12-19-2003

Partly Conditional Survival Models for Longitudinal Data

Yingye Zheng

Fred Hutchinson Cancer Research Center, yzheng@fhcrc.org

Patrick Heagerty

University of Washington, heagerty@u.washington.edu

1. Introduction

Methods that can quantify the risk of death or disease as a function of current or past covariate measures can provide medical predictions which are used to guide patient care. Altman and De Stavola (1994) comment: "In clinical practice individual data are routinely collected at frequent time points after entry to a study... but are rarely examined in relation to survival. Yet a key clinical question is that of prognosis, and a means of updating prognosis on the basis of the latest observations on a patient would be valuable to many clinicians." (page 321). For example, cystic fibrosis patients are registered for lung transplantation once their pulmonary function has declined to less than 30% of the patient's expected value as determined by gender and age (Davis 1997). Similarly, in cancer screening the advent of new molecular assays has lead to a number of measurements that may have the ability to signal cancer onset. For prostate cancer the serum measurement PSA has been studied (Etzioni, Pepe, Longton, Hu and Goodman 1999; Slate and Turnbull 2000) and for ovarian cancer the marker CA-125 has been used as a possible early indicator of disease (Skates, Pauler and Jacobs 2001). These examples illustrate the common biomedical data structure where a longitudinal measurement is taken at select follow-up times, and the scientific question focuses on the correlation between the longitudinal series and the time until a key clinical event.

One approach that links the time until an event to time-dependent covariates is the proportional hazards model of Cox (1972). Details of the time-varying covariate proportional hazards model are given in monographs by Kalbfleisch and Prentice (1980), Cox and Oakes (1984), and Andersen, Borgan, Gill and Keiding (1995). Briefly, in a time-varying proportional hazards model the instantaneous risk of death is modeled as a function of the current value of the measured covariate. Let $Z_i(t)$ denote the value of the covariate for subject i at time t, and let T_i denote the time until a major clinical endpoint (ie. disease or death). Classical survival analysis methods for time-varying covariates can be used to model the instantaneous risk of death, or hazard defined as $\lambda[t \mid H_i^Z(t)] = \lim_{\delta \to 0} \frac{1}{\delta} P[T_i \in [t, t+\delta) \mid H_i^Z(t); T_i \geq t]$ where we condition on the entire covariate history $H_i^Z(t) = \{ Z_i(u) : u \leq t \}$. The hazard may depend on additional aspects of the measured history beyond the current measurement. However, for simplicity of presentation we assume that the hazard depends only on the present value of the marker. The proportional hazards model assumes

$\lambda[t \mid Z_i(t)] = \lambda_0(t) \cdot \exp[\beta \cdot Z_i(t)]$, where $\lambda_0(t)$ represents a baseline hazard function.

The use of the time-varying covariate model typically assumes that $Z_i(t)$ is available for all possible times. However, in practice we almost never observe $Z_i(t)$ continuously in time. Rather, we commonly measure the covariate process at discrete times $s_{i1}, s_{i2}, \ldots, s_{in_i}$. Altman and De Stavola (1994) discuss some of the practical issues associated with discrete covariate measurement. Prentice (1982) discusses issues of bias that result from mismeasurement of covariates, and recent literature has developed models that require a joint model for covariate process and failure time (Pawitan and Self 1993; Tsiatis, DeGruttola and Wulfsohn 1995; Faucett and Thomas 1996; Wulfsohn and Tsiatis 1997; Henderson, Diggle and Dobson 1997).

The time-varying covariate hazard model is particularly useful for estimating regression parameters. However, unlike Cox models without time-varying covariates, estimates of survival probabilities are generally difficult to obtain, and require a model for the covariate process. In practice we may monitor a patient's vital status through time s and have available a discretely measured covariate history $H_i^Z(s) = \{ Z_i(s_{ik}) : s_{ik} \leq s \}$ and seek to estimate patient prognosis on the basis of observed data. For example, when using a model for $\lambda[t \mid H_i^Z(t)]$ estimating $P[T_i > t \mid H_i^Z(s), 0 \leq s < T_i]$ would require either knowledge of the future values of $Z_i(t)$, or integration over the conditional distribution of the future covariate process given the history: $\{ Z_i(u) : u > s \mid H_i^Z(s) \}$. A general method is desired that can directly structure and estimate $P[T_i > t \mid Z_i(s), 0 \leq s < T_i]$, for any pair of survival and measurement times, (s, t) where $s < t$.

In this manuscript we introduce a semiparametric method, which we call a "partly conditional survival model", for estimating the prognostic effect of longitudinal measurements on survival without relying on multivariate assumptions regarding the longitudinal marker proccess. Here we address the question concerning how well a longitudinal covariate measured at, or up to time s, $Z_i(s)$, predicts the risk of the occurrence of an important clinical event, T_i such as diagnosis or death, by any future time t. Specifically, we are interested in the conditional probabilities $P[T_i > t \mid Z_i(s), 0 \leq s < T_i]$ that can be computed for any time $t > s$ but condition only on the marker value through time s. In order to characterize a general conditional survival distribution we focus our model on hazard functions of

3
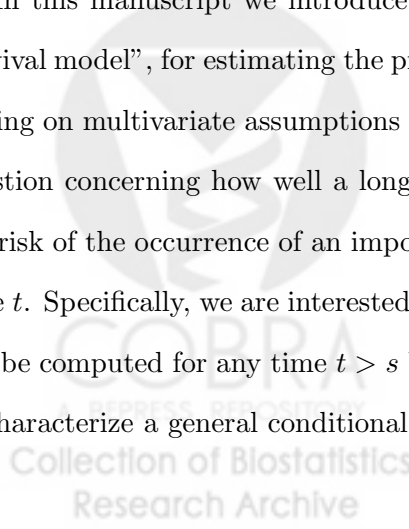
the form

$$\lambda[\,t \mid Z_i(s), 0 \le s < T_i] = \lim_{\delta \to 0} \frac{1}{\delta} \cdot P\left\{\, T_i \in [t, t + \delta) \mid Z_i(s), 0 \le s < T_i; \ t \le T_i\right\} \quad . \qquad (1.1)$$

The basic idea is that we can consider the marker value at time s as a "frozen" or baseline measurement rather than explicitly as a realization of a continuous time stochastic process. Also, in a typical survival model, the event time T_i is usually the time from study entry to the occurrence of the event. To estimate how well a marker can predict subsequent survival, we suggest modeling the time scale in terms of years since measurement, $T_i - s$. As a result there can be multiple derived event times for each individual corresponding to her/his repeatedly measured marker values, $T_{ik} = T_i - s_{ik}$, where s_{ik} is the kth marker measurement time. Thus we cast the problem of using longitudinal marker values for predicting survival within the general framework of multivariate survival models. Our partly conditional survival model is similar to marginal Cox regression models (Wei, Lin and Weissfeld 1989; Lee, Wei and Amato 1992) in the sense that we do not make any parametric assumptions about the dependence among the survival times from one individual. However, the proposed model is 'partly conditional', since when modelling hazards we do not condition on the full and dynamic covariate history, $H_i^Z(t)$, but rather on a static covariate subset, $H_i^Z(s)$ for fixed $s < t$.

One issue that arises with the use of partly conditional models is the need to allow regression parameters to depend on both the time of measurement for the predictor and the time of measurement for the outcome. Pepe, Heagerty and Whitaker (1999) give examples where the linear predictor for the mean at time t conditional on covariate information through time s takes the form $\beta_0(t, s) + \beta_1(t, s) \cdot Z_i(s)$. In specific examples a varying coefficient model of the form $\beta_1(t, s) = \beta_1(t - s)$ may be used which assumes that the association between the outcome and the covariate depends only on their time separation. In the survival setting we also anticipate the need for time-varying coefficient models since a time-varying measure may not satisfy the standard proportional hazards assumption. A Cox model with coefficient functions such as $\beta_1(t - s)$ may be adopted allowing the association between the marker measured at time s, $Z_i(s)$, and the hazard of death at future times t, to change as the distance, $t - s$, increases.

A formal definition of our partly conditional survival model is given in section 2. In section 3 we

4

describe estimation procedures and characterize large sample properties. In section 4 we discuss simulations that evaluate both coverage probabilities and efficiency of the proposed estimation methods. In section 5 we illustrate the new methods by analyzing a well known data set from HIV research, the Multicenter Aids Cohort Study (MACS) data.

In summary, we have the following specific goals:

(1) To obtain survival predictions, $P[T_i > t \mid Z_i(s), 0 \leq s < t]$, for survival beyond time t conditional on survial to, and the marker process at, time s for any pair of times (s, t) with $s < t$.

(2) To estimate regression functions that characterize the association between $Z_i(s)$ and the hazard of death at time t, recognizing that the regression coefficient of $Z_i(s)$ may depend on both s and t.

(3) To make efficient use of the available longitudinal data by using all marker measurements.

2. Partly Conditional Models in Survival Analysis

2.1 Notation

Let T_i be the time to diagnosis (or failure) for subject i. We assume that T_i may be censored at time C_i, and therefore we only observe $X_i = \min(T_i, C_i)$ and an associated censoring indicator Δ_i where $\Delta_i = 1$ if $X_i = T_i$ and 0 otherwise. Also, assume that each subject in the study has a time-dependent covariate measured $K_i \leq K$ times during follow-up, where K is relatively small compared to the total number of subjects n. Let $\mathbf{Z}_{ik}^T = [\mathbf{Z}_{i \cdot a}^T, Z_{ikb}, s_{ik}]$ denote a vector of covariates associated with subject i measured at time s_{ik}, where $\mathbf{Z}_{i \cdot a}$ denotes a vector of baseline covariates such as treatment or gender, while Z_{ikb}, or equivalently $Z_i(s_{ik})$, denotes a time-varying marker value measured at time s_{ik}.

For the longitudinal analysis setting we need to explicitly state the model assumptions regarding both measurement and missingness. First, we assume that the censoring time, C_i, independent of the survival time T_i Second, we assume that the measurement times, s_{ik}, are independent of the longitudinal marker process, $Z_i(s_{ik})$, and the survival time T_i. Finally, we assume that subjects may have missing marker measurements, but we assume that any such missingness is completely at random (MCAR). In the discussion section we comment on approaches that can relax these measurement and missingness mechanism assumptions.

In the partly conditional survival approach we focus analysis on derived survival times. Corre-

sponding to $\mathbf{Z}_{ik}$, let T_{ik} denote the time from s_{ik} to T_i, and C_{ik} denote the time from s_{ik} to C_i for censored T_i: $T_{ik} = T_i - s_{ik}$ and $C_{ik} = C_i - s_{ik}$. For failure time T_{ik}, one observes a bivariate vector (X_{ik}, Δ_{ik}), where $X_{ik} = min(T_{ik}, C_{ik})$. $\Delta_{ik} = 1$ if $X_{ik} = T_{ik} > 0$ and 0 otherwise. To specify hazards we convert to the derived time scale, $t^* = t - s_{ik}$, which measures the follow-up time since measurement of the marker. We use the standard counting process notation where $N_{ik}(t^*) = I(X_{ik} \leq t^*, \Delta_{ik} = 1)$, and write $dN_{ik}(t*)$ for the increment $N_{ik}[(t^* + dt)-] - N_{ik}(t^*)$. The at-risk process is defined as $R_{ik}(t^*) = I(X_{ik} \geq t^*, T_i > s_{ik})$. In situations where the covariate at time s_{ik} is not measured due to an MCAR mechanism we modify the at-risk process definition: $R_{ik}(t^*) = I(X_{ik} \geq t^*, T_i > s_{ik}) \times O_{ik}$ where $O_{ik} = 1$ if the kth covariate measurement is available, and $O_{ik} = 0$ otherwise. We assume that the random vectors $(\mathbf{X}_i, \boldsymbol{\Delta}_i, \mathbf{R}_i, \mathbf{Z}_i)$ are independent and identically distributed with $\mathbf{Z}_i$ bounded. In addition, we assume the censoring time C_{ik} is independent of T_{ik} conditional on $\mathbf{Z}_{ik}$. Since the measurement time, s_{ik}, is part of $\mathbf{Z}_{ik}$ the conditional independence of $T_{ik} = T_i - s_{ik}$ and $C_{ik} = C_i - s_{ik}$ follows from the assumption of independence for T_i and C_i given the marker process and baseline covariates.

2.2 Partly Conditional Survival Model

We now propose a class of methods that may be used to estimate the survival probability conditional on a longitudinal marker value, $P[T_i > t \mid \mathbf{Z}_{ik}, T_i > s_{ik}]$, or equivalently $P[T_{ik} > t^* \mid \mathbf{Z}_{ik}, T_i > s_{ik}]$, which we call "partly conditional Cox regression models". We define the partly conditional hazard function $\lambda_{ik}(t^*)$ for the derived survival outcomes T_{ik} as

$$\lambda_{ik}(t^* \mid \mathbf{Z}_{ik}, 0 \leq s_{ik} \leq T_i) = \lim_{\delta \to 0} \frac{1}{\delta} \cdot P(t^* \leq T_{ik} < t^* + \delta \mid T_{ik} \geq t^*, \mathbf{Z}_{ik}, T_i \geq s_{ik}) \ .$$

A regression model for the hazard can take the general form: $\lambda_{ik}(t^* \mid \mathbf{Z}_{ik}, 0 \leq s_{ik} \leq T_i) = g\left[\lambda_0(t^*, s), \boldsymbol{\beta}(t^*, s)^T \mathbf{Z}_{ik}\right]$ where $g(\lambda, \eta)$ is a link function. The baseline hazard, $\lambda_0(t^*, s)$, and the regression coefficient, $\boldsymbol{\beta}(t^*, s)$, may be functions of both the time since measurement, t^*, and measurement time, s (or equivalently t and s). The model is "partly conditional" since rather than conditioning on the entire covariate history through time t, the model conditions on the partial history measured through time s. Although we present a general form we will focus on the proportional hazards model.

With the partly conditional model we address several complications that frequently arise when using survival analysis with longitudinal measurements. First, we anticipate that the phenomenon of non-proportional hazards may be more frequently encountered when covariates are updated over time. To this end, we specify a time-varying coefficient Cox model which can be used without imposing any functional form on the coefficient functions. Second, the time at which the measurement is taken, s_{ik}, may be associated with the predictive capacity of $Z_i(s_{ik})$ on survival.

Here we briefly describe some examples of partly conditional survival models that model both $Z_i(s_{ik})$ and the measurement time s_{ik}. The simplest approach would be to create separate regression models for survival beyond key measurement times. Such a stratified method would define G "measurement intervals" $\mathcal{I}_1, ..., \mathcal{I}_G$, that partition time $[t_0, \max(T_i)]$ and then adopt the G models:

$$\lambda_{ik}(t^* \mid \mathbf{Z}_{ik}, 0 \leq s_{ik} \leq T_i, s_{ik} \in \mathcal{I}_g) = \lambda_{0g}(t^*)exp\left[\boldsymbol{\alpha}_g^T \mathbf{Z}_{i \cdot a} + \beta_g(t^*)Z_{ikb}\right], t^* > 0. \tag{2.1}$$

Here $\lambda_{0g}(t^*)$ is an unspecified baseline hazard function, whereas $\boldsymbol{\alpha}_g$ and $\beta_g(t^*)$ are unknown regression parameters that are unique for each measurement interval. If only at most one measurement per subject fell into $\mathcal{I}_g$ then separate baseline Cox models could be estimated with standard methods using the measurement Z_{ikb}, where $s_{ik} \in \mathcal{I}_g$, as a baseline measurement for interval g, and the residual lifetime, $T_i - s_{ik}$ as the outcome, after restricting to those subjects who survive long enough to have a measurement recorded in $\mathcal{I}_g$. Such models are sometimes referred to as "landmark analyses" when only a small number of measurement times are chosen to identify the time origins for analysis. The major limitiation to a fully stratified approach is the lack of parsimony with completely unstructured baseline hazard and relative risks as a function of the measurement time s. The partly conditional models that we propose allow adoption of smooth variation in the coefficient functions $\beta_g(t^*)$ and the baseline hazards by explicitly modelling the measurement time s.

A second model assumes that the prognostic capacity of the longitudinal marker is the same regardless of the time at which it is measured. However, the model may still allow different baseline hazard functions for different measurment intervals $\mathcal{I}_g$:

$$\lambda_{ik}(t^* \mid \mathbf{Z}_{ik}, 0 \leq s_{ik} \leq T_i) = \lambda_{0g}(t^*)exp\left[\boldsymbol{\alpha}^T \mathbf{Z}_{i \cdot a} + \beta(t^*)Z_{ikb}\right], t^* > 0 \ . \tag{2.2}$$

Alternatively, rather than assuming completely separate baseline hazards, $\lambda_{0g}(t^*)$, we may use the

time of measurement as a covariate in a model with a common unspecified baseline hazard. A general
form for this model is

$$\lambda_{ik}(t^* \mid \mathbf{Z}_{ik}, 0 \leq s_{ik} \leq T_i) = \lambda_0(t^*) exp\left\{ \boldsymbol{\alpha}_1^T \mathbf{Z}_{i \cdot a} + \boldsymbol{\alpha_2}^T [f_j(s_{ik})]_{j=1}^p + \beta(t^*) Z_{ikb} \right\}, t^* > 0 \qquad (2.3)$$

where $[f_j(s)]_{j=1}^p = [f_1(s), f_2(s), \dots, f_p(s)]$, and $f_j(s_{ik})$ represent basis functions for some parametric
but flexible function of s_{ik}, such as a cubic spline with fixed knots. The modelling procedure thus
does not rely on the partition of measurement stratum anymore.

Estimation procedures for partly conditional models depend on whether we allow the influence of
covariates to vary with time, $\beta(t^*)$, or assume a constant covariate effect $\beta(t^*) = \beta$. We distinguish
among three classes of partly conditional models: parametric; non-parametric; and partly parametric
hazard models. For a parametric hazard model, the effects of all covariates, including the longitudinal
marker, are assumed constant. For a non-parametric varying-coefficient hazard model, the effects of
all covariates vary with time, and are estimated non-parametrically. Finally, for a partly parametric
hazard model, the influence of only a few covariates varies non-parametrically over time, while the
remaining covariates have time-invariant effects.

3.   Estimation

To estimate the regression parameters and the baseline hazard function we propose use of "working
independence" estimating equations applied to the derived failure time data $(X_{ik}, \Delta_{ik}, R_{ik}, \mathbf{Z}_{ik})$. Since
we have chosen to directly model the partly conditional hazard function for these multiple correlated
failure times, a likelihood-based estimation approach would be analytically and computationally dif-
ficult. First, in a likelihood approach a joint model would be required for the event time and the
repeated measures process. Parameterization of the joint model in terms of the partly conditional
hazards would be analytically difficult as the regression structure we adopt is for pairwise marginal
distributions induced by the joint model. Second, a likelihood-based approach would generally re-
quire proper parametric specification of the longitudinal covariate distribution, and the validity of the
induced partly conditional regression estimates would depend on correct marker model specification.
As an alternative, we develop a direct estimating equation approach that proves computationally
simple and yields consistent estimators under correct specification of the partly conditional regres-

8

sion structure without reliance on any distributional assumptions for the marker process. Sandwich variance estimators permit valid asymptotic inference.

In the subsections below we discuss the estimation of regression parameters for three specific model classes. We first discuss estimation under a standard proportional hazards assumption, and then discuss relaxation to allow a non-parametric varying-coefficient specification. Finally, we discuss a model that allows both parametric and non-parametric components.

3.1 Parametric Proportional Hazards Model

For the situation where the proportional hazard assumption holds, estimation procedures for the partly conditional models are similar to those for marginal survival models (Wei et al. 1989; Lee et al. 1992). We assume that the baseline hazard is a function of the time since measurement, $t^* = t - s$, (hereafter referred to as the identical baseline model), or more generally, a function of both t^* and s (hereafter referred to as the stratified baseline model), and the effect of marker on the failure time is constant over time. Specifically, the hazard function for the ith subject and the failure corresponding to the measurement at s_{ik} is

$$\lambda_{ik}(t^* \mid \mathbf{Z}_{ik}, 0 \le s_{ik} \le T_i) = \lambda_0(t^*) exp(\boldsymbol{\beta}^T \mathbf{Z}_{ik})$$

or in the stratified model

$$\lambda_{ik}(t^* \mid \mathbf{Z}_{ik}, 0 \le s_{ik} \le T_i) = \lambda_{0g}(t^*) exp(\boldsymbol{\beta}^T \mathbf{Z}_{ik})$$

where either $\lambda_0(t^*)$ or $\lambda_{0g}(t^*)$, $g = 1, ..., G$ denote unspecified baseline hazard functions. The unknown parameter $\boldsymbol{\beta}$ can be obtained by solving the "working independence" estimating equation:

$$\sum_{i}^{n} \sum_{k}^{K} \int_{0}^{\tau} \left[\mathbf{Z}_{ik} - \bar{\mathbf{Z}}(u) \right] dN_{ik}(u) = 0$$

where $\tau < inf\{t^* : E[R_{ik}(t^*)] = 0\}$, $\bar{\mathbf{Z}}(u) = \sum_{g=1}^{G} \mathcal{S}_g^{(1)}(\boldsymbol{\beta}, u) / \sum_{g=1}^{G} \mathcal{S}_g^{(0)}(\boldsymbol{\beta}, u)$ under the identical baseline model and $\bar{\mathbf{Z}}(u) = \sum_{g=1}^{G} \mathcal{S}_g^{(1)}(\boldsymbol{\beta}, u) 1(s_{ik} \in \mathcal{I}_g) / \sum_{g=1}^{G} \mathcal{S}_g^{(0)}(\boldsymbol{\beta}, u) 1(s_{ik} \in \mathcal{I}_g)$ under the stratified baseline model, with

$$\mathcal{S}_g^{(j)}(\beta, u) = \frac{1}{n} \sum_{l=1}^{n} \sum_{m=1}^{K} R_{lm}(u) exp(\boldsymbol{\beta}^T \mathbf{Z}_{lm}) \mathbf{Z}_{lm}^{\otimes j} \cdot 1(s_{lm} \in \mathcal{I}_g) \ .$$

9

For a column vector a, $a^{\otimes 0}$ refers to a scalar 1, $a^{\otimes 1}$ refers to the vector a and $a^{\otimes 2}$ refers to the matrix aa^T.

Large sample distributional theory and robust variance-covariance estimation can be developed along the lines of the marginal survival model of Lee et al. (1992) for the identical baseline models, or Wei et al. (1989) for the stratified baseline models.

3.2 Non-parametric Varying-Coefficient Hazard Model

The partly conditional regression model decouples the marker measurement time, s_{ik}, from the time scale for the hazard model, $t^* = t - s_{ik}$ for $t > s_{ik}$. Therefore, the hazard ratio corresponding to the longitudinal marker, $\lambda_{ik}(t^* \mid \mathbf{Z}_{i \cdot a}, Z_{ikb} = (z+1), s_{ik})/\lambda_{ik}(t^* \mid \mathbf{Z}_{i \cdot a}, Z_{ikb} = z, s_{ik}) = \mathrm{HR}(t^*)$ may not be constant over time, and methods that allow relaxation of the standard proportional hazards assumption will be important in practice. For example, Hastie and Tibshirani (1993) studied a varying-coefficient model of the form:

$$\lambda_{ik}(t^* \mid \mathbf{Z}_{ik}) = \lambda_{0g}(t^*)exp\left[\boldsymbol{\beta}(t^*)^T \mathbf{Z}_{ik}\right] \tag{3.1}$$

A parametric spline basis can be adopted to characterize $\boldsymbol{\beta}(t^*)$, or non-parametric smoothing methods can be used. Here we modify local linear estimation described by Cai and Sun (2003) for use in the partly conditional setting. The idea of local linear estimation is that for a neighborhood around each time point t^*, $u \in \mathcal{N}(t^*, \epsilon)$, by Taylor series approximation, we have

$$\boldsymbol{\beta}(u) \approx \boldsymbol{\beta}(t^*) + \boldsymbol{\beta}'(t^*)(u - t^*) \ .$$

Based on a local "working independence" partial likelihood function, we can estimate $\beta(t^*)$ using a weighted estimating equation:

$$\sum_{i}^{n} \sum_{k}^{K} \int_0^\tau K_h(u - t^*)\left[\tilde{\mathbf{Z}}_{ik}(1, u - t^*) - \bar{\mathbf{Z}}(u)\right] dN_{ik}(u) = 0 \tag{3.2}$$

where $K(\cdot)$ is a kernel function with bounded support on $[-1, 1]$, h is the bandwidth, $K_h(x) = K(x/h)/h$, and $\tilde{\mathbf{Z}}_{ik}(1, u - t^*) = \mathbf{Z}_{ik} \otimes (1, \ u - t^*)$ with $\otimes$ denotes the Kronecker product. Under the stratified baseline model, $\bar{\mathbf{Z}}(u) = \sum_{g=1}^{G} \mathcal{S}_g^{(1)}[\boldsymbol{\beta}(t^*), u]1(s_{ik} \in \mathcal{I}_g)/\sum_{g=1}^{G} \mathcal{S}_g^{(0)}[\boldsymbol{\beta}(t^*), u]1(s_{ik} \in \mathcal{I}_g)$, with

$$\mathcal{S}_g^{(j)}[\beta(t^*), u] = \frac{1}{n} \sum_{l=1}^{n} \sum_{m=1}^{K} R_{lm}(u)exp[\boldsymbol{b}(t^*)^T \tilde{Z}_{lm}(1, u - t^*)]\tilde{Z}_{lm}(1, u - t^*)^{\otimes j} \cdot 1(s_{lm} \in \mathcal{I}_g)$$

10

where $\boldsymbol{b}(t^*) = [\boldsymbol{b}_0(t^*), \boldsymbol{b}_1(t^*)] = [\boldsymbol{\beta}(t^*), \boldsymbol{\beta}'(t^*)]$. The coefficient function $\boldsymbol{\beta}(t^*)$ is then estimated for each t^* using $\widehat{\boldsymbol{\beta}}(t^*) = \widehat{\boldsymbol{b}}_0(t^*)$.

Cai and Sun (2003) show the pointwise consistency and asymptotic normality of $\widehat{\boldsymbol{\beta}}(t^*)$ in the univariate case. We modify their results for the partly conditional setting. The consistency of $\widehat{\boldsymbol{\beta}}(t^*)$ in the multivariate setting can be established in the same way as in the marginal parametric hazard model discussed in the previous section and thus the proof is omitted here. By imposing a stronger condition, we establish the uniform consistency of $\boldsymbol{\beta}(t^*)$ in Theorem 3.1. The result is useful for the derivation of the large sample distribution for the survival functions and for estimators in the partly conditional survival models presented in the next section. We also derive the large sample distribution of $\boldsymbol{\beta}(t^*)$ in Theorem 3.2 for derived data.

Theorem 3.1 (Uniform Consistency of $\boldsymbol{\beta}(t^*)$) *Assume regularity conditions given in the Appendix, let $0 < t_1^* < t_2^* < \tau$ be fixed numbers, and assume that as $n \to \infty$, $h \to 0$ and $nh^2 \to \infty$, then as $n \to \infty$,*

$$\sup_{t^* \in [t_1^*, t_2^*]} |\widehat{\boldsymbol{\beta}}(t^*) - \boldsymbol{\beta}(t^*)| \to_p 0$$

Theorem 3.2 (Asymptotic normality) *Under the above conditions,*

$$\sqrt{nh} \left[\widehat{\boldsymbol{\beta}}(t^*) - \boldsymbol{\beta}(t^*) - \Delta(t^*, h, n) \right] \to N(0, \Sigma^{-1}(t^*) \Pi \Sigma^{-1}(t^*))$$

where $\Delta(t^*, h, n) = \frac{h^2 \mu_2}{2} \boldsymbol{\beta}''(t^*) + o_p(h^2)$ denotes the finite sample bias, and $\mu_2 = \int u^2 K_1(u) du$. We give the definition of $\Sigma(t^*)$ and Π in the proof found in the Appendix.

Following Cai and Sun (2003), the theoretical optimal bandwidth can be obtained by minimizing the asymptotic weighted mean integrated squared error. In the multivariate situation, this quantity depends on the robust variance estimator and the second derivative of the coefficient function at point t^*, which are unknown in advance. Data-dependent procedures for selecting the optimal bandwidth have been suggested in the literature for nonparametric function estimation (Hall and Carroll 1989; Eubank and Speckman 1993; Ducharme, Gannoun, Guertin and Jéquier 1995). Further research on adapting data-driven automatic procedures to the multivariate survival setting is warranted, and in practice a sensitivity analysis is suggested.

3.3 Partly Parametric Hazard Model

Finally we consider a class of models that accommodates the time-varying effect of a longitudinal marker in addition to other covariates whose effects are assumed independent of time. For presentation in this section, we only consider identical baseline hazard models. These models are essentially partly parametric hazard models as described by McKeague and Sasieni (1994) and have the following form for the hazard function:

$$\lambda_{ik}(t^* \mid \mathbf{Z}_{ik}, 0 \le s_{ik} \le T_i) = \lambda_0(t^*)exp\left\{\boldsymbol{\alpha}_1^T \mathbf{Z}_{i\cdot a} + \boldsymbol{\alpha}_2^T [f_j(s_{ik})]_{j=1}^p + \beta(t^*)Z_{ikb}\right\}, t^* > 0 \qquad (3.3)$$

To estimate the parameters under this model, we use two sets of estimating equations. Let $\boldsymbol{\theta} = [\boldsymbol{\alpha}, \beta(t^*)]$, where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2]$ is a vector of time-invariant coefficients for $\mathbf{Z}_{ika} = [\mathbf{Z}_{i\cdot a}, f_1(s_{ik}), \ldots, f_p(s_{ik})]$, and $\beta(t^*)$ is the time-varying coefficient for Z_{ikb}. In general one can obtain $\widehat{\boldsymbol{\theta}}$ by simultaneously solving the pair of estimating equations for $\boldsymbol{\alpha}$ and $\beta(t^*)$. A standard approach to solving these equations would involve backfitting (Hastie and Tibshirani 1995) and therefore require iterative solution to the equations. A computationally simple "one-step" alternative can yield an asymptotically equivalent estimator. The alternative estimator is defined by the following steps:

(i) Fit a nonparametric Cox model with all the covariates $\mathbf{Z}_{ik}$. Here the model is of a purely time-varying form: $\lambda_{ik}(t^* \mid \mathbf{Z}_{ik}, 0 \le s_{ik} \le T_i) = \lambda_0(t^*)exp\left[\boldsymbol{\alpha}^T(t^*)\mathbf{Z}_{ika} + \beta(t^*)Z_{ikb}\right]$. The resulting estimator $\tilde{\beta}(t^*)$ is consistent for $\beta(t^*)$ (see Zheng (2002) for details).

(ii) Fit a parametric Cox model with $\mathbf{Z}_{ika}$ as covariate, using a time-dependent offset $\tilde{\beta}(t^*)Z_{ikb}$. The estimator $\widehat{\boldsymbol{\alpha}}$ obtained from this model is consistent for $\boldsymbol{\alpha}$ (see Zheng (2002) for details).

(iii) Fit a time-varying coefficient Cox model with Z_{ikb}, using offset $\widehat{\boldsymbol{\alpha}}^T \mathbf{Z}_{ika}$. The new estimate $\widehat{\beta}(t^*)$ is the final estimate for $\beta(t^*)$.

3.4 Predictive Survival Functions with a Partly Conditional Survival Model

Here we present an estimation procedure for predicting the survival function for patients with a marker measurement z_b obtained at a certain specific measurement time s, in addition to a vector of time-invariant covariates $\mathbf{z}_a$. Let $P(T_i > t^* + s \mid \mathbf{Z}_i = \mathbf{z}_0, s) = S(t^* \mid \mathbf{Z}_i = \mathbf{z}_0, s) = e^{-\Lambda(\mathbf{z}_0, s)}$. $\Lambda_0(t^*)$

12

can be estimated by the natural Breslow-type estimator: $\widehat{\Lambda}_0(t^*) = \int_0^{t^*} \frac{1}{\widehat{J}_{n,0}^*(u)} d\bar{N}(u)$, where

$$\widehat{J}_{n,0}^*(t^*) = \sum_{i=1}^n \sum_{k=1}^K R_{ik}(t^*) exp\left\{\boldsymbol{\alpha}_1^T \mathbf{Z}_{i\cdot a} + \boldsymbol{\alpha}_2^T [f_j(s_{ik})]_{j=1}^p + \beta(t^*) Z_{ikb}\right\}$$

then

$$\widehat{S}(t^* \mid \mathbf{Z}_i = \mathbf{z}_0, s) = exp\left(-\int_0^{t^*} exp\left\{\boldsymbol{\alpha}_1^T \mathbf{z}_a + \boldsymbol{\alpha}_2^T [f_j(s)]_{j=1}^p + \beta(t^*) z_b\right\} d\widehat{\Lambda}_0(u)\right) \quad.$$

4. Simulations

We conduct simulations to investigate several aspects of the proposed partly conditional survival estimation procedure. First, we consider the small and large sample bias (and coverage) of the partly conditional regression estimates. Second, estimation using the derived survival data requires careful attention to mechanisms that lead to unbalanced cluster sizes (i.e., death and censoring). Theoretical results imply that the proposed risk-set based estimating equations provide valid inference. However, a second goal of our simulation study is to empirically demonstrate that the proposed methods do indeed lead to asymptotically unbiased estimation when the number of contributing observations per cluster is stochastic. Third, when the longitudinal marker values are exchangeable, a single marker measurement per subject can be selected and standard independent data methods can be used to estimate the regression parameters of interest. We compare the asymptotic relative efficiency of our multivariate procedure to the simple univariate analysis.

Our partly conditional model does not completely specify the joint distribution for the event time and the longitudinal marker process. Additional distributional assumptions are required to construct a joint distribution that can be used to simulate data. In the related multivariate survival literature it has been noted that it is surprisingly difficult to construct joint distributions that satisfy the marginal proportional hazards assumptions (Wei et al. 1989; Yang and Ying 2001). We show that it is possible to construct a valid joint distribution where derived survival times simultaneously satisfy the partly conditional hazards assumption.

4.1 Data Generation

Below we outline one algorithm for simulating marker and event time data that has the desired semi-parametric structure. Consider data with a single binary treatment group indicator, $Z_{i\cdot a} = 0$ or

<div align="center">13</div>

$Z_{i \cdot a} = 1$, and a single longitudinal marker $Z_i(s_k)$ measured at a common set of times s_1, s_2, ..., s_K. Data $\{ T_i,\ Z_{i \cdot a},\ \mathbf{Z}_{ib} = \text{vec}[Z_i(s_k)] \}$ are generated as follows:

(1) Generate a time independent binary covariate, $Z_{i \cdot a} \sim \text{bernoulli}(0.5)$.

(2) Generate $Z_{ib0} = b_i + \sum_{j=1}^m log(V_{ij})/\gamma_2$ where $b_i \sim \mathcal{N}(\mu, \sigma^2)$ and $V_{ij} \sim P(\rho)$, independent positive stable random variables with index ρ (Hougaard 1986).

(3) Generate a failure time T_i as: $H(T_i) = -(\gamma_1 Z_{i \cdot a} + \gamma_2 Z_{ib0}) + \epsilon_i$, where ϵ_i is an extreme value random variable, and $H(t)$ can be arbitrarily specified as long as it leads to an increasing cumulative hazard. For the Cox model $H(t) = log[\Lambda_0(t)]$.

(4) Let $Z_i(s_k) = Z_{ib0} - log(V_{ik})/\gamma_2$. This creates a form of exchangeable marker measurements. For a partly conditional model, we only include for analysis those $Z_i(s_k)$ with $s_k < T_i$. Based on properties of positive stable random random variables it can be shown that (1)-(4) leads to partly conditional hazards of the form:

$$\lambda[t^*|Z_{i \cdot a}, Z_i(s_k), T_i > s_k] = \lambda_0(t^* + s_k)\rho[\Lambda_0(t^* + s_k)]^{(\rho-1)}exp[\rho\gamma_1 Z_{i \cdot a} + \rho\gamma_2 Z_i(s_k)]. \qquad (4.1)$$

Using this construction the hazard for $T_{ik} = T_i - s_k$ will generally depend on s_k and therefore stratified models similar to those considered by Wei et al. (1989) would be appropriate. However, if we choose $\Lambda_0(t) = (t/a)^{1/\rho}$ then $\lambda_0(t + s_k)\rho[\Lambda_0(t + s_k)]^{(\rho-1)} = 1/a$, and thus a common baseline hazard obtains. By varying ρ and σ we can create marker measurements with differing amounts of within- and between-person variation. For similarity to the MACS CD4 data analyzed in section 5, we used $\mu = 600$ and $\sigma = 30$. For the induced partly conditional hazards model the regression coefficients are $\alpha = \rho \cdot \gamma_1$ and $\beta = \rho \cdot \gamma_2$.

4.2 Parametric Hazard Model

We evaluate samples with $n = 100, 500, 1000$ clusters using a uniform censoring distribution to obtain approximately $0\%, 25\%, 50\%$ and 75% censoring. We consider a study where the markers can be measured up to 10 times per subject. We perform 500 replications for each simulation scenario. The parameter for the positive stable random variables is chosen to be 0.6. Although we considered different values of the regression parameters, we only present results using $\alpha = -2$ and $\beta = -0.02$ (Note: we use this coefficient value (scale) since we create simulations to approximate analysis of CD4

14

where the marker ranges from less than 200 to greater than 1000).

With clusters of size 10, the average number of observations per subject ranges from 1.7 to 6.1, depending on the censoring level. Table 1 presents the simulations results for a partly conditional model using a single common baseline hazard. The relative bias is less than 4.3% for all situations, and tends to decrease with increasing sample size. Coverage probabilities using the robust variance estimator are very close to the nominal 95% level, whereas coverage probabilities from a naive independence model are generally below the nominal level.

4.3 Non-parametric Varying-coefficient Hazard Model

We also use simulation studies to assess the performance of local linear estimation for a time-varying coefficient partly conditional Cox model. We simulated data according to the scenario described in section 4.1, and use a local linear Cox model with Epanechnikov kernel. We investigate the performance at two different time points, namely, $t^* = 25$ and $t^* = 75$. At each time point, we choose bandwidth h such that 30% of the data points are included in the local linear estimation. As a result, at $t^* = 25$, the average bandwidth h is about 15, whereas at $t^* = 75$, the average bandwidth h is about 30. The actually bandwidth may vary slightly from sample to sample. The results presented in Table 2 indicate that estimates can be obtained with small bias using the partly conditional approach. We find -1.17% bias at $t^* = 25$ and -3.06% bias at $t^* = 75$. In addition, using robust standard errors to create 95% pointwise confidence intervals yielded empirical converge rates of 93.2% and 95.2% at $t^* = 25, 75$ respectively.

Finally, we also compare the multivariate approach with a valid univariate procedure that randomly chooses a single marker measurement for analysis. In Table 2 we display the bias and the standard deviation for $\widehat{\beta}(t^*)$ at $t^* = 25$ and $t^* = 75$. The estimates based on the partly conditional model have a variance that is 1/2.44 times the variance of the univariate estimator for $t^* = 25$ and 1/3.43 times smaller for $t^* = 75$, demonstrating the potential gain in efficiency through use of all marker measurements.

15

5. Example

Here we apply the partly conditional survival model to data from the Multicenter AIDS Cohort Study (MACS) which was reported in detail by Kaslow, Ostrow, Detels, Phair, Polk and Rinaldo (1987). Of the 5622 homosexual/bisexual men enrolled, 3426 were seronegative at baseline and 479 of these became seropositive between 1984 and 1996. Because we focus on the relationship between T-cell levels and AIDS diagnosis, we adopt the 1987 CDC definition of AIDS, which relies on symptoms rather than CD4 lymphocyte counts to define AIDS. Under this definition, 211 seroconverters developed AIDS during the study period. The mean time from seroconversion to the onset of AIDS among these subjects with observed times is 72 months (sd= 28 months, median=71 months). The present analysis uses data from the 438 seroconverters with dates of seroconversion known to within ±4.5 months. These subjects have an average of 13 measurements per person ($N = 3807$ total observations).

The objective of the present analysis is to investigate the relationship between a biomarker such as CD4 count and the risk of AIDS. We define the measurement time (s) as the time from seroconversion to the time that the CD4 count is recorded, and survival time as the time from CD4 count measurement to the AIDS diagnosis time ($t^* = t - s$). We seek to quantify the predictive value of serially measured CD4 counts. In all analyses we use $CD4(s)$ equal to the raw CD4 count divided by 300 (approximate standard deviation).

5.1 Partly Conditional Regression Function Estimation

We start by investigating the simplest model that assumes a common baseline hazard but allows a time-varying coefficient (model 1): $\lambda_{ik}(t^*) = \lambda_0(t^*)exp[\beta(t^*)CD4_i(s_{ik})]$. Subsequent analysis will relax this model to allow dependence on the measurement time s. For estimation we use all CD4 measurements after seroconversion and before AIDS diagnosis as the time-varying predictor, $CD4(s)$. Estimates of the function $\beta(t^*)$ are obtained by fitting a partly conditional Cox model using local linear estimation. We use the Epanechnikov kernel $K(u) = 0.75(1 - U^2)_+$ with a bandwidth of 30. The bandwidth is selected to ensure that we have substantial data available at each data point for stable estimation. We also consider other values of the bandwidth to assess the sensitivity of the results to this choice. Eubank and Speckman (1993) suggest use of an undersmoothed bandwidth

h, i.e., one that satisfies $n^{1/3}h \rightarrow 0$, so that the inherent bias (given as $\Delta(t^*, h, n)$ in section 3.2) is negligible asymptotically. The function $\beta(t^*)$ is estimated at the grid points $t^* = 4 \times j$ months, $j = 1, ..., 32$. In Figure 1 the estimate of $\beta(t^*)$ shows a strong time trend, with diminishing relative risk as t^* increases. For example, at any time $0 < s < T$, for two individuals whose CD4 difer by 300, the log relative hazard is the highest immediately after the measurement is taken ($t^* = 1$), with $\widehat{\beta}(1) \approx -3$, and then attenuates steadily over the next 60 months to $\widehat{\beta}(60) \approx -1$. Finally, the predictive capacity of CD4 wanes to nearly 0 when the measurement is more than 60 months old.

To explore the potential gain in efficiency that arises from using all of the longitudinal data we compare results from estimation using only a single randomly selected CD4 measurement per individual. Figure 1 also shows the local linear estimates using these 438 independent observations. We find similar point estimates yet narrower confidence intervals when all 3807 longitudinal measurements are used to estimate $\beta(t^*)$. As suggested by our simulation studies, estimation based on all the available longitudinal data is apparently more efficient.

Next we investigate whether the effect of CD4 depends on the measurement time s by adopting a model where both the baseline hazard and the form of the coefficient function, $\beta(t^*)$ may depend on s (model 2). We create 3 groups of data $(X_{ik}, \Delta_{ik}, Z_i(s_{ik}), s_{ik})$ based on the CD4 measurement time, s_{ik}: group 1 is comprised of CD4 measurements within the first year after seroconversion ($0 \leq s \leq 12$, n=392 and N=686); group 2 is comprised of measurements between the second and the third year ($24 < s \leq 36$, $n = 321$ and $N = 577$); and group 3 contains data from between the fourth and fifth year post-seroconversion ($48 < s \leq 60$, $n = 229$ and $N = 396$). Figure 2 shows the coefficient functions estimated separately for these three groups. It appears that CD4 measured earlier after seroconversion loses its predictive power more rapidly than CD4 measured at later times. For example, for CD4 observed within the first year the log relative hazard at $t^* = 4$ months is $\widehat{\beta}(t^* = 4) \approx -3$, and then quickly decays to $\widehat{\beta}(t^*) = -1$ by 20 months. In contrast, for measurements taken in the fifth year after conversion we find somewhat weaker short-term association, $\widehat{\beta}(t^* = 4) \approx -2$, but a longer follow-up time is required before the log relative hazard decays to -1, with the point estimate crossing -1 after 30 months of follow-up. Essentially, this analysis approach estimates the coefficient function $\beta(t^*, s)$ in the model $\lambda_0(t^*, s)exp[\beta(t^*, s)CD4(s)]$. If a parametric form for $\beta(t^*, s)$ were adopted the

17

simplifying assumption $\beta(t^*, s) = \beta(t^*)$ used in model 1 could be formally tested.

We also consider two intermediate models that differ in the way they model the measurement time s while retaining a common estimate for $\beta(t^*)$. Model 3 assumes different, but unspecified, baseline functions for CD4 measured at different followup times. Since subjects in MACS are followed semi-annually, we divide observations into $G = 11$ strata. An observation belongs to the gth stratum if it is measured within the gth year since seroconversion. Specifically, the model takes the form $\lambda_{ik}(t^*) = \lambda_{0g}(t^*)exp[\beta(t^*)CD4_i(s_{ik})]$. Alternatively, model 4 uses both the CD4 count and the measurement time as covariates in a partly parametric Cox model, namely, $\lambda_{ik}(t^*) = \lambda_0(t^*)exp\left\{\boldsymbol{\alpha}^T[f_j(s)]_{j=1}^p + \beta(t^*)CD4_i(s_{ik})\right\}$. We specify a flexible parametric model for measurement time, s, using natural cubic spline basis functions, $f_j(s)$, with a single knot at $s = 48$. We also conduct analyses with a single knot at $s = 36$ and with a pair of knots at $s = 16, 36$ respectively. These alternative choices for $f_j(s)$ result in very minor differences in the estimated coefficient function $\widehat{\beta}(t^*)$. Since later follow-up times tend to have shorter survival time due to administrative censoring, in order to obtain stable estimates, we restrict the analysis to a time interval of $0 \le t - s \le 60$. Figure 2 shows the estimated coefficient functions for CD4 along with 95% pointwise confidence intervals from both the stratified and the partly parametric models. The two methods give very similar estimated coefficient functions.

5.2 Partly Conditional Survival Function Estimation

One important objective of our partly conditional survival model is to estimate the updated survival probability $P[T_i > t \mid CD4_i(s), T_i > s]$ for an arbitrary pair of survival and measurement times, (s, t), where $s < t$. Figure 3 shows the predictive survival probabilities given in terms of years since measurement, $P[T_i > t^* + s \mid CD4_i(s), T_i > s]$, based on models 2 and 4 described above. Recall that in our application the survival probability is equivalent to the probability of being free of AIDS. Figure 3 illustrates estimated probabilities for two hypothetical individuals: one with a high CD4 value (725); and one with a low CD4 value (340). We also consider measurement times of one year post seroconversion, $s = 12$, and the third, $s = 36$, and fifth year, $s = 60$, after seroconversion. In Figure 3 panel (a) we use model 2 which allows the baseline hazard and the coefficient function

to depend on the measurement time s. We see that for both individuals, the chance of being free of AIDS decreases steadily with time, but the individual with a higher CD4 value is less likely to develop AIDS during the follow-up period. Furthermore, the predictive survival functions appear to depend on the time at which CD4 count is measured. For example, an individual with a CD4 value of 340 measured at the 1st year after seroconversion has a chance of developing AIDS within 4 years of approximately 30%, whereas if the same value of CD4 is obtained at 3 years post seroconversion, then his chance of getting AIDS within the next 4 years is approximately 60%.

To assess the sensitivity of survival predictions to the choice of model we also display estimates based on a more structured model that assumes a common coefficient function, and uses the measurement time s as a covariate (model 4). In Figure 3 panel (b) we find estimated survival probabilities that are qualitatively similar to those from the less structured model, but specific estimates differ, particularly for longer follow-up times.

If the ultimate goal is to create accurate predictions then the trade-off between a potentially less biased but more variable approach (model 2) and a less variable but potentially biased approach (model 4) may be empirically evaluated using cross-validation methods if a meaningful measure of the discrepancy between data and prediction can be adopted. Unfortunately there is no well accepted summary of predictive model accuracy for survival models, although alternatives have been proposed (Schemper and Henderson 2000; Heagerty, Lumley and Pepe 2000). In addition, empirical evaluation of accuracy for varying-coefficient models would be computationally demanding.

6.   Discussion

We have proposed a new approach that can quantify the risk of a key clinical event at time t as a function of the marker process accumulated through time s, for any pair of times (s, t) with $s < t$. In contrast to the standard time-varying covariate regression model for the event time, our method decouples the time scale for modeling the hazard from the time scale for accrual of available longitudinal covariate information, and thus directly facilitates the calculation of quantities such as $P[T_i > t | Z_i(s), 0 \le s < t]$ without assumptions regarding the longitudinal marker distribution.

One important feature of our partly conditional model is that we allow regression parameters to depend on both the time of measurement for the predictor and the time of measurement for the

outcome. In specific applications a varying coefficient model of the form $\beta(t, s) = \beta(t - s)$ may be used which assumes that the association between the survival outcome and the covariate depends only on their time separation. For estimation, we extended local linear estimation for the univariate Cox model (Cai and Sun 2003) to the partly conditional setting, and we provided detailed estimation procedures for three classes of partly conditional survival models. One issue that arises with use of a partly conditional hazard model is the need to model the measurement time s. We have introduced alternative models that differ in the way the effect of the measurement time, s, is specified. In general, the choice of model may depend on the aim of the study and the specific features of the data structures such as the frequency and spacing of visits. Our example analysis explored the extent to which results were sensitive to model choice. Further research to develop methods for model checking and to define appropriate criteria for selecting models would be useful.

In this article we make fairly strong assumptions regarding both measurement timing and marker missingness. In particular, we assume that each individual provides a sequence of measurements at either a set of fixed times, or at times that occur in a completely random fashion. However, in some observational studies, individuals are not necessarily followed at scheduled intervals. Furthermore, the timing at which individuals are measured may depend on the previous value of the marker measurement. In the repeated measures setting with outcome-dependent follow-up it has been demonstrated that potential bias could be associated with the use of an estimating equation approach (Lipsitz, Fitzmaurice, Ibrahim, Richard and Steven 2002). Further work is needed to evaluate the robustness of our estimation procedure to the timing assumptions, or to develop a more general method that can relax the measurement timing assumptions.

Instead of employing a likelihood-based estimation procedure, we develop non-parametric and semi-parametric methods. One advantage of using a semi-parametric approach is that it provides a computationally simple and robust solution. One potential weakness is that our methods are based a working independence assumption, and as such, may be less efficient than a full-likelihood approach. Future work that compares our semi-parametric approach with specific likelihood-based alternatives in terms of both efficiency and robustness would be valuable.

20

REFERENCES

Altman, D. G., and De Stavola, B. L. (1994), "Practical problems in fitting a proportional hazards model to data with updated measurements of the covariates," *Statistics in Medicine*, 13, 301–341.

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1995), *Statistical models based on counting processes* Springer-Verlag Inc.

Cai, Z., and Sun, Y. (2003), "Local linear estimation for time-dependent coefficients in Cox's regression models," *Scandinavian Journal of Statistics*, to appear, 1–10.

Cox, D. R. (1972), "Regression models and life-tables (with discussion)," *Journal of the Royal Statistical Society, Series B, Methodological*, 34, 187–220.

Cox, D. R., and Oakes, D. O. (1984), *Analysis of survival data* Chapman & Hall Ltd.

Davis, P. B. (1997), "The decline and fall of pulmonary function in cystic fibrosis: New models, new lessons," *The Journal of Pediatrics*, 131, 789–790.

Ducharme, G. R., Gannoun, A., Guertin, M.-C., and Jéquier, J.-C. (1995), "Reference values obtained by kernel-based estimation of quantile regressions," *Biometrics*, 51, 1105–1116.

Etzioni, R., Pepe, M., Longton, G., Hu, C., and Goodman, G. (1999), "Incorporating the time dimension in receiver operating characteristic curves: A case study of prostate cancer," *Medical Decision Making*, 19, 242–251.

Eubank, R. L., and Speckman, P. L. (1993), "Confidence bands in nonparametric regression," *Journal of the American Statistical Association*, 88, 1287–1301.

Faucett, C. L., and Thomas, D. C. (1996), "Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach," *Statistics in Medicine*, 15, 1663–1685.

Hall, P., and Carroll, R. J. (1989), "Variance function estimation in regression: The effect of estimating the mean," *Journal of the Royal Statistical Society, Series B, Methodological*, 51, 3–14.

Hastie, T., and Tibshirani, R. (1993), "Varying-coefficient models," *Journal of the Royal Statistical Society, Series B*, 55, 757–796.

Hastie, T., and Tibshirani, R. (1995), "Generalized additive models for medical research," *Statistical Methods in Medical Research*, 4, 187–196.

Heagerty, P. J., Lumley, T., and Pepe, M. S. (2000), "Time-dependent ROC curves for censored survival data and a diagnostic marker," *Biometrics*, 56(2), 337–344.

Henderson, R., Diggle, P., and Dobson, A. (1997), "A joint model for survival and longitudinal data measured with error," *Biometrics*, 53, 330–339.

Hougaard, P. (1986), "Survival models for heterogeneous populations derived from stable distributions (Corr: V75 p395)," *Biometrika*, 73, 387–396.

Kalbfleisch, J. D., and Prentice, R. L. (1980), *The statistical analysis of failure time data* John Wiley & Sons.

Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F., and Rinaldo, C. R. A. (1987), "The Multicentre AIDS Cohort Study: rationale, organization, and selected characteristics of the participants," *American Journal of Epidemiology*, 1(126), 310–318.

Lee, E. W., Wei, L. J., and Amato, D. A. (1992), Cox-type regression analysis for large numbers of small groups of correlated failure time observations (Disc: p247),, in *Survival analysis: State of the art*, pp. 237–247.

Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J. G., Richard, G., and Steven, L. (2002), "Parameter Estimation in Longitudinal Studies with Outcome-Dependent Follow-Up," *Biometrics*, 58(1), 621–630.

McKeague, I. W., and Sasieni, P. D. (1994), "A partly parametric additive risk model," *Biometrika*, 81, 501–514.

Newey, W. K., and McFadden, D. (1994), Large Sample Estimation and Hypothesis Testing,, in *Handbook of Econometrics, Volume IV*, pp. 2113–2241.

Pawitan, Y., and Self, S. (1993), "Modeling disease marker processes in AIDS," *Journal of the American Statistical Association*, 88, 719–726.

Pepe, M. S., and Couper, D. (1997), "Modeling partly conditional means with longitudinal data," *Journal of the American Statistical Association*, 92, 991–998.

Pepe, M. S., Heagerty, P., and Whitaker, R. (1999), "Prediction using partly conditional time-varying coefficients regression models," *Biometrics*, 55, 944–950.

Prentice, R. L. (1982), "Covariate measurement errors and parameter estimation in a failure time regression model (Corr: V71 p219)," *Biometrika*, 69, 331–342.

Schemper, M., and Henderson, R. (2000), "Predictive accuracy and explained variation in Cox regression," *Biometrics*, 56(1), 249–255.

Skates, S. J., Pauler, D. K., and Jacobs, I. J. (2001), "Screening based on the risk of cancer calculation from Bayesian hierarchical changepoint and mixture models of longitudinal markers," *Journal of the American Statistical Association*, 96(454), 429–439.

Slate, E. H., and Turnbull, B. W. (2000), "Statistical models for longitudinal biomarkers of disease onset," *Statistics in Medicine*, 19(4), 617–637.

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.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989), "Regression analysis of multivariate incomplete failure time data by modeling marginal distributions," *Journal of the American Statistical Association*, 84, 1065–1073.

Wulfsohn, M. S., and Tsiatis, A. A. (1997), "A joint model for survival and longitudinal data measured with error," *Biometrics*, 53, 330–339.

Yang, Y., and Ying, Z. (2001), "Marginal proportional hazards models for multiple event-time data," *Biometrika*, 88(2), 581–586.

Zheng, Y. (2002), *Semiparametric Methods for Longitudinal Diagnostic Accuracy* PhD Thesis. Univerisity of Washington, Seattle.

Table 1: Simulation Results for Identical Baseline Models, K = 10. The partly conditional hazard model is $\lambda_0(t^*)exp[\alpha Z_{i \cdot a} + \beta Z_i(s_{ik})]$.

n	% censored	AN	RBias $\times 10^2$	SE $\times 10^{-3}$	CP(N)	CP(R)	RBias $\times 10^2$	SE $\times 10^{-5}$	CP(N)	CP(R)
				$\alpha = -2$				$\beta = -0.02$		
100	0	4.2	1.63	11.31	0.656	0.906	0.13	3.55	0.838	0.942
100	25	3.7	2.18	13.95	0.690	0.924	0.87	4.22	0.888	0.958
100	50	1.5	1.47	14.59	0.940	0.932	2.07	5.95	0.952	0.940
100	75	1.0	2.08	22.39	0.958	0.914	4.21	9.45	0.954	0.924
200	0	4.2	0.63	7.17	0.678	0.930	0.05	2.67	0.794	0.936
200	25	3.7	0.26	9.16	0.724	0.944	0.54	2.97	0.858	0.940
200	50	1.5	0.26	10.58	0.918	0.926	0.68	3.99	0.936	0.930
200	75	1.0	0.61	15.13	0.940	0.934	2.01	5.78	0.956	0.930
500	0	4.2	0.33	4.54	0.652	0.928	0.09	1.53	0.818	0.966
500	25	3.7	0.20	6.27	0.698	0.922	0.07	1.92	0.836	0.932
500	50	1.5	0.47	6.19	0.930	0.948	0.33	2.48	0.934	0.932
500	75	1.0	0.18	9.30	0.940	0.934	0.84	3.48	0.946	0.924
1000	0	4.2	0.20	2.96	0.700	0.956	0.08	1.19	0.782	0.930
1000	25	3.7	0.04	3.83	0.712	0.946	0.01	1.42	0.838	0.944
1000	50	1.5	0.66	4.30	0.942	0.960	0.08	1.63	0.952	0.956
1000	75	1	0.36	6.27	0.960	0.962	0.23	2.33	0.948	0.942

Note: AN is the average number of measurements per subject that is used in the partly conditional model. RBias (Relative Bias) is the sampling mean of the ratio $|\hat{\beta} - \beta_0|/\beta_0$. SE is the sampling mean of the robust standard error estimator for $\hat{\beta}$. CP(N) and CP(R) are the coverage probabilities of the 95% confidence intervals corresponding to the naive and robust variance estimates.

Table 2: Simulation Results for $\beta(t^*) = -0.02$ with Identical Baseline Models. n=200, K=10. The partly conditional hazard model is $\lambda_0(t^*)exp[\alpha Z_{i \cdot a} + \beta(t^*) \cdot Z_i(s_{ik})]$.

t	RBias $\times 10^2$	SE_{emp} $\times 10^{-3}$	SE_{est} $\times 10^{-3}$	CP	RBias $\times 10^2$	SE_{emp} $\times 10^{-3}$	SE_{est} $\times 10^{-3}$	CP(R)	EF
	Univariate Model				Partly Conditional Model				
25	-3.525	2.247	2.547	0.938	-1.170	1.630	1.427	0.932	2.442
75	8.035	4.681	3.801	0.938	-3.055	2.203	2.052	0.952	3.432

Note: RBias (Relative Bias) is the sampling mean of the ratio $|\hat{\beta}(t^*) - \beta_0(t^*)|/\beta_0(t^*)$. SE_{est} is the mean of the standard error estimates; SE_{emp} is the standard error of the estimates of β; EF is the relative efficiency of multivariate model vs. univariate model $(\hat{\sigma}^2_{univ}/\hat{\sigma}^2_{mult})$.
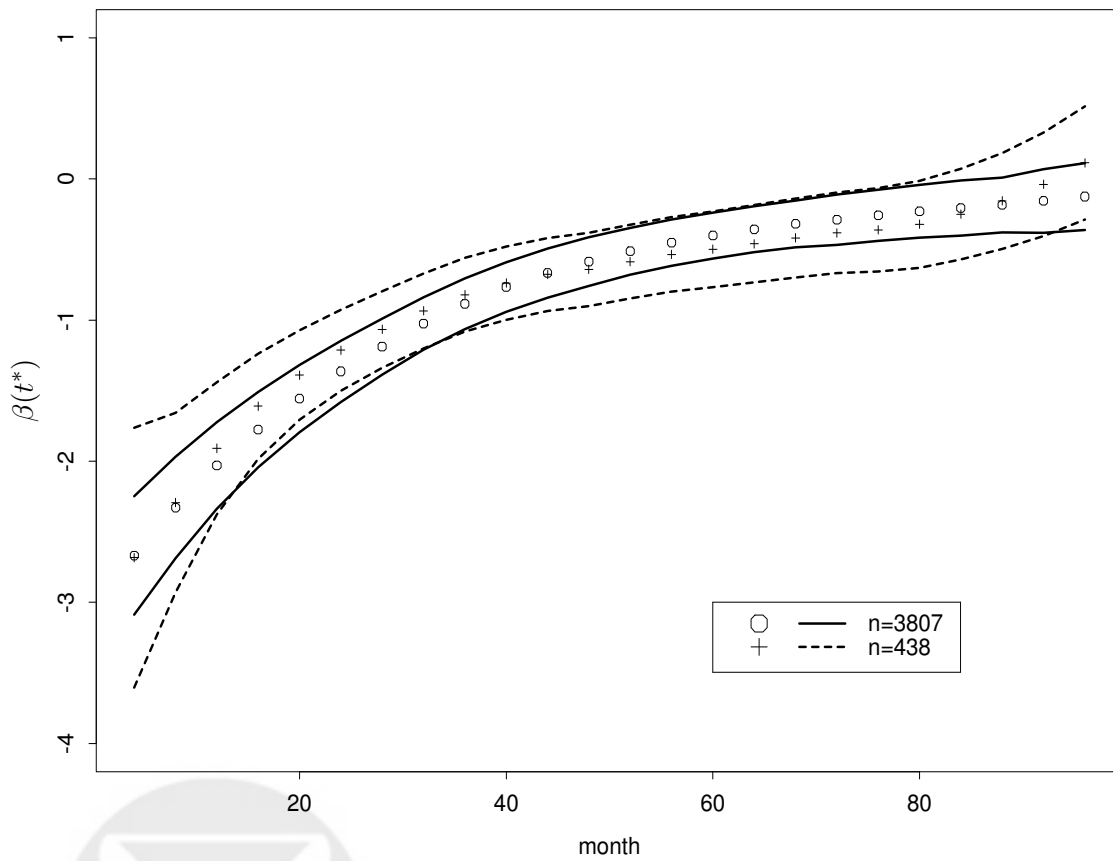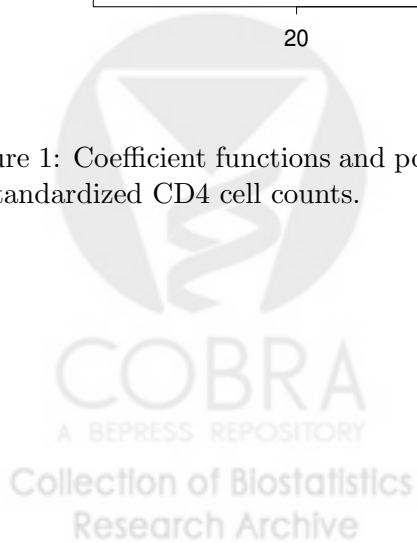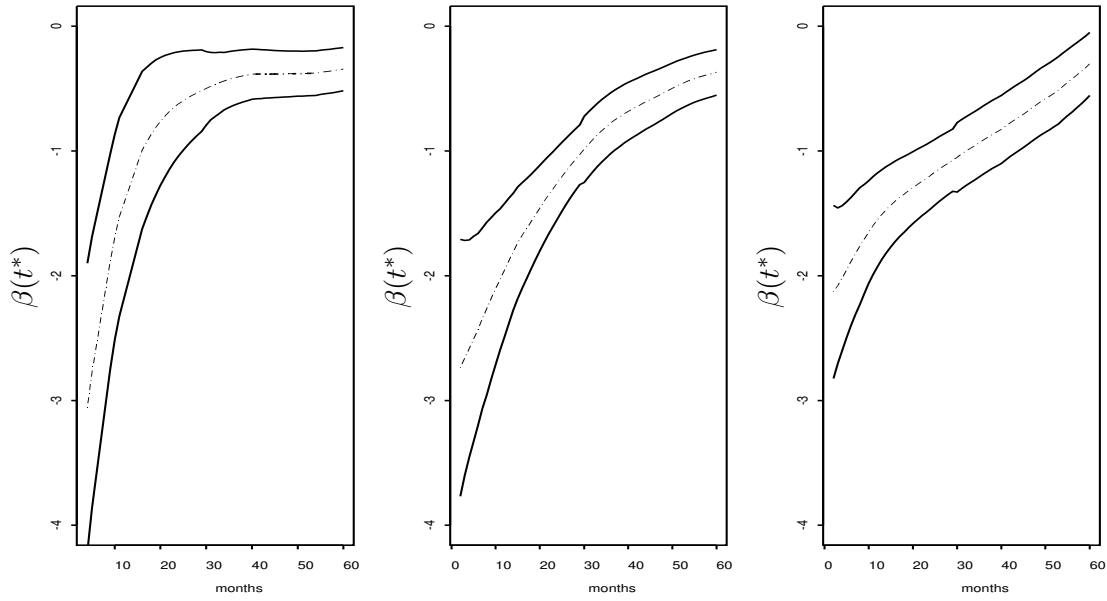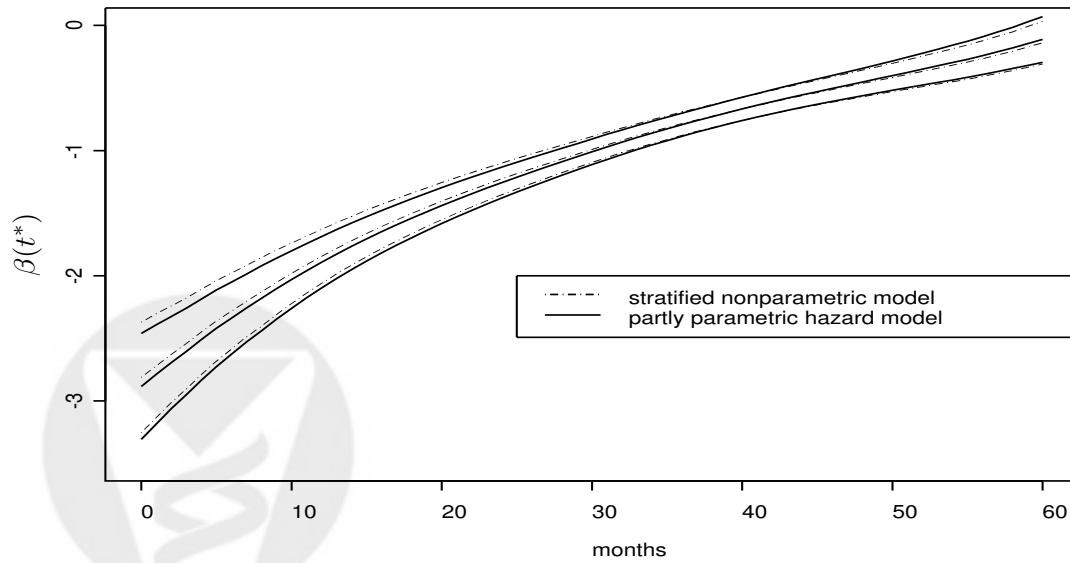
24

Figure 1: Coefficient functions and pointwise 95% confidence intervals for the time-varying coefficient of standardized CD4 cell counts.

25

(a) Coefficient functions and pointwise 95% confidence intervals for standardized CD4 cell counts measured within the first ($s \leq 1$), between the second and the third ($1 < s \leq 3$), and between the fourth and fifth year ($3 < s \leq 5$) following seroconversion based on separate non-parametric partly conditional Cox models.

(b) Coefficient functions and pointwise 95% confidence intervals for standardized CD4 cell counts based on a non-parametric and a partly parametric Cox model.

Figure 2: Coefficient functions for standardized CD4 cell counts using different partly conditional survival models.
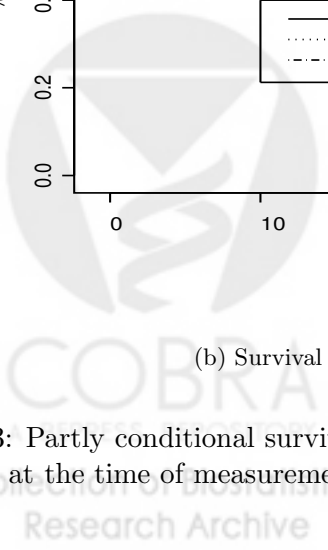
26
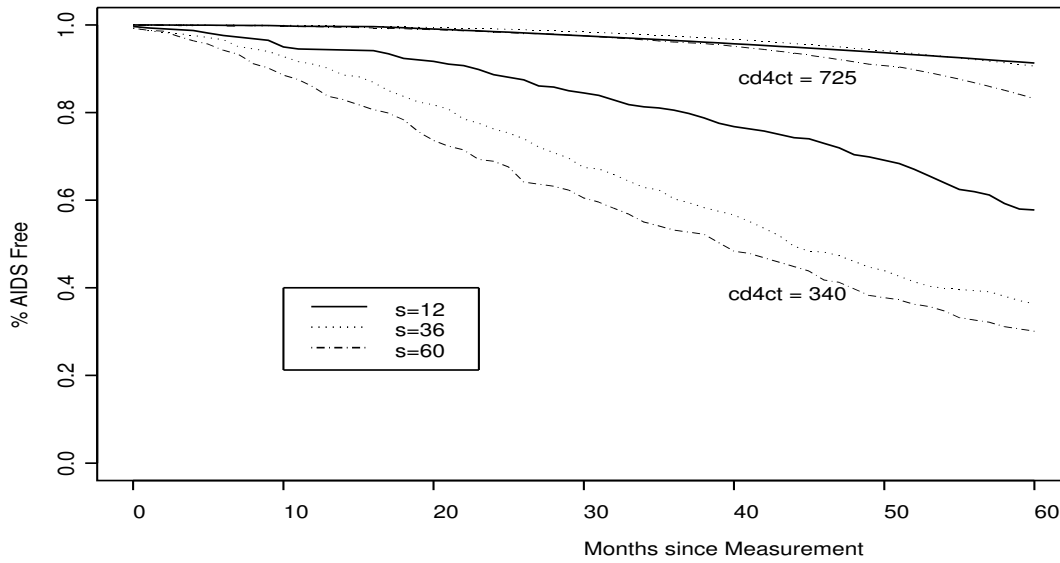
(a) Survival estimates based on separate non-parametric Cox models fit to $s \leq 1$, $1 < s \leq 3$, and $3 < s \leq 5$.
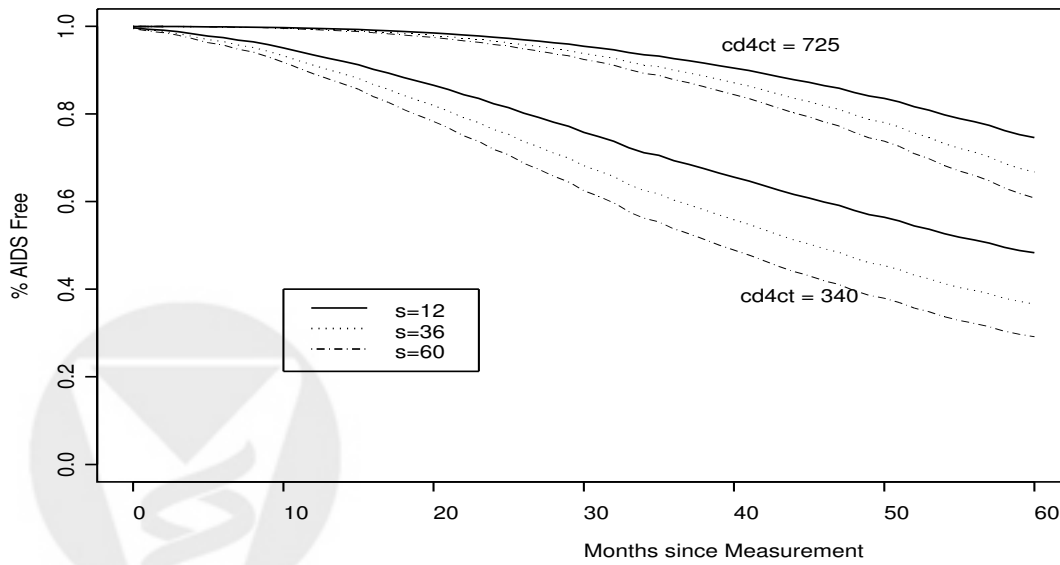
(b) Survival estimates based on a partly parametric Cox model.

Figure 3: Partly conditional survival functions for hypothetical subjects with CD4 cell counts of 725 and 340 at the time of measurement.

Appendix A. Large Sample Properties for Proposed Estimators

A.1 Preliminary

Following the notation of Cai and Sun (2003), we denote $P(t^*|\mathbf{Z}) = E[I(X \geq t^*)|\mathbf{Z} = \mathbf{z}]$, $Q_j = E\{P[t^*|\mathbf{Z}]\lambda(t^*|\mathbf{Z})\mathbf{Z}^{\otimes j}\}$, and $\Sigma(t) = Q_2(t) - Q_1(t)^T Q_1(t)/Q_0(t)$. In addition, let $\mu_j = \int u^j K(u) du$ and $\nu_j = \int u^j K^2(u) du$ for $0 \leq j \leq 2$. We require the same conditions as in Condition A of Cai and Sun (2003).

We also introduce some notation that is used within this section. For any fixed $\boldsymbol{b}_0(t^*)$, let $\hat{\boldsymbol{b}}(t^*)$ be the estimator that solves equation 3.2. Suppose $\hat{\boldsymbol{b}}(t^*)$ is of dimension $q \times 2$. Let $\mathbf{H} = diag\{\mathbf{I_q}, h\mathbf{I_q}\}$, $\boldsymbol{\gamma}(t^*) = \mathbf{H}[\boldsymbol{b}(t^*) - \boldsymbol{b}_0(t^*)]$, $\hat{\boldsymbol{\gamma}}(t^*) = \mathbf{H}[\hat{\boldsymbol{b}}(t^*) - \boldsymbol{b}_0(t^*)]$ and $\tilde{\mathbf{U}}_{ik}(1, \ u - t^*) = \mathbf{H}^{-1}\tilde{\mathbf{Z}}_{ik}(1, u - t^*)$. For $0 \leq j \leq 2$, $1 \leq g \leq G$ and $n_g = \sum_{i=1}^{n} \sum_{k=1}^{K} 1(s_{ik} \in \mathcal{I}_g)$,

$$\mathcal{S}_g^{(j)}(u) = n_g^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} R_{ik}(u) exp\left[\boldsymbol{b}(t^*)^T \tilde{\mathbf{Z}}_{ik}(1, u - t^*)\right] \tilde{\mathbf{U}}_{ik}(1, u - t^*)^{\otimes j} 1(s_{ik} \in \mathcal{I}_g)$$

$$\mathcal{S}_g^{(j)}[\boldsymbol{\gamma}(t^*), u] = n_g^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} R_{ik}(u) exp\left[\boldsymbol{b}(t^*)^T \tilde{\mathbf{Z}}_{ik}(1, u - t^*) + \boldsymbol{\gamma}(t^*)^T \tilde{\mathbf{U}}_{ik}(1, u - t^*)\right] \tilde{\mathbf{U}}_{ik}(1, u - t^*)^{\otimes j} 1(s_{ik} \in \mathcal{I}_g)$$

$$\mathcal{S}_g^{(j)*}(u) = n_g^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} R_{ik}(u) exp\left[\boldsymbol{\beta}_0(u)^T \mathbf{Z}_{ik}\right] \tilde{\mathbf{U}}_{ik}(1, u - t^*)^{\otimes j} 1(s_{ik} \in \mathcal{I}_g)$$

$$\tilde{\mathcal{S}}_g^1(u) = n_g^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} R_{ik}(u) exp\left[\boldsymbol{b}(t^*)^T \tilde{\mathbf{Z}}_{ik}(1, u - t^*)\right] \mathbf{Z}_{ik} 1(s_{ik} \in \mathcal{I}_g)$$

Furthermore, Let $\boldsymbol{s}_g^{(j)}(u)$, $\boldsymbol{s}_g^{(j)}[\boldsymbol{\gamma}(t^*), u]$, $\boldsymbol{s}_g^{(j)*}(u)$, and $\tilde{\boldsymbol{s}}_g^{(1)}(u)$ be the expected values of $\mathcal{S}_g^{(j)}(u)$, $\mathcal{S}_g^{(j)}[\boldsymbol{\gamma}(t^*), u]$, $\mathcal{S}_g^{(j)*}(u)$, and $\tilde{\mathcal{S}}_g^{(1)}(u)$. For example, for $\mathcal{S}_g(u) = n_g^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} R_{ik}(u) h(u, \mathbf{Z}_{ik})$, $\boldsymbol{s}_g(u) = E[P(u|\mathbf{Z}) h(u, \mathbf{Z})]$.

A.2 Proof of theorem 3.1

Let the filtration $\mathcal{F}_{t^*}$ be the statistical information accruing up to time $[0, t^*]$, i.e.,

$$\mathcal{F}_{t^*} = \sigma\{\mathbf{Z}_{ik}, N_{ik}(u), R_{ik}(u), i = 1, ..., n, k = 1, ..., K_i, 0 \leq u \leq t^*\}$$

Then by the independent censoring assumption,

$$M_{ik}(t^*) = N_{ik}(t^*) - \int_0^{t^*} R_{ik}(u)\lambda_{ik}[u|\mathbf{Z}_{ik}]du \tag{A.1}$$

is an $\mathcal{F}_{t^*}$-martingale, with $\lambda_{ik}[u|\mathbf{Z}_{ik}]du$ as defined in equation 3.1.

28

Now, define $\bar{N}(u) = n_g^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} N_{ik}(u) 1(s_{ik} \in \mathcal{I}_g)$, for any $t^* \in [t_1^*, t_2^*]$, let $\ell[\boldsymbol{b}(t^*)]$ denote a local "working independence" partial likelihood function, then the process

$$\ell[\boldsymbol{\gamma}(t^*), t^*] - \ell(0, t^*) = \int_0^{\tau} K_h(u - t^*) n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \boldsymbol{\gamma}(t^*)^T \tilde{\mathbf{Z}}_{ik}(1, u - t^*) dN_{ik}(u)$$

$$- \int_0^{\tau} K_h(u - t^*) log \left\{ \frac{\mathcal{S}_g^{(0)}[\boldsymbol{\gamma}(t^*), u]}{\mathcal{S}_g^{(0)}(0, u)} \right\} d\bar{N}(u)$$

has compensator

$$A_n[\boldsymbol{\gamma}(t^*), t^*] = \int_0^{\tau} K_h(u - t^*) \left\{ \boldsymbol{\gamma}(t^*)^T \mathcal{S}_g^{(1)*}(u) - log \left\{ \frac{\mathcal{S}_g^{(0)}[\boldsymbol{\gamma}(t^*), u]}{\mathcal{S}_g^{(0)}(0, u)} \right\} \mathcal{S}_g^{(1)*}(u) \right\} \lambda_0(u) du,$$

with their difference

$$X_n[\boldsymbol{\gamma}(t^*), t^*] = \int_0^{\tau} K_h(u - t^*) \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \left\{ \boldsymbol{\gamma}(t^*)^T \tilde{\mathbf{U}}_{ik}(1, u - t^*) - log \left\{ \frac{\mathcal{S}_g^{(0)}[\boldsymbol{\gamma}(t^*), u]}{\mathcal{S}_g^{(0)}(0, u)} \right\} \right\} dM_{ik}(u),$$

being a local square integrable martingale. By modifying the arguments in Cai and Sun (2003) for the multivariate case, we can show that $A_n[\boldsymbol{\gamma}(t^*), t^*]$ converges to a strictly concave function $A[\boldsymbol{\gamma}(t^*), t^*]$ with a maximum at point $\boldsymbol{\gamma}(t^*) = 0$. In addition, the concave function $\ell[\boldsymbol{\gamma}(t^*), t^*] - \ell(0, t^*)$ converges to $A[\boldsymbol{\gamma}(t^*), t^*]$, and thus $\hat{\boldsymbol{\beta}}(t^*) \to_p \boldsymbol{\beta}_0(t^*)$ pointwise for any $0 < t^* < \tau$.

To show that the convergence is uniform, we first show that $A_n[\boldsymbol{\gamma}(t^*), t^*]$ is stochastically equicontinuous. Let $\boldsymbol{\gamma}^*(t^*)$ denotes $\mathbf{H}[\boldsymbol{b}^*(t^*) - \boldsymbol{b}_0(t^*)]$, where $\boldsymbol{b}^*(t^*)$ lies between $\boldsymbol{b}_0(t^*)$ and $\boldsymbol{b}^*(t^*)$. We note that

$$A_n[\boldsymbol{\gamma}(t^*), t^*] = \boldsymbol{\gamma}(t^*) \int_0^{\tau} K_h(u - t^*) \left\{ \mathcal{S}_g^{(1)*}(u) - \left\{ \frac{\mathcal{S}_g^{(0)}[\boldsymbol{\gamma}^*(t^*), u]}{\mathcal{S}_g^{(0)}[\boldsymbol{\gamma}^*(t^*), u]} \right\} \mathcal{S}_g^{(1)*}(u) \right\} \lambda_0(u) du$$

$$\leq \sup_{t^* \in [t_1^*, t_2^*]} |\boldsymbol{\gamma}(t^*)| V(K) \left\{ \mathcal{S}_g^{(1)*}(t^*) - \left\{ \frac{\mathcal{S}_g^{(0)}[\boldsymbol{\gamma}^*(t^*), t^*]}{\mathcal{S}_g^{(0)}[\boldsymbol{\gamma}^*(t^*), t^*]} \right\} \mathcal{S}_g^{(1)*}(t^*) \right\} \int \lambda_0(s) ds$$

Since V(K) is bounded, $\int \lambda_0(s) ds$ is finite, $\mathcal{S}_g^{(1)*}(t^*)$, $\mathcal{S}_g^{(0)}[\boldsymbol{\gamma}^*(t^*), t^*]$ and $\mathcal{S}_g^{(0)}(\boldsymbol{\gamma}^*(t^*), t^*)$ are bounded on $\mathcal{B} * [0, \tau]$, we have $A_n[\boldsymbol{\gamma}(t^*), t^*] \leq \sup_{t^* \in [t_1^*, t_2^*]} |\boldsymbol{\gamma}(t^*)| \cdot C$, with C bounded in probability. It then follows that $\sup_{t^* \in [t_1^*, t_2^*]} |A_n[\hat{\boldsymbol{b}}(t^*), t^*] - A[\boldsymbol{b}_0(t^*), t^*]| \to_p 0$ by Lemma 2.9 of Newey and McFadden (1994).

Finally to show the uniform consistency of $\boldsymbol{\beta}(t^*)$, we only need to focus on the first p components

of $X_n[\boldsymbol{\gamma}(t^*), t^*]$, which is,

$$
X_n^*[\boldsymbol{\beta}(t^*), t^*] \;=\; \int_0^\tau K_h(u - t^*) \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K \left\{ [\boldsymbol{\beta}(t^*) - \boldsymbol{\beta}_0(t^*)]^T \mathbf{Z}_{ik} - log \frac{\mathcal{S}_g^{(0)}[\boldsymbol{\gamma}(t^*), u]}{\mathcal{S}_g^{(0)}(0, u)} \right\} dM_{ik}(u)
$$

$$
\equiv\; \int_0^\tau K_h(u - t^*) dH(u)
$$

with

$$
H(u) = \int_0^u n^{-1} \sum_{i=1}^n \sum_{k=1}^K \left\{ [\boldsymbol{\beta}(v) - \boldsymbol{\beta}_0(v)]^T \mathbf{Z}_{ik} - log \frac{\mathcal{S}_g^{(0)}[\boldsymbol{\gamma}(t^*), v]}{\mathcal{S}_g^{(0)}(0, v)} \right\} dM_{ik}(v),
$$

a locally square integrable martingale since $n^{-1} \sum_{i=1}^n \sum_{k=1}^K \left\{ [\boldsymbol{\beta}(t^*) - \boldsymbol{\beta}_0(t^*)]^T \mathbf{Z}_{ik} - log \left\{ \frac{\mathcal{S}_g^{(0)}[\boldsymbol{\gamma}(t^*), u]}{\mathcal{S}_g^{(0)}(0, u)} \right\} \right\}$
is predictable. Because we assume K_h to be of bounded variation, then for $t^* \in [t_1^*, t_2^*]$ and n large enough, we have

$$
|X_n^*[\boldsymbol{\beta}(t^*), t^*]| \leq 2h^{-1} V(K_h) \sup_{u \in [0,s]} |H(u)| \tag{A.2}
$$

where $V(K_h)$ denotes the total variation of K_h. By Lenglart's inequality (Andersen et al. (1995), equation 2.5.18), we get for any $\delta, \eta > 0$

$$
P\left\{ \sup_{u \in [0,s]} \left| h^{-1} H(u) \right| > \eta \right\} \leq
$$

$$
\frac{\delta}{\eta^2} + P\left\{ (nh)^{-2} \int_0^s \sum_{i=1}^n \sum_{k=1}^K \left\{ [\boldsymbol{\beta}(t^*) - \boldsymbol{\beta}_0(t^*)]^T \mathbf{Z}_{ik} - log \frac{\mathcal{S}_g^{(0)}[\boldsymbol{\gamma}(t^*), u]}{\mathcal{S}_g^{(0)}(0, u)} \right\}^2 R_{ik}(u) \lambda_{ik}[u | \mathbf{Z}_{ik}(u)] du > \delta \right\}
$$

where $n^{-1} \int_0^s \sum_{i=1}^n \sum_{k=1}^K \left\{ [\boldsymbol{\beta}(t^*) - \boldsymbol{\beta}_0(t^*)]^T \mathbf{Z}_{ik} - log \left\{ \frac{\mathcal{S}_g^{(0)}[\boldsymbol{\gamma}(t^*), u]}{\mathcal{S}_g^{(0)}(0, u)} \right\} \right\}^2 R_{ik}(u) \lambda_{ik}[u | \mathbf{Z}_{ik}(u)] du$ converges in probability to a finite function. Thus the right hand side of A.2 converges to zero in probability if $nh^2 \to \infty$ as $n \to \infty$. This shows that $\sup_{t^* \in [t_1^*, t_2^*]} |X_n^*[\boldsymbol{\beta}(t^*), t^*]| \to_p 0$.

Coupling this result with the pointwise consistency of $\hat{\boldsymbol{\beta}}(t^*)$ gives uniform consistency. We see here that the bandwidth must tend toward zero more slowly to obtain uniform consistency than to obtain pointwise consistency.

A.3   Proof of theorem 3.2

The proof of 3.2 consists of the following main steps.

STEP 1. Consistency: $\hat{\boldsymbol{\gamma}}(t^*) \to_p 0$. This is shown in theorem 3.1.

STEP 2. Asymptotic normality of $\frac{\partial \ell[\boldsymbol{\gamma}(t^*), \tau]}{\partial \boldsymbol{\gamma}(t^*)} |_{\boldsymbol{\gamma}(t^*) = 0} = \ell'(0, \tau)$.

30

Now,

$$
\begin{aligned}
\ell'(0,\tau) &= \int_0^\tau K_h(u-t^*)n^{-1}\sum_{i=1}^n\sum_{k=1}^K\left[\tilde{\mathbf{U}}_{ik}(u,u-t^*)-\frac{\mathcal{S}_g^{(1)}(u)}{\mathcal{S}_g^{(0)}(u)}\right]dN_{ik}(u)\\
&= \int_0^\tau K_h(u-t^*)n^{-1}\sum_{i=1}^n\sum_{k=1}^K\left[\tilde{\mathbf{U}}_{ik}(u,u-t^*)-\frac{\mathcal{S}_g^{(1)}(u)}{\mathcal{S}_g^{(0)}(u)}\right]dM_{ik}(u)\\
&\quad + \int_0^\tau K_h(u-t^*)n^{-1}\sum_{i=1}^n\sum_{k=1}^K\left[\tilde{\mathbf{U}}_{ik}(u,u-t^*)-\frac{\mathcal{S}_g^{(1)}(u)}{\mathcal{S}_g^{(0)}(u)}\right]R_{ik}(u)exp[\boldsymbol{\beta}_0(u)^T\mathbf{Z}_{ik}]\lambda_0(u)du\\
&= W_n + B_n,
\end{aligned}
$$

First note that

$$
\begin{aligned}
B_n &= \int_0^\tau K_h(u-t^*)n^{-1}\sum_{i=1}^n\sum_{k=1}^K\left[\tilde{\mathbf{U}}_{ik}(u,u-t^*)-\frac{\mathcal{S}_g^{(1)}(u)}{\mathcal{S}_g^{(0)}(u)}\right]\\
&\qquad R_{ik}(u)\left\{exp[\boldsymbol{\beta}_0(u)^T\mathbf{Z}_{ik}]-exp[\boldsymbol{b}_0(t^*)^T\tilde{\mathbf{Z}}_{ik}(1,u-t^*)]\right\}\lambda_0(u)du\\
&= \int_0^\tau K_h(u-t^*)n^{-1}\sum_{i=1}^n\sum_{k=1}^K\left[\tilde{\mathbf{U}}_{ik}(u,u-t^*)-\frac{\mathcal{S}_g^{(1)}(u)}{\mathcal{S}_g^{(0)}(u)}\right]\\
&\qquad R_{ik}(u)exp[\boldsymbol{\beta}_0(u)^T\mathbf{Z}_{ik}]\left\{\frac{1}{2}(u-t^*)^2\boldsymbol{\beta}_0''(t^*)\mathbf{Z}_{ik}\lambda_0(u)du + o_p[(u-t^*)^2]\right\}\\
&= \frac{\boldsymbol{\beta}_0''(t^*)h^2}{2}\int_{-1}^1 K(w)(w^2,\ w^3)^T dw \otimes\\
&\qquad n^{-1}\sum_{i=1}^n\sum_{k=1}^K\left[\mathbf{Z}_{ik}-\frac{\tilde{\mathcal{S}}_g^{(1)}(t^*)}{\mathcal{S}_g^{(0)*}(t^*)}\right]R_{ik}(t^*)exp[\boldsymbol{\beta}(t^*)^T\mathbf{Z}_{ik}]\mathbf{Z}_{ik}\lambda_0(t^*)dt^* + o_p(1)\\
&\doteq \frac{1}{2}\boldsymbol{\beta}''(t^*)h^2\int_{-1}^1 K(w)(w^2,\ w^3)^T dw \otimes \Sigma(t^*) \equiv b_n
\end{aligned}
$$

We now consider the process $n^{-1/2}V_n(\tau) = \sqrt{nh}W_n(\tau)$. In the independent situation, it is a local square integrable martingale and thus one can apply Rebolledo's Central Limit Theorem to show asymptotic normality. When the event times are possibly correlated, we first approximate $V_n(\tau)$ with $\tilde{V}_n(\tau)$,

$$
\tilde{V}_n(\tau) = \sum_{i=1}^n\sum_{k=1}^K\sqrt{h}\int_0^\tau K_h(u-t^*)\left[\tilde{\mathbf{U}}_{ik}(1,u-t^*)-\frac{\boldsymbol{s}_g^{(1)}(u)}{\boldsymbol{s}_g^{(0)}(u)}\right]dM_{ik}(u)
$$

Note that $\tilde{V}_n(\tau)$ is a sum of n independent and identically distributed random vectors, whose ith

31

element is

$$\tilde{V}_i(\tau) = \sum_{k=1}^{K} \sqrt{h} \zeta_{ik}(\tau)$$

with

$$\zeta_{ik}(\tau) = \int_0^\tau K_h(u - t^*) \left[\tilde{\mathbf{U}}_{ik}(1, u - t^*) - \frac{\mathcal{S}_g^{(1)}(u)}{\mathcal{S}_g^{(0)}(u)} \right] dM_{ik}(u).$$

To prove the asymptotic normality by multivariate Central Limit Theorem, we need to check that $E\left\{ [\tilde{V}_i(\tau)]^2 \right\} < \infty$. now,

$$E\left\{ [\tilde{V}_i(\tau)]^2 \right\} = hE \left(\sum_{k=1}^{K} \sum_{l=1}^{K} \zeta_{ik} \zeta_{il} \right) \leq h \sum_{k=1}^{K} \sum_{l=1}^{K} \left[E\zeta_{ik}^2 E\zeta_{il}^2 \right]^{1/2}$$

The last inequality follows from Cauchy-Schwarz inequality. Now, $\zeta_{ik}(v)$ is a local square integrable martingale with predictable variation process, i.e.,

$$
\begin{aligned}
E[\zeta_{ik}(v)^2] &= var[\zeta_{ik}(v)] = \langle \zeta_{ik}, \zeta_{ik} \rangle(v) \\[2mm]
&= \left\{ \int_0^v K_h^2(u - t^*) \left[\tilde{\mathbf{U}}_{ik}(1, u - t^*) - \frac{\mathcal{S}_g^{(1)}(u)}{\mathcal{S}_g^{(0)}(u)} \right]^{\otimes 2} R_{ik}(u) \lambda[u|\mathbf{Z}_{ik}(u)] du \right\} \\[2mm]
&= \frac{1}{h} \begin{pmatrix} 1 & w \\ w & w^2 \end{pmatrix} K^2(w) dw \otimes \left[Q_0(t^*) Q_2(t^*) - Q_1(t^*)^{\otimes 2} \right] / Q_0(t^*) + o_p(1) \\[2mm]
&= \frac{1}{h} \varsigma \otimes \Sigma(t^*) + o_p(1)
\end{aligned}
$$

It then follows that $E\left\{ [\tilde{V}_i(\tau)]^2 \right\} \leq h \frac{1}{h} \varsigma \otimes \Sigma(t^*) < \infty$. By Multivariate Central Limit Theorem, the process $n^{-1/2} V_n(v) = \sqrt{nh} W_n(v)$ converges to a normal random vector with mean 0 and covariance Π, which can be consistently estimated by

$$\hat{\Pi} = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{l=1}^{K} \hat{\zeta}_{ik} \hat{\zeta}_{il}$$

with

$$
\begin{aligned}
\hat{\zeta}_{ik}[\hat{\boldsymbol{b}}(t^*)] &= \Delta_{ik} K_h(X_{ik} - t^*) \left\{ \mathbf{Z}_{ik} - \frac{\sum_{g=1}^{G} \mathcal{S}_g^{(1)}[\hat{\boldsymbol{b}}(t^*); X_{ik}] 1(s_{ik} \in \mathcal{I}_g)}{\sum_{g=1}^{G} \mathcal{S}_g^{(0)}[\hat{\boldsymbol{b}}(t^*); X_{ik}] 1(s_{ik} \in \mathcal{I}_g)} \right\} \otimes \psi(X_{ik}) \\[2mm]
&- \sum_{j=1}^{n} \frac{\Delta_{jk} R_{ik}(X_{jk}) exp[\hat{\boldsymbol{\beta}}(X_{ik})^T \mathbf{Z}_{ik}]}{\sum_{g=1}^{G} \mathcal{S}_g^{(0)}[\hat{\boldsymbol{b}}(t^*); X_{jk}] 1(s_{ik} \in \mathcal{I}_g)} \times \left\{ \mathbf{Z}_{ik} - \frac{\sum_{g=1}^{G} \mathcal{S}_g^{(1)}[\hat{\boldsymbol{b}}(t^*); X_{jk}] 1(s_{ik} \in \mathcal{I}_g)}{\sum_{g=1}^{G} \mathcal{S}_g^{(0)}[\hat{\boldsymbol{b}}(t^*); X_{jk}] 1(s_{ik} \in \mathcal{I}_g)} \right\} \\[2mm]
&\times K_h(X_{jk} - t^*) \otimes \psi(X_{jk}),
\end{aligned}
$$

32

Here $\psi(u)$ denotes $[1, \ (u - t^*)/h]$, and for $j = 0, 1$,

$$\mathcal{S}_g^{(j)}[\hat{\boldsymbol{b}}(t^*); u] = \sum_{i=1}^n \sum_{k=1}^K R_{ik}(u) exp[\hat{\boldsymbol{b}}(t^*)^T \tilde{\mathbf{Z}}_{ik}(1, u - t^*)] \mathbf{Z}_{ik}^{\otimes j} 1(s_{ik} \in \mathcal{I}_g)$$

Therefore, in STEP 2 we establish that:

$$\sqrt{nh}[\ell'(0, \tau) - b_n(\tau)] \to_d N[0, \Pi(\tau)] \tag{A.3}$$

STEP 3. for any $\boldsymbol{\gamma}(t^*)^* \to_p 0$, $\ell''[\boldsymbol{\gamma}(t^*)^*, \tau] \to_p \Xi(\tau)$, a finite quantity.

We have for any consistent estimator $\boldsymbol{\gamma}(t^*)^* \to_p 0$, since all the involved random variables are bounded by assumption, it can be shown that

$$\ell''[\boldsymbol{\gamma}(t^*), \tau] = \ell''(0, \tau) + o_p(1).$$

and

$$
\begin{aligned}
\ell''(0, \tau) &= \int_0^\tau K_h(u - t^*) n^{-1} \sum_{i=1}^n \sum_{k=1}^K \frac{\mathcal{S}_g^{(2)}(u) \boldsymbol{s}_g^{(0)}(u) - \boldsymbol{s}_g^{(1)}(u) \boldsymbol{s}_g^{(1)}(u)^T}{\boldsymbol{s}_g^{(0)}(u)^2} dN_{ik}(u) \\
&= \int_0^\tau K_h(u - t^*) n^{-1} \sum_{i=1}^n \sum_{k=1}^K \frac{\mathcal{S}_g^{(2)}(u) \boldsymbol{s}_g^{(0)}(u) - \boldsymbol{s}_g^{(1)}(u) \boldsymbol{s}_g^{(1)}(u)^T}{\boldsymbol{s}_g^{(0)}(u)^2} dM_{ik}(u) \\
&\quad + \int_0^\tau K_h(u - t^*) n^{-1} \sum_{i=1}^n \sum_{k=1}^K \frac{\mathcal{S}_g^{(2)}(u) \boldsymbol{s}_g^{(0)}(u) - \boldsymbol{s}_g^{(1)}(u) \boldsymbol{s}_g^{(1)}(u)^T}{\boldsymbol{s}_g^{(0)}(u)^2} \mathcal{S}^*(u) \lambda_0(u) du \\
&= I + II,
\end{aligned}
$$

By considering the second moment of I, we have

$$I = O_p[(nh)^{-1}]$$

It can be shown that

$$II = \Sigma(t^*) \otimes \int K(w) \begin{pmatrix} 1 & w \\ w & w^2 \end{pmatrix} dw + o_p(1)$$

Therefore,

$$\ell''(0, \tau) \to_p \Sigma(t^*) \otimes \int K(w) \begin{pmatrix} 1 & w \\ w & w^2 \end{pmatrix} dw \equiv \Xi(\tau) \tag{A.4}$$

STEP 4. Asymptotic normality of $\hat{\boldsymbol{\gamma}}(t^*)$.

It follows from a Taylor series expansion around 0 that

$$0 = \ell'[\hat{\boldsymbol{\gamma}}(t^*), \tau] = \ell'(0, \tau) + \ell''[\hat{\boldsymbol{\gamma}}(t^*)^*, \tau] \hat{\boldsymbol{\gamma}}(t^*),$$

33

where $\hat{\boldsymbol{\gamma}}(t^*)^*$ lies between 0 and $\hat{\boldsymbol{\gamma}}(t^*)$. By A.3 and A.4 and Slutsky's theorem, we have

$$\sqrt{nh}\left[\hat{\boldsymbol{\gamma}}(t^*) + \Xi(\tau)^{-1}b_n(\tau)\right] \to_d N[0, \Xi(\tau)^{-1}\Pi(\tau)\Xi(\tau)^{-1}]$$

when K is symmetric, i.e., when t^* is an interior point of $[0, \tau]$, $\int_{-1}^{1} K(u)du = 1$ and $\int_{-1}^{1} uK(u)du = 0$, thus the distribution of $\hat{\boldsymbol{\beta}}(t^*)$ can be simplified as:

$$\sqrt{nh}\left[\hat{\boldsymbol{\beta}}(t^*) - \boldsymbol{\beta}_0(t^*) - \frac{h^2\mu_2}{2}\boldsymbol{\beta}''(t^*) + o_p(h^2)\right] \to N[0, \Sigma^{-1}(t^*)\Pi^q\Sigma^{-1}(t^*)]$$

where Π^q is the upper $q \times q$ submatrix of Π. A consistent estimator for $\Sigma(t^*)$ is given in Cai and Sun (2003). In the multivariate situation, an estimator takes the following form:

$$\hat{\Sigma}(t^*) = \sum_{i=1}^{n}\sum_{k=1}^{K}\Delta_{ik}K_h(X_{ik} - t^*)\frac{G_g^{(0)}[u, \hat{\boldsymbol{\beta}}(t^*)]G_g^{(2)}[u, \hat{\boldsymbol{\beta}}(t^*)] - \{G_g^{(1)}[u, \hat{\boldsymbol{\beta}}(t^*)]\}^{\otimes 2}}{\{G_g^{(0)}[u, \hat{\boldsymbol{\beta}}(t^*)]\}^2}$$

with

$$G_g^{(j)}[u, \hat{\boldsymbol{\beta}}(t^*)] = \sum_{i=1}^{n}\sum_{k=1}^{K}R_{ik}(u)exp[\hat{\boldsymbol{\beta}}(t^*)^T\mathbf{Z}_{ik}]\mathbf{Z}_{ik}^{\otimes j}1(s_{ik} \in \mathcal{I}_g).$$