12-30-2005

Semiparametric Approaches for Joint modeling of Longitudinal and Survival Data with Time Varying Coefficients

Xiao Song
University of Washington, songx@u.washington.edu

C.Y. Wang
Fred Hutchinson Cancer Research Center, cywang@fhcrc.org

1. INTRODUCTION

In clinical trials, information is often collected on a time-to-event (e.g. "survival") and time-dependent and time-independent covariates. An example is given by AIDS Clinical Trials Group (ACTG) Protocol 175, a randomized trial to compare zidovudine alone, zidovudine plus didanosine, zidovudine plus zalcitabine, or didanosine alone, in HIV-infected subjects on the basis of time to progression to AIDS or death (Hammer et al., 1996). Between December 1991 and October 1992, 2467 subjects were recruited and followed until November 1994. CD4 count, as a reflection of immune status, was measured for each participant about every 12 weeks after randomization. It is well known that observations of CD4 count are subject to substantial biological variation and measurement error.

In survival analysis, a routine objective is to characterize the relationship between survival and the covariates. Standard inference procedures usually require the true values of the covariates at the event times. However, covariates like CD4 count may be subject to substantial measurement error. In addition, longitudinal covariates are usually collected intermittently and may not observed at the event times. Naive approaches that ignore measurement error might lead to biased estimation and misleading inference (Prentice, 1982; Tsiatis and Davidian 2001; Song and Huang 2005b). A popular approach to dealing with the aforementioned measurement error is to use a joint model, which assumes that the longitudinal observations follow a mixed effects model and the survival time depends on the random effects of the mixed effects model through a survival model. The mixed effects model may be viewed as an extension of the standard additive measurement error model for time-independent covariates and the joint modeling approaches may be applied to repeated measured time-independent covariates as well (Wang, Wang and Wang, 2000).

In joint modeling, the most widely used survival model is the proportional hazards model. Various approaches have been proposed under this framework, including the regression calibration (e.g. Pawitan and Self, 1993; Tsiatis et al., 1995; Bycott and Taylor, 1998; Dafni and Tsiatis, 1998), likelihood based

approaches (e.g. DeGruttola and Tu, 1994; Wulfsohn and Tsiatis, 1997; Faucett and Thomas, 1996; Henderson et al., 2000; Xu and Zeger, 2001a; Song, Davidian and Tsiatis, 2002b), corrected score (Wang, 2005) and conditional score (Tsiatis and Davidian, 2001; Song, Davidian and Tsiatis, 2002a) approaches. However, the proportional hazards assumption may be too restrictive in practice. For example, in AIDS studies, it is well known that the effect of the anti-retrovival treatments may decay after some time. More flexible models are needed to characterize such covariate effects.

An appealing alternative is the time-varying-coefficient proportional hazards model, which allows the effect of coefficients to vary over time. In the case of no measurement error, this model has been studied by Zucker and Karr (1990) using the penalized partial likelihood method and by Murphy and Sen (1991) using the histogram sieve method. However, both approaches involve complicated optimization procedures over a high dimensional parameter space. Recently, Winnett and Sasieni (2003) took an iterated residual approach based on the Schoenfeld residuals for the standard proportional hazards model. They showed the estimator is consistent, but did not provide the asymptotic distribution. One appealing alternative is the local partial likelihood approach proposed by Cai and Sun (2003). The estimator is shown to be consistent and asymptotically normal. In addition, it is the maximum point of a concave function and is thus simple to compute. Tian, Zucker and Wei (2005) investigated a similar estimating procedure and proposed a resampling method to construct confidence bands for the time-varying coefficients over a properly selected time interval. However, at our best knowledge, no approach exists to dealing with measurement error or joint modeling under the time-varying-coefficient proportional hazards framework.

In this paper, based on the local partial likelihood method, we propose two approaches for the varying coefficient proportional hazards model when longitudinal covariates are measured with error. One is a corrected score approach, and the other is a conditional score approach. Both approaches require no distributional assumptions on the underlying true covariates. The asymptotic properties of the estimators

are derived based on martingale and empirical process theories. Since time-independent covariates with repeated measurements can be viewed as a special case of longitudinal covariates, this subsumes the case of time-independent covariates measured with error. We organize the paper as follows. In Section 2, we give the model definition. The corrected score and conditional score estimators are proposed in Sections 3 and 4. The finite sample performances of the estimators are assessed by simulation studies in Section 5 and illustrated by an application to the ACTG 175 data in Section 6. We conclude with discussions in Section 7.

2. MODEL DEFINITION

Let T denote the failure time and C denote the censoring time. The observed survival data are $V = \min(T, C)$, and $\Delta = I(T \leq C)$, where $I(\cdot)$ is the indicator function. Let $\boldsymbol{X}(u) = \{X_1(u), \ldots, X_p(u)\}$ denote p covariates at time u, which include both time-dependent and time-independent covariates. Suppose that the covariate process $X_k(u)$ is not observed directly; rather, longitudinal measurements of $X_k(u)$, $W_k = (W_{k1}, \ldots, W_{km_k})^T$ are observed at times $t_k = (t_{k1}, \ldots, t_{km_k})$. For time-independent covariates that are exactly measured, $m_k = 1$ and $W_k = X_k$. Let $\boldsymbol{W} = (W_1^T, \ldots, W_p^T)^T$, $\tau = (t_1, \ldots, t_p)$, $m = (m_1, \ldots, m_p)$. Suppose $\{(T_i, C_i, V_i, \Delta_i, \boldsymbol{W}_i, \boldsymbol{X}_i, \tau_i, m_i) : i = 1, \ldots, n\}$ are independent and identically distributed samples of $(T, C, V, \Delta, \boldsymbol{W}, \boldsymbol{X}, \tau, m)$ and the observed data set is $\{(V_i, \Delta_i, \boldsymbol{W}_i, \tau_i, m_i) : i = 1, \ldots, n\}$.

Assume that the longitudinal covariate processes follow the linear mixed effects models,

$$X_{ik}(u) = \alpha_{ik}^T f_k(u),$$

$$W_{ikj} = X_i(t_{ikj}) + e_{ikj},$$

where $f_k(u)$ is a known q_k-dimensional function of u, and α_{ik} is a q_k-dimensional random effect, $j = 1, \ldots, m_k$, and f_k and α_{ik} may be different for each k, $k = 1, \ldots, p$. This allows flexible modeling of the time trajectory of each covariate via polynomial or spline models. The random effects α_{ik} may be correlated

3

across k. In fact, no distributional assumption is placed on $\alpha_i = (\alpha_{i1}^T, \ldots, \alpha_{ip}^T)^T$ $(q \times 1)$, $q = \sum_k q_k$, nor is one needed. For time-independent covariates, α_{ik} is a scalar and $f_k(u) = 1$.

The errors e_{ikj} are assumed to be normally-distributed with mean zero and variance σ_{kk}^2 that may reflect both biological variation and measurement error. For time-independent covariates measured with no error, $e_{ikj} = 0$. For simplicity, we assume that the available measurements are sufficiently separated in time that serial correlation associated with within-subject biological variation is negligible; however, this assumption can be relaxed as discussed in Section 7. We allow measurements on different covariates at the same time to be correlated. More formally, we may write for $k, k' = 1, \ldots, p$, $j = 1, \ldots, m_{ik}$, and $j' = 1, \ldots, m_{ik'}$, $\mathrm{cov}(e_{ikj}, e_{ik'j'}) = \sigma_{kk'}I(t_{ikj} = t_{ik'j'})$. Here $\sigma_{kk'}$ is the covariance between errors from covariates k and k' measured at the same time point, reflecting correlation of components of within-subject biological variation, the measurement error, or both; this formulation subsumes the case $\sigma_{kk'} = 0$ for all $k \neq k'$. Let $e_i = (e_{i1}^T, \ldots, e_{iK}^T)^T$, where $e_{ik} = (e_{ik1}, \ldots, e_{ikm_{ik}})^T$. We assume that e_i is independent of $(T_i, C_i, \alpha_i, t_i)$.

A time-varying-coefficient proportional hazards model is assumed for the relationship between the hazard of failure and the covariates, under which the hazard for subject i equals

$$
\begin{aligned}
\lambda_i(u|X) &= \lim_{du \to 0} du^{-1} \Pr(u \leq T_i < u + du | T_i \geq u, \alpha_i, C_i, t_i(u), e_i(u)) \\
&= \lambda_0(u) \exp\left\{\beta_0^T(u) X_i(u)\right\}.
\end{aligned}
\tag{1}
$$

Here $\lambda_0(u)$ is an unspecified baseline hazard; $\beta_0(u)$ is a $p \times 1$ vector of regression parameters; $t_i(u) = (t_{ikj} < u; k = 1, \ldots, p)$ denotes the observation times before u; and $e_i(u) = \{e_{ikj} : t_{ikj} < u, k = 1, \ldots, p, j = 1, \ldots, m_{ik}\}$. Model (1) is different from the standard proportional hazards model in the regression coefficients, which may vary over time. This equation makes explicit the assumption that censoring, timing of measurements, and covariate measurement errors are noninformative. Our interest focuses on estimation

4

of $\beta_0(u)$.

3. CORRECTED SCORE ESTIMATOR

For any given $t > 0$, let $\beta(u) \approx \beta(t) + \beta'(t)(u - t)$ be the linear approximation of $\beta(u)$ at t, where

$\beta'(t) = d\beta(t)/dt$. Let $b_T = (b_{T0}^T, b_{T1}^T)^T = \{\beta_0^T(t), \beta_0'^T(t)\}^T$, and $b = (b_0^T, b_1^T) = \{\beta^T(t), \beta'^T(t)\}^T$. For any

vector c, define $c^{\otimes r} = 1, c, cc^T$ for $c = 0, 1, 2$ respectively. When the true covariate processes $X_i(u)$ are

known, we consider the local estimating function for b_T,

$$\tilde{U}_I(b) = (nH)^{-1} \sum_{i=1}^{n} \int_0^L K_h(u - t) \left\{ X_i(u, u - t) - \frac{\tilde{G}_{I,1}(b, u, u - t)}{\tilde{G}_{I,0}(b, u, u - t)} \right\} dN_i(u). \tag{2}$$

Here L is a fixed time; $H = diag(I_p, hI_p)$, where I_p denotes a $p \times p$ identity matrix, h is the bandwidth;

$K_h(u) = h^{-1}K(h^{-1}u)$, where $K(\cdot)$ is a kernel density; $X_i(u, u - t) = u_t \otimes X_i(u)$, where $\otimes$ is the operator

for Kronecker product and $u_t = (1, u - t)^T$; For $r = 0, 1, 2$, $\tilde{G}_{I,r}(b, u, u - t) = n^{-1} \sum_{i=1}^{n} \tilde{G}_{I,ri}(b, u, t)$, where

$\tilde{G}_{ri}(b, u, t) = Y_i(u) X_i^{\otimes r}(u, u - t) \exp\left\{ b^T X_i(u, u - t) \right\}$; $N_i(u) = I(V_i \leq u, \Delta_i = 1, m_{ik}(u) \geq q_k, k = 1, \ldots, p)$

is the counting process, $Y_i(u) = I(V_i \geq u, m_{ik}(u) \geq q_k, k = 1, \ldots, p)$ is the at risk process, and $m_{ik}(u)$ is the

number of the observations before time u for the kth covariate. Estimating function (2) is similar to Cai and

Sun's local partial likelihood score function (Cai and Sun, 2003), only with the replacements of $N_i^*(u) = $

$I(V_i \leq u, \Delta_i = 1)$ by $N_i(u)$ and $Y_i^*(u) = I(V_i \geq u)$ by $Y_i(u)$. Noting that $N_i(u) = N_i^*(u)I\{m_{ik}(u) \geq$

$q_k, k = 1, \ldots, p\}$ and $Y_i(u) = Y_i^*(u)I\{m_{ik}(u) \geq q_k, k = 1, \ldots, p\}$, it is easy to see that (2) is a weighted

local partial likelihood score function with predictable weights $I\{m_{ik}(u) \geq q_k, k = 1, \ldots, p\}$. Using these

weights facilitates the construction of a corrected score, as will become clear below. We call the estimator

obtained from the equation $\tilde{U}_I(b) = 0$, say $\hat{b}_I$, the ideal estimator since it is unachievable when $X_i(u)$ is

not observed.

In practice, the true covariate process $X_i(u)$ is unknown. Intuitively, a naive estimator can be obtained

by replacing $X_i(u)$ by its ordinary least square estimate in (2). However, the naive estimator is biased

as indicated by the simulation studies in Section 5. One useful technique is the corrected score method (Stefanski, 1989; Nakamura, 1990), which removes the bias through correction of the "score" function.

To implement the corrected score approach, we substitute $X_{ik}(u)$ by its ordinary square estimator $\hat{X}_{ik}(u)$ based on the longitudinal observations before time u rather than on all the longitudinal observations as in naive regression. This ensures the predictability of $\hat{X}_{ik}(u)$, which is important for the consistency of the corrected score estimator. The same technique has been used for the conditional and corrected score approaches under the standard proportional hazards framework (Tsiatis and Davidian, 2001; Song et al., 2002a; Wang, 2005). The estimation of $\hat{X}_{ik}(u)$ is possible only if $m_{ik}(u) \geq q_k$, which clarifies the inclusion of this condition in the definitions of $Y_i(u)$ and $N_i(u)$. Define $\omega = \{\sigma_{kk'} : k \geq k'\}$, the distinct parameters characterizing the variances and covariances of the errors. For now, we assume ω is known. Let $\hat{X}_i(u) = \{\hat{X}_{i1}^T(u), \ldots, \hat{X}_{ip}^T(u)\}^T$. For $k = 1, \ldots, p$, define $F_{ik} = [f_k(t_{ik1}), \ldots, f_k(t_{ikm_{ik}})]^T$, $(m_{ik} \times q_k)$, and let 0_{q_k} be a $q_k \times 1$ vector of zeros, and $I_{ikk'}$ be the $(m_{ik} \times m_{ik'})$ matrix whose (j, j') entry is $I(t_{ikj} = t_{ik'j'})$, for $j = 1, \ldots, m_{ik}$ and $j' = 1, \ldots, m_{ik'}$. Let

$$
F(u) = \begin{bmatrix} f_1^T(u) & 0_{q_2}^T & \cdots & 0_{q_p}^T \\ 0_{q_1}^T & f_2^T(u) & \cdots & 0_{q_p}^T \\ \vdots & \vdots & \ddots & \vdots \\ 0_{q_1}^T & 0_{q_2}^T & \cdots & f_p^T(u) \end{bmatrix}, \quad (p \times q).
$$

Since $X_i(u) = F(u)\alpha_i$, following Song et al. (2002a), conditional on $\{\alpha_i, t_i\}$, the covariance of $\hat{X}_i(u)$ is equal to $\Sigma_i(u) = F(u)D_i(\omega)F^T(u)$, where

$$
D_i(\omega) = \begin{bmatrix} D_{i11}(\omega) & D_{i12}(\omega) & \cdots & D_{i1p}(\omega) \\ D_{i21}(\omega) & D_{i22}(\omega) & \cdots & D_{i2p}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ D_{ip1}(\omega) & D_{ip2}(\omega) & \cdots & D_{ipp}(\omega) \end{bmatrix}, \quad (q \times q),
$$

6

and $D_{ikk'}(\omega) = \sigma_{kk'}\{F_{ik}^T F_{ik}\}^{-1} F_{ik}^T I_{ikk'} F_{ik'} \{F_{ik'}^T F_{ik'}\}^{-1}, (q_k \times q_{k'})$.

We now derive the corrected score estimator. Suppose h is fixed. When n goes to infinity, by the functional law of large number (Andersen and Gill, 1982), the weighted local partial likelihood score function $\tilde{U}_I(b)$ converges in probability to

$$\tilde{U}_I(b) = H^{-1} \int_0^L K_h(u-t) \left[E\{X_i(u, u-t)\} - \frac{E\left\{\tilde{G}_{I,1i}(b, u, u-t)\right\}}{E\left\{\tilde{G}_{I,0i}(b, u, u-t)\right\}} \right] dE\{N_i(u)\}.$$

Let $\hat{X}_i(u, u-t) = u_t \otimes \hat{X}_i(u)$, $\Sigma_i(u, u-t) = var\left\{\hat{X}_i(u, u-t)|\alpha_i, t_i(u)\right\} = (u_t u_t^T) \otimes \Sigma_i(u)$, and

$$\hat{G}_{CR,ri}(b, u, u-t) = Y_i(u)\hat{X}_i^{\otimes r}(u, u-t) \exp\left\{ b^T \hat{X}_i(u, u-t) - \frac{1}{2} b^T \Sigma_i(u, u-t)b \right\}$$

for $r = 0, 1, 2$. With some algebra, we can show that

$$\frac{E\left\{\hat{G}_{CR,1i}(b, u, u-t)\right\}}{E\left\{\hat{G}_{CR,0i}(b, u, u-t)\right\}} = \frac{E\left\{\tilde{G}_{I,1i}(b, u, u-t)\right\}}{E\left\{\tilde{G}_{I,0i}(b, u, u-t)\right\}} + E\left\{\Sigma_i(u, u-t)b\right\}.$$

This suggests that the bias of the naive estimating function can be removed by adding a consistent estimator of $E\left\{\Sigma_i(u, u-t)b\right\}$. Thus the local corrected score estimating equation is

$$
\begin{aligned}
\hat{U}_{CR}(b) &= (nH)^{-1} \sum_{i=1}^n \int_0^L K_h(u-t)\bigg\{ \hat{X}_i(u, u-t) + \Sigma_i(u, u-t)b \\
&\qquad - \frac{\hat{G}_{CR,1}(b, u, u-t)}{\hat{G}_{CR,0}(b, u, u-t)} \bigg\} dN_i(u) = 0,
\end{aligned}
\tag{3}
$$

where $\hat{G}_{CR,r}(b, u, u-t) = n^{-1} \sum_{i=1}^n \hat{G}_{CR,ri}(b, u, u-t)$ $(r = 0, 1, 2)$. In fact, when h goes to 0 at a certain rate, we can also show that $\tilde{U}_I(b)$ and $\tilde{U}_{CR}(b)$ still converge to a common limit. The arguments are sketched in Appendix A. When there is no measurement error, the corrected score estimating equation (3) reduces to the weighted local partial likelihood score equation (2).

Cai and Sun (2003) showed the consistency and asymptotic normality of the local partial likelihood estimator based on the concavity of the local partial likelihood and the martingale theory. However, unlike the local partial likelihood score function $\hat{U}_I(b)$, the corrected score estimating function $\hat{U}_{CD}(b)$ in (3) is

not the derivative of a concave function. Therefore the arguments of Cai and Sun (2003) cannot be applied directly for the corrected score estimator. Here, using both the martingale and empirical process theories, we derive their asymptotic properties under the regularity conditions stated in Appendix A, which are slightly stronger than those given in Cai and Sun (2003). Write $\mu_r = \int u^r K(u) du$, and

$$Q_\mu = \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix}.$$

Let $\beta''(t)$ be the second derivative of $\beta(t)$. A point $t \in [0, L]$ is called an interior point if $t \in [h, L - h]$. The results are summarized in the following theorems with the arguments sketched in Appendix A.

Theorem 1. Under Conditions A1–A10, as $n \to \infty$, a solution to the corrected score estimating equation (3), say $\hat{b}$, exists uniquely in a neighborhood of b_T with probability 1. Moreover, $H(\hat{b} - b_T)$ converges in probability to 0.

Theorem 2. Under Conditions A1–A10, when t is an interior point, as $n \to \infty$,

$$(nh)^{1/2} \left[H\left(\hat{b} - b_T\right) - \frac{1}{2}h^2 \left\{ Q_\mu^{-1} \left(\mu_2, \mu_3\right)^T \right\} \otimes \beta_0''(t) \right]$$

is asymptotically normal with mean zero and the variance can be consistently estimated by $\hat{\Gamma}^{-1}(t)\hat{\Omega}(t, \hat{b})\hat{\Gamma}^{-1}(t)$, where $\hat{\Gamma}(t)$ and $\hat{\Omega}(t, b)$ are given in Appendix A. The asymptotic variance is larger than that of the ideal estimator $\hat{b}_I$.

By analogy to Cai and Sun (2003), we can derive the theoretically optimal bandwidth based on Theorem 2, which minimizes the asymptotical mean integrated square error. However, this depends on the unknown quantities such as $\beta''(t)$ and the asymptotic variance. Future research is needed to develop appropriate bandwidth selection methods.

8

When t is a boundary point on $[0, L]$, similar results exist as those in Theorem 2. Specifically, for $t = ch$, $0 < c < 1$, μ_r is replaced by $\mu_{r,c} = \int_{-c}^{1} uK(u)du$; for $t = (L - ch)$, μ_r is replaced by $\mu_{r,c}^* = \int_{-1}^{c} uK(u)du$. The proofs are similar and omitted for brevity.

4. CONDITIONAL SCORE ESTIMATOR

An alternative technique to deal with measurement error is the conditional score method. The key idea of the conditional score approach is to "condition away" the nuisance random effects based on some "sufficient statistics." Given $Y_i(u) = 1$, α_i and $t_i(u)$, the conditional distribution of $dN_i(u)$ is Bernoulli with the probability $\lambda_0(u)du \exp\{\beta^T(u)X_i(u)\}$, which can be approximated by $\lambda_0(u)du \exp\{b^T X_i(u, u - t)\}$ through the linear expansion of $\beta_0(u)$ at t. Then, with similar arguments as those given by Tsiatis and Davidian (2001) and Song and Huang (2005a), we can show that the "sufficient statistic" for the nuisance parameter α_i is $S_i(b, u, u - t) = \hat{X}_i(u, u - t) + \Sigma_i(u, u - t)bdN_i(u)$. Conditional on this sufficient statistics, the conditional intensity process can be written as

$$
\begin{aligned}
\lambda_i(u|s, t) &= \lim_{du \to 0} du^{-1} \Pr\{dN_i(u) = 1 | S_i(b, u, u - t) = s, t_i(u), Y_i(u)\} \\
&\approx \lambda_0(u)\hat{G}_{CD,0i}(b, u, u - t).
\end{aligned}
$$

Here

$$
\hat{G}_{CD,ri}(b, u, u - t) = Y_i(u)S_i^{\otimes r}(b, u, u - t) \exp\left\{b^T S_i(b, u, u - t) - \frac{1}{2}b^T \Sigma_i(u, u - t)b\right\}
$$

for $r = 0, 1$. Thus we may consider the following estimating equations,

$$
\sum_i^n \int_0^L K_h(u, u - t)S_i^T(b, u, u - t)\{dN_i(u) - \lambda_0(u)\hat{G}_{CD,0i}(b, u, u - t)du\} = 0, \tag{4}
$$

$$
\sum_i^n \{dN_i(u) - \lambda_0(u)\hat{G}_{CD,0i}(b, u, u - t)du\} = 0. \tag{5}
$$

Solving for $\lambda_0(u)$ from (5) and substituting it into equation (4), we get the local conditional score estimating equation,

$$\hat{U}_{CD}(b) = (nH)^{-1} \sum_{i=1}^{n} \int_0^L K_h(u-t) \left\{ S_i(b, u, u-t) - \frac{\hat{G}_{CD,1}(b, u, u-t)}{\hat{G}_{CD,0}(b, u, u-t)} \right\} dN_i(u) = 0, \qquad (6)$$

where for $r = 0, 1$, $\hat{G}_{CD,r}(b, u, u-t) = n^{-1} \sum_{i=1}^{n} \hat{G}_{CD,ri}(b, u, u-t)$. When there is no measurement error, the conditional score estimating equation (6) also reduces to the weighted local partial likelihood score equation (2).

Instead of deriving the asymptotical properties directly for the conditional score estimator, we show that the conditional score estimator is asymptotic equivalent to the corrected score estimator, with the proof given in Appendix B.

Theorem 3. Under Conditions A1–A10, as $n \to \infty$, a solution to the conditional score estimating equation (6), say $\tilde{b}$, exists uniquely in a neighborhood of b_T with probability 1. In addition, $(nh)^{1/2} H(\tilde{b} - \hat{b}) = o_p(1)$.

In practice, the error variance parameter ω is unknown, however, it can be estimated using the methods of moments estimator $\hat{\omega}$ as given in Song et al. (2002a). Since $\hat{\omega}$ is a regular linear estimator for ω, the consistency and the asymptotic normality of the conditional score and corrected score estimators remain and the variances can be estimated by the sandwich technique.

5. SIMULATION STUDIES

We conducted several simulation experiments to assess the performance of the conditional score and corrected score estimators. Under the standard proportional hazards model with constant regression coefficients, the performance of these two estimators have been compared in Song and Huang (2005a) and Wang (2005), and the conditional score estimator performs better than the corrected score estimator in the case of small sample and large measurement error.

For simplicity, we consider the case when there is a single covariate $X(u)$. Under the first scenario, for each subject i, $X_i(u) = \alpha_{i0} + \alpha_{i1}u$, where $(\alpha_{i0}, \alpha_{i1})$ are jointly normal with mean $(4.173, -0.0103)$ and variance D, which has distinct elements $(D_{11}, D_{12}, D_{22}) = (1.24, -0.0114, 0.003)$. The longitudinal observations of $X(u)$ are measured at $u = \{0, 2, 4, 8, 16, 24, 32, 40, 48, 56, 64, 72, 80\}$, with a 10% missing rate after $u = 16$. The error e_{ij} is generated from a normal distribution with mean 0 and variance 0.1 or 0.2. The true regression coefficient $\beta(u) = -1$ and the baseline hazard $\lambda_0(u) = 1$. Censoring is generated from an exponential distribution with mean 110 and truncated at $u = 80$, leading to a censoring rate of 36%.

Under the second scenario, the covariate $X_i(u) = \alpha_{i0}$ is measured twice at the baseline $u = 0$, where α_{i0} is normal with mean 1 and variance 1. The error e_{ij} is also generated from a normal distribution with mean 0 and variance 0.1 or 0.2. The true regression coefficient $\beta(u) = u$ and the baseline hazard $\lambda_0(u) = 1$. Censoring is generated from an exponential distribution with mean 2 and truncated at $u = 2$. The censoring rate is 32%.

For each scenario, 1000 Monte Carlo data sets are simulated with $n = 300$. For each data set, we fit the above model four ways: (i) using the "ideal" approach in which the true values of $X_i(u)$ at each failure time were used (Cai and Sun's method); (ii) using the naive regression; (iii) using the conditional score estimator; (iv) using the corrected score estimator. For all methods, 95% Wald confidence interval for $\beta(t)$ is constructed. We use the Epanechnikov kernel $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$ here and for the application in Section 6.

We estimate the regression coefficient $\beta(t)$ at $t = 10, 20, 30, 40$ with $h = 20, 40, 60$ for the first scenario and at $t = 0.25, 0.50, 0.75, 1.0$ with $h = 0.4, 0.8, 1.2$ for the second scenario. The results are given in Tables 1–4. In all cases, the conditional score and corrected score estimators show negligible bias close to that of the unachievable 'ideal' estimator. In contrast, the naive regression can yield biased estimates and

11

coverage probabilities well below the nominal level. The empirical standard deviations are close to the standard errors except for a few cases for the corrected score estimator, which is caused by some extreme outliers. The corrected score approach fails to find solution for some datasets while the conditional score estimator works well. This issue worsens with decrease in h. This conforms to the relative behaviors of the two estimators under the standard proportional hazards model. Both methods have coverage probabilities close to the nominal level and works similarly among the wide ranges of the bandwidth. The selection of the bandwidth does not show obvious impact on any of these estimators. Similar results were observed by Cai and Sun (2003) for the local partial likelihood estimator.

We have also conducted simulations with a skewed bimodal mixture of normal distributions for the random effects and observed similar results. With increase in sample size, both the conditional score and corrected score methods show improvement while the naive approach performs worse.

6. APPLICATION TO ACTG 175 DATA

To demonstrate the utility of the methods for investigating association between error contaminated covariates and clinical endpoint, we apply the methods to the ACTG 175 data. We are interested in assessing the effect of CD4 count and treatment on time to AIDS or death. Figure 1 presents $\log_{10}$-transformed CD4 profiles for 10 randomly selected subjects and shows an apparent initial increase, with a peak at week 12, followed by an approximate linear decline. The logarithmic transformation is usually used for CD4 count to achieve approximate within subject normality and constant variance. Because only 9 events occurred before week 12, for simplicity, we consider the data including and after week 12. The trajectory of $\log_{10}$ CD4 seems to follow approximate straight-line relationship (Song and Huang, 2005b). Thus we assume $X_{i1}(u) = \alpha_{i10} + \alpha_{i11}u$ represents "inherent" $\log_{10}$(CD4 count) for subject i at time u. The estimate of the error variance is 0.0104. The primary analysis found zidovudine alone to be inferior to the other three therapies; thus, further investigations focused on two treatment groups, zidovudine alone and the combination

of the other three. We took the hazard for AIDS or death to be $\lambda_i(u) = \lambda_0(u) \exp\{\beta_1(u)X_{i1}(u) + \beta_2(u)X_{i2}\}$, where $X_{i2} = I(\text{treatment} \neq \text{zidovudine})$.

We estimated the regression parameters using the naive regression and the conditional score approaches. Results are shown in Figure 2 for $h = 60, 72, 84$. The estimates for $\beta_1(u)$ are negative with magnitude decreasing over time. The estimates for $\beta_2(u)$ tend to decrease after randomization before reaching a negative minimum around week 20 and then increase to close to zero, which indicates that the treatment reaches its maximum effect within several months, gradually decays thereafter and diminishes eventually. Thus, in this case, it seems inappropriate to use constant regression coefficients as in the standard proportional hazards model. The estimates are not sensitive to the different choices of bandwidth h except for u close to 0. The corrected score approach fails to find a solution at many time points and the results are omitted. From our numerical simulation experience, the corrected score estimator is more likely to encounter the small sample problem than the conditional score estimator.

7. DISCUSSION

We have proposed two semiparametric estimators for the time-varying-coefficient proportional hazards model with covariates measured with error. The estimators are asymptotically equivalent, but the conditional score estimator has better finite sample performance than the corrected score estimator. The asymptotic normality of the estimators justifies the computation of the pointwise Wald confidence intervals. However, for time-varying-coefficients, it may be more appealing to derive the confidence bands over an interval. A resampling method similar to that proposed by Tian et al. (2005) may be used and will be investigated in the future. Another interesting topic for future research is to develop appropriate methods for bandwidth selection.

The model as presented here may be extended to more complicated situations. We have focused on the case when the errors are independent across time. However, noting that this assumption is only used

in computing the variance of the least square estimator $\hat{X}_i(u)$, it can be relaxed to the cases of other error correlation structures, such as the exponential correlation model (Diggle et al., 2002, p.56), as long as we can derive a consistent estimator for the variance of $\hat{X}_i(u)$. When some coefficients do not vary over time, it is more efficient to use a model with mixed constant and time-varying coefficients. The conditional score and corrected score estimators may be derived similarly except that the slopes for the constant coefficients are zero. In addition, the survival model can be generalized to include the random effects instead of the true covariates by analogy to Song et al. (2002a) and Wang (2005).

ACKNOWLEDGEMENTS

APPENDIX A: PROOF OF THEOREM 1

Let $\mathcal{N}(b_0)$ be a compact neighborhood of b_0. Let $e_{X,ik}(u) = \left\{ F_{ik}^T F_{ik} \right\}^{-1} F_{ik} e_{ik}(u)$, $e_{X,i}(u) = (e_{1i}(u), \ldots, e_{ip}(u))^T$, $\nu_r = \int u^r K^2(u) du$, $H^* = diag(1, h)$, and $S_i^*(b, u, u-t) = \hat{X}_i(u) + \Sigma_i(u) \left\{ b_0 + (u-t)b_1 \right\} dN_i(u)$. Write

$$
\hat{G}_{I,r}^*(b, u, u-t) = n^{-1} \sum_{i=1}^n Y_i(u) X_i^r(u) \exp\left\{ b_0^T X_i(u) + b_1^T (u-t) X_i(u) \right\},
$$

$$
\hat{G}_{CR,r}^*(b, u, u-t) = n^{-1} \sum_{i=1}^n Y_i(u) \hat{X}_i^r(u) \exp\left\{ b_0^T \hat{X}_i(u) + b_1^T (u-t) \hat{X}_i(u) - \frac{1}{2} b^T \Sigma_i(u, u-t) b \right\},
$$

$$
\hat{G}_{CD,r}^*(b, u, u-t) = n^{-1} \sum_{i=1}^n Y_i(u) S_i^{*r}(b, u, u-t)
$$
$$
\times \exp\left\{ b_0^T S_i^*(b, u, u-t) + b_1^T (u-t) S_i^*(b, u, u-t) - \frac{1}{2} b^T \Sigma_i(u, u-t) b \right\},
$$

$$
\hat{G}_{II,r}^*(\beta_0, u) = n^{-1} \sum_{i=1}^n Y_i(u) X_i^r(u) \exp\left\{ \beta_0^T(u) X_i(u) \right\},
$$

$$
\hat{G}_{CRI,r}(\beta, u, u-t) = n^{-1} \sum_{i=1}^n Y_i(u) \hat{X}_i^r(u, u-t) \exp\left\{ \beta^T(u) X_i(u) \right\},
$$

$$
\hat{G}_{CRI,r}^*(\beta, u) = n^{-1} \sum_{i=1}^n Y_i(u) \hat{X}_i^r(u) \exp\left\{ \beta^T(u) X_i(u) \right\},
$$

$$\hat{G}^*_{\Sigma,r}(b^{\#},b,u,u-t) = n^{-1}\sum_{i=1}^{n} Y_i(u)\left[\Sigma_i(u)\left\{b_0^{\#}+(u-t)b_1^{\#}\right\}\right]^{\otimes r}$$
$$\times \exp\left\{b^T\hat{X}_i(u,u-t)-\frac{1}{2}b^T\Sigma_i(u,u-t)b\right\},$$

$$\hat{G}^*_{\Sigma I,r}(b,\beta,u,u-t) = n^{-1}\sum_{i=1}^{n} Y_i(u)\left[\Sigma_i(u)\left\{b_0+(u-t)b_1\right\}\right]^{\otimes r}\exp\left\{\beta^T(u)X_i(u)\right\},$$

$$\hat{G}^*_{X\Sigma}(b,u,u-t) = n^{-1}\sum_{i=1}^{n} Y_i(u)\hat{X}_i(u)\left[\Sigma_i(u)\left\{b_0+(u-t)b_1\right\}\right]^T$$
$$\times \exp\left\{b^T\hat{X}_i(u,u-t)-\frac{1}{2}b^T\Sigma_i(u,u-t)b\right\},$$

$$\hat{G}^*_{X\Sigma I}(b,\beta,u,u-t) = n^{-1}\sum_{i=1}^{n} Y_i(u)\hat{X}_i(u)\left[\Sigma_i(u)\left\{b_0+(u-t)b_1\right\}\right]^T\exp\left\{\beta^T(u)X_i(u)\right\}.$$

The corresponding expectations are represented by the same notations with $\hat{G}$ replaced by G.

To derive the asymptotical properties, we assume the following regularity conditions.

A1. $P(V \geq L) > 0$.

A2. $P(V = u) = 0$ for $u \in [0, L]$.

A3. $K(\cdot)$ has a compact support, say $[-1, 1]$, and is bounded and has a continuous derivative on $[-1, 1]$. Assume $|K(u)| \leq M_K$, and Q_μ is positive definite.

A4. $X(u)$ has a continuous derivative $X'(u)$ for $u \in [0, L]$.

A5. There exists a random variable B such that $\sup_{u \in \mathcal{N}(t)} \|X(u)\| \leq B$, $\sup_{u \in \mathcal{N}(t)} \|e_X(u)\| \leq B$, and
$\sup_{u \in \mathcal{N}(t)} \|\Sigma(u)\| \leq B$, $E\left[\exp\left\{4\sup_{u \in \mathcal{N}(t)}(\|\beta(u)\| + \|\beta'(u)\| + \|\beta(u)\|^2 + \|\beta'(u)\|^2 + 1)B\right\}\right] < \infty$.

A6. $G^*_{I,r}(b,u,u-t)$, $G^*_{II,r}(\beta_0,u)$, $G^*_{CR,r}(b,u,u-t)$, $G^*_{CRI,r}(\beta,u)$, $G^*_{\Sigma,r}(b^{\#},b,u,u-t)$, $G^*_{X\Sigma}(b,u,u-t)$, $G^*_{X\Sigma I}(b,\beta,u,u-t)$ are continuous at $u \in \mathcal{N}(t)$, a compact neighborhood around t, for $b \in \mathcal{N}(b_0)$ and $r = 0, 1, 2$.

A7. $\beta(u)$ has a continuous second derivative $\beta''(u)$ for $u \in \mathcal{N}(t)$.

A8. $\int_0^L \lambda_0(u)du < \infty$, $\lambda_0(t) > 0$, and $\lambda_0(u)$ is continuous for $u \in \mathcal{N}(t)$.

A9. The matrix $\Gamma_0^*(t) = G_{II,2}^*(\beta_0, t) - G_{II,1}^{*\otimes 2}(\beta_0, t)/G_{II,0}^*(\beta_0, t)$ is positive definite.

A10. $h = o(1)$, $nh^5 = O(1)$, and $nh \to \infty$ as $n \to \infty$.

First, we give some lemmas. Let $W_{ik}(u) = \{W_{ik}(t_{ikj}) : t_{ikj} \leq u\}$, and $W_i(u) = \{W_{i1}(u), \ldots, W_{ip}(u)\}$.

Define $\mathcal{F}_i(u) = \{N_i(s), Y_i(s), X_i(s), W_i(s) : s \leq u\}$. Then $M_i(u) = N_i(u) - \int \lambda_0(s)Y_i(s) \times \exp\{\beta_0(s)X_i(s)\}ds$

is a martingale with respect to $\mathcal{F}_i(u)$. Write $c'(u, b)$ as the partial derivative of $c(u, b)$ with respective to

u, and $\widehat{\mathcal{E}}$ be the empirical average operator such that $\widehat{\mathcal{E}}c = \widehat{\mathcal{E}}c_i = n^{-1}\sum_{i=1}^n c_i$.

Lemma 1. Suppose that $g_i(u)$ is a predictable process with respect to the filtration $\mathcal{F}_i(u)$ and has a deriva-

tive $g_i'(u)$ for $u \in \mathcal{N}(t)$. If $\sup_{u \in \mathcal{N}(t)} E\left[g_i^2\{u\} + g_i'^2\{u\}\right] < \infty$, then

$$E'\left[g_i(t)N_i(t)\right] - E\left\{g_i'(t)N_i(t)\right\} = \lambda_0(t)E\left[g_i(t)Y_i(t)\exp\left\{(\beta_0(t)X_i(t))\right\}\right].$$

Specifically, if $g_i(t) = 1$, $E'\left[N_i(t)\right] = \lambda_0(t)G_{II,0}^*(\beta_0, t)$.

Proof. Note that $M_i(u)$ is a martingale with respect to the filtration $\mathcal{F}_i(u)$, and

$$E\left[g_i(u)M_i(u)|\mathcal{F}_i(u-)\right] = g_i(u)E\left[M_i(u)|\mathcal{F}_i(u-)\right] = 0$$

$$E\left[g_i'(u)M_i(u)|\mathcal{F}_i(u-)\right] = g_i'(u)E\left[M_i(u)|\mathcal{F}_i(u-)\right] = 0.$$

Hence

$$0 = E\left[g_i(t)M_i(t)\right] = E\left\{g_i(t)N_i(t)\right\} - E\left\{g_i(t)\int_0^t \lambda_0(u)Y_i(u)\exp\left\{(\beta_0(u)X_i(u)\right\}du\right\} \tag{A.1}$$

and

$$0 = E\left[g_i'(t)M_i(t)\right] = E\left\{g_i'(t)N_i(t)\right\} - E\left\{g_i'(t)\int_0^t \lambda_0(u)\exp\left\{(\beta_0(u)X_i(u)\right\}du\right\}. \tag{A.2}$$

16

Taking derivative with respect to t in equation (A.1), under Condition A8, we have

$$E'\left[g_i(t)N_i(t)\right] = E\left\{g_i'(t)\int_0^t \lambda_0(u)\exp\left\{(\beta(u)X_i(u)\right\}du\right\} + E\left[g_i(t)\lambda_0(t)Y_i(t)\exp\left\{(\beta_0(t)X_i(t)\right\}\right].$$

This, together with (A.2), completes the proof.

In the following lemmas, let $c_i(u,w,b)$ be a random process for the ith subject on $R = \{(u,w,b) : u \in \mathcal{N}(t), w \in [-1,1], b \in \mathcal{N}(b_0)\}$, and $g_n(u,w,b)$ be a random process on R and $g(u,w,b)$ be a non-random process on R.

From the functional strong law of large number of Anderson and Gill (1982), we have the following lemma.

Lemma 2. Suppose $\sup_{(u,w,b)\in R} E\left\|c_i(u,w,b)\right\| < \infty$. Then

$$\sup_{(u,w,b)\in R}\left\|n^{-1}\sum_{i=1}^n Y_i(u)c_i(u,w,b) - E\left\{Y_i(u)c_i(u,w,b)\right\}\right\| \overset{a.s.}{\to} 0.$$

Lemma 3. Suppose $\sup_{(u,w,b)\in R}\left|g_n(u,w,b) - g(u,w,b)\right| \overset{a.s.}{\to} 0$ and $\sup_{(u,w,b)\in R}\left|g(u,w,b)\right| < \infty$. Then, for $r \geq s \geq 0$,

$$\sup_{b\in\mathcal{N}(b_0)}\left|\int_0^L K_h(u-t)h^{-s}(u-t)^r g_n\left\{u,h^{-1}(u-t),b\right\}du - I(r=s)\int_0^L K(w)h^{r-s}w^r g(t+hw,w,b)dw\right| \overset{a.s.}{\to} 0.$$

In the special case when $g_n(u,w,b) = g_n(u,b)$ and $g(u,w,b) = g(u,b)$, if $g(u,b)$ is continuous at $u = t$, then

$$\sup_{b\in\mathcal{N}(b_0)}\left|\int_0^L K_h(u-t)h^{-s}(u-t)^r g_n(u,b)du - I(r=s)\int_0^L K(w)h^{r-s}w^r dw g(t,b)\right| \overset{a.s.}{\to} 0.$$

Proof. $\int K_h(u-t)h^{-s}(u-t)^r g_n\left\{u,h^{-1}(u-t),b\right\}du = \int K(w)h^{r-s}w^r g_n(t+hw,w,b)dw$. And for n large

enough, under Condition A3,

$$\sup_{b\in\mathcal{N}(b_0)}\left|\int_0^L K(w)h^{r-s}w^r g_n(t+hw,w,b)dw - \int_0^L K(w)h^{r-s}w^r g(t+hw,w,b)dw\right|$$

$$\leq \sup_{(u,w,b)\in R}|g_n(u,w,b)-g(u,w,b)|\int_0^L K(w)h^{r-s}w^r dw$$

$$= \sup_{(u,w,b)\in R}|g_n(u,w,b)-g(u,w,b)\,|\,h^{r-s}\int_0^L K(w)w^r dw$$

$$\overset{a.s.}{\to} 0.$$

If $\sup_{u\in\mathcal{N}(t),w\in[-1,1],b\in\mathcal{N}(b_0)}\|g(u,w,b)\|<\infty,\ \ \text{for } r>s,$

$$\sup_{b\in\mathcal{N}(b_0)}\left|\int_0^L K(w)h^{r-s}w^r g(t+hw,w,b)dw\right|$$

$$\leq \sup_{(u,w,b)\in R}|g(u,b)|\,h^{r-s}\int_0^L K(w)w^r dw$$

$$\to\quad 0.$$

In the special case when $g_n(u,w,b)=g_n(u,b)$ and $g_n(u,w,b)=g_n(u,b)$, if $g(u,b)$ is continuous at $u=t$,

$$\sup_{b\in\mathcal{N}(b_0)}\left|\int_0^L K(w)h^{r-s}w^r g(t+hw,b)dw - \int_0^L K(w)h^{r-s}w^r g(t,b)dw\right|$$

$$\leq \sup_{(u,b)\in\mathcal{N}(t)\times\mathcal{N}(b_0)}|g(u,b)|\ \int_0^L K(w)h^{r-s}w^{r-s}dw$$

$$\to\quad 0.$$

Thus the result follows.

Lemma 4. If $g(u,w,b)$ has bounded and continuous partial derivatives with respect to u and w for $u\in\mathcal{N}(t)$

and $w\in[-1,1]$, and $\sup_{(u,w,b)\in R}|g(u,w,b)-g(t,w,b)|\to 0$, then for $r\geq s\geq 0$,

$$\sup_{b\in\mathcal{N}(b_0)}\left|\int_0^L K_h(u-t)h^{-s}(u-t)^r g\{u,h^{-1}(u-t),b\}d\widehat{\mathcal{E}}N(u)\right.$$

$$\left.-I(r=s)\int_0^L K(w)w^r g(t,w,b)dwE'N(t)\right|\overset{a.s.}{\to} 0.$$

18

Proof. Let $g'_1(u,w,b)$ and $g'_2(u,w,b)$ be the derivatives $g(u,w,b)$ with respect to u and w, respectively.

By integration by parts, for $r \geq s$ and n large enough,

$$\sup_{b \in \mathcal{N}(b_0)} \left| \int_0^L K_h(u-t)h^{-s}(u-t)^r g\left\{u, h^{-1}(u-t), b\right\} d\hat{\hat{\mathcal{E}}}N(u) \right.$$

$$\left. - \int_0^L K_h(u-t)h^{-s}(u-t)^r g(u, h^{-1}(u-t), b) dEN(u) \right|$$

$$\leq \sup_{b \in \mathcal{N}(b_0)} K_h(u-t)h^{-s}(u-t)^r g\left\{u, h^{-1}(u-t), b\right\} \left| \hat{\mathcal{E}}N(u) - EN(u) \right| \Big|_{u=t-h}^{u=t+h}$$

$$+ \sup_{b \in \mathcal{N}(b_0)} \int_0^L \left| \hat{\mathcal{E}}N(u) - EN(u) \right|$$

$$\times \left(\left[K'_h\left\{h^{-1}(u-t)\right\} h^{-s-2}(u-t)^r + K_h(u-t)h^{-s}r(u-t)^{r-1} \right] \left| g\left\{u, h^{-1}(u-t), b\right\} \right| \right.$$

$$\left. + K_h(u-t)h^{-s}(u-t)^r \left[g'_1\left\{u, h^{-1}(u-t), b\right\} + h^{-1} g'_2\left\{u, h^{-1}(u-t), b\right\} \right] \right) du$$

$$\leq \sup_{u \in \mathcal{N}(t)} \left| \hat{\mathcal{E}}N(u) - EN(u) \right|$$

$$\times \left(\sup_{(u,w,b) \in R} \left\{ |g(u,w,b)| + \left| g'_1(u,w,b) \right| + \left| g'_2(u,w,b) \right| \right\} \right.$$

$$\left. \times \left[M_K O(h^{r-s+1}) + \int_0^L \left\{ \left| K'(w) \right| h^{r-s-1}w^a + K(w)h^{r-s-1}rw^{r-1} \right\} + K(w) \left(h^{r-s}w^r + h^{r-s-1}w^r \right) dw \right] \right)$$

$$= \sup_{u \in \mathcal{N}(t)} \left| \hat{\mathcal{E}}N(u) - EN(u) \right| \sup_{(u,w,b) \in R} \left\{ |g(u,w,b)| + \left| g'_1(u,w,b) \right| + \left| g'_2(u,w,b) \right| \right\} O(h^{r-s-1})$$

$$= o_p(1),$$

where the last step follows from $\sup_{u \in \mathcal{N}(t)} \left| \hat{\mathcal{E}}N(u) - EN(u) \right| = O_p(n^{-1/2})$ and assumptions A3. Note that

$$\int K_h(u-t)h^{-s}(u-t)^r g(u, h^{-1}(u-t), b) dEN(u) = \int K_h(u-t)h^{-s}(u-t)^r g(u, h^{-1}(u-t), b) E'N(u) du$$

$$= \int K(w)h^{r-s}w^r g(t+hw, w, b) E'N(t+hw) dw.$$

For $r = s$,

$$\sup_{b \in \mathcal{N}(b_0)} \left| \int K(w)w^r g(t+hw, w, b) E'N(t+hw) dw - \int K(w)w^r g(t, w, b) E'N(t) dw \right|$$

$$\leq \sup_{w \in [-1,1], b \in \mathcal{N}(b_0)} \left| \int K(w)w^r g(t+hw, w, b) E'N(t+hw) dw - \int K(w)w^r g(t+hw, w, b) E'N(t) dw \right|$$

$$+ \sup_{w\in[-1,1], b\in\mathcal{N}(b_0)} \left| \int K(w)w^r g(t+hw, w, b)E'N(t)dw - \int K(w)w^r g(t, w, b)E'N(t)dw \right|$$

$$\leq \sup_{(u,w,b)\in R} |g(u, w, b)| \sup_{w\in[-1,1]} \left| E'N(t+w) - E'N(t) \right| \int K(w)w^r dw$$

$$+ \sup_{w\in[-1,1], b\in\mathcal{N}(b_0)} |g(t+hw, w, b) - g(t, w, b)| \, |E'N(t)| \int K(w)w^r dw$$

$$\rightarrow \quad 0$$

For $r > s$,

$$\sup_{b\in\mathcal{N}(b_0)} \left| \int K(w)h^{r-s}w^r g(t+hw, w, b)E'N(t+hw)dw \right|$$

$$\leq \sup_{(u,w,b)\in R} |g(u, w, b)| \, h^{r-s} \int K(w)w^r dw$$

$$\rightarrow \quad 0.$$

The result then follows.

Lemma 5. If $\sup_{(u,w,b)\in R} |g_n(u, w, b) - g(u, w, b)| \overset{a.s.}{\rightarrow} 0$, then for $r \geq s \geq 0$,

$$\sup_{b\in\mathcal{N}(b_0)} \left| \int_0^L K_h(u-t)h^{-s}(u-t)^r \left[g_n\{u, h^{-1}(u-t), b\} - g\{u, h^{-1}(u-t), b\} \right] d\widehat{\mathcal{E}}N(u) \right| \overset{a.s.}{\rightarrow} 0.$$

Proof.

$$\sup_{b\in\mathcal{N}(b_0)} \left| \int_0^L K_h(u-t)h^{-s}(u-t)^r \left[g_n\{u, h^{-1}(u-t), b\} - g\{u, h^{-1}(u-t), b\} \right] d\widehat{\mathcal{E}}N(u) \right|$$

$$= \sup_{b\in\mathcal{N}(b_0)} \left| \int_0^L K(w)h^{r-s}w^r \left\{ g_n(u, w, b) - g(u, w, b) \right\} d\widehat{\mathcal{E}}N(t+hw) \right|$$

$$\leq \sup_{(u,w,b)\in R} |g_n(u, w, b) - g(u, w, b)| \, h^{r-s} \int_0^L K(w) d\widehat{\mathcal{E}}N(t+hw)$$

$$= \sup_{(u,w,b)\in R} |g_n(u, w, b) - g(u, w, b)| \, h^{r-s} \int_0^L K_h(u-t) d\widehat{\mathcal{E}}N(u)$$

$$\overset{a.s.}{\rightarrow} 0,$$

with the last step follows from Lemma 4.

Lemma 6. If $\sup_{(u,w,b)\in R} |g_n(u,w,b) - g(u,w,b)| \overset{a.s.}{\to} 0$, and $\sup_{(u,w,b)\in R} |g(u,w,b)| < \infty$, then for $r \geq s+1 \geq 0$, then

$$\sup_{b\in\mathcal{N}(b_0)} \left| \int_0^L K_h^2(u-t)h^{-s}(u-t)^r g_n(u, h^{-1}(u-t), b)du \right.$$
$$\left. -I(r=s+1) \int_0^L K^2(w)w^r g(t+hw, w, b)dw \right| \overset{a.s.}{\to} 0.$$

In the special case when $g_n(u,w,b) = g_n(u,b)$, and $g(u,w,b) = g(u,b)$, if $g(u,b)$ is continuous at $u = t$,

$$\sup_{b\in\mathcal{N}(b_0)} \left| K_h^2(u-t)h^{-s}(u-t)^r g_n(u,b)du - I(r=s+1) \int_0^L K(w)w^r dw g(t,b) \right| \overset{a.s.}{\to} 0.$$

The proof is similar to that for Lemma 3 and is thus omitted.

Lemma 7. Suppose that $c_i(u,w,b)$ is predictable with respect to $\mathcal{F}_i(u)$ and has a continuous partial derivative $c_i'(u,w,b)$ with respect to u almost everywhere for $u \in [0, L]$. If for some $\eta > 0$,

$$\sup_{u\in\mathcal{N}(t), w\in[-\eta,\eta], b\in\mathcal{N}(b_0)} E\left[\left| c_i(u, u-t, b) \right| + \left| \frac{\partial c_i(u, u-t, b)}{\partial u} \right| \right] < \infty,$$

then for $r \geq s \geq 0$,

$$\sup_{b\in\mathcal{N}(b_0)} \left| \int_0^L K_h(u-t)h^{-s}(u-t)^r n^{-1} \sum_{i=1}^n c_i(u, u-t, b)dN(u) \right.$$
$$\left. -I(r=s) \int_0^L K(w)w^r dw \lambda_0(t) E\left[c_i(t, 0, b)Y_i(t) \exp\left\{ (\beta_0(t)X_i(t)) \right\} \right] \right| \overset{a.s.}{\to} 0.$$

Proof. Note that

$$\int_0^L K_h(u-t)h^{-s}(u-t)^r \widehat{\mathcal{E}}c(b, u, u-t)dN(u)$$
$$= \int_0^L K_h(u-t)h^{-s}(u-t)^r d\widehat{\mathcal{E}}c(b, u, u-t)N(u) - \int_0^L K_h(u-t)h^{-s}(u-t)^r \widehat{\mathcal{E}} \frac{\partial c(b, u, u-t)}{\partial u} N(u)du.$$

With similar arguments as those for Lemma 4, we have

$$\sup_{b\in\mathcal{N}(b_0)} \left| \int_0^L K_h(u-t)h^{-s}(u-t)^r d\widehat{\mathcal{E}}c(b, u, u-t)N(u) - I(r=s) \int_0^L K(w)w^r dw E'\{c(b, u, u-t)N(u)\} \right| \overset{a.s.}{\to} 0,$$

where $E'\{c(b, u, u - t)N(u)\}$ is the partial derivative of $E\{c(b, u, u - t)N(u)\}$ with respect to u. And from Lemma 3, we have

$$\sup_{b \in \mathcal{N}(b_0)} \left| \int K_h(u - t)h^{-s}(u - t)^r \widehat{\mathcal{E}} \frac{\partial c(b, u, u - t)}{\partial u} N(u)du - I(r = s) \int K(w)w^r dw E\{ \frac{\partial c(b, u, u - t)}{\partial u} N(u)\} \right| \overset{a.s.}{\to} 0.$$

Then the result follows from Lemma 1.

We now show the asymptotic properties for the corrected score estimator $\hat{b} = (\hat{b}_0, \hat{b}_1)$.

Consistency

Let $\alpha = (\alpha_0^T, \alpha_1^T)^T = H(b - b_T)$. Then $\hat{\alpha} = H(\hat{b} - b_T)$ is a solution to $\tilde{U}_{CR}(\alpha) = \hat{U}_{CR}(H^{-1}\alpha + b_T)$ and $\hat{\alpha}_I = H(\hat{b}_I - b_T)$ is a solution to $\tilde{U}_I(\alpha) = \hat{U}_I(H^{-1}\alpha + b_T)$. Let $H^* = diag(1, h)$. Note that $\tilde{U}_{CR}(\alpha)$ can be rewritten as

$$\int_0^L K_h(u - t) \left(H^{*-1}u_t \right) \otimes \left\{ \widehat{\mathcal{E}} \hat{X}_i(u)dN_i(u) + \widehat{\mathcal{E}} \Sigma_i(u)(u_t \otimes I_p)^T \left(H^{-1}\alpha + b_T \right) dN_i(u) \right.$$

$$\left. - \frac{G_{CR,1}^*(H^{-1}\alpha + b_T, u, u - t)}{G_{CR,0}^*(H^{-1}\alpha + b_T, u, u - t)} \widehat{\mathcal{E}} dN_i(u) \right\} \tag{A.3}$$

$$+ \int_0^L K_h(u - t) \left(H^{*-1}u_t \right) \otimes \left\{ \frac{\hat{G}_{CR,1}^*(H^{-1}\alpha + b_T, u, u - t)}{\hat{G}_{CR,0}^*(H^{-1}\alpha + b_T, u, u - t)} \right.$$

$$\left. - \frac{G_{CR,1}^*(H^{-1}\alpha + b_T, u, u - t)}{G_{CR,0}^*(H^{-1}\alpha + b_T, u, u - t)} \right\} \widehat{\mathcal{E}} dN_i(u). \tag{A.4}$$

Condition A1 implies that $G_{CR,0}^*(H^{-1}\alpha + b_T, u, u - t)$ is uniformly bounded below. Then under Condition A5, by Lemma 2, (A.4) converges almost surely to 0 uniformly for $\alpha \in \mathcal{N}(0)$. With an application of Lemmas 2, 4 and 7, Conditions A4–A6 imply that (A.3) converges uniformly to

$$(\mu_0, \mu_1)^T \otimes \lambda_0(t)G_{CRI,1}^*(\beta_0, t) + \int_0^L K(w)(1, w)^T \otimes \lambda_0(t)G_{\Sigma I,1}^*(\alpha + b_T^*, \beta_0, t, w)dw$$

$$- \int_0^L K(w)(1, w)^T \otimes \frac{G_{CR,1}^*(\alpha + b_T^*, t, w)}{G_{CR,0}^*(\alpha + b_T^*, t, w)} \lambda_0(t)G_{II,0}^*(\beta_{0,}, t, 0)dw, \tag{A.5}$$

22

where $b_T^* = (\beta_0^T(t), 0_p^T)^T$. With some algebra, we can show that

$$G_{CR,0}^*(b, u, u - t) = G_{I,0}^*(b, u, u - t),$$

$$G_{CR,1}^*(b, u, u - t) = G_{I,1}^*(b_0, u, u - t) + E\left[\Sigma(u)\left\{b_0 + (u - t)b_1\right\}\right]G_{I,0}^*(b, u, u - t),$$

$$G_{CR,2}^*(b, u, u - t) = G_{I,2}^*(b, u, u - t) + G_{I,1}^*(b, u, u - t)E^T\left\{\Sigma(u)\left\{b_0 + (u - t)b_1\right\}\right\}$$

$$+E\left\{\Sigma(u)\left\{b_0 + (u - t)b_1\right\}\right\}G_{I,1}^{*T}(b, u, u - t) \tag{A.6}$$

$$+G_{I,0}^*(b, u, u - t)E^{\otimes 2}\left\{\left[\Sigma(u)\left\{b_0 + (u - t)b_1\right\}\right] + \Sigma(u)\right\},$$

$$G_{\Sigma I,1}^*(\alpha + b_T^*, \beta_0, t, w) = E\left\{\Sigma(t)\left(\alpha_0 + b_{T0} + w\alpha_1\right)\right\}G_{II,0}^*(\beta_0,, t),$$

$$G_{CRI,r}^*(\beta_0, t) = G_{I,r}^*(b_T, t, 0) = G_{II,r}^*(\beta_0, t).$$

Using these equations, (A.5) can be simplified as

$$U(\alpha) = (\mu_0, \mu_1)^T$$

$$\otimes \lambda_0(t)\left\{G_{II,1}^*(\beta_0, t) - \int_0^L K(w)(1, w)^T \otimes \frac{G_{I,1}^*(\alpha + b_T^*, t, w)}{G_{I,0}^*(\alpha + b_T^*, t, w)}dw G_{II,0}^*(\beta_0, t)\right\}.$$

Thus $\tilde{U}_{CR}(\alpha)$ converges almost surely to $U(\alpha)$ uniformly for $\alpha \in \mathcal{N}(0)$. It is easy to see that $U(\alpha)$ is the derivative of a concave function and is equal to zero at $\alpha = 0$. Therefore $\hat{\alpha} \xrightarrow{p} 0$; that is, $H\left(\hat{b} - b_T\right) \xrightarrow{p} 0$.

Since $U(\alpha)$ is the derivative of a concave function, the asymptotic uniqueness of the corrected score estimator follows.

Asymptotic Normality

We first consider $(nh)^{1/2}\hat{U}_{CR}(\beta_T)$. It is easy to see that

$$(nh)^{1/2}\hat{U}_{CR}(\beta_T) =$$

$$(nh)^{1/2}n^{-1}\sum_{i=1}^n \int_0^L H^{-1}K_h(u - t)\left\{\hat{X}_i(u, u - t) + \Sigma_i(u, u - t)b_T\right.$$

$$\left. - \frac{\hat{G}_{CR,1}(b_T, u, u - t)}{\hat{G}_{CR,0}(b_T, u, u - t)}\right\}dM_i(u) \tag{A.7}$$

23

$$+ (nh)^{1/2} n^{-1} \sum_{i=1}^{n} \int_{0}^{L} H^{-1} K_h(u-t) \bigg\{ \hat{X}_i(u, u-t) + \Sigma_i(u, u-t) b_T$$

$$- \frac{\hat{G}_{CR,1}(b_T, u, u-t)}{\hat{G}_{CR,0}(b_T, u, u-t)} \bigg\} \lambda_0(u) Y_i(u) \exp \left\{ \beta_0^T(u) X_i(u) \right\} du. \tag{A.8}$$

Note that (A.8) can be written as $\xi_{n1}(\beta_T) + \xi_{n2}(\beta_T)$, where

$$\begin{aligned}
\xi_{n1}(\beta_T) &= (nh)^{1/2} \int_0^L \left(H^{*-1} u_t \right) \otimes K_h(u-t) \lambda_0(u) \bigg\{ G^*_{CRI,1}(\beta_0, u) \\
&\quad + G^*_{\Sigma I,1}(b_T, \beta_0, u, u-t) - \frac{G^*_{CR,1}(b_T, u, u-t)}{G^*_{CR,0}(b_T, u, u-t)} G^*_{CRI,0}(\beta_0, u) \bigg\} du, \\
\xi_{n2}(\beta_T) &= h^{1/2} \int_0^L \left(H^{*-1} u_t \right) \otimes K_h(u-t) \lambda_0(u) n^{1/2} J_n(b_T, u, u-t) du,
\end{aligned}$$

and

$$\begin{aligned}
J_n(b_T, u, u-t) &= \left\{ \hat{G}^*_{CRI,1}(\beta_0, u) - G^*_{CRI,1}(\beta_0, u) \right\} \\
&\quad + \left\{ \hat{G}^*_{\Sigma I,1}(b_T, \beta_0, u, u-t) - G^*_{\Sigma I,1}(b_T, \beta_0, u, u-t) \right\} \\
&\quad + \left\{ \frac{\hat{G}^*_{CR,1}(b_T, u, u-t)}{\hat{G}^*_{CR,0}(b_T, u, u-t)} \hat{G}^*_{CRI,0}(\beta_0, u) - \frac{G^*_{CR,1}(b_T, u, u-t)}{G^*_{CR,0}(b_T, u, u-t)} G^*_{CRI,0}(\beta_0, u) \right\}.
\end{aligned}$$

By the functional delta method, the empirical process $n^{1/2} J_n(b_T, u, u-t)$ converges to a Gaussian process $\mathcal{J}$. Now, using the strong embedding theorem (Shorack and Wellner, 1986, p. 47–48), there exists a new probability space such that it converges to $\mathcal{W}$ almost surely. Together with that $h = o(1)$, we have $\xi_{n2}(\beta_T) = o_p(1)$. Using (A.6), with some algebra, we can show that

$$\begin{aligned}
\xi_{n1}(\beta_T) &= (nh)^{1/2} \int_0^L \left(H^{*-1} u_t \right) K_h(u-t) \lambda_0(u) \bigg[G^*_{II,1}(\beta_0, u) - G^*_{I,1}(b_T, u, u-t) \\
&\quad + \frac{G^*_{I,1}(b_T, u, u-t)}{G^*_{I,0}(b_T, u, u-t)} \left\{ G^*_{I,0}(b_T, u, u-t) - G^*_{II,0}(\beta_0, u) \right\} \bigg] du. \tag{A.9}
\end{aligned}$$

With an application of the Taylor series expansion, under Condition A7, we can show that, for $r = 1, 2$,

$$G^*_{II,r}(\beta_0, u) - G^*_{I,r}(b_T, u, u-t) = \frac{1}{2}(u-t)^2 G^*_{II,r+1}(\beta_0, u) \beta_0''(t) + o_p(h^2).$$

24

Substituting this into equation (A.9) gives that

$$\xi_{n1}(\beta_T) = (nh)^{1/2} \frac{1}{2} h^2 (\mu_2, \mu_3)^T \otimes \lambda_0(t) \Gamma_0^*(t) \beta_0''(t) + o_p(1).$$

Then we only need to show that $\varsigma_n(v) = (A.7)$ converges to a Gassian process. This can be proved by Rebolledo's central limit theorem. Using Lemmas 2 and 6, we can show that the predictable variation process

$$
\begin{aligned}
\langle \varsigma_n, \varsigma_n \rangle &= n^{-1} h \sum_{i=1}^{n} \int_0^L \Bigg[K_h^2(u-t) \left(H^{*-1} u_t \right)^{\otimes 2} \otimes \\
&\qquad \left\{ \hat{X}_i(u) + \Sigma_i(u) \{ b_{T0} + (u-t) b_{T1} \} - \frac{\hat{G}_{CR,1}^*(b_T, u, u-t)}{\hat{G}_{CR,0}^*(b_T, u, u-t)} \right\}^{\otimes 2} \\
&\qquad \times \lambda_0(u) Y_i(u) \exp \{ \beta_0(u) X_i(u) \} \Bigg] du \\
&= Q_\nu \otimes \lambda_0(t) \Bigg[G_{CRI,2}^*(\beta_0, t) + G_{\Sigma I,2}^*(b_T, t, 0) + \frac{G_{CR,1}^{*\otimes 2}(b_T, t, 0)}{G_{CR,0}^{*2}(b_T, t, 0)} G_{II,0}^*(\beta_0, t) \\
&\qquad + G_{X\Sigma I}^*(b_T, \beta_0, t, 0) + G_{X\Sigma I}^{*T}(b_T, \beta_0, t, 0) \\
&\qquad - G_{CRI,1}^*(\beta_0, t) \frac{G_{CR,1}^{*T}(b_T, t, 0)}{G_{CR,0}^*(b_T, t, 0)} - \frac{G_{CR,1}^*(b_T, t, 0)}{G_{CR,0}^*(b_T, t, 0)} G_{CRI,1}^{*T}(\beta_0, t) \\
&\qquad - G_{\Sigma I,1}^*(b_T, \beta_0, t, 0) \frac{G_{CR,1}^{*T}(b_T, t, 0)}{G_{CR,0}^*(b_T, t, 0)} - \frac{G_{CR,1}^*(b_T, t, 0)}{G_{CR,0}^*(b_T, t, 0)} G_{\Sigma I,1}^{*T}(b_T, \beta_0, t, 0) \Bigg] \\
&\quad + o_p(1).
\end{aligned}
\tag{A.10}
$$

With some straightforward algebra, it can be shown that

$$
\begin{aligned}
G_{CRI,2}^*(\beta_0, t) &= G_{II,2}^*(\beta_0, t) + G_{II,0}^*(\beta_0, t) E \{ \Sigma(t) \}, \\
G_{\Sigma I,2}^*(b, t, 0) &= G_{II,0}^*(\beta_0, t, 0) E \left[\Sigma(t) b b^T \Sigma^T(t) \right], \\
G_{X\Sigma I}^*(b, \beta_0, t, 0) &= G_{II,1}^*(\beta_0, t) E^T \left[\Sigma(t) b_0 \right].
\end{aligned}
\tag{A.11}
$$

Applying (A.6), (A.11) and Lemma 1, (A.10) can be written as

$$\langle \varsigma_n, \varsigma_n \rangle = \Omega(t) + o_p(1), \tag{A.12}$$

25

where

$$\Omega(t) = Q(\nu) \otimes \lambda_0(t) \left(G_{II,2}^*(\beta_0, t) - \frac{G_{II,1}^{*\otimes 2}(\beta_0, t)}{G_{II,0}^*(\beta_0, t)} + G_{II,0}^*(\beta_0, t) \left[E\{\Sigma(t)\} + var\{\Sigma(t)b_T\} \right] \right).$$

Then we just need to verify the Lindeberg Condition. Let

$$\varsigma_{ni}(u) = \hat{X}_i(u) + \Sigma_i(u)\{b_{T0} + (u-t)b_{T1}\} - \frac{\hat{G}_{CR,1}^*(b_T, u, u-t)}{\hat{G}_{CR,0}^*(b_T, u, u-t)}.$$

Given any $\varepsilon > 0$, the Lindeberg Condition can be written as

$$n^{-1}h \sum_{i=1}^n \int_0^L \left[K_h^2(u-t)\varsigma_{ni}^2(u)I\left\{ n^{-1/2}h^{1/2}K_h(u-t)|\varsigma_{ni}(u)| > \varepsilon \right\} \right.$$
$$\left. \times \lambda_0(u)Y_i(u)\exp\{\beta_0^T(u)X_i(u)\} \right] du \xrightarrow{p} 0, \tag{A.13}$$

$$n^{-1}h \sum_{i=1}^n \int_0^L \left[(h^{-1}u_t)^2 K_h^2(u-t)\varsigma_{ni}^2(u)I\left\{ n^{-1/2}h^{1/2}h^{-1}u_t K_h(u-t)|\varsigma_{ni}(u)| > \varepsilon \right\} \right.$$
$$\left. \times \lambda_0(u)Y_i(u)\exp\{\beta_0^T(u)X_i(u)\} du \right] \xrightarrow{p} 0. \tag{A.14}$$

Note that (A.13) implies (A.14). Therefore it is sufficient to show (A.13). This follows from

$$\Pr\left(\sup_{u\in\mathcal{N}(t)} \left| n^{-1/2}h^{1/2}K_h(u-t)\varsigma_{ni}(u) \right| > \varepsilon \right) \le \Pr\left(\sup_{u\in\mathcal{N}(t)} |\varsigma_{ni}(u)| > \frac{n^{1/4}\varepsilon}{M_K} \right),$$

which, for n large enough, is equal to 0 from Condition A5 and Lemma 2. Therefore we have

$$(nh)^{1/2}\left[\hat{U}_{CR}(b_T) - \frac{1}{2}h^2(\mu_2, \mu_3)^T \otimes \lambda_0(t)\Gamma_0^*(t)\beta_0''(t) \right] \xrightarrow{d} N(0, \Omega(t)). \tag{A.15}$$

By a Taylor series expansion,

$$0 = (nh)^{1/2}\hat{U}_{CR}(\hat{b}) = (nh)^{1/2}\tilde{U}_{CR}(\hat{\alpha}) = (nh)^{1/2}\tilde{U}_{CR}(0) + (nh)^{1/2}\frac{\partial\tilde{U}_{CR}(\alpha^*)}{\partial\alpha^T}\hat{\alpha},$$

where α^* lies between 0 and $\hat{\alpha}$. This implies that

$$(nh)^{1/2}H\left(\hat{b} - b_0\right) = \left\{ -\frac{\partial\tilde{U}_{CR}(\alpha^*)}{\partial\alpha^T} \right\}^{-1}(nh)^{1/2}\hat{U}_{CR}(b_0). \tag{A.16}$$

26

Note that $\partial \tilde{U}_{CR}(\alpha)/\partial \alpha^T$ can be written as

$$n^{-1} \sum_{i=1}^{n} \int_0^L K_h(u-t) \left(H^{*-1}u_t\right)^{\otimes 2} \otimes \Sigma_i(u)dM_i(u) \tag{A.17}$$

$$-\int_0^L K_h(u-t) \left(H^{*-1}u_t\right)^{\otimes 2} \otimes \lambda_0(u)\hat{G}^*_{\Sigma I,1}(\beta_0,u,u-t)du \tag{A.18}$$

$$-\int_0^L K_h(u-t) \left(H^{*-1}u_t\right)^{\otimes 2}$$

$$\otimes \left[\frac{\hat{G}^*_{CR,2}(H^{-1}\alpha+b_T,u,u-t) - \hat{G}^*_{X\Sigma}(H^{-1}\alpha+b_T,u,u-t)}{\hat{G}^*_{CR,0}(H^{-1}\alpha+b_T,u,u-t)}\right.$$

$$-\frac{\hat{G}^*_{CR,1}(H^{-1}\alpha+b_T,u,u-t)}{\hat{G}^{*2}_{CR,0}(H^{-1}\alpha+b_T,u,u-t)}$$

$$\left.\times \left\{\hat{G}^{T*}_{CR,1}(H^{-1}\alpha+b_T,u,u-t) - \hat{G}^{*T}_{\Sigma,1}(H^{-1}\alpha,u,u-t)\right\}\right]d\hat{\mathcal{E}}N(u) \tag{A.19}$$

$$+o_p(1).$$

By the Lenglart's Inequality, we can show that the martingale (A.17) converges to 0 as its predictable variation process converges to 0. Moreover, uniformly for $\alpha \in \mathcal{N}(0)$,

$$(A.18) \xrightarrow{p} Q_\mu \otimes \lambda_0(t)G^*_{II,0}(\beta_0,t,0)E\left[\Sigma(t)\right],$$

and

$$(A.19) \xrightarrow{p} -\int_0^L K(w) \begin{pmatrix} 1 & w \\ w & w^2 \end{pmatrix} \otimes \left[\frac{G^*_{CR,2}(\alpha+b_T^*,t,w) - G^*_{X\Sigma}(\alpha+b_T^*,t,w)}{\hat{G}^*_{CR,0}(\alpha+b_T^*,t,w)}\right.$$

$$\left.-\frac{G^*_{CR,1}(\alpha+b_T^*,t,w)}{G^{*2}_{CR,0}(\alpha+b_T^*,t,w)}\left\{G^{T*}_{CR,1}(\alpha+b_T^*,t,w) - G^{*T}_{\Sigma,1}(\alpha+b_T^*,t,w)\right\}\right]dwE'\left\{N(t)\right\}. \tag{A.20}$$

From (A.6), the right side of (A.20) can be simplified as

$$-\Gamma(\alpha,t) - Q_\mu \otimes \lambda_0(t)G^*_{II,0}(\beta_0,t,0)E\left[\Sigma(t)\right],$$

where

$$\Gamma(\alpha, t) = \int_0^L K(w) \begin{pmatrix} 1 & w \\ w & w^2 \end{pmatrix}$$

$$\otimes \left[\frac{G_{I,2}^*(\alpha + b_T^*, t, w)}{\hat{G}_{I,0}^*(\alpha + b_T^*, t, w)} - \frac{G_{I,1}^*(\alpha + b_T^*, t, w)G_{I,1}^{*T}(\alpha + b_T^*, t, w)}{G_{I,0}^{*2}(\alpha + b_T^*, t, w)} \right] dw$$

$$\times \lambda_0(t) G_{II,0}^*(\beta_0, t, 0).$$

Hence

$$\frac{\partial \tilde{U}_{CR}(\alpha)}{\partial \alpha^T} \xrightarrow{p} -\Gamma(\alpha, t)$$

uniformly for $\alpha \in \mathcal{N}(0)$. This, coupled with the consistency of $\hat{\alpha}$ and the continuity of $\Gamma(\alpha, t)$ at α obtained

from Condition A6, implies that

$$\frac{\partial \tilde{U}_{CR}(\alpha^*)}{\partial \alpha^T} \xrightarrow{p} \Gamma_0(t) = \Gamma(0, t) = Q_\mu \otimes \lambda_0(t) \Gamma_0^*(t). \tag{A.21}$$

Combining (A.15), (A.16) and (A.21) and the Conditions A3, A8, A9 and A10, we have

$$(nh)^{1/2} \left[H\left(\hat{b} - b_0\right) - \frac{1}{2} h^2 \left\{ Q_\mu^{-1} \left(\mu_2, \mu_3\right)^T \right\} \otimes \beta_0''(t) \right] \xrightarrow{d} N\left(0, \Gamma_0^{-1}(t)\Omega(t) \left\{ \Gamma_0^{-1}(t) \right\}^T \right).$$

Note that $\Omega(t) = \Gamma_0(t) + \Gamma_1(t)$, where

$$\Gamma_1(t) = Q_\mu \otimes G_{II,0}^*(b_0, t, 0) \left[E\left\{ \Sigma(t) \right\} + var\left\{ \Sigma(t)b_0 \right\} \right].$$

Therefore

$$\Gamma_0^{-1}(t)\Omega(t) \left\{ \Gamma_0^{-1}(t) \right\}^T = \Gamma_0^{-1}(t) + \Gamma_0^{-1}(t)\Gamma_1(t) \left\{ \Gamma_0^{-1}(t) \right\}^T.$$

Note that $\Gamma_0^{-1}(t)$ is the asymptotic variance for the ideal estimator when there is no measurement error.

Hence the corrected score estimator has a larger variance than that of the ideal estimator as we expect.

A consistent estimator of $\Omega(t)$ is

$$
\begin{aligned}
\hat{\Omega}(t,\hat{b}) &= n^{-1}\sum_{i=1}^{n}\int_0^L hK_h^2(u-t)\left(H^{*-1}u_t\right)^{\otimes 2} \\
&\quad \otimes\left\{\hat{X}_i(u)+\Sigma_i(u)\left\{\hat{b}_0+(u-t)\hat{b}_1\right\}-\frac{\hat{G}_{CR,1}^*(\hat{b},u,u-t)}{\hat{G}_{CR,0}^*(\hat{b},u,u-t)}\right\}^{\otimes 2}dN_i(u).
\end{aligned}
$$

The consistency of $\hat{\Omega}(t,\hat{b})$ follows from (A.12) and $\hat{\Omega}(t,b_T)-\langle\varsigma_n,\varsigma_n\rangle=\chi_n+o_p(1)$, where

$$
\begin{aligned}
\chi_n &= n^{-1}\sum_{i=1}^{n}\int_0^L hK_h^2(u-t)\left(H^{*-1}u_t\right)^{\otimes 2} \\
&\quad \otimes\left\{\hat{X}_i(u)+\Sigma_i(u)\left\{b_{T0}+(u-t)b_{T1}\right\}-\frac{\hat{G}_{CR,1}^*(b_T,u,u-t)}{\hat{G}_{CR,0}^*(b_T,u,u-t)}\right\}^{\otimes 2}dM_i(u).
\end{aligned}
$$

We can show the martingale $\chi_n=o_p(1)$ by the Lenglart's Inequality. In addition, $\Omega(t,\hat{b})-\Omega(t,b_T)=o_p(1)$ follows from $H(\hat{b}-b_T)=o_p(1)$. Likewise, we can show that a consistent estimator for $\Gamma_0(t)$ is $\hat{\Gamma}(t)=-\partial\tilde{U}_{CR}(\alpha^*)/\partial\alpha^T=-\partial\hat{U}_{CR}(\hat{b})/\partial b^T H^{-1}$. Therefore a consistent estimator for the variance of $n^{-1}(\hat{\beta}-\beta_T)$ is $\hat{\Gamma}^{-1}(t)\hat{\Omega}(t,\hat{b})\left\{\hat{\Gamma}^{-1}(t)\right\}^T$.

APPENDIX B: PROOF OF THEOREM 3

First we show that

$$
(nh)^{1/2}\left\{\hat{U}_{CD}(b,t)-\hat{U}_{CR}(b,t)\right\}=o_p(1) \tag{A.22}
$$

uniformly for $b\in\mathcal{N}(b_0)$. Note that under Condition A2, $G_{CD,r}^*(b,u,u-t)=G_{CR,r}^*(b,u,u-t)$ for $r=0,1$. Thus

$$
\begin{aligned}
&(nh)^{1/2}\left\{\hat{U}_{CD}(b,t)-\hat{U}_{CR}(b,t)\right\} \\
&= -(nh)^{1/2}(nH)^{-1}\sum_{i=1}^{n}\int_0^L K_h(u-t) \\
&\quad \times\left\{\frac{\hat{G}_{CD,1}(b,u,u-t)}{\hat{G}_{CD,0}(b,u,u-t)}-\frac{\hat{G}_{CR,1}(b,u,u-t)}{\hat{G}_{CR,0}(b,u,u-t)}\right\}dN_i(u) \\
&= -h^{1/2}\int_0^L K_h(u-t)\left(H^{*-1}u_t\right)\otimes\left\{\frac{\hat{G}_{CD,1}^*(b,u,u-t)}{\hat{G}_{CD,0}^*(b,u,u-t)}-\frac{G_{CD,1}^*(b,u,u-t)}{G_{CD,0}^*(b,u,u-t)}\right\}
\end{aligned}
$$

29

$$\times n^{1/2} d\widehat{\mathcal{E}}M(u) \tag{A.23}$$

$$+h^{1/2}\int_0^L K_h(u-t)\left(H^{*-1}u_t\right)\otimes\left\{\frac{\hat{G}^*_{CR,1}(b,u,u-t)}{\hat{G}^*_{CR,0}(b,u,u-t)}-\frac{G^*_{CR,1}(b,u,u-t)}{G^*_{CR,0}(b,u,u-t)}\right\}$$

$$\times n^{1/2}d\widehat{\mathcal{E}}M(u) \tag{A.24}$$

$$-h^{1/2}\int_0^L K_h(u-t)\left(H^{*-1}u_t\right)$$

$$\otimes n^{1/2}\left\{\frac{\hat{G}^*_{CD,1}(b,u,u-t)}{\hat{G}^*_{CD,0}(b,u,u-t)}-\frac{G^*_{CD,1}(b,u,u-t)}{G^*_{CD,0}(b,u,u-t)}\right\}du \tag{A.25}$$

$$-h^{1/2}\int_0^L K_h(u-t)\left(H^{*-1}u_t\right)$$

$$\otimes n^{1/2}\left\{\frac{\hat{G}^*_{CR,1}(b,u,u-t)}{\hat{G}^*_{CR,0}(b,u,u-t)}-\frac{G^*_{CR,1}(b,u,u-t)}{G^*_{CR,0}(b,u,u-t)}\right\}du. \tag{A.26}$$

Note that under Condition A5, $\sup_{u\in\mathcal{N}(t),b\in\mathcal{N}(b_0)}\|\hat{G}^*_{CD,r}(b,u,u-t)-G^*_{CD,r}(b,u,u-t)\|=o_p(1)$, and $n^{1/2}d\widehat{\mathcal{E}}M(u)$ converges to a Gaussian process. By the strong embedding theorem and Lemma 6, we have $\sup_{b\in\mathcal{N}(b_T)}\|(A.23)\|=o_p(1)$. Similarly, we can show that $\sup_{b\in\mathcal{N}(b_T)}\|(A.24)\|=o_p(1)$, $\sup_{b\in\mathcal{N}(b_T)}\|(A.25)\|=o_p(1)$, and $\sup_{b\in\mathcal{N}(b_T)}\|(A.26)\|=o_p(1)$. Therefore, $\sup_{b\in\mathcal{N}(b_T)}\|(A.22)\|=o_p(1)$.

Likewise, we can show that

$$\left\{\frac{\partial U_{CD}(b,t)}{\partial b}-\frac{\partial U_{CR}(b,t)}{\partial b}\right\}H^{-1}=o_p(1). \tag{A.27}$$

Then $(nh)^{1/2}H(\tilde{b}-\hat{b})=o_p(1)$ follows from (A.22), (A.27) and a Taylor series expansion.

REFERENCES

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

Bycott, P. and Taylor, J. (1998), "A Comparison of Smoothing Techniques for CD4 Data Measured With Error in a Time-Dependent Cox Proportional Hazards Model," *Statistics in Medicine*, 17, 2061–2077.

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

Dafni, U.G. and Tsiatis, A.A. (1998), "Evaluating Surrogate Markers of Clinical Outcome Measured With Error," *Biometrics*, 54, 1445–1462.

DeGruttola, V. and Tu, X.M. (1994), "Modeling Progression of CD-4 Lymphocyte Count and its Relationship to Survival Time," *Biometrics*, 50, 1003–1014.

Diggle, P.J., Heagerty, P.J., Liang, K.-Y. and Zeger, S.L. (2002), *Analysis of Longitudinal Data*, New York: Oxford.

Faucett, C.J. and Thomas, D.C. (1996), "Simultaneously Modeling Censored Survival Data and Repeatedly Measured Covariates: a Gibbs Sampling Approach," *Statistics in Medicine*, 15, 1663–1685.

Henderson, R., Diggle, P., and Dobson, A. (2000), "Joint Modeling of Longitudinal Measurements and Event Time Data," *Biostatistics*, 4, 465–480.

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

Pawitan, Y. and Self, S. (1993), "Modeling Disease Marker Processes in AIDS," *Journal of the American Statistical Association*, 83, 719–726.

Prentice, R. (1982), "Covariate Measurement Errors and Parameter Estimates in a Failure Time Regression Model," *Biometrika*, 69, 331–42.

Song, X., Davidian, M. and Tsiatia, A.A. (2002a), "An Estimator for the Proportional Hazards Model With Multiple Longitudinal Covariates Measured With Error," *Biostatistics*, 3, 511–528.

——— (2002b), "A Semiparametric Likelihood Approach to Joint Modeling of Longitudinal and Time-to-Event Data," *Biometrics*, 58, 742–753.

Song, X. and Huang, Y. (2005a), "On Corrected Score Approach for Proportional Hazards Model With Covariate Measurement Error," *Biometrics* 61, 702–714.

——— (2005b), "A Corrected Pseudo-score Approach for Additive Hazards Model With Longitudinal

Covariates Measured With Error," *Lifetime Data Analysis*, in press.

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

Tsiatis, A.A. and Davidian, M. (2001), "A Semiparametric Estimator for the Proportional Hazards Model With Longitudinal Covariates Measured With Error," *Biometrika*, 88, 447–458.

Tsiatis, A.A., DeGruttola, V. and Wulfsohn, M.S. (1995), "Modeling the Relationship of Survival to Longitudinal Data Measured With Error: Applications to Survival and CD4 Counts in Patients With AIDS," *Journal of the American Statistical Association*, 90, 27–37.

Wang, C.Y. (2005), "Corrected Score Estimator for Joint Modeling of Longitudinal and Failure Time Data," *Statistica Sinica*, to appear.

Wang, C.Y., Wang, N, and Wang, S. (2000), "Regression Analysis When Covariates are Regression Parameters of a Random Effect Model for Observed Longitudinal Measurements," *Biometrics*, 56, 487–495.

Winnett, A., and Sasieni, P. (2003), "Iterated Residuals and Time-Varying Covariate Effects in Cox Regression," *Journal of the Royal Statistical Society*, Ser. B, 65, 473–488.

Wulfsohn, M.S. and Tsiatis, A.A. (1997), "A joint Model For Survival and Longitudinal Data Measured With Error," *Biometrics*, 53, 330–339.

Xu, J. and Zeger, S.L. (2001), "Joint Analysis of Longitudinal Data Comprising Repeated Measures and Times to Events," *Applied Statistics*, 50, 375–387.

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

http://biostats.bepress.com/uwbiostat/paper273

Table 1. Simulation results in the case of a single time-dependent covariate with $\beta_0(t) = 1$ and $\sigma^2 = 0.1$. I, "ideal" approach; NR, naive regression; CD, conditional score; CR, corrected score; B, bias; SD, empirical standard deviation across simulated data sets ; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

t		10				20				30				40			
		B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP
$h = 20$	I	−0.006	0.334	0.310	0.926	−0.008	0.101	0.096	0.940	−0.013	0.095	0.086	0.931	−0.022	0.126	0.113	0.928
	NR	0.035	0.319	0.297	0.923	0.047	0.096	0.092	0.897	0.036	0.090	0.083	0.897	0.015	0.120	0.110	0.912
	CD	−0.010	0.364	0.334	0.922	−0.012	0.114	0.107	0.945	−0.017	0.107	0.097	0.930	−0.027	0.137	0.123	0.924
	CR	−0.011	0.371	0.343	0.929	−0.017	0.116	0.109	0.945	−0.023	0.110	0.099	0.930	−0.034	0.141	0.126	0.927
$h = 40$	I	−0.006	0.184	0.169	0.932	−0.009	0.097	0.092	0.939	−0.012	0.075	0.073	0.947	−0.015	0.092	0.086	0.930
	NR	0.053	0.183	0.163	0.910	0.044	0.096	0.089	0.889	0.034	0.074	0.070	0.890	0.022	0.091	0.083	0.907
	CD	−0.011	0.212	0.186	0.919	−0.014	0.114	0.103	0.931	−0.016	0.087	0.081	0.936	−0.019	0.104	0.093	0.931
	CR	−0.014	0.216	0.191	0.921	−0.019	0.116	0.105	0.931	−0.022	0.089	0.082	0.933	−0.026	0.107	0.095	0.929
$h = 60$	I	−0.002	0.142	0.137	0.932	−0.006	0.094	0.089	0.928	−0.010	0.077	0.071	0.927	−0.015	0.089	0.082	0.935
	NR	0.062	0.140	0.132	0.912	0.050	0.091	0.086	0.885	0.037	0.074	0.068	0.865	0.023	0.088	0.080	0.904
	CD	−0.004	0.164	0.152	0.931	−0.009	0.106	0.100	0.937	−0.014	0.086	0.079	0.935	−0.020	0.100	0.089	0.937
	CR	−0.007	0.166	0.156	0.937	−0.013	0.108	0.102	0.938	−0.019	0.088	0.081	0.936	−0.026	0.103	0.091	0.936

Table 2. Simulation results in the case of a single time-dependent covariate with $\beta_0(t) = 1$ and $\sigma^2 = 0.2$. I, "ideal" approach; NR, naive regression; CD, conditional score; CR, corrected score; B, bias; SD, empirical standard deviation across simulated data sets ; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

t		10				20				30				40			
		B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP
$h = 20$	I	−0.006	0.334	0.310	0.926	−0.008	0.101	0.096	0.940	−0.013	0.095	0.086	0.931	−0.022	0.126	0.113	0.928
	NR	0.074	0.308	0.286	0.913	0.093	0.093	0.089	0.773	0.076	0.087	0.081	0.789	0.047	0.117	0.107	0.871
	CD	−0.011	0.401	0.361	0.925	−0.019	0.131	0.120	0.946	−0.023	0.122	0.108	0.934	−0.023	0.122	0.108	0.934
	CR	0.006	0.584	0.426	0.920	−0.027	0.139	0.125	0.941	−0.042	0.199	0.118	0.933	−0.052	0.197	0.147	0.919
$h = 40$	I	−0.006	0.184	0.169	0.932	−0.009	0.097	0.092	0.939	−0.012	0.075	0.073	0.947	−0.015	0.092	0.086	0.930
	NR	0.105	0.179	0.158	0.863	0.091	0.095	0.086	0.759	0.074	0.073	0.068	0.750	0.055	0.090	0.081	0.823
	CD	−0.015	0.241	0.205	0.915	−0.020	0.132	0.116	0.922	−0.022	0.101	0.091	0.938	−0.024	0.116	0.101	0.924
	CR	−0.018	0.253	0.238	0.933	−0.030	0.136	0.121	0.926	−0.034	0.105	0.095	0.931	−0.039	0.123	0.110	0.935
$h = 60$	I	−0.002	0.142	0.137	0.932	−0.006	0.094	0.089	0.928	−0.010	0.077	0.071	0.927	−0.015	0.089	0.082	0.935
	NR	0.115	0.137	0.128	0.837	0.096	0.088	0.083	0.745	0.077	0.072	0.066	0.742	0.056	0.087	0.078	0.826
	CD	−0.008	0.186	0.169	0.919	−0.014	0.122	0.112	0.933	−0.019	0.097	0.089	0.931	−0.026	0.111	0.097	0.934
	CR	−0.015	0.202	0.196	0.955	−0.024	0.127	0.118	0.937	−0.031	0.101	0.092	0.933	−0.039	0.115	0.105	0.947

Table 3. Simulation results in the case of a single time independent covariate with $\beta_0(t) = t$ and $\sigma^2 = 0.1$. I, "ideal" approach; NR, naive regression; CD, conditional score; CR, corrected score; B, bias; SD, empirical standard deviation across simulated data sets ; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

t		0.25				0.5				0.75				1.0			
		B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP
$h = 0.4$	I	0.007	0.087	0.086	0.947	0.011	0.104	0.100	0.937	0.025	0.164	0.149	0.934	0.062	0.301	0.262	0.911
	NR	−0.006	0.084	0.084	0.953	−0.017	0.100	0.097	0.936	−0.025	0.157	0.144	0.919	−0.021	−0.021	0.251	0.896
	CD	0.007	0.088	0.089	0.954	0.011	0.109	0.104	0.939	0.030	0.179	0.160	0.930	0.085	0.356	0.294	0.916
	CR	0.007	0.088	0.089	0.954	0.012	0.109	0.105	0.938	0.034	0.181	0.162	0.930	0.061	1.374	0.313	0.923
$h = 0.8$	I	0.000	0.085	0.083	0.937	0.004	0.086	0.084	0.941	0.012	0.131	0.124	0.938	0.032	0.224	0.204	0.935
	NR	−0.013	0.083	0.081	0.938	−0.024	0.082	0.081	0.932	−0.037	0.125	0.120	0.923	−0.046	0.217	0.195	0.908
	CD	0.000	0.088	0.085	0.942	0.006	0.089	0.088	0.958	0.016	0.140	0.133	0.936	0.043	0.251	0.222	0.928
	CR	0.000	0.088	0.085	0.942	0.008	0.090	0.089	0.957	0.021	0.142	0.135	0.938	0.061	0.264	0.231	0.936
$h = 1.2$	I	0.002	0.084	0.083	0.943	0.009	0.082	0.081	0.945	0.017	0.124	0.121	0.937	0.028	0.183	0.178	0.951
	NR	−0.010	0.082	0.081	0.946	−0.020	0.081	0.079	0.938	−0.031	0.125	0.117	0.912	−0.041	0.186	0.172	0.913
	CD	0.003	0.087	0.085	0.946	0.012	0.090	0.086	0.939	0.024	0.142	0.130	0.942	0.040	0.211	0.192	0.933
	CR	0.003	0.087	0.085	0.946	0.015	0.091	0.087	0.938	0.031	0.145	0.133	0.939	0.054	0.219	0.198	0.934

Table 4. Simulation results in the case of a single time independent covariate with $\beta_0(t) = t$ and $\sigma^2 = 0.2$. I, "ideal" approach; NR, naive regression; CD, conditional score; CR, corrected score; B, bias; SD, empirical standard deviation across simulated data sets ; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

| t		0.25				0.5				0.75				1.0			
		B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP
$h = 0.4$	I	0.007	0.087	0.086	0.947	0.011	0.104	0.100	0.937	0.025	0.164	0.149	0.934	0.062	0.301	0.262	0.911
	NR	−0.017	0.082	0.082	0.954	−0.042	0.097	0.095	0.916	−0.069	0.151	0.140	0.879	−0.094	0.281	0.240	0.866
	CD	0.008	0.091	0.091	0.951	0.013	0.113	0.109	0.944	0.036	0.196	0.172	0.930	0.114	0.428	0.332	0.924
	CR	0.008	0.091	0.091	0.951	0.014	0.114	0.110	0.942	0.048	0.269	0.178	0.926	0.058	2.476	0.486	0.801
$h = 0.8$	I	0.000	0.085	0.083	0.937	0.004	0.086	0.084	0.941	0.012	0.131	0.124	0.938	0.032	0.224	0.204	0.935
	NR	−0.024	0.081	0.079	0.926	−0.050	0.079	0.079	0.895	−0.080	0.121	0.116	0.858	−0.113	0.210	0.188	0.855
	CD	0.000	0.090	0.087	0.944	0.008	0.094	0.093	0.956	0.020	0.151	0.142	0.941	0.055	0.282	0.242	0.930
	CR	0.001	0.090	0.087	0.943	0.011	0.096	0.094	0.954	0.031	0.157	0.147	0.943	0.028	1.649	0.273	0.905
$h = 1.2$	I	0.002	0.084	0.083	0.943	0.009	0.082	0.081	0.945	0.017	0.124	0.121	0.937	0.028	0.183	0.178	0.951
	NR	−0.022	0.080	0.079	0.944	−0.047	0.080	0.077	0.892	−0.074	0.123	0.114	0.858	−0.102	0.183	0.166	0.864
	CD	0.003	0.089	0.087	0.947	0.015	0.097	0.091	0.937	0.030	0.157	0.140	0.933	0.050	0.237	0.207	0.919
	CR	0.003	0.089	0.088	0.947	0.021	0.100	0.093	0.935	0.045	0.165	0.146	0.936	0.079	0.251	0.222	0.926
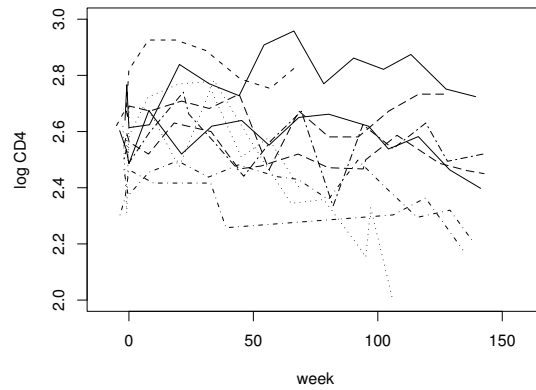
Figure 1. Trajectories of log CD4 for 10 randomly selected subjects.
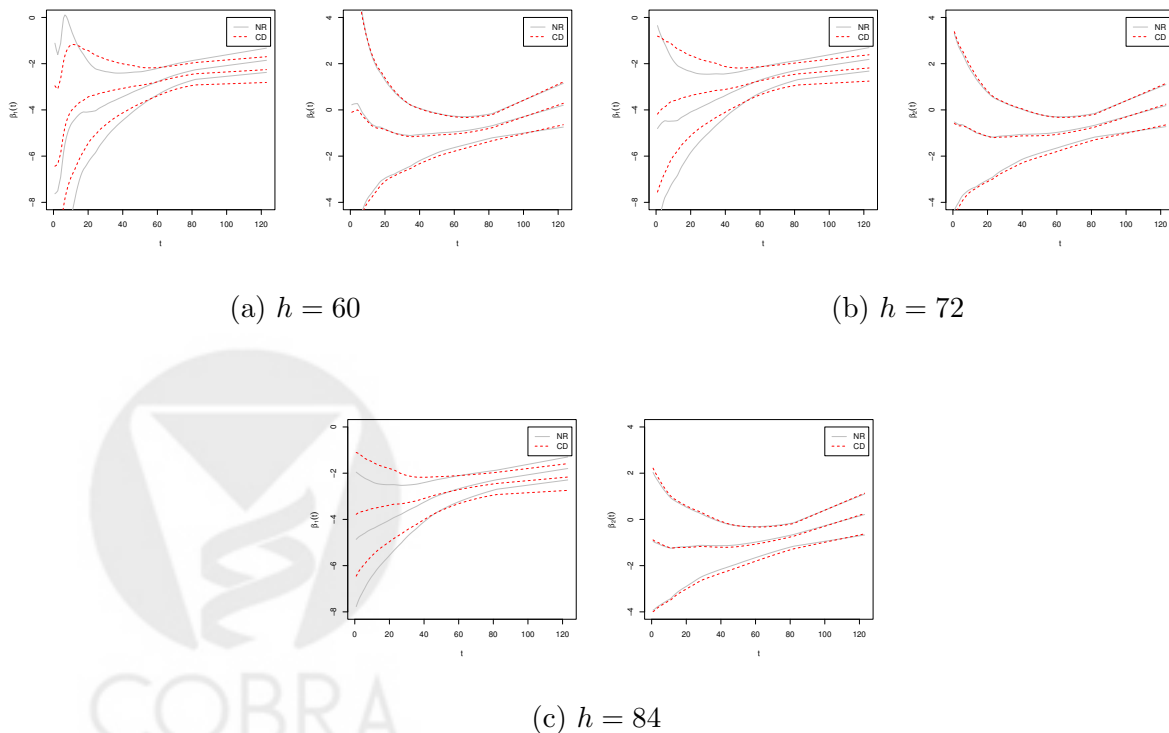
(a) $h = 60$

(b) $h = 72$

(c) $h = 84$

Figure 2. Estimation of $\beta(t)$ for the ACTG 175 data including CD4 and treatment with $h = 60, 72$ and 80. For each h, the left panel is for $\beta_1(t)$, and the right panel is for $\beta_2(t)$. NR, naive regression; CD, conditional score. 95% confidence bands are shown with the outer curves, the estimates themselves are shown with the center curves.

37