On the proportional hazards model with last observation carried forward covariates

Hongyuan Cao[1] and Jason P. Fine[2]

Abstract

Standard partial likelihood methodology for the proportional hazards model with time-dependent covariates requires knowledge of the covariates at the observed failure times, which is not realistic in practice. A simple and commonly used estimator imputes the most recently observed covariate prior to each failure time, which is known to be biased. In this paper, we show that a weighted last observation carried forward approach may yield valid estimation. We establish the consistency and asymptotic normality of the weighted partial likelihood estimators and provide a closed form variance estimator for inference. The estimator may be conveniently implemented using standard software. Interestingly, the convergence rate of the estimator is slower than the parametric rate achieved with fully observed covariates but the same as that obtained with all lagged covariate values. Simulation studies provide numerical support for the theoretical findings. Data from an Alzheimer's study illustrate the practical utility of the methodology.

Keywords: Convergence rates; Kernel weighted estimation; Last value imputation; Partial likelihood; Time-varying covariates.

[1] Department of Statistics, Florida State University, Tallahassee, Florida 32306, U.S.A.
[2] Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514, U.S.A.

1 Introduction

In clinical trials and epidemiological studies, covariates are often collected longitudinally. Incorporation of these covariates into survival analysis is challenging since these covariates are only observed at a finite number of time points. As an example, the proportional hazards model (Cox, 1972) requires covariate values at each failure time. Very rarely do the failure times coincide with the covariate observation times (Liu and Craig, 2006).

These issues may be understood more precisely by representing the event history using counting processes. In the failure time setting, $N(t)$ indicates whether an event has occurred by time t and $Z(\cdot)$ is a p-dimensional covariate process. The proportional hazards model specifies the hazard function for $N(t)$ conditionally on the history of $Z(r), r \leq t$ as

$$\lambda\{t \mid Z(r), r \leq t\} = \lambda_0(t) e^{\beta_0^T Z(t)}, \tag{1.1}$$

where $\lambda_0(\cdot)$ is an unspecified function and β_0 is a vector of unknown regression parameters.

A simple and frequently used estimator with time-dependent covariates is to impute the last observed covariate value prior to each failure time. Such analyses are problematic (Molenberghs et al., 2002; Molnar et al., 2008). First, it is assumed that the longitudinal covariate does not change from the time of the last measurement. Second, no distinction is made between those subjects who had a valid measurement and those subjects with imputed values, artificially increasing the amount of information in the data. These issues can induce substantial biases in parameter estimates and lead to inaccurate inferences (Andersen and Liestol , 2003).

To circumvent these problems, likelihood based approaches such as joint modeling have been proposed as more principled methods for analysis (Tsiatis et al., 1995; Henderson et al., 2000; Xu and Zeger, 2001; Rizopoulos, 2012). Commonly, the time-dependent covariate follows a linear mixed effects model with normal measurement error and the hazard function depends on the underlying random effects of this covariate process. Inference may be based on the joint likelihood of the survival and longitudinal data under parametric assumptions on the random effects. As an alternative, Tsiatis and Davidian (2001) proposed a semiparametric conditional score estimator for the covariate effect that requires no assumptions on the distribution of the random effects. Such methods impose stringent modeling assumptions and the inferences they produce are highly dependent on untestable and often implicit assumptions regarding the distribution of the unobserved measurements. Previous numerical work has shown that such estimators may be quite biased with model misspecification (Cao et al., 2015).

In this paper, we propose an intuitively appealing weighting approach which retains the simplicity of last observation carried forward imputation. The main idea is that the further the last observation is from the current failure time, the less it should contribute to the estimating equation. This is handled formally by weighting the last observation as a decreasing function of the time between the most recently observed covariates and the failure event. Models for the underlying covariate process and for the dependence structure between that process and the event history process are unspecified, unlike the joint modeling approach. Cao et al. (2015) proposed similar weighting approaches using all backward lagged covariates. We adapt these techniques to obtain valid estimation employing only the most recently observed covariate, denoted weighted last observation carried forward. This method may be implemented in standard software for the proportional hazards model permitting time-dependent weights, which is not possible when using all lagged covariate measurements.

The paper is organized as follows. In section 2, we discuss the proposed weighted last observation carried forward estimator and corresponding asymptotic properties and inferences. Interestingly, the proposed estimators converge more slowly than the usual parametric rates which are achieved with fully observed covariates but at the same rate as in Cao et al. (2015). Section 3 reports simulation studies which evidence little loss of efficiency versus using all previously observed covariates. Application to an Alzheimer's dataset illustrates the practical utility of the methodology in Section 4. Concluding remarks are given in Section 5. Proofs of results from Section 2 are relegated to the Appendix.

2 Estimation and inference

2.1 Notation and last observation carried forward

Let T be the failure time and let C be the corresponding censoring variable. We assume that censoring is coarsened at random such that T and C are conditionally independent given $Z(\cdot)$ (Heitjan and Rubin, 1991). Let $[\{T_i, Z_i(\cdot), C_i\}, i = 1, \ldots, n]$ be n independent copies of $\{T, Z(\cdot), C\}$. The longitudinal covariates are observed at M_i observation times $R_{ik} \leq X_i, k = 1, \ldots, M_i$, where $X_i = \min(T_i, C_i)$, and M_i is assumed finite with probability one. The p-dimensional covariate process may include both time-independent and time-dependent covariates, under the restriction that the time-dependent covariates are observed at the same time points within individuals. The measurement times R_{ik} are assumed independent of the measurements $Z_i(R_{ik}), k = 1, \ldots, M_i$. The observed data consist of the n independent realizations $\{X_i, \Delta_i, Z_i(R_{ik}), R_{ik}, k = 1, \ldots, M_i\}, i = 1, \ldots, n$, where $\Delta_i = 1$ if $X_i = T_i$ and 0 otherwise.

To present the estimators, we adopt the counting process notation, where $N_i(t) = I(X_i \leq t, \Delta_i = 1)$ and $Y_i(t) = I(X_i \geq t)$. With fully observed covariates, the partial likelihood for model (1.1) is

$$L_n(\beta) = \prod_{i=1}^{n} \prod_{t \geq 0} \{ \frac{e^{\beta^T Z_i(t)}}{\sum_{j=1}^{n} Y_j(t) e^{\beta^T Z_j(t)}} \}^{\Delta N_i(t)}, \tag{2.2}$$

$$\text{where} \qquad \Delta N_i(t) = \begin{cases} 1 & \text{if} \quad N_i(t) - N_i(t-) = 1 \\ 0 & \text{otherwise.} \end{cases}$$

The log partial likelihood is:

$$
\begin{aligned}
l_n(\beta) &= n^{-1}\log L_n(\beta) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\Big[\beta^T Z_i(u) - \log\{\sum_{j=1}^{n} Y_j(u)e^{\beta^T Z_j(u)}\}\Big]dN_i(u) \\
&= n^{-1}\sum_{i=1}^{n}\Delta_i\Big[\beta^T Z_i(X_i) - \log\{\sum_{j=1}^{n} Y_j(X_i)e^{\beta^T Z_j(X_i)}\}\Big], \qquad (2.3)
\end{aligned}
$$

where τ is a prespecified time point such that $\mathrm{pr}(X > \tau) > 0$. Because $Z_i(u), i = 1, \ldots, n$, are not observed continuously, $l_n(\beta)$ is not computable from the observed data.

To use the last observation carried forward, in (2.3), $\beta^T Z_i(X_i)$ is replaced by $\beta^T Z_i(s_i)$ and $\log\{\sum_{j=1}^{n} Y_j(X_i)e^{\beta^T Z_j(X_i)}\}$ is replaced by $\log\{\sum_{j=1}^{n} Y_j(X_i)e^{\beta^T Z_j(s_i)}\}$, where $s_i = \max\{x \leq X_i, x \in (R_{i1}, \ldots, R_{iM_i})\}, i = 1, \ldots, n$. This method assumes that the subject's covariate does not change from the most recent observation time and does not account for the variability inherent in this imputation. These assumptions may not hold in practice and violations can confound covariates with time, which in turn can bias estimates of covariate effects and their standard errors. As a result, the magnitude and even the direction of bias from last observation carried forward is extremely difficult, if not impossible, to determine *a priori.*

2.2 Weighted last observation carried forward

We propose to remedy this bias by adopting a weighting strategy, downweighing imputed values which are far in time from the failure event. To be specific, for a sample of n independent subjects, the weighted log partial likelihood is

$$
\begin{aligned}
&l_n^*(\beta) = \qquad (2.4) \\
&n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\int_0^{\tau} J(u,r)\Big(\beta^T Z_i(r) - \log[\sum_{j=1}^{n}\int_0^{\tau} J(u,s)Y_j(u)e^{\beta^T Z_j(s)}dN_i^*(s)]\Big)dN_i^*(r)dN_i(u),
\end{aligned}
$$

where $J(u,r) = K_h(u-r)I\{r \le u, \int_r^u dN_i^*(t) = 0\}$, denoting the weighted last observation, $N_i^*(t) = \sum_{k=1}^{M_i} I(R_{ik} \le t)$ is a realization of $N^*(t)$, the counting process for the covariate observation times, $K_h(t) = K(t/h)/h, h$ is the bandwidth and the kernel function $K(t)$ is a symmetric probability density with mean 0, and bounded first derivative. In simulation studies and real data analysis, we use Epanechnikov kernel $K(x) = 0.75(1-x^2)_+$ due to its good empirical performance (Fan and Gijbels, 1996). If $x > h$, $K_h(x) = K(x.h)/h = 0$. Consequently, if the distance between the last observed covariate and the failure time is greater than h for a subject, this subject's failure time does not contribute to the estimating equation. This subject contributes to the estimating equation via $\sum_{j=1}^n \int_0^\tau J(u,s)Y_j(u)e^{\beta^T Z_j(s)}dN_i^*(s)$ at other observed failure times. When $h \to \infty$, the proposed weighted last observation carried forward reduces to the last observation carried forward and bias will incur. As sample size $n \to \infty$, $h \to 0$ to ensure that bis is negligible. h strikes a balance between the bias and the variability. Smaller h produces smaller bias yet larger variability. On the other hand, larger h results larger bias and smaller variability. In practice, we choose h to minimize the mean squared error.

Define $\hat{\beta}$ to be the maximizer of $l_n^*(\beta)$. This estimator is a root of the score function $U_n(\beta) = 0$, where

$$U_n(\beta) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \int_0^\tau J(u,r)\{Z_i(r) - \bar{Z}(\beta,u)\}dN_i(u)dN_i^*(r), \tag{2.5}$$

where

$$S_n^{(k)}(\beta,u) = n^{-1}\sum_{j=1}^n \int_0^\tau J(u,s)Y_j(u)Z_j(s)^{\otimes k}e^{\beta^T Z_j(s)}dN_j^*(s),$$

and $\bar{Z}(\beta,u) = S_n^{(1)}(\beta,u)/S_n^{(0)}(\beta,u)$, $k = 0,1,2, a^{\otimes 0} = 1, a^{\otimes 1} = a$, and $a^{\otimes 2} = aa^T$. It can be seen from (2.5) that different individuals receive different weights inside the integral in $U_n(\beta)$ depending on the time between the most recently observed longitudinal covariate and the observed failure event. This is also reflected in $\bar{Z}(\beta, u)$. Regarding the computation, (2.4) is concave in β and therefore there exists a unique root of (2.5). Once the kernel function K has been chosen and the bandwidth has been fixed, the estimating equation can be solved using a standard Newton-Raphson method, with good convergence properties. Standard software for the proportional hazards model accommodating time-dependent covariates and time-dependent weights may be used for these computations. If for certain subjects, there are no covariate observations before their observed failure time, the subject's failure time does not contribute to the estimating equation (2.5). Such subjects still contribute to (2.5) via $\bar{Z}(\beta, u)$ at other observed failure times.

The weighted last observation carried forward method is not a special case of the backward lagged covariates approach (Cao et al., 2015). In the weighted log partial likelihood function (2.4), $I\{r \leq u, \int_r^u dN_i^*(t) = 0\}$ precludes the contribution of other covariates except for the last observed covariate into the estimating equation. On the other hand, all covariates prior to the failure time contribute to the estimating equation in the backward lagged covariates approach. Even if the bandwidth h is extremely small, the backward lagged covariates approach cannot guarantee to include only last observed covariate.

If $M_i = 1$, the two partial likelihoods are the same. In the backward lagged covariate approach, for each failure time, all backward lagged covariates contribute to the partial likelihood and their effects are aggregated by summation weighted by the difference between the longitudinal observation time and the failure time. In the general case, one cannot get weighted last observation approach from backward lagged covariates approach or vice versa.

2.3 Statistical inference and asymptotic properties

To state our key results, additional notation and regularity conditions are needed. Denote $E\{dN_i(t) \mid \mathcal{F}_s, s \leq t\} = Y_i(t)e^{\beta_0^T Z_i(t)}\lambda_0(t)dt$, where $\lambda_0(t)$ is assumed twice continuously differentiable and strictly positive for $t \in [0, \tau]$, and $\mathcal{F}_t$ is the filtration, which includes all information in $\{N_i(s), Y_i(s), Z_i(s), s \leq \min(t, X_i)\}$, as well as the measurement times up to time $t, i = 1, \ldots, n$. For $u > r$, the measurement times are allowed to depend on covariates through

$$E\Big[dN_i^*(r) I\{N_i^*(u) - N_i^*(r+) = 0\} \mid Z_i(r)\Big] = \lambda^*\{r, u; Z_i(r)\} dr. \tag{2.6}$$

This assumption is weaker than that specified in Cao et al. (2015). Denote

$$s^{(k)}(\beta, t) = E\Big[Y_i(t) Z_i(t)^{\otimes k} e^{\beta^T Z_i(t)} \lambda^*\{t, t; Z_i(t)\}\Big].$$

The following results provide the limiting distribution of the proposed estimator.

Theorem 1 *Under (C1)-(C6) specified in the Appendix, we have*

$$(nh)^{1/2} A(\beta_0)(\hat{\beta} - \beta_0) \xrightarrow{d} N\{0, \Sigma(\beta_0)\}, \tag{2.7}$$

where

$$A(\beta_0) = \int_0^\tau \{s^{(2)}(\beta_0, t) - \frac{s^{(1)}(\beta_0, t)^{\otimes 2}}{s^{(0)}(\beta_0, t)}\} \lambda_0(t) dt,$$

β_0 *is the true regression coefficient and the asymptotic variance*

$$\Sigma(\beta_0) = \int_0^\infty K(x)^2 dx \int_0^\tau \{s^{(2)}(\beta_0, t) - \frac{s^{(1)}(\beta_0, t)^{\otimes 2}}{s^{(0)}(\beta_0, t)}\} \lambda_0(t) dt.$$

In practice, inference is conducted based on the estimating equation (2.5). The first moment of $U_n(\hat{\beta})$ is 0 and we can estimate the variance of $U_n(\hat{\beta})$ by

$$\hat{\Sigma} = n^{-2} \sum_{i=1}^n \Big(\int_0^\tau \int \Big[J(u, r)\{Z_i(r) - \bar{Z}(\beta, u)\} \Big] dN_i^*(r) dN_i(u) \Big)^{\otimes 2} |_{\beta = \hat{\beta}}.$$

By Taylor expansion, the variance of $\hat{\beta}$ can be estimated by

$$\{\frac{\partial U_n(\beta)}{\partial \beta}|_{\beta=\hat{\beta}}\}^{-1}\hat{\Sigma}\{\frac{\partial U_n(\beta)}{\partial \beta}|_{\beta=\hat{\beta}}\}^{-1}.$$

We show the validity of this approach in the following corollary.

Corollary 1 *Under conditions (C1)-(C6) specified in the Appendix, the sandwich formula*

$$\{-\frac{\partial U_n(\beta)}{\partial \beta}|_{\beta=\hat{\beta}}\}^{-1}\hat{\Sigma}\{-\frac{\partial U_n(\beta)}{\partial \beta}|_{\beta=\hat{\beta}}\}^{-1}$$

consistently estimates the variance of $\hat{\beta}$.

The validity of the weighted last observation carried forward method in Theorem 1 depends on an appropriate choice of bandwidth. The bias is of order $O(h)$ as shown in the Appendix. Therefore, the allowable range of valid bandwidths is $(n^{-1}, n^{-1/3})$ as specified in condition (C6) in the Appendix. With $h = o(n^{-1/3})$, we achieve a rate of convergence $o(n^{1/3})$. This rate of convergence is the same as the half kernel approach in Cao et al. (2015) but slower than the joint modeling approaches where strong modeling assumptions on the joint distribution of the covariate process and event times facilitates likelihood based inferences which may achieve parametric rates of convergence for the regression parameter β.

Following Cao et al. (2015), we propose a data adaptive bandwidth selection procedure. The idea is to minimize the mean squared error, where the bias and variance are calculated separately. From (A15), we know bias is of order h. We first regress $\hat{\beta}(h)$ on h in a reasonable range of h to obtain the slope estimate $\hat{C}$. To obtain the variance, we split the data randomly into two parts and obtain regression coefficient estimates $\hat{\beta}_1(h)$ and $\hat{\beta}_2(h)$ based on each half sample. The variance of $\hat{\beta}(h)$ is estimated by $\hat{V}(h) = \{\hat{\beta}_1(h) - \hat{\beta}_2(h)\}^2/4$. Using both $\hat{C}$ and $\hat{V}(h)$, we thus calculate the mean squared error as $\hat{C}^2h^2 + \hat{V}(h)$. Finally, we select the optimal bandwidth h minimizing this mean squared error.

3 Simulation studies

We conducted extensive simulation studies to compare the performance of the proposed estimator and the half kernel estimator in Cao et al. (2015). The simulated model is exactly the same as that in Cao et al. (2015). Specifically, we generated $1,000$ datasets, each consisting of $100, 400$ or 900 subjects. The total number of covariate observation times for each subject was Poisson distributed with intensity rate 8. The covariate observation times were generated from uniform distribution $\mathcal{U}(0,1)$. The covariate process was generated through a piecewise constant function

$$Z(t) = \sum_{i=1}^{20} I\{(i-1)/20 \leq t < i/20\} z_i,$$

where $z = (z_1, \ldots, z_{20})^T$ follows a unit variance multivariate normal distribution with mean 0 and correlation $e^{-|i-j|/20}, i, j = 1, \ldots, 20$. The survival time was simulated from model (1.1) with $\lambda_0(t) = 2$ and $\beta_0 = 1.5$. The censoring time was generated from a uniform distribution with lower bound 0 and upper bound giving censoring percentages of 15% and 50%. The results for other choices of the model parameters were rather similar and thus omitted.

For both estimators, the kernel function is the Epanechnikov kernel, which is $K(x) = 0.75(1-x^2)_+$. We employ bandwidths in the range $(n^{-1}, n^{-1/3})$ and the automatic bandwidth selection described in the Appendix and Cao et al. (2015). Similar results were obtained using other kernel functions.

Table 1 summarizes the main findings over $1,000$ simulations. We observe that the weighted last observation carried forward estimator performs satisfactorily in terms of bias, variance, and coverage probability, particularly with larger sample sizes. Compared with the half kernel approach, the proposed estimator has similar biases and loses little efficiency, generally less than 10%, with the empirical variances of the two estimators in good agreement at larger sample sizes. This finding can be explained heuristically: as the sample size increases, the weight assigned to the most recent covariate observation tends to dominate those from earlier measurements. The advantage of the weighted last observation carried forward estimator is that it is considerably easier to implement in practice using standard software and is valid under weaker assumptions.

Table 1: Weighted last observation carried forward and half kernel comparison with different censoring rate

		Weighted LOCF					Half Kernel				
n	BD	Bias	SD	SE	MSE	CP(%)	Bias	SD	SE	MSE	CP(%)
Censoring rate is 15%											
100	$n^{-0.7}$	0.059	0.501	0.419	0.254	92	0.051	0.502	0.420	0.255	91
	$n^{-0.6}$	0.025	0.367	0.340	0.135	93	0.019	0.384	0.346	0.148	91
	auto	0.039	0.427	0.386	0.184	94	0.042	0.434	0.385	0.190	93
400	$n^{-0.7}$	0.026	0.304	0.291	0.093	93	0.040	0.299	0.321	0.091	93
	$n^{-0.6}$	-0.005	0.232	0.222	0.054	94	-0.009	0.222	0.242	0.049	93
	auto	0.000	0.295	0.258	0.087	91	0.028	0.298	0.263	0.090	93
900	$n^{-0.7}$	0.006	0.248	0.246	0.062	94	0.008	0.247	0.253	0.061	95
	$n^{-0.6}$	0.001	0.177	0.175	0.031	95	0.006	0.181	0.188	0.033	94
	auto	0.010	0.223	0.213	0.050	96	0.014	0.244	0.216	0.060	94
Censoring rate is 50%											
100	$n^{-0.7}$	0.139	0.734	0.540	0.558	91	0.169	0.740	0.543	0.576	90
	$n^{-0.6}$	0.050	0.529	0.441	0.282	93	0.084	0.584	0.451	0.348	91
	auto	-0.035	0.434	0.452	0.190	90	0.047	0.595	0.500	0.356	90
400	$n^{-0.7}$	0.026	0.367	0.358	0.135	94	0.046	0.365	0.423	0.135	92
	$n^{-0.6}$	0.027	0.286	0.275	0.083	93	0.008	0.277	0.301	0.077	93
	auto	0.053	0.367	0.325	0.137	94	0.035	0.378	0.332	0.144	95
900	$n^{-0.7}$	0.019	0.323	0.299	0.105	93	0.033	0.305	0.327	0.094	93
	$n^{-0.6}$	0.006	0.226	0.216	0.051	93	0.007	0.220	0.236	0.048	94
	auto	0.043	0.289	0.265	0.085	95	0.020	0.306	0.260	0.094	93

Note: "LOCF" represents last observation carried forward, "BD" represents different bandwidths, "Bias" is the empirical bias, "SD" is the sample standard deviation, "SE" is the average of the standard error estimates, "MSE" is the mean squared error and "CP" represents the coverage probability of the 95% confidence interval for $\hat{\beta}$.

Per the request of a referee, we have provided additional simulations comparing our approach and the half kernel approach with two covariates, one time-dependent covariate and one time-independent covariate, to see the performance of our method in a multivariate regression case. The simulation set up is exactly the same as that in Cao et al. (2015). The results are summarized in Table 2. The results of half kernel and weighted LOCF are fairly comparable.

Table 2: Power comparison with both time-dependent and time-independent covariate

	Weighted LOCF					Half Kernel				
n	Bias	SD	SE	CP	Power(%)	Bias	SD	SE	CP	Power(%)
$\beta_1 = 0, \beta_2 = 0.5$										
100	-0.032	0.255	0.251	92	8	-0.005	0.295	0.250	91	8
400	0.012	0.210	0.194	93	6	-0.001	0.190	0.188	94	5
900	-0.025	0.142	0.165	96	3	0.003	0.166	0.163	93	6
$\beta_1 = -0.3, \beta_2 = 0.5$										
100	-0.001	0.280	0.270	92	18	0.021	0.280	0.253	92	24
400	0.026	0.194	0.191	92	31	-0.003	0.201	0.191	93	37
900	0.019	0.154	0.168	94	40	0.011	0.176	0.164	93	43
$\beta_1 = -0.15, \beta_2 = 0.5$										
100	0.012	0.265	0.246	94	11	0.008	0.281	0.250	91	12
400	-0.007	0.210	0.202	93	14	-0.006	0.203	0.192	93	14
900	0.026	0.163	0.165	94	18	0.005	0.166	0.164	94	15
$\beta_1 = -0.3, \beta_2 = 0$										
100	0.016	0.255	0.251	95	24	0.005	0.272	0.250	92	24
400	0.006	0.200	0.189	94	38	0.015	0.202	0.186	93	35
900	0.011	0.164	0.162	95	43	0.003	0.164	0.162	94	45

Note: The notations are the same as in Table 1.

To see how mean squared error changes with different bandwidth for the weighted last observation carried forward and the half kernel approach, we plot the mean squared error as a function of the bandwidth for sample size $n = 400$ and censoring rate 21.75%. As can be seen from Figure 1,the optimal bandwidth for half kernel and weighted last observation carried forward are close by.

4 Alzheimer's data example

We now illustrate the proposed inferential procedure in Section 2 with a comparison to the last observation carried forward and half kernel approach on data from an Alzheimer's study. This is a longitudinal population study of common chronic health problems of older persons, in a biracial neighborhood in Chicago from 1993 to 2006. Their demographics are recorded at baseline and they are longitudinally followed for clinical evaluation of Alzheimer's disease.

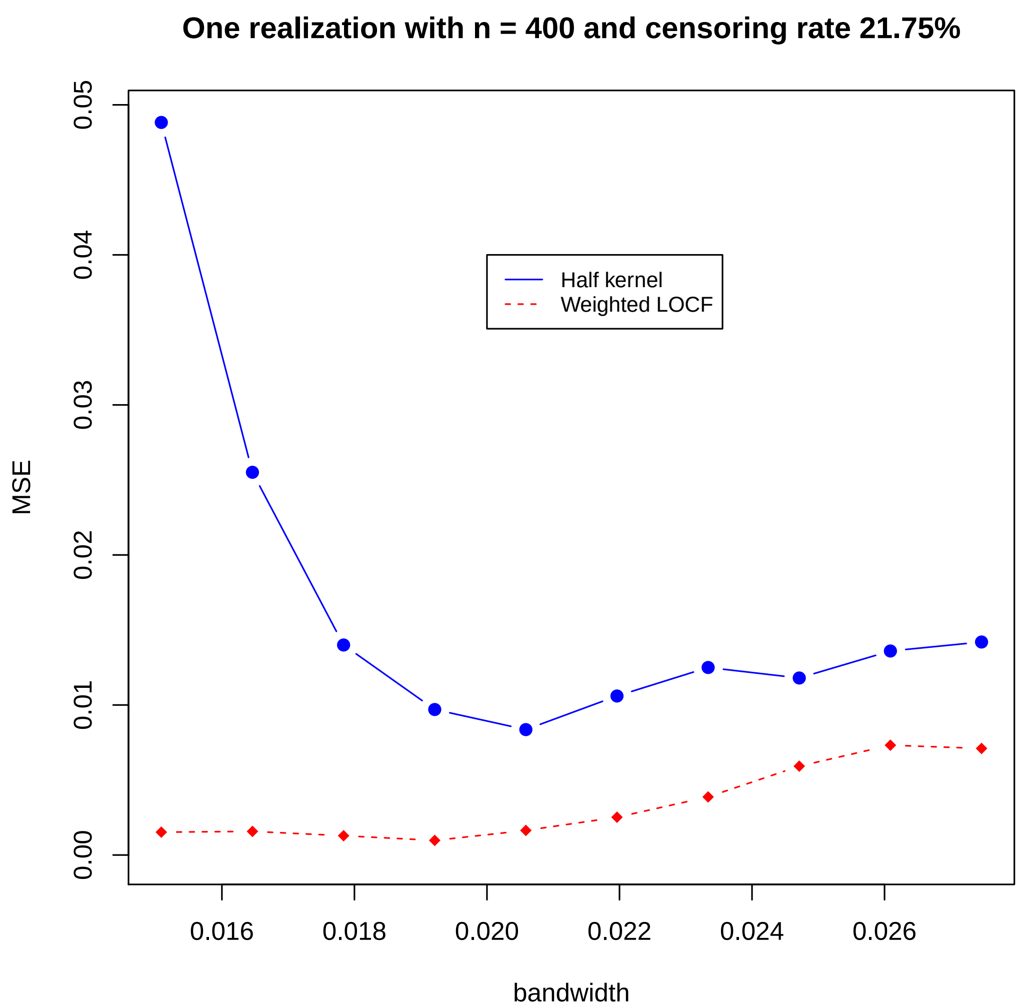

Figure 1: MSE as a function of bandwidth

Their ages range from 60 to 100. For each patient, the time origin is the first visit on study with the event time being the time since the first visit. We investigate the relationship between mortality and the longitudinal predictor mean corpuscular volume, the average volume of red cells in a specimen. Since the majority of subjects are caucasians (96%), our analysis is based on caucasians only. $2,209$ persons were used for analysis with 59.74% censoring. Details of the study design, methods and medical implications can be found in Bienias et al. (2003).

We use estimating equation (2.5) with bandwidths $h = 2(Q_3 - Q_1)n^{-\gamma}$, where Q_3 is the 0.75 quantile and Q_1 is the 0.25 quantile of the longitudinal measurement times, n is the number of persons and $\gamma = 0.6$ or 0.7. This effectively scales the time alignment to be consistent with our simulations. The results are summarized in Table 3 with fixed bandwidths and data adaptive bandwidth. Results based on last observation carried forward, half kernel approach and baseline mean corpuscular volume are presented for comparison.

We can see the negative association between time dependent mean corpuscular volume and mortality using weighted last observation carried forward and half kernel approach, which are statistically significant at the 0.05 level. The similarity of the two analyses, including the agreement of the standard errors, matches the results of the simulation studies. In contrast, last observation carried forward suggests a weak positive association, which is similar to results produced by using the baseline mean corpuscular volume. The positive association between mean corpuscular volume and mortality has recently been established using baseline only observations (Yoon et al., 2016). It is very interesting that time dependent mean corpuscular suggests a negative association. Further studies are needed to validate these findings.

5 Concluding remarks

In this paper, we proposed a weighted last observation carried forward approach for the proportional hazards model with time-dependent covariates. The newly proposed estimator

Table 3: Summary statistics for $\hat{\beta}$ based on (2.5).

	Weighted LOCF			LOCF	Half Kernel			Base
$h(\gamma)$	0.129 (0.6)	0.060 (0.7)	0.129 (auto)		0.129 (0.6)	0.060 (0.7)	0.069 (auto)	
$\hat{\beta}$	-0.066	-0.094	-0.066	0.014	-0.066	-0.094	-0.0860	0.016
HR	0.936	0.911	0.936	1.014	0.936	0.911	0.918	1.017
SE($\hat{\beta}$)	0.030	0.051	0.030	0.008	0.030	0.051	0.049	0.011
z-value	-2.193	-1.842	-2.193	1.840	-2.193	-1.842	-1.739	1.483
p-value	0.028	0.060	0.028	0.066	0.028	0.065	0.082	0.138

Note: "LOCF" represents last observation carried forward, "Half Kernel" represents the method that uses backward lagged covariates in Cao et al. (2015), "Base" represents analysis based on baseline time-independent mean corpuscular volume, "HR" represents hazard ratio, and "SE" represents standard error of the estimator.

is shown to be valid under weaker assumptions with little efficiency loss and is much faster to compute compared to the half kernel approach (Cao et al., 2015). Numerical studies corroborate our theoretical results and the proposed method can be conveniently implemented using standard software. While we focus our analysis on the proportional hazards model, our approach could also be used for other purposes such as additive hazards model, with additional development.

In practice, it may happen that time-dependent covariates are observed at different time points within individuals. For such scenarios, our proposed weighting methods would not be applicable. Alternative weighting methods could potentially be developed which incorporate that the time-dependent covariates are observed at different time points. This is beyond the scope of the current paper but is an interesting and important topic for future research.

For time-dependent longitudinal covariates, joint modeling is commonly used. Joint modeling consists of two sub-models: a longitudinal sub-model (such as a linear mixed effects model) and a time-to-event sub-model (such as a Cox proportional hazards model) which are linked using an association structure that quantifies the relationship between the outcomes of interest. If interest lies on the dynamics of the longitudinal process, such as accounting for informative dropout or the link between the outcomes, joint modeling will be used and the proposed method is not applicable. If interest lies on including the longitudinal variable as a time-dependent covariate in a time-to-event model, both joint modeling and the proposed method can be used, though investigators prefer one or another depending on the research question of interest. If investigators would like to use the slope of the population trajectory alone or in conjunction with the current value, joint modeling is preferred. On the other hand, if the longitudinal process is very difficult to model or the working linear mixed model is a mis-specified model, the weighted last observation carried forward is preferred. Joint modeling is more efficient when modeling assumptions in the joint modeling approach are satisfied.

Time-dependent covariates are pervasive in various disciplines. An alternative two stage modeling approach would be to apply kernel smoothing methods to the observed part of covariate $Z_i(t)$ for individual i to get an estimated curve $\{\hat{Z}_i(t) : t \leq X_i\}$, and to replace the missing $Z_i(t_j)$ with $\hat{Z}_i(t_j)$ for $t_j \leq X_i$. As the uncertainty inherent in $\hat{Z}_i(t_j)$ is ignored in the second stage, this approach may induce bias. Comparison between the two stage approach and the proposed method is beyond the current paper and warrant additional research.

A Appendix: Conditions of Theorem 1

We collect the required conditions of Theorem 1 below.

(C1) $\{N_i(\cdot), Y_i(\cdot), Z_i(\cdot)\}(i = 1, \dots, n)$ are independent and identically distributed.

(C2) $\text{pr}(C \geq \tau) > 0$ and $\text{pr}(T \geq \tau) > 0$.

(C3) For $r < u$, (2.6) is satisfied. $N(\tau)$ and $N^*(\tau)$ are bounded by finite constants, $\lambda_0(t)$ is twice continuously differentiable and $E\Big[Z(s)Y(t)e^{\beta_0^T Z(t)}\lambda^*\{s, t; Z(s)\}\Big]$ is twice continuously differentiable for $s, t \in [0, \tau]^{\otimes 2}$.

(C4) For $i = 1, \dots, n$, Z_i has bounded total variation, where $|Z_{ij}(0)| + \int_0^\tau |dZ_{ij}(t)| \leq D$ for all $j = 1, \dots, p$, where Z_{ij} is the jth component of Z_i and D is a finite constant.

(C5) $A(\beta_0) \equiv \int_0^\tau E\big[\{Z(t) - \frac{s^{(1)}(\beta_0,t)}{s^{(0)}(\beta_0,t)}\}^{\otimes 2} Y(t) e^{\beta_0^T Z(t)} \lambda^*\{t, t; Z(t)\}\big]\lambda_0(t)dt$ is a positive definite matrix.

(C6) $K(z)$ is a symmetric probability density function with mean 0 and bounded first derivative. In addition, $K(z)$ satisfies $\int_{-\infty}^{\infty} K(z)^2 dz < \infty$. Moreover, $nh \to \infty$ and $nh^3 \to 0$ as $n \to \infty$.

Conditions (C1) and (C2) are standard for the proportional hazards model. For $r < u$, the condition (C3) requires conditionally independent observation times in which the expectation of the counting process of measurement times is conditionally independent of the failure time given the observed covariates. This assumption is weaker than that in Cao et al. (2015). In (C3), the assumption of bounded $N(t)$ and $N^*(t)$ is also conventional. Conditions (C4) and (C5) guarantee finiteness and positive definiteness of the estimator's variance-covariance matrix. Condition (C6) indicates the restriction on the kernel and bandwidths. The following theorem, which is proved in the Appendix, states the asymptotic properties of $\hat{\beta}$ from $U_n(\beta)$ in (2.5).

B Appendix: Proofs of Theorems

This appendix includes the proofs of Theorem 1 and Corollary 1.

B.1 Proof of Theorem 1

Our main tools are empirical processes (van der Vaart & Wellner, 1996). First we need the following proposition:

Proposition 1 *Under (C1)-(C6), for any compact neighbourhood $\mathcal{B}$ of β_0, we have*

$$lim_{n \to \infty} sup_{0 \le t \le \tau, \beta \in \mathcal{B}} ||S_n^{(k)}(\beta, t) - s^{(k)}(\beta, t)|| = 0 \quad a.s. \quad for \quad k = 0, 1, 2. \tag{A8}$$

Proof: This follows from Theorem 37 of Pollard (1984) and the observation that $S_n^{(k)}(\beta, t)$

is Lipschitz continuous in $\beta \in \mathcal{B}$. $\square$

The key idea is to establish the following relationship

$$\begin{aligned}
& \sup_{|\beta-\beta_0|<M(nh)^{-1/2}} \left|(nh)^{1/2}U_n(\beta) - (nh)^{1/2}[U_n(\beta_0) - E\{U_n(\beta_0)\}] + (nh)^{1/2}A(\beta_0)(\beta - \beta_0)\right| \\
= \ & Dn^{1/2}h^{3/2} + o_p\{1 + (nh)^{1/2}|\beta - \beta_0|\}, \qquad \text{(A9)}
\end{aligned}$$

where $A(\beta_0)$ is given in Theorem 1 and D is a constant.

To obtain (A9), first, using $\mathcal{P}_n$ and $\mathcal{P}$ to denote the empirical measure and true probability measure respectively, we obtain

$$\begin{aligned}
(nh)^{1/2}U_n(\beta) &= (nh)^{1/2}(\mathcal{P}_n - \mathcal{P}) \int_0^\tau \int_0^\tau J(u,r)\{Z(r) - \frac{S_n^{(1)}(\beta,u)}{S_n^{(0)}(\beta,u)}\}dN^*(r)dN(u) \\
&+ (nh)^{1/2}E\left[\int_0^\tau \int_0^\tau J(u,r)\{Z(r) - \frac{S_n^{(1)}(\beta,u)}{S_n^{(0)}(\beta,u)}\}dN^*(r)dN(u)\right] \qquad \text{(A10)} \\
&= I + II, \qquad \text{(A11)}
\end{aligned}$$

where $J(u,r) = K_h(u-r)I\{r \le u, \int_r^u dN^*(t) = 0\}$, kernel weighting of the last observed covariate and failure time.

We now calculate the second term on the right-hand side of (A10). From Proposition 1, it follows that

$$\begin{aligned}
II &= (nh)^{1/2} \int_0^\tau \int_0^\tau \frac{1}{h}K(\frac{u-r}{h})E\Big[Z(r)Y(u)e^{\beta_0^T Z(u)}\lambda^*\{r,u,Z(r)\}\Big]\lambda_0(u)drdu \\
&- (nh)^{1/2} \int_0^\tau \int_0^\tau \frac{1}{h}K(\frac{u-r}{h})\{\bar{z}(\beta,u) + o(1)\}E\Big[Y(u)e^{\beta_0^T Z(u)}\lambda^*\{r,u,Z(r)\}\Big]\lambda_0(u)drdu,
\end{aligned}$$

where

$$\bar{z}(\beta, u) = \frac{s^{(1)}(\beta, u)}{s^{(0)}(\beta, u)}.$$

After change of variable and incorporating (C3) and (C6), we obtain

$$\begin{aligned} II &= (nh)^{1/2} \int_0^\tau E\Big[Y(r)Z(r)e^{\beta_0^T Z(r)} \lambda^*\{r, r; Z(r)\}\Big] \lambda_0(r) dr \\ &- (nh)^{1/2} \int_0^\tau \bar{z}(\beta, r) s^{(0)}(\beta_0, r) \lambda_0(r) dr + O(n^{1/2} h^{3/2}). \end{aligned} \tag{A12}$$

Following a Taylor expansion, we have

$$\begin{aligned} \bar{z}(\beta, r) &= \bar{z}(\beta_0, r) + \frac{\partial \bar{z}(\beta, r)}{\partial \beta}|_{\beta=\beta_0} (\beta - \beta_0) + o(|\beta - \beta_0|) \\ &= \bar{z}(\beta_0, r) + \{\frac{s^{(2)}(\beta_0, r)}{s^{(0)}(\beta_0, r)} - \frac{s^{(1)}(\beta_0, r)^{\otimes 2}}{s^{(0)}(\beta_0, r)^{\otimes 2}}\}(\beta - \beta_0) + o(|\beta - \beta_0|). \end{aligned}$$

Plug this into (A12), we obtain

$$\begin{aligned} II &= (nh)^{1/2} \int_0^\tau s^{(1)}(\beta_0, r) \lambda_0(r) dr - (nh)^{1/2} \int_0^\tau s^{(1)}(\beta_0, r) \lambda_0(r) dr + O(n^{1/2} h^{3/2}) \\ &- (nh)^{1/2} \int_0^\tau \{s^{(2)}(\beta_0, r) - \frac{s^{(1)}(\beta_0, r)^{\otimes 2}}{s^{(0)}(\beta_0, r)}\} \lambda_0(r) dr (\beta - \beta_0) + o\{(nh)^{1/2} |\beta - \beta_0|\} \\ &= -(nh)^{1/2} A(\beta_0)(\beta - \beta_0) + D n^{1/2} h^{3/2} + o\{(nh)^{1/2} |\beta - \beta_0|\}, \end{aligned} \tag{A13}$$

where D is a constant and

$$\begin{aligned} A(\beta_0) &= \int_0^\tau \{s^{(2)}(\beta_0, t) - \frac{s^{(1)}(\beta_0, t)^{\otimes 2}}{s^{(0)}(\beta_0, t)}\} \lambda_0(t) dt \\ &= \int_0^\tau E\Big[\{Z(t) - \frac{s^{(1)}(\beta_0, t)}{s^{(0)}(\beta_0, t)}\}\{Z(t) - \frac{s^{(1)}(\beta_0, t)}{s^{(0)}(\beta_0, t)}\}^T Y(t) e^{\beta_0^T Z(t)} \lambda^*\{t, t, Z(t)\}\Big] \lambda_0(t) dt. \end{aligned}$$

The matrix $A(\beta_0)$ is non-singular by assumption (C5). For the first term on the right-hand side of (A10), we consider the class of functions

$$\Big\{h^{1/2}\int_0^\tau\int_0^\tau J(u,r)\{Z(r)-\frac{S_n^{(1)}(\beta,u)}{S_n^{(0)}(\beta,u)}\}dN^*(r)dN(u):|\beta-\beta_0|<\epsilon\Big\}$$

for a given constant ϵ. Note that the functions in this class are Lipschitz continuous in β and the Lipschitz constant is uniformly bounded by

$$M_1\int_0^\tau\int_0^\tau h^{1/2}K_h(u-r)dN^*(r)dN(u),$$

which has finite second moment and M_1 is the upper bound of $\frac{s^{(2)}(\beta,t)}{s^{(0)}(\beta,t)}-\{\frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)}\}^{\otimes 2}$. Therefore, this class is P-Donsker class by the Jain-Marcus theorem (van der Vaart & Wellner, 1996). As the result, we obtain that the first term in the right-hand side of (A10) for $|\beta-\beta_0|<M(nh)^{-1/2}$ is equal to

$$\begin{aligned}&(nh)^{1/2}(\mathcal{P}_n-\mathcal{P})\int_0^\tau\int_0^\tau J(u,r)\{Z(r)-\frac{S_n^{(1)}(\beta_0,u)}{S_n^{(0)}(\beta_0,u)}\}dN^*(r)N(u)+o_p(1)\\=\;&(nh)^{1/2}\Big[U_n(\beta_0)-E\{U_n(\beta_0)\}\Big]+o_p(1).\end{aligned}\tag{A14}$$

Combining (A10), (A13) and (A14), we obtain (A9). Consequently,

$$\begin{aligned}&(nh)^{1/2}A(\beta_0)(\hat\beta-\beta_0)+O_p(n^{1/2}h^{3/2})+o_p\{1+(nh)^{1/2}|\hat\beta-\beta_0|\}\\=\;&(nh)^{1/2}[U_n(\beta_0)-E\{U_n(\beta_0)\}].\end{aligned}\tag{A15}$$

On the other hand, as the subjects are independent identically distributed, we calculate

$$
\begin{aligned}
& \mathrm{Var}\{(nh)^{1/2}U_n(\beta_0)\} = h\mathrm{Var}\Big\{\int_0^\tau \int_0^\tau J(u,r)\{Z(r) - \bar{Z}(\beta_0,u)\}dN^*(r)dN(u)\Big\} \\
= \ & hE \iint_0^\tau \iint_0^\tau J(u_1,r_1)J(u_2,r_2)\{Z(r_1) - \bar{Z}(\beta_0,u_1)\}\{Z(r_2) - \bar{Z}(\beta_0,u_2)\} \\
& dN^*(r_1)dN^*(r_2)dN(u_1)dN(u_2) \\
- \ & h\Big(E\int_0^\tau \int_0^\tau J(u,r)\{Z(r) - \bar{Z}(\beta_0,u)\}dN^*(r)dN(u)\Big)^2 \\
= \ & A - B
\end{aligned}
$$

We next show that $B = o(h)$. By Proposition 1, we have

$$
\begin{aligned}
& E\int_0^\tau \int_0^\tau J(u,r)\{Z(r) - \bar{Z}(\beta_0,u)\}dN^*(r)dN(u) \\
= \ & E\int_0^\tau \int_0^\tau J(u,r)\{Z(r) - \bar{z}(\beta_0,u)\}dN^*(r)dN(u) + o(1),
\end{aligned}
$$

where

$$
\bar{z}(\beta_0,u) = \frac{s^{(1)}(\beta_0,u)}{s^{(0)}(\beta_0,u)}
$$

and

$$
s^{(k)}(\beta,t) = E\Big[Y(t)Z(t)^{\otimes k}e^{\beta^T Z(t)}\lambda^*\{t,t;Z(t)\}\Big].
$$

Taking conditional expectation, we have

$$
\begin{aligned}
& E\int_0^\tau \int_0^\tau J(u,r)\{Z(r) - \bar{z}(\beta_0,u)\}dN^*(r)dN(u) \\
= \ & E\int_0^\tau \int_0^\tau Z(r)Y(u)e^{\beta_0^T Z(u)}\lambda_0(u)\lambda^*\{r,u,Z(r)\}drdu \\
- \ & E\int_0^\tau \int_0^\tau \frac{s^{(1)}(\beta_0,u)}{s^{(0)}(\beta_0,u)}Y(u)e^{\beta_0^T Z(u)}\lambda_0(u)\lambda^*\{r,u,Z(r)\}drdu \\
= \ & I_1 - I_2
\end{aligned}
$$

After a change of variable and Taylor expansion, we obtain

$$\begin{aligned} I_1 &= E\int_0^\tau \int_{-\infty}^{+\infty} K(z)dzZ(r)Y(r)e^{\beta_0^T Z(r)}\lambda_0(r)\lambda^*\{r,r,Z(r)\}dr + o(1) \\ &= \int_0^\tau s^{(1)}(\beta_0,r)\lambda_0(r)dr + o(1). \end{aligned}$$

Similarly,

$$\begin{aligned} I_2 &= E\int_0^\tau \int_{-\infty}^{+\infty} K(z)dz\frac{s^{(1)}(\beta_0,r)}{s^{(0)}(\beta_0,r)}Y(r)e^{\beta_0^T Z(r)}\lambda_0(r)\lambda^*\{r,r,Z(r)\}dr + o(1) \\ &= \int_0^\tau s^{(1)}(\beta_0,r)\lambda_0(r)dr + o(1). \end{aligned}$$

Consequently, $B = o(h)$. Now we decompose A into four parts

$$\begin{aligned} A &= hE\int_{u_1\neq u_2}\int_{r_1\neq r_2} J(u_1,r_1)J(u_2,r_2)\{Z(r_1)-\bar{Z}(\beta_0,u_1)\}\{Z(r_2)-\bar{Z}(\beta_0,u_2)\} \\ &\quad dN^*(r_1)dN^*(r_2)dN(u_1)dN(u_2) \\ &+ hE\int_{u_1\neq u_2}\int_0^\tau J(u_1,r)J(u_2,r)\{Z(r)-\bar{Z}(\beta_0,u_1)\}\{Z(r)-\bar{Z}(\beta_0,u_2)\} \\ &\quad dN^*(r)dN(u_1)dN(u_2) \\ &+ hE\int_0^\tau\int_{r_1\neq r_2} J(u,r_1)J(u,r_2)\{Z(r_1)-\bar{Z}(\beta_0,u)\}\{Z(r_2)-\bar{Z}(\beta_0,u)\} \\ &\quad dN^*(r_1)dN^*(r_2)dN(u) \\ &+ hE\int_0^\tau\int_0^\tau J(u,r)^2\{Z(r)-\bar{Z}(\beta_0,u)\}^{\otimes 2}dN^*(r)dN(u) \\ &= A_1 + A_2 + A_3 + A_4. \end{aligned}$$

It is easy to see that $A_1 = O(h), A_2 = O(h), A_3 = O(h)$. After change of variables and Taylor expansion, we obtain

$$A_4 = \int_0^\infty K(x)^2dx\int_0^\tau \{s^{(2)}(\beta_0,t) - \frac{s^{(1)}(\beta_0,t)^{\otimes 2}}{s^{(0)}(\beta_0,t)}\}\lambda_0(t)dt + O(h) + O\{(nh)^{-1}\}.$$

Therefore

$$\mathrm{Var}\{(nh)^{1/2}U_n(\beta_0)\} \to \Sigma(\beta_0),$$

where

$$\Sigma(\beta_0) = \int_0^\infty K(x)^2 dx \int_0^\tau \{s^{(2)}(\beta_0, t) - -\frac{s^{(1)}(\beta_0, t)^{\otimes 2}}{s^{(0)}(\beta_0, t)}\}\lambda_0(t)dt.$$

To prove the asymptotic normality, we verify that the Lyapunov condition holds. Define

$$\psi_i = (nh)^{1/2} n^{-1} \int_0^\tau \int_0^\tau J(u, r)\{Z_i(r) - \bar{Z}(\beta_0, u)\}dN_i^*(r)dN_i(u).$$

Similar to the calculation of $\Sigma(\beta_0)$,

$$\sum_{i=1}^n E\Big(|\psi_i - E\psi_i|^3\Big) = nO\{(nh)^{3/2}n^{-3}h^{-2}\} = O\{(nh)^{-1/2}\}.$$

Thus,

$$(nh)^{1/2}\Big[U_n(\beta_0) - E\{U_n(\beta_0)\}\Big] \to N\{0, \Sigma(\beta_0)\}.$$

Combining with (A15), we finish the proof of Theorem 1. □

B.2 Proof of Corollary 1

To begin with, we have

$$-\frac{\partial U_n(\beta)}{\partial \beta} = n^{-1}\sum_{i=1}^n \int_0^\tau \int_0^\tau J(u, r)dN_i^*(r)\Big\{\frac{S_n^{(2)}(\beta, u)}{S_n^{(0)}(\beta, u)} - \frac{S_n^{(1)}(\beta, u)^{\otimes 2}}{S_n^{(0)}(\beta, u)^2}\Big\}dN_i(u).$$

Using a similar argument as for equation (A14), we show

$$\Big\{\int_0^\tau \int_0^\tau J(u, r)dN^*(r)\Big\{\frac{S_n^{(2)}(\beta, u)}{S_n^{(0)}(\beta, u)} - \frac{S_n^{(1)}(\beta, u)^{\otimes 2}}{S_n^{(0)}(\beta, u)^2}\Big\}dN(u) : |\beta - \beta_0| < \epsilon\Big\}$$

is a P-Glivenko-Cantelli class. Therefore, $\sup_{|\beta-\beta_0|<\epsilon}|\frac{\partial U_n(\beta)}{\partial\beta}|_{\beta=\hat{\beta}} - E\{\frac{\partial U_n(\beta)}{\partial\beta}|_{\beta=\hat{\beta}}\}| \to 0$ in probability. Since $\hat{\beta}$ is consistent for β_0, by continuous mapping theorem, $\frac{\partial U_n(\beta)}{\partial\beta}|_{\beta=\hat{\beta}}$ converges in probability to $-A(\beta_0)$. Similarly, let

$$\hat{\Sigma}(\beta) = n^{-2}\sum_{i=1}^{n}\Big[\int_0^\tau\int_0^\tau K(u,r)\{Z_i(r)-\bar{Z}(\beta,u)\}dN_i^*(r)dN_i(u)\Big]^{\otimes 2},$$

then $\sup_{|\beta-\beta_0|<\epsilon}|\hat{\Sigma}(\beta) - E\{\hat{\Sigma}(\beta)\}| \to 0$ in probability. On the other hand,

$$E\{\hat{\Sigma}(\beta)\} = n^{-1}E\Big[\int_0^\tau\int_0^\tau K(u,r)\{Z_i(r)-\bar{Z}(\beta,u)\}dN_i^*(r)dN_i(u)\Big]^{\otimes 2}.$$

After change of variables, and by (C3),

$$E\{\hat{\Sigma}(\beta)\} \quad = \quad \frac{1}{nh}\int_0^\infty K(z)^2dz\int_0^\tau\{s^{(2)}(\beta_0,u) - \frac{s^{(1)}(\beta_0,u)^{\otimes 2}}{s^{(0)}(\beta_0,u)}\}du.$$

Therefore,

$$(nh)\hat{\Sigma} \xrightarrow{p} \Sigma(\beta_0) \quad \text{as} \quad nh\to\infty.$$

The consistency of variance estimate follows.

References

Andersen, P. and Liestol, K. (2003), Attenuation caused by infrequently updated covariates in survival analysis. *Biostatistics* **4**, 633-649.

Bienias, J., Beckett, L. A., Bennett, D. A., Wilson, R. S. and Evans, D. A. (2003). Design of the Chicago health and aging project. *J. Alzheimers Dis.* **5**, 349–355.

Cao, H., Churpek, M., Zeng, D. and Fine, J. P. (2015). Analysis of the proportional hazards model with sparse longitudinal covariates. *J. Amer. Statist. Assoc.* **110**, 1187–1196.

Cox, D. R. (1972). Regression models and life-tables. *J. R. Stat. Soc. Ser. B,* **34**, 187–220.

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

Heitjan, D. and Rubin, D. (1991). Ignorability and coarse data. *Ann. Statist.* **19**, 2244–2253.

Henderson, R., Diggle, P. and Dobson, P. (2000). Joint modeling of longitudinal measurements and event time data. *Biostatistics* **1**, 465–480.

Liu, Y. and Craig, B. A. (2006). Incorporating time-dependent covariates in survival analysis using the LVAR method. *Statist. Med.* **25**, 1729–1740.

Molenberghs, G., Williams, P. L. and Lipsitz, S. R. (2002). Prediction of survival and opportunities infections in HIV-infected patients: a comparison of imputation methods of incomplete CD4 counts. *Statist. Med.* **21**, 1387–1408.

Molnar, F.J., Hutton, B. and Fergusson, D. (2008). Does analysis using "last observation carried forward" introduce bias in dementia research? *CMAJ.* **179**, 751–753.

Pollard, D. (1984). *Convergence of stochastic processes.* Springer, New York.

Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: with applications in R.* Chapman and Hall/CRC Biostatistics Series.

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., De Gruttola, 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. *J. Amer. Statist. Assoc.* **90**, 27–37.

van der Vaart A. and Wellner, J. (1996). *Weak convergence and empirical processes.* Springer, New York.

Xu, J. and Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics* **50**, 375–387.

Yoon, H.-J., Kim, K., Nam, Y.-S., Yun, J.-M and Park, M (2016) Mean corpuscular volume levels and all-cause and liver cancer mortality. *Clin. Chem. Lab. Med.* 1434–6621.