Ann Inst Stat Math (2009) 61:197–213
DOI 10.1007/s10463-007-0129-1

Varying-coefficient model for the occurrence rate function of recurrent events

Chin-Tsang Chiang · Mei-Cheng Wang

Received: 13 June 2006 / Revised: 2 October 2006 / Published online: 25 April 2007

Abstract This article mainly considers the recurrent event process with independent censoring mechanism through a more flexible varying-coefficient model. The smoothing estimators for the varying-coefficient functions are also proposed via maximizing the kernel weight version of the log-partial likelihood function with respect to the coefficients at each time point. For the selection of appropriate bandwidths and the construction of confidence intervals, the consistent empirical smoothing estimators for the covariance functions of the estimators and a bias correction method are considered. As for the baseline effect function of recurrent events in the population, two different smoothing estimation methods are suggested and investigated. In this study, the asymptotic properties of the proposed smoothing estimators are derived. The finite sample properties of our methods are examined through a Monte Carlo simulation. Moreover, the procedures are applied to a recurrent sample of AIDS link to intravenous experiences (ALIVE) cohort study.

Keywords Independent censoring · Kernel · Partial likelihood function · Rate function · Recurrent event · Smoothing estimator · Varying-coefficient model

1 Introduction

In longitudinal follow-up studies, recurrent event data in which individuals experience multiple events repeatedly over time have been widely analyzed and studied.

C.-T. Chiang (✉)
Department of Mathematics, National Taiwan University, Taipei 106, Taiwan, ROC
e-mail: chiangct@math.ntu.edu.tw

M.-C. Wang
Department of Biostatistics, The Johns Hopkins University, Baltimore, MD 21205, USA

Statistical analyses for this type of data are required to estimate the occurrence rate function of the recurrent event process $N(t)$, and to detect the effects of possible covariates $\mathbf{X}(t) = (X_1(t), \ldots, X_k(t))^T$ on the recurrent event process. For the loss to follow up, the recurrent event process may be terminated at time Y, which is the censoring time, during the study period $[0, T_0]$. Based on the above setting, the recurrent event data, say, $\{(N_i(\cdot), Y_i, \mathbf{X}_i(\cdot)); 1 \le i \le n\}$ are collected from n independent selected individuals.

Recent statistical methods for the rate function without using the information of covariates can be tracked back to the non-parametric procedures of Andersen et al. (1993), Lawless and Nadeau (1995), and Nelson (1995). When covariates are considered, Pepe and Cai (1993) modeled the recurrent event process with independent censoring mechanism through the semi-parametric regression model

$$E[\mathrm{d}N_i(t)|\mathbf{X}_i(t)] = \exp(\boldsymbol{\beta}^T\mathbf{X}_i(t))\mathrm{d}\mu_0(t), \tag{1}$$

where $\boldsymbol{\beta}^T = (\beta_1, \ldots, \beta_k)$ is a $k \times 1$ time-independent regression parameter vector, and $\mu_0(t)$ is an unknown continuous function. Lin et al. (2000) further provided a rigorous justification of the marginal model through the empirical process theory. For related research which takes into account informative censoring on the recurrent event process, it can be referred to the works of Lancaster and Intrator (1998), and Wang et al. (2001).

Although the semi-parametric regression model (1) was used in many empirical biomedical and epidemiological cohort studies, the constant effects of covariates on the recurrent event process are sometimes unreasonable in practical applications. Motivated by an epidemiological example of detecting some demographical variables on the frequency of inpatient cares among HIV-negative intravenous drug users, we consider a more flexible varying-coefficient model for each recurrent event process as

$$E_i[\mathrm{d}N_i(t)|\mathbf{X}_i(t)] = \exp(\boldsymbol{\beta}^T(t)\mathbf{X}_i(t))\lambda_{0i}(t)\mathrm{d}t \tag{2}$$

with

$$E[\mathrm{d}N_i(t)|\mathbf{X}_i(t)] = \exp(\boldsymbol{\beta}^T(t)\mathbf{X}_i(t))\lambda_0(t)\mathrm{d}t,$$

where $E_i[\cdot|\mathbf{X}_i(t)]$ is the expectation conditioning on $\mathbf{X}_i(t)$ and some other random variables of the ith subject, which can be considered as the random effect or frailty, $\lambda_{0i}(t)$'s are subject-specific baseline effects, and the coefficients $\boldsymbol{\beta}(t) = (\beta_1(t), \ldots, \beta_k(t))^T$ and $\lambda_0(t)$ are assumed to be smooth functions of t. Here, $\lambda_0(t) = E[\lambda_{0i}(t)]$ is defined to be the baseline effect of recurrent events at time t in the population. It can be found that model (2) avoids the complexity of modeling and allows the effects of covariates to be time-dependent. Under the validity of the above varying-coefficient model, an independent censoring condition is further assumed in this study for the recurrent event process:

(A1) Conditioning on $\mathbf{X}_i(t)$, $N_i(t)$ is independent of Y_i.

Note that neither of uniformly bounded nor distributional assumptions is made on the recurrent event process $\{N_i(t)\}$. In the theoretical development of this paper, a very mild condition, the bounded third moment of each recurrent event process, is assumed. With the above independent censoring assumption on the recurrent event process, a class of smoothing estimation methods are proposed for the varying-coefficient functions $\boldsymbol{\beta}(t)$. In this study, our research efforts will focus on the discussion of the kernel smoothing estimators. The consistent empirical smoothing estimators for the covariance functions of the estimators and a bias correction method are also considered in the bandwidth selection and the construction of confidence intervals. For the estimation of the baseline effect function $\lambda_0(t)$, two smoothing estimation methods are considered and are further explored through a theoretical viewpoint and a practical implementation.

In Sect. 2, the smoothing estimation methods are proposed. A bias correction method is also suggested for bandwidth selection. Section 3 derives the asymptotic properties of the smoothing estimators, and proposes the procedures to construct the approximated confidence intervals for the varying coefficient functions. The finite sample properties of the estimators and the proposed procedures are investigated through a Monte Carlo simulation in Sect. 4. In Sect. 5, our methods are applied to the data set of AIDS link to intravenous experiences (ALIVE) cohort study. Finally, the proofs are placed in the Appendix.

2 Estimation

In this section, we propose the smoothing estimation methods for the varying-coefficient functions $\boldsymbol{\beta}(t)$ and the baseline effect function $\lambda_0(t)$ in (2). Moreover, the consistent empirical smoothing estimators for the covariance functions of the estimators are suggested. The bias correction method of Schucany (1995) is also extended in this data setting to estimate the dominant bias terms of the estimators.

2.1 Estimation of $\boldsymbol{\beta}(t)$

Let $S_j(\boldsymbol{\beta}(t), u) = n^{-1}\sum_{i=1}^{n} Y_i(u)\mathbf{X}_i^{\otimes j}(u)\ \exp(\boldsymbol{\beta}^T(t)\mathbf{X}_i(u)),\ j = 0, 1, 2,$ with $\mathbf{X}^{\otimes 0}(u) = 1$, $\mathbf{X}^{\otimes 1}(u) = \mathbf{X}(u)$, and $\mathbf{X}^{\otimes 2}(u) = \mathbf{X}(u)\mathbf{X}^T(u)$. When the subject-specific baseline effect functions are set to be equal, model (2) will be reduced to

$$E_i[\mathrm{d}N_i(t)|\mathbf{X}_i(t)] = \exp(\boldsymbol{\beta}^T(t)\mathbf{X}_i(t))\lambda_0(t)\mathrm{d}t. \tag{3}$$

Under the validity of model (3), our proposed estimator $\widehat{\boldsymbol{\beta}}_{h_t}(t) = \big(\widehat{\beta}_{1h_t}(t), \ldots, \widehat{\beta}_{kh_t}(t)\big)^T$ for $\boldsymbol{\beta}(t)$ can be obtained by maximizing the kernel weight version of the log-partial likelihood function

$$l(\boldsymbol{\beta}(t); h_t) = \sum_{i=1}^{n}\int_0^{T_0} [\boldsymbol{\beta}^T(t)\mathbf{X}_i(u) - \ln(nS_0(\boldsymbol{\beta}(t), u))]Y_i(u)K_2\left(\frac{t-u}{h_t}\right)\mathrm{d}N_i(u), \tag{4}$$

where $Y_i(u) = 1_{[Y_i \geq u]}$, $K_l(\frac{t-u}{h_t}) = \frac{1}{h_t}\alpha_l(T_0, \frac{t-u}{h_t})K(\frac{t-u}{h_t})$ is the lth order boundary kernel function of Gasser and Müller (1978) with adjustment for the boundary time T_0, which satisfies $\gamma_{0,l}(t, h_t) = 1$, $\gamma_{j,l}(t, h_t) = 0$ for $1 \leq j \leq l-1$, and $\gamma_{l,l}(t, h_t) < \infty$ with $\gamma_{j,l}(t, h_t) = \int_{(t-T_0)/h_t}^{t/h_t} u^j \alpha_l\ (T_0, u)K(u)\mathrm{d}u$, h_t is a positive valued bandwidth, and $K(\cdot)$ is a kernel density function. In practical implementation, $\alpha_l(T_0, u)$ is often assigned to be the lth order polynomial function of u. Differentiating $l(\boldsymbol{\beta}(t); h_t)$ with respect to $\boldsymbol{\beta}(t)$, the kernel weight score function of $\boldsymbol{\beta}(t)$ is derived to be

$$U(\boldsymbol{\beta}(t); h_t) = \sum_{i=1}^{n} \int_0^{T_0} \left(\mathbf{X}_i(u) - \bar{\mathbf{X}}(\boldsymbol{\beta}(t), u)\right) Y_i(u) K_2\left(\frac{t-u}{h_t}\right) \mathrm{d}N_i(u), \tag{5}$$

where $\bar{\mathbf{X}}(\boldsymbol{\beta}(t), u) = S_1(\boldsymbol{\beta}(t), u)/S_0(\boldsymbol{\beta}(t), u)$. The maximum partial likelihood smoothing estimator $\widehat{\boldsymbol{\beta}}_{h_t}(t)$ is then defined to be the solution of $U(\boldsymbol{\beta}(t); h_t)$, i.e. $U(\widehat{\boldsymbol{\beta}}_{h_t}(t); h_t) = \mathbf{0}$. Besides the proposed kernel smoothing method, the local polynomial estimation procedure can be developed by using the polynomial functions within a neighborhood of t to approximate the non-parametric functions $\beta_j(t)$, $j = 1, \ldots, k$, in (4). Moreover, the smoothing spline estimation method of Zuuker and Karr (1990) and polynomial splines of Gray (1992), which are applied in the survival analysis, can be extended to our data setting. Recent research concerning the time-varying coefficient effects in the Cox model can also track back to the works of Murphy and Sen (1991), Martinussen and Scheike (2002), Cai and Sun (2003), Winnett and Sasieni (2003), and Tian et al. (2005). In the succeeding sections, we focus on the discussion of the kernel smoothing estimators.

Under model (2) and some regularity conditions assumed in the next section, we will show that the random vector $(nh_t)^{1/2}\left(\widehat{\boldsymbol{\beta}}_{h_t}(t) - \boldsymbol{\beta}(t)\right)$ asymptotically converges to a multivariate normal distribution with mean vector $\mathbf{b}(t)$ and variance–covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\beta}(t), t)\delta_2(t, h_t)$, where

$$\mathbf{b}(t) = \sum_{\{i+j=1, i, j>=0\}} \boldsymbol{\Sigma}(\boldsymbol{\beta}(t), t)\left(\lambda_0^{(i)}(t)\mathbf{M}_{11}^{(j)}(\boldsymbol{\beta}(t), t) + \frac{\lambda_0(t)\mathbf{M}_{i+1, j+1}^{(1)}(\boldsymbol{\beta}(t), t)}{2}\right)\gamma_{2,2}(t, h),$$

and

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}(t), t) = \left(E\left[\mathbf{Z}^{\otimes 2}(\boldsymbol{\beta}(t), t)Y(t)\lambda_0(t)\exp(\boldsymbol{\beta}^T(t)\mathbf{X}(t))\right]\right)^{-1}$$

with $\delta_j(t, h_t) = \int_{(t-T_0)/h_t}^{t/h_t} |\alpha_2(T_0, u)K(u)|^j \mathrm{d}u$,

$$\mathbf{M}_{ij}(\boldsymbol{\beta}(t), u) = E\left[\mathbf{Z}(\boldsymbol{\beta}(t), u)Y(u)(\boldsymbol{\beta}^{(j)T}(t)\mathbf{X}(u))^i \exp(\boldsymbol{\beta}^T(t)\mathbf{X}(u))\right],$$

$\mathbf{Z}(\boldsymbol{\beta}(t), u) = \mathbf{X}(u) - \bar{\mathbf{x}}(\boldsymbol{\beta}(t), u)$, $\lambda^{(i)}(t)$ and $\boldsymbol{\beta}^{(i)}(t)$ are separately the ith derivatives of $\lambda(t)$ and $\boldsymbol{\beta}(t)$, $\mathbf{M}_{ij}^{(1)}(\boldsymbol{\beta}(t), u)$ being the first derivative of $\mathbf{M}_{ij}(\boldsymbol{\beta}(t), u)$ with respect to u, and

$$\bar{\mathbf{x}}(\boldsymbol{\beta}(t), u) = \frac{E[Y(u)\mathbf{X}(u)\exp(\boldsymbol{\beta}^T(t)\mathbf{X}(u))]}{E[Y(u)\exp(\boldsymbol{\beta}^T(t)\mathbf{X}(u))]}.$$

As it is well known, the smoothing estimator $\widehat{\boldsymbol{\beta}}_{h_t}(t)$ is not an unbiased estimator of $\boldsymbol{\beta}(t)$. To remedy this problem, the differences of the second order kernel estimator $\widehat{\boldsymbol{\beta}}_{h_t}(t)$ and the fourth order kernel estimator $\widetilde{\boldsymbol{\beta}}_{h_t}(t)$, which is computed from (5) with the fourth order boundary kernel function $K_4(\cdot)$, at some selected bandwidths are used to estimate $\mathbf{b}(t)$. Let $\widehat{\mathbf{b}}(t)$ be the estimator of $\mathbf{b}(t)$. Our bias adjusted estimator for $\boldsymbol{\beta}(t)$ is suggested to be $\bar{\boldsymbol{\beta}}_{h_t}(t) = \widehat{\boldsymbol{\beta}}_{h_t}(t) - \widehat{\mathbf{b}}(t)h_t^2$. For the covariance functions of $\widehat{\boldsymbol{\beta}}_{h_t}(t)$, instead of directly estimating the unknown quantities in $\boldsymbol{\Sigma}(\boldsymbol{\beta}(t), t)$, we propose the empirical consistent smoothing estimators, say, $\mathrm{V}(\bar{\boldsymbol{\beta}}_{h_t}(t), t)$ as

$$\mathrm{V}(\bar{\boldsymbol{\beta}}_{h_t}(t), t) = n^{-1}\mathrm{V}_2^{-1}(\bar{\boldsymbol{\beta}}_{h_t}(t), t)\mathrm{V}_1(\bar{\boldsymbol{\beta}}_{h_t}(t), t)\mathrm{V}_2^{-1}(\bar{\boldsymbol{\beta}}_{h_t}(t), t), \tag{6}$$

where

$$\mathrm{V}_1(\bar{\boldsymbol{\beta}}_{h_t}(t), t) = n^{-1}\sum_{i=1}^{n}\int_0^{T_0}\left(\mathbf{X}_i(u) - \bar{\mathbf{X}}(\bar{\boldsymbol{\beta}}_{h_t}(t), u)\right)^{\otimes 2} Y_i(u)K_2^2\left(\frac{t-u}{h_t}\right)\mathrm{d}N_i(u),$$

$$\mathrm{V}_2(\bar{\boldsymbol{\beta}}_{h_t}(t), t) = n^{-1}\sum_{i=1}^{n}\int_0^{T_0} S(\bar{\boldsymbol{\beta}}_{h_t}(t), u)Y_i(u)K_2\left(\frac{t-u}{h_t}\right)\mathrm{d}N_i(u),$$

and

$$S(\boldsymbol{\beta}(t), u) = \left(\frac{n^{-1}\sum_{i=1}^{n}(\mathbf{X}_i(u) - \bar{\mathbf{X}}(\boldsymbol{\beta}(t), u))^{\otimes 2}Y_i(t)\exp(\boldsymbol{\beta}^T(t)\mathbf{X}_i(u))}{S_0(\boldsymbol{\beta}(t), u)}\right).$$

The reason of doing so can be explained from the derivation for the asymptotic properties of $\widehat{\boldsymbol{\beta}}_{h_t}(t)$ in the next section. Although the well-known bootstrap procedures can be considered to estimate the covariance functions or approximate the sampling quantities of $\widehat{\boldsymbol{\beta}}_{h_t}(t)$, it was found to be computational inefficient in implementation from a practical point of view. This is due to a mass of computation works in the bootstrap analogues of $\widehat{\boldsymbol{\beta}}_{h_t}(t)$.

2.2 Estimation of $\lambda_0(t)$

For the baseline effect function $\lambda_0(t)$, two different smoothing estimation methods are proposed. From the varying-coefficient model (2), we get the equality

$$E\left[\frac{\mathrm{d}N_i(t)}{\exp(\boldsymbol{\beta}^T(t)\mathbf{X}_i(t))}\middle|\mathbf{X}_i(t)\right] = \lambda_0(t)\mathrm{d}t. \tag{7}$$

Under the independent censoring assumption (A1), $Y.(t) = \sum_{i=1}^{n} Y_i(t)$ indicates the size of a random sample from the risk population which includes subjects who are still at risk at t. Thus, $W_i(t) = Y_i(t)/Y.(t)$ can be treated as the weight function for the ith recurrent event process $N_i(t)$ in the establishment of an estimation criterion. Using the property $E[W_i(t)|\mathbf{X}_i(t)] = n^{-1}$ and the independent censoring assumption (A1), it implies that

$$E\left[\frac{nW_i(t)\mathrm{d}N_i(t)}{\exp(\boldsymbol{\beta}^T(t)\mathbf{X}_i(t))}\bigg|\mathbf{X}_i(t)\right] = \lambda_0(t)\mathrm{d}t. \tag{8}$$

By substituting the estimator $\bar{\boldsymbol{\beta}}_{h_t}(t)$ for unknown parameter function $\boldsymbol{\beta}(t)$ in (8), the first smoothing estimator, say, $\widehat{\lambda}_{0h_t}(t)$ is proposed to be the minimizer of the following sum of squares

$$D_1(\lambda_0(t); h_t) = \sum_{i=1}^{n}\int_0^{T_0}\left(\frac{nW_i(u)\mathrm{d}N_i(u)}{\exp(\bar{\boldsymbol{\beta}}_{h_t}^T(u)\mathbf{X}_i(u))\mathrm{d}u} - \lambda_0(t)\right)^2 K_2\left(\frac{t-u}{h_t}\right)\mathrm{d}u. \tag{9}$$

Instead of using the estimator $\widehat{\boldsymbol{\beta}}_{h_t}(t)$ in criterion (9), we do this mainly to avoid the influence of the dominant bias $\mathbf{b}(t)h_t^2$ of $\widehat{\boldsymbol{\beta}}_{h_t}(t)$ on the moments of $\widehat{\lambda}_{0h_t}(t)$. Differentiating the sum of squares $D_1(\lambda_0(t); h_t)$ with respect to $\lambda_0(t)$, $\widehat{\lambda}_{0h_t}(t)$ is derived to be

$$\widehat{\lambda}_{0h_t}(t) = \sum_{i=1}^{n}\int_0^{T_0}\frac{W_i(u)K_2(\frac{t-u}{h_t})\mathrm{d}N_i(u)}{\exp(\bar{\boldsymbol{\beta}}_{h_t}^T(u)\mathbf{X}_i(u))}. \tag{10}$$

Multiplying both sides of model (2) by $Y_i(t)$ and taking expectation, one gets the equality

$$E[Y_i(t)\mathrm{d}N_i(t)] = E[Y_i(t)\exp(\boldsymbol{\beta}^T(t)\mathbf{X}_i(t))]\lambda_0(t)\mathrm{d}t \tag{11}$$

By using the above equality and substituting the consistent estimator $S_0(\bar{\boldsymbol{\beta}}_{h_t}(t), t)$ for $E[Y_i(t)\exp(\boldsymbol{\beta}^T(t)\mathbf{X}_i(t))]$, another smoothing estimator can be obtained via minimizing the sum of squares

$$D_2(\lambda_0(t); h_t) = \sum_{i=1}^{n}\int_0^{T_0}\left(\frac{Y_i(u)\mathrm{d}N_i(u)}{S_0(\bar{\boldsymbol{\beta}}_{h_t}(u), u)\mathrm{d}u} - \lambda_0(t)\right)^2 K_2\left(\frac{t-u}{h_t}\right)\mathrm{d}u \tag{12}$$

with respect to $\lambda_0(t)$. The solution of (12) leads to the smoothing estimator

$$\widetilde{\lambda}_{0h_t}(t) = \sum_{i=1}^{n}\int_0^{T_0} K_2\left(\frac{t-u}{h_t}\right)\frac{Y_i(u)\mathrm{d}N_i(u)}{nS_0(\bar{\boldsymbol{\beta}}_{h_t}(u), u)}. \tag{13}$$

As one can observe, the estimator $\widetilde{\lambda}_{0h}(t)$ in (13) can also be computed by smoothing the Aalen–Breslow type estimator

$$\widetilde{\Lambda}_{0h_t}(t) = \sum_{i=1}^{n} \int_0^{T_0} \frac{Y_i(u)\mathrm{d}N_i(u)}{nS_0(\bar{\boldsymbol{\beta}}_{h_t}(u), u)}$$

of $\Lambda_0(t) = \int_0^t \lambda_0(u)\mathrm{d}u$ as below.

$$\widetilde{\lambda}_{0h_t}(t) = \int_0^{T_0} K_2\left(\frac{t-u}{h_t}\right)\mathrm{d}\widetilde{\Lambda}_{0h_t}(u). \tag{14}$$

When the covariates are not considered, one can derive from the equalities (8) and (11) that both estimators are the same. Further comparisons between the two estimators will be made through the asymptotic behaviors and a simulation study.

3 Asymptotic properties

The asymptotic normalities of the estimators for the varying-coefficients $\boldsymbol{\beta}(t)$ and the baseline effect $\lambda_0(t)$ are derived in this section. For the asymptotic properties of $\widehat{\boldsymbol{\beta}}_{h_t}(t)$, we first expand the term $\bar{\mathbf{X}}(\widehat{\boldsymbol{\beta}}_{h_t}(t), u)$ as

$$\bar{\mathbf{X}}(\widehat{\boldsymbol{\beta}}_{h_t}(t), u) = \bar{\mathbf{X}}(\boldsymbol{\beta}(t), u) - S(\boldsymbol{\beta}^*(t), u)\left(\widehat{\boldsymbol{\beta}}_{h_t}(t) - \boldsymbol{\beta}(t)\right), \tag{15}$$

where $\boldsymbol{\beta}^*(t)$ is on the line segment between $\boldsymbol{\beta}(t)$ and $\widehat{\boldsymbol{\beta}}_{h_t}(t)$. Under the assumption (A1) and the following conditions, $\widehat{\boldsymbol{\beta}}_{h_t}(t)$ can be shown to converge in probability to $\boldsymbol{\beta}(t)$ by paralleling the proof of Lin et al. (2000).

- (A2) $\lambda_0(t)$ and $\beta_l(t)$, $l = 1, \ldots, k$, are twice differentiable and bounded.
- (A3) $|X_l(0)| + \int_0^{T_0} |\mathrm{d}X_l(u)| \leq c_0$ for $l = 1, \ldots, k$, where c_0 is a positive constant.
- (A4) The elements of $\boldsymbol{\Sigma}(\boldsymbol{\beta}(t), u)$ are continuous with respect to u and $\boldsymbol{\Sigma}(\boldsymbol{\beta}(t), t)$ is positive definite for $t \in [0, T_0]$.

By further applying the law of large numbers to $\bar{\mathbf{X}}(\boldsymbol{\beta}(t), u)$ and $\mathrm{V}_2(\boldsymbol{\beta}(t), t)$, $U(\widehat{\boldsymbol{\beta}}_{h_t}(t); h_t)$ and $\widehat{\boldsymbol{\beta}}_{h_t}(t)$ can be expressed as

$$\begin{aligned} n^{-1}U(\widehat{\boldsymbol{\beta}}_{h_t}(t); h_t) &= n^{-1}U(\boldsymbol{\beta}(t); h_t) + \mathrm{V}_2(\boldsymbol{\beta}(t), t)\left(\widehat{\boldsymbol{\beta}}_{h_t}(t) - \boldsymbol{\beta}(t)\right)\left(1 + o_p(1)\right) \\ &= n^{-1}U(\boldsymbol{\beta}(t); h_t) + \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}(t), t)\left(\widehat{\boldsymbol{\beta}}_{h_t}(t) - \boldsymbol{\beta}(t)\right)\left(1 + o_p(1)\right) \end{aligned} \tag{16}$$

and

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{h_t}(t) - \boldsymbol{\beta}(t) &= -\boldsymbol{\Sigma}(\boldsymbol{\beta}(t), t)\left(n^{-1}U(\boldsymbol{\beta}(t); h_t)\right)\left(1 + o_p(1)\right) \\ &= n^{-1}\sum_{i=1}^{n} \xi_i(t)\left(1 + o_p(1)\right), \end{aligned} \tag{17}$$

where $\xi_i(t) = -\boldsymbol{\Sigma}(\boldsymbol{\beta}(t), t) \int_0^{T_0} \mathbf{Z}_i(\boldsymbol{\beta}(t), u) Y_i(u) K_2(\frac{t-u}{h_t}) \mathrm{d}N_i(u)$. Let $\mathbf{a} = (a_1, \ldots, a_k)^T$ be any constant vector and define $\widehat{\beta}_{ah_t}(t) - \beta_a(t) = \mathbf{a}^T(\widehat{\boldsymbol{\beta}}_{h_t}(t) - \boldsymbol{\beta}(t))$. It follows that

$$\widehat{\beta}_{ah_t}(t) - \beta_a(t) = n^{-1} \sum_{i=1}^{n} \eta_i(t) \left(1 + o_p(1)\right) \tag{18}$$

with $\eta_i(t) = \mathbf{a}^T \xi_i(t)$. By the Berry–Eséen theorem for the triangle array of independent random variables, we can get the inequality

$$\sup_z \left| P\left(\frac{\sum_{i=1}^n (\eta_i(t) - E[\eta_i(t)])}{\sqrt{nV(\eta_i(t))}} \le z \right) - \Phi(z) \right| \le \mathrm{d}(t) \frac{(nE[|\eta_i(t) - E[\eta_i(t)]|^3])}{(nV(\eta_i(t))^{3/2})}, \tag{19}$$

where $\mathrm{d}(t)$ is a positive constant independent of n and $\eta_i(t)$. The moments of $\eta_i(t)$ in (19) are derived in Lemma 1 under the following further conditions:

(A5) The elements of $\mathbf{M}_{11}^{(1)}(\boldsymbol{\beta}(t), u)$, $\mathbf{M}_{12}(\boldsymbol{\beta}(t), u)$, and $\mathbf{M}_{21}(\boldsymbol{\beta}(t), u)$ are continuous with respect to u.

(A6) $E[dN(u)dN(v)dN(w)|\mathbf{X}(u)\mathbf{X}(v)\mathbf{X}(w)]$ is uniformly bounded for all $u, v, w \in [0, T_0]$.

When each recurrent event process is assumed to be uniformly bounded by a positive constant, assumption (A6) is automatically satisfied.

Lemma 1 *Suppose that assumptions* (A1) *through* (A6) *are satisfied and* $E[Y(T_0)] > 0$,

$$E[\eta_i(t)] = \mathbf{a}^T \mathbf{b}(t) h_t^2 (1 + o(1)), \tag{20}$$

$$V[\eta_i(t)] = \left(\mathbf{a}^T \boldsymbol{\Sigma}(\boldsymbol{\beta}(t), t) \mathbf{a}\right) \delta_2(t, h_t) h_t^{-1} (1 + o(1)), \tag{21}$$

and

$$E[|\eta_i(t) - E[\eta_i(t)]|^3] \le c_1(t) \delta_3(t, h_t) h_t^{-2} (1 + o(1)), \tag{22}$$

where $c_1(t)$ *is a non-negative constant.*

Proof See Appendix. □

From (19), Lemma 1, and the Cramér–Wold device, one can show the asymptotic normality of $\widehat{\boldsymbol{\beta}}_{h_t}(t)$ in the following theorem.

Theorem 1 *Suppose that the assumptions in Lemma* 1 *are satisfied and* $h_t = n^{-1/5} h_{t0}$ *for some positive constant* h_{t0}. *When* n *converges to infinity,*

$$(nh_t)^{1/2} \left(\widehat{\boldsymbol{\beta}}_{h_t}(t) - \boldsymbol{\beta}(t)\right) \xrightarrow{d} N_k\left(\mathbf{b}(t), \delta_2(t, h_t) \boldsymbol{\Sigma}(\boldsymbol{\beta}(t), t)\right). \tag{23}$$

When the recurrent event process is characterized via a non-stationary Poisson process or is assumed to be uniformly bounded by a positive constant without placing any distributional assumption, the asymptotic behavior of $\widehat{\boldsymbol{\beta}}_{h_t}(t)$ in Theorem 1 is still valid. Using the consistent estimators $\bar{\boldsymbol{\beta}}_{h_t}(t)$ and $\widehat{\mathrm{V}}(\bar{\boldsymbol{\beta}}_{h_t}(t), t)$ in (6) of $\boldsymbol{\beta}(t)$ and the covariance of $\widehat{\boldsymbol{\beta}}_{h_t}(t)$, one can construct an approximated $(1-\alpha)$ confidence interval with bias adjustment for each coefficient of $\boldsymbol{\beta}(t)$ via

$$\bar{\beta}_{j,h_t}(t) \pm Z_{1-\alpha}(\mathrm{V}(\bar{\boldsymbol{\beta}}_{h_t}(t), t)_{j,j})^{1/2}, \quad j = 1, \ldots, k, \tag{24}$$

where Z_p is the $100p$ percentile point of a standard normal distribution.

By expanding the exponential term $\exp(\bar{\boldsymbol{\beta}}_{h_t}^T(u)\mathbf{X}_i(u))$ with respect to $\boldsymbol{\beta}(u)$ and applying the law of large numbers, $\widehat{\lambda}_{0,h_t}$ can be expressed as

$$\begin{aligned}
\widehat{\lambda}_{0h_t}(t) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{T_0} \frac{Y_i(u)K_2(\frac{t-u}{h_t})\mathrm{d}N_i(u)}{E[Y(u)]\exp(\boldsymbol{\beta}^T(u)\mathbf{X}_i(u))}\left(1+O_p(n^{-1/2})\right) \\
&\quad - \int_0^{T_0} \lambda_0(u)\frac{\mathbf{xy}^T(u)}{E[Y(u)]}\left(\bar{\boldsymbol{\beta}}_{h_t}(u) - \boldsymbol{\beta}(u)\right) K_2\left(\frac{t-u}{h_t}\right)\mathrm{d}u\left(1+o_p(1)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^{T_0} \frac{Y_i(u)K_2(\frac{t-u}{h_t})\mathrm{d}N_i(u)}{E[Y(u)]\exp(\boldsymbol{\beta}^T(u)\mathbf{X}_i(u))}\left(1+O_p(n^{-1/2})\right) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}\int_0^{T_0} \lambda_0(u)\frac{\mathbf{xy}^T(u)}{E[Y(u)]}\varsigma_i(u)K_2\left(\frac{t-u}{h_t}\right)\mathrm{d}u\left(1+o_p(1)\right),
\end{aligned} \tag{25}$$

where $\mathbf{xy}(u) = E[Y(u)\mathbf{X}(u)]$ and $\varsigma_i(u)$ is same with the definition of $\xi_i(u)$ in (17) except that the fourth order boundary kernel function $K_4(\cdot)$ is substituted for $K_2(\cdot)$. Let

$$\begin{aligned}
\zeta_i(t) = -\lambda_0(t) &+ \int_0^{T_0} \frac{Y_i(u)K_2(\frac{t-u}{h_t})\mathrm{d}N_i(u)}{E[Y(u)]\exp(\boldsymbol{\beta}^T(u)\mathbf{X}_i(u))} \\
&- \int_0^{T_0} \lambda_0(u)\frac{\mathbf{xy}^T(u)}{E[Y(u)]}\varsigma_i(u)K_2\left(\frac{t-u}{h_t}\right)\mathrm{d}u.
\end{aligned}$$

One can re-express $\widehat{\lambda}_{0h_t}(t)$ in (25) as

$$\widehat{\lambda}_{0h_t}(t) - \lambda_0(t) = n^{-1}\sum_{i=1}^{n}\zeta_i(t)\left(1+o_p(1)\right). \tag{26}$$

Using the Berry–Esséen theorem and the moments of $\zeta_i(t)$ below, it is straightforward to show that, when n converges to infinity,

$$(nh_t)^{1/2}\left(\widehat{\lambda}_{0h_t}(t) - \lambda_0(t)\right) \xrightarrow{d} N\left(b_{\lambda_0}(t), \sigma^2_{\lambda_0 1}(t)\right). \tag{27}$$

Lemma 2 *Suppose that the assumptions in Lemma* 1 *are satisfied.*

$$E[\zeta_i(t)] = b_{\lambda_0}(t)h_t^2(1+o(1)), \tag{28}$$

$$V[\zeta_i(t)] = \sigma^2_{\lambda_0 1}(t)h_t^{-1}\left(1+o(1)\right), \tag{29}$$

and

$$E[|\zeta_i(t)-E[\zeta_i(t)]|^3] \le c_2(t)\delta_3(t,h_t)h_t^{-2}(1+o(1)), \tag{30}$$

where $c_2(t)$ *is a non-negative constant,* $b_{\lambda_0}(t) = \lambda_0^{(2)}(t)\gamma_{2,2}(t,h_t)$, *and*

$$\sigma^2_{\lambda_0 1}(t) = \frac{\lambda_0^2(t)}{(E[Y(t)])^2}\left(E\left[\frac{Y(t)}{\lambda_0(t)\exp(\boldsymbol{\beta}^T(t)\mathbf{X}(t))}\right] + \mathbf{xy}^T(t)\boldsymbol{\Sigma}(\boldsymbol{\beta}(t),t)\mathbf{xy}(t)\right)\delta_2(t,h_t)$$
$$+\frac{2\lambda_0^2(t)}{(E[Y(t)])^2}\mathbf{xy}^T(t)\boldsymbol{\Sigma}(\boldsymbol{\beta}(t),t)(\mathbf{xy}(t)-E[Y(t)]\bar{\mathbf{x}}(\boldsymbol{\beta}(t),t))\delta_3^*(t,h_t)$$

with $\delta_3^*(t,h_t) = \int_{(t-T_0)/h_t}^{t/h_t}\int_{(t-T_0)/h}^{t/h_t}\alpha_2(T_0,u-v)\alpha_2(T_0,u)\alpha_2(T_0,v)K(u-v)K(u)$ $K(v)\mathrm{d}u\mathrm{d}v$.

Since the proof for Lemma 2 is similar to that of Lemma 1, it is omitted here. Similarly, using the expression

$$\widetilde{\lambda}_{0h_t}(t) - \lambda_0(t) = n^{-1}\sum_{i=1}^{n}\eta_i(t)\left(1+o_p(1)\right) \tag{31}$$

with

$$\eta_i(t) = -\lambda_0(t) + \int_0^{T_0}\frac{Y_i(u)K_2(\frac{t-u}{h_t})\mathrm{d}N_i(u)}{E[Y(u)\exp(\boldsymbol{\beta}^T(u)\mathbf{X}(u))]}$$
$$-\int_0^{T_0}\lambda_0(u)\bar{\mathbf{x}}^T(\boldsymbol{\beta}(u),u)\varsigma_i(u)K_2\left(\frac{t-u}{h_t}\right)\mathrm{d}u,$$

the asymptotic normality of $\widetilde{\lambda}_{0h_t}(t)$ can be derived, see Theorem 2 below.

Theorem 2 *Suppose that the assumptions in Theorem 1 are satisfied. When n converges to infinity,*

$$(nh_t)^{1/2}\left(\widetilde{\lambda}_{0h_t}(t)-\lambda_0(t)\right) \xrightarrow{d} N\left(b_{\lambda_0}(t),\sigma^2_{\lambda_0 2}(t)\right), \tag{32}$$

where

$$\sigma^2_{\lambda_0 2}(t) = \left(\frac{\lambda_0(t)}{E[Y(t)\exp(\boldsymbol{\beta}^T(t)\mathbf{X}(t))]} + \lambda_0^2(t)\bar{\mathbf{x}}^T(\boldsymbol{\beta}(t),t)\boldsymbol{\Sigma}(\boldsymbol{\beta}(t),t)\bar{\mathbf{x}}(\boldsymbol{\beta}(t),t)\right)\delta_2(t,h_t).$$

We observe that both of $\widehat{\lambda}_{0h_t}(t)$ and $\widetilde{\lambda}_{0h_t}(t)$ have the same asymptotic bias but the different asymptotic variances. When the varying-coefficients $\boldsymbol{\beta}(t)$ are known, the dominant variances of $\widehat{\lambda}_{0h_t}(t)$ and $\widetilde{\lambda}_{0h_t}(t)$ in (27) and (32) will be reduced separately to

$$\sigma^{*2}_{\lambda_0 1}(t) = \frac{\lambda_0(t)\delta_2(t, h_t)}{(E[Y(t)])^2} E\left[\frac{Y(t)}{\exp(\boldsymbol{\beta}^T(t)\mathbf{X}(t))}\right](nh_t)^{-1} \tag{33}$$

and

$$\sigma^{*2}_{\lambda_0 2}(t) = \frac{\lambda_0(t)\delta_2(t, h_t)}{E\left[Y(t)\exp(\boldsymbol{\beta}^T(t)\mathbf{X}(t))\right]}(nh_t)^{-1}. \tag{34}$$

By using the Hölder's inequality, we get

$$E\left[\frac{Y(t)}{\exp(\boldsymbol{\beta}^T(t)\mathbf{X}(t))}\right] E\left[Y(t)\exp(\boldsymbol{\beta}^T(t)\mathbf{X}(t))\right] \geq (E[Y(t)])^2,$$

and, hence, $\sigma^{*2}_{\lambda_0 1}(t) \geq \sigma^{*2}_{\lambda_0 2}(t)$. Since the varying-coefficients $\boldsymbol{\beta}(t)$ are unknown and need to be estimated, there is no apparent superiority for either estimators. This phenomenon will also be indicated in the next simulation study. For the consideration of efficiency in implementation, the estimator $\widehat{\lambda}_{0h_t}(t)$ takes this advantage and is slightly faster in computation speed.

4 Monte Carlo simulation

The simulated recurrent event data are generated from 500 independent individuals $\{(N_i(\cdot), Y_i, X_i(\cdot))\}_{i=1}^{500}$ with

$$E_i[\mathrm{d}N_i(t)|X_i(t)] = Z_i(t)\phi_0(t)\exp(\beta(t)X_i(t))\mathrm{d}t, \quad t \in [0, 5].$$

Here, the covariate $X(t)$ and the random variable $Z(t)$ are set separately to be time independent random variables, say, X and Z with the Bernoulli distribution $Bernoulli(0.5)$ and the uniform distribution $U(0.9, 1.1)$. Moreover, $\phi_0(t)$ and $\beta(t)$ are assigned to be $2.5 + (t-3)^3/27$ and $-0.3\sqrt{t}$ in the above model. Under the independent censoring assumption (A1) and the model setting, the baseline effect $\lambda_0(t)$ can be derived to be $\phi_0(t)$. As for the censoring time, conditioning on the covariate X, it is designed to be distributed as the truncated exponential density

$$f_{Y|X}(y|x) = \frac{2\exp(2y)}{\exp(10) - \exp(2+x)}, \quad y \in [1 + 0.5x, 5].$$

The recurrent event data are repeatedly generated 1,000 times. For each simulated data, the smoothing estimators $\widehat{\beta}_{h_t}(t)$, $\widehat{\lambda}_{0h_t}(t)$ and $\widetilde{\lambda}_{0h_t}(t)$ are computed separately by the solution of $U(\beta(t); h_t)$ in (5), (10) and (13). In this numerical study, $\alpha_2(T_0, \frac{t-u}{h_t})$

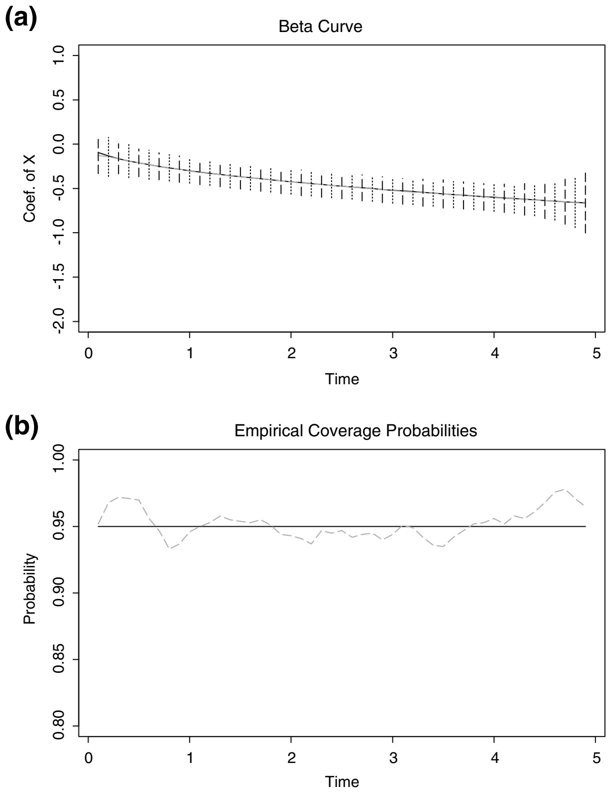

Fig. 1 **a** The curve $\beta(t)$ (*solid curve*), the averages of 500 estimated curves $\widehat{\beta}_h(t)$ (*dashed curve*), ± 1.96 standard error bars (*dashed line*) of $\widehat{\beta}_h(t)$, and ± 1.96 averages of 500 estimated standard error bars (*dotted line*) of $\widehat{\beta}_h(t)$. **b** The nominal level (*solid line*) and the empirical coverage probability curve (*dashed curve*)

and $K(\cdot)$ are assigned to be the linear function and the normal density. For the bandwidth selection, the local bandwidth selection criterion for $\widehat{\beta}_{h_t}(t)$ is obtained via minimizing the estimated mean squared errors of $\widehat{\beta}_{h_t}(t)$:

$$\widehat{\text{MSE}}(\widehat{\beta}_{h_t}(t)) = \widehat{b}^2(t)h_t^4 + V(\bar{\beta}_{h_t}(t), t). \tag{35}$$

Moreover, an approximated 95% confidence interval for $\beta(t)$ can be constructed via (24).

Figure 1a reveals the true curve, the averages and ± 1.96 standard error bars of 1,000 estimated curves, and ± 1.96 averages of 1,000 standard error curves, respectively. It appears that the estimated curves are very close to the corresponding true curves. The empirical coverage probabilities for $\beta(t)$ and the horizontal line of 0.95 nominal level are shown in Fig. 1b. We can see that the estimated coverage probabilities are around the nominal level with the average of 0.952 at 49 equally spaced time points. In Fig. 2, no apparent difference in the standard errors of $\widehat{\lambda}_{0h_t}(t)$ and $\widetilde{\lambda}_{0h_t}(t)$ is found. However, the computation speed of $\widehat{\lambda}_{0h_t}(t)$ is slightly faster.

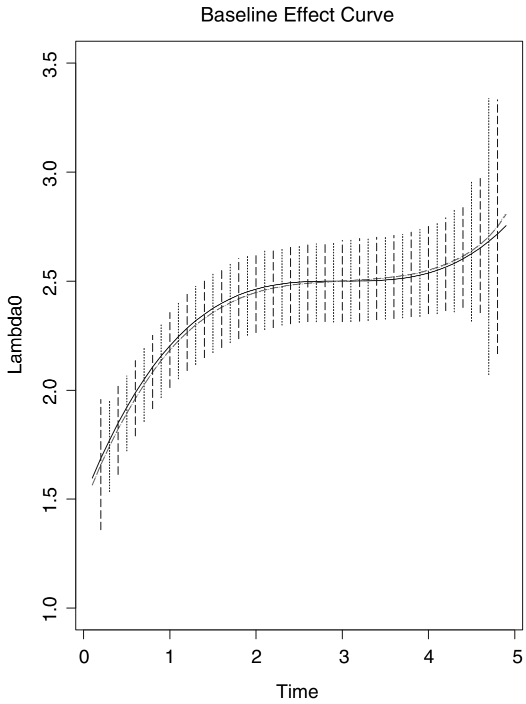

Fig. 2 The baseline effect curve $\lambda_0(t)$ (*solid curve*), the averages and ± 1.96 standard error bars of 1,000 estimated curves $\widehat{\lambda}_{0h}(t)$ (*dotted curve*) and $\widetilde{\lambda}_{0h}(t)$ (*dashed curve*)

5 Application to the ALIVE cohort study

In this section, we apply the more flexible varying-coefficient model (2) to characterize the influence of some demographical variables on the repeated hospitalizations for HIV-negative intraveneous drug users. The proposed smoothing estimation procedures are used to evaluate the time-evolution effects of these variables. Here, the analyzed recurrent event data are mainly collected from the Baltimore site of the ALIVE cohort study, which involves 451 HIV-negative intravenous drug users who entered study before August 1, 1993. Details of this study can be found in the work of Vlahov et al. (1991).

Measurements considered here for each patient consist of the time lengths from August 1, 1993 to the dates of inpatient admissions and to the last visit, race indicator for black and non-black people, gender, and age on August 1, 1993. There are about 92% black and 68.3% male patients aged 19.93 to 68.22 with the median of the age being 39.11. The age group, which is defined to be 0 if he or she is younger than the median entering age of 39.11 years old and 1 otherwise, will be used in the analysis. Since there are very low proportion of non-black people among the patients, this variable will not be used in the succeeding analysis. The objective here is to detect the effects of gender and age on the hospitalization rate.

Based on the varying-coefficient model (2), the time effects of gender and age on the hospitalization rate at each time point can be computed separately by the solution of $U(\beta(t); h_t)$ in (5). As in the simulation study, the bandwidth selection procedure (35)

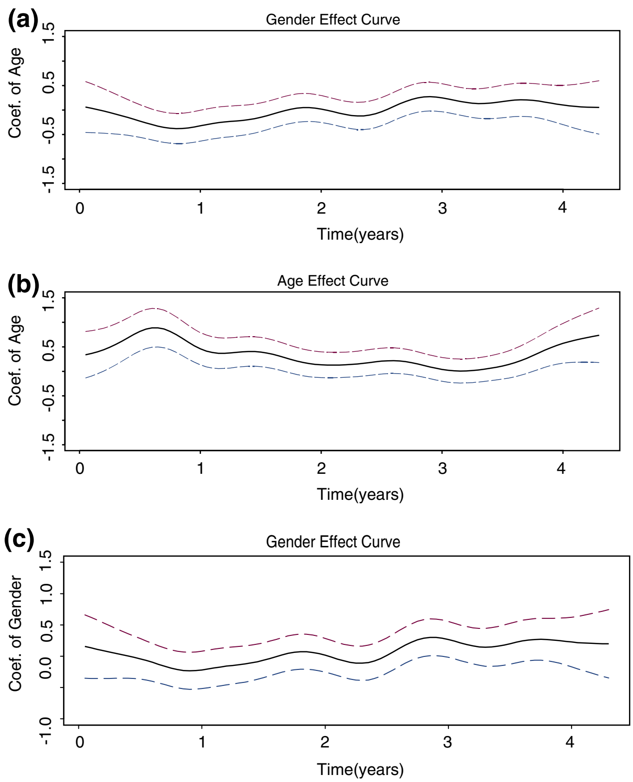

Fig. 3 The *solid curves* represent the estimated effects and the *dashed curves* show the corresponding 95% pointwise confidence intervals

and the approximated confidence interval (24) are used to select local bandwidths for the estimators and to construct a confidence interval for each effect curve. Figure 3a,b show the estimated curves and the corresponding 95% confidence intervals. It is indicated from Fig. 3a that no significant effect is detected for gender. However, in Fig. 3b, an older age group tends to have a positive association with the hospitalization rate and appears to be significant roughly at the first 1.65 years since August 1, 1993 and a half year before the end of study.

Acknowledgments The first author's research was partially supported by the National Science Council grant 93-2118-M-002-001- (Taiwan), and the second author's research was partially supported by the National Institute of Health grants R01 HD38209 and R01 MH56639 (USA). We would like to thank the referee for valuable comments. We are also grateful to Dave Vlahov and Strathdee at Johns Hopkins University for providing the anonymous ALIVE data. provision of the data was supported by National Institute on Drug Abuse grants DA04334 and DA08009.

Appendix

Proof of Lemma 1

Let $\mathbf{a}^{*T}(t) = \mathbf{a}^T \boldsymbol{\Sigma}(\boldsymbol{\beta}(t), t)$. The expectation of $E[\eta_i(t)]$ can be expressed as

$$
\begin{aligned}
E[\eta_i(t)] &= -\mathbf{a}^{*T}(t)E\left[\int_0^{T_0} \mathbf{Z}(\boldsymbol{\beta}(t),u)Y(u)K_2\left(\frac{t-u}{h_t}\right)\mathrm{d}N(u)\right] \\
&= -\mathbf{a}^{*T}(t)\int_0^{T_0} \lambda_0(u)E[\mathbf{Z}(\boldsymbol{\beta}(t),u)Y(u)\exp(\boldsymbol{\beta}^T(u)\mathbf{X}(u))]K_2\left(\frac{t-u}{h_t}\right)\mathrm{d}u
\end{aligned} \tag{36}
$$

By the Taylor expansion and assumptions (A2)–(A4), one can derive that

$$
\begin{aligned}
&\int_0^{T_0} \lambda_0(u)E[\mathbf{Z}(\boldsymbol{\beta}(t),u)Y(u)\exp(\boldsymbol{\beta}^T(u)\mathbf{X}(u))]K_2\left(\frac{t-u}{h_t}\right)\mathrm{d}u \\
&= \int_{t-T_0/h_t}^{t/h_t} \lambda_0(t-h_tv)E[\mathbf{Z}(\boldsymbol{\beta}(t),t-h_tv)Y(t-h_tv) \\
&\quad \times \exp(\boldsymbol{\beta}^T(t-h_tv)\mathbf{X}(t-h_tv))]K_2^*(v)\mathrm{d}v \\
&= \int_{t-T_0/h_t}^{t/h_t} E[\mathbf{Z}(\boldsymbol{\beta}(t),t-h_tv)Y(t-h_tv) \\
&\quad \times \exp(\boldsymbol{\beta}^T(t)\mathbf{X}(t-h_tv))(1-\boldsymbol{\beta}^{(1)T}(t)\mathbf{X}(t-h_tv)hv \\
&\quad + \frac{(\boldsymbol{\beta}^{(1)T}(t)\mathbf{X}(t-h_tv))^2+\boldsymbol{\beta}^{(2)T}(t)\mathbf{X}(t-h_tv)}{2}(h_tv)^2)] \\
&\quad \times (\lambda_0(t)-\lambda_0^{(1)}(t)h_tv)K_2^*(v)\mathrm{d}v \\
&\quad \cdot(1+o(1)) \\
&= h_t\int_{t-T_0/h_t}^{t/h_t} (\lambda_0(t)-\lambda_0^{(1)}(t)h_tv\mathbf{M}_{11}(\boldsymbol{\beta}(t),t-h_tv)vK_2^*(v)\mathrm{d}v(1+o(1)) \\
&\quad + h_t^2\lambda_0(t)\int_{t-T_0/h_t}^{t/h_t} \frac{\mathbf{M}_{21}(\boldsymbol{\beta}(t),t-h_tv)+\mathbf{M}_{12}(\boldsymbol{\beta}(t),t-h_tv)}{2} \\
&\quad \times v^2K^*(v)\mathrm{d}v(1+o(1)) \\
&= (\lambda_0^{(1)}(t)\mathbf{M}_{11}(\boldsymbol{\beta}(t),t)+\lambda_0(t)\left(\mathbf{M}_{11}^{(1)}(\boldsymbol{\beta}(t),t)+\frac{\mathbf{M}_{21}(\boldsymbol{\beta}(t),t)+\mathbf{M}_{12}(\boldsymbol{\beta}(t),t)}{2}\right) \\
&\quad \cdot\gamma_{2,2}(t,h_t)h_t^2,
\end{aligned} \tag{37}
$$

where $K_2^*(v)=\alpha_2(T_0,v)K(v)$. Substituting (37) into (36), (20) is then obtained.

For the variance of $\eta_i(t)$, it can be derived by considering $E[\eta_i^2(t)]$ which is shown to be

$$
\begin{aligned}
E[\eta_i^2(t)] &= \mathbf{a}^{*T}(t)E\left[\int_0^{T_0} \mathbf{Z}^{\otimes 2}(\boldsymbol{\beta}(t),u)Y(u)K_2^2\left(\frac{t-u}{h_t}\right)\mathrm{d}N(u)\right]\mathbf{a}^*(t) \\
&\quad + \mathbf{a}^{*T}(t)E\left[\int_{u\neq v} \mathbf{Z}(\boldsymbol{\beta}(t),u)Y(u)\mathbf{Z}^T(\boldsymbol{\beta}(t),v)Y(v)\lambda_{\mathbf{X}}(u,v)K_2\left(\frac{t-u}{h_t}\right)\right.
\end{aligned}
$$

$$\times K_2\left(\frac{t-v}{h_t}\right)\mathrm{d}u\mathrm{d}v\Bigg]\mathbf{a}^*(t)$$
$$= \mathbf{a}^{*T}(t)E\left[\int_0^{T_0}\mathbf{Z}^{\otimes 2}(\boldsymbol{\beta}(t),u)Y(u)K_2^2\left(\frac{t-u}{h_t}\right)\mathrm{d}N(u)\right]\mathbf{a}^*(t)$$
$$+\mathbf{a}^{*T}(t)\rho(t,t)\mathbf{a}^*(t), \tag{38}$$

where $E[\mathrm{d}N(u)\mathrm{d}N(v)|\mathbf{X}(u),\mathbf{X}(v)] = \lambda_{\mathbf{X}}(u,v)\mathrm{d}u\mathrm{d}v$ and $\rho(t,t) = E[\int_{u\neq v}\mathbf{Z}(\boldsymbol{\beta}(t), t-h_tu)Y(t-h_tu)\mathbf{Z}^T(\boldsymbol{\beta}(t),t-h_tv)Y(t-h_tv)\lambda_{\mathbf{X}}(t-h_tu,t-h_tv)]K_2(u)K_2(v)\mathrm{d}u\mathrm{d}v$. By assumptions (A5)–(A6), the first term in (38) cab be derived to be

$$\mathbf{a}^{*T}(t)E\left[\int_0^{T_0}\mathbf{Z}^{\otimes 2}(\boldsymbol{\beta}(t),u)Y(u)K_2^2\left(\frac{t-u}{h_t}\right)\mathrm{d}N(u)\right]\mathbf{a}^*(t)$$
$$= \mathbf{a}^{*T}(t)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}(t),t)\mathbf{a}^*(t)\delta_2(t,h_t)h_t^{-1}(1+o(1)), \tag{39}$$

and thus,

$$E[\eta_i^2(t)] = \mathbf{a}^{*T}(t)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}(t),t)\mathbf{a}^*(t)\delta_2(t,h_t)h_t^{-1}(1+o(1)). \tag{40}$$

From (40) and (20), on can show that (21) holds. Finally, along the same lines as the derivation of $E[\eta_i^2(t)]$, the statement of (22) follows.

References

Andersen, P. K., Borgan, O., Gill, R. D., Keiding, N. (1993). *Statistical models based on counting processes*. 76. New York: Springer.

Cai, Z., Sun, Y. (2003). Local linear estimation for time-dependent coefficients in Cox's regression models. *Scandinavian Journal of Statistics, 30*, 93–111.

Gasser, Th., Müller, H.-G. (1978). *Kernel estimation of regression functions. smoothing techniques for curve estimation* In: Th. Gasser, M. Rosenblatt, (Eds.), *Spring Lecture Notes in Mathematics* 757 (pp. 23–68). Berlin: Springer.

Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association, 87*, 942–951.

Lancaster, T., Intrator, O. (1998). Panel data with survival: hospitalization of HIV-positive patients. *Journal of the American Statistical Association*, *93*, 46–53.

Lawless, J. F., Nadeau, C. (1995). Some simple robust method for the analysis of recurrent events. *Technometrics, 37*, 158–168.

Lin, D.Y., Wei, L. J., Yang, I., Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society, B62*, 711–730.

Martinussen, T., Scheike, T. H. (2002). A flexible additive multiplicative hazard model. *Biometrika, 89*, 283–298.

Murphy, S., Sen, P. (1991). Time-dependent coefficients in a Cox-type regression model. *Stochastic Processes an Their Applications, 39*, 153–180.

Nelson, W. B. (1995). Confidence limits for recurrence data-applied to cost or number of product repairs. *Technometrics, 37*, 147–157.

Pepe, M. S., Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association*, *88*, 811–820.

Schucany, W. R. (1995). Adaptive bandwidth choice for kernel regression. *Journal of the American Statistical Association, 90*, 535–540.

Tian, L., Zucker, D., Wei, L. J. (2005). On the Cox model with time-varying regression coefficients. *Journal of the American Statistical Association, 100*, 172–183.

Wang, M.-C., Qin, J., Chiang, C.-T. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association, 96*, 1057–1065.

Vlahov, D., Anthony, J. C., Muñov, A., Margolick J., Nelson K. E., Celentano D. D., Solomon, L., Polk, B. F. (1991). The ALIVE study: a longitudinal study of HIV-1 infection in intravenous drug users: description of methods. *The Journal of Drug Issues, 21*, 759–776.

Winnett, A., Sasieni, P. (2003). Iterated residuals and time-varying covariate effect in Cox regression. *Journal of the Royal Statistical Society, B65*, 473–488.

Zuuker, D. M., Karr, A. F. (1990). Nonparametric survival analysis with time-dependent covariates: a penalized partial likelihood approach. *The Annals of Statistics, 18*, 329–353.