Statistica Sinica **20** (2010), 853-869

SEMIPARAMETRIC REGRESSION WITH TIME-DEPENDENT COEFFICIENTS FOR FAILURE TIME DATA ANALYSIS

Zhangsheng Yu and Xihong Lin

Indiana University and Harvard School of Public Health

Abstract: We propose a working independent profile likelihood method for the semiparametric time-varying coefficient model with correlation. Kernel likelihood is used to estimate time-varying coefficients. Profile likelihood for the parametric coefficients is formed by plugging in the nonparametric estimator. For independent data, the estimator is asymptotically normal and achieves the asymptotic semiparametric efficiency bound. We evaluate the performance of proposed nonparametric kernel estimator and the profile estimator, and apply the method to the western Kenya parasitemia data.

Key words and phrases: Clustered survival data, efficiency, estimating equation, kernel smoothing, marginal model, profile likelihood, sandwich estimator.

1. Introduction

The Cox proportional hazard model has been widely used for analysis of independent censored failure time data, where covariate effects are often modeled parametrically. This parametric assumption is sometimes undesirable in practice, as the functional form of the covariate effect is often unknown and can be complex. Several extensions of the Cox model have been proposed to model the covariate effects nonparametrically (Fan, Gijbels and King (1997)) and to allow regression coefficients to vary with time (Hastie and Tibshirani (1993); Cai and Sun (2003)).

Clustered failure time data have emerged in the last decade. Examples include familial studies, where the survival times of multiple family members are clustered within the same family. A common feature of clustered failure time data is that observations within the same cluster are likely to be correlated. Statistical methods for analysis of clustered failure time data in the last decade have been mainly focused on multivariate analogs of the Cox model, where covariate effects are modeled parametrically. In this paper, we propose a flexible semiparametric regression model with time-varying regression coefficients for clustered survival data, where some regression coefficients are assumed to be constant while some are assumed to be time-varying. Such models are particularly useful in situations when exposure effects or treatment effects diminish over time.

One motivating example is the western Kenya parasitemia study (McElroy, Beier, Oster, Onyango, Lin, Beedle and Hoffmann (1997)). This study enrolled 542 children from 309 households and followed them over time for the onset of parasitemia. The risk factors of major interest are baseline parasitemia density (BPD) and exposure to mosquito bites (BITE). Other covariates include age and gender. McElroy et al. (1997) assumed the effect of BPD to be constant over time. As shown in Figure 5.2, the effect of BPD diminishes over time. Examination of the data shows that it is reasonable to assume constant effects of BITE, age, and gender. It is hence desirable to consider a semiparametric time-varying coefficient model, where the regression coefficient of BPD is allowed to be time varying and nonparametric. A further complication of this study, and of the estimation problem, is that survival times of the children within the same family are likely to be correlated, due to similar genetic and environmental factors.

For independent data, varying-coefficient models were first proposed by Hastie and Tibshirani (1993). The asymptotic properties of kernel estimation in such models were studied by Cai and Sun (2003). Gray (1992) considered smoothing spline estimation in such models. Extensions to semiparametric regression, where some covariate effects are modeled parametrically, were considered by several authors. Tian, Zucker, and Wei (2005) and Fan, Lin and Zhou (2006) estimated the nonparametric component using kernel smoothing and the parametric component using weighted estimators or the profile estimators, respectively. Ahmad, Leelahanon, and Li (2005) studied efficient estimation in these semiparametric models, where the nonparametric component was estimated using splines, which are computationally more intensive compared to the kernel method. Other related works include Zucker and Karr (1990), and Winnett and Sasieni (2003).

Several authors have considered nonparametric regression for clustered failure time data. Yu and Lin (2008) considered weighted local polynomial kernel estimating equations for clustered failure time data, and showed the most efficient local polynomial kernel estimator is obtained by ignoring the within-cluster correlation. Cai, Fan, Zhou and Zhou (2007) considered nonparametric regression in varying coefficient models for multivariate survival data and proposed kernel smoothing assuming independence. We consider in this paper a semiparametric model here, where one covariate effect is modeled nonparametrically using a time-varying nonparametric function, and other covariates are modeled parametrically.

Motivated by the Western Kenya data, we consider a semiparametric time-varying coefficient model for clustered survival data. In view of the results of

Yu and Lin (2008), we consider survival models with only time-varying coefficients and a common baseline hazard, and propose estimation of nonparametric functions using the working independence kernel partial likelihood constructed by ignoring within-cluster correlation. We study the asymptotic properties of these nonparametric estimators. We next consider semiparametric models, where the effects of some covariates are modeled parametrically. We propose a working independence kernel profile partial likelihood method for estimating nonparametric time-varying coefficients and parametric regression coefficients. We show that the profile estimator is $\sqrt{n}$-consistent and asymptotically normal. For independent data, we derive the semiparametric efficient score and show the profile estimators of regression coefficients are semiparametric efficient. To our knowlege, this is the first paper that shows the semiparametric efficiency for the profile-kernel estimator in semiparametric varying coefficient models. We evaluate the finite sample performance of the proposed methods using simulations and illustrate their application using the western Kenya parasitemia data.

The remainder of this paper is organized as follows. We described the model and the estimation procedure in Section 2. In Section 3, we study the asymptotic properties of the proposed estimator, and derive the efficiency score. In Section 4, we evaluate the performance of the proposed estimators by simulations. We apply the proposed estimation procedure to the western Kenya parasitemia study in Section 5, followed by some discussions in Section 6. Technical details of theory derivation are included in an appendix as a supplement to the online version available at `http://www.stat.sinica.edu.tw/statistica`.

2. Models and Estimation

In this section we consider nonparametric and semiparametric regression models for time-varying coefficient models for clustered survival models with a common baseline hazard.

2.1. Nonparametric time-varying coefficient models

In this section we propose working independence kernel estimation for nonparametric time-varying coefficient models. Suppose one observes $(T_{ij}, X_{ij}, \Delta_{ij})$ for the jth subject of the ith cluster $(i = 1, \ldots, n; j = 1, \ldots, J_i)$. Here $T_{ij} = \min(T^*_{ij}, C_{ij})$, where T^*_{ij} is the failure time and C_{ij} is the censoring time; $X_{ij}(t)$ is the covariate vector $\{X_{1ij}(t), \ldots, X_{pij}(t)\}^T$, and $\Delta_{ij} = I(T^*_{ij} \le C_{ij})$ is the censoring indicator. Let $Y_{ij}(t) = I(Y_{ij} \ge t)$ be the at-risk process, and $N_{ij}(t)$ be the observed failure counting process. The T_{ij} within the same cluster i might be correlated, while observations from different clusters are assumed to be independent. Failure times $\{T^*_{ij}\}$ are independent from censoring time $\{C_{ij}\}$, conditional on covariates.

We assume that the marginal hazard for T_{ij} follows

$$\lambda_{ij}(t; X_{ij}) = Y_{ij}(t)\lambda_0(t)\exp\{X_{ij}(t)^T\beta(t)\}, \tag{2.1}$$

where $\beta(t) = \{\beta_1(t), \ldots, \beta_p(t)\}^T$ is a vector of time-varying coefficients and $\beta_k(t)$ is a nonparametric smooth function with a continuous second derivative, and $\lambda_0(t)$ is a common baseline hazard. The hazard $\lambda_{ij}(t; X_{ij})$ is the instantaneous failure rate at time t for the ijth subject, conditional on all the information of the ijth subject prior to t. The random process

$$M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(u)e^{X_{ij}(u)^T\beta(u)}\lambda_0(u)du$$

is a martingale with respect to the marginal filtration, but not to the joint filtration.

We propose estimating the nonparametric time-varying coefficient using the local polynomial kernel method assuming working independence. For simplicity we focus here on local linear kernel estimators, extension to a local pth order kernel estimator are straightforward. Specifically, to estimate $\beta_k(\cdot)$ at time t, one can approximate $\beta_k(u)$ using a linear Taylor expansion around t as $\beta_k(u) \approx b_{0k} + b_{1k}(u-t)$. Then $\hat\beta_k(t) = \hat b_{0k}$. Let $b = (b_{01}, \ldots, b_{0p}, b_{11}, \ldots, b_{1p})^T$ and $\widetilde X_{ij}(u,t) = (1, u-t)^T \otimes X_{ij}$, with $\otimes$ being the Kronecker product. In view of the results of Yu and Lin (2008), we propose the working independence kernel partial likelihood by ignoring within-cluster correlation for estimating b as

$$\begin{aligned} p\ell_N(b,t) = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^{J_i}\int_0^\tau K_h(u-t)\Big[&\widetilde X_{ij}(u,t)^T b \\ &- \log\Big\{\sum_{l=1}^n\sum_{r=1}^{J_l} Y_{lr}(u)e^{\widetilde X_{lr}(u,t)^T b}\Big\}\Big]dN_{ij}(u), \end{aligned} \tag{2.2}$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K(\cdot)$ is a kernel function, and h is a bandwidth. The kernel function $K(\cdot)$ is usually chosen to be a unimodal probability density function. In the numerical study presented latter, we use the Epanechnikov kernel $K(t) = \frac{1}{4}(1-t^2)I_{\{|t|<1\}}$. Similar to the Cox partial likelihood, one can derive this kernel partial likelihood by profiling out the baseline hazard from a kernel full likelihood. It reduces to that in Cai and Sun (2003) when the cluster size $J_i = 1$, i.e., all observations are independent.

The concavity of (2.2) as a function of b can be easily verified. One can estimate b by maximizing the kernel partial likelihood (2.2) or by solving the following kernel estimating equations, which are the derivative of the partial likelihood (2.2) with respect to b:

$$0 = U_N(\mathbf{b}, t) = \sum_{i=1}^{n} U_{Ni}(b, t) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{J_i} \int_0^{\tau} K_h(u - t)$$
$$\times \Big\{ \widetilde{X}_{ij}(u, t) - \frac{\sum_{l=1}^{n} \sum_{r=1}^{J_l} Y_{lr}(u) \widetilde{X}_{lr}(u, t) e^{\widetilde{X}_{lr}(u,t)^T b}}{\sum_{l=1}^{n} \sum_{r=1}^{J_l} Y_{lr}(u) e^{\widetilde{X}_{lr}(u,t)^T b}} \Big\} dN_{ij}(u). \quad (2.3)$$

We term equations (2.3) the working independence kernel estimating equations and the resulting estimators $\widehat{\mathbf{b}}$ the working independence kernel estimators, reflecting the fact that the kernel partial likelihood ignores the within-cluster correlation. Then $\hat{\beta}_k(t) = \hat{b}_{0k}$ and $\hat{\beta}'_k(t) = \hat{b}_{1k}$ $(k = 1, \ldots, p)$. We denote the resulting kernel estimator of $\beta(t)$ as $\widehat{\beta}(t)$.

Equations (2.3) can be solved using the Newton-Raphson method. We estimate the covariance of $\widehat{b}$ using the sandwich estimator $\widehat{V}(\widehat{\mathbf{b}}) = \Omega_1^{-1}\Omega_2\Omega_1^{-1}$, where $\Omega_1 = -\partial U_N(b,t)/\partial b|_{b=\widehat{\mathbf{b}}}$, $\Omega_2 = \sum_{i=1}^{n} U_{Ni}(b,t)^{\otimes 2}|_{b=\widehat{\mathbf{b}}}$, and $A^{\otimes 2} = AA^T$. The covariance matrix of $\{\widehat{\beta}_1(t), \ldots, \widehat{\beta}_p(t)\}$ can then be estimated by $\Delta_1^T \widehat{V}(\widehat{\mathbf{b}})\Delta_1$, where $\Delta_1 = (1, \ldots, 1, 0, \ldots, 0)^T$.

2.2. Semiparametric models

In this section we consider a semiparametric regression model where some covariate effects are modeled nonparametrically using varying coefficients, and other covariates are modeled parametrically using parametric regression coefficients. We propose to first estimate nonparametric time-varying coefficients using the local kernel method and then estimate parametric regression coefficients using a profile likelihood method assuming working independence.

Following the notation in Section 2.1, besides $(T_{ij}, X_{ij}, Z_{ij}, \Delta_{ij})$, suppose we also observe covariates Z_{ij}, a $p \times 1$ covariate vector associated with parametric regression coefficients γ. Observed failure times T_{ij} are assumed to have the marginal hazard

$$\lambda_{ij}(t; X_{ij}, Z_{ij}) = Y_{ij}(t)\lambda_0(t) \exp\{X_{ij}\beta(t) + Z_{ij}^T\gamma\}. \quad (2.4)$$

For simplicity, the covariates Z_{ij} are assumed to be time-independent. This model corresponds to a useful setting in our western Kenya data application with a time-varying coefficient for a baseline covariate (BPD) and time-independent coefficients for the other covariates. Similarly, we allow failure times T_{ij} to be correlated within the same cluster and assume independent censoring. We assume a scalar covariate X_{ij} with a scalar time-varying coefficient $\beta(t)$, extensions of the proposed method to the vector X_{ij} case are straightforward.

To estimate both $\beta(t)$ and γ, we propose a working independence kernel profile partial likelihood method by ignoring the within-cluster correlation. The estimation procedure is a two-step iterative procedure. For simplicity, we consider a local kernel estimator of $\beta(t)$ that approximates $\beta(u)$ at a target point u using its linear expansion about t as $\widetilde{\beta}(u) = b_0 + b_1(u-t)$. Let $\widetilde{X}_{ij}(u,t) = (1, u-t)^T \otimes X_{ij}$.

The kernel profile estimation steps are as follows.

(i) For given γ, $b = (b_0, b_1)^T$ is estimated by maximizing the working independence kernel partial likelihood

$$p\ell_1(\mathbf{b}, t; \gamma) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J_i}\int_0^{\tau} K_h(u-t) \times\Bigg[\widetilde{X}_{ij}(u,t)^T b + Z_{ij}^T\gamma - \log\Bigg\{\sum_{l=1}^{n}\sum_{r=1}^{J_l} Y_{lr}(u) e^{\widetilde{X}_{lr}(u,t)^T b + Z_{lr}^T\gamma}\Bigg\}\Bigg] dN_{ij}(u). \quad (2.5)$$

Then the resulting kernel estimator of $\beta(t)$ is $\widehat{\beta}(t,\gamma) = \hat{b}_0$, which is a function of γ.

(ii) After obtaining the kernel estimator $\widehat{\beta}(t,\gamma)$, γ is estimated by maximizing the working independence profile partial likelihood

$$p\ell_2\{\gamma, \widehat{\beta}(u,\gamma)\} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J_i}\int_0^{\tau}\Bigg[X_{ij}\widehat{\beta}(u,\gamma) + Z_{ij}^T\gamma - \log\Bigg\{\sum_{l=1}^{n}\sum_{r=1}^{J_l} Y_{lr}(u) e^{X_{lr}\widehat{\beta}(u,\gamma) + Z_{lr}^T\gamma}\Bigg\}\Bigg] dN_{ij}(u). \quad (2.6)$$

We iterate between these two steps until convergence and obtain the kernel nonparametric estimator $\widehat{\beta}(t)$ and the profile estimator $\widehat{\gamma}$. The concavity of the partial likelihood in these two steps can be easily verified. Note that the profile method for estimating γ differs from the backfitting method in that $\widehat{\beta}(u,\gamma)$ in Step (ii) is considered to be a function of γ instead of being held fixed at the previous iteration when maximizing (2.6), as done in backfitting. Hence as we show in Section 3, $\sqrt{n}$-consistency of γ using the profile method only requires regular smoothing of $\beta(t)$. The backfitting estimator of $\widehat{\gamma}$ on the other hand often requires undersmoothing of $\beta(t)$ to achieve $\sqrt{n}$-consistency.

Some calculations show that the working independence kernel and profile partial likelihood estimators of $\{\beta(t), \gamma\}$ solve the following estimating equations.

Given γ, the working independence kernel estimating equations for estimating $\beta(t)$ is

$$0 = U_1(b,t;\gamma) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J_i}\int_0^{\tau} K_h(u-t)\Big\{\widetilde{X}_{ij}(u,t) - \frac{\sum_{l=1}^{n}\sum_{r=1}^{J_l} Y_{lr}(u)X_{lr}(u,t)e^{\widetilde{X}_{lr}(u,t)^T\mathbf{b}+Z_{lr}^T\gamma}}{\sum_{l=1}^{n}\sum_{r=1}^{J_l} Y_{lr}(u)e^{\widetilde{X}_{lr}(u,t)^T b+Z_{lr}^T\gamma}}\Big\}dN_{ij}(u). \tag{2.7}$$

We denote the kernel estimator of $\beta(t)$ by $\widehat{\beta}(t,\gamma)$.
The working independence profile estimating equation for γ is

$$\begin{aligned}0 &= U_2\{\gamma,\widehat{\beta}(u,\gamma)\} = \sum_{i=1}^{n} U_{2i}\{\gamma,\widehat{\beta}(u,\gamma)\}\\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J_i}\int_0^{\tau}\Big\{X_{ij}\widehat{\beta}_{\gamma}(u,\gamma)+Z_{ij}\\ &\quad -\frac{\sum_{l=1}^{n}\sum_{r=1}^{J_l} Y_{lr}(u)(X_{lr}\widehat{\beta}_{\gamma}(u,\gamma)+Z_{lr})e^{X_{lr}\widehat{\beta}(u,\gamma)+Z_{lr}^T\gamma}}{\sum_{l=1}^{n}\sum_{r=1}^{J_l} Y_{lr}(u)e^{X_{lr}\widehat{\beta}(u,\gamma)+Z_{lr}^T\gamma}}\Big\}dN_{ij}(u),\end{aligned} \tag{2.8}$$

where $\widehat{\beta}_{\gamma}(t,\gamma)$ is the derivative of $\widehat{\beta}(t,\gamma)$ with respect to γ.

The Newton-Raphson algorithm can be used to solve these estimating equations. To carry this out, one also has to calculate the estimator $\widehat{\beta}_{\gamma}(u,\gamma)$ and the matrix estimator $\widehat{\beta}_{\gamma\gamma}(u,\gamma)$, the first and second derivatives of $\widehat{\beta}(u,\gamma)$ with respect to γ, by solving the corresponding derivative equations of $U_1(\gamma,t)$ with respect to γ.

The covariance estimators for the kernel estimator $\widehat{\beta}(t)$ and the parametric regression coefficients $\widehat{\gamma}$ can be obtained using the sandwich method. Specifically, the sandwich estimator for the variance of $\widehat{\beta}(t)$ is similar to that given in Section 2.1 except that $X_{ij}\beta(t)$ needs to be replaced by $X_{ij}\beta(t)+Z_{ij}^T\gamma$. The sandwich estimator for $\widehat{\gamma}$ is $\Omega_{\gamma 1}^{-1}\Omega_{\gamma 2}\Omega_{\gamma 1}^{-1}$, where $\Omega_{\gamma 2}=\sum_{i=1}^{n} U_{2i}(\gamma,\widehat{\beta}(u,\gamma))^{\otimes 2}|_{\gamma=\widehat{\gamma}}$, and $\Omega_{\gamma 1}=\partial U_2(\gamma,\widehat{\beta}(u,\gamma))/\partial\gamma^T|_{\gamma=\widehat{\gamma}}$.

2.3. Estimation of the baseline hazard function

The baseline hazard can be estimated using the Breslow-Aalen type estimator for the cumulative baseline hazard $\Lambda_0(t)$. For the semiparametric model (2.4), after obtaining the kernel and profile estimators $\widehat{\beta}(t)$ and $\widehat{\gamma}$, the cumulative baseline hazard can be estimated by

$$\widehat{\Lambda}_0(t) = \sum_{i=1}^{n}\sum_{j=1}^{J_i}\frac{\Delta_{ij}1\{T_{ij}\le t\}}{\sum_{i'=1}^{n}\sum_{l=1}^{J_i} e^{X_{i'l}\widehat{\beta}(T_{i'l})+Z_{i'l}^T\widehat{\gamma}}1\{T_{i'l}\ge t\}}.$$

To estimate the baseline hazard $\lambda_0(t)$ itself, one can use a smoothing technique as suggested by many authors, such as Fan, Gijbels and King (1997), via

$$\widehat{\lambda}_0(t) = \int K^{\lambda}_{h'}(u-t)d\widehat{\Lambda}_0(u) = \frac{1}{h'}\sum_{i=1}^{n}\sum_{j=1}^{J_i}\frac{\Delta_{ij}K^{\lambda}((T_{ij}-t)/h')}{\sum_{i'=1}^{n}\sum_{l=1}^{J_i} e^{X_{i'l}\widehat{\beta}(T_{i'l})+Z_{i'l}^T\widehat{\gamma}}1\{T_{i'l}\geq t\}}, \tag{2.9}$$

where K^{λ} is a kernel function and h' is a bandwidth. The asymptotic properties of the baseline hazard estimator are complicated, and are not pursued here, as this is not our primary focus. The baseline hazard estimator for the nonparametric model (2.1) can be formed in a similar way.

3. Asymptotic Properties

In this section we study the asymptotic properties of the nonparametric kernel estimator in Section 3.1 and the profile-kernel estimators in Section 3.2. We further derive, for independent data, the semiparametric efficient score of parametric regression coefficients γ, and, we show the profile-kernel estimator $\widehat{\gamma}$ is semiparametric efficient in Section 3.3.

3.1. Nonparametric varying-coefficient models

We first consider the nonparametric model (2.1).Without loss of generality, assume an equal cluster size, $J_i = J$, unequal cluster sizes can be easily handled by including dummy observations with censoring times of 0. Let $P_{Nj}(t|x) = P(Y_j > t|X_j(t) = x)$ and
$Q_{Nr}(t) = E[\sum_{j=1}^{J} P_{Nj}(t|X_j(t))X_j(t)^{\otimes r}\exp\{X_j(t)^T\beta(t)\}\lambda_0(t)]$, $r = 0, 1, 2$. For simplicity, the subscript i is suppressed. Let

$$\Sigma_N(t) = Q_{N2}(t) - \frac{Q_{N1}(t)Q_{N1}(t)^T}{Q_{N0}(t)}. \tag{3.1}$$

Here the subscript N stands for the nonparametric model. We use subscript s to stand for the semiparametric model in the following subsection. Let $v_j = \int s^j K^2(u)du$, $j = 0, 1, 2$, $\beta_0(t)$ be a vector of true values of the time-dependent coefficients at time t. We introduce Conditions A to ensure Theorem 1.

Conditions A

1. *The kernel function $K(s)$ is a bounded symmetric function with a bounded support, bandwidth $h \to 0$, $nh \to \infty$, as $n \to \infty$, and $nh^5 = O(1)$.*
2. *The covariate $X_k(t)$ is bounded, and the time-dependent coefficient $\beta_k(t)$ satisfyies $sup_{t\in[0,\tau]}|\beta_k(t)| < B$ for a positive constant B, and has a continuous second derivative for all $k = 1, \ldots, K$.*

3. $\Sigma_N(t)$ *is positive definite for each* $t \in [0, \tau]$.
4. $P(Y_{ij}(t) = 1$, *for all* $t \in [0, \tau]) > 0$ *for each* i, j.

Denote by $\beta_0(t)$ the true value of $\beta(t)$. Theorem 1 states the asymptotic consistency and normality of the local linear kernel estimator of $\beta(t)$. The proof is in Appendix 1.

Theorem 1. *Under A, the working independence local kernel estimator* $\widehat{\beta}(t)$ *obtained by maximizing the local kernel partial likelihood* (2.2) *converges in probability to* $\beta_0(t)$, *and*

$$\sqrt{nh}\left\{\widehat{\beta}(t) - \beta_0(t) - \frac{h^2}{2}\mu_2\beta_0^{(2)}(t)\right\} \rightarrow N\{0, \Sigma_N(t)^{-1}v_0\} \tag{3.2}$$

in distribution, where $\mu_2 = \int u^2 K(u)du$, *and* $\Sigma_N(t)$ *is defined in* (3.1).

For independent data, $\Sigma_N(t)$ reduces to that of Cai et al. (2003), and the asymptotic results are the same as Theorem 2 there. One can easily show that the optimal bandwidth is of order $h = O(n^{-1/5})$. The asymptotic bias of $\widehat{\beta}(t)$ is of order h^2; the bias and rate of convergence under the optimal bandwidth are of order $n^{-2/5}$ and $n^{2/5}$, respectively.

3.2. Semiparametric varying-coefficient models

In this subsection, we study the asymptotic properties of the working independence kernel and profile estimators for the semiparametric model (2.4). Additional notation is needed. Let

$$Q_{sr}(t,\gamma) = E\Big[\sum_{j=1}^{J} Y_j(t)X_j^r e^{X_j\beta(t,\gamma)+Z_j^T\gamma}\lambda_0(t)\Big],\ r = 0, 1, 2,$$

$$\Sigma_s(t,\gamma) = Q_{s2}(t,\gamma) - \frac{Q_{s1}(t,\gamma)^2}{Q_{s0}(t,\gamma)}. \tag{3.3}$$

We use $\beta(t,\gamma), \beta_\gamma(t,\gamma)$ to denote the asymptotic limit of $\widehat{\beta}(t,\gamma), \widehat{\beta}_\gamma(t,\gamma)$, which are obtained by solving the first and second derivative of (2.7) with respect to γ. Denote by $\beta_0(t)$ and γ_0 the true values of $\beta(t)$ and γ. Regularity conditions are given in Conditions B in Appendix 2.1

Theorem 2. *Under B, the local linear kernel estimator* $\widehat{\beta}(t)$ *for* $\beta(t)$ *in model* (2.4), *obtained by maximizing the local partial likelihood* (2.5), *converges in probability to* $\beta_0(t)$, *and*

$$\sqrt{nh}\left\{\widehat{\beta}(t) - \beta_0(t) - \frac{h^2}{2}\mu_2\beta_0^{(2)}(t)\right\} \rightarrow N\{0, \Sigma_s(t,\gamma_0)^{-1}v_0\} \tag{3.4}$$

in distribution, where $\Sigma_s(t,\gamma)$ *is defined in* (3.3).

The proof of Theorem 2 is similar to Theorem 1 and hence is omitted. The estimator for the nonparametric component in the semiparametric model has the same convergence rate and bias as that of nonparametric model (2.1). The asymptotic variance is different since it involves the parametric coefficient γ_0. The bias term does not involve γ_0 since $\Sigma_s(t,\gamma_0)$ is canceled in the calculation.

The parametric component γ is often of greater interest in semiparametric model (2.4) for ease of interpretation. In order to obtain the asymptotic properties of $\widehat{\gamma}$, one has to study $\beta_\gamma(t,\gamma_0)$, the asymptotic limit of $\widehat{\beta}_\gamma(t,\gamma_0)$. The related calculations are presented in Appendix 3.

Theorem 3. *Under B, the working independence profile-kernel estimator $\widehat{\gamma}$ in Model* (2.4)*, obtained by solving the profile-kernel estimating equations* (2.8) *converges in probability to γ_0, and $\sqrt{n}(\widehat{\gamma}-\gamma_0)\rightarrow N\{0,\Sigma_{p2}^{-1}\Sigma_{p1}\Sigma_{p2}^{-1}\}$ in distribution, where*

$$\Sigma_{p2}=\int_0^\tau \Big[E\{(X_1\beta_\gamma(u,\gamma_0)+Z_1)^{\otimes 2}|T_1=u,\Delta_1=1\}$$

$$-\{E\{X_1\beta_\gamma(u,\gamma_0)+Z_1|T_1=u,\Delta_1=1\}\}^{\otimes 2}\Big]\sum_{j=1}^{J}E\{Y_j(u)e^{X_j\beta(u)+Z_j^T\gamma_0}\}\lambda_0(u)du,$$

$$\Sigma_{p1}=E\Big[\sum_{j=1}^{J}\int_0^\tau Z_j+X_j\beta_\gamma(u,\gamma_0)-E\{X_j\beta_\gamma(u,\gamma_0)+Z_j|T=u,\Delta=1\}dM_j(u)\Big]^{\otimes 2},$$

where the subscript i is suppressed.

We sketch the proof of Theorem 3 in Appendix 2. The results are consistent with Result 1.c in Lin and Carroll (2001) for the working independence estimators for non-censored data. No undersmoothing is required for the $\sqrt{n}$ rate convergence of $\widehat{\gamma}$. Condition B.7, which assumes the same marginal distribution of observations in one cluster, is necessary for the calculation of the asymptotic variance. We consider this assumption reasonable under the common baseline hazard model, and for the western Kenya parasitemia data.

3.3. The efficient score functions

In this section we are interested in studying the semiparametric efficiency bound of the parametric estimator for model (2.4), and whether the profile-kernel estimator $\widehat{\gamma}$ reaches the efficiency bound.

We first consider the situation when all observations are independent. We study the semiparametric efficient score functions and the efficiency bound of the semiparametric estimator. Using this result, we show that the profile-kernel estimator is semiparametric efficient. When observations are independent, we can show that the semiparametric efficient score for γ of a single observation is

$$U^*_\gamma(T,X,Z,\Delta) = \int_0^\infty [Z+X\beta_\gamma(u,\gamma_0)-E\{Z+X\beta_\gamma(u,\gamma_0)|T=u,\Delta=1\}]dM(u), \tag{3.5}$$

where

$$\beta_\gamma(u,\gamma_0) = -\frac{E\{XZ|T=t,\Delta=1\}-E\{X|T=t,\Delta=1\}E\{Z|T=t,\Delta=1\}}{E\{X^2|T=1,\Delta=1\}-E\{X|T=t,\Delta=1\}^2}.$$

The detailed calculations of $\beta_\gamma(u,\gamma_0)$ are given in Appendix 3. The proof for this semiparametric efficient score is similar to that of Sasieni (1992b) and is omitted. Some calculations show that the semiparametric efficiency bound of γ is

$$\begin{aligned}\nu_{\gamma\gamma} = \int_0^\infty \Big[&E\{(X\beta_\gamma(u,\gamma_0)+Z)^{\otimes 2}|T=u,\Delta=1\}\\ &-\{E[X\beta_\gamma(u,\gamma_0)+Z|T=u,\Delta=1]\}^{\otimes 2}\Big]E\{Y(u)e^{X\beta(u)+Z^T\gamma_0}\}\lambda_0(u)du.\end{aligned}$$

Theorem 4 states, for independent data, that the profile kernel estimator $\widehat{\gamma}$ is semiparametric efficient and its variance reaches the semiparametric efficiency bound. The proof is straightforward using Theorem 3 and $\Sigma_{p1} = \nu_{\gamma\gamma}$, and is omitted.

Theorem 4. *Under B, for independent data, $\Sigma_{p2} = \Sigma_{p1} = \nu_{\gamma\gamma}$, hence the profile-kernel estimator of $\widehat{\gamma}$ under model* (2.4) *reaches the semiparametric efficient bound.*

For correlated clustered data, calculations of the efficient score function are much more complicated. The parametric results of Cai and Prentice (1995) show that accounting for within-cluster correlation using weighted estimating equations often provides little efficiency except for the case when the correlation is very strong. We expect the working independence profile estimator of γ to have a similar property in semiparametric varying coefficient models, and hence recommend its use in practice, especially in view of its simplicity.

4. Simulation Study

In this section, we evaluate the finite sample performance of the proposed estimators for the semiparametric model (2.4) using simulations. We present results for both correlated data and independent data. Results for the nonparametric model are similar and are omitted.

We assumed that bivariate survival times (T_{i1},T_{i2}) follow the Clayton model

$$S(t_1,t_2;x_{ij},z_{ij}\phi) = \Big\{S_1(t_1;x_{i1},z_{i1})^{-1/\phi}+S_2(t_2;x_{i2},z_{i2})^{-1/\phi}-1\Big\}^{-\phi},$$

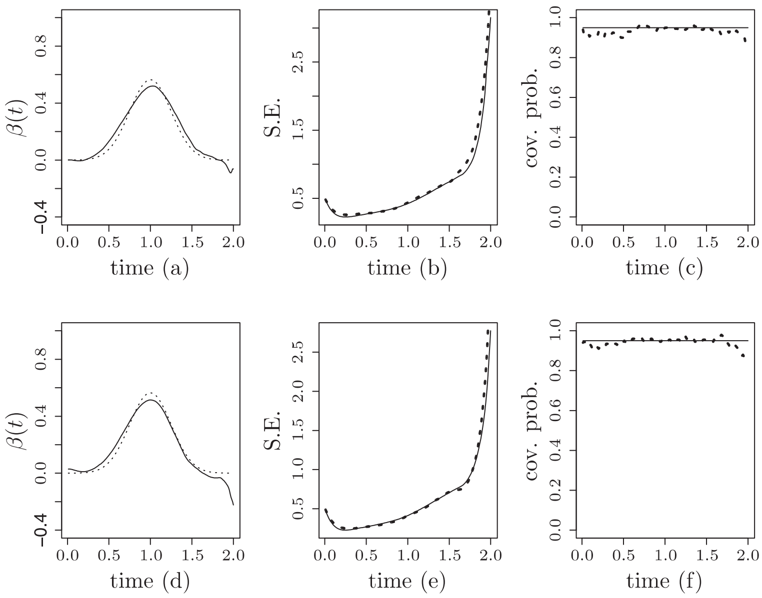

Figure 4.1. Estimation of $\beta(t)$ for the correlated semiparametric model. Top panel: correlated model; bottom panel: independent model. (a) and (c) Estimator of $\beta(t)$. – true value, $\cdots$ kernel estimator, bandwidth=0.35; (b) and (d) Empirical and estimated standard error. – empirical SE, $\cdots$ estimated SE; (c)and (f) empirical coverage of 95% confidence intervals.

where the marginal survival function is

$$S_j(t; x_{ij}, z_{ij}) = \exp\{-\int_0^t \lambda_{ij}(u; x_{ij}, z_{ij})du\}, \tag{4.1}$$

and $\lambda_{ij}(t; x_{ij}, z_{ij}) = \lambda_0(t)\exp\{x_{ij}\beta(t) + z_{1ij}\gamma_1 + z_{2ij}\gamma_2\}$. The correlation between T_{i1} and T_{i2} decreases as ϕ increases. The time-varying coefficient was set as $\beta(t) = \pi^{-1/2}\exp\{-16(t-1)^2/2\}$, proportional to the normal density function with mean 1 and variance 1/4. This unimodal bell shape function is shown in Figure 4.1 (a). The baseline hazard was $\lambda_0(t) = 0.8$. The covariate x_{ij} was generated as i.i.d. $Uniform(0,1)$. The covariate z_1 was generated as a binary random variable with $Pr(z_1 = 1) = 0.5$ and correlated with x_{ij} with correlation about 0.5. The covariate z_2 was generated as an i.i.d. normal random variable with mean 0 variance 0.25, and independent of other covariates. The true coefficients were $(\gamma_1, \gamma_2) = (1, 0.5)$. To generate (T_{i1}, T_{i2}) under the Clayton model, we first generated u_{i1}, u_{i2} independently from $Uniform(0,1)$, and then calculated $T_{i1} = S_1^{-1}(u_{i1}; x_{i1}, z_{i1})$. We evaluated the inverse function S_1^{-1} numerically. Then T_{i2} was generated using

$$T_{i2} = S_2^{-1}[\{S_1(T_{i1})^{-1/\phi}u_{i2}^{-1/(1+\phi)} + 1 - S_1(T_{i1})^{-1/\phi}\}^{-\phi}],$$

Table 4.1. Simulation results for the parametric regression coefficient estimator

model	parameter	true	estimator	empirical SE	estimated SE	coverage
correlated	γ_1	1	1.014	0.127	0.129	95.5%
($\phi = 0.5$)	γ_2	0.5	0.512	0.094	0.102	96.5%
independent	γ_1	1	0.995	0.117	0.127	98.0%
	γ_2	0.5	0.494	0.102	0.099	95.0%

which can be derived from the conditional distribution of T_{i2} given T_{i1}. We set $\phi = 0.5$, which corresponds to a strong correlation. Censoring times were generated from i.i.d. $\sim \exp(0.7)$, and independent from failure times and covariates. The maximum follow-up time was 2, and the censoring percentage was about 20%. We generated 200 data sets with each having 250 clusters.

We took bandwidths $h = 0.25, 0.30, 0.35$. The results presented here are for $h = 0.35$, the results for $h = 0.25, 0.30$ are similar and are omitted. We used the Epanechnikov kernel density. The top panel of Figure 4.1 (a) compares the mean function of the estimated time-dependent coefficient of $\beta(t)$ with its true value. The estimated function $\widehat{\beta}(t)$ was close to the true function, showing that our kernel method worked well. Figure 4.1 (b) shows the point-wise empirical standard error close the point-wise mean estimated standard error. Except for the region close to the maximum follow-up time, which results from a small number of events; the sandwich method worked well. Figure 4.1 (c) shows the empirical coverage probabilities of the point-wise 95% confidence intervals constructed using the sandwich estimator, and very close to the 95% nominal value. The mean coverage probability over the time interval $[0, 2]$ was 93.2%.

Table 4.1 gives the simulation results for the profile-kernel estimators of the parametric regression coefficients γ_1, γ_2. The average point estimates were very close to their true values, the estimated standard errors were very close to the empirical standard errors. The empirical coverage probabilities of the 95% confidence intervals were 95.5% for $\widehat{\gamma}_1$ and 96.5% for $\widehat{\gamma}_2$.

We ran another set of simulations for independent data with the same marginal distributions and the same censoring time distributions. One hundred data sets were generated with 500 observations in each. The bottom panel of Figure 4.1 shows a similar pattern of the time-dependent coefficient kernel estimate $\widehat{\beta}(t)$ as in the correlated model. The mean coverage probability of the 95% confidence intervals was 93.4%. Table 4.1 shows the results for the profile-kernel parametric coefficient estimate $(\widehat{\gamma}_1, \widehat{\gamma}_2)$. These results are similar to the correlated data case and show that the profile-kernel estimates worked well with little bias, and the estimated standard errors agreed with their empirical counterparts.

We ran additional simulations with different levels of censoring and correlation, and binary X_{ij}. The results were similar and are omitted.

5. Application

We applied the proposed method to the analysis of the western Kenya parasitemia data set (McElroy et al. (1997)) that was introduced in Section 1. This study enroled 607 children. 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. Children were examined at 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. They were then followed up for 10 weeks for the time to parasitemia infection. Observations from the same family were likely to be correlated due to similar environmental and genetic factors. We are interested in studying the risk factors associated with time to the onset of parasitemia. The risk factors of interest included baseline parasitemia density, age, gender, and daily mean of mosquito bites. Following McElroy et al. (1997), we log-transformed the baseline parasitemia density and denoted the new variable by $LNBPD$. The daily mean of mosquito bites was also quartic-root transformed, and denoted by covariate $BITE$.

Preliminary examination of the $LNBPD$ effect in Figure 5.2 shows a nonlinear effect, while the other covariate effects were found to be linear. We hence considered a semiparametric varying coefficient model, where the effect of $LNBDP$ is allowed to be time-varying and the effects of $BITE$, AGE, GENDER to be parametric. We fit the model using the proposed working independence kernel profile method. We examined several choices of the bandwidth for estimating $\beta(t)$ and found $h = 20$ fit the data well.

Figure 5.2 shows $LNBPD$ significantly positively associated with the risk of parasitemia before day 26; after Day 30, the LNBPD effect plateaued at a level close to 0 and was not significant. This result is not surprising since the effect of baseline parasitemia density on the time of onset of parasitemia is likely to diminish over time. Table 5.2 shows the parametric regression coefficient estimates. The dose of mosquito bites significantly increased the risk of parasitemia. Older children had a higher risk of developing parasitemia. Gender has no significant effect.

6. Discussion

In this paper, we propose estimation in semiparametric time-varying coefficient models for clustered survival data using kernel and profile methods under

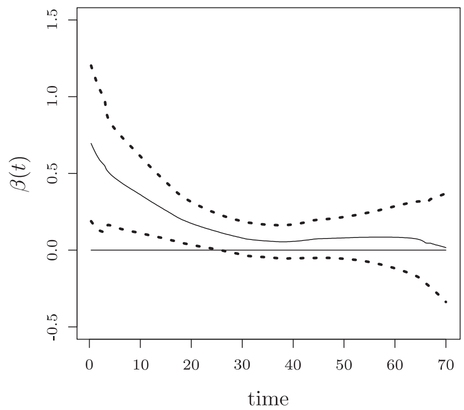

Figure 5.2. The kernel estimator and 95% confidence interval of the varying coefficient function of LNBPD for the western Kenya parasitemia data, bandwidth $h = 25$

Table 5.2. Parametric covariates' effects for the western Kenya parasitemia data.

Covariate	BITE	AGE (Month)	GENDER
$\hat{\gamma}$ *(SE)*	0.3221(0.1085)	0.0362 (0.0253)	0.0285 (0.0807)

working independence. We show that the profile-kernel estimator of the parametric regression coefficients is consistent and $\sqrt{n}$-consistent. We further show that it is semiparametric efficient when data are independent. The simulation results suggest that the proposed method performs well in finite samples.

For semiparametric models with independent data, Tian et al. (2005) proposed an estimator $\hat{\gamma} = \int_{h_n}^{\tau-h_n} w(s)\hat{\gamma}(s)ds / \int_{h_n}^{\tau-h_n} w(s)ds$, where h_n is the bandwidth and $w(t)$ is the inverse of the upper left submatrix of $I^{-1}(\beta(t), \gamma(t))$. They showed that $\hat{\gamma}$ is semiparametric efficient. Although this is an non-iterative estimator, cross-validation is often time consuming, perhaps more so than the profile estimator. No simulation was performed to evaluate the finite sample performance of this estimator. We expect that undersmoothing might be needed for estimating $\beta(t)$ and the integrated γ estimator might be subject to a larger finite sample variance. It is of future interest to compare the small sample performance of these estimates. It is also worth noting that, in our simulation, it only took about 5 to 6 iterations between steps (2.5) and (2.6) to converge which is not an insurmountable increase.

It is important to develop a data driven bandwidth selection tool, though cross validation method is more commonly used. For the nonparametric regression of time-dependent coefficient for survival data, little work has been done on data-driven methods for choosing an optimal bandwidth. Tian et al. (2005)

proposed a a K-fold cross validation method, but its performance was not evaluated.

One might consider improving the efficiency of the nonparametric kernel estimator $\beta(t)$ by extending the seemingly unrelated (SUR) kernel method (Wang (2003)) to censored data, and improving the efficiency of the parametric regression parameter γ by incorporating rates in a similar spirit to Cai and Prentice (1995). The parametric regression results of Cai and Prentice (1995) and Gray and Li (2002) for clustered survival data show that the improvement in efficiency by incorporating weights in estimating equations is often small. Alternative approaches are likely to be needed.

References

Ahmad, I., Leelahanon, S. and Li, Q. (2005). Efficient estimation for a semiparametric partial linear varying coefficient model. *Ann. Statist.* **33**, 258-283.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100-1120.

Cai, J., Fan, J., Zhou, H. and Zhou, Y. (2007). Hazard models with varying-coefficients for multivariate failure time data. *Ann. Statist*, **35** (to appear).

Cai, J. and Prentice, R. L. (1995). Estimating equations for hazard ratio parameters based on correlated failure time data. *Biometrika* **82**, 151-164.

Cai, Z. and Sun, Y. (2003). Local linear estimation for time-dependent coefficients in Cox's regression model. *Scand. J. Statist.* **30**, 93-111.

Carothers, N. L. (2000). *Real Analysis.* Cambridge Press.

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

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

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

Fan, J., Lin, H. and Zhou, Y. (2006). Local partial-likelihood estimation for lifetime data. *Ann. Statist.* **34**, 290-325.

Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with application to breast cancer prognosis. *J. Amer. Statist. Assoc.* **87**, 942-951.

Gray and Li (2002). Optimal weights for marginal proportional hazards analysis of clustered failure time data. *Lifetime Data Analysis***8**, 4-19.

Hastie, T. J. and Tibshirani, R. (1993). Varying-coefficient models(with discussion). *J. Roy. Statist. Soc. B.* **55**, 757-796.

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

McElroy, P. D., Beier, J. C., Oster, C. N., Onyango, F. K., Lin, X., Beedle, C. and Hoffmann, 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.

Sasieni, P. (1992a). Information bounds for conditional hazard ratio in a nested family of regression models. *J. R. Statist. Soc. B* **54**, 617-635.

Sasieni, P. (1992b). Information bounds for the additive and multiplicative intensity models. *In Survival Analysis: State of Art, Kluwer Academic Publishers Group*, 249-263.

Spiekerman, C. F. and Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *J. Am. Statist. Assoc.* **93**, 1164-1175.

Tian, L., Zucker, D., and Wei, L. J. (2005). On the Cox model with time-varying coefficients. *J. Am. Statist. Assoc.* **100**, 172-183.

Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika* **90**, 43-52.

Winnett, A. and Sasieni, P. (2003). Iterated residuals and time-varying covariate effects in Cox regression. *J. Roy. Statist. Soc. Ser. B.* **65**, 473-488.

Yu, Z. and Lin, X. (2008). Nonparametric regression using local kernel estimating equations for correlated failure time data. *Biometrika* **95**, 123-137.

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

Division of Biostatistics, Indiana University School of Medicine, Indianapolis, IN 46202, U.S.A.

E-mail: yuz@iupui.edu

Department of Biostatistics, Harvard School of Public Health, , 677 Huntington Avenue, Boston, MA 02115, U.S.A.

E-mail: xlin@hsph.harvard.edu

(Received April 2008; accepted February 2009)