boldsymbol{u}_1(\eta, 0; x) = \Gamma(x) \exp\{\psi(x)\} - \exp\{\eta(x)\} \int_0^\tau \frac{P(t|x)}{s_0(t;\eta)} s_0(t;\psi)\lambda_0(t)dt.$$

Lemma 2 shows that there exists one unique solution $\psi(\cdot)$ to $\boldsymbol{u}_1(\eta, 0; \cdot) = 0$ in $\mathcal{C}_0$.

Step (ii). Show that, for fixed continuous functions $\eta(\cdot)$ and $\gamma(\cdot)$,

$$\sup_{x\in[0,1]} \|U(\eta, \gamma; x) - \boldsymbol{u}(\eta, \gamma; x)\| = o_p(1). \tag{A.14}$$

Lemma 4 implies

$$\sup_{t\in[0,\tau],\ x\in[0,1]} \|S_{n1}(t; \eta, \gamma, x) - s_1(t; \eta, \gamma, x)\| = o_p(1). \tag{A.15}$$

Let $S^*_{n1}(t; x) = n^{-1} \sum_{i=1}^n \omega_i(x) N_i(t)$ and $s^*_1(t; x) = \exp\{\psi(x)\} \int_0^t P(s|x)\lambda_0(s)ds(1, 0)^\top$. The same argument leading to (A.15) gives

$$\sup_{t\in[0,\tau],\ x\in[0,1]} \|S^*_{n1}(t; x) - s^*_1(t; x)\| = o_p(1). \tag{A.16}$$

Let $\bar{N}(t) = n^{-1} \sum_{i=1}^n N_i(t)$. It follows from the uniform law of large numbers (Pollard,1990, p. 41) that

$$\sup_{t\in[0,\tau]} |\bar{N}(t) - \int_0^t s_0(s;\psi)\lambda_0(s)ds| = o_p(1), \tag{A.17}$$

and

$$\sup_{t\in[0,\tau]} |S_{n0}(t;\eta) - s_0(t;\eta)| = o_p(1). \tag{A.18}$$

Note that $U(\eta, \gamma; x) = S^*_{n1}(\tau; x) - \int_0^\tau S_{n1}(t; \eta, \gamma, x)[S_{n0}(t; \eta)]^{-1} d\bar{N}(t)$. Hence, (A.14) follows from (A.15)-(A.18).

Step (iii). Show that

$$\sup_{x\in[0,1],\ \eta,\gamma\in\mathcal{B}_n} \|U(\eta, \gamma; x) - \boldsymbol{u}(\eta, \gamma; x)\| = o_p(1). \tag{A.19}$$

For any $\varepsilon > 0$, let $\mathcal{H}$ denote the finite set of points mD/M, $m = 0, \pm 1, ..., M$, where M is an integer such that $D/M < \varepsilon/3$ ($\mathcal{H}$ is an ε-net for the linear interval

$[-D, D]$). Choose k and n large enough so that $d(1/k+b_n) < \varepsilon/3$. Let $\mathcal{V}$ consist of those elements of the space of continuous functions on $[0,1]$ that are linear on each subinterval $I_i = [(i-1)/k, i/k], i = 1, ..., k$, and assume values in $\mathcal{H}$ at points $i/k, i = 0, 1, ..., k$. The size of the set $\mathcal{V}$ is $(2M+1)^{k+1} = O((1/\varepsilon)^{1/\varepsilon})$. For any $\eta(\cdot) \in \mathcal{B}_n$, there exists an element $\varphi(\cdot)$ of $\mathcal{V}$ such that $|\eta(i/k) - \varphi(i/k)| < \varepsilon/3,\ i = 0, 1, ..., k$. Since $\varphi(\cdot)$ is linear on each subinterval I_i, it follows that $\sup_{x\in[0,1]} |\eta(x) - \varphi(x)| < \varepsilon$. That is, for all n sufficiently large, $\mathcal{V}$ is an ε-net for $\mathcal{B}_n$. As a result,

$$\log N(\varepsilon, \mathcal{B}_n, ||\cdot||_\infty) \leq O(\frac{1}{\varepsilon}\log(\frac{1}{\varepsilon})) = o(n), \tag{A.20}$$

where $N(\varepsilon, \mathcal{B}_n, ||\cdot||_\infty)$ is the covering number with respect to the norm $||\cdot||_\infty$ of the class $\mathcal{B}_n$. Hence (A.19) follows from the uniform law of large numbers (Pollard, 1990, p. 39), (A.14) and (A.20).

Step (iv). Show that

$$P\{\hat{\psi} \in \mathcal{B}_n,\ \hat{\gamma} \in \mathcal{B}_n\} \to 1. \tag{A.21}$$

We first show that

$$\sup_{x\in[0,1]} |\hat{\gamma}(x)| = O_p(b_n). \tag{A.22}$$

It follows from (A.1) that

$$\frac{n^{-1}\sum_{i=1}^n \int_0^\tau \omega_{i2}(x) dN_i(t)}{n^{-1}\sum_{i=1}^n \int_0^\tau \omega_{i1}(x) dN_i(t)} = \frac{\int_0^\tau \tilde{S}_{n2}(\hat{\gamma}, t; x)[S_{n0}(t;\hat{\psi})]^{-1} d\bar{N}(t)}{\int_0^\tau \tilde{S}_{n1}(\hat{\gamma}, t; x)[S_{n0}(t;\hat{\psi})]^{-1} d\bar{N}(t)}, \tag{A.23}$$

where $\tilde{S}_{nk}(\alpha, t; x) = n^{-1}\sum_{i=1}^n \omega_{ik}(x) Y_i(t) \exp\{\alpha(X_i - x)/h\}((X_i - x)/h)^{k-1},\quad k = 1, 2.$ Lemma 4 implies that the left hand side of (A.23) is $O_p(b_n)$. Similarly, we also have

$$\sup_{x\in[0,1],\ t\in[0,\tau],\ \alpha\in[\sigma_m,\sigma_M]} \Big|\tilde{S}_{nk}(\alpha, t; x) - \frac{P(t|x)\xi_k(x,\alpha)}{\mu_2(x)\mu_0(x) - \mu_1^2(x)}\Big| = O_p(b_n),\quad k = 1, 2, \tag{A.24}$$

where $[\sigma_m, \sigma_M]$ is bounded and $\xi_k(x, \alpha)$ is given by (A.2). Then, Lemma 1 (ii) ensures that for any $\sigma_0 > 0$, the right hand side of (A.23) does not converge to 0 in probability on the set $A_n = \{\sup_{x\in[0,1]} |\hat{\gamma}(x)| \geq \sigma_0 > 0\}$. Since the left hand side of (A.23) is $O_p(b_n)$, it follows that $P(A_n) \to 0$ as $n \to \infty$, which implies $\sup_{x\in[0,1]} |\hat{\gamma}(x)| = o_p(1)$. By the Taylor expansion, it is seen that $\sup_{x\in[0,1]} \left|\xi_2(x,\hat{\gamma})/\xi_1(x,\hat{\gamma})\right| \leq c \cdot \sup_{x\in[0,1]} |\hat{\gamma}(x)|(1 + o_p(1))$, where c is a positive constant. (A.22) now follows from (A.24) and by comparing the order of the left hand side of (A.23) with that of the right side.

Observe that $U_1(\hat{\psi},\hat{\gamma};x) = 0$, $\hat{\psi}(0) = 0$ and $\hat{\gamma}(x) = O_p(b_n)$ for $x \in [0,1]$. We have $\sup_{x\in[0,1]} |\hat{\psi}(x)| = O_p(1)$ and $\int_0^\tau \tilde{S}_{n1}(\hat{\gamma},t;0)[S_{n0}(t;\hat{\psi})]^{-1} d\bar{N}(t) = n^{-1}\sum_{i=1}^n \int_0^\tau \omega_{i1}(0) dN_i(t)$, which implies $S_{n0}(t;\hat{\psi})$ is strictly positive for all sufficiently large n. It follows from $U_1(\hat{\psi},\hat{\gamma};x) = 0$ that

$$\begin{aligned}
\exp\{\hat{\psi}(x)\} - \exp\{\hat{\psi}(y)\} = [G_n(x)]^{-1}\Big[H_n(x) - H_n(y) \\
- \int_0^\tau (I_n(x;t) - I_n(y;t)) \exp\{\hat{\psi}(y)\}[S_{n0}(t;\hat{\psi})]^{-1} d\bar{N}(t)\Big],
\end{aligned} \tag{A.25}$$

where $H_n(x) = n^{-1}\sum_{i=1}^n \omega_{i1}(x) N_i(\tau)$, $\quad G_n(x) = \int_0^\tau I_n(x;t)[S_{n0}(t;\hat{\psi})]^{-1} d\bar{N}(t)$, and $I_n(x;t) = n^{-1}\sum_{i=1}^n \omega_{i1}(x) Y_i(t) \exp\{\hat{\gamma}(x)(X_i - x)/h\}$. It can be checked from (C2), (C5)-(C6) that

$$H_n(x) = \Gamma(x)\exp\{\psi(x)\} + O_p(b_n) \tag{A.26}$$

uniformly over $x \in [0,1]$. By the Taylor expansion, we have

$$I_n(x,t) = \frac{1}{n}\sum_{i=1}^n \omega_{i1}(x) Y_i(t) + \hat{\gamma}(x) \cdot \frac{1}{n}\sum_{i=1}^n \omega_{i1}(x) Y_i(t)(X_i - x)/h + o_p(b_n)$$

uniformly over $x \in [0,1]$ and $t \in [0,\tau]$. Analogous to (A.26),

$$I_n(x,t) = P(t|x) + O_p(b_n) \tag{A.27}$$

uniformly over $x \in [0,1]$ and $t \in [0,\tau]$. It can also be shown that there exist some positive constants c_1, c_2 and c_3 such that $c_1 + o_p(1) \leq S_{n0}(t;\hat{\psi}) \leq c_2 + o_p(1)$ and $[G_n(x)]^{-1} = c_3 + o_p(1)$ uniformly over $x \in [0,1]$ and $t \in [0,\tau]$. Therefore, (A.25)-(A.27) and the mean value theorem ensure

$$|\hat{\psi}(x) - \hat{\psi}(y)| \leq c|x - y| + O_p(b_n), \tag{A.28}$$

for all $x, y \in [0,1]$. Consequently, (A.21) follows from (A.22) and (A.28).

Once $(i)-(iv)$ are established, using the Arzela-Ascoli theorem and (A.21), 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 $x \in [0,1]$, $(\hat{\psi}_{nk}, \hat{\gamma}_{nk}) \rightarrow (\psi^*, \gamma^*)$ in probability with $\psi^*(\cdot) \in \mathcal{C}_0$. Note that

$$\boldsymbol{u}(\psi^*, \gamma^*; x) = [\boldsymbol{u}(\psi^*, \gamma^*; x) - \boldsymbol{u}(\hat{\psi}_{n_k}, \hat{\gamma}_{n_k}; x)] + [\boldsymbol{u}(\hat{\psi}_{n_k}, \hat{\gamma}_{n_k}; x) - U(\hat{\psi}_{n_k}, \hat{\gamma}_{n_k}; x)].$$

It follows from (A.19) and $U(\hat{\psi}, \hat{\gamma}; \cdot) \equiv 0$ that the left hand side converges to 0 uniformly over $x \in [0,1]$. As a result, $\boldsymbol{u}(\psi^*, \gamma^*; \cdot) = 0$. Since $\boldsymbol{u}(\eta, \gamma; \cdot) = 0$ has a unique root $(\psi(\cdot), 0)$ by the result of step (i), we have $\psi^*(\cdot) \equiv \psi(\cdot)$ on $[0,1]$, which ensures the uniform consistency of $\hat{\psi}(\cdot)$. This completes the proof of Theorem 1.

A.3. Proof of Theorem 2

For convenience of notation, denote

$$\begin{aligned} c_n &= \sup_{x \in [0,1]} |\hat{\psi}(x) - \psi(x)|, \quad d_n = \sup_{x \in [0,1]} |h\hat{\psi}'(x) - h\psi'(x)|, \quad \text{and} \\ e_n &= h^2 + (nh)^{-1/2} (\log n)^{1/2}. \end{aligned}$$

Recall that $b_n = h + (nh)^{-1/2} (\log n)^{1/2}$, and $s_0(t) = E\{P(t|X) \exp\{\psi(X)\}\}$.

This proof is divided into the four steps.

Step 1. Show that uniformly over $x \in [0,1]$,

$$\begin{aligned} &\hat{\psi}(x) - \psi(x) - \frac{1}{n} \sum_{j=1}^{n} \exp\{\psi(X_j)\}[\hat{\psi}(X_j) - \psi(X_j)][\int_0^\tau Y_j(t) P(t|x) (\Gamma(x) s_0(t))^{-1} \lambda_0(t) dt] \\ &\quad = [\Gamma(x) \exp\{\psi(x)\}]^{-1} U_1(\psi, h\psi'; x) + O_p(c_n^2 + e_n^2 + b_n(c_n + e_n)). \qquad \text{(A.29)} \end{aligned}$$

Write

$$\begin{aligned} U(\hat{\psi}, h\hat{\psi}'; x) - U(\psi, h\psi'; x) &= \int_0^\tau \left[S_{n0}(t; \hat{\psi}) - S_{n0}(t; \psi) \right] \frac{S_{n1}(t; \hat{\psi}, h\hat{\psi}', x)}{S_{n0}(t; \psi) S_{n0}(t; \hat{\psi})} d\bar{N}(t) \\ &\quad + \int_0^\tau \left[S_{n1}(t; \psi, h\psi', x) - S_{n1}(t; \hat{\psi}, h\hat{\psi}', x) \right] [S_{n0}(t; \psi)]^{-1} d\bar{N}(t). \qquad \text{(A.30)} \end{aligned}$$

Let $\zeta_n(x) = (\hat{\psi}(x) - \psi(x), h(\hat{\psi}'(x) - \psi'(x)))^\top$. Using a Taylor expansion and the consistency of $(\hat{\psi}, h\hat{\psi}')$, we have, uniformly over $x \in [0,1]$ and $t \in [0,\tau]$,

$$S_{n1}(t; \psi, h\psi', x) = P(t|x)\exp\{\psi(x)\}(1,0)^\top + O_p(b_n), \tag{A.31}$$

$$\begin{aligned} & S_{n1}(t; \hat{\psi}, h\hat{\psi}', x) - S_{n1}(t; \psi, h\psi', x) \\ & = P(t|x)\exp\{\psi(x)\}\zeta_n(x) + O_p(c_n^2 + d_n^2 + e_n^2 + b_n(c_n + d_n + e_n)), \end{aligned} \tag{A.32}$$

and

$$\begin{aligned} & [S_{n0}(t; \hat{\psi}) - S_{n0}(t; \psi)][S_{n0}(t; \hat{\psi})]^{-1} \\ & = \frac{1}{s_0(t)} \frac{1}{n} \sum_{i=1}^n \exp\{\psi(X_i)\}[\hat{\psi}(X_i) - \psi(X_i)]Y_i(t) + O_p(c_n^2). \end{aligned} \tag{A.33}$$

Lemma 4 implies

$$\sup_{x \in [0,1]} \left\| U(\psi, h\psi'; x) \right\| = O_p(e_n). \tag{A.34}$$

Note that

$$U(\hat{\psi}, h\hat{\psi}'; x) - U(\psi, h\psi'; x) \equiv -U(\psi, h\psi'; x). \tag{A.35}$$

It follows from the second component in (A.31)-(A.35) that

$$d_n = O_p(c_n^2 + b_n c_n + e_n).$$

Furthermore, it is seen that

$$d_n^2 = O_p(c_n^2 + e_n^2) \quad \text{and} \quad b_n d_n = O_p(b_n(c_n + e_n)). \tag{A.36}$$

Consequently, (A.29) follows from (A.36) and the first component in (A.30)-(A.35).

Step 2. Show that

$$\begin{aligned} & \sup_{x \in [0,1]} \Big| \frac{1}{n} \sum_{i=1}^n \exp\{\psi(X_i)\}[\hat{\psi}(X_i) - \psi(X_i)][\int_0^\tau Y_i(t) P(t|x)(\Gamma(x) s_0(t))^{-1} \lambda_0(t) dt] \\ & - \int_0^\tau \Phi(u|x)(\hat{\psi}(u) - \psi(u)) du \Big| = O_p(c_n^2 + e_n^2 + b_n(c_n + e_n)) + o_p(n^{-1/2}). \end{aligned} \tag{A.37}$$

By letting $x = X_i$, multiplying $n^{-1}\exp\{\psi(X_i)\}[\int_0^\tau Y_i(t)P(t|x)(s_0(t))^{-1}\lambda_0(t)dt]$ on both sides of (A.29), and taking sum over $i = 1, \cdots, n$, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\exp\{\psi(X_i)\}[\hat{\psi}(X_i)-\psi(X_i)][\int_0^\tau Y_i(t)P(t|x)(s_0(t))^{-1}\lambda_0(t)dt] \\
&\quad -\frac{1}{n}\sum_{j=1}^{n}\exp\{\psi(X_j)\}[\hat{\psi}(X_j)-\psi(X_j)][\int_0^\tau Y_j(t)A_n(x,t)(s_0(t))^{-1}\lambda_0(t)dt] \\
&= \frac{1}{n}\sum_{i=1}^{n}[\Gamma(X_i)]^{-1}\int_0^\tau Y_i(t)P(t|x)(s_0(t))^{-1}\lambda_0(t)dt\cdot U_1(\psi, h\psi'; X_i) \\
&\quad + O_p(c_n^2+e_n^2+b_n(c_n+e_n)),
\end{aligned} \tag{A.38}$$

where $A_n(x,t) = n^{-1}\sum_{i=1}^n P(t|X_i)[\Gamma(X_i)]^{-1}\exp\{\psi(X_i)\}\int_0^\tau Y_i(\upsilon)P(\upsilon|x)(s_0(\upsilon))^{-1}\lambda_0(\upsilon)d\upsilon$. On the other hand, changing x into u, multiplying $[\int_0^\tau P(t|u)P(t|x)(s_0(t))^{-1}\lambda_0(t)dt]$ $\exp\{\psi(u)\}f(u)$ on both sides of (A.29), and integrating from 0 to 1, we obtain

$$\begin{aligned}
&\int_0^1 \exp\{\psi(u)\}f(u)(\hat{\psi}(u)-\psi(u))[\int_0^\tau P(t|u)P(t|x)(s_0(t))^{-1}\lambda_0(t)dt]du \\
&\quad -\frac{1}{n}\sum_{j=1}^{n}\exp\{\psi(X_j)\}[\hat{\psi}(X_j)-\psi(X_j)][\int_0^\tau Y_j(t)A(x,t)(s_0(t))^{-1}\lambda_0(t)dt] \\
&= \int_0^1[\Gamma(u)]^{-1}[\int_0^\tau P(t|u)P(t|x)(s_0(t))^{-1}\lambda_0(t)dt]f(u)U_1(\psi, h\psi'; u)du \\
&\quad + O_p(c_n^2+e_n^2+b_n(c_n+e_n)),
\end{aligned} \tag{A.39}$$

where $A(x,t) = \int_0^1 P(t|u)\exp\{\psi(u)\}f(u)[\Gamma(u)]^{-1}\int_0^\tau P(\upsilon|u)P(\upsilon|x)(s_0(\upsilon))^{-1}\lambda_0(\upsilon)d\upsilon du$. Applying the inequality (7.10) of Pollard (1990), we have

$$\sup_{t\in[0,\tau],\ x\in[0,1]}\Big|A_n(x,t)-A(x,t)\Big| = O_p(n^{-1/2}). \tag{A.40}$$

Subtracting (A.38) by (A.39), it follows from (A.40) that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\exp\{\psi(X_i)\}[\hat{\psi}(X_i)-\psi(X_i)][\int_0^\tau Y_i(t)P(t|x)(s_0(t))^{-1}\lambda_0(t)dt] \\
&\quad -\int_0^1 \exp\{\psi(u)\}f(u)(\hat{\psi}(u)-\psi(u))[\int_0^\tau P(t|u)P(t|x)(s_0(t))^{-1}\lambda_0(t)dt] \\
&= \frac{1}{n}\sum_{i=1}^{n}[\Gamma(X_i)]^{-1}[\int_0^\tau Y_i(t)P(t|x)(s_0(t))^{-1}\lambda_0(t)dt]U_1(\psi, h\psi'; X_i) \\
&\quad -\int_0^1[\Gamma(u)]^{-1}[\int_0^\tau P(t|u)P(t|x)(s_0(t))^{-1}\lambda_0(t)dt]f(u)U_1(\psi, h\psi'; u)du \\
&\quad + O_p(c_n^2+e_n^2+b_n(c_n+e_n)) + o_p(n^{-1/2}).
\end{aligned} \tag{A.41}$$

Lemma 5 shows that the first two terms on the right hand side of (A.41) are $o_p(n^{-1/2})$. Therefore, (A.37) holds.

Step 3. Show that

$$c_n = O_p(e_n). \tag{A.42}$$

Combining (A.29) with (A.37), we have

$$\begin{aligned}\hat{\psi}(x) - \psi(x) - \int_0^\tau \Phi(u|x)(\hat{\psi}(u) - \psi(u))du &= [\Gamma(x)\exp\{\psi(x)\}]^{-1}U_1(\psi, h\psi'; x) \\ &+O_p(c_n^2 + e_n^2 + b_n(c_n + e_n)) + o_p(n^{-1/2}),\end{aligned} \tag{A.43}$$

uniformly over $x \in [0, 1]$. Denote $\mathcal{F}_0 = \{\phi(\cdot) : \phi(\cdot) = \Phi(\cdot|x) - \Phi(\cdot|0),\ x \in [0, 1]\}$. Observe that $\mathcal{F}_0 \subset \mathcal{D}_0$. Lemma 3 implies that there exists $\tilde{\mathcal{F}}_0$ such that $\tilde{\mathcal{F}}_0 = \{g(\cdot) : g(\cdot)$ satisfies (A.7), $\phi(\cdot) \in \mathcal{F}_0\}$. Moreover, $\sup_{g\in\tilde{\mathcal{F}}_0} \sup_{x\in[0,1]} |g(x)| = O(1)$ holds. For $g(\cdot) \in \tilde{\mathcal{F}}_0$ and $\phi(\cdot) \in \mathcal{F}_0$ satisfying (A.7), by multiplying $\Gamma(x)\exp\{\psi(x)\}f(x)g(x)$ on both sides of (A.43) and integrating from 0 to 1, we get

$$\begin{aligned}\int_0^1 \phi(u)(\hat{\psi}(u) - \psi(u))du &= \int_0^1 g(u)f(u)U_1(\psi, h\psi'; u)du \\ &\quad +O_p(c_n^2 + e_n^2 + b_n(c_n + e_n)) + o_p(n^{-1/2}),\end{aligned}$$

uniformly over $g(\cdot) \in \tilde{\mathcal{F}}_0$. Hence we have

$$\begin{aligned}\sup_{\phi\in\mathcal{F}_0} |\int_0^1 \phi(u)(\hat{\psi}(u) - \psi(u))du| &= \sup_{g\in\tilde{\mathcal{F}}_0} \Big|\int_0^1 g(u)f(u)U_1(\psi, h\psi'; u)du\Big| \\ &\quad +O_p(c_n^2 + e_n^2 + b_n(c_n + e_n)) + o_p(n^{-1/2}).\end{aligned} \tag{A.44}$$

(A.34) and (A.44) imply

$$\sup_{x\in[0,1]} \Big|\int_0^1 [\Phi(u|x) - \Phi(u|0)][\hat{\psi}(u) - \psi(u)]du\Big| = O_p(e_n + c_n^2 + b_n c_n). \tag{A.45}$$

It can be shown from (A.43), together with the identifiability condition $\hat{\psi}(0) = \psi(0) = 0$, that

$$\begin{aligned}&\int_0^\tau \Phi(u|0)(\hat{\psi}(u) - \psi(u))du \\ &= -[\Gamma(0)]^{-1}U_1(\psi, h\psi'; 0) + O_p(c_n^2 + e_n^2 + b_n(c_n + e_n)) + o_p(n^{-1/2}).\end{aligned}$$

Consequently,

$$
\begin{aligned}
&\hat{\psi}(x) - \psi(x) - \int_0^\tau [\Phi(u|x) - \Phi(u|0)](\hat{\psi}(u) - \psi(u))du \\
&= [\Gamma(x)\exp\{\psi(x)\}]^{-1}U_1(\psi, h\psi'; x) - [\Gamma(0)]^{-1}U_1(\psi, h\psi'; 0) \\
&\quad + O_p(c_n^2 + e_n^2 + b_n(c_n + e_n)) + o_p(n^{-1/2}),
\end{aligned} \tag{A.46}
$$

uniformly over $x \in [0,1]$. Again by (A.34), (A.45) and (A.46), it is seen that $c_n = O_p(e_n + c_n^2 + b_n c_n) + o_p(n^{-1/2})$, which implies $c_n = O_p(e_n)$ and (A.42).

Step 4. Show that (9) in Section 3 holds for each fixed point $x \in (0,1)$.

We find from (A.42) and (A.43) that

$$
\begin{aligned}
&\hat{\psi}(x) - \psi(x) - \int_0^\tau \Phi(u|x)(\hat{\psi}(u) - \psi(u))du \\
&= [\Gamma(x)\exp\{\psi(x)\}]^{-1}U_1(\psi, h\psi'; x) + O_p(e_n^2 + b_n e_n) + o_p(n^{-1/2}),
\end{aligned} \tag{A.47}
$$

uniformly over $x \in [0,1]$. It is seen that $e_n^2 + b_n e_n = o_p(h^2 + (nh)^{-1/2})$ by condition (C7). To establish (9), we need to derive the asymptotic expansion of $U_1(\psi, h\psi'; x)$. To this end, let $S_{n1}(t;x)$ and $s_1(t;x)$ be the first entries of $S_{n1}(t;\psi, h\psi', x)$ and $s_1(t;\psi, 0, x)$, respectively. Recall 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

$$
U_1(\psi, h\psi'; x) = V_n(x) + B_n(x), \tag{A.48}
$$

where

$$
\begin{aligned}
V_n(x) &= \frac{1}{n}\sum_{i=1}^n \int_0^\tau \Big[\omega_{i1}(x) - \frac{S_{n1}(t;x)}{S_{n0}(t;\psi)}\Big] dM_i(t) \quad \text{and} \\
B_n(x) &= \frac{1}{n}\sum_{i=1}^n \int_0^\tau \Big[\omega_{i1}(x) - \frac{S_{n1}(t;x)}{S_{n0}(t;\psi)}\Big] Y_i(t)\exp\{\psi(X_i)\}\lambda_0(t)dt.
\end{aligned}
$$

The martingale central limit theorem implies that

$$
(nh)^{1/2}V_n(x) = (nh)^{1/2}\cdot\frac{1}{n}\sum_{i=1}^n \int_0^\tau \Big[\omega_{i1}(x) - \frac{s_1(t;x)}{s_0(t;\psi)}\Big] dM_i(t) + o_p(1), \tag{A.49}
$$

which is asymptotically normal with mean zero and variance

$$\sigma_0^2(x) = \Gamma(x)\exp\{\psi(x)\}\frac{\int_{\mathcal{Z}(x,h)}[\mu_2(x) - u\mu_1(x)]^2 K^2(u)du}{f(x)[\mu_2(x)\mu_0(x) - \mu_1^2(x)]^2}. \tag{A.50}$$

On the other hand, it follows from a Taylor expansion that, for $|X_i - x| \le h$,

$$\exp\{\psi(X_i)\} - \exp\{\psi(x) + \psi'(x)(X_i - x)\} = \frac{1}{2}\exp\{\psi(x)\}\psi''(x)(X_i - x)^2[1 + o_p(1)].$$

Thus,

$$B_n(x) = \frac{1}{2}h^2\Gamma(x)\exp\{\psi(x)\}\psi''(x)\frac{\mu_2^2(x) - \mu_1(x)\mu_3(x)}{\mu_2(x)\mu_0(x) - \mu_1^2(x)} + o_p(h^2). \tag{A.51}$$

From (A.50)-(A.51), we find that for fixed $x \in (0, 1)$ and all large n,

$$\begin{aligned} \sigma_0^2(x) &= \Gamma(x)\exp\{\psi(x)\}\nu/f(x), \\ B_n(x) &= \frac{1}{2}h^2\mu\Gamma(x)\exp\{\psi(x)\}\psi''(x) + o_p(h^2). \end{aligned}$$

Therefore, (9) follows.

A.4. Proof of Proposition 1

The proof follows the same lines as that of Theorem 2. The key difference is the calculation of $\sigma_0^2(x)$ and $B_n(x)$ in (A.50) and (A.51) when $x = x_n$. Let

$$\tilde{\mu}_c = \frac{\mu_{2,c}^2 - \mu_{1,c}\mu_{3,c}}{\mu_{2,c}\mu_{0,c} - \mu_{1,c}^2} \quad \text{and} \quad \tilde{\nu}_c = \frac{\int_{-1}^{c}[\mu_{2,c} - u\mu_{1,c}]^2 K^2(u)du}{[\mu_{2,c}\mu_{0,c} - \mu_{1,c}^2]^2}, \tag{A.52}$$

where $\mu_{k,c} = \int_{-1}^{c} u^k K(u)du, k = 0, \cdots, 3$. Then, some straightforward calculations yield

$$\begin{aligned} \sigma_0^2(x_n) &= \Gamma(0)\tilde{\nu}_c/f(0)(1 + o(1)) \\ \text{and} \qquad B_n(x_n) &= \frac{1}{2}h^2\tilde{\mu}_c\Gamma(0)\psi''(0)(1 + o(1)) + o_p(h^2). \end{aligned}$$

Therefore, (11) holds.

A.5. Proof of Theorem 3

For a fixed function $\phi(\cdot) \in \mathcal{D}_0$, let $g(\cdot) \in \mathcal{C}_0$ satisfy the integral equation (A.7), see Lemma 3. First, we derive the asymptotic variance of $\int_0^1 \hat{\psi}(x)\phi(x)dx$. Note that $nh^4 \to 0$ and $nh^2/(\log n)^2 \to \infty$. From (A.7) and (A.47), we have

$$\int_0^1 \big[\hat{\psi}(x) - \psi(x)\big]\phi(x)dx = \int_0^1 U_1(\psi, h\psi'; x)f(x)g(x)dx + o_p(n^{-1/2}). \qquad \text{(A.53)}$$

Let $S_{ng}(t) = n^{-1}\sum_{i=1}^n Y_i(t)\exp\{\psi(X_i)\}g(X_i)$. It follows that

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

Note that $nh^4 \to 0$. Applying the martingale central limit theorem, we have

$$n^{1/2}\int_0^1 U_1(\psi, h\psi'; x)f(x)g(x)dx \xrightarrow{\mathcal{L}} N(0, \tilde{\sigma}^2), \qquad \text{(A.54)}$$

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

$$\tilde{\sigma}^2 = E\Big\{\int_0^\tau [g(X) - \bar{g}(t)]^2 P(t|X)\exp\{\psi(X)\}\lambda_0(t)dt\Big\}. \qquad \text{(A.55)}$$

(A.53) and (A.54) entail

$$n^{1/2}\int_0^1 [\hat{\psi}(x) - \psi(x)]\phi(x)dx \xrightarrow{\mathcal{L}} N(0, \tilde{\sigma}^2). \qquad \text{(A.56)}$$

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 $g(\cdot)$ is the solution to the integral equation (A.7) for a given $\phi(\cdot) \in \mathcal{D}_0$. Let $\theta_0 = 0$ be the true value of θ. The score for this parametric submodel at θ_0 is

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

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

$$\sqrt{n}(\tilde{\theta} - \theta_0) \to N(0, \tilde{\sigma}^{-2}). \qquad \text{(A.57)}$$

Observe that

$$\int_0^1 [\psi(x;\tilde{\theta}) - \psi(x;\theta_0)]\phi(x)dx = (\hat{\theta} - \theta_0)\int_0^1 g(x)\phi(x)dx,$$

$$\text{and} \qquad \int_0^1 g(x)\phi(x)dx = E\Big\{\int_0^\tau g(X)[g(X) - \bar{g}(t)]dN(t)\Big\} = \tilde{\sigma}^2. \qquad (A.58)$$

Thus, it follows from (A.57)-(A.58) that

$$n^{1/2}\int_0^1 [\psi(x;\tilde{\theta}) - \psi(x;\theta_0)]\phi(x)dx \xrightarrow{\mathcal{L}} N(0,\tilde{\sigma}^2). \qquad (A.59)$$

This, together with (A.56) and (A.59), 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 in Bickel, Klaassen, Ritov and Wellner (1993, p.46), $\int_0^1 \hat{\psi}(x)\phi(x)dx$ is asymptotically efficient for the estimation of $\int_0^1 \psi(x)\phi(x)dx$. The proof of Theorem 3 is complete.

A.6. 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

$$\phi(x) = \int_0^\tau b(t)\exp\{\psi(x)\}P(t|x)\lambda_0(t)/s_0(t)dtf(x).$$

Let g be the solution of (A.7), which can be re-expressed 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,

$$\operatorname{cov}\Big(\int_0^\tau [g(X) - \bar{g}(t) - b(t)/s_0(t)]dM(t),\ \int_0^\tau g(X)dM(t)\Big) = 0. \qquad (A.60)$$

To show the asymptotic normality of $\int_0^\tau b(t)d\hat{\Lambda}_0(t)$, write

$$\begin{aligned}&\int_0^\tau b(t)d\hat{\Lambda}_0(t) - \int_0^\tau b(t)d\Lambda_0(t)\\ =\ &\int_0^\tau \frac{b(t)}{\sum_{j=1}^n \exp\{\hat{\psi}(X_j)\}Y_j(t)}\sum_{i=1}^n [dN_i(t) - \exp\{\hat{\psi}(X_i)\}Y_i(t)\lambda_0(t)dt]\end{aligned}$$

$$
\begin{aligned}
&= \int_0^\tau \frac{b(t)}{\sum_{j=1}^n \exp\{\hat\psi(X_j)\}Y_j(t)} \sum_{i=1}^n dM_i(t) \\
&\quad - \int_0^\tau \frac{b(t)}{\sum_{j=1}^n \exp\{\hat\psi(X_j)\}Y_j(t)} \sum_{i=1}^n [\exp\{\hat\psi(X_i)\} - \exp\{\psi(X_i)\}]Y_i(t)\lambda_0(t)dt \\
&= \int_0^\tau \frac{b(t)}{\sum_{j=1}^n \exp\{\psi(X_j)\}Y_j(t)} \sum_{i=1}^n dM_i(t) \\
&\quad - \sum_{i=1}^n [\hat\psi(X_i) - \psi(X_i)] \int_0^\tau \frac{b(t)\exp\{\psi(X_i)\}Y_i(t)\lambda_0(t)}{\sum_{j=1}^n \exp\{\hat\psi(X_j)\}Y_j(t)} dt + o_P(n^{-1/2}) \\
&= \frac{1}{n}\int_0^\tau \frac{b(t)}{s_0(t)} \sum_{i=1}^n dM_i(t) \\
&\quad - \frac{1}{n}\sum_{i=1}^n [\hat\psi(X_i) - \psi(X_i)] \int_0^\tau \frac{b(t)\exp\{\psi(X_i)\}Y_i(t)\lambda_0(t)}{s_0(t)} dt + o_P(n^{-1/2}) \\
&= \frac{1}{n}\int_0^\tau \frac{b(t)}{s_0(t)} \sum_{i=1}^n dM_i(t) \\
&\quad - \int_0^1 [\hat\psi(x) - \psi(x)] \int_0^\tau \frac{b(t)\exp\{\psi(x)\}P(t|x)\lambda_0(t)}{s_0(t)} dt\, f(x)dx + o_P(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^n \int_0^\tau [\frac{b(t)}{s_0(t)} - g(X_i) + \bar g(t)]dM_i(t) + o_P(n^{-1/2}),
\end{aligned}
$$

where the last equality is due to (A.53)-(A.54). As a result,

$$
\sqrt{n}\Big(\int_0^\tau b(t)d\hat\Lambda_0(t) - \int_0^\tau b(t)d\Lambda_0(t)\Big) \to N(0,\sigma^2),
$$

where

$$
\begin{aligned}
\sigma^2 &= \mathrm{var}\Big(\int_0^\tau [\frac{b(t)}{s_0(t)} - g(X) + \bar g(t)]dM(t)\Big) \\
&= \mathrm{cov}\Big(\int_0^\tau [\frac{b(t)}{s_0(t)} - g(X) + \bar g(t)]dM(t),\ \int_0^\tau [\frac{b(t)}{s_0(t)} + \bar g(t)]dM(t)\Big) \qquad \text{by (A.60)} \\
&= \mathrm{cov}\Big(\int_0^\tau [\frac{b(t)}{s_0(t)}]dM(t),\ \int_0^\tau [\frac{b(t)}{s_0(t)} + \bar g(t)]dM(t)\Big) \\
&= \int_0^\tau \frac{b(t)}{s_0(t)}[\frac{b(t)}{s_0(t)} + \bar g(t)]E(\exp\{\psi(X)\}Y(t))\lambda_0(t)dt \\
&= \int_0^\tau b(t)[\frac{b(t)}{s_0(t)} + \bar g(t)]\lambda_0(t)dt,
\end{aligned}
$$

where the third equality is by the orthogonality of $\int_0^\tau [g(X)-\bar g(t)]dM(t)$ with $\int_0^\tau h(t)dM(t)$ for any nonrandom bounded function h, and the last equality is by the definition of $s_0(t)$.

Consider a parametric submodel with parameter θ, such that $\Lambda_0(\cdot;\theta) = \Lambda_0(\cdot) + \theta A(\cdot)$ and $\psi(\cdot;\theta) = \psi(\cdot) - \theta g(\cdot)$, where the derivative of $A(t)$, denoted as $a(t)$, is chosen to be $\lambda_0(t)[b(t)/s_0(t) + \bar{g}(t)]$. Let $\theta_0 = 0$ be the true value of θ. Then the score for this parametric submodel at θ_0 is

$$\int_0^\tau [-g(X) + a(t)/\lambda_0(t)]dM(t) = \int_0^\tau [\frac{b(t)}{s_0(t)} - g(X) + \bar{g}(t)]dM(t),$$

whose variance is σ^2, as shown above. And the maximum likelihood estimator of θ, denoted as $\hat{\theta}$, satisfies

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

Therefore

$$\sqrt{n}[\int_0^\tau b(t)d\Lambda_0(t;\hat{\theta}) - \int_0^\tau b(t)d\Lambda_0(t;\theta_0)] = \sqrt{n}(\hat{\theta} - \theta_0)\int_0^\tau b(t)a(t)dt$$

$$\to N(0,\ \sigma^{-2}(\int_0^\tau b(t)a(t)dt)^2).$$

The definitions of $a(t)$ and σ^2 imply that $\sigma^2 = \int_0^\tau b(t)a(t)dt$. Hence, the asymptotic variance of $\int_0^\tau b(t)d\Lambda_0(t;\hat{\theta})$ is σ^2. In summary, the asymptotic variance of $\int_0^\tau b(t)d\hat{\Lambda}_0(t)$, which is also σ^2, achieves the asymptotic variance of the efficient estimator in a parametric submodel. Therefore, the estimator is semiparametric efficient. The proof of Theorem 4 is complete.

Appendix B: Convergence of the iterative algorithm

We prove the convergence of the iterative algorithm for the case of local constant fitting (i.e., when $p = 0$ in (4) and $\beta \equiv 0$ in the estimating equation (6)) under conditions (C2)-(C4) and (C7). In addition, we assume that given the censoring time C, the density function of X is continuous and positive on [0,1]. Without loss of generality, we assume that $\tilde{T}_1 < ... < \tilde{T}_n$ and that the observed failure times are $\tilde{T}_{i_1} < ... < \tilde{T}_{i_L}$, where L is the number of the observed failure times. It can be checked that the solution of $\alpha(x)$ is determined by the solutions of $\alpha(X_1), ..., \alpha(X_n)$. Thus, we just consider the solutions of $\alpha(X_1), ..., \alpha(X_n)$, which satisfy the following estimating equations under local constant fitting:

$$\sum_{i=1}^{n}\int_0^\infty \Big[K_i(X_l) - \frac{\sum_{j=1}^n K_j(X_l)\exp\{\alpha(X_l)\}Y_j(t)}{\sum_{j=1}^n \exp\{\alpha(X_j)\}Y_j(t)}\Big]dN_i(t) = 0, \quad l = 1, ..., n. \qquad \text{(B.1)}$$

Let $\alpha_j^{(m)} = \exp\{\alpha^{(m)}(X_j)\}$ for $j = 1, ..., n$ and $m = 1, 2, ...$ For convenience and identifiability, we assume that $\sum_{i\in\mathcal{D}} \alpha_i^{(m)} = 1$ in any iteration step m, which is equivalent to the identifiability condition in the proposed algorithm, where $\mathcal{D} = \{j : \tilde{T}_j \geq \tilde{T}_{i_L}\}$. Define $d(x) = \sum_{i=1}^n K_i(x)\delta_i$ and $b_i(x) = \sum_{j=1}^n K_j(x)I(\tilde{T}_j \geq \tilde{T}_i)\delta_i$ for $i = 1, ..., n$. Let $J = i_L$. Using (B.1) and the proposed iterative procedure, we have that $\alpha_l^{(m+1)} = G_l(\alpha_1^{(m)}, ..., \alpha_{J-1}^{(m)})$, $l = 1, ..., J-1$, where

$$G_l(\alpha_1^{(m)}, ..., \alpha_{J-1}^{(m)}) = d(X_l)\Big(\sum_{i=1}^{J-1}\frac{b_i(X_l)}{\sum_{j=1}^{J-1}\alpha_j^{(m)}I(\tilde{T}_j \geq \tilde{T}_i) + 1} + \sum_{j=J}^{n} b_j(X_l)\Big)^{-1}. \qquad \text{(B.2)}$$

Let M be the integer part of $2/h$. Take $x_k = kh/2$, $k = 0, 1, ..., M$, and $x_{M+1} = 1$. For $x \in [0, 1]$ and all x_k, $k = 0, 1, ..., M+1$, define

$$D_x = \cup_{j=J}^n\{|X_j - x| \leq h\} \quad \text{and} \quad \tilde{D}_k = \cup_{j=J}^n\{|X_j - x_k| \leq h/2\}.$$

Let $\tilde{D}_k^c$ denote the complementary set of $\tilde{D}_k$. Then

$$\begin{aligned}
P\Big(\cap_{x\in[0,1]} D_x\Big) &\geq P\Big(\cap_{k=0}^{M+1}\tilde{D}_k\Big) \geq 1 - \sum_{k=0}^{M+1} P(\tilde{D}_k^c) \\
&= 1 - \sum_{k=0}^{M+1} P\Big(\cap_{j=J}^n \{|X_j - x_k| > h/2\}\Big) \\
&= 1 - \sum_{k=0}^{M+1}\Big(\prod_{j=J}^{n} P\{|X_j - x_k| > h/2\}\Big) \\
&\geq 1 - 2(1 + h^{-1})\Big(1 - P\{|X_j - x_k| \leq h/2\}\Big)^{n-J+1} \\
&= 1 - 2(1 + h^{-1})\Big(1 - f(\xi_k^*)h\Big)^{n-J+1},
\end{aligned}$$

where ξ_k^* lies between $x_k - h/2$ and $x_k + h/2$. In the sequel, let $c > 0$ denote a generic constant, which may be different at different places. Note that for n large enough,

$n-J+1=O(n)$ and $1-x\le\exp\{-x\}$ for $x\ge 0$. Hence under (C4) and (C7), we get that for n large enough,

$$\begin{aligned}&P\Big(\cap_{x\in[0,1]}D_x\Big)\ge 1-2(1+n)\exp\{-cnh\}\\&\ge 1-2(1+n)\exp\{-c\log n\}=1-n^{-c}.\end{aligned}\tag{B.3}$$

Similarly to (B.3), we have, for some large c,

$$P\Big\{\sum_{j=J}^{n}b_j(X_l)>0\text{ for all }l=1,...,n\Big\}\ge P\Big(\cap_{x\in[0,1]}D_x\Big)\ge 1-n^{-c}.$$

Thus, it follows from Borel-Cantelli lemma that with probability one,

$$\max_{1\le l\le J-1}\alpha_l^{(m+1)}\le c^*\quad\text{for some constant }c^*>0.$$

Define $\Omega=\{z=(z_1,...,z_{J-1})^\top\in R^{J-1},0\le z_j\le c^*,\ j=1,...,J-1\}$, and set $G(z)=(G_1(z),G_2(z),\cdots,G_{J-1}(z))^\top$. Since $G(z)$ is a continuous map from Ω to Ω, it follows from the Brouwer fixed point theorem that there exists $z^*=(z_1^*,...,z_{J-1}^*)^\top$ in Ω such that $z^*=G(z^*)$. In addition, it can be seen from (B.1) that $z_l^*>0$ for $l=1,...,J-1$.

Next, we show that the convergence of the iterative algorithm. Write $\alpha_l^{(m)}=p_l^{(m)}z_l^*$ for $l=1,...,J-1$. It follows from (B.2) that

$$p_l^{(m+1)}=\frac{d(X_l)}{z_l^*}\Big(\sum_{i=1}^{J-1}\frac{b_i(X_l)}{\sum_{j=1}^{J-1}p_j^{(m)}z_j^*I(\tilde T_j\ge\tilde T_i)+1}+\sum_{j=J}^{n}b_j(X_l)\Big)^{-1}.\tag{B.4}$$

Denote $r^{(m)}=\max_{1\le l\le J-1}p_l^{(m)}$. If there exists a positive number m such that $r^{(m)}\le 1$, then taking $m\to\infty$ in (B.4) and using $z^*=G(z^*)$, we have

$$p_l^{(m+1)}\le\frac{d(X_l)}{z_l^*}\Big(\sum_{i=1}^{J-1}\frac{b_i(X_l)}{\sum_{j=1}^{J-1}z_j^*I(\tilde T_j\ge\tilde T_i)+1}+\sum_{j=J}^{n}b_j(X_l)\Big)^{-1}=1,$$

which means that $r^{(m+1)}\le 1$ and $\overline{\lim}_{m\to\infty}r^{(m)}\le 1$.

If $r^{(m)}>1$ for all m, we have

$$p_l^{(m+1)}<\frac{d(X_l)}{z_l^*}\Big(\sum_{i=1}^{J-1}\frac{b_i(X_l)}{r^{(m)}\sum_{j=1}^{J-1}z_j^*I(\tilde T_j\ge\tilde T_i)+r^{(m)}}+\frac{\sum_{j=J}^{n}b_j(X_l)}{r^{(m)}}\Big)^{-1}=r^{(m)},$$

which yields that $r^{(m+1)} < r^{(m)}$ and $\lim_{m\to\infty} r^{(m)} = r$ exists. Suppose $r > 1$, a contradiction appears as $r < r$ by taking $m \to \infty$ in (B.4). Therefore,

$$\overline{\lim}_{m\to\infty} \max_{1\le l\le\ J-1} p_l^{(m)} \le 1.$$

In a similar manner, we can show

$$\underline{\lim}_{m\to\infty} \min_{1\le l\le J-1} p_l^{(m)} \ge 1.$$

Thus, $\lim_{m\to\infty} p_l^{(m)} = 1$ for $l = 1, ..., J-1$, which implies that $\alpha_l^{(m)}$ is convergent for $l = 1, ..., J-1$. Now it follows from (B.1) that $\alpha_l^{(m+1)} = G_l(\alpha_1^{(m)}, \cdots, \alpha_{J-1}^{(m)})$ for $l = J, ..., n-1$, where

$$G_l(\alpha_1^{(m)}, \cdots, \alpha_{J-1}^{(m)}) = d_l\Big(\sum_{i=1}^{J-1} \frac{b_i(X_l)}{\sum_{j=1}^{J-1} \alpha_j^{(m)} I(\tilde{T}_j \ge \tilde{T}_{(i)}) + 1} + \sum_{j=J}^{n} b_j(X_l)\Big)^{-1}.$$

Hence, for $l = J, ..., n-1$, $\alpha_l^{(m+1)}$ are determined by $\alpha_l^{(m)}, l = 1, ..., J-1$, and they are convergent. Therefore, the proposed iterative algorithm for local constant fitting converges and the solution of equation (B.1) is unique, which completes the proof.

References

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

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), Elsevier, Amsterdam, 217-231.

Chen, K., Jin, Z. and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89**, 659-668.

Chen, S. and Zhou, L. (2007). Local partial likelihood estimation in proportional hazards regression. *The Annals of Statistics* **35**, 888-916.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society* B **4**, 187-220.

Crowley, J. and Hu, M. (1977). Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association* **72**, 27-36.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association* **87**, 998-1004.

Fan, J. and Gijbels, I. (1994). Censored regression: local linear approximations and their applications. *Journal of the American Statistical Association* **89**, 560-570.

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

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

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

Gentleman, R. and Crowley, J. (1991). Local full likelihood estimation for the proportional hazards model. *Biometrics* **47**, 1283-1296.

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

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

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.

Kooperberg, C., Stone, C. J. and Truong, Y. (1995). Hazard regression. *Journal of the American Statistical Association* **90**, 78-94.

Miller, R. and Halpern, J. (1982). Regression with censored data. *Biometrika* **69**, 521-531.

O'Sullivan, F. (1988). Nonparametric estimation of relative risk using splines and cross-validation. *SIAM Journal on Scientific and Statistical Computing* **9**, 531-542.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

Pollard, D. (1990). *Empirical Processes: Theory and Applications.* Institute of Mathematical Statistics, Hayward, California.

Speckman, P. (1988). Kernel smoothing in partial linear models, *Journal of the Royal Statistical Society* B **50**, 413-436.

Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation. *Journal of the American Statistical Association* **82**, 559-567.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes: with applications to statistics.* Springer, New York.

Table 1. Summary of the WMISE for the estimators in Models 1-2.

Two initial bandwidths h_0 are specified for CZ in the second row.

bandwidth	Model 1			Model 2		
h	GPL	FGK	CZ $h_0 = 0.7(1.25h)$	GPL	FGK	CZ $h_0 = 0.7(1.25h)$
0.30	0.0953	0.1966	0.0875(0.2382)	0.1279	0.3337	0.1259(0.1478)
0.40	0.0684	0.1325	0.0676(0.1420)	0.0770	0.1589	0.0760(0.0905)
0.50	0.0529	0.1014	0.0597(0.1150)	0.0515	0.1035	0.0542(0.0647)
0.60	0.0443	0.0825	0.0589*(0.0877)	0.0387	0.0733	0.0456(0.0544)
0.70	0.0387	0.0718	0.0643(0.0812)	0.0318	0.0571	0.0442*(0.0503)
0.80	0.0342	0.0657	0.0757(0.0757)	0.0273	0.0502	0.0474(0.0474)
0.90	0.0302	0.0615	0.0914(0.0693)	0.0243	0.0476	0.0534(0.0426)
1.00	0.0264*	0.0580*	0.1111(0.0667*)	0.0219*	0.0462*	0.0660(0.0391*)

Note: $*$ stands for the minimum WMISE.

Table 2. Summary of the WMISE for the estimators in Models 3-4.

Two initial bandwidths are specified for CZ in the second row.

bandwidth	Model 3			Model 4		
h	GPL	FGK	CZ $h_0 = 0.88(1.25h)$	GPL	FGK	CZ $h_0 = 0.88(1.25h)$
0.30	0.0812	0.1686	0.0795(0.2228)	0.1057	0.2335	0.1021(0.1278)
0.40	0.0606	0.1200	0.0674*(0.1182)	0.0707	0.1261	0.0698(0.0771)
0.50	0.0500	0.0974	0.0693(0.1001)	0.0516	0.0884	0.0548(0.0609)
0.60	0.0443	0.0852	0.0830(0.0926)	0.0424	0.0700	0.0522*(0.0541*)
0.70	0.0413	0.0790	0.1090(0.0922*)	0.0381	0.0642*	0.0577(0.0543)
0.80	0.0393	0.0768	0.1440(0.0929)	0.0367	0.0665	0.0690(0.0573)
0.90	0.0373	0.0761	0.2128(0.0943)	0.0365	0.0722	0.0814(0.0575)
1.00	0.0351*	0.0753*	0.2570(0.0935)	0.0363*	0.0776	0.0988(0.0601)

Note: $*$ stands for the minimum WMISE.

Table 3. Summary of the WMISE for the estimators in Models 5-6.

Two initial bandwidths h_0 are specified for CZ in the second row.

bandwidth	Model 5			Model 6		
h	GPL	FGK	CZ $h_0 = 1.0(1.25h)$	GPL	FGK	CZ $h_0 = 1.0(1.25h)$
0.20	0.4488	13.534	7.5286(20.992)	–	–	–
0.25	0.2561*	6.8074	3.4412(5.1061)	10.076	34.514	13.913(49.969)
0.30	0.2604	4.2181	1.2381(2.1611)	0.7640	23.466	8.1067(29.580)
0.35	0.3182	2.7164	0.4213(1.4629)	0.6053*	12.494	3.1060(18.531)
0.40	0.4197	1.2760	0.2989*(0.6269)	0.7650	6.4053	1.2757(2.2496)
0.45	0.5559	1.0147	0.3075(0.3707)	1.1254	3.3545	0.6061(1.2189)
0.50	0.7187	0.8514	0.3357(0.3459*)	1.1765	2.0220	0.5896*(0.9380)
0.55	0.9018	0.7468	0.3783(0.3762)	1.4021	1.3998	0.6706(0.6774*)
0.60	1.1001	0.6826	0.4336(0.4487)	1.6296	1.0955	0.7395(0.7030)
0.65	1.3097	0.6463	0.5109(0.5569)	1.8554	0.9115	0.8715(0.8118)
0.70	1.5270	0.6341*	0.5924(0.6979)	2.0800	0.8156	1.1139(1.0025)
0.75	1.7503	0.6433	0.6999(0.8788)	2.3021	0.7777*	1.3086(1.2632)
0.80	1.9768	0.6743	0.8073(1.0984)	2.5218	0.7936	1.5003(1.6218)

Note: $*$ stands for the minimum WMISE.

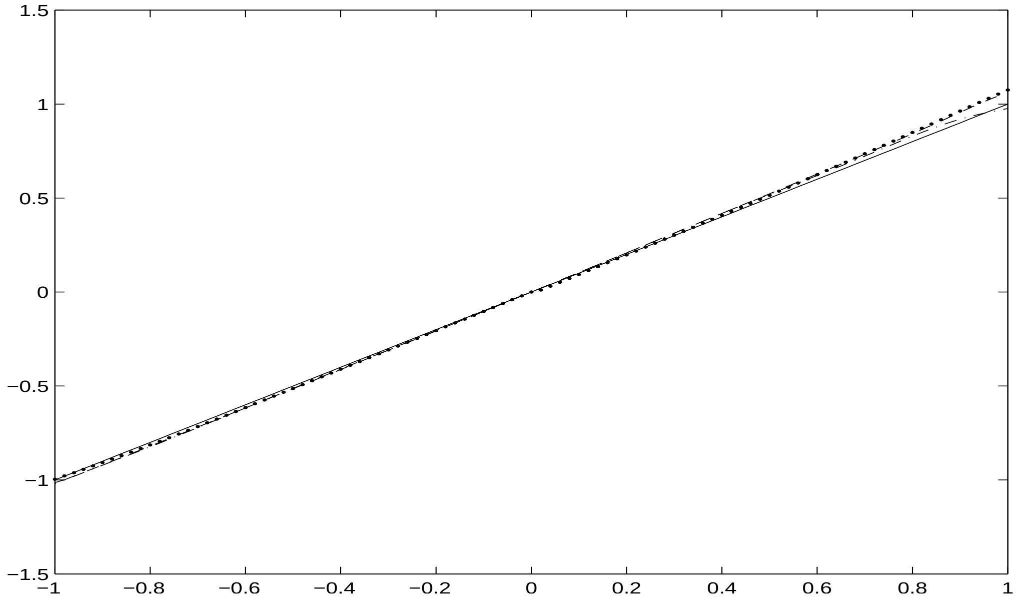

Figure 1. Biases of the estimators in Model 1. We take $h_{GPL} = h_{FGK} = h_{CZ} = 1.0$ ($h_0 = 1.25$). The solid curve is the true function $\psi(x)$, the dashed curve is based on GPL, the dotted curve is based on FGK, and the dash-dotted line is based on CZ.

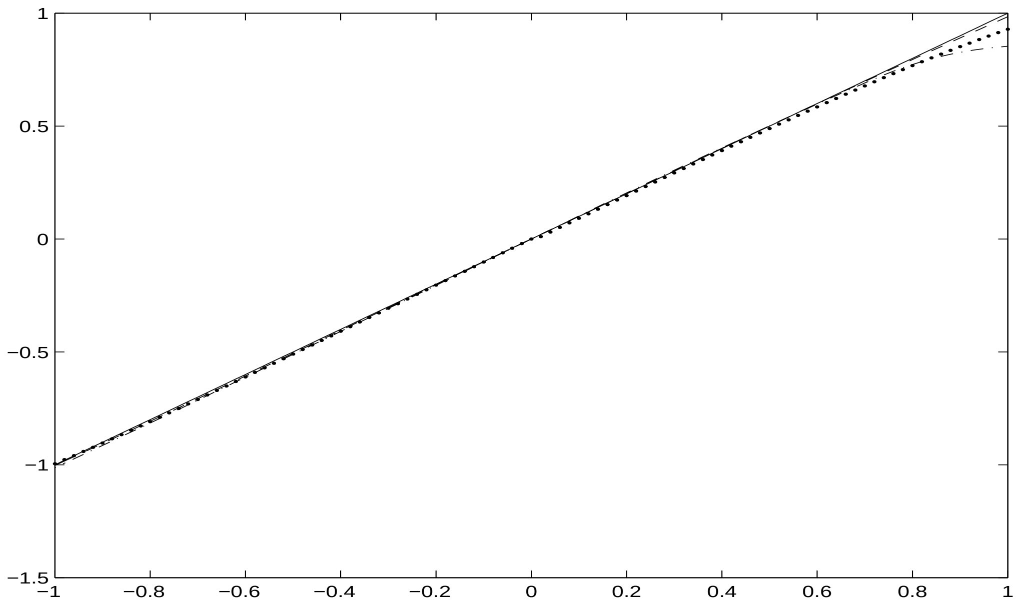

Figure 2. Biases of the estimators in Model 2. We take $h_{GPL} = h_{FGK} = h_{CZ} = 1.0$ ($h_0 = 1.25$). The solid curve is the true function $\psi(x)$, the dashed curve is based on GPL, the dotted curve is based on FGK, and the dash-dotted line is based on CZ.

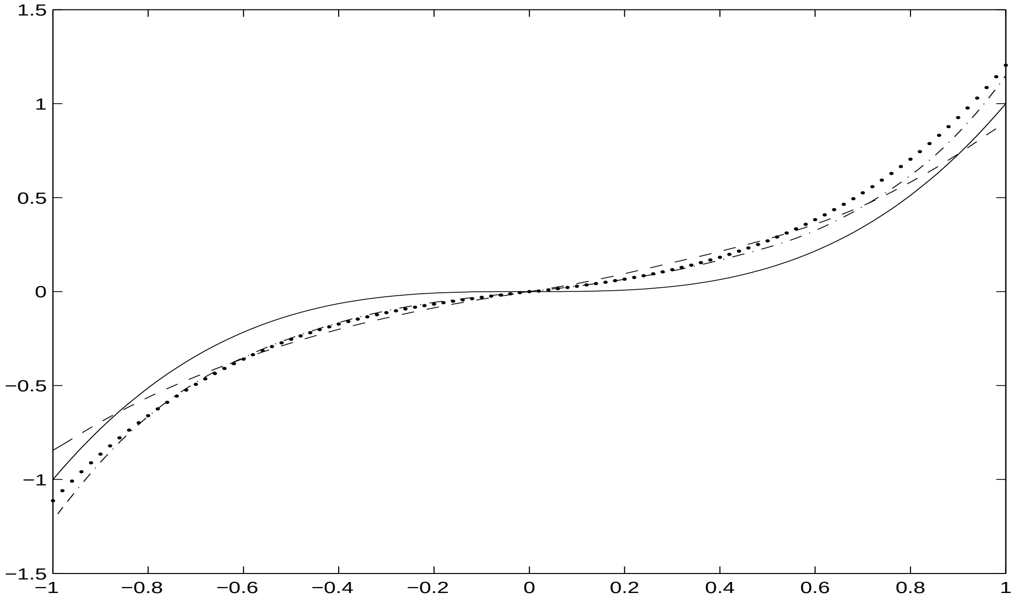

Figure 3. Biases of the estimators in Model 3. We take $h_{GPL} = h_{FGK} = 0.70, h_{CZ} = 0.40$ ($h_0 = 0.8$). The solid curve is the true function $\psi(x)$, the dashed curve is based on GPL, the dotted curve is based on FGK, and the dash-dotted line is based on CZ.

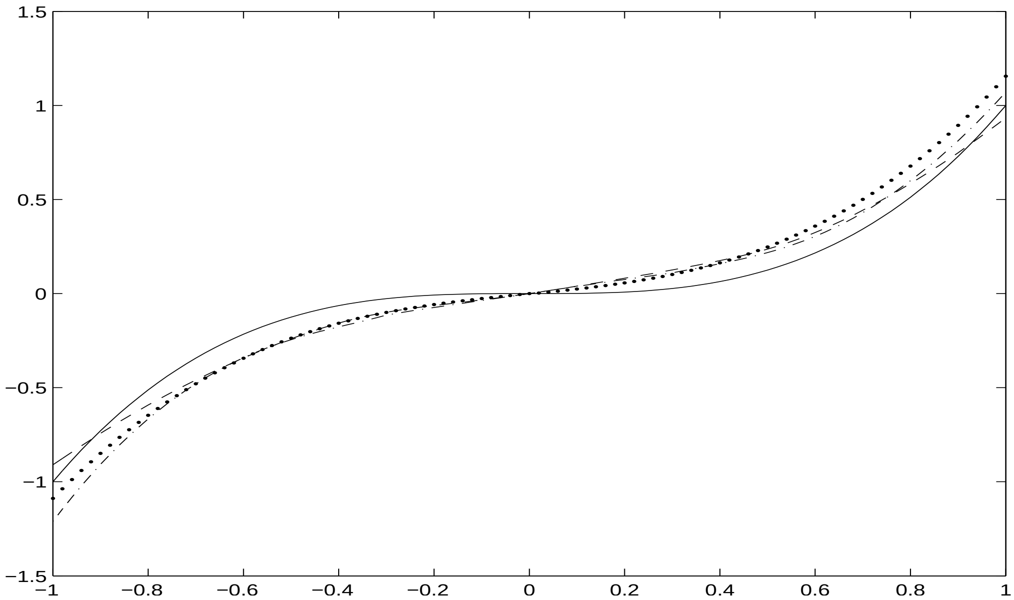

Figure 4. Biases of the estimators in Model 4. We take $h_{GPL} = h_{FGK} = 0.70, h_{CZ} = 0.60$ ($h_0 = 0.8$). The solid curve is the true function $\psi(x)$, the dashed curve is based on GPL, the dotted curve is based on FGK, and the dash-dotted line is based on CZ.

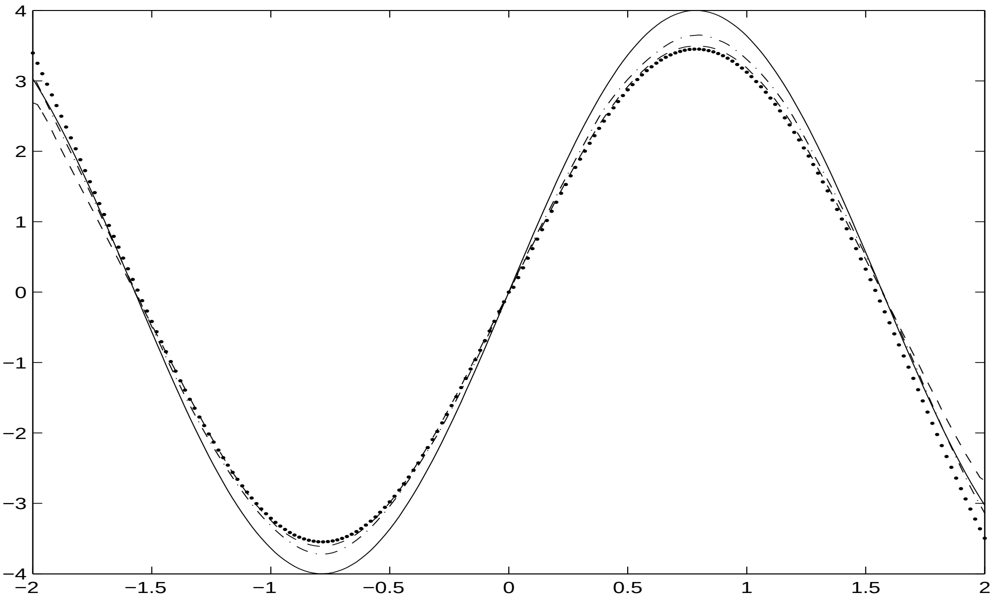

Figure 5. Biases of estimators in Model 5. We take $h_{GPL} = 0.25, h_{FGK} = 0.70, h_{CZ} = 0.40$ ($h_0 = 0.70$). The solid curve is the true function $\psi(x)$, the dashed curve is based on GPL, the dotted curve is based on FGK, and the dash-dotted line is based on CZ.

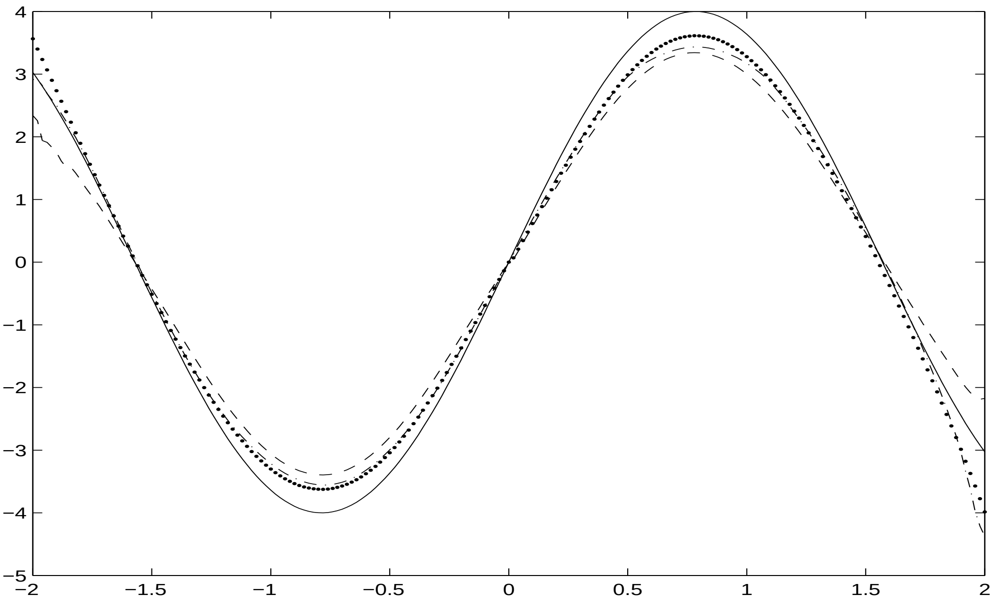

Figure 6. Biases of the estimators in Model 6. We take $h_{GPL} = 0.35, h_{FGK} = 0.75, h_{CZ} = 0.50$ ($h_0 = 0.70$). The solid curve is the true function $\psi(x)$, the dashed curve is based on GPL, the dotted curve is based on FGK, and the dash-dotted line is based on CZ.

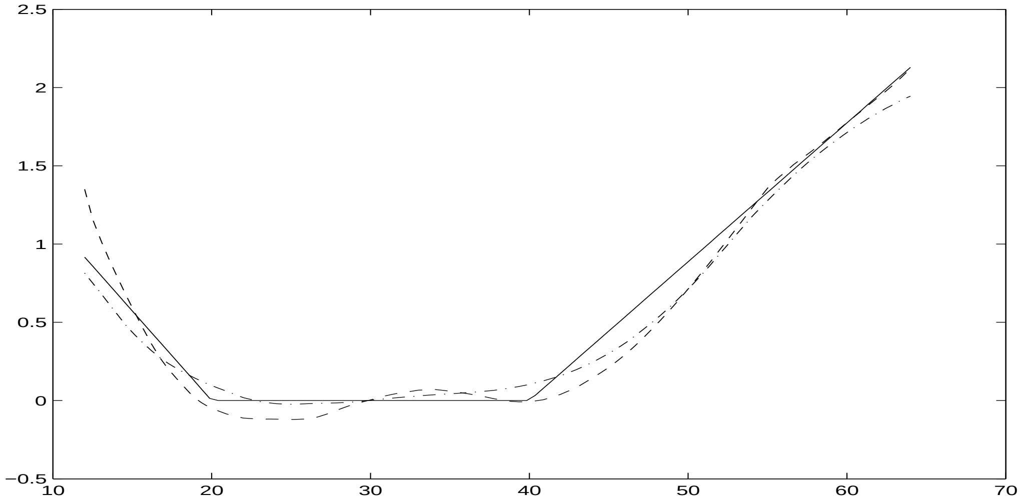

Figure 7. Estimates of $\psi(\cdot)$ for the Stanford Heart Transplant Data (the dashed curves is based on GPL with bandwith h=7, the dashed-dotted curve is based on GPL with bandwith h=10, and the solid curve is based on partial likelihood with a piecewise linear function joined at age 20 and 40 and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

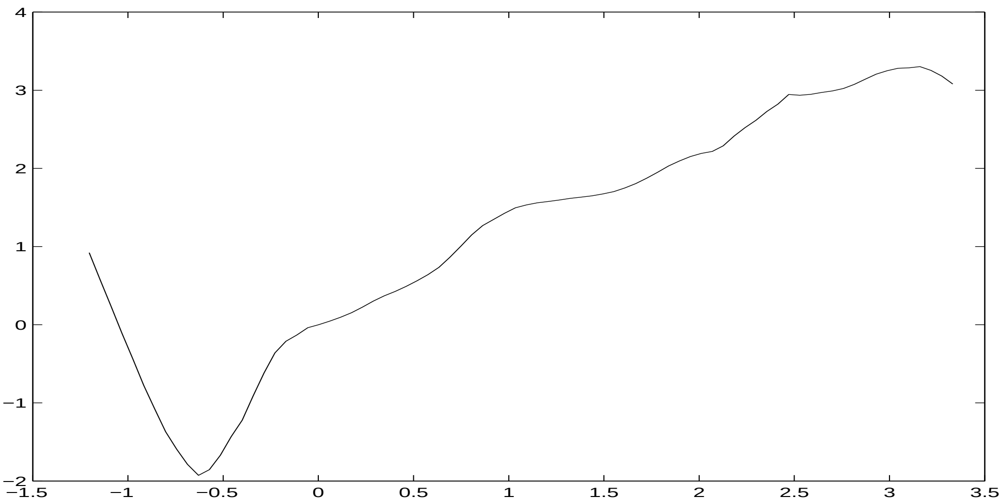

Figure 8. Estimate of $\psi(\cdot)$ for the PBC data.