Harvard University

Harvard University Biostatistics Working Paper Series

Year 2008 *Paper* 83

Nonparametric Regression Using Local Kernel Estimating Equations for Correlated Failure Time Data

Zhangsheng Yu* Xihong Lin†

*University of Michigan, zyuz@umich.edu

†Harvard School of Public Health, xlin@hsph.harvard.edu

Nonparametric Regression Using Local Kernel Estimating Equations for Correlated Failure Time Data

By Zhangsheng Yu
Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.
zyuz@umich.edu
AND
Xihong Lin
Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, U.S.A.
xlin@hsph.harvard.edu

August 22, 2006

Summary

We study nonparametric regression for correlated failure time data. Kernel estimating equations are used to estimate nonparametric covariate effects. Independent and weighted kernel estimating equations are studied. The derivative of the nonparametric function is first estimated and the nonparametric function is then estimated by integrating the derivative estimator. We show that the nonparametric kernel estimator is consistent for any arbitrary working correlation matrix and its asymptotic variance is minimized by assuming working independence. We evaluate the performance of the proposed kernel estimator using simulation studies, and apply the proposed method to the western Kenya parasitemia data.

Some key words: Asymptotic bias and variance, Clustered survival data, Efficiency, Estimating equation, Kernel smoothing, Marginal model, Sandwich Estimator.

1 Introduction

Correlated failure time data arise frequently in health sciences research, such as familial data and recurrent event data. Statistical research for modeling correlated failure time data has been mainly focused on regression with parametric covariate effects in the last two decades, e.g., see Wei, Lin, Weissfeld (1989), Cai and Prentice (1995), Gray and Li (2002), among others. The latter authors found that accounting for the within-cluster correlation improves the efficiency of parameter estimation. In this paper, we refer such a model with parametric covariate effects as a parametric model even though a nonparametric baseline hazard is assumed. For a comprehensive review of parametric modeling of correlated failure time data, see Kalbfleisch and Prentice (2002, Ch 10).

In practice, such a parametric assumption might not always be desirable, since some covariate effects might be complicated and their functional forms might not be known in advance. A motivating example is the western Kenya Parasitemia study (McElroy, et al, 1997). Parasitemia is an indicator for potential malaria. This study involved 542 children from 309 households and followed them over time for the occurrence of parasitemia. A scientific question of interest is to investigate the baseline age effect on the onset of parasitemia. However, the baseline age effect appears nonlinear and somewhat complicated (see Figure 3). Analysis of this data set is hence challenged by the fact that the survival outcomes from the children within the same family are likely to be correlated and it is desirable to model the baseline age effect nonparametrically.

Considerable work has been done on nonparametric regression for univariate survival data using kernel and spline methods. Tibshirani and Hastie (1987) proposed kernel smoothing using the local partial likelihood. Fan, et al. (1997) studied the theoretical properties of the local kernel estimator. Dabroska (1987) and Li et al. (1995) developed a two-dimensional local nonparametric kernel estimator of time and covariates without posing a proportional hazard assumption. Hastie and Tibshirani (1990), O'Sullivan(1988) and Gray (1992) developed smoothing spline methods in proportional hazard models. However, there is little literature on nonparametric regression for multivariate failure time data.

In this paper, we consider nonparametric regression estimation of a single covariate for censored multivariate failure time data. We assume a marginal proportional hazard model and propose local polynomial kernel estimating equations. We consider both working independence and weighted kernel estimating equations. These extend parametric estimating equations of Wei, et al. (1989), Lee, et al. (1992), and Cai and Prentice (1995, 1997) by introducing local polynomial kernel regression techniques. We derive the asymptotic bias and variance of the kernel estimator, and show that the most efficient kernel estimator using weighted kernel estimating equations is obtained by ignoring the within-cluster correlation. This result is significantly different from those in parametric regression where accounting for correlation improves efficiency (Cai and Prentice, 1995, 1997; Gray and Li, 2002). However, this result is consistent with the kernel smoothing results of Lin and Carroll (2000) in uncensored longitudinal data, where they found the most efficient kernel GEE estimator is obtained by ignoring the within-cluster correlation. Unlike longitudinal data, one has to deal with censoring in censored multivariate failure time data. Further, a unique difficulty in kernel smoothing for censored multivariate failure time data under marginal proportional hazard models is that the nonparametric function is not directly estimable from the kernel estimating equations in the presence of the unspecified baseline hazard. We hence use the kernel estimating equations to first estimate the derivative of the nonparametric function and then construct the estimator of the nonparametric function by integrating the estimator of the derivative.

The remaining of this paper is organized as following. We introduce the nonparametric covariate model in §2 and propose working independence and weighted local polynomial kernel estimating equations for multivariate failure time data assuming a common baseline hazard in §3. We study the asymptotic properties of the proposed kernel estimators in §4. We extend the results to allow for different baseline hazards in §5. We evaluate the performance of the proposed method using a simulation study in §6 and apply it to the western Kenya parasitemia data in §7, with conclusions in §8.

2 The Nonparametric Marginal Model for Multivariate Failure Time Data

Let T_{ij} and C_{ij} be the underlying failure and censoring times of the jth observation in the ith cluster ($j = 1, \cdots, J_i$, $i = 1, \cdots, n$). We assume the cluster size $J_i < \infty$ as the number of clusters n goes to infinity. The observed data are $(X_{ij}, \delta_{ij}, Z_{ij})$, where X_{ij} is the observed event time, i.e., $X_{ij} = \min(T_{ij}, C_{ij})$; $\delta_{ij} = I(T_{ij} \leq C_{ij})$ is a censoring indicator; Z_{ij} is a scalar time-independent covariate. Let $Y_{ij}(s) = I(X_{ij} \geq s)$ be the at-risk process. We assume that T_{ij} might be correlated within the same cluster. Observations from different clusters are assumed to be independent. The $\{T_{ij}\}_{j=1\cdots J_i}$ are assumed to be independent of the $\{C_{ij}\}_{j=1\cdots J_i}$ conditioned on the $\{Z_{ij}\}_{j=1\cdots J_i}$, i.e., independent censoring.

We assume T_{ij} follows a marginal proportional hazard model with the effect of Z_{ij} modeled nonparametrically as,

$$\lambda_{ij}(t) = \lambda_0(t) exp\{\theta(Z_{ij})\}, \tag{1}$$

where $\lambda_{ij}(t)$ is the hazard for the jth observation in the ith cluster, $\lambda_0(t)$ is an unspecified baseline hazard, and $\theta(z)$ is an unknown smooth nonparametric function to model the effect of the covariate Z. We propose estimating $\theta(z)$ using working independent and weighted local polynomial kernel estimating equations.

3 The Local Polynomial Kernel Estimators of $\theta(z)$

3.1 The Local Pseudo Partial Likelihood Kernel Estimator

We propose in this section a kernel estimator of $\theta(z)$ in model (1) for multivariate failure time data by maximizing a local pseudo partial likelihood. To estimate $\theta(z)$ at a target point z, local polynomial kernel regression techniques approximate $\theta(Z)$ for any Z in the neighborhood of z by a pth order polynomial as

$$\theta(Z) \approx \beta_0 + Z(z)^T\beta = \beta_0 + \beta_1(Z - z) + \cdots + \beta_p(Z - z)^p,$$

where $\beta_j = \theta^{(j)}(z)/j!$, $Z(z) = \{(Z - z), \cdots, (Z - z)^p\}^T$, and $\beta = (\beta_1, \cdots, \beta_p)^T$. Then the local pth order polynomial kernel estimator of $\theta(z)$ is $\widehat{\theta}(z) = \widehat{\beta}_0$.

Denote by h a bandwidth and $K(\cdot)$ a symmetric kernel density function which, without loss of generality, has mean 0 and variance 1. We propose the following local pseudo partial likelihood for kernel estimation of $\theta(z)$

$$\begin{aligned} &\ell_I(\beta, z) \qquad\qquad (2)\\ = &\ \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J_i} K_h(Z_{ij}-z)\delta_{ij}\left[Z_{ij}(z)^T\beta - log\left\{\sum_{i'=1}^{n}\sum_{l=1}^{J_j} K_h(Z_{i'l}-z)Y_{i'l}(X_{ij})e^{Z_{i'l}(z)^T\beta}\right\}\right], \end{aligned}$$

where $K_h(s) = h^{-1}K(s/h)$. The local pseudo partial likelihood (2) ignores the within-cluster correlation, and can be viewed as a nonparametric kernel extension of the parametric pseudo partial likelihood of Lee, et al. (1992) for multivariate failure time data. For univariate censored survival data ($J_i = 1$), equation (2) reduces to the local partial likelihood of Fan, et al. (1997).

In traditional local polynomial kernel smoothing, $\theta(z)$ is estimated by $\widehat{\theta}(z) = \widehat{\beta}_0$ by maximizing the local loglikelihood. However, the intercept β_0 is not directly estimable from the local pseudo partial likelihood (2) since it is canceled out. This is in the same spirit of the partial likelihood in parametric regression, where the baseline hazard $\lambda_0(t)$ and the intercept are eliminated. Denote by $\widehat{\beta} = (\widehat{\beta}_1, \cdots, \widehat{\beta}_p)^T$ the maximum local partial likelihood estimator obtained by maximizing (2). Since $\beta_1 = \theta^{(1)}(z)$, we can estimate $\widehat{\theta}_I^{(1)}(z) = \widehat{\beta}_1$ directly from (2). Note that $\theta(z)$ is identifiable up to a constant in model (1). We hence impose the identifiability constraint $\theta(a) = 0$ for some constant a. It follows that we can estimate $\widehat{\theta}(z)$ by $\widehat{\theta}_I(z) = \int_a^z \widehat{\theta}_I^{(1)}(s)ds$. The Trapezoidal rule can be used to approximate the integral. A similar approach was used by Tibshirani and Hastie (1987) and Fan, et al. (1997) in kernel smoothing for univariate censored survival data. Note that $\theta^{(1)}(z)$ instead of $\theta(z)$ measures the covariate effect at z. To see this, consider the parametric model $\theta(z) = \beta_1 z$. Then $\theta^{(1)}(z) = \beta_1$, which captures the effect of Z. Hence one can test $H_0 : \theta^{(1)}(z) = 0$ for the effect of Z. This argues estimation of the derivative $\theta^{(1)}(z)$ is of more practical interest.

Differentiation of (2) with respect to β gives the local pseudo partial score equation of β as

$$
\begin{aligned}
0 &= U_I(\beta, z) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{J_i} U_{I,ij}(\beta, z) \qquad (3) \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{J_i} K_h(Z_{ij} - z)\delta_{ij} \left\{ Z_{ij}(z) - \frac{\sum_{i'=1}^{n} \sum_{l=1}^{J_j} K_h(Z_{i'l} - z) Y_{i'l}(X_{ij}) Z_{i'l} e^{Z_{i'l}(z)^T \beta}}{\sum_{i'=1}^{n} \sum_{l=1}^{J_j} K_h(Z_{i'l} - z) Y_{i'l}(X_{ij}) e^{Z_{i'l}(z)^T \beta}} \right\}.
\end{aligned}
$$

We term this estimating equation as the working independence kernel estimating equation, and the resulting kernel estimators $\widehat{\theta}_I^{(1)}(z)$ and $\widehat{\theta}_I(z)$ as the working independence kernel estimator, reflecting the fact that the local pseudo partial likelihood ignores the within-cluster correlation. The theoretical properties of $\widehat{\theta}_I^{(1)}(z)$ and $\widehat{\theta}_I(z)$ are discussed in §4.

Equation (3) can be solved using the Newton-Raphson algorithm. The covariance of $\widehat{\beta}_I$ can be estimated using the sandwich estimator $\widehat{V}_I(\widehat{\beta}) = \Omega_1^{-1}\Omega_2\Omega_1^{-1}$, where $\Omega_1 = \partial U_I(\beta, z)/\partial \beta^T\big|_{\beta=\widehat{\beta}_I}$ and $\Omega_2 = \sum_{i=1}^{n}\{\sum_{j=1}^{J_i} U_{I,ij}(\widehat{\beta}_I, z)\}^{\otimes 2}$ and $\otimes$ denotes the outer product, i.e., $A^{\otimes 2} = AA^T$. It follows that $\text{var}\{\widehat{\theta}_I^{(1)}(z)\} = \Delta_1^T \widehat{V}_I(\widehat{\beta}_I)\Delta_1$, where $\Delta_1 = (1, 0, \cdots, 0)^T$. Since calculation of $\widehat{\theta}_I(z)$ using numerical integration involves a weighted sum of $\{\theta_I^{(1)}(z_j)\}$ at a set of grid points $\{z_j\}$, estimation of the variance $\widehat{\theta}_I(z)$ requires the covariance estimators of $\{\theta_I^{(1)}(z_j)\}$ at the grid points $\{z_j\}$ and is complicated. A bootstrap method can be used.

3.2 The Weighted Local Polynomial Kernel Estimator

The local pseudo partial likelihood method in §3.1 ignores the within-cluster correlation. For parametric regression, to improve efficiency, Cai and Prentice (1995, 1997) proposed a set of weighted estimating equations by extending the working independence estimating equations of Wei, et al. (1989) and Lee, et al. (1992) by incorporating a working correlation matrix in a similar fashion to the GEE method of Liang and Zeger (1986) used in longitudinal data. They found that the resulting weighted parameter estimators are more efficient than the working independence estimator when the correlation is strong. It is of interest to explore whether an introduction of such a working correlation matrix in the working independence kernel estimating equation (3) could improve the efficiency of the kernel estimator of $\theta^{(1)}(z)$ and $\theta(z)$.

To proceed, rewrite the working independent kernel estimating equation (3) using the Martingale notation. Specifically, Let $N_{ij}(t) = \delta_{ij} I(X_{ij} \le t)$, and

$$M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(s)\lambda_0(s)e^{\theta(Z_{ij})}ds, \qquad (4)$$

where $M_{ij}(t)$ is a martingale with respect to the filtration $\mathcal{F}_{t,j} = \vee_{i=1}^n \mathcal{F}_{t,ij}$ when $\theta(Z_{ij})$ equals to the true value, and $\mathcal{F}_{t,ij} = \sigma\{N_{ij}(s), Y_{ij}(s), X_{ij}(s); 0 \le s \le t\}$, but not with respect to the joint filtration $\mathcal{F}_t = \vee_{i=1}^n \vee_{j=1}^{J_i} \mathcal{F}_{t,ij}$. The working independence kernel estimating equation (3) can be rewritten as

$$U_I(\beta, z) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau Z_i(z) K_{ih} d\widetilde{M}_i(s),$$

where τ is the length of study followup, $Z_i(z) = \{Z_{i1}(z), \cdots, Z_{iJ_i}(z)\}^T$, $K_{ih} = diag\{K_h(Z_{i1} - z), \cdots, K_h(Z_{iJ_i} - z)\}$, $\widetilde{M}_i(t) = \{\widetilde{M}_{i1}(t), \cdots, \widetilde{M}_{iJ_i}(t)\}^T$, and

$$\widetilde{M}_{ij}(t) = N_{ij}(t) - \int_0^t \frac{Y_{ij}(s)e^{Z_{ij}(z)^T\beta}\sum_{i'=1}^n\sum_{l=1}^{J_i} K_h(Z_{i'l} - z)dN_{i'l}(s)}{\sum_{i'=1}^n\sum_{l=1}^{J_i} K_h(Z_{i'l} - z)Y_{i'l}(s)e^{Z_{i'l}(z)^T\beta}}.$$

We now propose the weighted local partial likelihood kernel estimating equation

$$\begin{aligned} 0 &= = U_W(\beta, z) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau Z_i(z) K_{ih}^{1/2} W_i K_{ih}^{1/2} d\widetilde{M}_i(s) \qquad (5) \\ &= \frac{1}{n}\sum_{i=1}^n \sum_{l=1}^{J_i} \int_0^\tau \sum_{j=1}^{J_i} Z_{ij}(z) Q_{ijl}(z) dN_{il}(s) \\ &- \frac{1}{n}\sum_{i=1}^n\sum_{l=1}^{J_i}\int_0^\tau \frac{\{\sum_{i'=1}^n\sum_{j=1}^{J_i}\sum_{r=1}^{J_i} Y_{i'r}(s)Z_{i'j}(z)Q_{i'jr}(z)e^{Z_{i'r}(z)^T\beta}\}}{\sum_{i'=1}^n\sum_{r=1}^{J_i} Y_{i'r}(s)e^{Z_{i'r}(z)^T\beta}K_h(Z_{i'r} - z)} K_h(Z_{il} - z)dN_{il}(s), \end{aligned}$$

where $Q_{ijl}(z) = K_h^{1/2}(Z_{ij} - z) w_{jl}^i K_h^{1/2}(Z_{il} - z)$, and W_i is a $p \times p$ bounded weight matrix that attempts to account for the within-cluster correlation, w_{jl}^i is the (j, l)th element of W_i that might depend on Z_i and a correlation parameter vector ϕ. Following Cai and Prentice (1995), one could set W_i to be the inverse of the correlation matrix of the martingale $\{M_{ij}(X_{ij})\}_{j=1,\cdots,J_i}$. If W_i is an identity matrix I, the weighted kernel estimating equation (5) reduces to the working independence kernel estimating equation (3). Moment estimators of ϕ can be constructed using a similar method to that of Prentice and Cai (1992).

Denote by $\widehat{\beta}_W$ the solution of the weighted local kernel estimating equation (5). The weighted local kernel estimator of $\theta^{(1)}(z)$ is $\widehat{\theta}_W^{(1)}(z) = \widehat{\beta}_{W,1}$ and the weighted local kernel estimator of $\theta(z)$ is $\widehat{\theta}_W(z) = \int_a^z \widehat{\theta}_W^{(1)}(s)ds$. One can approximate the integral using a similar integration technique described in §3.1. The variance of $\widehat{\theta}_W^{(1)}(z)$ can be estimated by a sandwich estimator in the same way as that of the working independent estimator $\widehat{\theta}_I^{(1)}(z)$ except that $U_W(\beta, z)$ is used in calculating Ω_1 and Ω_2. The asymptotic properties of $\widehat{\theta}_W^{(1)}(z)$ are studied in §4. Note that in this case, the marginal filtration is modified to be $\mathcal{F}_{t,j} = \vee_{i=1}^n \mathcal{F}_{t,ij}$, where $\mathcal{F}_{t,ij} = \sigma\{N_{ij}(s), Y_{ij}(s); 0 \le s \le t\} \vee \{N_{il}(0), Y_{il}(0), X_{il}(0), l = 1, \cdots, J_i\}$. A bootstrap method can be used to calculate the variance of $\widehat{\theta}_W(z)$. Note that the confidence interval of $\widehat{\theta}_W(z)$ becomes wider as z increases over its support due to the accumulation of errors when one integrates the estimator over $[0, z]$ using a summation of $\widehat{\theta}^{(1)}(z)$ at grid points.

4 Asymptotic Properties of the Kernel Estimators

As discussed in §3, $\theta^{(1)}(z)$ is estimable directly from the kernel estimating equations (3) and (5) and is more useful to measure the effect of Z, e.g., $\theta^{(1)}(z) = 0$ indicates no effect of Z. We hence focus our asymptotic investigation on the estimators of $\theta^{(1)}(z)$. Specifically, we study in this section the asymptotic properties of the local pseudo partial likelihood based working independence kernel estimator $\widehat{\theta}_I^{(1)}(z)$ proposed in §3.1 and the weighted local polynomial kernel estimator $\widehat{\theta}_W^{(1)}(z)$ proposed in §3.2.

Although the working independence kernel estimating equation (3) is a special case of the weighted kernel estimating equation (5) when the weight matrix is an identity matrix, its asymptotic properties can be obtained through the likelihood analysis with less strict conditions. We hence present the asymptotic results of the two estimators separately. Our major findings are that, in contrast to the parametric regression results of Cai and Prentice (1995, 1997), the most efficient local polynomial kernel estimator is obtained by ignoring the within-cluster correlation and assuming working independence in the weighted local polynomial kernel estimating equation (5), i.e., by setting $W_i = I$.

4.1 The Asymptotic Properties of the Working Independence Kernel Estimator

We study in this section the asymptotic properties of the working independence local kernel estimator $\widehat{\theta}_I^{(1)}(z)$. Without loss of generality, we assume in our asymptotic analysis $J_i = J$. For simplicity, we suppress the subscript i. Let $H = diag(h, \cdots, h^p)$, $\tilde{u} = (u, \cdots, u^p)^T$, $\tilde{\nu}_r = \int \tilde{u} K^r(u) du$ $(r = 1, 2)$, $\bar{P}_j(t|z) = P(X_j > t | Z_j = z)$ and $\bar{\Lambda}_j(t, z) = \int_0^t \bar{P}_j(s|z)\lambda_0(s) ds$ for $j = 1, \cdots, J$. Assuming the kernel function $K(u)$ is symmetric with mean 0 and variance 1, Let β^0 be the true value of β, where $\beta = (\beta_1, \cdots, \beta_p)^T$ and $\beta_l = \theta^{(l)}(z)/l!$ $(l = 1, \cdots, p)$. Let $\theta_0(z)$ be the true function of $\theta(z)$. Theorem 1 gives the asymptotic normality results of the working independence kernel estimator $\widehat{\beta}_I$. We assume the number of clusters $n \to \infty$ while the cluster size J is finite, and $h \to 0$ and $nh \to \infty$.

Theorem 1 *Under the regularity conditions A in Appendix A, the working independence local kernel estimator $\widehat{\beta}_I$ obtained by maximizing the local pseudo partial likelihood (2) has the following asymptotic properties:*

(1) $H(\widehat{\beta}_I - \beta^0)$ converges in probability to 0.

(2) The asymptotic distribution of $\widehat{\beta}_I$ satisfies

$$\sqrt{nh}\left\{H(\widehat{\beta}_I - \beta^0) - \frac{\theta_0^{(p+1)}(z)}{(p+1)!} D^{-1} c h^{p+1}\right\} \to N\{0, V_I(z)\}$$

in distribution, where $c = \int K(u) u^{p+1}(\tilde{u} - \tilde{\nu}_1) du$, $D = \int \tilde{u}\tilde{u}^T K(u) du - \tilde{\nu}_1 \tilde{\nu}_1^T$, and $B = \int K^2(u)(\tilde{u} - \tilde{\nu}_1)(\tilde{u} - \tilde{\nu}_1)^T du$, $f_j(z)$ is the density of Z_j, and

$$V_I(z) = \frac{D^{-1} B D^{-1}}{\sum_{j=1}^J f_j(z) e^{\theta_0(z)} \bar{\Lambda}_j(\tau, z)}.$$

The proof of Theorem 1 is given in Appendix B. Theorem 1 gives the joint asymptotic distribution of the working independence kernel estimators of the derivatives $\{\widehat{\theta}_I^{(1)}(z), \cdots, \widehat{\theta}_I^{(p)}(z)\}$. One can easily obtain from Theorem 1 the asymptotic bias and variance of $\widehat{\theta}^{(1)}(z)$, which are given in Corollary 1.

Corollary 1 *Under the regularity conditions of Theorem 1, The asymptotic properties of $\widehat{\theta}_I^{(1)}(z)$ are as follows.*

(a) The asymptotic bias of $\widehat{\theta}_I^{(1)}(z)$ is

$$E\{\widehat{\theta}_I^{(1)}(z)\} - \theta_0^{(1)}(z) = \frac{h^p}{(p+1)!}\theta_0^{(p+1)}(z)\Delta_1^T D^{-1}c + o(h^p),$$

where $\Delta_1 = (1, 0, \cdots, 0)^T$, e.g., when $p = 2$,

$$E\{\widehat{\theta}_I^{(1)}(z)\} - \theta_0^{(1)}(z) = \frac{h^2}{6}\theta_0^{(3)}(z)\int u^4 K(u)du$$

(b) The asymptotic variance of $\widehat{\theta}_I^{(1)}(z)$ is

$$\text{var}\{\widehat{\theta}_I^{(1)}(z)\} = \frac{1}{nh^3}\frac{\int u^2 K^2(u)du}{\sum_{j=1}^J f_j(z)e^{\theta_0(z)}\bar{\Lambda}_j(\tau, z)}.$$

When $J = 1$, one can easily show that the results in Theorem 1 and Corollary 1 reduce to those in Fan et al.(1997) by using their equation (1.3)

$$e^{\theta(z)} = \frac{E\{\delta|Z = z\}}{E\{\int_0^X \lambda_0(s)ds|Z = z\}}.$$

4.2 Asymptotic Properties of the Weighted Local Kernel Estimator

In this subsection, we study the asymptotic properties of the weighted local kernel estimator $\widehat{\theta}_W^{(1)}(z)$ that solves the weighted kernel estimating equation (5), which incorporates the within-cluster correlation in the weight matrix. The asymptotic properties of $\widehat{\beta}_W$ are given in Theorem 2.

Theorem 2 *Under the conditions A and B, the weighted local kernel estimator $\widehat{\beta}_W$ has the following asymptotic properties:*

(a) $H(\widehat{\beta}_W - \beta^0)$ converges to 0 in probability.

(b) The asymptotic distribution of $\widehat{\beta}_W$ satisfies

$$\sqrt{nh}\left\{H(\widehat{\beta}_W - \beta^0) - \frac{\theta_0^{(p+1)}(z)}{(p+1)!}D^{-1}ch^{p+1}\right\} \to N\{0, V_W(z)\}$$

in distribution, where D and c were defined in Theorem 1, and

$$V_W(z) = e^{-\theta_0(z)}\left\{\sum_{j=1}^{J} f_j(z)\bar{w}_{jj}(z,z)\bar{\Lambda}_j(\tau,z)\right\}^{-2} \quad (6)$$

$$\times\left[\sum_{j=1}^{J} f_j(z)\bar{w}_{jj}^2(z,z)\bar{\Lambda}_j(\tau,z)A_1 + \left\{\int_0^\tau \frac{(\sum_{j=1}^J f_j(z)\bar{w}_{jj}(z,z)\bar{P}_j(s|z))^2}{\sum_{j=1}^J f_j(z)\bar{P}_j(s|z)}\lambda_0(s)ds\right\} A_2\right],$$

and $\bar{w}_{jj}(z,z) = E\{w_{jj}(Z)|Z_j = z\}$, $A_1 = D^{-1}\{\int \tilde{u}\tilde{u}^T K^2(u)du\}D^{-1}$, and

$$A_2 = D^{-1}\Big\{ - \tilde{\nu}_1\tilde{\nu}_2^T - \tilde{\nu}_2\tilde{\nu}_1^T + \tilde{\nu}_1\tilde{\nu}_1^T \int K^2(u)du\Big\}D^{-1}.$$

The proof of Theorem 2 is given in Appendix C. One can easily derive from Theorem 2 the asymptotic bias and variance of $\widehat{\theta}_W^{(1)}(z)$, and the results are given in Corollary 2.

Corollary 2 *Under conditions A and B in appendix A, $\widehat{\theta}_W^{(1)}(z)$ has the following asymptotic properties:*

(a) The asymptotic bias of $\widehat{\theta}_W^{(1)}(z)$ is

$$E\{\widehat{\theta}_W^{(1)}(z)\} - \theta_0^{(1)}(z) = \frac{h^p}{(p+1)!}\theta_0^{(p+1)}(z)\Delta_1^T D^{-1}c + o(h^p),$$

where $\Delta_1 = (1, 0, \cdots, 0)^T$, e.g., when $p = 2$,

$$E\{\widehat{\theta}_W^{(1)}(z)\} - \theta_0^{(1)}(z) = \frac{h^2}{6}\theta_0^{(3)}(z)\int u^4K(u)du$$

(b) The asymptotic variance of $\widehat{\theta}_W^{(1)}(z)$ is

$$\text{var}\{\widehat{\theta}_W^{(1)}(z)\} = \frac{1}{nh^3}R(z)\int u^2K^2(u)du, \quad (7)$$

where

$$R(z) = \frac{\sum_{j=1}^J \bar{\Lambda}_j(\tau,z)\bar{w}_{jj}(z,z)^2 f_j(z)}{e^{\theta_0(z)}\left\{\sum_{j=1}^J \bar{\Lambda}_j(\tau,z)\bar{w}_{jj}(z,z)f_j(z)\right\}^2}. \quad (8)$$

(c) The optimal bandwidth h that minimizes the asymptotic weighted integrated mean squared error is

$$h_{opt} = \Big[\frac{1}{n}\frac{3\int R(z)q(z)dz \int u^2K^2(u)du}{2p\int\{b_0(z)\}^2q(z)dz}\Big]^{\frac{1}{2p+3}},$$

where $b_0(z) = \theta_0^{(p+1)}(z)\Delta_1^T D^{-1}c/(p+1)!$*, and* $q(z)$ *is some weight function used in the calculating the integrated mean square error.*

The results in Corollary 2 show that the asymptotic bias of the weighted local kernel estimator $\widehat{\theta}_W^{(1)}(z)$ is the same as that of the working independence local kernel estimator $\widehat{\theta}_I^{(1)}(z)$, and does not depend on the weight matrix W. By examining the variance of $\widehat{\theta}_W^{(1)}(z)$, we are interested in identifying an optimal weight matrix W that gives the most efficient weighted kernel estimator $\widehat{\theta}_W^{(1)}(z)$ by minimizing $\text{var}\{\widehat{\theta}_W^{(1)}(z)\}$. Theorem 3 states our main result.

Theorem 3 *The asymptotic variance of* $\widehat{\theta}_W^{(1)}(z)$ *in (7) is minimized by* $W = I$*, i.e., by assuming working independence.*

The proof of Theorem 3 is straightforward by directly applying the Cauchy-Schwartz inequality to $R(z)$ in (8), which is minimized when $\bar{w}_{j,j}(z,z)^2 = \bar{w}_{j,j}(z,z)$, i.e., $\bar{w}_{j,j}(z,z) = 1$. An identity matrix satisfies this condition. Note that the minimizer of $\text{var}\{\widehat{\theta}_W^{(1)}(z)\}$ is not necessarily unique.

The result in Theorem 3 shows the most efficient weighted local kernel estimator for multivariate survival data is obtained by ignoring the within-cluster correlation. This result is significantly different from that of Cai and Prentice (1997) and Gray and Li (2002) in parametric regression, where the regression coefficient estimator accounting for the within-cluster correlation in the weight matrix W is more efficient than the working independent estimator. However, our result is consistent with that of Lin and Carroll (2000), who showed that the working independence local kernel estimator is most efficient in nonparametric regression for uncensored longitudinal data within the GEE kernel estimating equation framework.

An intuitive explanation of this seemingly "counter-intuitive" result is the local property of the local polynomial kernel estimator. Specifically, since the cluster size J is finite as $n \to \infty$, the probability of having more than one observation from the same cluster in the neighborhood of a target point z goes to 0 as the bandwidth $h \to 0$ and $n \to \infty$. Hence the optimal strategy is to ignore the within-cluster correlation.

5 Extension to the Nonparametric Stratified Hazard Model

We have mainly focused on the nonparametric model with a common baseline hazard in §2–§4. In settings where there are distinct features among observations within the same cluster, it might be desirable to have a stratified hazard model to allow different baseline hazards for different observations. For example, Wei, et al. (1989) discussed an AID example where different baseline hazards corresponding to different disease stages were assumed. We briefly describe an extension of our results to this setting. The nonparametric stratified hazard model takes the form

$$\lambda_{ij}(t) = \lambda_{0j}(t) exp\{\theta(Z_{ij})\}, \tag{9}$$

where $\lambda_{0j}(t)$ is the baseline hazard for the jth observation of each cluster $(j = 1, \cdots, J)$.

At a target point z, the local pseudo partial likelihood, ignoring the within-cluster correlation, is

$$\ell_I^*(\beta, z) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{J} K_h(Z_{ij} - z)\delta_{ij}[Z_{ij}(z)^T \beta - log\{nS_{n,0}^{(j)}(\beta, X_{ij})\}], \tag{10}$$

where $S_{n,r}^{(j)}(\beta, s) = \frac{1}{n}\sum_{i'=1}^{n} K_h(Z_{i'j} - z)Y_{i'j}(s)Z_{i'j}(z)^{\otimes r} e^{Z_{i'j}(z)^T\beta}$ for $r = 0, 1$. It follows that the local pseudo partial likelihood score equation is

$$U_I^*(\beta, z) \;\; = \;\; n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{J} K_h(Z_{ij} - z)\delta_{ij} \left\{ Z_{ij}(z) - \frac{S_{n,1}^{(j)}(\beta, X_{ij})}{S_{n,0}^{(j)}(\beta, X_{ij})} \right\}$$

Note that the sum in $S_{n,r}^{(j)}(\cdot)$ is over at-risk subjects in the jth stratum only. In the common baseline hazard model (3), one sums over at-risk subjects from all strata. One can show that the asymptotic properties of the resulting working independence kernel estimator $\widehat{\theta}_I^{(1)}(z)$ are similar to those in Theorem 1 except that the matrices $V_I(z)$ need to be modified. The covariance of $\widehat{\theta}_I^{(1)}(z)$ can be estimated using a similar sandwich estimator.

Similarly, the weighted local polynomial kernel estimating equation (11) can be modified under the nonparametric stratified hazard model as

$$U_W^*(\beta,z) = \sum_{i=1}^n \int_0^\tau Z_i(z)^T K_{ih}^{1/2} W_i K_{ih}^{1/2} dM_i^*(s) \tag{11}$$
$$= n^{-1}\sum_{i=1}^n \sum_{j=1}^J \sum_{l=1}^J \int_0^\tau Z_{ij}(z) Q_{ijl} dN_{il}(s)$$
$$-n^{-1}\sum_{i=1}^n \sum_{j=1}^J \sum_{l=1}^J \int_0^\tau \frac{\{\sum_{i'=1}^n Y_{i'l}(s) Z_{ij}(z) Q_{i'jl} e^{Z_{i'l}(z)^T\beta}\} K_h(Z_{il}-z) dN_{il}(s)}{S_j^{(0)}(\beta,s)},$$

where $Q_{ijl} = K_h^{1/2}(Z_{ij}-z) w_{j,l}^i K_h^{1/2}(Z_{il}-z)$, $M_i^*(t) = \{M_{i1}^*(t), \cdots, M_{iJ}^*(t)\}^T$, and

$$M_{ij}^*(t) = N_{ij}(t) - \int_0^t \frac{Y_{ij}(s) e^{Z_{ij}(z)^T\beta} \sum_{i'=1}^n K_h(Z_{i'j}-z) dN_{i'j}(s)}{\sum_{i'=1}^n K_h(Z_{i'j}-z) Y_{i'j}(s) e^{Z_{i'j}(z)^T\beta}}.$$

The asymptotic properties of the weighted kernel estimator $\widehat{\theta}_W^{(1)}(z)$ are similar to those stated in Theorem 2 and are omitted.

6 Simulation Study

We evaluate in this section using simulation studies the finite sample performance of the working independence kernel estimator $\widehat{\theta}_I^{(1)}(z)$ and the weighted kernel estimator $\widehat{\theta}_W^{(1)}(z)$. We consider the common baseline hazard model (3), and assume bivariate survival times (T_{i1}, T_{i2}) $(i = 1, \cdots, n)$ follow the Clayton model

$$F_i(t_1, t_2; z_1, z_2, \phi) = \left(exp\left[t_1 exp\{\theta(z_1)\}/\phi\right] + exp\left[t_2 exp\{\theta(z_2)\}/\phi\right] - 1\right)^{-\phi},$$

where $\theta(z) = 0.1\,\{2f(z,8,8) + f(z,5,5)\}$, and $f(z,a,b)$ is the density function of the beta distribution. The function $\theta(z)$ has has a unimodal bell shape, and $\theta^{(1)}(z)$ has a shape similar to a sine function (see Figure 1). The correlation of T_1 and T_2 decreases to 0 as ϕ increase to ∞. This model assumes the marginal distribution of T_{ij} $(j = 1, 2)$ is an exponential distribution with the hazard function $exp\{\theta(z_{ij})\}$ and the baseline hazard is a constant.

The covariates (z_{i1}, z_{i2}) were generated by assuming that they were independent and identically distributed uniform(0,1) random variables. To generate (T_{i1}, T_{i2}) under the

Clayton model, we first generated u_{i1} and u_{i2} independently from uniform(0,1), and set

$$\begin{aligned} T_{i1} &= -log(1-u_{i1})e^{-\theta(z_{i1})} \\ T_{i2} &= \phi log\left\{(1-a)+a(1-u_{i2})^{-(1+\phi)^{-1}}\right\} e^{-\theta(z_{i2})}, \end{aligned}$$

where $a=(1-u_{i1})^{-\phi^{-1}}$ with $\phi=0.5$, which corresponding to strong correlation between T_{i1} and T_{i2}. For example, when $\theta(z_1)=\theta(z_2)=0.6$, without censoring, the correlation between T_1 and T_2 is 0.83. The setting is similar to that of Cai and Prentice(1995) except that a nonparametric covariate function $\theta(\cdot)$ is used. Censoring times C_{ij} were generated from a random variable which follows an exponential distribution with mean 3.5 and a maximum follow-up time 4. Censoring times were generated independently from the failure times and the covariates. The censoring proportion was about 18%.

For each simulated data set, we calculated both the working independence kernel estimators $\{\widehat{\theta}_I^{(1)}(z), \widehat{\theta}_I(z)\}$ and the weighted kernel estimators $\{\widehat{\theta}_W^{(1)}(z), \widehat{\theta}_W(z)\}$. The Newton-Raphson method was used to solve the local polynomial kernel estimating equations (3) and (5). For the weighted kernel estimating equation (5), following Cai and Prentice (1995), we use the weight which is the inverse of correlation matrix of $M_{i1}(X_{i1})$ and $M_{i2}(X_{i2})$ calculated as

$$corr\{M_{i1}(X_{i1}), M_{i2}(X_{i2})|Z_{i1}, Z_{i2}, C_{i1}, C_{i2}\} = \frac{cov\{M_{i1}(X_{i1}), M_{i2}(X_{i2})|Z_{i1}, Z_{i2}, C_{i1}, C_{i2}\}}{([1-exp\{-C_{i1}e^{\theta(Z_{i1})}\}][1-exp\{-C_{i2}e^{\theta(Z_{i2})}\}])^{1/2}}$$

where

$$\begin{aligned} & Cov\{M_{i1}(X_{i1}), M_{i2}(X_{i2})|Z_{i1}, Z_{i2}, C_{i1}, C_{i2}\} \\ = \; & F_i(C_{i1}, C_{i2}; \phi) - 1 + \int_0^{C_{k1}} F_i(t_1, C_{i2}; \phi) exp\{\theta(Z_{i1})\} dt_1 \\ & + \int_0^{C_{i2}} F_i(C_{i1}, t_2; \phi) exp\{\theta(Z_{i2})\} dt_2 + \int_0^{C_{i1}} \int_0^{C_{i2}} F_i(t_1, t_2; \phi) exp\{\theta(Z_{i1})\} exp\{\theta(Z_{i2})\} dt_2 \end{aligned}$$

We used the local linear kernel estimators ($p=1$) with the Epanechnikov kernel and ran 500 replications. We estimated $\theta^{(1)}(z)$ at 81 equally spaced grid points in $[0.1, 0.9]$. Estimation was performed assuming the bandwidth h equal to 0.15, 0.2, 0.25.

The left panel of Figure 1 plots the true curve of $\theta^{(1)}(z)$, the average of the working independence kernel estimator $\widehat{\theta}_I^{(1)}(z)$, and the weight kernel estimator $\widehat{\theta}_W^{(1)}(z)$, each over

500 replications. The bandwidth was set as 0.15. Both estimators are very close to the true function. The right panel of Figure 1 compares the true curve of $\theta(z)$, and the working independence and weighted kernel estimators of $\theta(z)$ calculated by integrating the estimators of $\theta^{(1)}(z)$ using the Trapezoidal rule. One can see that the empirical biases in both the working independence kernel estimator and the weighted kernel estimator are very small. The results using the other two choices of bandwidths ($h = 0.2, 0.25$) are similar and are not reported here.

Figure 2(a) compares the empirical standard errors of the working independence local kernel estimator $\widehat{\theta}_I^{(1)}(z)$ and the weighted local kernel estimator $\widehat{\theta}_W^{(1)}(z)$. The empirical standard error of the weighted local kernel estimator is very close to that of the working independence local kernel estimator. This empirical finding supports our theoretical result in Theorem 3 that the weighted local kernel estimator does not improve the efficiency of the working independence local kernel estimator. This result strongly contrasts the parametric result of Cai and Prentice (1995), where they found that, in this high correlation setting, the weighted estimator is much more efficient than the working independence estimator in parametric regression. Similar results were found when using the bandwidth 0.20 and 0.25. Figure 2(b) and (c) compare the empirical and estimated standard errors of the working independence kernel estimator and the weighted kernel estimator respectively. The estimated standard errors agree with their empirical counterparts well. This suggests that the sandwich standard error estimator works well in finite samples. Figure 2(d) shows that the empirical coverage probabilities of the 95% confident intervals constructed by using the sandwich standard error estimators. The average empirical coverage probability is 92% for the working independence kernel estimator and 94% for the weighted kernel estimators, both are close to the nominal 95% level. We also performed simulations using different levels of censoring and correlation, and observed similar results.

7 Application to the Western Kenya Parasitemia Data

We applied the proposed method to the analysis of the western Kenya parasitemia data described in §1. Parasitemia is an indicator for potential malaria. The western Kenya parasitemia study (McElroy, et al, 1997) enrolled 607 children aged from six months to six years between February 1986 and July 1987. Parasitemia is highly prevalent in Africa, and 94% of children in the study were affected with parasitemia at enrollment. At the date of enrollment, regardless of his/her parasitemia status, each child received a treatment of sulfadoxine and pyrimethamine to eliminate the parasitemia infection. Their blood films were examined two weeks after enrollment. Children with positive blood films were excluded from the study to minimize the chance that a recurrent parasitemia was caused by drug sulfadoxine/pyrimethamine resistance. This resulted in 542 children from 309 households. The outcome variable was time to the first recurrence of parasitemia. Each child was followed up to 84 days. For details of the study, see McElroy, et al. (1997).

In this paper, we are interested in studying the effect of baseline age (AGE) on the risk of the first recurrence of parasitemia. Preliminary examination of the data indicated somewhat complicated nonlinear effect of age (see Figure 3). It is hence desirable to explore the relationship between age and the hazard of parasitemia recurrence nonparametrically and let the data determine its functional form. Since the children from the same household were likely to share similar genetic factors and a similar living environment, their outcomes were likely to be correlated. We hence considered the proportional hazard model with a nonparametric function of age as in (1) to account for the within-family correlation and to model the effect of age nonparametrically.

In view of our theoretical and simulation results that the weighted local kernel estimator does not improve efficiency of the nonparametric local kernel estimator compared with the working independence kernel estimator, we analyzed the data by assuming working independence. Ties of the observed failure times were handled using Efron's method (Kalbfleisch and Prentice, 2002). We examined several choices of the bandwidth and found the choice of the bandwidth $h = 1.5$ for AGE < 2 and $h = 1.6$ for AGE ≥ 2 fit the data well. This choice reflected the observation that the observed age values were more spare for AGE ≥ 2. The right panel of Figure 3 gives the estimated working kernel estimator of the derivative $\theta^{(1)}$(AGE). The left panel of Figure 3 shows the estimated curves assuming θ(AGE) to be piecewise linear, quadratic, cubic and nonparametric, where the nonparametric estimator of θ(AGE) was calculated by integrating $\widehat{\theta}^{(1)}$(AGE) using the Trapezoidal rule. We here set $\theta(0.5) = 0$ for the sake of identifiability of θ(AGE). Note that $H_0 : \theta^{(1)}(\text{AGE} = s) = 0$ tests for the effect of age at AGE$= s$.

Figure 3 show that the linear model would not fit the data. The effect of AGE could not be captured by quadratic and cubic models either. The cubic model generated an artifical curvature when age is large. Examination of the nonparametric curves of θ(AGE) and $\theta^{(1)}$(AGE) shows that the hazard rate increased with AGE for children less than < 2 years old and became constant after age 2. Specially, for AGE$<$2, $\widehat{\theta}^{(1)}$(AGE) was positive. Especially when age was between 0.5 and 1, the confidence interval of $\theta^{(1)}$(AGE) did not cover 0, indicating the risk of parasitemia significantly increased with age in early years. After AGE 2, $\hat{\theta}^{(1)}$(AGE) was close to 0 and its confidence interval covered 0, indicating the risk of parasitemia did not vary with age significantly. To validate this finding, we then fit a piece-wise linear age model with a knot at 2. The fitted piecewise curve of θ(AGE) was given in the left panel of Figure 3 and supported the finding of increasing age effect before 2 and reaching a plateau after 2.

8 Discussions

We propose nonparametric regression for multivariate survival data using the pseudo local partial likelihood based working independence kernel estimating equations and the weighted local partial likelihood kernel estimating equations survival data. Our results show that in contrast to the parametric regression results of Cai and Prentice (1995, 1997), ignoring within-cluster correlation gives the most efficient local kernel estimator for multivariate failure time data. This result is supported by both our theoretical and numerical investigation. It is consistent with the findings of Lin and Carroll (2000) of nonparametric regression for longitudinal data.

We assume in our theoretical investigation that the correlation matrix of $\{M_{ij}(X_{ij})\}_{j=1\cdots J_i}$ is known. In practice, it is often estimated. It is of future research interest to study the asymptotic properties of the local weighted kernel estimator $\widehat{\theta}_W^{(1)}(z)$ with the weight function being estimated. We conjecture that such an estimator will not improve the efficiency either. Since the correlation parameter ϕ in the weight matrix W_i is estimated at the parametric rate, one would expect that the asymptotic variance of the nonparametric estimator $\widehat{\theta}_W^{(1)}(z)$ will not be affected. It is also of future research to develop a data driven method to select the bandwidth of a kernel estimator of the nonparametric derivative function.

We consider in this paper a scalar covariate whose effect is modeled nonparametrically. In practice, multiple covariates are often available. It is of future research interest to extend this nonparametric model to a semiparametric model, where some covariate effects are modeled parametrically while other covariate effects are modeled nonparametrically. Estimation in such a model could proceed using the profile kernel method, along the lines of Lin and Carroll (2001).

Acknowledgment

This work was supported by an R01 grant from the U.S. National Cancer Institute. We thank Bin Nan and Jianwen Cai for their helpful discussions. We also thank Peter McElroy for providing the West Kenya parasitemia data.

Appendix A Regularity Conditions and Lemmas

Before we state the asymptotic results, we introduce some notation and state regularity conditions in this section. For a target point z in the interior of the support of Z, let $U_{ij} = H^{-1}Z_{ij}(z)$, β^0 and $\theta_0(z)$ be the true values, and $\alpha = H(\beta - \beta^0)$. Hence $\alpha = 0$ if and only if $\beta = \beta^0$,

$$S_{n,r}(\alpha, t) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{J} K_h(Z_{ij} - z) Y_{ij}(t) e^{Z_{ij}^T \beta^0 + U_{ij}^T \alpha} U_{ij}^{\otimes r},$$

$$s_r(\alpha, t) = \lim_{n \to \infty} E\{S_{n,r}(\alpha, t)\} = \sum_{j=1}^{J} f_j(z) \bar{P}_j(t|z) \int e^{\tilde{u}^T \alpha} \tilde{u}^{\otimes r} K(u) du,$$

where $r = 0, 1, 2$, $\tilde{u}$ was defined in §4.1, $\otimes$ is the outer product defined as $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = a^T a$ for a vector a. It follows that $s_1(0, t)/s_0(0, t) = \tilde{\nu}_1$, where $\tilde{\nu}_1$ was defined in §4.1. Further let

$$\bar{S}_{n,r}(t) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{J} K_h(Z_{ij} - z) Y_{ij}(t) e^{\theta_0(Z_{ij})} U_{ij}^{\otimes r}$$

$$\bar{s}_r(t) = \lim_{n \to \infty} E\{\bar{S}_{n,r}(t)\} = \sum_{j=1}^{J} f_j(z) \bar{P}_j(t|z) e^{\theta_0(z)} \int \tilde{u}^{\otimes r} K(u) du.$$

We assume conditions A for Theorem 1, which states the properties of the working independent local kernel estimator $\widehat{\theta}_I(z)$. When $J = 1$, the data become independent cross-sectional data and these conditions reduce to those for the local partial likelihood on page 1685 of Fan et al. (1997).

Conditions A *(1)* $P\{Y_{ij}(t) = 1, \text{for } t \in [0, \tau]\} > 0$ *for all* i, j; *(2) The kernel function* $K(\cdot)$ *is a bounded continuous symmetric density function with mean 0 and variance 1 and a compact support, and* $\bar{P}_j(t|z)$ *is equicontinuous at target point* z; *(3)* $\theta(z)$ *has a continuous* $(p+1)$*th derivative in the neighborhood of* z; *(4) The density* $f_j(z)$ *is continuous at* z *and* $f_j(z) > 0$ *has a bounded support for all* j; *(5)* $n \to \infty$, $h \to 0$, $nh \to \infty$ *and* nh^{2p+3} *is bounded.*

Let $\bar{w}_{jl}(z, z) = E\{w_{jl}^i | Z_{ij} = z, Z_{il} = z\}$, and for $j, l = 1, \cdots, J$,

$$T_{n,r}^l(\alpha, t) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{J} U_{ij}^{\otimes r} K_h^{1/2}(Z_{ij} - z) w_{jl}^i K_h^{1/2}(Z_{il} - z) Y_{il}(t) e^{Z_{il}^T \beta^0 + U_{il}^T \alpha} \quad (r = 1, 2)$$

$$T_{n,r}(\alpha, t) = \sum_{l=1}^{J} T_{n,r}^l(\alpha, t)$$

$$\widetilde{T}_r^l(\alpha, t) = \bar{w}_{ll}(z, z) \bar{P}_l(t|z) f_l(z, z) \int e^{\tilde{u}^T \alpha} \tilde{u}^{\otimes r} K(u) du,$$

We assume conditions B for Theorems 2 and 3, which state the properties of the weighted local kernel estimator.

Conditions B *(1) The weight* w_{jl} *is a known bounded function of* Z*; (2)* $S_{n,r}(\alpha, t)$ *converges in probability to* $s_r(\alpha, t)$ *uniformly over* $B \times [0, \tau]$*, where* B *is the support of* α*; (3)* $T^l_{n,r}(\alpha, t)$ *converges in probability to* $\widetilde{T}^l_r(\alpha, t)$ *uniformly on* $B \times [0, \tau]$ *for* $r = 1, 2$*; (4)* $E\{Y_j(t)|Z\} = E\{Y_j(t)|Z_j\}$ *for* $j = 1, \cdots, J$*; (5)* $E\{M_j(t)M_l(t)|Z\} = E\{M_j(t)M_l(t)|Z_j, Z_l\}$*, for* $j, l = 1, \cdots, J$*.*

The following lemmas are used in the proofs of Theorems 1 and 2. We name Lemma 1 in Fan et al. (1997) as Lemma 1 in this paper for the convenience of reference. The proofs of Lemmas 2-4 are straightforward and are omitted. Lemma 2 can be proved by using theorem 37 in Chapter II of Pollard(1984). Lemma 3 can be proved following the proof of Lemma A1 of Spiekerman and Lin (1998). Lemma 4 follows directly from Lemma 1 and 2.

Lemma 2 *Suppose* $G_r(\cdot)$ $(r = 1, 2)$ *are bounded functions and have a compact support,* $g(\cdot)$ *is a continuous function on a compact domain* $\mathcal{D}$ *of dimension* J*. Define*

$$\begin{aligned} c_n(t) &= h^{-1}n^{-1}\sum_{i=1}^{n} Y_{ij}(t)g(Z_{i1}, \cdots, Z_{iJ})\{G_{1h}(Z_{ij} - z)G_{2h}(Z_{il} - z)\}^{1/2}, \\ c(t) &= \bar{P}_j(t|z)g_1(z, z)f_{jl}(z, z)\int G_1(u)^{1/2}du \int G_2(u)^{1/2}du, \end{aligned}$$

where $g_1(z, z) = E\{g(Z_1, \cdots, Z_J)|Z_j = z, Z_l = z\}$*,* $G_{rh}(u) = G_r(u/h)/h$*. Then* $c_n(t)$ *converges in probability uniformly to* $c(t)$ *for* $t \in [0, \tau]$ *if* $h \to 0$*,* $nh^2/log(n) \to \infty$*, and* $0 < \tau \leq \infty$*.*

In the proof of the convergence of $T^l_{n,1}$ in Lemma 4 we use Lemma 2 and set $G_r(\cdot)$ to be $G_1(u) = u^2K(u)$ and $G_2(u) = K(u)e^{2\tilde{u}^T\alpha}$.

Lemma 3 *If* $\{f_n(\cdot)\}$ *is a sequence of random functions on* $[0, \tau]$ *with* $\int_0^\tau |df_n(s)| = O_p(1)$ *and* $\|f_n(t)\| = o_p(1)$*, then for* $j = 1, \cdots, J$*,* $\|\sqrt{h/n}\int_0^t f_n(s)d\widetilde{M}_{\cdot j}(s)\| \to 0$ *in probability, where* $\|\cdot\|$ *is the supreme norm over* t*, and* $\widetilde{M}_{\cdot j}(s) = \sum_{i=1}^n K_h(Z_{ij} - z)M_{ij}(s)$*.*

Lemma 4 *The random variables* $S_{n,r}(t, \alpha) \to s_r(t, \alpha)$*;* $\bar{S}_{n,r}(t) \to \bar{s}_r(t)$*;* $T^l_{n,r}(\alpha, t) \to \widetilde{T}^l_r(\alpha, t)$ *and hence* $T_{n,r}(\alpha, t) \to \sum_{l=1}^J \widetilde{T}^l_r(\alpha, t)$*, in probability uniformly for* $t \in [0, \tau]$ *and* $r = 0, 1, 2$*.*

Appendix B Proof of Theorem 1

Reparametrize (2) to $\ell_I(\alpha, z)$ using α, one can prove Theorem 1 by following the structure of the proof of Andersen and Gill (1982) in three steps: (i) Show that $\widehat{\alpha}_I = H(\widehat{\beta}_I - \beta^0)$ converges in probability to 0; (ii) Show that $\sqrt{nh}\{U_I(\alpha,z)|_{\alpha=0} - b(z)\}$ converges in distribution to $N(0,\Sigma)$, where $b(z) = O(h^{p+1})$; (iii) For $\widehat{\alpha}_I^* \to 0$ in probability, $\partial U_I(\alpha,z)/\partial\alpha|_{\alpha=\widehat{\alpha}_I^*}$ converges in probability to Σ_I, where Σ_I is a positive definite matrix. Note that $U_I(\alpha,z) = \partial\ell_I(\alpha,z)/\partial\alpha^T$. If these three results are obtained, the proof of the asymptotic distribution of $\widehat{\alpha}_I$ follows from the arguments in Fan et al. (1997). The proofs of (ii) and (iii) are a special case of those for the weighted estimator $\widehat{\alpha}_W$ if assuming working independence. We hence skip these proofs to save space and only give the proof of (i).

Using Lemmas 3 and 4 and following a similar approach of Fan et al. (1997), one can show that the pseudo local partial likelihood can be rewritten as

$$\begin{aligned}
\ell_I(\alpha,z) - \ell_I(0,z) &= \int_0^\tau \bar{S}_{n,1}(t)^T \alpha\lambda_0(t)dt - \int_0^\tau log\left\{\frac{S_{n,0}(\alpha,t)}{S_{n,0}(0,t)}\right\}\bar{S}_{n,0}(t)\lambda_0(t)dt \\
&\quad + \int_0^\tau n^{-1}\sum_{i=1}^n\sum_{j=1}^J K_h(Z_{ij}-z)\left[U_{ij}^T\alpha - log\left\{\frac{S_{n,0}(\alpha,t)}{S_{n,0}(0,t)}\right\}\right]dM_{ij}(t) \\
&= A_1(\alpha,\tau) - A_2(\alpha,\tau) + o_p(1)
\end{aligned}$$

Note the third integral converges to 0 in probability from Lenglart Inequality, and

$$\begin{aligned}
A_1(\alpha,\tau) &= \sum_{j=1}^J f_j(z)e^{\theta_0(z)}\bar{\Lambda}_j(s,z)\tilde{\nu}_1^T\alpha, \\
A_2(\alpha,\tau) &= \sum_{j=1}^J f_j(z)e^{\theta_0(z)}\bar{\Lambda}_j(t,z)log\{\int e^{\tilde{u}^T\alpha}K(u)du\}.
\end{aligned}$$

One can show that $A_1(\alpha,\tau) - A_2(\alpha,\tau)$ is a concave function of α with the maximizer $\alpha = 0$. Therefore, using Corollary II.2 of Andersen and Gill (1982), $H(\hat{\beta}_I - \beta^0) \to 0$ in probability.

Appendix C Proof of Theorem 2

The equations (5) are reparametrized to $U_W(\alpha,z) = 0$ using $\alpha = H(\beta - \beta^0)$. The asymptotic properties of its solution $\widehat{\alpha}_W$ can be shown in similar three steps to those used for the working independence kernel estimator $\widehat{\alpha}_I$ in Appendix B. For step (i), in absence of the likelihood, estimating equation techniques need to be used. Specifically, to prove step (i), we use the four conditions extended by Cai and Prentice (1995) from Foutz (1977) to prove the consistency of the solutions of the weighted local kernel estimating equations. These conditions can be verified under conditions A and B. To save space, we omit the proofs. Detail proofs are available upon request. We provide below proofs of steps (ii) and (iii). The proof of normality (step ii) uses the Cramer-Wald device and the Linderberg-Feller's Central Limit Theorem (Durrett, 1995).

C.1 Proof of Step (ii)

We prove in this section step (ii), $\sqrt{nh}\{U_W(\alpha,z)|_{\alpha=0}-b(z)\}$ converges to $N(0,\Sigma_W)$. We write $U_W(\alpha,z)|_{\alpha=0}$ as

$$\begin{aligned} U_W(0,z) &= n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{J}\int_0^\tau \Big\{\sum_{j=1}^{J}U_{ij}Q_{ijl}-\frac{T_{n,1}(0,t)K_h(Z_{il}-z)}{S_{n,0}(0,t)}\Big\}dM_{il}(t) \\ &\quad + n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{J}\int_0^\tau \Big\{\sum_{j=1}^{J}U_{ij}Q_{ijl}-\frac{T_{n,1}(0,t)K_h(Z_{il}-z)}{S_{n,0}(0,t)}\Big\}Y_{il}(t)e^{\theta_0(Z_{il})}\lambda_0(t)dt \\ &= U_n+B_n \end{aligned}$$

Some calculations show that

$$\sqrt{nh}B_n=\sqrt{nh}b(z)+o_p(\sqrt{nh^{2p+3}}), \tag{A. 1}$$

where

$$\begin{aligned} b(z) &= \Big\{\int(\tilde{u}-\tilde{\nu}_1)u^{p+1}K(u)du\Big\}\times\sum_{j=1}^{J}\frac{e^{\theta_0(z)}\theta_0^{(p+1)}(z)h^{p+1}}{(p+1)!}f_j(z)\bar{w}_{jj}(z,z)\bar{\Lambda}_j(\tau,z) \\ &= O(h^{p+1}). \end{aligned}$$

Write the first term, $\sqrt{nh}U_n$ as

$$\sqrt{nh}U_n \;=\; \sqrt{nh}U_{n1}-\sqrt{nh}U_{n2}$$

where

$$\begin{aligned} U_{n1} &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{l=1}^{J}\int_0^\tau\left\{U_{ij}Q_{ijl}-\frac{e_j(t,z)K_h(Z_{il}-z)}{s_0(0,t)}\right\}dM_{il}(t) \\ U_{n2} &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{l=1}^{J}\int_0^\tau\left\{\frac{T_{n,1}^{j}(0,t)}{S_{n,0}(0,t)}-\frac{e_j(t,z)}{s_0(0,t)}\right\}K_h(Z_{il}-z)dM_{il}(t), \end{aligned}$$

where $e_j(t,z)=\bar{P}_j(t|z)f_j(z)\bar{w}_{j,j}(z,z)\tilde{\nu}_1$. ¿From conditions A and Lemmas 3 and 4,

$$\sqrt{nh}U_{n2}=o_p(1) \tag{A. 2}$$

To prove the asymptotic normality of $\sqrt{nh}U_{n1}$, write $U_{n1} = n^{-1}\sum_{i=1}^n U_{n1i}$. We use the Cramer-Wald device and need to show that $\sqrt{h/n}c^T\sum_{i=1}^n U_{ni1}$ converges to a normal random variable for all $\mathbf{c} \in R^p$. To proceed, we first replace $e_j(t,z)$ in U_{n1i} by $\tilde{e}_{jl}(t,z) = \sum_{r=1}^J e_r(t,z)$ for $j = l$, and 0 otherwise. Then check the three conditions of Linderberg-Feller's central limit theorem (P. 116 of Durrett, 1995). The mean 0 and Lindeberg conditions are easy to check and their proofs are omitted. The third condition is the finite limit variance condition, that is, $lim_{n\to\infty}\sum_{i=1}^n \mathrm{var}(\sqrt{h/n}c^T U_{n1i}) < \infty$. This can be verified by checking the limit of the covariance matrix $\lim_{n\to\infty} E\{\sqrt{h}U_{n1i}\}^{\otimes 2} = \Sigma_W$ is finite. Some calculations show that

$$\begin{aligned}\Sigma_W &= e^{\theta_0(z)}\sum_{j=1}^J f_j(z)\bar{w}_{jj}^2(z,z)\bar{\Lambda}_j(\tau,z)\int \tilde{u}\tilde{u}^T K^2(u)du \qquad \text{(A. 3)}\\ &+e^{\theta_0(z)}\Big[\int_0^\tau \frac{(\Big\{\sum_{j=1}^J f_j(z)\bar{w}_{jj}(z,z)\bar{P}_j(t|z)\Big\}^2}{\sum_{j=1}^J f_j(z)\bar{P}_j(t|z)}\lambda_0(t)dt\Big]\Big[-\tilde{\nu}_1\tilde{\nu}_2^T - \tilde{\nu}_2\tilde{\nu}_1^T + \tilde{\nu}_1\tilde{\nu}_1^T\int K^2(u)du\Big]\end{aligned}$$

It follows that

$$\sqrt{nh}U_{n1} \to N(0,\Sigma_W) \qquad \text{(A. 4)}$$

in distribution. Write

$$\sqrt{nh}\{U_W(0,z) - b(z)\} = \sqrt{nh}U_{n1} + \sqrt{nh}U_{n2} + \sqrt{nh}\{B_n - b(z)\}.$$

From (A. 1), (A. 2), (A. 4), using Slusky's theorem and condition A(5), we have $\sqrt{nh}\{U_W(0,z) - b(z)\} \to N(0,\Sigma_W)$ in distribution.

C.2 Proof of Step (iii)

Suppose $\widehat{\alpha}_W^*$ lies between 0 and $\widehat{\alpha}_W$. Hence $\widehat{\alpha}_W^* \to 0$ in probability. Let $I_{n,\hat{\alpha}_W^*} = \partial U_W(\alpha,z)/\partial\alpha^T|_{\alpha=\hat{\alpha}_W^*}$. We need to show that $I_{n,\hat{\alpha}_W^*} \to \Sigma_{W1}$ for some Σ_{W1}. By a Taylor expansion, $I_{n,\hat{\alpha}_W^*} = I_{n,0} + I_{n,\hat{\alpha}_W^{**}}^{(1)}(\hat{\alpha}_W^* - 0)$, where $\widehat{\alpha}_W^{**}$ lies between $\widehat{\alpha}_W^*$ and 0 and $I^{(1)}(\cdot)$ denotes the first derivative. Since $I_{\hat{\alpha}_W^{**}}$ is bounded, the second term is $o_p(1)$. It follows that $I_{n,\hat{\alpha}_W^*} = I_{n,0} + o_p(1)$. Some calculations show that

$$I_{n,0} = -n^{-1}\sum_{i=1}^n\sum_{l=1}^J\int_0^\tau \frac{T_{n,2}(0,t)S_{n,0}(0,t) - T_{n,1}(0,t)S_{n,1}(0,t)^T}{\{S_{n,0}(0,t)\}^2}K_h(Z_{il}-z)dN_{il}(t).$$

¿From Lemma 4, we have

$$\frac{T_{n,2}(0,t)S_{n,0}(0,t) - T_{n,1}(0,t)S_{n,1}(0,t)^T}{\{S_{n,0}(0,t))\}^2} = \frac{\sum_{j=1}^J \bar{P}_j(t|z)f_j(z)\bar{w}_{jj}(z,z)}{\sum_{j=1}^J \bar{P}_j(t|z)f_j(z)}D + o_p(1)$$

uniformly over $[0,\tau]$, where D was the positive definite matrix defined in Theorem 1. Some calculations show that $I_{n,0} = \Sigma_{W1} + o_p(1)$, where

$$\Sigma_{W1} = -De^{\theta_0(z)}\sum_{j=1}^{J} f_j(z)\bar{w}_{j,j}(z,z)\bar{\Lambda}_j(\tau,z) + o_p(1).$$

It follows that $I_{n,\hat{\alpha}_W^*} = \Sigma_{W1} + o_p(1)$.

C.3 Proof of Asymptotic Normality of $\widehat{\alpha}_W$

Take a Taylor expansion

$$U_W(\hat{\alpha}_W, z) = U_W(0,z) + I_{n,\hat{\alpha}_W^*}(\hat{\alpha}_W - 0),$$

where $\hat{\alpha}_W^*$ lies between 0 and $\hat{\alpha}_W$. We have $\hat{\alpha}_W = -I_{n,\hat{\alpha}_W^*}^{-1}U_W(0,z)$. Using step (i) and the results in Sections C.1 and C.2, we have

$$\sqrt{nh}\{\hat{\alpha}_W + \Sigma_{W1}^{-1}b(z)\} \rightarrow N(0, \Sigma_{W1}^{-1}\Sigma_W\Sigma_{W1}^{-1}).$$

Some calculations show that the asymptotic bias and variance can be simplified as $\Sigma_{W1}^{-1}b = -\{\theta^{(p+1)}(z)/(p+1)!\}D^{-1}ch^{p+1}$, where D and c are given in Theorem 1, and the covariance can be simplified as $\Sigma_{W1}^{-1}\Sigma_W\Sigma_{W1}^{-1} = V_W(z)$, where $V_W(z)$ is given in equation (6).

Reference

Andersen, P.K. & Gill, R. D. (1982). Cox's Regression Model for Counting Processes: A Large Sample Study. *Ann. Statist.* **10**, 1100-1120.

Cai, J. & Prentice R. L. (1995). Estimating Equations for Hazard Ratio Parameters Based on Correlated Failure Time Data. *Biometrika* **82**, 151-164.

Cai, J. & Prentice R. L. (1997). Regression Estimation Using Multivariate Failure Time Data and a Common Baseline Hazard Function Model. *Lifetime Data Anal.* **3**, 197-213.

Dabrowska, D. M. (1987). Non-parametric regression with censored survival time data. *Scand. J. Statist.* **14**, 181-197.

Durrett, R. (1995). Probability Theory and Examples. *Duxbury Press*, 170.

Fan, J., Gijbels, I. & King, M. (1997). Local Likelihood and Local Partial Likelihood in Hazard Regression. *Ann. Statist.*, **25**, 1661-1690.

Foutz, R. V. (1977). On the unique consistent solution to the likelihood equations. *J. Am. Statist. Assoc.*, **72**, 147-148.

Gray, R. (1992). Flexible methods for analyzing survival data using splines, with application to breast cancer prognosis. *J. Am. Statist. Assoc.*, **87**, 942-51.

Gray, R., & Li, Y. (2002). Optimal weight functions for marginal proportional hazards analysis of clustered failure time data. *Lifetime Data Anal.*, **8(1)**, 5-19.

Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statist. Sci.*, **1**, 297-310.

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

Kalbfleisch, J. D., & Prentice, R. L. (2002). The statistical analysis of failure time data. *John Wiley & Sons New York; Chichester.*

Lee, E. W., Wei, L. J., & Amato, D. A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. *Survival Analysis: State of the Arts*, 237-247.

Li, G & Doss, H. (1995). An approach to nonparametric regression for life history data using local linear fitting. *Ann. Statist.* **23** 787-823.

Liang, K. Y. & Zegger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.

Lin, X. & Carroll, R. J. (2000). Nonparametric Function Estimation for Clustered Data

When the Predictor is Measured Without/With Error. *J. Am. Statist. Assoc.*, **95**, 520-534.

Lin, X. & Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations, *J. Am. Statist. Assoc.*, **96**, 1045-1056. k

Mcelroy, P. D., Beier, J. C., Oster, C. N., Onyango, F. K., Lin X., Beedle, C. & Hoffman S. L. (1997). Dose- and Time-dependent relations between infective anopheles inoculation and outcomes of plasmodium falciparum parasitemia among children in Western Kenya. *Am. J. Epidemiol.*, **145**, 945-956.

O'Sullivan, F. (1988). Nonparametric estimation of relative risk using splines and cross-validation. *SIAM J. Sci. Stat. Comput.* **9**, 531-542.

O'Sullivan, F. (1993). Nonparametric estimation in the Cox model. *Ann. Statist.* **21**, 124-145.

Pollard, D. (1984). Convergence of Stochastic Process. *New York: Springer*, 43-52 and 179-180.

Prentice, R. & Cai, J. (1992). Covariate and Survivor Function Estimation Using Censored Multivariate Failure Time Data. *Biometrika*, **79**, 495-512.

Spiekerman, C. F. & Lin D. Y. (1998). Marginal Regression Models for Multivariate Failure Time Data. *J. Am. Statist. Assoc.*, **93**, 1164-1175.

Tibshirani, R. & Hastie T. (1987). Local likelihood estimation. *J. Am. Statist. Assoc*, **82**, 559-567.

Wei , L. J., Lin, D. Y., & Weissfeld, L. (1989). Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. *J. Am. Statist. Assoc.*, **84**, 1065-1073.

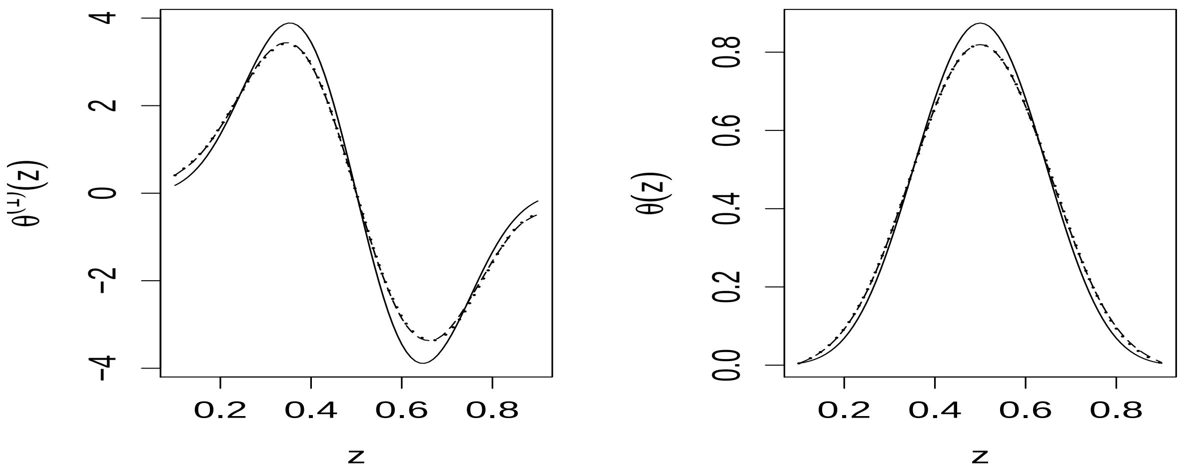

Figure 1: The left panel compares the average of the working independence local kernel estimates, the average of the weighted local kernel estimates of $\theta^{(1)}(z)$, with the true derivative curve $\theta^{(1)}(z)$. The right panel gives the corresponding comparison of the estimates of $\theta(z)$. The bandwidth is set as $h = 0.15$: $\cdots$ the working independence local kernel estimator; $-\quad-$ the weighted local kernel estimator; —— the true curve.

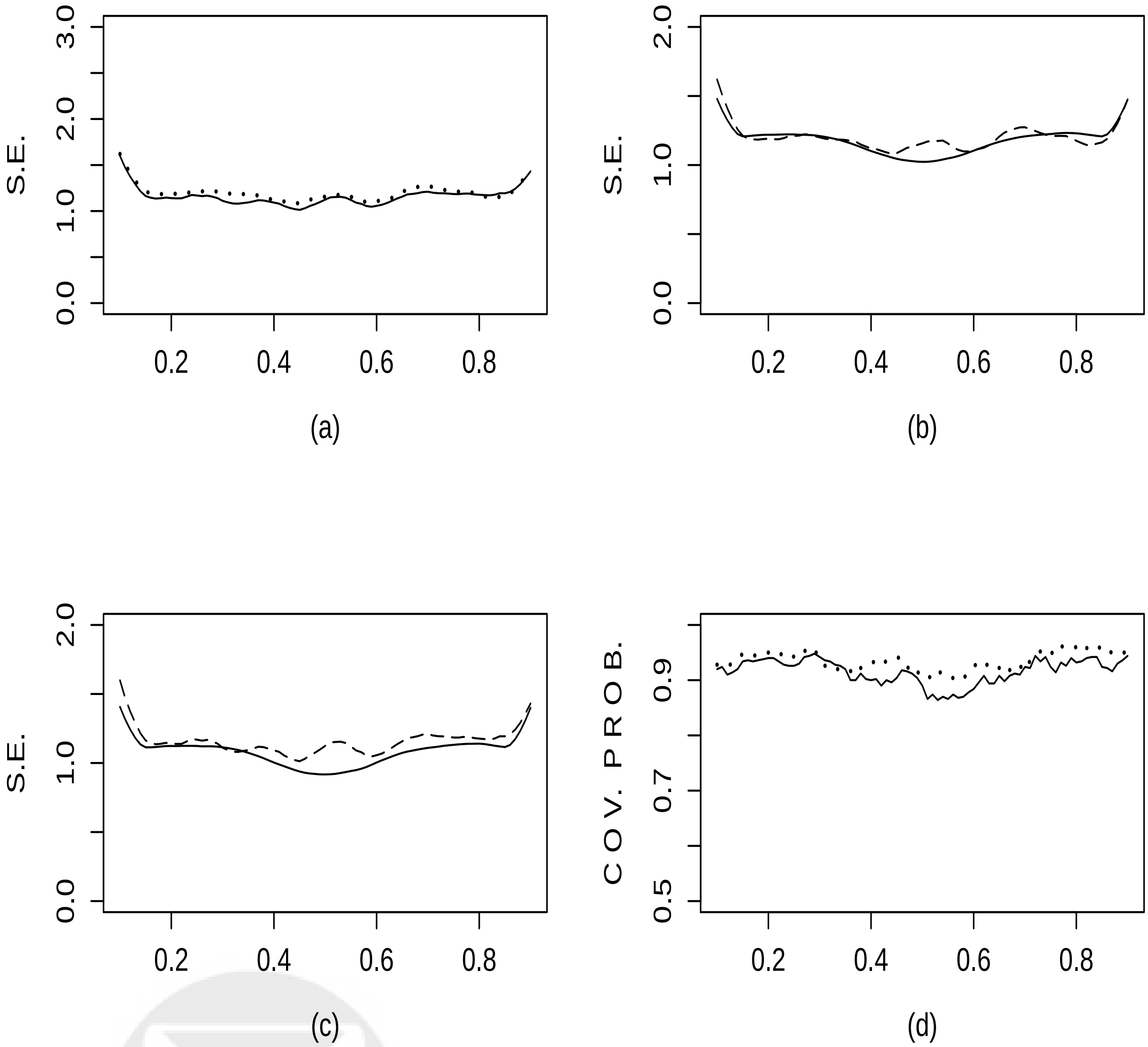

Figure 2: (a) Comparison of the empirical SEs of the working independent kernel estimate $\widehat{\theta}_I^{(1)}(z)$ and the weighted kernel estimates $\widehat{\theta}_W^{(1)}(z)$: $\cdots$ $\widehat{\theta}_I^{(1)}(z)$, — $\widehat{\theta}_W^{(1)}(z)$. (b) Comparison of the empirical and sandwich SE estimates of the working independence kernel estimator $\widehat{\theta}_I^{(1)}(z)$: $\cdots$ Estimated sandwich standard error; — the empirical standard error. (c) Comparison of the empirical and sandwich SEs of the weighted kernel estimator $\widehat{\theta}_W^{(1)}(z)$: $\cdots$ the estimated sandwich standard error; — the empirical standard error. (d) Empirical coverage probabilities of the 95% estimated point-wise confidence intervals: $\cdots$ $\theta_I^{(1)}(z)$; — $\theta_W^{(1)}(z)$. The bandwidth was set as $h = 0.15$.

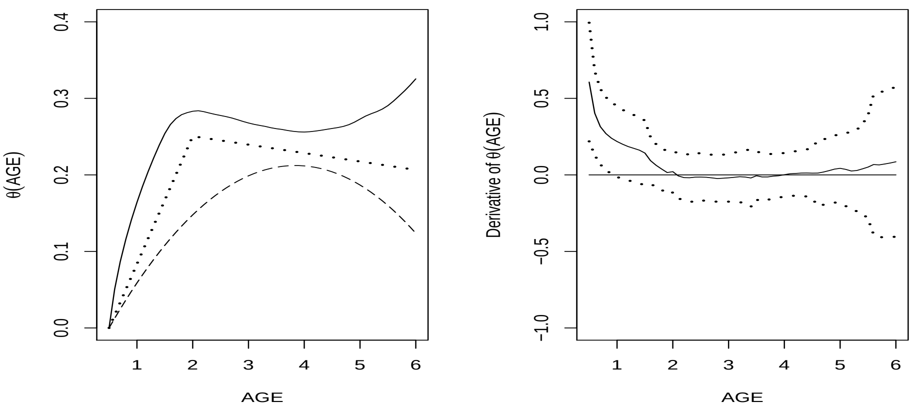

Figure 3: The left panel plots the parametric and nonparametric estimators of θ(AGE) assuming working independence: — kernel fit; $\cdots$ piecewise linear fit; $- \quad -$ quadratic fit; $\cdot - \cdot$ cubic fit. The right panel plots the working independence kernel estimator of $\theta^{(1)}$(AGE) and its corresponding point-wise CIs: $\cdots$ the 95% point-wise confidence interval; — the working independence kernel estimate of $\theta^{(1)}$(AGE).