Statistics in Medicine

Research Article

Received 25 February 2015, Accepted 26 April 2016 Published online 29 May 2016 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6995

Time-varying coefficients models for recurrent event data when different varying coefficients admit different degrees of smoothness: application to heart disease modeling

Ehsan Eshaghi, Hossein Baghishani*† and Davood Shahsavani

We consider a class of semiparametric marginal rate models for analyzing recurrent event data. In these models, both time-varying and time-free effects are present, and the estimation of time-varying effects may result in non-smooth regression functions. A typical approach for avoiding this problem and producing smooth functions is based on kernel methods. The traditional kernel-based approach, however, assumes a common degree of smoothness for all time-varying regression functions, which may result in suboptimal estimators if the functions have different levels of smoothness. In this paper, we extend the traditional approach by introducing different bandwidths for different regression functions. First, we establish the asymptotic properties of the suggested estimators. Next, we demonstrate the superiority of our proposed method using two finite-sample simulation studies. Finally, we illustrate our methodology by analyzing a real-world heart disease dataset. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: consistency; convergence rate; different bandwidths; kernel function; recurrent event data; semiparametric model

1. Introduction

In many research disciplines, such as medicine, engineering, and environmental sciences, the data collected for statistical analysis include interesting events that repeatedly occur in time. Such events are called recurrent events, and the corresponding data are called recurrent event data [1]. Disease relapses, recurrent opportunistic infections in human immunodeficiency virus patients [2], repeated transient ischemic attacks in patients with cerebrovascular disease [3], recurrent pyogenic infections in chronic granulomatous disease [4–6], repeated asthma attacks in children [7, 8], and repeated failures in software systems [9, 10] or industrial equipment [11] are examples of such events. Other examples include tumor metastases, myocardial infarctions, and heartbeat patterns in patients with arrhythmias, who may experience too fast, too slow, or irregular heartbeat patterns. These changes in heartbeat rhythm can happen suddenly and unexpectedly and sometimes may result in the death of the patient.

In regression models for analyzing survival data, it is common to assume that the covariates are independent of time. However, this assumption is often violated in recurrent event data. Therefore, evaluating the temporal effects of the covariates is of interest in these applications. Survival semiparametric regression models provide a flexible method to analyze such data [1].

1.1. Related works

Several semiparametric conditional [12, 13] and marginal [14] regression models have been proposed to analyze recurrent event data. Researchers widely use conditional hazard regression models [15] to

Department of Statistics, Shahrood University of Technology, Shahrood, Iran
**Correspondence to: Hossein Baghishani, Department of Statistics, Shahrood University of Technology, Shahrood, Iran.*
†E-mail: hbaghishani@shahroodut.ac.ir

describe the dependency of survival times on covariates. The Cox proportional hazard model [16] is a well-known example of such models. In practice, using the mean or total number of event occurrences is often more interpretable than hazard rate for analysis of recurrent event data. Murphy and Sen [17], Hastie and Tibshirani [18], Grambsch and Thearneau [19], and Fahrmeir and Klinger [20] studied the time-varying intensity models. Some authors have also used regression models for the mean and rate functions [5, 21–23]. Lin *et al.* [5] used a Cox-type link function to analyze marginal mean and rate models. Martinussen *et al.* [24] proposed an efficient procedure for estimating the time-varying and time-constant effects in a general semiparametric multiplicative intensity model. Zucker and Karr [25] proposed a penalized partial likelihood approach for fitting a Cox regression type model with time-varying regression coefficients for right censored data. Cai and Sun [26] and Tian *et al.* [27] used a local partial likelihood technique for this model. Amorim *et al.* [28] used regression splines to fit time-dependent coefficient models for the recurrent event data. Recently, Sun *et al.* [6] proposed a model with a mixture of time dependent and invariant coefficients in which the non-parametric component of the model is estimated by using kernel methods. In their proposed approach, a common smoothing parameter is used to estimate baseline hazard and all time-varying regression coefficients.

1.2. Motivation

In practice, each time-varying coefficient has its own degree of smoothness. Therefore, to obtain the optimal estimators, in the general case, it is necessary to use different bandwidths for different coefficients. This becomes especially important, and the gain could be substantial, as the number of functions that need to be estimated increases. Fan and Zhang [29] proposed a two-step estimating procedure for a linear-varying coefficient model with different bandwidths. However, their work suffers from a few limitations. First, they developed their approach for linear models only. Second, for p time-varying regression coefficients, $p \geqslant 2$, they assumed that only one coefficient is smoother than the rest, and all other functions have equal, smaller degrees of smoothness. They demonstrated that considering these two different bandwidths is always more reliable than considering a fixed bandwidth for all functions. Nevertheless, in general, this is still a limitation if the functions have several different degrees of smoothness. Third, in real applications, we typically do not know in advance which regression function is smoother than the others.

We develop the work of Sun *et al.* [6] by considering different smoothing parameters for different functions in a kernel-based approach. Our proposed method avoids the limitations of Fan and Zhang [29] but retains its strong advantages: each time-varying coefficient has its own degrees of smoothness, and the estimation method could be extended to a broad class of survival semiparametric models. We also show that all coefficients of the model can be estimated consistently at the rate $\sqrt{n}$, where n is the sample size.

1.3. Outline

This paper is organized as follows. In Section 2, we describe the heart disease dataset used in the analysis. Section 3 presents the proposed method: We consider different bandwidths for different regression functions, and incorporate Taylor expansions and kernel methods into our estimation procedure for the resulting model, which leads to improved performance. We also present a simple iterative algorithm for estimating the proposed model and an adaptive algorithm for selecting the optimal bandwidth. Section 4 provides our theoretical results, establishing the consistency and asymptotic normality of our estimators. In Section 5, we evaluate the performance of our approach using two simulation examples. In Section 6, we apply our method to a real-world application in analyzing heart disease data from Mashhad, Iran. Section 7 concludes the paper with a discussion. Finally, two technical appendices summarize the notation used in our theoretical results, as well as the missing proofs.

2. Heart disease data

Heart disease is the number one cause of mortality around the world. This disease is the third leading causes of burden of disease in 2030, and also, it kills more people than cancer in the world [30]. Consequently, improving therapies to reduce the mortality rate has been an important research topic. One of the treatments for patients with a high risk of heart diseases, particularly heart failure, is heart defibrillator. This device is a small battery-powered impulse generator that is implanted in patients who are at risk of sudden death due to a dangerously fast heartbeat (ventricular fibrillation) or a chaotic heartbeat (ventricular tachycardia). Heart defibrillators work by detecting irregular rhythms and intervening abnormal heartbeats: These devices continuously monitor the patients' heartbeat and deliver extra beats or electrical shocks to restore a normal heart rhythm when necessary. In this study, we consider two types of heart defibrillators: implantable cardioverter defibrillator (ICD) and cardiac resynchronization therapy defibrillator (CRTD). Depending on what the heart needs, an ICD can have one or two wires that called leads. One lead goes in the right ventricle (a single-chamber type of ICD), and if the person needs a second lead, it will be placed in the right atrium (a two-chamber type of ICD). A CRTD system adds a third, attaching a lead to the left ventricle so it can help both sides beat in synch and thus pump more efficiently.

In this work, we studied heart defibrillator-implanted patients with heart disease, in Ghaem Hospital of Mashhad[‡], affiliated with Mashhad University of Medical Sciences. The dataset was collected by the clinical research center of Ghaem hospital between 2005 and 2008. The data includes a total of 44 patients, followed until death or censoring. The patients in this study were, on average, 48 years old, with 40% of them more than 52 years old.

Before implanting the heart defibrillator, the duration of ventricular depolarization (the so-called QRS duration) was measured in the patients' electrocardiogram with a digital caliper with an accuracy of 1 mm. In order to increase the accuracy of the study, this duration was measured twice, and the average was recorded. The duration, amplitude, and morphology of the QRS are useful in diagnosing cardiac arrhythmias, conduction abnormalities, ventricular hypertrophy, myocardial infarction, electrolyte derangements, and other states of heart disease [31]. The following information were recorded for each patient: age, sex, type of decision for embedded heart defibrillator (primary or secondary prevention), and a family history of sudden cardiac death in first-degree relatives. Before implanting the heart defibrillator, the QTc interval duration in the electrocardiogram of patients was calculated by the Bazzet formula [32]. After embedding the heart defibrillator, patients were regularly followed up with for 1, 3, and 6 months. In follow-up visits, patients were reviewed by an electrophysiologist, and the cases of arrhythmias were recorded by the software. We focus on the patients' age, QRS duration, and type of heart defibrillators as the covariates for modeling the rate at which the defibrillator device delivers electrical shocks to the heart of each patient.

Treatment depends on the type and severity of arrhythmia in each patient. In some cases, no treatment is necessary. Treatment options include medications, lifestyle changes, invasive therapies, electrical devices, or surgery. In this paper, we apply our new method to the aforementioned dataset. We show that given the QRS duration, the patients' age, and the type of the heart defibrillator device that is implanted in that patient, we can infer the usefulness or harmfulness of the given heart defibrillator device for the patient in question, and can decide whether to implant this type of heart defibrillator for other, new patients or not.

3. Methodology

Consider n subjects that are observed over time. Let $\tilde{N}_i(t)$ and $N_i(t)$ denote the number of failures and observable events of an individual over the interval $(0, t]$, respectively. Suppose that $N_i(t) = \tilde{N}_i(t \wedge C_i)$ where $a \wedge b = \min(a, b)$, and C_i denotes the follow-up or censoring time. Because of the censoring, $N_i(\cdot)$ may be less than $\tilde{N}_i(\cdot)$. We also let $\mathbf{X}_i(\cdot)$ and $\mathbf{Z}_i(\cdot)$ denote the vectors of covariate processes with dimensions p and q, respectively. The at-risk process for each subject is denoted by $Y_i(t) = I(C_i \geqslant t)$, $i = 1, \ldots, n$, where $I(\cdot)$ is the indicator function.

We consider the following marginal regression model with time-varying coefficients [6]:

$$E\left\{d\tilde{N}_i(t)|\mathbf{X}_i(t), \mathbf{Z}_i(t)\right\} = \exp\left\{\boldsymbol{\beta}_0(t)^T\mathbf{X}_i(t) + \boldsymbol{\omega}_0^T\mathbf{Z}_i(t)\right\} d\lambda_0(t), \tag{1}$$

where $\boldsymbol{\beta}_0(t)$ and $\boldsymbol{\omega}_0$ are unknown p and q-vector of time-varying and time-independent coefficients, respectively. Also, $\lambda_0(t)$ is an unknown continuous baseline mean function. To derive more stable estimators, it is preferable to use the cumulative regression functions $\mathbf{B}_0(t) = \int_0^t \boldsymbol{\beta}_0(u)du$ [24, 33]. The functions $\mathbf{B}_0(t)$ and the regression parameter $\boldsymbol{\omega}_0$ can be estimated consistently at the rate $n^{1/2}$ and further lead to a uniform asymptotic description of the estimators, which is needed when one is curious to examine hypotheses about $\boldsymbol{\beta}_0(t)$ [24, 33]. If one is interested in $\boldsymbol{\beta}_0(t)$ directly, methods such as kernel estimation may be used based on the estimate of $\mathbf{B}_0(t)$ [6].

[‡]*Mashhad is the second largest city in Iran, located in the north-east of the country*

Statistics in Medicine

We define $\mathbf{N}(t) = (N_1(t), \ldots, N_n(t))^T$, $N_0(t) = n^{-1}\sum_{i=1}^n N_i(t)$, $X(t) = (\mathbf{X}_1(t), \ldots, \mathbf{X}_n(t))^T$, and $Z(t) = (\mathbf{Z}_1(t), \ldots, \mathbf{Z}_n(t))^T$. We also define the counting process by

$$M_i(t) = N_i(t) - \int_0^t Y_i(u) E\left\{d\tilde{N}_i(u)|\mathbf{X}_i(u), \mathbf{Z}_i(u)\right\}. \tag{2}$$

Under model (1), $M_i(t)$ is a zero-mean stochastic process. For given $(\boldsymbol{\beta}(t), \boldsymbol{\omega})$, we can estimate $\lambda_0(t)$ by the Breslow estimator

$$\hat{\lambda}_0(t; \boldsymbol{\beta}, \boldsymbol{\omega}) = \int_0^t S_0(u; \boldsymbol{\beta}, \boldsymbol{\omega})^{-1} dN_0(u), \tag{3}$$

where $S_0(u; \boldsymbol{\beta}, \boldsymbol{\omega}) = \frac{1}{n}\sum_{i=1}^n \phi_i(u; \boldsymbol{\beta}, \boldsymbol{\omega})$ and $\phi_i(t; \boldsymbol{\beta}, \boldsymbol{\omega}) = Y_i(t)\exp\left\{\boldsymbol{\beta}_0(t)^T\mathbf{X}_i(t) + \boldsymbol{\omega}_0^T\mathbf{Z}_i(t)\right\}$.

3.1. Estimation

To estimate $\boldsymbol{\beta}_0(t)$ and $\boldsymbol{\omega}_0$, by using generalized estimating equation methods [34] and substituting (3) back into (2), we have

$$X(t)^T d\mathbf{M}(t) = 0; \quad 0 \leqslant t \leqslant \tau, \qquad \int_0^\tau Z(t)^T d\mathbf{M}(t) = 0,$$

where $d\mathbf{M}(t) = (dM_1(t), \ldots, dM_n(t))$, and τ is a prespecified constant. For estimating $\boldsymbol{\beta}_0(t)$ and $\boldsymbol{\omega}_0$, a Newton–Raphson algorithm around the current estimate $(\boldsymbol{\beta}^l(t), \boldsymbol{\omega}^l)$ is then applied. Thus, the updating equations are

$$\begin{aligned}&\left\{\left(\boldsymbol{\beta}^{l+1}(t) - \boldsymbol{\beta}^l(t)\right) + \Lambda_{xx}^l(t)^{-1}\Lambda_{xz}^l(t)\left(\boldsymbol{\omega}^{l+1} - \boldsymbol{\omega}^l\right)\right\} S_0^l(t)^{-1} dN_0(t)\\ &\quad = n^{-1}\Lambda_{xx}^l(t)^{-1}\acute{X}^l(t)^T d\mathbf{N}(t),\end{aligned} \tag{4}$$

and

$$\begin{aligned}&\int_0^\tau \left\{\Lambda_{zz}^l(t)\left(\boldsymbol{\omega}^{l+1} - \boldsymbol{\omega}^l\right) + \Lambda_{zx}^l(t)\left(\boldsymbol{\beta}^{l+1}(t) - \boldsymbol{\beta}^l(t)\right)\right\} S_0^l(t)^{-1} dN_0(t)\\ &\quad = n^{-1}\int_0^\tau \acute{Z}^l(t)^T d\mathbf{N}(t),\end{aligned} \tag{5}$$

where $\Lambda_k(t; \boldsymbol{\beta}, \boldsymbol{\omega})$, $\Lambda_k^l(t) = \Lambda_k(t; \boldsymbol{\beta}^l, \boldsymbol{\omega}^l)(k = xx, xz, zx, zz)$, $S_0^l(t)$, $\acute{X}^l(t)$, and $\acute{Z}^l(t)$ are defined in Appendix A. Substituting (4) into (5) and solving for $\boldsymbol{\omega}^{l+1}$, we obtain the iteration step $\boldsymbol{\omega}^{l+1} = \Upsilon(\boldsymbol{\omega}^l)$, where

$$\Upsilon(\boldsymbol{\omega}^l) = \boldsymbol{\omega}^l + \frac{A^l(\tau)^{-1}}{n}\int_0^\tau \left[\acute{Z}^l(t)^T - \Lambda_{zx}^l(t)\Lambda_{xx}^l(t)^{-1}\acute{X}^l(t)^T\right] d\mathbf{N}(t), \tag{6}$$

$A^l(\tau) = A(\tau; \boldsymbol{\beta}^l, \boldsymbol{\omega}^l)$ and

$$A(\tau; \boldsymbol{\beta}, \boldsymbol{\omega}) = \int_0^\tau \left[\Lambda_{zz}(t; \boldsymbol{\beta}, \boldsymbol{\omega}) - \Lambda_{zx}(t; \boldsymbol{\beta}, \boldsymbol{\omega})\Lambda_{xx}(t; \boldsymbol{\beta}, \boldsymbol{\omega})^{-1}\Lambda_{zx}(t; \boldsymbol{\beta}, \boldsymbol{\omega})^T\right] \frac{dN_0(t)}{S_0(t; \boldsymbol{\beta}, \boldsymbol{\omega})}.$$

As Martinussen *et al.* [24] have noted, it is not a good idea to try to iterate towards a solution for each time point t, because the information about any particular time point is limited and even consistency cannot be obtained. To stabilize the solution, smoothing is required.

For simplicity, as discussed in Sun *et al.* [6], $\beta^l(t)$ and $\theta_0^l(t)$ are taken to be a simple kernel estimators of $\boldsymbol{\beta}_0(t)$ and $\theta_0(t) = d\lambda_0(t)/dt$ based on $\mathbf{B}^l(t)$ and $\lambda_0^l(t)$, respectively where $\mathbf{B}^l(t)$ is the estimate of $\mathbf{B}(t)$ after the lth iteration and $\lambda_0^l(t) = \hat{\lambda}\left(t; \boldsymbol{\beta}^l, \boldsymbol{\omega}^l\right)$. Sun *et al.* [6] proposed an estimation method by which only one common smoothing parameter h is used to fit the baseline hazard and all time-varying coefficient functions. However, the estimation of these functions may be further improved by using different degrees of smoothness. In the following section, we propose our method for smoothing time-varying coefficients while allowing each coefficient to have its own bandwidth.

3.2. *Different bandwidths*

Let $(h_1, h_2, \ldots, h_p)$ and h_0 be the bandwidths for $\boldsymbol{\beta}_0(t) = (\beta_{01}(t), \beta_{02}(t), \ldots, \beta_{0p}(t))$ and $\theta_0(t)$, respectively. We define

$$\mathbf{K} = \begin{pmatrix} K\left(\frac{u-t}{h_1}\right) & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & K\left(\frac{u-t}{h_p}\right) \end{pmatrix},$$

and

$$H = \begin{pmatrix} h_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & h_p \end{pmatrix},$$

where $\mathbf{K}$ is a diagonal matrix of symmetric kernel functions with different bandwidths. We now enforce smoothness for the underlying regression coefficients through the estimation of $\boldsymbol{\beta}^l(t)$ and also $\theta_0(t)$, that is,

$$\boldsymbol{\beta}^l(t) = \int H^{-1}\mathbf{K}d\mathbf{B}^l(u), \tag{7}$$

and

$$\theta_0^l(t) = \int h_0^{-1}K\left(\frac{u-t}{h_0}\right) d\lambda_0^l(u).$$

Using the updated coefficients $\boldsymbol{\omega}^{l+1}$ and replacing $dN_0(t)/S_0^l(t)$ in (4), we have $\mathbf{B}^{l+1}(t) = \Psi(\mathbf{B}^l)(t)$, where

$$\begin{aligned} \Psi(\mathbf{B}^l)(t) = &\int_0^t \boldsymbol{\beta}^l(u)du + n^{-1}\int_0^t \theta_0^l(u)^{-1}\Lambda_{xx}^l(u)^{-1}\acute{X}^l(u)^T d\mathbf{N}(u) \\ &- \int_0^t \Lambda_{xx}^l(u)^{-1}\Lambda_{xz}^l(u)\left(\boldsymbol{\omega}^{l+1} - \boldsymbol{\omega}^l\right) du. \end{aligned} \tag{8}$$

Given an initial estimates $\left(\hat{\boldsymbol{\beta}}^0(t), \hat{\boldsymbol{\omega}}^0\right)$, we obtain the estimation procedure summarized in Table I.

3.3. *Bandwidth selection*

Bandwidth selection is the most important characteristic of kernel estimation methods. This topic has been widely studied by many researchers; see, for example, Sheather and Jones [35], Ruppert, Sheather and Wand [36] and Marron [37]. Two usual approaches for selecting the optimal bandwidths are using plug-in methods and cross-validation [38]. On the one hand, cross-validation methods usually suffer from two problems: large running times, as well as high variance and undersmoothness. On the other hand, there is strong evidence in favor of plug-in methods in comparison with cross-validation; see, for example, Park and Marron [39] and Jones, Marron and Sheather [40]. Hence, in this section, we focus on plug-in methods for bandwidth selection.

Table I. The steps of estimating procedure.

Estimation iterative steps
0. Choose initial estimates $\left(\hat{\boldsymbol{\beta}}^0(t), \hat{\boldsymbol{\omega}}^0\right)$. Set $l = 0$. (These estimates can be easily obtained by using the method proposed by [6]).
1. Compute the Breslow estimator (3) based on initial estimates and smooth it to obtain $\theta_0^l(t)$.
2. Use equations (6) and (8) to obtain updated estimates $\boldsymbol{\omega}^{l+1}$ and $\mathbf{B}^{l+1}(t)$, respectively.
3. Smooth $\mathbf{B}^{l+1}(t)$ using (7) to obtain $\boldsymbol{\beta}^{l+1}(t)$.
4. Increase l by 1 and repeat Steps 1 to 3 until the algorithm is judged to have converged.

Katkovnik and Shmulevich [41] proposed a plug-in type method with a varying adaptive bandwidth based on the so-called intersection of confidence intervals (ICI) rule [42, 43]. As they have mentioned, one of the several attractive properties of the ICI rule is that its quality is close to the quality that one could achieve if the smoothness of the original function was known in advance.

We develop the proposed method of Katkovnik and Shmulevich [41] to determine the bandwidths $\left(h_0, h_1, \dots, h_p\right)$. This method is much faster than cross-validation. For example, in the first simulation study of Section 5 comprising 36,403 grid points and $p = 2$ time-varying coefficients (three bandwidths), the algorithm requires $\sim$ 47 s of CPU time to choose optimal bandwidths. But in the cross-validation method, for only one grid point, the associated CPU time is $\sim$ 23 seconds. It is particularly useful in practical situations when we have a large number of variables. The algorithm is as follows:

Algorithm 1. Adaptive Bandwidth Selection

1 $\bar{L} \Leftarrow -\infty$ and $\bar{U} \Leftarrow \infty$
2 while ($\bar{L} \leqslant \bar{U}$) and ($i \leqslant J$) do
$L \Leftarrow \hat{B}_{\bar{H}_i}(t) - \Gamma . n^{-2} \sum_{i=1}^{n} \hat{\zeta}_i(r) \hat{\zeta}_i(t)^T$
$U \Leftarrow \hat{B}_{\bar{H}_i}(t) + \Gamma . n^{-2} \sum_{i=1}^{n} \hat{\zeta}_i(r) \hat{\zeta}_i(t)^T$
$\bar{L} \Leftarrow \max(\bar{L}, L)$ and $\bar{L} \Leftarrow \min(\bar{U}, U)$
$i \Leftarrow i + 1$
end while
3 $\bar{H}_{opt} \Leftarrow \bar{H}_{i-1}$.

In this algorithm, $\bar{H}$ is the matrix of bandwidth grids in which each column contains the grid points for the corresponding bandwidth parameters; each row contains a possible combination of bandwidths values, and J is the total number of grid points. The estimate of the cumulative regression functions for the ith row of $\bar{H}$ is denoted by $\hat{B}(t)_{\bar{H}_i}$. Using a pilot estimate of $B(t)$ with constant bandwidth, the corresponding covariance function of $\hat{B}(t)_{\bar{H}_i}$ is $n^{-2} \sum_{i=1}^{n} \hat{\zeta}_i(r) \hat{\zeta}_i(t)^T$. Finally, Γ is a design parameter of the algorithm, and the selection of its value is discussed in [41].

4. Asymptotic properties

We present the asymptotic properties of the proposed estimators by three theorems in Appendix B. Especially, Theorem B.1 shows that $\hat{\boldsymbol{\omega}}$ is asymptotically normal and achieves a convergence rate of order $n^{-1/2}$ which is the optimal rate for parametric estimation. Theorem B.2 shows that $\hat{\mathbf{B}}(t)$ is an asymptotically Gaussian process. Hence, although the bandwidths for time-varying coefficients are different, $\hat{\mathbf{B}}(t)$ still enjoys convergence rate of $n^{-1/2}$, under certain regularity conditions. Theorem B.3 shows that $\hat{\lambda}_0(t)$ is an asymptotically Gaussian process as well.

5. Simulation study

We conduct two simulation experiments to investigate the finite-sample properties of the proposed estimators. The first example is the one used in Sun *et al.* [6] to make a comparison between our proposed method and their results. The second simulated example is motivated by the number of coefficients to be considered for analyzing heart disease data. For each simulated example, 1000 independent replications were generated.

Example 1
In this example, a marginal rates model is considered as follows:

$$E\left\{d\tilde{N}_i(t)|\mathbf{X}_i(t), Z_i(t)\right\} = \exp\left\{-0.5 + 0.5\cos(2t - 1.75)X_{i1}(t) + 0.7\left(\sqrt{t} - 1\right)X_{i2}(t) + 0.5Z_i(t)\right\}dt, \quad i = 1, \dots, n, \tag{9}$$

where the event times were simulated from a Poisson process. The covariates X_{i1} and X_{i2} were generated independently from the standard normal distribution, Z_i is a Bernoulli random variable with success probability 0.5, and the follow-up time C_i was generated from a uniform distribution $U(2, 5)$,

Table II. Simulation results for the estimates $\hat{B}_1(t)$ and $\hat{B}_2(t)$ of the cumulative time-varying regression coefficient functions in the first example.

Selected bandwidths	h_0	h_1	h_2	ISB_1	ISB_2	$MISE_1$	$MISE_2$
Equal	0.1	0.5	0.5	0.0031	0.0320	0.1394	0.1775
Different	0.1	0.5	0.2	0.0033	0.0132	0.1407	0.1775
Optimal	0.08	0.05	0.01	0.0002	0.0229	0.2018	0.2028

Table III. Simulation results for the estimate of time-independent coefficient with $\omega_0 = 0.5$ in the first example.

Selected bandwidths	h_0	h_1	h_2	Bias	SSE	SEE	CP
Equal	0.1	0.5	0.5	0.0014	0.0957	0.0992	0.9500
Different	0.1	0.5	0.2	0.0020	0.0966	0.0998	0.9500
Optimal	0.08	0.05	0.01	−0.0069	0.1185	0.1205	0.9460

which produced approximately three observed events per subject on average. We considered a sample size $n = 200$, $\tau = 4.5$, three sizes of bandwidth and Epanechnikov kernel $K(x) = 0.75(1 - x^2)I(|x| \leqslant 1)$. The cumulative regression functions are $B_1(t) = \int_0^t 0.5\cos(2s - 1.75)ds$ and $B_2(t) = \int_0^t 0.7\left(\sqrt{s} - 1\right)ds$.

Our main aim here is to study the capability of unequal and equal bandwidths for estimating time-varying parameters $B_1(t)$ and $B_2(t)$. To do this, following Sun *et al.* [6], we used equal bandwidths $h_1 = 0.5$ and $h_2 = 0.5$ (setting 1) for smoothing the coefficients. Then, we repeated the same task with different bandwidths $h_1 = 0.5$ and $h_2 = 0.2$ (setting 2) to compare the effect of equal and unequal bandwidths. In both cases, we set the bandwidth for $\theta_0(t)$ to $h_0 = 0.1$. Moreover, we obtained the optimal bandwidths $h_0 = 0.08$, $h_1 = 0.05$, and $h_2 = 0.01$ (setting 3) by applying the adaptive bandwidth selection algorithm.

We calculated the integrated squared bias (ISB) and the mean integrated squared error (MISE) for $B_i(t)$, $i = 1, 2$ for the aforementioned three settings. The results are presented in Table II. The values of ISBs and MISEs reveal that the proposed estimators of the cumulative regression coefficients are generally more precise than estimators of Sun *et al.* [6]. Especially, the results for $\hat{B}_2(t)$ are improved markedly when h_2 decreases from 0.5 to 0.2. As the third line of Table II shows, our proposed method based on the optimal bandwidths produces better bias results for $B_1(t)$, but a slightly increased MISE.

Table III presents the results of estimating the time-independent regression coefficient $\omega_0 = 0.5$ in model (9). The table shows the biases (Bias) which is the difference between the sample means of the point estimates $\hat{\omega}$ and the true value, the sampling means of the estimated standard errors (SEE) of $\hat{\omega}$, the sampling standard errors (SSE) of $\hat{\omega}$, and the 95% empirical coverage probabilities (CP) for ω_0. As it is evident from the results reported in Table III, the first two settings perform well regarding bias and standard errors. It is clear that the estimates are unbiased, and the empirical and estimated standard errors are close to each other. The empirical coverage probabilities in these two settings are 0.95, which are reasonable and show that asymptotic normal distribution provides a good description of the variability of $\hat{\omega}$.

To investigate the performance of estimators of time-varying regression coefficients and their asymptotic standard errors, we computed the point-wise estimates and empirical coverage probabilities for $\hat{B}_1(t)$ and $\hat{B}_2(t)$ at 100 grid points that are depicted in Figures 1–4.

Figure 1 shows the true cumulative regression curves for $B_1(t)$, panel (a), and $B_2(t)$, panel (b), and their estimates by considering settings 1 and 2. As it is apparent in panel (a), the performance of both equal and unequal bandwidths are nearly the same for $B_1(t)$, whereas for $B_2(t)$, panel (b), the estimate provided by different bandwidths has a lower bias. Figure 2 compares the performance of optimal different bandwidths (setting 3) and different bandwidths (setting 2), in which the true cumulative regression functions and their estimates were shown in the same manner as Figure 1. It can be seen, in panel (a), that the optimal bandwidths for $B_1(t)$ result in more accurate and nearly perfect estimates. However, for $B_2(t)$, panel (b), both approaches are well behaved, and there is no significant difference between them.

The empirical coverage probabilities for $B_1(t)$ and $B_2(t)$ are depicted in Figures 3 and 4, respectively. In both figures, the panels (a)–(c) are devoted to the settings 1–3, respectively. One sees in Figure 3, panel (c), that the difference between the coverage probabilities for $B_1(t)$ provided by optimal bandwidths and the nominal level, 0.95, for the whole time period is negligible. In Figure 3, we see that the coverage

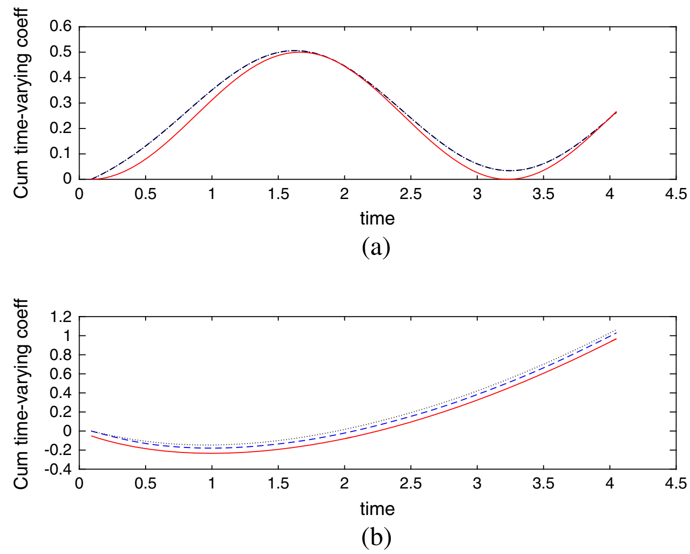


Figure 1. Plots of the cumulative regression curves and their estimates: (a) $B_1(t)$ and (b) $B_2(t)$. The solid lines are the real cumulative regression functions; the dot-dashed lines are the average estimates for the cumulative regression functions when using equal bandwidths based on 1000 replications (setting 1), and the dashed lines are the average estimates obtained by using different bandwidths (setting 2).

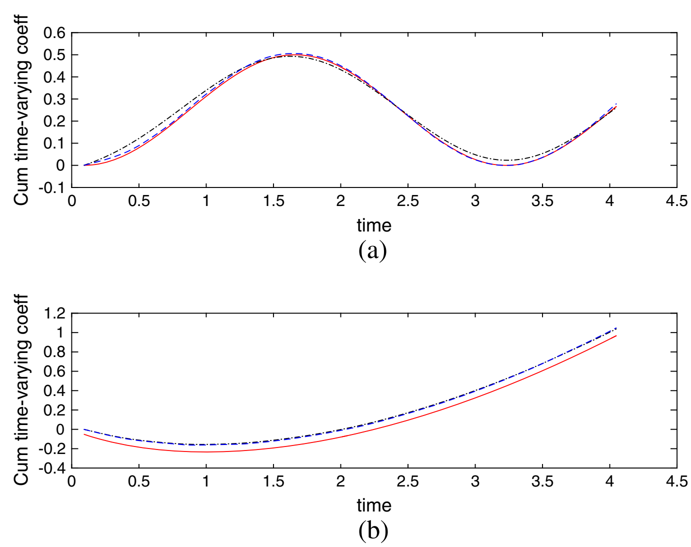


Figure 2. Plots of the cumulative regression curves and their estimates: (a) $B_1(t)$ and (b) $B_2(t)$. The solid lines are the real cumulative regression functions; the dot-dashed lines are the average estimates for the cumulative regression functions using different bandwidths (setting 2) based on 1000 replications, and the dashed lines are the average estimates obtained by using the optimal bandwidths values (setting 3).

probabilities for $B_1(t)$ provided by the other two settings, panels (a) and (b), are not as good an approximation as the output of setting 3 for time points within $[0, 1]$. As Figure 4, panels (b) and (c) shows, the results for different bandwidths (settings 2 and 3) are in agreement; indeed one detects practically no difference between their empirical coverage probabilities for $B_2(t)$ and the nominal level. From panel (a), however, the equal bandwidths method is not a good descriptor of the variability for $\hat{B}_2(t)$. In summary, the additional flexibility from our methodology resulted in substantial improvements in the asymptotic standard errors of the estimators.

Example 2

In the second example, we consider four time-varying regression coefficient functions as follows:

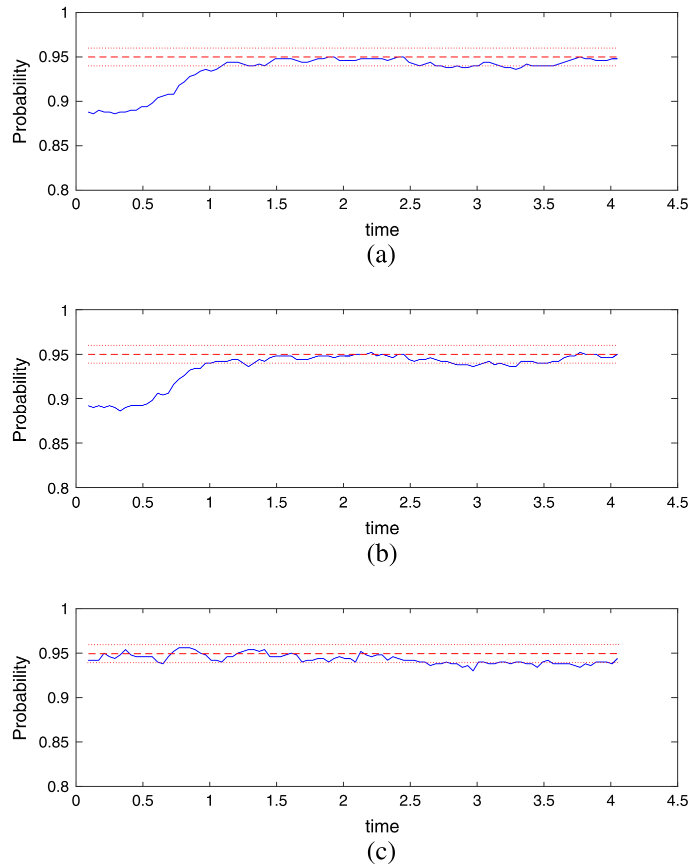


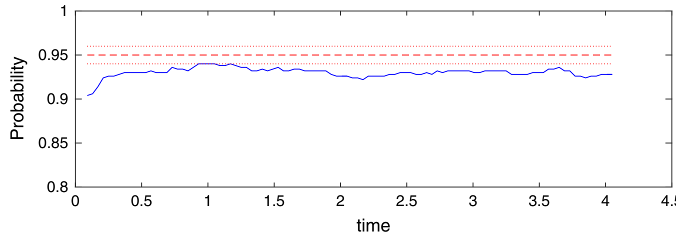
Figure 3. Coverage probabilities for $\hat{B}_1(t)$: (a) Equal bandwidths (setting 1), (b) Different bandwidths (setting 2), and (c) Optimal bandwidths (setting 3).

$$E\left\{d\tilde{N}_i(t)|\mathbf{X}_i(t), Z_i(t)\right\} = \exp\Big\{ -0.5 + 0.5\cos(2t - 1.75)X_{i1}(t) + 0.7\left(\sqrt{t} - 1\right)X_{i2}(t) + \sin(3t)X_{i3}(t) + \cos(5t)X_{i4}(t) + 0.5Z_i(t)\Big\}\, dt, \quad i = 1, \ldots, n,$$

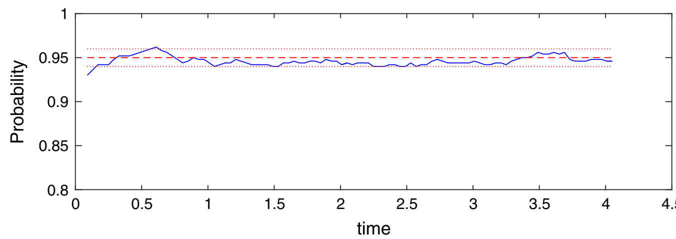
where $X_{i1}, \ldots, X_{i4}$, $i = 1, 2, \ldots, n$, were generated from standard normal distributions, and Z_i is the same as in the first example. In this example, all time-varying coefficients, except the second one, oscillate with time. In practice, we may have many coefficients with different degrees of smoothness. Therefore, this example demonstrates the power of our proposed method. To estimate all coefficients, we considered a sample size of $n = 200$. Similar to the first example, we investigated three settings for selecting bandwidths: for equal bandwidths, we considered $h_i = 0.6$ for $i = 0, 1, \ldots, 4$, whereas for different bandwidths we used $h_0 = 0.3, h_1 = 0.6, h_2 = 0.4, h_3 = 0.45$, and $h_4 = 0.5$. We also computed the optimal bandwidths, $h_0 = 0.13, h_1 = 0.13, h_2 = 0.14, h_3 = 0.12$, and $h_4 = 0.19$, similar to the first example, by using our adaptive bandwidth selection algorithm (Section 3.3). Table IV summarizes the results, showing that our proposed method results in a major improvement and enjoys lower bias and MISE for time-varying coefficients.

Comparing Tables II and IV, it can be seen that by increasing the number of time-varying regression coefficients, our proposed method outperforms the method of Sun *et al.* [6]. It makes our method particularly suitable in practical situations where there is a large number of time-varying coefficients. Table V presents the simulation results on estimating the time-independent regression coefficient $\omega_0 = 0.5$. The results are similar to those from Table III, so we do not elaborate on the details again.

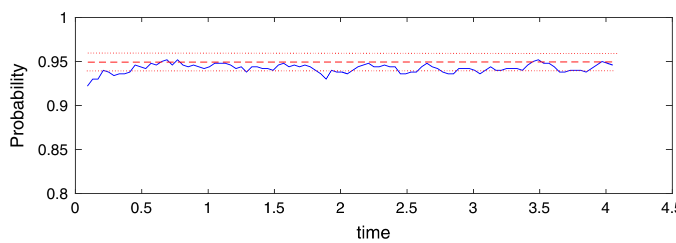
Statistics in Medicine



(a)



(b)



(c)

Figure 4. Coverage probabilities for $\hat{B}_2(t)$: (a) Equal bandwidths (setting 1), (b) Different bandwidths (setting 2), and (c) Optimal bandwidths (setting 3).

Table IV. Simulation results for the estimates $\hat{B}_1(t)$, $\hat{B}_2(t)$, $\hat{B}_3(t)$, and $\hat{B}_4(t)$ of the cumulative time-varying regression coefficient functions in the second example.

Bandwidths	h_0	h_1	h_2	h_3	h_4	ISB_1	ISB_2	ISB_3	ISB_4	$MISE_1$	$MISE_2$	$MISE_3$	$MISE_4$
Equal	0.6	0.6	0.6	0.6	0.6	0.0110	0.0251	0.0175	0.3200	0.1158	0.1219	0.1607	0.5623
Different	0.3	0.6	0.4	0.45	0.5	0.0121	0.0059	0.0159	0.2721	0.1130	0.1046	0.1512	0.4847
Optimal	0.13	0.13	0.14	0.12	0.19	0.0095	0.0034	0.0308	0.0396	0.1818	0.1747	0.1926	0.2235

Table V. Simulation results for the estimate of time-independent coefficient with $\omega_0 = 0.5$ in the second example.

Selected bandwidths	h_0	h_1	h_2	h_3	h_4	Bias	SSE	SEE	CP
Equal	0.6	0.6	0.6	0.6	0.6	−0.0053	0.0943	0.0928	0.9200
Different	0.3	0.6	0.4	0.45	0.5	−0.0052	0.0904	0.0889	0.9300
Optimal	0.13	0.13	0.14	0.12	0.19	−0.0034	0.1003	0.0831	0.9000

6. Application to heart disease data

In this section, we present the application of our method to the analysis of the real-world heart disease dataset introduced in Section 2. The data include a total of 44 patients as mentioned before. We focus on four covariates, namely QRS (time to ventricular depolarization), age of the patients, and type of the heart defibrillator device recoded into two dummy variables: ICD1 (coded 1 for single-chamber and 0

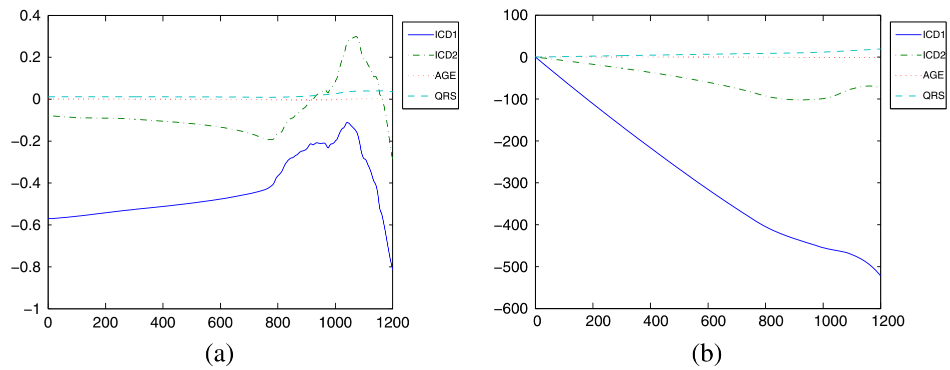


Figure 5. Estimates of (a) the coefficient functions and (b) the cumulative coefficient functions, in model (10) using optimal bandwidths.

for CRTD) and ICD2 (coded 1 for two-chamber and 0 for CRTD). For the ith patient, we denote the above variables by $W_i(t)$, $V_i(t)$, $X_i(t)$, and $Z_i(t)$, respectively. The number of defibrillator shocks to the corresponding patient's heart is denoted by $\tilde{N}_i(t)$.

We start with the following time-varying marginal regression model:

$$E\left\{d\tilde{N}_i(t)|W_i(t), V_i(t), X_i(t), Z_i(t)\right\} = \exp\left\{\beta_1(t)X_i(t) + \beta_2(t)Z_i(t) + \beta_3(t)V_i(t) + \beta_4(t)W_i(t)\right\} d\lambda_0(t). \tag{10}$$

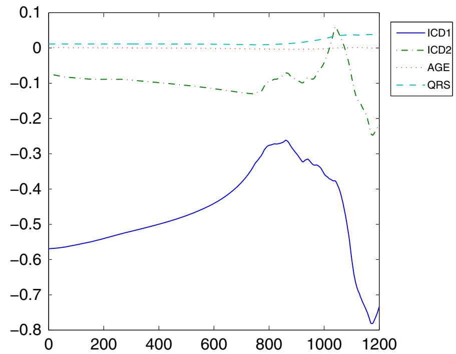
First, we should check whether $\beta_1(t)$, $\beta_2(t)$, $\beta_3(t)$, and $\beta_4(t)$ are time-varying. To that end, we computed the proposed estimators by using the Epanechnikov kernel with optimal bandwidths $h_0 = 300$, $h_1 = 791$, $h_2 = 784$, $h_3 = 772$, and $h_4 = 740$ days, resulting in smooth regression coefficients for ICD1, ICD2, age, and QRS, respectively. To obtain these optimal bandwidths, we used our adaptive Algorithm 1, presented in Section 3.3. To verify the significance of time-variation property in the coefficients, we calculated the test statistics $\mathscr{T}_1$ and $\mathscr{T}_2$ of Sun *et al.* [6] for $B_i(t) = \int_0^t \beta_i(s)ds$ $(i = 1, 2, 3, 4)$, using their proposed simulation technique based on 10,000 replications. The p-values of $\mathscr{T}_1$ and $\mathscr{T}_2$ for all cumulative coefficients were found to be less than 0.002, implying that $\beta_1(t)$, $\beta_2(t)$, $\beta_3(t)$, and $\beta_4(t)$ are indeed varying over time.

Figure 5 shows the estimates of the regression coefficients as a function of time (measured in days), together with their cumulative versions.

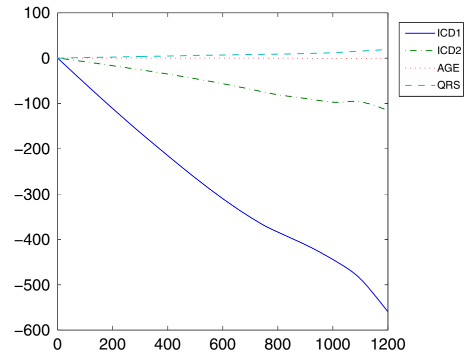
The negative values of $\hat{\beta}_1(t)$ throughout the time range in panel (a) indicate that, on average, ICD with single chamber delivers fewer beats or electrical shocks than the CRTD. During the first 34 months of the study, the differences between the number of delivered shocks from a single-chamber type of ICD and a CRTD shrink continuously. Afterwards, their differences start to grow and keep growing until the end of the study. Similarly, during the first 26 months of the study, the differences between the number of shocks generated by a two-chamber type of ICD and a CRTD initially grow, but then start to shrink. During a short period, around the 130th week of the study, these two devices have no difference and deliver an equal number of shocks. Afterwards, for the next 6 months, the differences grow again; at this period, the CRTD has better performance than ICD with two chambers. Then, differences shrink again. Note that a two-chamber ICD almost always delivers fewer shocks than a CRTD.

As Figure 5 shows, the differences between the two-chamber type of ICD and CRTD are less than the differences between the single-chamber and CRTD. Therefore, a single-chamber ICD delivered fewer beats or electrical shocks (and, in particular, fewer inappropriate beats) compared with the other type of ICD and CTRD. The figure also shows that the effect of age during the first 30 months of the study is approximately constant; thereafter, the effect increases: an older age leads to more shocks or beats. QRS behaves similarly as a function of age. Therefore, the longer the duration of the QRS, the larger the expected number of shocks or beats.

The estimates of regression coefficients using equal bandwidths $h_i = 740; i = 0, 1, 2, 3, 4$, and their cumulative versions, are shown in Figure 6. The results are substantially different compared with those obtained by using different bandwidths. For example, using the traditional method (with equal



(a)



(b)

Figure 6. Estimates of (a) the coefficient functions and (b) the cumulative coefficient functions in model (10) using equal bandwidths.

bandwidths), the second peak in the estimated number of shocks delivered by single-chamber devices is not present; hence, unlike the estimates obtained by our proposed method, the differences between the effects of the single-chamber type of ICD and CRTD increase.

6.1. Conclusion

For patients at risk of a fatally fast or chaotic heartbeat, a heart defibrillator device can be helpful for preventing sudden death. However, this device may not have a positive result on every patient. It is, therefore, crucial to correctly select patients for whom a heart defibrillator device might be helpful. Our analysis of the data from patients suffering from heart disease shows that all interesting covariates, including age, QRS duration, and type of the heart defibrillator device affect the frequency of shocks delivered to the patients' hearts. We estimated the time-varying effects of regression functions as well. Thus, if the values of covariates for a new patient are given, we can use the results of our analysis to make a primary decision on whether using a heart defibrillator device is recommended for that patient, and if so, what type of heart defibrillator should be prescribed.

7. Discussion

Many researchers have studied time-varying coefficient models under different conditions: for example, Martinussen *et al.* [24], Winnett and Sasieni [44], Cai and Sun [26], Tian *et al.* [27], and Sun *et al.* [6]. Despite the extensive literature, considering different smoothness parameters for time-varying coefficients has received little attention. This study is an effort to incorporate this feature in a class of time-varying coefficient models for recurrent events data. Our results match the findings of Sun *et al.* [6] in situations where the number of time-varying coefficients is small. However, our proposed approach significantly outperforms their method in estimating the parameters and their uncertainties as the number of functions that need to be estimated increases.

Under some regularity conditions, we showed that the suggested estimators are root-n consistent and asymptotically normally distributed. To obtain these asymptotic results, we do not need to make strong regularity assumptions: the required conditions, except Condition (C5), are equivalent to the conditions used by Sun *et al.* [6].

The main advantage of our approach is its flexibility and efficiency when the number of time-varying functions grows. Our simulation results demonstrate that the proposed method provides more accurate estimates for both time-varying and time-independent regression coefficients. This distinction becomes more evident as the number of regression functions increases. It is particularly important in practical situations, where we usually have a large number of time-varying coefficients. Despite being more general, our method remains easy to implement. Nevertheless, our approach comes with its own limitations. For instance, in the adaptive bandwidth selection algorithm, the optimum bandwidth depends on the initial choice of bandwidth. As an alternative, one could use spline regression for smoothing the coefficients. Adapting this approach to the model we considered here is left for future study. Another limitation, which also exists in the method of Sun *et al.* [6], is that our proposed method assumes the censoring and event processes are independent conditional on the covariates. This assumption does not hold if censoring were caused by informative dropouts or failure events. Nevertheless, this is a standard assumption made in the literature, and further discussion about this problem is beyond the scope of this paper.

One important issue not addressed here is the inherent problem with local constant kernel estimators that we used for smoothing regression coefficients: As one of the reviewers of this manuscript pointed out, from a practical point of view, this approach could lead to a substantial bias at boundary points. One possible solution is to use biased reduction: by correcting the bias, we may further improve the coverage accuracy for time-varying functions. While the idea of bias reduction has been widely used for kernel density estimation, each bias reduction method has its own drawbacks. For example, Härdle and Bowman [45] applied asymptotic expansions of the kernel estimators for correcting the bias. Their approach requires two levels of smoothing that can be problematic in practice. Härdle and Marron [46] proposed the application of bootstrap to reduce the bias of kernel estimators. However, their approach still requires two levels of smoothing. Another alternative for bias reduction is to employ high-order kernels. However, according to Scott [47]:

> Care should be taken in their actual use... . In practice, second- and fourth-order methods are probably the most one should consider, as kernels beyond the order 4 provide little further reduction in MISE.

As Park *et al*. [48] noticed, usually the extension of the previous bias reduction methods to kernel-based nonparametric regression is not straightforward. Examples of work in this area include Hall [49], Linton and Nielsen [50], Park *et al*. [48], and Choi *et al*. [51]. We do not study bias correction for kernel estimators in our proposed model in this paper; progress in this direction is left for future work.

Appendix A. Notations

In the following, we introduce the notation needed in the presentation of the asymptotic properties of the proposed estimators. Let

$$\Theta(t, \boldsymbol{\beta}, \boldsymbol{\omega}) = diag\{\phi_i(t; \boldsymbol{\beta}, \boldsymbol{\omega})\},$$

$$\mathbf{S}_x(t) = \mathbf{S}_x(t; \boldsymbol{\beta}, \boldsymbol{\omega}) = \frac{1}{n}\sum_{i=1}^{n} \phi_i(t; \boldsymbol{\beta}, \boldsymbol{\omega})\mathbf{X}_i(t),$$

$$\mathbf{S}_z(t) = \mathbf{S}_z(t; \boldsymbol{\beta}, \boldsymbol{\omega}) = \frac{1}{n}\sum_{i=1}^{n} \phi_i(t; \boldsymbol{\beta}, \boldsymbol{\omega})\mathbf{Z}_i(t),$$

$$\boldsymbol{\Lambda}_x(t) = \boldsymbol{\Lambda}_x(t; \boldsymbol{\beta}, \boldsymbol{\omega}) = \frac{\mathbf{S}_x(t; \boldsymbol{\beta}, \boldsymbol{\omega})}{S_0(t; \boldsymbol{\beta}, \boldsymbol{\omega})},$$

$$\boldsymbol{\Lambda}_z(t) = \boldsymbol{\Lambda}_z(t; \boldsymbol{\beta}, \boldsymbol{\omega}) = \frac{\mathbf{S}_z(t; \boldsymbol{\beta}, \boldsymbol{\omega})}{S_0(t; \boldsymbol{\beta}, \boldsymbol{\omega})},$$

$$\acute{X}(t) = \{X(t) - \bar{X}(t)\},$$

$$\acute{Z}(t) = \{Z(t) - \bar{Z}(t)\},$$

where $\bar{X}(t) = \bar{X}(t; \boldsymbol{\beta}, \boldsymbol{\gamma})$ is the $n \times p$ matrix with rows $\boldsymbol{\Lambda}_x(t; \boldsymbol{\beta}, \boldsymbol{\omega})$, and $\bar{Z}(t) = \bar{Z}(t; \boldsymbol{\beta}, \boldsymbol{\gamma})$ is the $n \times q$ matrix with rows $\boldsymbol{\Lambda}_z(t; \boldsymbol{\beta}, \boldsymbol{\omega})$. Let also

$$\Lambda_{xx}(t) = \Lambda_{xx}(t; \boldsymbol{\beta}, \boldsymbol{\omega}) = \frac{1}{n}\left\{X(t) - \bar{X}(t)\right\}^T \Theta(t; \boldsymbol{\beta}, \boldsymbol{\omega})\left\{X(t) - \bar{X}(t)\right\},$$

$$\Lambda_{zx}(t) = \Lambda_{zx}(t; \boldsymbol{\beta}, \boldsymbol{\omega}) = \frac{1}{n}\left\{Z(t) - \bar{Z}(t)\right\}^T \Theta(t; \boldsymbol{\beta}, \boldsymbol{\omega})\left\{X(t) - \bar{X}(t)\right\},$$

$$\Lambda_{xz}(t) = \Lambda_{zx}(t; \boldsymbol{\beta}, \boldsymbol{\omega}) = \frac{1}{n}\left\{X(t) - \bar{X}(t)\right\}^T \Theta(t; \boldsymbol{\beta}, \boldsymbol{\omega})\left\{Z(t) - \bar{Z}(t)\right\},$$

$$\Lambda_{zz}(t) = \Lambda_{zz}(t; \boldsymbol{\beta}, \boldsymbol{\omega}) = \frac{1}{n}\left\{Z(t) - \bar{Z}(t)\right\}^T \Theta(t; \boldsymbol{\beta}, \boldsymbol{\omega})\left\{Z(t) - \bar{Z}(t)\right\},$$

$$\Lambda^l_{xx}(t) = \Lambda_{xx}\left(t;\boldsymbol{\beta}^l,\boldsymbol{\omega}^l\right), \qquad \Lambda^l_{zx}(t) = \Lambda_{zx}\left(t;\boldsymbol{\beta}^l,\boldsymbol{\omega}^l\right), \qquad \Lambda^l_{zz}(t) = \Lambda_{zz}\left(t;\boldsymbol{\beta}^l,\boldsymbol{\omega}^l\right)$$
$$\Lambda^l_{xx}(t) = \Lambda_{xx}\left(t;\boldsymbol{\beta}^l,\boldsymbol{\omega}^l\right), \qquad S^l_0(t) = S_0\left(t;\boldsymbol{\beta}^l,\boldsymbol{\omega}^l\right), \qquad \Theta^l(t) = \Theta\left(t;\boldsymbol{\beta}^l,\boldsymbol{\omega}^l\right),$$
$$\acute{X}^l(t) = \left\{X(t) - \bar{X}\left(t;\boldsymbol{\beta}^l,\boldsymbol{\omega}^l\right)\right\}, \qquad \acute{Z}^l(t) = \left\{Z(t) - \bar{Z}\left(t;\boldsymbol{\beta}^l,\boldsymbol{\omega}^l\right)\right\},$$

where l denotes the lth iteration step in the estimation procedure.

Further, let $s_j(t;\boldsymbol{\beta},\boldsymbol{\omega})(j = 0,x,z)$ and $\mathbf{e}_k(t;\boldsymbol{\beta},\boldsymbol{\omega})(k = x,z,xx,zx,xz,zz)$ denote the limits of $S_j(t;\boldsymbol{\beta},\boldsymbol{\omega})\,(j = 0,x,z)$ and $\boldsymbol{\Lambda}_k(t;\boldsymbol{\beta},\boldsymbol{\omega})\,(k = x,z,xx,zx,xz,zz)$, respectively, as $n \to \infty$. Set $s_j(t) = s_j(t;\boldsymbol{\beta}_0,\boldsymbol{\omega}_0)\,(j = 0,x,z)$ and similarly define $\mathbf{e}_k(t)(k = x,z,xx,zx,xz,zz)$.

Appendix B. Proofs

Here, we present the asymptotic properties of the proposed estimators. For any vector v, let $\|v\| = (v^T v)^{1/2}$ denote the Euclidean norm of v. First the following regularity conditions are needed to establish the asymptotic properties of the estimators:

(C1) $\boldsymbol{\beta}_0(t)$ and $\lambda_0(t)$ are three times continuously differentiable for $t \in [0,\tau]$.
(C2) $\mathbf{X}_i(t)$ and $\mathbf{Z}_i(t)$ are of bounded variation on $[0,\tau]$.
(C3) $e_{xx}(t)$ is non-singular for $t \in [0,\tau]$.
(C4) $s_j(t;\boldsymbol{\beta},\boldsymbol{\omega})$ is uniformly continuous with respect to $(t;\boldsymbol{\beta},\boldsymbol{\omega}) \in [0,\tau]\times\mathscr{B}\times\Theta$, where $\mathscr{B}$ is a compact set of $\mathfrak{R}^p$ that includes a neighborhood of $\boldsymbol{\beta}_0(t)$ for $t \in [0,\tau]$, Θ is a compact set of $\mathfrak{R}^q$ including $\boldsymbol{\omega}$, and j denotes 0, x or z.
(C5) $K(\cdot)$ is a symmetric and continuous kernel function on a compact support such that

$$\int \mathbf{K}d\mathbf{w} = \int \begin{pmatrix} K\left(w_1\right) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & K\left(w_p\right) \end{pmatrix} d\mathbf{w} = \underset{-}{1}, \qquad \int K(w)dw = 1,$$

where $\underset{-}{1}$ is the p-vector of ones. Also, for $j = 0,1,\dots,p$, with $1/8 < \alpha_j < 1/4$, we have

$$h_0 = O(n^{-\alpha_0}), h_1 = O(n^{-\alpha_1}), \dots, h_p = O(n^{-\alpha_p}).$$

Remark B.1
Conditions (C1)–(C4) are equivalent to conditions of Sun *et al*. [6]. Condition (C1) imposes the smoothness constraints needed to estimate the parameters of the model. In practical situations, condition (C2) usually holds. Condition (C3) ensures that the limit of the information matrix of the parameters is nonsingular.

Remark B.2
Condition (C5) is modified according to our given multiple bandwidths. This condition is crucial for $\sqrt{n}$-rate of convergence for the parameter estimators.

Let h denote the element of $\left(h_0,h_1,\dots,h_p\right)$ whose convergence rate is $n^{-\alpha}$ where $\alpha = \min\left(\alpha_0,\alpha_1,\dots,\alpha_p\right)$. By this definition, we have

$$h_0 = O(n^{-\alpha}), h_1 = O(n^{-\alpha}), \dots, h_p = O(n^{-\alpha}).$$

Indeed h is the bandwidth with the slowest rate. The asymptotic properties of the estimators are presented in the following three theorems, and the proofs are relegated to the supplementary material.

Theorem B.1
Under the regularity conditions (C1)–(C5), for $\Upsilon(\boldsymbol{\omega})$ defined in (6), with probability approaching 1, there exist solution $\hat{\boldsymbol{\omega}}$ to equation $\Upsilon(\boldsymbol{\omega}) = \boldsymbol{\omega}$, such that $\|\hat{\boldsymbol{\omega}}-\boldsymbol{\omega}\| = O_p(n^{-1/2})$, and $n^{1/2}(\hat{\boldsymbol{\omega}}-\boldsymbol{\omega}_0)$ is asymptotically normal with zero mean and covariance matrix that can be consistently estimated by $n^{-1}\sum_{i=1}^n \hat{\boldsymbol{\vartheta}}_i(\tau)\hat{\boldsymbol{\vartheta}}_i(\tau)^T$, where

$$\hat{\boldsymbol{\vartheta}}_i(\tau) = \hat{A}(\tau)^{-1}\int_0^\tau \left[\left\{\mathbf{Z}_i(t) - \hat{\boldsymbol{\Lambda}}_z(t)\right\} - \hat{\Lambda}_{zx}(t)\hat{\Lambda}_{xx}(t)^{-1}\left\{\mathbf{X}_i(t) - \hat{\boldsymbol{\Lambda}}_x(t)\right\}\right] d\hat{M}_i(t),$$

in which $d\hat{M}_i(t) = dN_i(t) - Y_i(t)\exp\left\{\hat{\boldsymbol{\beta}}(t)^T\mathbf{X}_i(t) + \hat{\boldsymbol{\omega}}^T\mathbf{Z}_i(t)\right\}d\hat{\lambda}_0(t)$, $\hat{\lambda}_0(t) = \hat{\lambda}_0\left(t;\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\omega}}\right)$, $\hat{A}(\tau) = A\left(\tau,\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\omega}}\right)$ and $\hat{\Lambda}_k(t) = \Lambda_k\left(t;\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\omega}}\right)$ $(k = x, z, zx, xz, xx)$.

Theorem B.2
Under the regularity conditions (C1)–(C5), for $\Psi(\mathbf{B})(t)$ defined in (8), with probability approaching 1, there exist solution $\hat{\mathbf{B}}(t)$ to equation $\Psi(\mathbf{B})(t) = \mathbf{B}(t)$, such that $\sup_{0\leqslant t\leqslant\tau}\left\|\hat{\mathbf{B}}(t) - \mathbf{B}_0(t)\right\| = O_p(n^{-1/2})$, and the process $n^{1/2}\left(\hat{\mathbf{B}}(t) - \mathbf{B}_0(t)\right)$ converges weakly to a Gaussian process with zero mean, and a covariance function whose value at (r,t) can be consistently estimated by $n^{-1}\sum_{i=1}^n\hat{\boldsymbol{\zeta}}_i(r)\hat{\boldsymbol{\zeta}}_i(t)^T$, where

$$\hat{\boldsymbol{\zeta}}_i(t) = \int_0^t\hat{\Lambda}_{xx}(u)^{-1}\hat{\theta}_0(u)^{-1}\left\{\mathbf{X}_i(u) - \hat{\boldsymbol{\Lambda}}_x(u)\right\}d\hat{M}_i(u) - \int_0^t\hat{\Lambda}_{xx}(u)^{-1}\hat{\Lambda}_{zx}(u)^T du\hat{\boldsymbol{\vartheta}}_i(\tau),$$

and $\hat{\theta}_0(t) = \int h_0^{-1}K\left(\frac{u-t}{h_0}\right)d\hat{\lambda}_0(u)$.

Theorem B.3
Under the regularity conditions (C1)–(C5), for the Breslow-type estimator defined in (3), $\sup_{0\leqslant t\leqslant\tau}|\hat{\lambda}_0(t) - \lambda_0(t)| = O_p(n^{-1/4})$, and the process $n^{\frac{1}{2}}\left\{\hat{\lambda}_0(t) - \lambda_0(t) + \frac{1}{2}\int_0^t\mathbf{e}_x(u)^T\left(\int H^2\mathbf{K}_s\boldsymbol{\beta}_0''(u)\circ\mathbf{s}^2\circ d\mathbf{s}\right)d\lambda_0(u)\right\}$ converges weakly to a Gaussian process with zero mean and a covariance function whose value at (r,t) can be consistently estimated by $n^{-1}\sum_{i=1}^n\hat{\boldsymbol{\varphi}}_i(r)\hat{\boldsymbol{\varphi}}_i(t)$, where

$$\hat{\boldsymbol{\varphi}}_i(t) = \int_0^t\frac{d\hat{M}_i(u)}{\hat{S}_0(u)} - \int_0^t\hat{\boldsymbol{\Lambda}}_z(u)^T d\hat{\lambda}_0(u)\hat{\boldsymbol{\vartheta}}_i(\tau) - \int_0^t\hat{\boldsymbol{\Lambda}}_x(u)^T\hat{\theta}_0(u)d\hat{\boldsymbol{\zeta}}_i(u).$$

In Theorem B.3, the notation $\circ$ denotes the Hadamard product known as the entrywise product. For two matrices, say A and B, of the same dimensions, the Hadamard product A $\circ$ B is a matrix of the same dimension in which the (i,j)th element is the multiplication of the (i,j)-th element of A by the (i,j)th element of B. Moreover, $\mathbf{s}^2 = \left(s_1^2, s_2^2, \ldots, s_p^2\right)^T$, $\boldsymbol{\beta}_0''(t)$ is the second derivative of $\boldsymbol{\beta}_0(t)$, $d\mathbf{s} = \left(ds_1, ds_2, \ldots, ds_p\right)^T$, $\mathbf{K}_s$ is a diagonal matrix of $\left(K\left(s_1\right), K\left(s_2\right), \ldots, K\left(s_p\right)\right)$, $H^2 = HH^T$, and $\hat{S}_0(t) = S_0\left(t;\hat{\boldsymbol{\beta}}(t),\hat{\boldsymbol{\omega}}\right)$.

Acknowledgements

We would like to thank the management of Ghaem Hospital of Mashhad for providing the heart disease data used in this investigation. The authors also thank Dr. Shaojun Guo for providing relevant MATLAB programs and Mr. Pooria Joulani for reading a draft of the manuscript and providing useful comments.

References

1. Cook RJ, Lawless JF. *The Statistical Analysis of Recurrent Event*. Springer: New York, 2007.
2. Li QH, Lagakos SW. Use of Wei-Lin–Weissfeld method for the analysis of a recurring and a terminating events. *Statistics in Medicine* 1997; **16**:925–940.
3. Hobson RW, Weiss DG, Fields WS, Goldstone J, Moore WS, Towne JB, Wright CB. The veterans affaires cooperative study group. Effect of carotid endarterectomy for asymptotic carotid stenosis. *The New England Journal of Medicine* 1993; **328**:221–227.
4. Fleming TR, Harrington DP. *Counting Processes and Survival Analysis*. Wiley: New York, 1991.
5. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate function of recurrent events. *Journal of the Royal Statistical Society Series B* 2000; **62**:711–730.
6. Sun L, Zhou X, Guo S. Marginal regression models with time-varying coefficients for recurrent event data. *Statistics in Medicine* 2011; **30**:2265–2277.
7. Çiğşar C, Lawless JF. Assessing transient carryover effects in recurrent even processes, with application to chronic health conditions. *The Annals of Applied Statistics* 2012; **6**:1641–1663.
8. Duchateau L, Janssen P, Kezic I, Fortpied C. Evolution of recurrent asthma event rate over time in frailty models. *Journal of the Royal Statistical Society Series C Applied Statistics* 2003; **52**:355–363.
9. Ascher H, Feingold H. *Repairable Systems Reliability: Modeling, Inference, Misconceptions and Their Causes*, Lecture Notes in Statistics, vol. 7. Dekker: New York, 1984.

10. Lindqvist BH. On the statistical modeling and analysis of repairable systems. *Statistical Science* 2006; **21**:532–551.
11. Baker RD. Data-based modeling of the failure rate of repairable equipment. *Lifetime Data Analysis* 2001; **7**:65–83.
12. Andersen PK, Gill RD. Cox's regression models for counting processes: a large sample study. *The Annals of Statistics* 1982; **10**:1100–1120.
13. Prentice R, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika* 1981; **68**:373–379.
14. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* 1989; **84**:1065–1073.
15. Martinussen T, Scheike TH. *Dynamic Regression Models for Survival Data*. Springer: New York, 2006.
16. Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society Series B* 1972; **34**:187–220.
17. Murphy SA, Sen PK. Time-dependent coefficients in a Cox-type regression model. *Stochastic Processes and their Applications* 1991; **39**:153–180.
18. Hastie T, Tibshirani R. Varying coefficient model (with discussion). *Journal of the Royal Statistical Society Series B* 1993; **55**:757–796.
19. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; **81**:515–526.
20. Fahrmeir L, Klinger A. A nonparametric multiplicative hazard model for event history analysis. *Biometrika* 1998; **85**: 581–592.
21. Lawless JF, Nadeau C. Some simple robust methods for the analysis of recurrent events. *Technometrics* 1995; **37**:158–168.
22. Lin DY, Wei LJ, Ying Z. Semiparametric transformation models for point processes. *Journal of the American Statistical Association* 2001; **96**:620–628.
23. Pepe MS, Cai J. Some graphical displays and marginal regression analyses for recurrent failure times and time-dependent covariates. *Journal of the American Statistical Association* 1993; **88**:811–820.
24. Martinussen T, Scheike TH, Skovgaard IM. Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models. *Scandinavian Journal of Statistics* 2002; **29**:57–74.
25. Zucker DM, Karr AF. Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *The Annals of Statistics* 2002; **18**:329–353.
26. Cai Z, Sun Y. Local linear estimation for time-dependent coefficients in Cox's regression models. *Scandinavian Journal of Statistics* 2003; **30**:93–111.
27. Tian L, Zucker D, Wei LJ. On the Cox model with time-varying regression coefficients. *Journal of the American Statistical Association* 2005; **100**:172–183.
28. Amorim LD, Cai JW, Zeng DL, Barreto ML. Regression splines in the time-dependent coefficient rates models for recurrent event data. *Statistics in Medicine* 2008; **27**:5890–5906.
29. Fan JQ, Zhang WY. Statistical estimation in varying coefficient models. *The Annals of Statistics* 1999; **27**:1491–1518.
30. Mathers CD, Loncar D. Projection of global mortality and burden of disease from 2002 to 2030. *PLOS Medicine* 2006; **3**:2011–2030.
31. Ashley EA, Niebauer J. *Cardiology Explained*. Remedica: London, 2004.
32. Luo S, Michler K, Johnston P, Macfarlane PW. A comparison of commonly used QT correction formulae: the effect of heart rate on the QTc of normal ECGs. *PubMed* 2004; **37**:81–90.
33. Scheike TH, Martinusssen T. On estimation and tests of time-varying effects in the proportional hazard model. *Scandinavian Journal of Statistics* 2002; **31**:51–62.
34. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.
35. Sheather SJ, Jones MC. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B* 1991; **53**:683–690.
36. Ruppert D, Sheather SJ, Wand MP. An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association* 1995; **90**:1257–1270.
37. Marron JS. A personal view of smoothing and statistics. In *Statistical Theory and Computational Aspects of Smoothing*, Hardle W, Schimek M (eds). Physica-Verlag: Heidelberg, 1996; 1–9.
38. Loader CR. Bandwidth selection: classical or plug-in. *The Annals of Statistics* 1999; **27**:415–438.
39. Park BU, Marron JS. Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association* 1990; **85**:66–72.
40. Jones MC, Marron JS, Sheather SJ. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association* 1996; **91**:401–407.
41. Katkovnik V, Shmulevich I. Kernel density estimation with adaptive varying window size. *Pattern Recognition Letters* 2002; **23**:1641–1648.
42. Goldenshluger A, Nemirovsky A. On spatial adaptive estimation of nonparametric regression. *Mathematical Methods of Statistics* 1997; **6**:135–170.
43. Katkovnik V. A new method for varying adaptive bandwidth selection. *IEEE Transactions on Signal Processing* 1999; **47**:2567–2571.
44. Winnett A, Sasieni P. Iterated residuals and time-varying covariate effects in Cox regression. *Journal of the Royal Statistical Society Series B* 2003; **65**:473–488.
45. Härdle W, Bowman S. Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bounds. *Journal of the American Statistical Association* 1988; **83**:102–110.
46. Härdle W, Marron S. Bootstrap simultaneous error bars for nonparametric regression. *Annals of Statistics* 1991; **19**: 778–796.

47. Scott DW. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons: New York, 1992.
48. Park BU, Kim WC, Ruppert D, Jones MC, Signorini DF, Kohn R. Simple transformation techniques for improved nonparametric regression. *Scandinavian Journal of Statistics* 1997; **24**:145–163.
49. Hall P. On the bias of variable bandwidth curve estimators. *Biometrika* 1990; **77**:529–535.
50. Linton O, Nielsen JP. A multiplicative bias reduction method for nonparametric regression. *Statistics and Probability Letter* 1994; **19**:181–187.
51. Choi E, Hall P, Rousson V. Data sharpening methods for bias reduction in nonparametric regression. *Annals of Statistics* 2000; **28**:1339–1355.

Supporting Information

Additional supporting information may be found in the online version of this article at the publishers web site.