Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://amstat.tandfonline.com/loi/uasa20

Analysis of the Proportional Hazards Model With Sparse Longitudinal Covariates

Hongyuan Cao, Mathew M. Churpek, Donglin Zeng & Jason P. Fine

To cite this article: Hongyuan Cao, Mathew M. Churpek, Donglin Zeng & Jason P. Fine (2015) Analysis of the Proportional Hazards Model With Sparse Longitudinal Covariates, Journal of the American Statistical Association, 110:511, 1187-1196, DOI: 10.1080/01621459.2014.957289

To link to this article: http://dx.doi.org/10.1080/01621459.2014.957289

Analysis of the Proportional Hazards Model With Sparse Longitudinal Covariates

Hongyuan CAO, Mathew M. CHURPEK, Donglin ZENG, and Jason P. FINE

Regression analysis of censored failure observations via the proportional hazards model permits time-varying covariates that are observed at death times. In practice, such longitudinal covariates are typically sparse and only measured at infrequent and irregularly spaced follow-up times. Full likelihood analyses of joint models for longitudinal and survival data impose stringent modeling assumptions that are difficult to verify in practice and that are complicated both inferentially and computationally. In this article, a simple kernel weighted score function is proposed with minimal assumptions. Two scenarios are considered: half kernel estimation in which observation ceases at the time of the event and full kernel estimation for data where observation may continue after the event, as with recurrent events data. It is established that these estimators are consistent and asymptotically normal. However, they converge at rates that are slower than the parametric rates that may be achieved with fully observed covariates, with the full kernel method achieving an optimal convergence rate that is superior to that of the half kernel method. Simulation results demonstrate that the large sample approximations are adequate for practical use and may yield improved performance relative to last value carried forward approach and joint modeling method. The analysis of the data from a cardiac arrest study demonstrates the utility of the proposed methods. Supplementary materials for this article are available online.

KEY WORDS: Convergence rates; Cox model; Kernel weighted estimation.

1. INTRODUCTION

In biomedical and public health research, it is common to observe both longitudinal data, with repeated measurements of a variable at a number of time points, and event history data, in which times to recurrent or terminating events are recorded. In such studies, investigators may be interested in evaluating the effects of longitudinal covariates on the occurrence of events. The usual proportional hazards analysis may not be applicable when the time-dependent covariates are measured intermittently.

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. For single event data, the Cox 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 baseline hazard function and β_0 is a vector of unknown regression parameters. With recurrent event data, there may be multiple jumps in $N(t)$ and model (1.1) refers to the Andersen and Gill (1982) proportional intensity model. The standard partial likelihood analysis of the model (1.1) requires the full trajectory of the covariates. Similar issues arise with recurrent events when relaxing the intensity assumption (1.1) to the proportional rate model

$$E\{dN(t)|Z(t)\} = e^{\beta_0^T Z(t)} d\mu_0(t), \tag{1.2}$$

where $\mu_0(\cdot)$ is an unspecified function and β_0 is a vector of unknown regression parameters. The estimation procedures for models (1.1) and (1.2) require knowledge of $Z(t)$ at those event times where a subject is still under observation.

The simplest method for handling incompletely observed longitudinal covariates in the above models is to naïvely impute missing values using the last value carried forward approach. The missing values of $Z(r)$ may be replaced by the most recent observed values of $Z(u)$, $u \leq r$. This approach may be generalized to permit additional usage of lagged covariates, as discussed in Andersen and Liestol (2003). While these ad hoc imputation approaches are conceptually simple and may be implemented using standard software, they lack rigorous theoretical justification and may incur substantial bias. An alternative to these naïve techniques is to jointly model the longitudinal covariates and the event history data (Ibrahim, Chu, and Chen 2010). There has been considerable interest in modeling the dependence between these two processes via shared random effects (Hogan and Laird 1997). Under such assumptions, the joint distribution of $N(\cdot)$ and $Z(\cdot)$ may be fully specified (Degruttola and Tu 1994; Faucett and Thomas 1996; Henderson, Diggle, and Dobson 2000; Xu and Zeger 2001). To obtain more flexible modeling, Yao (2007) adopted a nonparametric functional principal component approach to model the longitudinal process and Cox model for the time-to-event outcome. The modeling assumptions are rather strong and the computation and inference are complicated, requiring full nonparametric maximum likelihood (Tsiatis, Degruttola, and Wulfsohn 1995; Wulfsohn and Tsiatis 1997; Zeng and Cai 2005; Dupuy, Grama, and Mesbah

Hongyuan Cao is Assistant Professor, Department of Statistics, University of Missouri-Columbia, Columbia, MO 65201 (E-mail: *caohong@missouri.edu*). Mathew M. Churpek is Assistant Professor, Department of Health Studies and Department of Medicine, University of Chicago, Chicago, IL 60637 (E-mail: *matthew.churpek@uchospitals.edu*). Donglin Zeng is Professor (E-mail: *dzeng@email.unc.edu*) and Jason P. Fine is Professor (E-mail: *jfine@email.unc.edu*), Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514. Donglin Zeng is partially supported by NIH grant R01CA082659 and P01 CA142538; Jason P. Fine is partially supported by NIH grant 5R01CA09483-07 and P01 CA142538. We thank Zhuang Ma for help with simulation and Trevor C Yuen for help with data preparation.

Journal of the American Statistical Association
September 2015, Vol. 110, No. 511, Theory and Methods
DOI: 10.1080/01621459.2014.957289

2006) or likelihood-motivated procedures, like the conditional score approach in Tsiatis and Davidian (2001). The theoretical justification depends critically on correct model specification, which may involve assumptions that are unverifiable from the observed data. A comprehensive review of the joint modeling approach is given in Tsiatis and Davidian (2004) and Rizopoulos (2012).

In this article, we develop simple, computationally efficient, and theoretically justified estimators for models (1.1) and (1.2) using intermittently collected longitudinal covariates that require minimal assumptions on the joint distribution of $N(\cdot)$ and $Z(\cdot)$. The main idea is to modify the naïve imputation approaches like those in Andersen and Liestol (2003) to obtain theoretically justified estimation procedures that are valid under weak assumptions. A kernel weighting scheme is used to downweight imputed covariate values in the partial likelihood, where those observations that are distant in time from the event time receive less weight. Such kernel weighting approach has been adopted by Cai and Sun (2003) and Tian, Zucker, and Wei (2005) for time-dependent coefficient in Cox model. However, there are fundamental differences between our work and the time-varying coefficient methodology. The estimation of time-varying regression parameters assumes that the covariate effect varies with respect to time while we assume that the covariate is a dynamic process with fixed coefficient. The smoothing methods employed by Cai and Sun (2003) and Tian, Zucker, and Wei (2005), which localize the partial likelihood in time are not applicable in our setting, where smoothing occurs at the individual level, as opposed to the population level, where the same weights are applied to all individuals. The dependence structure between the longitudinal measurements and the event history process is otherwise unspecified, in contrast to the joint models. With a suitable choice of the bandwidth, the estimators for the regression coefficients are consistent and asymptotically normal, with simple plug-in variance estimators. Interestingly, the optimal rates of convergence for β_0 are slower than the usual parametric rate with time-invariant covariates. For recurrent events data, one may include both forward and backward lagged covariates, employing covariate information observed after event times. Our theoretical results demonstrate that using all available covariate information yields an estimator that converges at $n^{2/5}$, while the estimator that only uses backward lagged covariates converges more slowly than $n^{2/5}$. These results are detailed in Sections 2 and 3.

In Section 2, we propose an estimation for the Cox model for single event data using half kernel smoothing with backward lagged covariates and present the corresponding theoretical findings. The results for full kernel smoothing including both forward and backward lagged covariates are given in Section 3. We report the results of our simulation studies in Section 4, exhibiting improved performance versus the last value carried forward approach and joint modeling method. The joint modeling approach exhibits efficiency gains when the joint model is correctly specified but may exhibit substantial bias and poor coverage under model misspecification. We then apply our method to data from a cardiac arrest study in Section 5. In this analysis, the joint modeling approach has convergence issues, with the corresponding results being somewhat unreliable. Concluding remarks are given in Section 6. Proofs of the results from Sections 2 and 3 are given in the Appendix provided in an online supplementary materials file.

2. HALF KERNEL ESTIMATION WITH BACKWARD LAGGED COVARIATES

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 $Z(\cdot)$ (Gill, van der Laan, and Robin 1997). 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 $R_{ik} \le X_i, k = 1, \ldots, M_i$, where $X_i = \min(T_i, C_i)$, and M_i is assumed finite with probability one such that the observed covariates are sparse. 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 timing of the measurements $R_{ik}, k = 1, \ldots, M_i$ is assumed exogenous in the sense that the decision to schedule a measurement is made independently of the measurement. 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 Δ_i equals 1 if $X_i = T_i$ and 0 otherwise.

Following Andersen and Liestol (2003), one may use backward lagged covariates in imputing missing covariates in the partial likelihood for estimation of β_0 in model (1.1). To ease the presentation, we adopt the counting process notation, where $N_i(t) = I(X_i \le t, \Delta_i = 1)$ and $Y_i(t) = I(X_i \ge t)$. If the covariate $Z_i(t)$ were fully observed for all $t < X_i$, then one might construct the following partial likelihood

$$L_n(\beta) = \prod_{i=1}^{n} \prod_{t \ge 0} \left\{ \frac{e^{\beta^T Z_i(t)}}{\sum_{j=1}^{n} Y_j(t) e^{\beta^T Z_j(t)}} \right\}^{\Delta N_i(t)}, \tag{2.3}$$

where

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

The log partial likelihood is

$$l_n(\beta) = n^{-1} \log L_n(\beta) = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \left[\beta^T Z_i(u) - \log\left\{ \sum_{j=1}^{n} Y_j(u) e^{\beta^T Z_j(u)} \right\} \right] dN_i(u),$$

where τ is a prespecified time point such that $\mathrm{pr}(X_i > \tau) > 0$. Because $Z_i(u), i = 1, \ldots, n$, are not observed continuously, $l_n(\beta)$ is not computable from the observed data. We propose using lagged covariate values with observation times smaller than u, where we downweight covariates at times distant from u in $dN_i(u)$ and $Y_i(u)$. This approach formalizes the lagging strategy in Andersen and Liestol (2003), with the kernel weighting enabling the use of all available lagged covariates. If the covariate observation times for a subject are all far away from the event time, then this subject may be disregarded in the calculation of partial likelihood but not in the last value carried forward approach, where the most recent observed covariate is used. With irregularly observed longitudinal covariates, as the number of subjects increases, borrowing strength across subjects, the longitudinal measurements times will become dense. To exploit this accumulation of information, one may "smooth" an individual's contributions to the partial likelihood based on the distance of their observed covariates to the time of interest. The resulting log partial likelihood up to time t is

$$l_n^*(\beta,t)=n^{-1}\sum_{i=1}^{n}\int_0^t\Bigg[\sum_{k=1}^{M_i}K_{h_n}(u-R_{ik})I(R_{ik}\le u)\times\Bigg\{\beta^T Z_i(R_{ik})-\log\sum_{j=1}^{n}\sum_{l=1}^{M_j}K_{h_n}(u-R_{jl})\times I(R_{jl}\le u)Y_j(u)e^{\beta^T Z_j(R_{jl})}\Bigg\}\Bigg]dN_i(u), \qquad (2.4)$$

where $K_{h_n}(t)=K(t/h_n)/h_n$, h_n is the bandwidth and the kernel function $K(t)$ is a symmetric probability density with support $[-1,1]$, mean 0, and bounded first derivative.

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

$$U_n(\beta)=n^{-1}\sum_{i=1}^{n}\int_0^\tau\Bigg[\int K_{h_n}(u-r)I(r\le u)\{Z_i(r)-\bar{Z}(\beta,u)\}dN_i^*(r)\Bigg]dN_i(u), \qquad (2.5)$$

$$\bar{Z}(\beta,t)=\frac{S_n^{(1)}(\beta,t)}{S_n^{(0)}(\beta,t)},$$

$$S_n^{(k)}(\beta,t)=n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{M_i}K_{h_n}(t-R_{ij})I(R_{ij}\le t)Y_i(t)Z_i(R_{ij})^{\otimes k}e^{\beta^T Z_i(R_{ij})},\quad k=0,1,2,$$

$a^{\otimes 0}=1$, $a^{\otimes 1}=a$, $a^{\otimes 2}=aa^T$. For $i=1,\ldots,n$, $N_i^*(t)=\sum_{k=1}^{M_i}I(R_{ik}\le t)$ is a realization of $N^*(t)$, the counting process for the covariate observation times. Observe that the smoothing leads to different weights for different individuals inside the integral in $U_n(\beta)$, which differs from smoothing methods for proportional hazards model with time-dependent regression parameters (Cai and Sun 2003; Tian, Zucker, and Wei 2005), where the same weights are applied to all individuals inside the integral. Since (2.4) is concave in β, there exists a unique root of the estimating function (2.5).

In addition, $E\{dN^*(t)\}=\lambda^*(t)dt$, and $E[dN(t)|\mathcal{F}_s, s\le t]=Y(t)e^{\beta_0^T Z(t)}\lambda_0(t)dt$, where $\lambda^*(t)$ and $\lambda_0(t)$ are twice continuous differentiable and strictly positive for $t\in[0,\tau]$, and $\mathcal{F}_t$ is the filtration, which includes all information in $N(s)$, $Y(s)$, and $Z(s)$ up to time t, as well as the measurement times.

To state our key results, additional notation and regularity conditions are needed. Denote

$$u_1(\beta)\equiv\int_0^\tau\lambda_0(u)\{s^{(1)}(\beta_0,u)-\bar{z}(\beta,u)s^{(0)}(\beta_0,u)\}du,$$

where $s^{(k)}(\beta,t)$ is the limit of $S_n^{(k)}(\beta,t)$, $k=0,1,2$. That is,

$$s^{(k)}(\beta,t)=2^{-1}E\{Y(t)Z(t)^{\otimes k}e^{\beta^T Z(t)}\}\lambda^*(t),\quad\text{and}$$

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

Let

$$v_1(\beta)\equiv 2^{-1}\int_0^\tau\lambda^*(u)\lambda_0(u)\frac{s^{(0)}(\beta_0,u)}{s^{(0)}(\beta,u)}*E\Bigg[Y(u)e^{\beta^T Z(u)}\Bigg\{Z(u)-\frac{s^{(1)}(\beta,u)}{s^{(0)}(\beta,u)}\Bigg\}^{\otimes 2}\Bigg]du$$

and let $\mathcal{B}$ be a compact set of R^p that includes a neighborhood of β_0. Assume that the following conditions hold:

(A1) The covariate process $Z(t)$ is left continuous, has right-hand limit, and is contained in a bounded subset with total variation bounded by a constant $c<\infty$ almost surely. Moreover, $E[\{Z(v)-\bar{Z}(\beta_0,u)\}Y(u)e^{\beta_0^T Z(v)}]$ is twice continuously differentiable for $u, v\in[0,\tau]^{\otimes 2}$.

(A2) $N^*(t)$ is independent of $N(t)$ and $Z(t)$. In addition, $\int_0^\tau\lambda_0(t)dt<\infty$ and $N^*(\tau)$ is bounded by a finite constant.

(A3) There exists a neighborhood $\mathcal{B}$ of β_0 such that

$$E\{\sup_{t\in[0,\tau],\beta\in\mathcal{B}}Y(t)||Z(t)Z(t)^T||e^{\beta^T Z(t)}\}<\infty. \qquad (2.6)$$

(A4) $v_1(\beta_0)$ is nonsingular.

(A5) $K(z)$ is a symmetric density function satisfying $\int K(z)^2dz<\infty$, $h_n\to 0$, $nh_n^3\to 0$, and $nh_n\to\infty$.

Condition (A1) posits a certain level of smoothness on $Z(t)$. It is worth emphasizing that joint modeling strategies (Tsiatis and Davidian 2004) generally imply at least this level of smoothness, if not stronger. The condition (A2) requires that the covariate observation process is independent of the covariates and the event history process. This is somewhat stronger than the missing at random assumption under which a valid likelihood might be constructed with joint modeling assumptions. Condition (A3) places mild restrictions on the variability of $Z(t)$, which would typically be satisfied in practice, and guarantees that $\hat{\beta}_n$ has finite variance in large samples. Condition (A4) ensures that the variance-covariance matrix is positive definite. Condition (A5) states the restrictions on the kernel bandwidths.

The following theorem, which is established in the Appendix, states the asymptotic properties of $\hat{\beta}_n$ based on solving (2.5) with kernel bandwidth selected to yield consistent estimation:

Theorem 1. Under conditions (A1)–(A5), $\hat{\beta}_n$ is consistent and the asymptotic distribution of $\hat{\beta}_n$ satisfies

$$\sqrt{nh_n}(\hat{\beta}_n-\beta_0)\to N\Big[0, W(\beta_0)^{-1}\Sigma(\beta_0)\{W(\beta_0)^{-1}\}^T\Big], \qquad (2.7)$$

where

$$W(\beta_0)=\int_0^\tau\Bigg[s^{(2)}(\beta_0,u)-\frac{s^{(1)}(\beta_0,u)^{\otimes 2}}{s^{(0)}(\beta_0,u)}\Bigg]\lambda_0(u)du,$$

and

$$\Sigma(\beta_0)=\int_0^\infty K(z)^2dz\int_0^\tau\Bigg[s^{(2)}(\beta_0,u)-\frac{s^{(1)}(\beta_0,u)^{\otimes 2}}{s^{(0)}(\beta_0,u)}\Bigg]\lambda_0(u)du.$$

For statistical inference, it is challenging to estimate the variance in (2.7) directly, owing to time-varying quantities that depend on unknown values of $Z(\cdot)$, which are not available in the intermittent longitudinal covariate observations. In practice, we employ estimating Equation (2.5) to estimate

$\Sigma(\beta_0)$ by $\hat{\Sigma} = n^{-2}\sum_{i=1}^{n}[\int_0^\tau \int_0^\infty K_{h_n}(u-r)I(r \le u)\{Z_i(r) - \bar{Z}(\hat{\beta}_n, u)\}dN_i^*(r)dN_i(u)]^{\otimes 2}$ and estimate the variance of $\hat{\beta}_n$ by the sandwich formula

$$\left(\frac{\partial U_n(\beta)}{\partial \beta}\bigg|_{\beta=\hat{\beta}_n}\right)^{-1}\hat{\Sigma}\left\{\left(\frac{\partial U_n(\beta)}{\partial \beta}\bigg|_{\beta=\hat{\beta}_n}\right)^{-1}\right\}^T.$$

Corollary 1. Under conditions (A1)–(A5), the sandwich formula consistently estimates the variance of $\hat{\beta}_n$.

Our method depends on an appropriate choice of bandwidth. Theoretically speaking, condition (A5) says that the bandwidth cannot be too small; otherwise, the variance will be quite large. On the other hand, to eliminate the asymptotic bias, one requires a small bandwidth. With $h_n = o(n^{-1/3})$, we achieve an optimal rate of convergence $o(n^{1/3})$. This result provides insight into the ad hoc procedure of Andersen and Liestol (2003). When tuning the partial likelihood estimation using backward lagged covariates to obtain a theoretically rigorous estimator, parametric convergence rates are not achievable. This contrasts with joint modeling approaches (Tsiatis and Davidian 2004), where strong modeling assumptions on the joint distribution of the covariate process and event times facilitate likelihood-based inferences, which may achieve parametric rates of convergence for the regression parameter β.

Numerical studies reported in Section 4 show that small bias may be achieved for bandwidths between $n^{-0.9}$ and $n^{-0.4}$, with stable variance estimation and confidence interval coverage for bandwidths larger than $n^{-0.7}$. Within this range, the bias diminishes as the sample size increases, as predicted by Theorem 1. In Section 4, an automatic bandwidth selection procedure is proposed, with both the corresponding model-based variance estimators and confidence intervals exhibiting good performance.

3. FULL KERNEL ESTIMATION WITH FORWARD AND BACKWARD LAGGED COVARIATES

If data continue to be collected on subjects for whom an event has occurred, as in the recurrent events case, we may use full kernel to impute missing values using both forward and backward lagged covariates. Andersen and Liestol (2003) investigated scenarios where observation terminates at the time of the first event, as with classical right censored data, hence, did not consider the use of forward lagged covariates. Let $N_i(t)$ be a recurrent event counting process and let $Y_i(t) = I(C_i \ge t)$ be the at risk process for subject i up to time C_i, $i = 1, \ldots, n$. Similarly to half kernel estimation, one may construct a smoothed partial likelihood score function using full kernel smoothing:

$$\tilde{U}_n(\beta, t) = n^{-1}\sum_{i=1}^{n}\int_0^t\left[\sum_{k=1}^{M_i} K_{h_n}(u - R_{ik})\{Z_i(R_{ik}) - \tilde{Z}(\beta, u)\}\right]dN_i(u), \quad (3.8)$$

where $\tilde{Z}(\beta, t) = \tilde{S}_n^{(1)}(\beta, t)/\tilde{S}_n^{(0)}(\beta, t)$ and

$$\tilde{S}_n^{(k)}(\beta, t) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{M_i} K_{h_n}(t - R_{ij})Y_i(t)Z_i(R_{ij})^{\otimes k}e^{\beta^T Z_i(R_{ij})},$$
$$k = 0, 1, 2.$$

One should recognize that the smoothed estimating function (3.8) is valid under both model (1.1) and under the weaker proportional rate model (1.2). The results provided below hold under the weaker assumption (1.2).

To state the main asymptotic findings for full kernel estimator $\tilde{\beta}_n$ solving (3.8), some additional notations and conditions are needed. For $i = 1, \ldots, n$, N_i, Y_i, and C_i are independent realizations of random variables N, Y, and C. Denote

$$\tilde{s}^{(k)}(\beta, t) = E[Y(t)Z(t)^{\otimes k}e^{\beta^T Z(t)}]\lambda^*(t), \quad (3.9)$$

where the definition of $\lambda^*(t)$ is given in Section 2. We assume that the following conditions are satisfied:

- (C1) $\{N_i(\cdot), Y_i(\cdot), Z_i(\cdot)\}(i = 1, \ldots, n)$ are independent and identically distributed.
- (C2) $\mathrm{pr}(C \ge \tau) > 0$, where τ is a predetermined constant.
- (C3) $N(\tau)$ and $N^*(\tau)$ are bounded by finite constants and $\mu_0(t)$ and $\lambda^*(t)$ are twice continuously differentiable.
- (C4) For $i = 1, \ldots, n$, Z_i have bounded total variation, where $|Z_{ij}(0)| + \int_0^\tau |dZ_{ij}(t)| \le K$ for all $j = 1, \ldots, p$, where Z_{ij} is the jth component of Z_i and K is a constant. In addition, $E\{Z(s)Y(t)e^{\beta_0^T Z(t)}\}$ is twice continuously differentiable for $s, t \in [0, \tau]^{\otimes 2}$.
- (C5) $A(\beta_0) \equiv \int_0^\tau E[\{Z(t) - \frac{\tilde{s}^{(1)}(\beta_0, t)}{\tilde{s}^{(0)}(\beta_0, t)}\}^{\otimes 2}Y(t)e^{\beta_0^T Z(t)}]\mu_0(t)dt$ is positive definite.
- (C6) $K(z)$ is a symmetric density function satisfying $\int_{-\infty}^{\infty} K(z)^2 dz < \infty$. In addition, $h_n \to 0$ and $nh_n \to \infty$.
- (C7) $nh_n^5 \to 0$.

Conditions (C1)–(C6) are similar in spirit to those for half kernel estimation in Section 2. Conditions (C1) and (C2) are standard for the proportional rate model (1.2). The assumption of bounded $N(t)$ in (C3) is also conventional with recurrent events over finite time intervals. Conditions (C4) and (C5) are full kernel analogs for model (1.2) of half kernel conditions (A3) and (A4) for model (1.1). These guarantee finiteness and positive definiteness of the full kernel estimator's variance-covariance matrix. The kernel requirements in (C6) are similar to those in (A5), with (C6) and (C7) indicating the allowable range of bandwidth for full kernel estimation is larger than that for half kernel estimation given in (A5). The implications of this weaker bandwidth requirement are discussed below.

The asymptotic properties of the full kernel estimator $\tilde{\beta}_n$ are detailed in the following theorem:

Theorem 2. Under conditions (C1)–(C6), the asymptotic distribution of $\tilde{\beta}_n$ satisfies

$$(nh_n)^{1/2}\{A(\beta_0)(\tilde{\beta}_n - \beta_0) + Dh_n^2\} \to N\{0, \tilde{\Sigma}(\beta_0)\}, \quad (3.10)$$

where

$$A(\beta_0) = -\int_0^\tau \left\{\tilde{s}^{(2)}(\beta_0, t) - \frac{\tilde{s}^{(1)}(\beta_0, t)^{\otimes 2}}{\tilde{s}^{(0)}(\beta_0, t)}\right\}\mu_0(t)dt,$$

β_0 is the true regression coefficient, and D is a constant vector, which can be found in the Appendix. The asymptotic variance

$$\tilde{\Sigma}(\beta_0) = \int_{-\infty}^{\infty} K(z)^2 dz \int_0^{\tau} \left\{ \tilde{s}^{(2)}(\beta_0, t) - \frac{\tilde{s}^{(1)}(\beta_0, t)^{\otimes 2}}{\tilde{s}^{(0)}(\beta_0, t)} \right\} \mu_0(t) dt.$$

Theorem 2 permits bandwidths yielding nonzero asymptotic biases in the standardized distribution of the estimator. If we further restrict the bandwidths under (C7), then the asymptotic bias vanishes. This result is stated in the following corollary:

Corollary 2. Under conditions (C1)–(C7), $\tilde{\beta}_n$ is consistent and converges to a mean zero normal distribution given in Theorem 2.

The variance estimate for $\tilde{\beta}_n$ may be obtained by expanding the estimating Equation (3.8). With $\frac{\partial \tilde{U}_n(\beta)}{\partial \beta}|_{\beta=\tilde{\beta}_n}$ estimating $A(\beta_0)$ and $\hat{\tilde{\Sigma}}(\tilde{\beta}_n) = n^{-2} \sum_{i=1}^{n} [\int_0^{\tau} \int_0^{\infty} K_{h_n}(u-r)\{Z_i(r) - \tilde{Z}(\tilde{\beta}_n, u)\} dN_i^*(r) dN_i(u)]^{\otimes 2}$ estimating $\tilde{\Sigma}(\beta_0)$, we can do valid inference.

Corollary 3. Under conditions (C1)–(C7), the sandwich formula consistently estimates the variance of $\tilde{\beta}_n$.

With full kernel estimation, the bias is of order $O(n^{1/2} h_n^{5/2})$. Taking $h_n = o(n^{-1/5})$, the maximum allowable bandwidth under (C7) giving negligible asymptotic bias, the estimator achieves $o(n^{2/5})$ rate of convergence. One can easily show that the convergence rates in Theorem 1 hold for half kernel estimation based on recurrent events data where only backward lagged covariates are used. Thus, full kernel estimation using both forward and backward lagged covariates yields an optimal convergence rate, which is superior to the optimal $o(n^{1/3})$ rate for half kernel estimation giving negligible asymptotic bias. The practical gains associated with using both forward and backward covariates are examined in the simulations in Section 4.

4. SIMULATION STUDIES

We first study the performance of the half kernel estimator using backward lagged covariates in estimating Equation (2.5) with classical right censored data. We generate 1000 datasets, each consisting of $n = 100$, 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 are generated from uniform distribution Unif(0, 1). The covariate process is generated through a piecewise constant function

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

where $(z_i)_{i=1}^{20}$ follows a unit variance multivariate normal distribution with mean 0 and correlation $e^{-|i-j|/20}$, $i, j = 1, \ldots, 20$. The survival time is simulated from model (1.1) with $\lambda_0(t) = 1$ and $\beta_0 = 1.5$. The censoring time is 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 are rather similar and are omitted.

Based on Theorem 1, to obtain a half kernel estimator with asymptotically negligible bias, we employ bandwidths in the range $(n^{-1}, n^{-1/3})$ when calculating $\hat{\beta}_n$ using the smoothed likelihood score function. The kernel function is the Epanechnikov kernel, which is $K(x) = 0.75(1-x^2)_+$. Further simulations (not reported) evidence that the use of other kernel function has little impact on the estimator's empirical performance.

Table 1 summarizes the main findings over 1000 simulations. We observe that as the sample size increases, the bias decreases and is small, that the empirical and model-based standard errors agree reasonably well, and that the coverage is close to the nominal 0.95 level. The performance improves with larger sample sizes.

We next study the properties of the full kernel estimator under model (1.1). The data generation is identical to that presented above, where covariate observation was terminated at the minimum of the failure and censoring times. To use full kernel estimation, if a failure occurs prior to censoring, then the covariate process continues to be observed beyond the event time until the right censoring time. This is equivalent to a recurrent events setup. Following Corollary 2, the allowable bandwidth range for full kernel estimation with negligible bias is $(n^{-1}, n^{-1/5})$. The results presented in Table 1 enable a direct assessment of improvements provided by full kernel estimation with forward and backward lagged covariates over half kernel estimation, which only employs backward lagged covariates.

Similarly to half kernel estimation, as the sample size increases, the bias is well controlled, the empirical and model-based standard errors agree reasonably well, and the empirical coverage probability is close to 0.95. As predicted by the theoretical developments, full kernel estimation yields empirical gains over half kernel estimation. With the same bandwidth and same sample size, the standard error is markedly diminished when using both forward and backward lagged covariates relative to using only backward lagged covariates, with the magnitude of the bias being comparable.

We also propose a strategy for automatic bandwidth selection. The idea is to minimize the mean squared error, where the bias and variance are calculated separately. For half kernel method, the bias is of order h_n as shown in the proofs in the Appendix. So we regress $\hat{\beta}(h_n)$ with 30 equally spaced h_n in the allowable range to get an estimate for the slope $\hat{C}$. To calculate the variance, we randomly split the data into two parts, and calculate $\hat{\beta}_1(h_n)$ and $\hat{\beta}_2(h_n)$, respectively. We then choose h_n to minimize $\hat{C}^2 h_n^2 + \{\hat{\beta}_1(h_n) - \hat{\beta}_2(h_n)\}^2/4$. For full kernel method, we use similar idea except that the bias is of order h_n^2 and we regress $\hat{\beta}(h_n)$ with h_n^2 in the allowable range to calculate the bias. The results are summarized in Table 1. From the table, the automatic bandwidth procedure performs well relative to the fixed bandwidth results, both for half and full kernel estimation.

With longitudinal covariates in time-to-event analysis, a naïve approach is the last value carried forward approach. If data at a particular time point are missing, then the backward lagged covariate observed at the most recent time point in the past is imputed for the missing value. Andersen and Liestol (2003) discussed bias reduction strategies, in which the backward lagged covariate is only imputed if it falls in a window around the time point of interest. Last value carried forward is conceptually simple and its ease of implementation has led to its use in practice. However, because backward lagged covariates are not weighted by their distance from the imputation time, such procedure lacks theoretical validity. To demonstrate that

Table 1. Simulation results with different censoring rate

		Censoring rate is 15%					Censoring rate is 50%				
n	BD	Bias	RB	SD	SE	CP(%)	Bias	RB	SD	SE	CP(%)
						Half kernel					
100	$n^{-0.7}$	0.051	0.034	0.420	0.502	91	0.169	0.113	0.543	0.740	90
	$n^{-0.6}$	0.019	0.013	0.346	0.384	91	0.084	0.056	0.451	0.584	91
	Auto	0.042	0.028	0.434	0.385	93	0.047	0.031	0.595	0.500	90
400	$n^{-0.7}$	0.040	0.026	0.299	0.321	93	0.046	0.031	0.365	0.423	92
	$n^{-0.6}$	−0.009	−0.006	0.222	0.242	93	0.008	0.006	0.277	0.301	93
	Auto	0.028	0.019	0.298	0.263	93	0.035	0.023	0.378	0.332	95
900	$n^{-0.7}$	0.008	0.005	0.247	0.253	95	0.033	0.022	0.305	0.327	93
	$n^{-0.6}$	0.006	0.004	0.181	0.188	94	0.007	0.005	0.220	0.236	94
	Auto	0.014	0.010	0.244	0.216	94	0.020	0.013	0.306	0.260	93
						Last value					
100		−0.157	−0.104	0.232	0.220	85	−0.086	−0.057	0.339	0.309	90
400		−0.180	−0.120	0.115	0.106	59	−0.130	−0.087	0.146	0.146	83
900		−0.178	−0.119	0.073	0.070	30	−0.137	−0.091	0.098	0.095	69
						Full kernel					
100	$n^{-0.7}$	0.005	0.003	0.294	0.316	93	0.025	0.017	0.354	0.383	93
	$n^{-0.6}$	−0.053	−0.035	0.242	0.247	93	0.011	0.007	0.298	0.305	94
	Auto	0.039	0.026	0.311	0.278	93	0.018	0.012	0.359	0.335	91
400	$n^{-0.7}$	−0.003	−0.002	0.204	0.209	94	0.015	0.010	0.247	0.247	95
	$n^{-0.6}$	−0.020	−0.014	0.159	0.160	95	−0.013	−0.008	0.192	0.192	94
	auto	−0.022	−0.015	0.185	0.186	94	0.013	0.009	0.235	0.222	94
900	$n^{-0.7}$	0.001	0.001	0.172	0.179	94	0.004	0.003	0.209	0.213	94
	$n^{-0.6}$	−0.018	−0.012	0.126	0.135	95	−0.003	−0.002	0.223	0.231	94
	auto	−0.004	−0.003	0.146	0.151	96	−0.005	−0.004	0.206	0.185	94
						Nearest value					
100		−0.196	−0.130	0.159	0.155	73	−0.190	−0.126	0.204	0.188	76
400		−0.220	−0.145	0.081	0.073	20	−0.219	−0.146	0.100	0.088	3
900		−0.218	−0.146	0.056	0.048	2	−0.220	−0.146	0.065	0.058	7

NOTE: "BD" represents different bandwidths, "Bias" is the empirical bias, "RB" is the "Bias" divided by the true β_0, "SD" is the sample standard deviation, "SE" is the average of the standard error estimates, and "CP" represents the coverage probability of the 95% confidence interval for $\hat{\beta}_n$.

this approach may lead to substantially biased inferences, we studied its properties under the above simulation setup.

The results in Table 1 exhibit that rather large bias may be incurred by the naïve last value carried forward analysis. Such biases do not attenuate as the sample size increases and the coverage probabilities may be much lower than the nominal 0.95 level. The coverage probability is worse with decreased censoring percentage. Heuristically, as the censoring rate decreases, more events are observed and the estimator's variance decreases, yielding lower coverage probabilities.

To make fair comparison with full kernel approach, we adopt a nearest value method. In this approach, the nearest observation that could be either backward lagged covariates or forward lagged covariates is used in the calculation of partial likelihood. The results are similar to last value carried forward as both methods are biased but the nearest value approach has smaller variability as seen in Table 1.

Per the request of a referee, we have provided additional simulations comparing our approach and last value carried forward 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 results presented in Table 2 indicate that last value carried forward does not generally control the Type I error and that there may be either gain or loss of power with last value carried forward versus our approach with multiple covariates. This depends in part on the direction of the bias of the last value carried forward estimates and in part on their variances. The simulation setup is similar to those in previous sections. The hazard function is generated from $h(t) = 2e^{\beta_1 X_1(t)+\beta_2 X_2}$, where $X_1(t)$ follows a multivariate normal distribution for 20 equally spaced piecewise constant function in (0, 1). It has mean $\mu(t) = 4\sin(2\pi t)$ and variance covariance matrix with 1 at diagonal and $e^{-|t_1-t_2|}$ off diagonal at time points t_1 and t_2. The time-independent covariate X_2 follows a standard normal distribution. We also employed binomial X_2, obtaining similar results that are omitted due to space constraints. In the simulation $h = n^{-0.7}$ and we use Wald statistics to test the hypothesis $\beta_1 = 0$. Four scenarios were investigated. In the first, $\beta_1 = 0, \beta_2 = 0.5$, which looks at the Type I error control for time-dependent covariate in the presence of time-independent covariate; the second scenario is $\beta_1 = -0.3, \beta_2 = 0, 5$, which looks at the power for testing time-dependent covariate with a time-independent covariate; the third is $\beta_1 = -0.15, \beta_2 = 0.5$, which looks at the power for testing time-dependent covariate at the presence of time-independent covariate with reduced signal strength; and the last scenario is $\beta_1 = -0.3, \beta_2 = 0$, which looks at the power for testing time-dependent covariate only.

Table 2. Simulation results comparing power

	Our approach					Last value carried forward				
n	Bias	SD	SE	CP	Power(%)	Bias	SD	SE	CP	Power(%)
					Case 1: $\beta_1 = 0$, $\beta_2 = 0.5$					
100	−0.017	0.263	0.244	92	7	0.137	0.080	0.073	56	44
400	−0.002	0.183	0.181	94	5	0.129	0.039	0.035	6	94
900	0.001	0.159	0.155	94	5	0.126	0.025	0.023	0	100
					Case 2: $\beta_1 = -0.3$, $\beta_2 = 0.5$					
100	0.015	0.266	0.247	91	24	0.282	0.062	0.065	1	5
400	0.012	0.190	0.186	94	36	0.278	0.033	0.032	0	11
900	0.012	0.167	0.161	94	47	0.274	0.021	0.021	0	22
					Case 3: $\beta_1 = -0.15$, $\beta_2 = 0.5$					
100	0.016	0.256	0.243	94	12	0.191	0.068	0.068	20	9
400	0.014	0.183	0.182	94	13	0.186	0.033	0.032	0	18
900	0.005	0.168	0.157	93	17	0.185	0.022	0.021	0	39
					Case 4: $\beta_1 = -0.3$, $\beta_2 = 0$					
100	0.017	0.284	0.251	91	24	0.268	0.064	0.065	2	8
400	0.004	0.190	0.185	94	38	0.268	0.031	0.032	0	16
900	0.006	0.166	0.164	95	45	0.269	0.021	0.021	0	33

As can be seen in Table 2, last value carried forward approach continues to evidence bias, with reduced variance. The Type I error is not controlled using last value carried forward approach when there are time-independent covariates. The power can either increase or decrease using last value carried forward approach as the bias can be either up or down when there are time-independent covariates. Our approach has better power when the model contains only time-dependent covariate, with the power improving as sample size increases.

Joint modeling of longitudinal and survival data has been proposed to incorporate the most commonly used first-choice assumptions from both subject areas. In the joint modeling, one assumes that there is a true, hypothetical unobserved value of the longitudinal outcome at time t, denoted by $m_i(t)$. That is, the observed covariate $Z_i(t)$ is assumed to be subject to measurement error. In contrast to the standard proportional hazards model, which assumes no measurement error, the hazard function for the event of interest is specified conditionally on $m_i(t)$ and not $Z_i(t)$. Specifically,

$$h(t|m_i(u), 0 \le u < t) = h_0(t)\exp\{\gamma^T w_i + \alpha m_i(t)\}, \quad (4.11)$$

where w_i is a vector of baseline covariates with a corresponding vector of regression coefficients γ. The parameter α quantifies the effect of the underlying longitudinal outcome on the risk of an event. A linear mixed effects model is specified for the longitudinal data:

$$Z_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \sim N(0, \sigma^2), \quad (4.12)$$

where $Z_i(t)$ is the observed covariate, $m_i(t)$ is assumed to follow a linear mixed model, and $\epsilon_i(t)$ is assumed independent of $m_i(t)$, with mean 0 and variance σ^2. To complete the model specification, the distribution of $\epsilon_i(t)$ must be specified up to σ^2, with normality commonly assumed. In the case that the measurement error $\sigma^2 = 0$, the standard proportional hazards specification that conditions on $Z_i(t)$ is obtained.

The joint modeling relies heavily on the underlying assumptions in (4.13) and may result in invalid inferences under model misspecification. Moreover, due to the complexity of the model specification, the procedure may be computationally unstable with small and moderate sample sizes. We compare the performance of joint models to our approach and last value carried forward under the simulation setup in Table 1. We fit the joint model using the R package JM (Rizopoulos 2010), assuming normal measurement error. Note that JM cannot accommodate $\sigma^2 = 0$. When we generate data from correctly specified models (4.12) and (4.13) with zero measurement error, as is assumed by the standard proportional hazards model, the program fails to converge.

We instead compared our proposed estimator with the joint modeling strategy using JM with small measurement error, giving approximately the same survival models. The longitudinal process is generated from the linear mixed model

$$Z_i(t_{ij}) = \beta_i + 1.1t_{ij} + \epsilon_{ij},$$

where random intercept $\beta_i \sim N(-0.01, 0.7^2)$ and independently, the measurement error $\epsilon_{ij} \sim N(0, 0.05^2)$. The number of measurements for each subject is Poisson distributed with intensity rate 8, and conditional on this, observation time $t_{ij} \sim \text{Unif}(0, 2)$. We then generate the survival time based on hazard function

$$h_i(t) = \exp\{\beta_i + 1.1t\}.$$

A uniformly distributed random variable is used to produce 15% censoring rate. For our method, which is based on the usual proportional hazards model conditioned on the observed covariates, the data-generation step for the event time is identical except we use Z_i in the hazard model. For estimation, we use automatic bandwidth selection approach introduced earlier. From Table 3, we see that both methods perform well in terms of bias, variance, and coverage probability in the correctly specified setup. The efficiency gains predicted from joint modeling are reflected in the smaller empirical standard errors.

Table 3. Summary statistics comparing our method and joint modeling method

	Our method					Joint modeling method				
n	Bias	RB	SD	SE	CP(%)	Bias	RB	SD	SE	CP(%)
					Correct model					
100	0.050	0.050	0.420	0.423	95	0.041	0.041	0.252	0.231	94
400	−0.037	−0.037	0.361	0.291	91	−0.008	−0.008	0.125	0.112	95
900	−0.032	−0.032	0.242	0.247	94	−0.008	−0.008	0.073	0.075	97
					Misspecified model					
100	0.015	0.010	0.330	0.267	92	−0.017	−0.011	0.299	0.248	89
400	0.018	0.012	0.204	0.174	92	−0.173	−0.115	0.154	0.111	56
900	0.012	0.008	0.177	0.142	93	−0.218	−0.145	0.133	0.070	32

NOTE: See Table 1.

Next we generate data when the longitudinal model (4.13) is misspecified. The covariate process is generated through

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

$(z_i)_{i=1}^{20}$ follows a normal mixture model $z = 0.4z_1 + 0.6z_2$, where z_1 is a unit variance multivariate normal distribution with mean -1 and correlation $e^{-|i-j|/20}$, and z_2 is also a unit variance multivariate normal distribution with mean 1.5 and correlation $2^{-|i-j|/20}$, $i, j = 1, \ldots, 20$. The survival time is simulated from model (1.1) with $\lambda_0(t) = 1$ and $\beta_0 = 1.5$. We use censoring rate $= 15\%$ to illustrate and the bandwidth selection is based on automatic procedure introduced earlier. The results in Table 3 demonstrate that our method continues to provide unbiased estimates, the model-based standard errors agree with the empirical standard errors, and our inferences provide coverage that agrees with the nominal level. On the other hand, JM exhibits substantial bias that does not diminish as the sample size increases, the empirical and model-based standard errors do not agree, and the coverage probability may be much less than the nominal level, particularly for larger sample sizes. In addition, under sample size $n = 900$, JM failed to converge in 20 datasets, with the results in Table 2 based on those datasets where JM converged.

5. CARDIAC ARREST STUDY

We now illustrate the proposed inferential procedure in Section 2 with a comparison to the last value carried forward approach and joint modeling method on data from a cardiac arrest study. A database of 58,132 patients who were hospitalized in the wards at the University of Chicago from November 2008 until August 2011 is used. During this period, there were 109 cardiac arrests in the hospital wards and we are interested in risk factors associated with cardiac arrest. Details of the study design, methods, and medical implications can be found in Churpek et al. (2012).

Patients in the general hospital wards have vital signs, such as heart rate, blood pressure, and respiratory rate, collected routinely every few hours, and studies have found that abnormal vital signs are common before cardiac arrest on the wards as a signal of worsening condition (Churpek et al. 2012). Importantly, the collection of vital signs for these patients is erratic, occurring at different time intervals for each patient. A statistical model that associates vital signs and time to cardiac arrest would yield improved detection of high-risk patients and earlier detection of clinical deterioration resulting in better patient outcomes.

To this end, we adopt model (1.1) to analyze the relationships between vital signs and time to cardiac arrest. Because heart rate has been shown to be positively correlated with cardiac arrest and is measured accurately in patients using an electronic monitor, we took the heart rate as the covariate and studied its effect on the time to occurrence of cardiac arrest. A last value carried forward analysis yields point estimate 0.041 with standard error 0.0042, which is highly statistically significant. However, because this analysis is ad hoc and lacks formal theoretical justification, it is worthwhile to assess potential biases using our proposed methods. We computed the half kernel estimates for model (1.1) with bandwidths $h_n = 3 * (Q_3 - Q_1) * n^{-\gamma}$, where Q_3 is the 0.75 quantile, Q_1 is the 0.25 quantile of the measurement times for heart rate, and $\gamma = 0.5, 0.6, 0.7$. We take n as total number of events of cardiac arrest after eliminating missing values, which is 107, due to the relatively low event rate. Thus, the effective sample size in this dataset, for example, the number of events, is comparable to the effective sample size in the simulation studies in Section 4, owing to the very high censoring rate. Parameter estimators were obtained from estimating function (2.5) with different choices of bandwidths, confirming the ad hoc results from the last value carried forward approach. The resulting estimates and standard errors are 0.029 and 0.0037 when $\gamma = -0.5$; 0.029 and 0.0040 when $\gamma = -0.6$; 0.030 and 0.0047 when $\gamma = -0.7$; and 0.030 and 0.0039 for automatic bandwidth selection procedure. We can clearly see the positive association between heart rate and time to cardiac arrest, which has been verified in medical studies (Churpek et al. 2012). For different choices of bandwidths, both point estimate and variance do not change much, which shows that our method is not sensitive to bandwidth selection. While the effect magnitude is somewhat diminished from the last value carried forward analysis, statistical significance is achieved at the 0.05 level for all bandwidth choices, confirming the ad hoc results.

We then fit the joint model using R package JM (Rizopoulos 2010) with random intercept. The point estimate was 0.048, but standard errors were not computable due to the lack of positive definiteness of the Hessian matrix at convergence. This raises questions about whether the point estimate is the actual maximizer of the full likelihood function used to estimate the joint model. In addition to these computational stability issues, the joint model required 2 hr computing time, while the proposed approach and the last value carried forward method required several minutes, on the same computer.

6. CONCLUDING REMARKS

We have presented kernel weighting methods for estimation of proportional intensity models (1.1) and (1.2) with intermittently observed longitudinal covariates. The weighting techniques formalize the ad hoc last value carried forward approach by reducing the impact of covariates measured distant in time from the missing values. One may view the half kernel estimator based on backward lagged covariates as a theoretically justified adaptation of the "windowing" idea in Andersen and Liestol (2003). Our theoretical results show that this approach yields an estimator that cannot achieve parametric rates of convergence, unlike joint modeling (Tsiatis and Davidian 2004), where much stronger modeling assumptions are invoked. Interestingly, we find that using forward lagged covariates observed after the occurrence of an event via full kernel estimation may lead to improved rates of convergence relative to half kernel estimation but which are still slower than the parametric rate. Whether parametric rates of convergence are achievable without strong joint modeling assumptions is unclear and merits further investigation.

Both smoothed and nonsmoothed covariates may be used in our estimation procedure. Our theoretical derivations assume that the probability that the covariates are observed when the event occurs is zero. Scenarios may arise in practice where covariates are observed at the time of events. The assumption is that the probability of this occurring has zero measure, such that the information in these covariates is asymptotically negligible and only the smoothed covariates contribute information. If the probability is nonzero, then in theory, the rate of convergence of the estimator is determined by covariates observed at event times and is the usual $\sqrt{n}$ rate.

Modeling the hazard function conditionally on the current value is the standard form of the proportional hazards model; see Therneau and Grambsch (2001) for a discussion of the proportional hazards model with time-dependent covariates. All of the standard software implements the proportional hazards model with time-dependent covariates based on the specification in which the current value of the time-dependent covariate is used. That said, there may be applications in which the relationship between the hazard and a complicated function of the covariate's trajectory, such as the trend, may be of interest. To conduct such analyses, more complicated models are needed, for example, joint models, in which the failure time and the time-dependent covariate are jointly modeled. The usual proportional hazards model that does not require modeling the time-dependent covariate may not be as amenable to capturing such covariate effects.

The standard form of the proportional hazards model is specified conditionally on the observed value of the covariate and does not permit measurement error. The goal of this article is to provide methods for fitting the standard proportional hazards model with sparsely observed time-dependent covariates in the absence of measurement error. We note that even with time-independent covariates the presence of measurement error invalidates the standard partial likelihood estimators and more complicated models and estimation procedures are needed. With time-dependent covariates, the presence of measurement error necessitates the use of joint models and simultaneous estimation of the longitudinal and survival models via maximum likelihood, which is complicated both computationally and inferentially.

We note that in general when employing a standard proportional hazards model with "internal" (or endogenous) time-dependent covariate it is not possible to predict survival based on $Z(t)$. Such "internal" covariates are measured on the individual being followed for the event of interest. For details, please see the discussion in Kalbfleisch and Prentice (2002). If the covariate is "external" (or exogenous), for example, not measured at the individual level, then prediction may be possible. Both the half and full kernel methods provide consistent and asymptotically normal estimates of the regression parameters in the proportional hazards model, regardless of whether the time-dependent covariates are "internal" or "external." If the covariates are "internal," then prediction is not possible, while if they are "external," prediction is possible. These results regarding prediction are true for the usual partial likelihood estimator with time-dependent covariates when the covariates are fully observed. For the case of "internal" covariates, if prediction is desired, alternative modeling strategies, like joint modeling, are needed.

While the joint modeling approach has certain modeling advantages over the standard Cox model, in addition to potential improvements in efficiency, these gains depend heavily on strong assumptions about the model for the longitudinal covariates and considerable care is needed in the implementation of these full likelihood methods. In the cardiac data analysis, computational problems resulted in questionable results, while simpler partial likelihood procedures converged reliably. Minimal assumptions are required for the longitudinal covariates for the validity of the kernel weighted partial likelihood estimators, which only use assumptions on the model for the failure time, leading to the bias variance trade-off evidenced in the simulations in Section 4.

Additional simulations were performed to assess whether our proposed method might be severely underpowered relative to the last value carried forward approach for testing the effect of the time-dependent covariate. Results (omitted) demonstrate that the relative power of the two procedures depends in part on the magnitude and direction of the bias for last value carried forward and in part on the improved efficiency of last value carried forward. In certain scenarios, where the bias is strongly toward the null, last value forward may lose considerable power relative to our proposed methods. In these simulations, the time-dependent covariate had a strong and nonlinear trend, where the mean function for the time-dependent covariate is a sinusoidal function, which oscillates strongly over the time interval of observation. Our proposed method performed well in this scenario with a moderate number of observation times for the time-dependent covariate. To further explore the efficiency issue, we conducted simulations in which the trajectory is completely observed. In this case, the naïve last value carried forward analysis is valid as covariate values are observed at all event times. This may be viewed as a "gold standard" analysis. In such settings, our proposed method is unbiased but unsurprisingly may incur substantial efficiency loss relative to the "gold standard," which is also unbiased. The efficiency loss in the simulations diminishes as the number of observations of the covariate process increases.

As mentioned previously, the methods of Cai and Sun (2003) and Tian, Zucker, and Wei (2005) for the proportional hazards model with time-dependent regression parameter apply identical kernel weights to all individuals when smoothing the partial likelihood. These methods are not directly applicable in our setting, where different weights are needed for different individuals. Moreover, the methods for time-dependent regression parameters with time-dependent covariates require that the trajectory of the time-dependent covariate is fully observed. If the covariate is sparsely observed, then the methods are not applicable. It would be of interest to generalize our smoothed partial likelihood approach for the standard formulation of the Cox model with time-independent regression parameter to the time-dependent proportional hazards model with sparsely observed time-dependent covariates. This is a topic for future research.

When the observation times are informative, as might occur when there is more frequent monitoring of high risk subjects, the usual assumption of independent observation times is violated and our methods are not valid. Relevant literature on this topic includes Sun et al. (2005), Liu, Huang, and O'Quigley (2008), and Sun et al. (2012), among others. Future work is needed to extend our methods to accommodate such informative observation times.

7. SUPPLEMENTARY MATERIALS

Supplementary materials contain relevant proofs for Sections 2 and 3.

[Received October 2013. Revised July 2014.]

REFERENCES

Andersen, P., and Gill, R. (1982), "Cox's Regression Models for Counting Processes: A Large-Sample Study," *The Annals of Statistics*, 10, 1100–1120. [1187]

Andersen, P., and Liestol, K. (2003), "Attenuation Caused by Infrequently Updated Covariates in Survival Analysis," *Biostatistics*, 4, 633–649. [1187,1188,1190,1191,1195]

Cai, Z., and Sun, Y. (2003), "Local Linear Estimation for Time-Dependent Coefficients in Cox's Regression Models," *Scandinavian Journal of Statistics*, 30, 93–111. [1188,1189,1196]

Churpek, M., Yuen, T., Park, S. Y., Hall, J. B., and Edelson, D. P. (2012), "Can Vital Signs Predict Cardiac Arrest on the Wards? A Nested Case-Control Study," *Chest*, 141, 1170–1176. [1194]

Degruttola, V., and Tu, X. M. (1994), "Modelling Progression of CD-4 Lymphocyte Count and its Relation to Survival Time," *Biometrics*, 50, 1003–1014. [1187]

Dupuy, J.-F., Grama, I., and Mesbah, M. (2006), "Asymptotic Theory for the Cox Model With Missing Time-Dependent Covariate," *The Annals of Statistics*, 34, 903–924. [1188]

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. [1187]

Gill, R. D., van der Laan, M. J., and Robins, J. M. (1997), "Coarsening at Random: Characterizations, Conjectures and Counterexamples," in *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, eds. D. Y. Lin and T. R. Fleming, New York: Springer, pp. 255–294. [1188]

Henderson, R., Diggle, P., and Dobson, A. (2000), "Joint Modelling of Longitudinal Measurements and Event Time Data," *Biostatistics*, 4, 465–480. [1187]

Hogan, J. W., and Laird, N. M. (1997), "Mixture Models for the Joint Distribution of Repeated Measures and Event Times," *Statistics in Medicine*, 16, 239–257. [1187]

Ibrahim, J. G., Chu, H., and Chen, L. M. (2010), "Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data," *Journal of Clinical Oncology*, 28, 2796–2801. [1187]

Kalbfleisch, J. D., and Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data*, New York: Wiley. [1195]

Liu, L., Huang, X., and O'Quigley, J. (2008), "Analysis of Longitudinal Data in the Presence of Informative Observational Times and a Dependent Terminal Event, With Application to Medical Cost Data," *Biometrics*, 64, 950–958. [1196]

Rizopoulos, D. (2010), "JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data," *Journal of Statistical Software*, 35, 1–33. [1193,1194]

——— (2012), *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*, Boca Raton, FL: Chapman and Hall/CRC. [1188]

Sun, J., Park, D.-H., Sun, L., and Zhao, X. (2005), "Semiparametric Regression Analysis of Longitudinal Data With Informative Observation Times," *Journal of the American Statistical Association*, 100, 882–889. [1196]

Sun, L., Song, X., Zhou, J., and Liu, L. (2012)m "Joint Analysis of Longitudinal Data With Informative Observation Times and a Dependent Terminal Event," *Journal of the American Statistical Association*, 107, 688–700. [1196]

Therneau, T. M., and Grambsch, P. M. (2001), *Modeling Survival Data: Extending the Cox Model*, New York: Springer. [1195]

Tian, L., Zucker, D., and Wei, L. J. (2005), "On the Cox Model With Time-Varying Regression Coefficients," *Journal of the American Statistical Association*, 100, 172–183. [1188,1189,1196]

Tsiatis, A., and Davidian, M. (2001), "A Semiparametric Estimator for the Proportional Hazards Model With Longitudinal Covariate Measured With Error," *Biometrika*, 88, 447–458. [1188]

——— (2004), "Joint Modelling of Longitudinal and Time-to-Event Data: An Overview," *Statistica Sinica*, 14, 809–834. [1188,1189,1190,1195]

Tsiatis, A., Degruttola, V., and Wulfsohn, M. S. (1995), "Modelling 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. [1187]

Wulfsohn, M. S., and Tsiatis, A. A. (1997), "A Joint Model for Survival and Longitudinal Data Measured With Error," *Biometrics*, 53, 330–339. [1187]

Xu, J., and Zeger, S. L. (2001), "Joint Analysis of Longitudinal Data Comprising Repeated Measures and Times to Events," *Journal of Applied Statistics*, 50, 375–387. [1187]

Yao, F. (2007), "Functional Principal Component Analysis for Longitudinal and Survival Data," *Statistica Sinica*, 17, 965–983. [1187]

Zeng, D., and Cai, J. (2005), "Asymptotic Results for Maximum Likelihood Estimators in Joint Analysis of Repeated Measurements and Survival Time," *Annals of Statistics*, 33, 2132–2163. [1187]