symmetry

MDPI

Article

A Symmetric Kernel Smoothing Estimation of the Time-Varying Coefficient for Medical Costs

Simeng Li [1,2], Dianliang Deng [2,*] and Yuecai Han [1]

[1] School of Mathematics, Jilin University, No. 2699 Qianjin Street, Changchun 130012, China; lsm18@mails.jlu.edu.cn (S.L.); hanyc@jlu.edu.cn (Y.H.)
[2] Department of Mathematics and Statistics, University of Regina, Regina, SK S4S 0A2, Canada
* Correspondence: deng@uregina.ca; Tel.: +1-306-585-4344

Abstract: In longitudinal studies, subjects are repeatedly observed at a set of distinct time points until the terminal event time. The time-varying coefficient model extends the parametric method and captures the dynamic trajectories of time-dependent covariate effects, thus enabling it to describe the potential relationship between the longitudinal variable and the observed time points. In this study, we propose a novel approach to the estimation of medical costs using a symmetric kernel smoothing method in the time-varying coefficient joint model. A smooth function of medical costs is derived by weighting the values of longitudinal data at all distinct observed time points via the combination of the kernel method and the inverse probability weighting method. For the simulation study, we first set up the true functions of time-varying coefficients; we then generated random samples for covariates and censored survival times. Subsequently, the longitudinal data of response variables could be produced. Further, numerical simulation experiments were conducted by using the proposed method and applying R code to the generated data. The estimated results for the parameters and non-parametric functions were compared with different settings. The numerical results illustrate that as the sample size increases, the bias and model-based standard errors decrease, and the performance improves with larger sample sizes. The estimates of functions in the model almost coincide with the true functions, as shown in the figures of the simulation study. Furthermore, the consistency of the obtained estimator is demonstrated via theoretical analysis, and a numerical simulation is performed to illustrate the performance of the proposed estimators. The proposed model is applied to a real-world data set acquired from a multicenter automatic defibrillator implantation trial (MADIT).

Keywords: varying coefficients; survival model; kernel function; medical cost; right censoring

check for updates

Citation: Li, S.; Deng, D.; Han, Y. A Symmetric Kernel Smoothing Estimation of the Time-Varying Coefficient for Medical Costs. *Symmetry* **2024**, *16*, 389. https://doi.org/10.3390/sym16040389

Academic Editor: Sergei D. Odintsov

Received: 21 February 2024
Revised: 16 March 2024
Accepted: 22 March 2024
Published: 26 March 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In medical cost studies, researchers need to use appropriate methods to evaluate the average medical cost of a patient across their whole life. Due to censoring mechanisms, the survival function is generally not identifiable. Bang and Tsiatis [1] introduced a class of weighted estimators that appropriately account for censoring; although extensive simulation studies showed that the estimators perform well in finite samples, even with heavily censored data, the estimator is not efficient, and the computations are complex. Lin et al. [2] partitioned the entire time period of interest into a number of small intervals and estimated the average total cost to minimize the bias induced by censoring. Furthermore, the estimators were proven to be asymptotically normal. Huang and Lovato [3] formulated weighted log-rank statistics in a marked point process framework and developed the asymptotic theory. The above methods have been applied to estimate the cumulative mean function. However, in addition to censoring mechanisms, data often consist of longitudinal outcomes, time-dependent/independent covariates, event times, and censoring times. In longitudinal studies, subjects are repeatedly observed at a set of distinct time points until the terminal event time. The joint model [4] contains both a longitudinal sub-model, to

express the effect on longitudinal measurements from time-dependent covariates, and a survival sub-model, to reveal the survival function association with the longitudinal part. Time-independent covariates are generally present in the survival sub-model due to the absence of repeated measurements.

Deng [5] considered a linear parametric regression model to describe longitudinal data and used the joint modeling technique and the inverse probability weighting method [6] to estimate the cumulative mean function. Although this method can also solve the problem of handling time-dependent covariates and right-censored time-to-event data, it may still be restrictive for capturing initial data.The main reason for this is that the effects of covariates on longitudinal outcomes are considered time-independent constants.

Thus, some researchers have focused on non-parametric longitudinal sub-models (e.g., Hoover et al. [7]; Zhao et al. [8]; Li et al. [9]; Do et al. [10]). There have been numerous approaches to estimating non-parametric estimators in the recent literature, such as kernel, smoothing spline, regression spline, and wavelet-based methods. For instance, Eubank and Speckman [11] proposed a well-behaved non-parametric kernel regression model in a small-sample study with bias-corrected confidence bands and proved the asymptotic properties. Wu et al. [12] minimized the local square criterion and obtained the asymptotic distributions for kernel estimators. However, when the covariate dimension is too high, the smooth estimation of general multivariate non-parametric regression may require a large sample size, and the smoothing results may be difficult to interpret.

Several time-varying coefficient longitudinal models have been considered in recent studies. For instance, You et al. [13] considered the time-varying coefficient as a polynomial basis regression spline, proposed a mixed-effects model for multiple longitudinal outcomes using the local polynomial method, and provided tuning parameters and variable selection. However, this model can only be used in continuous longitudinal outcomes. Observed data tend to be discrete in many applications. Moreover, time-independent covariates were not considered in their study [13]. The joint model with kernel smoothing varying coefficients in the longitudinal sub-model can be used to estimate the cumulative mean function with discrete data. Therefore, in this study, we estimate the cumulative mean function with time-dependent/independent covariates using the kernel method in a joint time-varying coefficient model based on right/interval censoring history process data. In our method, the estimator of the mean state function is unbiased at time points where all subjects are observed. The reason for this is that, after smoothing the function with the kernel method, the values of the original sample points are not changed. Thus, we can utilize all the known information for estimation without data bias. Moreover, in our simulation study, the estimator of the cumulative mean state function is almost equal to the value of the preassigned function at any time point.

The remainder of the paper is organized as follows: Section 2 establishes the joint varying coefficient model and proposes the estimators of time-varying coefficients based on our method. Section 3 demonstrates the feasibility of this method through numerical simulations. Section 4 applies the proposed model to a real-world data set from a multicenter automatic defibrillator implantation trial (MADIT). Section 5 discusses the influence of bandwidth h selection and concludes this paper. Finally, the appendices provide the proofs of the main results.

2. Estimation for a Time-Varing Coefficient Model

It is assumed that history process data are right-censored. Let T denote the terminal event time, and C denote the censoring event time. Let $Y(t)$ denote the state process which is related to the time-dependent covariate $\boldsymbol{X}(t)$ and time-independent covariate $\boldsymbol{W}$. The state process satisfies $Y(t) = 0$ when $t \geqslant T$. For $i = 1, 2, \ldots, m$, let T_i and C_i denote the true values of T and C for the ith subject. Further, $\delta_i = I(T_i \leqslant C_i)$ denotes the censoring indicator. Assume that censoring time is independent of terminal event time and the state process. Let $y_i(t)$, q-dimensional vector $\boldsymbol{x}_i(t) = (x_{i1}(t), x_{i2}(t), \ldots, x_{iq}(t))^T$, and

p-dimensional vector $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{ip})^T$ be the observed history of $Y(t)$, $\boldsymbol{X}(t)$, and $\boldsymbol{W}$ for the ith subject.

Now, the time-varying coefficient joint model can be formed as follows:

$$\begin{cases} y_i(t) = \boldsymbol{x}_i(t)^T \boldsymbol{\beta}(t) + \epsilon_i(t), \\ h_i(t) = h_0(t) \exp\{w_i\gamma + \alpha(y_i(t) - \epsilon_i(t))\}, \end{cases} \tag{1}$$

where $\boldsymbol{\beta}(t)$ is time-varying coefficient parameter, α is the association coefficient of the longitudinal outcome to the hazards for occurring event, $h_0(t)$ is the baseline hazards function, which is known in our model, and $\epsilon_i(t)$ is the random error with $\epsilon_i(t) \sim N(0, \sigma^2)$. Further, it is assumed that $\epsilon(t)$ is independent of terminal event T conditional on $X(t)$ and W.

2.1. Estimation of Time-Varying Coefficient in Longitudinal Model

We should notice that the observations for ith subject $(i = 1, 2, \ldots, m)$ are not continuous but only able to be obtained at some special times t_{ij}, $(j = 1, 2, \ldots, n_i)$. The state process $Y(t)$ can be the observed until the event time $T_i^* = \min(T_i, C_i)$ Thus, let $\boldsymbol{y}_i(t_{ij})$ be the observed state history. For the sake of illustration, assume $\boldsymbol{y}_i(t_{ij}) = 0$ when $t \geqslant T^*$.

For convenience, define sets $\Delta_i = \{t_{ij}, j = 1, 2, \ldots, n_i\}$ for $i = 1, 2, \ldots, m$ and $\Delta = \{t_{(k)}; k = 0, 1, 2, \ldots, N\}$, where $0 = t_{(0)} < t_{(1)} < t_{(2)} < \ldots < t_{(N)}$ are the observed distinct time points for all subjects. N denotes the number of all the distinct observed time points.

The estimator $\hat{\boldsymbol{\beta}}(t)$ of the time-varying coefficient $\boldsymbol{\beta}(t)$ can be calculated by minimizing the following equation:

$$L_N(t) = \sum_{k=1}^{N} W_{k,h}(t - t_{(k)}) \left\{ \sum_{i=1}^{m} [I\{T_i^* \geqslant t_{(k)}\}\{y_i(t_{(k)}) - \boldsymbol{x}_i(t_{(k)})^T \boldsymbol{\beta}(t)\}]^2 \right\}, \tag{2}$$

where

$$W_{k,h}(t - t_{(k)}) = \frac{K_h(t - t_{(k)})}{\sum_{k=1}^{N} K_h(t - t_{(k)})},$$

$K_h(\cdot)$ is the kernel weight function. The bandwidth, which is generally selected according to the observed time points, reveals the distance of t and $t_{(k)}$. There is a major distinction between the estimation from Equation (2) and the estimation in Wu [12] since the censoring mechanism, indicator $I\{T_i^* \geqslant t_{(k)}\}$, is necessary in order to avoid incomplete data due to censoring.

Remark 1. *We assume that $K_h(\cdot)$ is the Epanechnikov kernel function. The cross-validation method can be used to select the appropriate kernel function. The kernel function with the smallest induced error is considered the best. More details on kernel function selection can be found in [14].*

Equivalently, we write Equation (2) in the matrix form:

$$L_N(t) = \sum_{k=1}^{N} (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta}(t))^T \tilde{\mathbf{D}}_k(t) (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta}(t)), \tag{3}$$

where

$$\mathbf{Y}_k = \begin{pmatrix} y_1(t_{(k)}) \\ \vdots \\ y_m(t_{(k)}) \end{pmatrix},$$

$$\mathbf{X}_k = \begin{pmatrix} x_{11}(t_{(k)}) & \cdots & x_{1q}(t_{(k)}) \\ \vdots & \cdots & \vdots \\ x_{m1}(t_{(k)}) & \cdots & x_{mq}(t_{(k)}) \end{pmatrix},$$

and

$$\tilde{\mathbf{D}}_k(t) = \mathrm{diag}\Big(W_{k,h}(t - t_{(k)})I\{T_1^* \geqslant t_{(k)}\}, \ldots, W_{k,h}(t - t_{(k)})I\{T_m^* \geqslant t_{(k)}\}\Big),$$

It is assumed that $\left(\sum_{k=1}^{N} \mathbf{X}_{t_{(k)}}^T \tilde{\mathbf{D}}_k \mathbf{X}_{t_{(k)}}\right)$ is also invertible. By minimizing Equation (3), $\hat{\boldsymbol{\beta}}(t)$ can be expressed as the following q-dimensional column vector:

$$\hat{\boldsymbol{\beta}}(t) = \left(\sum_{k=1}^{N} \mathbf{X}_k^T \tilde{\mathbf{D}}_k(t) \mathbf{X}_k\right)^{-1} \left(\sum_{k=1}^{N} \mathbf{X}_k^T \tilde{\mathbf{D}}_k(t) \mathbf{Y}_k\right). \tag{4}$$

The traditional method for bandwidth selection is K-fold cross-validation (CV). Time-varying coefficients complicate the calculation. However, 'leave-one-subject-out' cross-validation proposed by Rice and Sliverman [15] can be used in such scenarios. In this case, the kernel weight function is $K(t - t_{(k)}; h)$, the estimator of time-varying coefficient $\boldsymbol{\beta}(t)$ is $\hat{\boldsymbol{\beta}}(t; h)$. Minimizing the following equation:

$$CV(h) = \sum_{k=1}^{N} \sum_{i=1}^{m} E[I\{T_i^* \geqslant t_{(k)}\}\{y_i(t_{(k)}) - \boldsymbol{x}_i(t_{(k)})^T \boldsymbol{\beta}^{(-i)}(t_{(k)}; h)\}]^2,$$

where $\hat{\boldsymbol{\beta}}^{(-i)}(t_{(k)}; h)$ is the kernel estimator computed with all measurements except the measurements of ith subject . Then, the cross-validation bandwidth can be obtained.

2.2. Estimation of Survival Model

Define $m_i(t) = y_i(t) - \epsilon_i(t) = \boldsymbol{x}_i(t)^T \boldsymbol{\beta}(t)$, and then the estimate of $m_i(t)$ is

$$\hat{m}_i(t) = \boldsymbol{x}_i(t)^T \hat{\boldsymbol{\beta}}(t).$$

The data related to survival sub-model consist of $\{(\boldsymbol{x}_i(t), \boldsymbol{w_i}, T_i^*, \delta_i,); i = 1, 2, \ldots, m\}$. For each given $t \in R$,

$$h_i(t) = h_0(t) \exp\{\boldsymbol{w}_i \boldsymbol{\gamma} + \alpha m_i(t)\},$$

The Cox partial maximum likelihood function [4] is

$$L_p(\boldsymbol{\gamma}, \alpha) = \prod_{i=1}^{m} \left[\frac{\exp\{\boldsymbol{w}_i \boldsymbol{\gamma} + \alpha m_i(T_i^*)\}}{\int_0^{T_i^*} \exp\{\boldsymbol{w}_i \boldsymbol{\gamma} + \alpha m_i(s)\} ds}\right]^{\delta_i}. \tag{5}$$

The log-partial likelihood function for Equation (5) is

$$\begin{aligned} LL(\boldsymbol{\theta}) &= \log L_p(\boldsymbol{\gamma}, \alpha) \\ &= \sum_{i=1}^{m} \delta_i \left[\boldsymbol{w}_i \boldsymbol{\gamma} + \alpha(m_i(T_i^*)) - \log \int_0^{T_i^*} \exp\{\boldsymbol{w}_i \boldsymbol{\gamma} + \alpha m_i(s)\} ds\right]. \end{aligned} \tag{6}$$

Then, replacing $m_i(t)$ by the estimator $\hat{m}_i(t)$, the estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\gamma}}, \hat{\alpha})$ can be obtained by maximizing Equation (6).

We define the survival function of T as:

$$S_T(t) = P(T > t).$$

The estimate of $S_T(t)$ can be obtained from the hazards function in joint model:

$$\hat{S}_T(t) = \exp\left\{-\int_0^t h_0(s) \exp\{\boldsymbol{w}^T \hat{\boldsymbol{\gamma}} + \hat{\alpha} \hat{m}(s)\} ds\right\}.$$

Theoretically, the estimate $\hat{S}_T(t)$ is related to the values of covariates $\boldsymbol{x}(s)$ for $s \in [0, t]$. Since $\boldsymbol{x}(t)$ generally is not continuously observed, the estimate $\hat{S}_T(t)$ can not be derived even at observed time points. $\hat{S}_T(t)$ can be calculated only if $\boldsymbol{x}(t)$ can be observed continuously. Thus, we replace $\hat{S}_T(t)$ with the Kaplan–Meier estimator [16]. Alternatively, in terms of the law of large numbers, the survival function can also be estimated as follows:

$$\hat{S}(t) = \frac{1}{m} \sum_{i=1}^{m} I\{T_i \geqslant t\}. \tag{7}$$

Moreover, the estimate of the survival function of $\hat{S}^*(t) = P(T^* > t)$ for T^* can be obtained in a similar way.

2.3. Estimation of Cumulative Mean State Function

Suppose $\nu(t)$ is the mean function of $Y(t)$, that is, given $X(t)$, $\nu(t) = E(Y(t) \mid X(t))$. Because of the censoring, the proposed estimator $\hat{\nu}(t)$ for the mean state function $\nu(t)$ at any time is as follows:

$$\hat{\nu}(t) = \frac{\frac{1}{m}\sum_{i=1}^{m} I\{T_i^* \geqslant t\}\hat{Y}_i(t)}{\hat{S}^*(t)}, \tag{8}$$

where $\hat{Y}_i(t) = \boldsymbol{x}_i(t)\hat{\boldsymbol{\beta}}(t)$ is the fitted value of $Y_i(t)$ at time point t.

Then, the cumulative mean function $\mu(t)$ for any time point t can be obtained as:

$$\hat{\mu}(t) = \int_0^t \hat{\nu}(s)ds. \tag{9}$$

2.4. The Asymptotic Property of Estimators

Here, we discuss the asymptotic property of estimators. To rigorously define the statements, we introduce various notations. Let $\mathcal{R}(t)$ denote the observed covariate processes, such as baseline information, study time, and so on, that is, $\mathcal{R}(t) = \{\boldsymbol{x}(t), \boldsymbol{w}\}$. Then, let $\bar{Z}_{\mathcal{R}}(t)$ denote the longitudinal covariate history prior to time t and $\bar{Z}_Y(t)$ denote the response history prior to time t, that is, $\bar{Z}_{\mathcal{R}}(t) = \{\mathcal{R}(s) : s < t\}$ and $\bar{Z}_Y(t) = \{Y(s) : s < t\}$. Furthermore, we use $\|\cdot\|$ to denote the Euclidean norm in real space and $\mathcal{R}'(t)$ to denote the derivative of $\mathcal{R}(t)$ with respect to time t. It is assumed that all observed time points t_{ij} are independent of each other and follow a distribution F_T with density f_T.

Assumption 1. *$X(t)$ is Lipschitz-continuous with order λ_0, $|E(X_l^T X_l) - E(X_s^T X_s)| \leqslant C_0 |t_{(l)} - t_{(s)}|^{\lambda_0}$ for any $t_{(l)}$ and $t_{(s)}$ in support of f_T and some $C_0 > 0$, and $\boldsymbol{\beta}(t)$ and f_T are Lipschitz-continuous with orders of $\lambda_1 > 0$ and $\lambda_2 > 0$, respectively.*

Assumption 2. *$o(t) = \lim_{\Delta t \to 0} E\{\epsilon(t + \Delta t)\epsilon(t)\}$ and $w(t) = E\{\epsilon^2(t)\}$ are continuous.*

Assumption 3. *$W_{k,h}(\cdot)$ is square-integrable that integrates to one and satisfies $\int uW_{k,h}(u)du = 0$, $m_1 = \int u^2 W_{k,h}(u)du < \infty$, and $m_2 = \int W_{k,h}^2(u)du < \infty$, while $h \to 0$ and $Nh \to \infty$ as $N \to \infty$.*

Assumption 4. *The bandwidth satisfies $h = N^{-1/5} M_0$ for some constant M_0.*

Assumption 5. $\lim_{n\to\infty} N^{-6/5} \sum_{i=1}^{n} n_i = U$ *for some* $0 \leqslant U < \infty$

Theorem 1. *Under Assumptions (1)–(5), the estimator $\hat{\boldsymbol{\beta}}(\tau)$ defined in Equation (4) is asymptotically multivariate normal for any $\tau \in [0, T]$ as $m \to \infty$.*

We summarize the asymptotic normality of $\hat{\boldsymbol{\beta}}(\tau)$ at a fixed point τ, where $\tau \in R$ in Appendix A.

Remark 2. *Most assumptions are similar regularity conditions as that in Wu et al. [12]. Assumptions 1 and 2 are rigorous statements for which* $\left(\sum_{k=1}^{N} \mathbf{X}_{t_{(k)}}^{T} \tilde{\mathbf{D}}_k \mathbf{X}_{t_{(k)}}\right)$ *is positive-definite and invertible asymptotically. Assumption 3 ensures that* $W_{k,h}(\cdot)$ *has a compact support on R. Assumptions 4 and 5 are results from finite moments in Assumptions 1 and 2. More discussions about the asymptotic risk for the kernel estimators can be found in [12].*

Theorem 2. *It is assumed that for* $t \in (0, T)$, $E(y_i(t) \mid \boldsymbol{x}_i(t)) = E(Y(t) \mid \mathbf{X}(t))$, *the estimator* $\hat{v}(t)$ *defined in Equation (8) is an unbiased estimator for any* $t \in \Delta$.

The proof can be found in Appendix B.

Based on Zeng and Cai [17], the following assumptions are imposed on the joint model.

Assumption 6. *For any* $t \in [0, T]$, *the covariate process* $\mathcal{R}(t)$ *is fully observed and conditional on* $\bar{Z}_{\mathcal{R}}(t)$, $\bar{Z}_Y(t)$, *and* $T \geqslant t$; *the distribution of* $\mathcal{R}(t)$ *depends only on* $\bar{Z}_{\mathcal{R}}(t)$. $\mathcal{R}(t)$ *is continuously differentiable in* $[0, T]$ *and* $max_{t\in[0,T]} \parallel \mathcal{R}'(t) \parallel < \infty$ *with a probability of one.*

Assumption 7. *The censoring time C depends only on* $\bar{Z}_{\mathcal{R}}(t)$ *and* $\mathcal{R}(t)$ *for any* $t < T$ *conditional on* $\bar{Z}_{\mathcal{R}}(t)$, $\bar{Z}_Y(t)$, $\mathcal{R}(t)$ *and* $T \geqslant t$.

Assumption 8. *Full-rank* $P(\boldsymbol{X}^T \boldsymbol{X})$ *is positive. Additionally, if there is an existing constant vector* $\boldsymbol{C}_0$ *satisfying* $\boldsymbol{X}(t)^T \boldsymbol{C}_0 = g(t)$ *for a deterministic function* $g(t)$ *for all* $t \in [0, T]$ *with a positive probability, then* $\boldsymbol{C}_0 = 0$ *and* $g(t) = 0$.

Assumption 9. *The true value of parameter* $\boldsymbol{\theta} = (\sigma^2, \gamma^T, \alpha)$ *satisfies* $\parallel \theta \parallel \leqslant Q_0, \sigma_0^2 > Q_0^{-1}$ *for a known positive constant* Q_0.

Assumption 10. *The baseline hazard function* $h_0(t)$ *is bounded and positive in* $[0, T]$.

Assumption 11. *There is an existing positive constant* $a > 0$ *satisfying* $S_T(T) \geqslant a$.

Remark 3. *Assumption 6 serves as a fundamental statement in joint models, indicating that the association between the history process and the survival time is due to observed covariate processes, such as baseline information, study time, and so on, denoted by* $\mathcal{R}(t)$. *Assumption 7 means that there exist some appropriate measures such that the intensity function of* $Nc(t)$ *exists. Assumption 8 is the identifiability assumption in a linear mixed-effects model. Assumptions 9–11 imply that, conditional on* $\bar{Z}_{\mathcal{R}}(t)$ *and* $\mathcal{R}(t)$, *the probability of a subject surviving after time* τ *is at least some positive constant. Theorem 3.1 in Zeng and Cai [17] states the strong consistency of the maximum likelihood estimator. More discussions on the assumptions can be found in [17].*

Theorem 3. *Under Assumptions (1)–(11), the estimators* $\hat{v}(t)$ *defined in Equation (8) and* $\hat{\mu}(t)$ *defined in Equation (9) are consistent for any* $t \in [0, T]$.

The proof can be found in Appendix C.

3. Simulation

In this section, some numerical results are presented. In our simulation, the joint model can be described as:

$$\begin{cases} Y(t) = \boldsymbol{x}(t)\boldsymbol{\beta}(t) + \epsilon(t), \\ h(t) = h_0(t)\exp\{\gamma w + \alpha(Y(t) - \epsilon(t))\}, \end{cases}$$

where $Y(t)$ is the state function, $\boldsymbol{x}(t) = (1, x_1(t), \cdots, x_p(t))$ is the vector of covariates for regression parameters, $\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t), \cdots, \beta_p(t))^T$ is the time-varying coefficients, w is the covariates for regression parameter γ, α is the association parameter, and $\epsilon(t) \sim N(0, 0.1)$.

The standard deviation (Std.dev) and the root mean square errors (RMSE) with $RMSE = [\frac{1}{N}\sum_{k=1}^{N}(\hat{g}(t_{(k)}) - g(t_{(k)}))^2]^{\frac{1}{2}}$ of overall estimates calculated by R 4.3.1 [18] are used to assess the quality of estimators. We summarize the steps in the following procedure:

1. Set the sample size n, the true function $\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t), \beta_2(t))$, the true value of parameters α, γ, and the rate of censoring r;
2. Generate a random sample $g_i \sim U[0,1]$;
3. Derive the random sample s_i of lifetime with the hazards function $h(t)$;
4. Generate a random sample of censoring $C_i \sim U[a,b]$;
5. Set $t_i = min\{v_i, c_i\}, \delta_i = I\{v_i \leq c_i\}$;
6. Generate the random sample of the time-dependent covariates $x_1(s), x_2(s)$, and the baseline hazards function $h_0(s)$.
7. For $s = \{1, 2, \ldots, t_i\}$, generate the response variables $y(s) = \beta_0 + \beta_1 x_1(s) + \beta_2 x_2(s)$.

Output the estimated function $\hat{\boldsymbol{\beta}}(t), \hat{\mu}(t)$, the estimated value of parameters $\hat{\alpha}$ and $\hat{\gamma}$, the bias and the std.err of parameters α and γ, and the RMSE of the estimated function $\hat{v}(t), \hat{\mu}(t)$.

We utilize packages of 'MASS', 'splines', 'survival', 'nlme', 'JM', 'lattice', 'mvtnorm', 'tibble' and 'ggplot2' in our simulation study.

Now, we consider the following scenarios:

Scenario 1: Set $\boldsymbol{x}(t) = (1, x_1, x_2)$, $x_1 = r(1.5\sin(t) + 1)$ where $r \sim Bernouli(0.5)$, $x_2 \sim U[1,2]$, $\boldsymbol{\beta}(t) = (\beta_0, \beta_1, \beta_2)^T$, $\beta_0(t) = 1.5t$, $\beta_1(t) = 1.2t^{0.5}$, $\beta_2(t) = t^{0.2}$, $w \sim U[0,1]$, $\gamma = 1.0$, $\alpha = 0.25$, and $h_0(t) = h_0$.

Scenario 2: Set $\boldsymbol{x}(t) = (1, x_1, x_2)$, $x_1 = r\log(t)$ where $r \sim N(1, 0.5)$, $x_2 \sim exp(0.5)$, $\boldsymbol{\beta}(t) = (\beta_0, \beta_1, \beta_2)^T$, $\beta_0(t) = t$, $\beta_1(t) = t^{0.5}$, $\beta_2(t) = \sin(t)$, $w \sim U[0,1]$, $\gamma = 1.0$, $\alpha = 0.15$, and $h_0(t) = h_0$.

The following is the form based on the joint model:

$$\begin{cases} Y(t) = \beta_0(t) + \beta_1(t)x_1(t) + \beta_2(t)x_2 + \epsilon(t), \\ h(t) = h_0(t)\exp\{\gamma w + \alpha(\beta_0(t) + \beta_1(t)x_1(t) + \beta_2(t)x_2\}. \end{cases}$$

After 1000 replications in R, we obtain the results of the estimation at different sample sizes. In each scenario, we control the censoring rate by about 25%.

In scenario 1, we set the true values of parameters as $\alpha = 0.25$ and $\gamma = 1.0$. In the scenario 2, we set the true values of parameters as $\alpha = 0.15$ and $\gamma = 1.0$. From the proposed estimators given in Equations (8) and (9), fitted values of the state function $v(t)$ and the cumulative mean function $\mu(t)$ for any time points can be computed.

Tables 1 and 2 present the estimates of the root of mean square errors (RMSEs) of the state function $v(t)$ and the cumulative mean function $\mu(t)$ with different numbers of time points N for the two scenarios. The RMSE is small when $N = 106$ because the estimators of $v(t)$ are unbiased at observed time points. Since the estimators are biased at other time points, the RMSE increases as N increases.

Table 1. The estimates of RMSE with different numbers of N for scenario 1.

Parameter	$N = 106$	$N = 221$	$N = 441$
$v(t)$	0.9019575	1.291602	1.762792
$\mu(t)$	11.42673	15.1283	27.13671

Table 2. The estimates of RMSE with different numbers of N for scenario 2.

Parameter	$N = 106$	$N = 221$	$N = 441$
$v(t)$	1.830297	2.434687	3.582806
$\mu(t)$	19.03169	45.26685	84.30626

Tables 3 and 4 summarize the main findings of fixed parameters. The results show that as the sample size increases, the bias and model-based standard errors decrease, which coincides with empirical results reasonably well. The performance improves with larger sample size. Note that it is common for the Std. Err. of γ to be very large, sometimes reaching as large as 0.2 in small samples for linear mixed-effects models (see Table 1 in [5]). In the semi-parametric model estimation based on polynomial regression, the Std. Err. of γ even reaches 0.3 (see Tables 1 and 2 in [19]). Compared with these methods, the Std. Err. of γ in our paper is less.

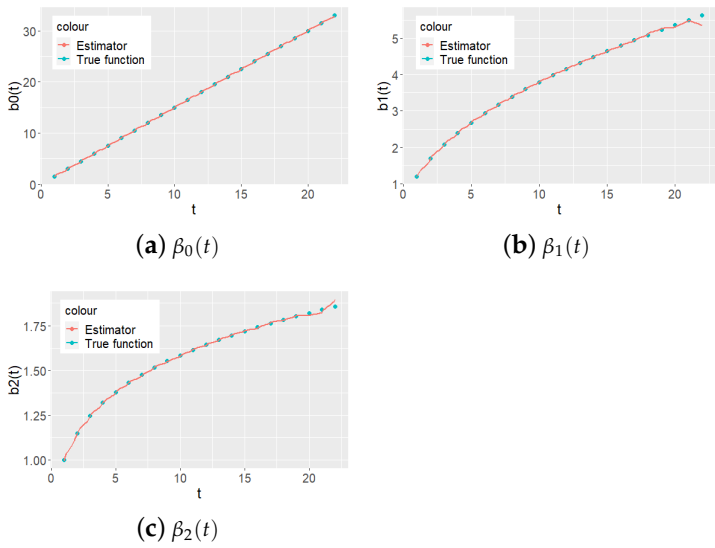
Table 3. The estimate results of the event process for scenario 1.

Parameter	True	$n = 125$		$n = 250$		$n = 500$	
		Bias	Std. Err.	Bias	Std. Err.	Bias	Std. Err.
α	0.25	−0.0345	0.0300	−0.0302	0.0212	−0.0300	0.0154
γ	1.00	−0.0119	0.2590	0.0275	0.1730	−0.0040	0.1260

Table 4. The estimate results of the event process for scenario 2.

Parameter	True	$n = 125$		$n = 250$		$n = 500$	
		Bias	Std. Err.	Bias	Std. Err.	Bias	Std. Err.
α	0.15	−0.0286	0.0311	−0.0230	0.0188	−0.0223	0.0116
γ	1.00	0.0219	0.1170	0.0208	0.1160	0.0235	0.1120

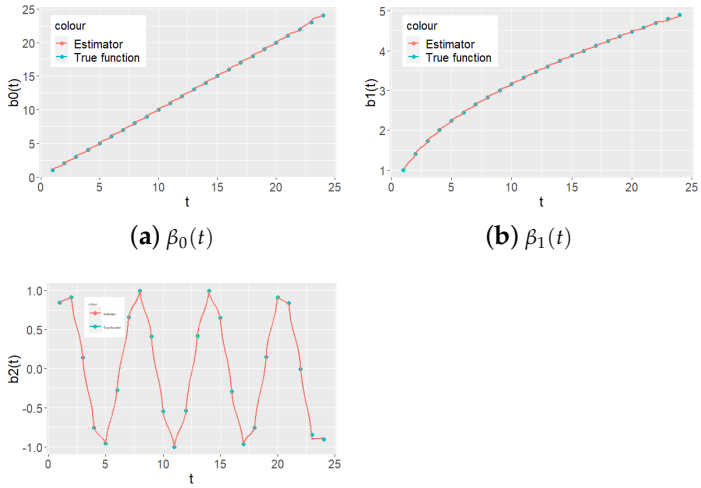
Figures 1 and 2 present the proposed estimated functions of $\boldsymbol{\beta}(t)$ as a continuous curve and true values of $\boldsymbol{\beta}(t)$ at time points $t \in \Delta$ as a series of points. Figures 3 and 4 present the estimated functions of $\boldsymbol{\beta}(t)$ using the local polynomial regression method. In Scenario 1, our method is not significantly superior to polynomial regression. However, in Scenario 2, the term regression method does not fit $\beta_2(t)$ well, and our method estimates $\hat{\beta}_2(t)$ close to the true value. This is because $\beta_2(t)$ is set up as a trigonometric function rather than a linear combination of power functions. It can be seen that while the local polynomial regression method is suitable for power basis functions, the kernel smoothing method fits the function better, even in such cases.



(**a**) $\beta_0(t)$ (**b**) $\beta_1(t)$

(**c**) $\beta_2(t)$

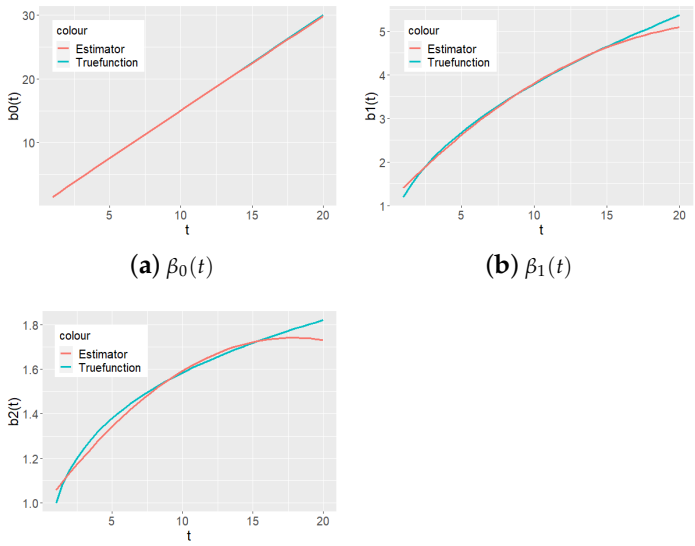
Figure 1. True values at observed time points and estimated function of $\boldsymbol{\beta}(t)$ for scenario 1.



(a) $\beta_0(t)$ **(b)** $\beta_1(t)$

(c) $\beta_2(t)$

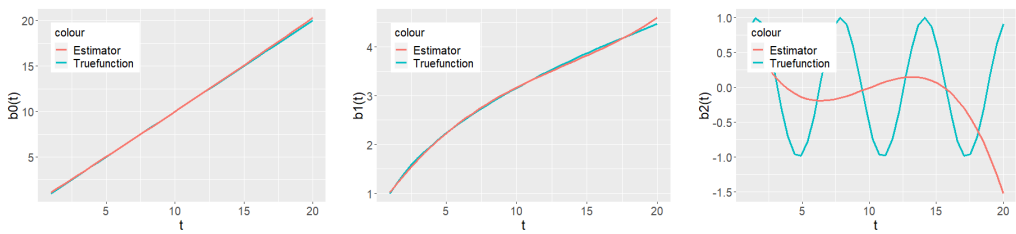
Figure 2. True values at observed time points and estimated function of $\boldsymbol{\beta}(t)$ for scenario 2.



(a) $\beta_0(t)$ **(b)** $\beta_1(t)$

(c) $\beta_2(t)$

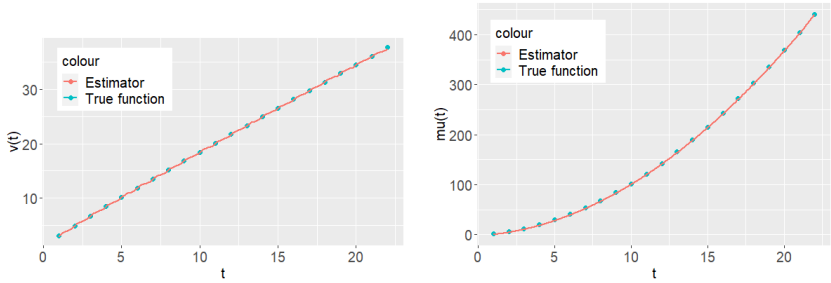
Figure 3. Polynomial regression estimator of $\boldsymbol{\beta}(t)$ for scenario 1.



(a) $\beta_0(t)$ **(b)** $\beta_1(t)$ **(c)** $\beta_2(t)$

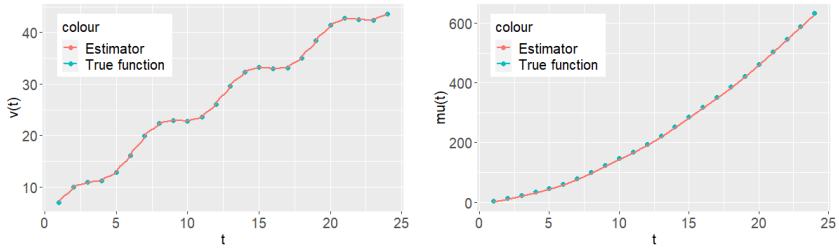
Figure 4. Polynomial regression estimator of $\boldsymbol{\beta}(t)$ for scenario 2.

Furthermore, Figures 5 and 6 show the true curves and fitted values of the state function $v(t)$ and the mean of cumulative mean function $\mu(t)$. In each figure, the estimated functions approximate to the values of true functions. Figures 7 and 8 show the results of the local polynomial regression method. Similar to the previous conclusion, the method we proposed works better.



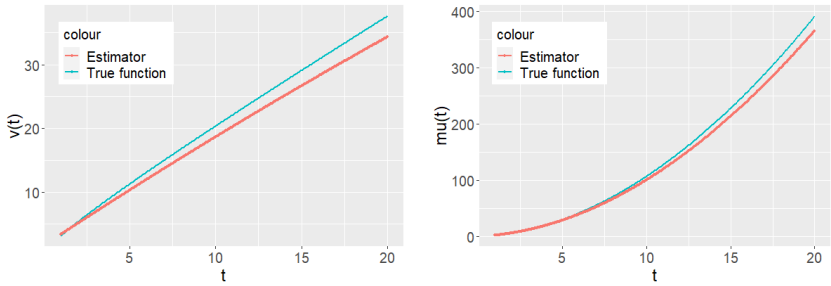
(a) $v(t)$ **(b)** $\mu(t)$

Figure 5. True values at observed time points and estimated function of $v(t)$ and $\mu(t)$ for scenario 1.



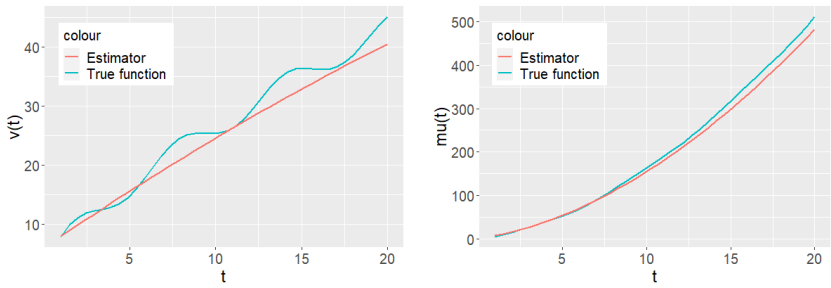
(a) $v(t)$ **(b)** $\mu(t)$

Figure 6. True values at observed time points and estimated function of $v(t)$ and $\mu(t)$ for scenario 2.



(a) $v(t)$ **(b)** $\mu(t)$

Figure 7. Polynomial regression estimator of $v(t)$ and $\mu(t)$ for scenario 1.



(a) $v(t)$ **(b)** $\mu(t)$

Figure 8. Polynomial regression estimator of $v(t)$ and $\mu(t)$ for scenario 2.

4. An Application to MADIT Data

In this section, we validate the proposed estimator with a real data set from a multicenter automatic defibrillator implantation trial (MADIT). MADIT data contain 181 subjects (patients) from 36 centers in the USA that were observed at a total of 134,853 discrete time points. With the 181 patients, 89 of them chose to implant cardiac defibrillators, while

another 92 did not. Throughout this section, we encode the 'implanted' group as $ICD = 1$ and 'not implanted' group as $ICD = 0$. Since the effect of treatment of whether to implant cardiac defibrillators (ICDs) did not directly induce any medical costs but actually affected the expected survival time, we consider it as the time-independent covariate in the survival sub-model.

The observed patients have six types of costs that can be observed daily from the start to the death time or censoring time: Type 1: hospitalization and emergency department visits; Type 2: outpatient tests and procedures; Type 3: physician/specialist visits; Type 4: community services; Type 5: medical supplies; Type 6: medication. These costs directly drive medical costs, thus they should be considered as time-dependent covariates in the longitudinal process sub-model. It should be pointed out that the whole observation contains quite a lot of data points, so the R-code cannot work for the daily cost data. Therefore, we merge the data by combining 12 days into one time unit.

These data also include the patients' ID codes, observed survival times in days, and death indicators.

Now, in the data set, we encode the total 181 patients as follows:

1. ID code (from 1 to 181);
2. Treatment code (1 for ICD and 0 for conventional);
3. Observation of survival time;
4. Death indicator(1 for death, 0 for censored);
5. Merged medical costs of type 1–6.

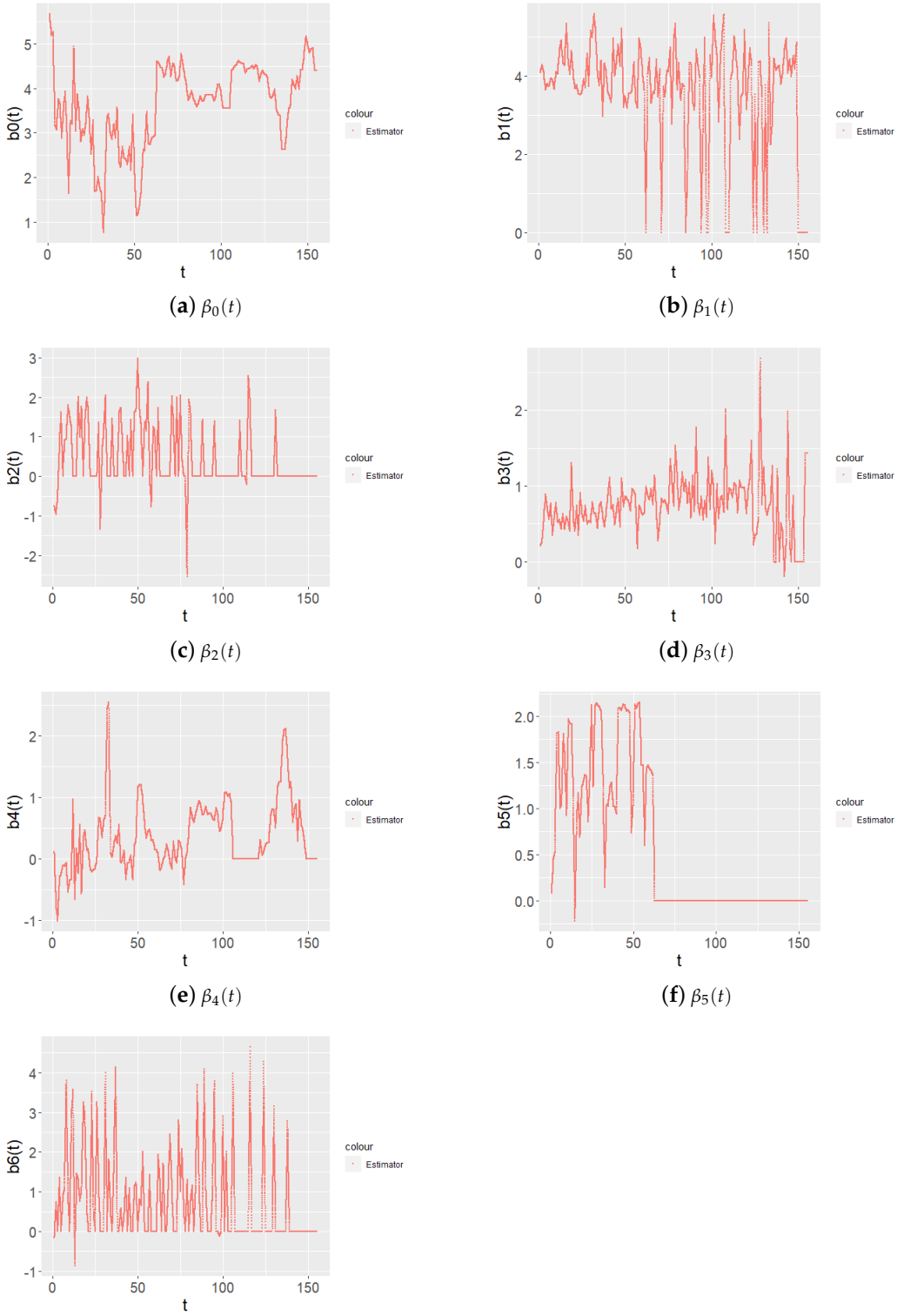
To analyze this data set, we describe the model as follows:

$$\begin{cases} Y_i(t) = \beta_0(t) + \beta_1(t)x_{1i}(t) + \beta_2(t)x_{2i}(t) + \beta_3(t)x_{3i}(t) + \beta_4(t)x_{4i}(t) \\ \qquad + \beta_5(t)x_{5i}(t) + \beta_6(t)x_{6i}(t) + \epsilon_i(t), \\ h_i(t) = h_0(t)\exp\{\gamma w_i + \alpha(Y_i(t) - \epsilon_i(t))\}, \end{cases}$$

where for $r = 1, 2, \ldots, 6$, $x_{ir}(t) = 1$ if type = 1, otherwise, $x_{ir}(t) = 0$, $w_i = 1$ for the ICD group and $w_i = 0$ for the not implanted group. The estimated parameters $\hat{\gamma}$ and $\hat{\alpha}$ in survival sub-model are obtained by Cox partial maximum likelihood method, they are always asymptotically normal. The estimates of γ and α in this paper are attained by calling 'coxph()' and 'method = peicewise-PH-GH' in the JM package. The Std. Err. and p-value are automatically produced.

In this case, $\boldsymbol{\beta}(t)$ can be considered the relationship between the natural logarithm of medical cost and time unit. The fitted curve is illustrated in Figure 9. Time-varying parameters more accurately describe the effect of covariates on state function over time, such as $\beta_5(t)$; after a certain time point, $x_5(5)$ does not affect the state function, although the in early stage it does.

Table 5 shows the estimates of the survival sub-model. The association effect α is positive, which corresponds to reality, that is, patients in a serious life condition require more medical attention. In other words, serious illness leads to higher medical costs, corresponding to lower survival rates. This further supports that our model is efficient and reasonable. However, due to the large p-value of the result, this conclusion is not specific, which should be further tested using a large sample in future research. The treatment effect γ is negative, which corresponds to reality, that is, an automatic defibrillator implantation trial can reduce the risk of death. In other words, the 'implanted' group (ICD = 1) has a lower risk of death. This conclusion is supported by the small p-value of the result.



(a) $\beta_0(t)$ **(b)** $\beta_1(t)$

(c) $\beta_2(t)$ **(d)** $\beta_3(t)$

(e) $\beta_4(t)$ **(f)** $\beta_5(t)$

(g) $\beta_6(t)$

Figure 9. Estimated function of $\beta(t)$.

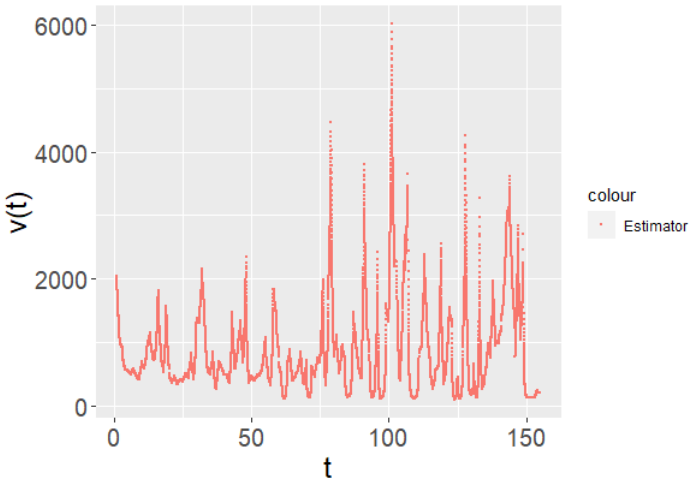
Table 5. Estimate results of the event process.

Parameter	Value	Std. Err.	*p*-Value
Treatment	−1.2219	0.3362	0.0003
Association effect	0.2519	0.2087	0.2274

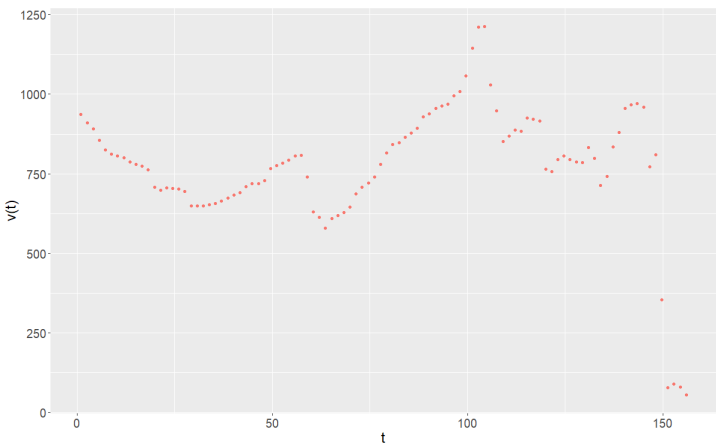
Table 6 presents the estimated values of the cumulative cost for 5 years and the total treatment period. Comparing Sub-figure(a) with Sub-figure(b) in Figures 10 and 11, we realize that the fitted points for the mean medical cost based on the current approach better describe the elaborate trajectories of medical costs, and the result for the fitted points of the cumulative mean medical cost based on the current approach exhibit a similar trend to Li's [19] result; however, the result is more accurate and shows continuous change, which, in turn, leads to improved understanding of real-world data.

Table 6. Estimated cumulative costs for 5 years and total period.

Year 1	Year 2	Year 3	Year 4	Year 5	Total
21,661.83	45,403.10	67,981.31	101,983.30	134,749.20	135,412.00

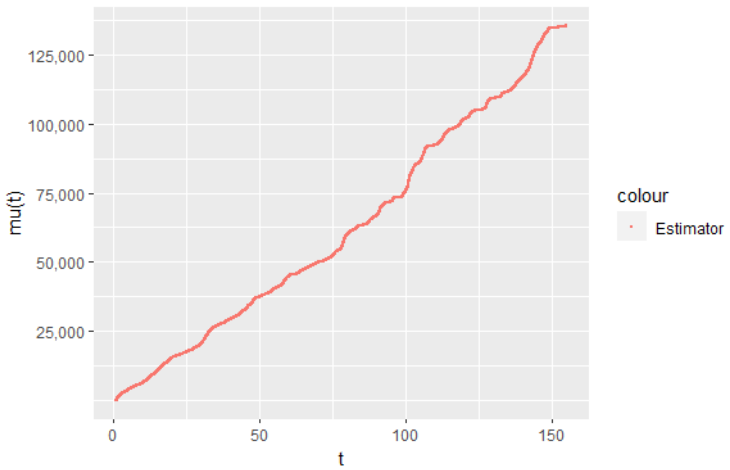


(**a**) Fitted values of cost based on current approach.

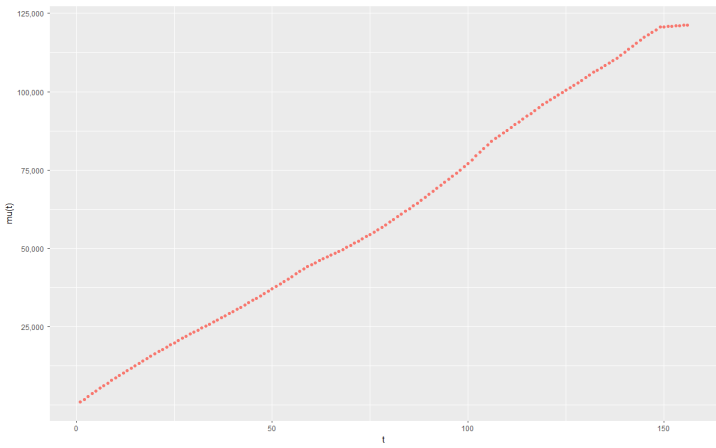


(**b**) Fitted points of cost based on the semi-parametric estimating approach.

Figure 10. Fitted points of cost based on current approach and the semi-parametric estimating approach.



(**a**) Fitted values of cumulative cost based on current approach.



(**b**) Fitted points of cumulative cost based on the semi-parametric estimating approach .

Figure 11. Fitted points of cumulative cost based on current approach and the semi-parametric estimating approach.

Compared with parametric/semi-parametric estimation, our method suggests a smoother time-varying mean medical cost function, which is no longer a straight line with an observable change rate, but exhibits more fluctuations over time. In practical situations, government agencies or insurance companies would not know which treatment was selected by a patient. The estimated cumulative mean function provides them with a more accurate statistical decision recommendation for macro allocation. For example, we should consider the additional changes in medical costs because they would suddenly increase or decrease at distinct time points.

Moreover, the estimated marginal survival function is smoother compared with the Kaplan–Meier estimator. The result is similar to research by Li's [19], as shown in Figure 12. The solid line is the estimated survival rate, and the dotted line represents the 95% confidence interval of the estimated survival rate.

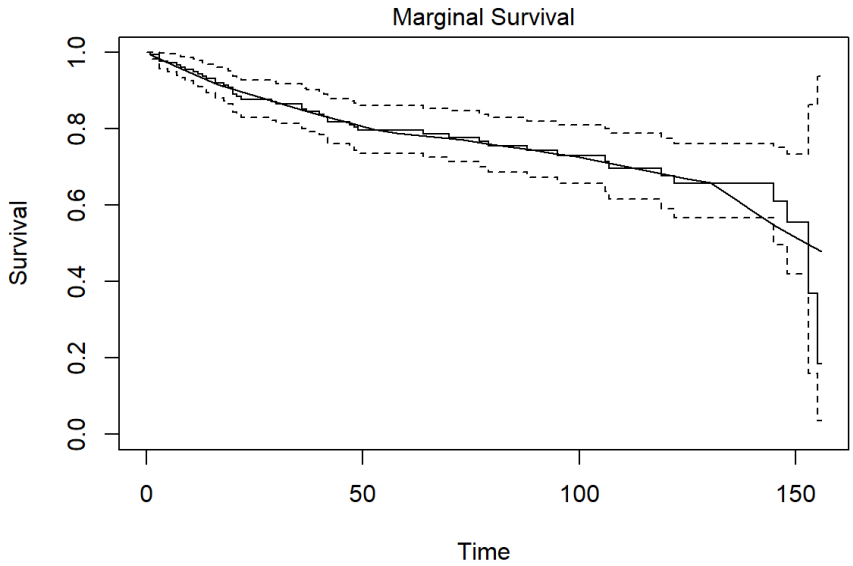


Figure 12. Marginal survival compared with Kaplan–Meier estimator.

5. Discussion

Generally, some estimate methods can lead to biased estimates in the model, and the time-varying coefficient model compensates for this shortcoming. In this study, we provide an estimation of the time-varying coefficient using the kernel function technique. When the covariates are continuous, the continuous state function can be obtained, allowing for the derivation of the continuous cumulative mean state function. In summary, our method is more precise because the kernel function smooths the estimator; more importantly, the proposed estimator at the original observed true points is unbiased. Theorem 2 demonstrates this perspective. In our numerical simulation and application analysis, we consider the observation interval to be a unit of 1, and we chose a bandwidth h of 1 in the kernel smoothing function. Notably, in cases where the observation time interval is not 1 or involves interval censoring, we must choose the appropriate bandwidth. Many studies on kernel function have proposed methods for choosing the bandwidth. In general, too large a bandwidth distorts data, and too small a bandwidth does not lead to a smooth continuous function. Moreover, as $h \to \infty$, $\boldsymbol{\beta}(t)$ turns out to be a series of constants, that is, a time-independent parametric estimator.

In this study, we do not consider the random effect coefficients in the longitudinal data model, and the correlation that may exist among repeated observations within individuals is ignored. If we consider the random effect coefficients, we will not obtain Theorem 2, but we can still obtain Theorem 3. In future studies, random effects should be considered in longitudinal models, which may yield better results.

Furthermore, the estimation of the time-varying coefficient using the kernel function technique can be extended to multiple longitudinal outcomes, and the justification for the survival sub-model would need a multi-dimensional association coefficient. The methods proposed in [20] may be applicable.

Author Contributions: Conceptualization, S.L., D.D. and Y.H.; methodology, S.L. and D.D.; software, S.L. and D.D.; validation, D.D. and Y.H.; formal analysis, S.L.; investigation, S.L.; resources, D.D.; data curation, S.L. and D.D.; writing—original draft preparation, S.L.; writing—review and editing, D.D.; visualization, Y.H.; supervision, D.D. and Y.H.; project administration, D.D.; funding acquisition, Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: Han's work is partially supported by the National Natural Science Foundation of China (grant number 11871244); Deng's work is partially supported by the Natural Sciences and Engineering Research Council of Canada.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. The Asymptotic Normality of $\hat{\beta}(\tau)$

For each $t_{(l)} \in \Delta$, denote

$$\boldsymbol{\beta}(t_{(l)}) = (\beta_1(t_{(l)}), \beta_2(t_{(l)}), \ldots, \beta_q(t_{(l)})).$$

For all $r, c = 0, \ldots, q$, denote

$$b_{rz}(t_{(l)}) = E[x_{ir}(t_{(l)}) I\{T_m^* \geqslant t_{(l)}\} x_{ic}(t_{(l)})];$$

$$\begin{aligned} u_r(t_{(l)}) =& M_0^{3/2} \sum_{z=1}^{q} m_1 [\beta_z^{'}(t_{(l)}) b_{rz}^{'}(t_{(l)}) f((t_{(l)}) \\ &+ \beta_z^{'}(t_{(l)}) b_{rz}(t_{(l)}) f^{'}(t_{(l)}) \\ &+ (1/2) \beta_z^{''}(t_{(l)}) b_{rz}(t_{(l)}) f(t_{(l)})]; \end{aligned}$$

$$\boldsymbol{G}(t_{(l)}) = (f(t_{(l)}))^{-1} E[\mathbf{X}_l^T \mathbf{X}_l]^{-1}(t_{(l)})(u_0(t_{(l)}), \ldots, u_q);$$

$$V_{rz}(t_{(l)}) = m^2(t_{(l)}) b_{rz}(t_{(l)}) f((t_{(l)}) m_2 + UM_0 o_\epsilon(t_{(l)}) b_{rz}(t_{(l)}) f^2((t_{(l)});$$

$$\boldsymbol{V}(t_{(l)}) = \begin{pmatrix} D_{00}(t_{(l)}) & \cdots & D_{0q}(t_{(l)}) \\ \vdots & \ldots & \vdots \\ D_{q0}(t_{(l)}) & \cdots & D_{Qq}(t_{(l)}) \end{pmatrix};$$

$$\boldsymbol{W}(t_{(l)}) = (f(t_{(l)}))^{-2} E[\mathbf{X}_l^T \mathbf{X}_l]^{-1} \boldsymbol{V}(t_{(l)}) E[\mathbf{X}_l^T \mathbf{X}_l]^{-1}.$$

Then,

$$(mh)^{1/2}(\hat{\boldsymbol{\beta}}(t_{(l)}) - \boldsymbol{\beta}(t_{(l)})) \to N(\boldsymbol{G}(t_{(l)}), \boldsymbol{W}(t_{(l)}))$$

Appendix B. Proof of Theorem 2

Proof of Theorem 2. For each $t_{(l)} \in \Delta$,

$$\hat{\boldsymbol{\beta}}(t_{(l)}) = \left(\sum_{k=1}^{N} \mathbf{X}_k^T \tilde{\mathbf{D}}_k(t_{(l)}) \mathbf{X}_k \right)^{-1} \left(\sum_{k=1}^{N} \mathbf{X}_k^T \tilde{\mathbf{D}}_k(t_{(l)}) \mathbf{Y}_k \right),$$

where

$$\tilde{\mathbf{D}}_k(t_{(l)}) = \mathrm{diag}\Big(W_{k,h}(t_{(l)} - t_{(k)}) I\{T_1^* \geqslant t_{(l)}\}, \ldots, W_{k,h}(t_{(l)} - t_{(k)}) I\{T_m^* \geqslant t_{(l)}\} \Big).$$

When $t_{(l)} = t_{(k)}$,

$$\begin{aligned} \tilde{\mathbf{D}}_l(t_{(l)}) &= \mathrm{diag}\Big(W_{k,h}(t_{(l)} - t_{(l)}) I\{T_1^* \geqslant t_{(l)}\}, \ldots, W_{k,h}(t_{(l)} - t_{(l)}) I\{T_m^* \geqslant t_{(l)}\} \Big) \\ &= \mathrm{diag}\Big(I\{T_1^* \geqslant t_{(l)}\}, \ldots, I\{T_m^* \geqslant t_{(l)}\} \Big), \end{aligned}$$

when $t_{(l)} \neq t_{(k)}$,

$$\begin{aligned} \tilde{\mathbf{D}}_k(t_{(l)}) &= \mathrm{diag}\Big(W_{k,h}(t_{(l)} - t_{(k)}) I\{T_1^* \geqslant t_{(l)}\}, \ldots, W_{k,h}(t_{(l)} - t_{(k)}) I\{T_m^* \geqslant t_{(l)}\} \Big) \\ &= \mathrm{diag}(0, \ldots, 0). \end{aligned}$$

$\hat{\boldsymbol{\beta}}(t_{(l)})$ can be written as

$$\begin{aligned}\hat{\boldsymbol{\beta}}(t_{(l)}) &= \left(\mathbf{X}_l^T \tilde{\mathbf{D}}_l(t_{(l)})\mathbf{X}_l\right)^{-1}\left(\mathbf{X}_l^T \tilde{\mathbf{D}}_l(t_{(l)})\mathbf{Y}_l\right)\\ &= \left(\sum_{i=1}^{m} \boldsymbol{x}_i^T(t_{(l)})\tilde{\mathbf{D}}_l(t_{(l)})\boldsymbol{x}_i(t_{(l)})\right)^{-1}\left(\sum_{i=1}^{m} \boldsymbol{x}_i^T(t_{(l)})\tilde{\mathbf{D}}_l(t_{(l)})y_i(t_{(l)})\right).\end{aligned}$$

Under the assumption that $\epsilon(t)$ is independent of $X(t)$, and $E\epsilon(t) = 0$. Then from Gauss–Markov theory,

$$\boldsymbol{E}\hat{\boldsymbol{\beta}}(t_{(l)}) = \boldsymbol{\beta}(t_{(l)}). \tag{A1}$$

Combining Equations (7) and (8), the estimator $\hat{v}(t_{(l)})$ of mean state function at time point $t_{(l)}$ is

$$\hat{v}(t_{(l)}) = \frac{\sum_{i=1}^{m} I\{T_i^* \geqslant t_{(l)}\}\hat{Y}_i(t_{(l)})}{\sum_{i=1}^{m} I\{T_i^* \geqslant t_{(l)}\}},$$

where $\hat{Y}_i(t_{(l)})$ is the fitted value of $Y_i(t_{(l)})$ at time point $t_{(l)}$ from the joint model:

$$\hat{Y}_i(t_{(l)}) = \boldsymbol{x}_i(t_{(l)})\hat{\boldsymbol{\beta}}(t_{(l)}).$$

By Assumption $E(y_i(t) \mid \boldsymbol{x}_i(t)) = E(Y(t) \mid \mathbf{X}(t))$, we have

$$\begin{aligned}v(t_{(l)}) = E(Y(t_{(l)}) \mid \mathbf{X}(t_{(l)})) &= \frac{\sum_{i=1}^{m} I\{T_i^* \geqslant t_{(l)}\}E\left[Y(t_{(l)}) \mid \mathbf{X}(t_{(l)})\right]}{\sum_{i=1}^{m} I\{T_i^* \geqslant t_{(l)}\}}\\ &= \frac{\sum_{i=1}^{m} I\{T_i^* \geqslant t_{(l)}\}E\left[y_i(t_{(l)}) \mid \boldsymbol{x}_i(t_{(l)})\right]}{\sum_{i=1}^{m} I\{T_i^* \geqslant t_{(l)}\}}\\ &= \frac{\sum_{i=1}^{m} I\{T_i^* \geqslant t_{(l)}\}\left[\boldsymbol{x}_i(t_{(l)})\boldsymbol{\beta}(t_{(l)})\right]}{\sum_{i=1}^{m} I\{T_i^* \geqslant t_{(l)}\}}.\end{aligned}$$

Thus, for any $t_{(l)} \in \Delta$, from Equation (A1), we have

$$\begin{aligned}E\hat{v}(t_{(l)}) - v(t_{(l)}) &= \frac{\sum_{i=1}^{m} I\{T_i^* \geqslant t_{(l)}\}E\left[\boldsymbol{x}_i(t_{(l)})\hat{\boldsymbol{\beta}}(t_{(l)})\right]}{\sum_{i=1}^{m} I\{T_i^* \geqslant t_{(l)}\}} - \frac{\sum_{i=1}^{m} I\{T_i^* \geqslant t_{(l)}\}\left[\boldsymbol{x}_i(t_{(l)})\boldsymbol{\beta}(t_{(l)})\right]}{\sum_{i=1}^{m} I\{T_i^* \geqslant t_{(l)}\}}\\ &= \frac{\sum_{i=1}^{m} I\{T_i^* \geqslant t_{(l)}\}\left[\boldsymbol{x}_i(t_{(l)})\right]}{\sum_{i=1}^{m} I\{T_i^* \geqslant t_{(l)}\}}\left(E\hat{\boldsymbol{\beta}}(t_{(l)}) - \boldsymbol{\beta}(t_{(l)})\right) = 0.\end{aligned}$$

□

This completes the proof of Theorem 2.

Appendix C. Proof of Theorem 3

Proof of Theorem 3. From Equation (8), we have

$$\begin{aligned}\hat{v}(t) - v(t) =& \frac{1}{m}\sum_{i=1}^{m}\frac{I\{T_i \geqslant t\}}{\hat{S}_T(t)}\hat{Y}_i(t) - EY(t)\\ =& \frac{1}{m}\sum_{i=1}^{m}\frac{I\{T_i \geqslant t\}}{\hat{S}_T(t)S_{T(t)}}\hat{Y}_i(t)[S_{T(t)} - \hat{S}_T(t)]\\ &+ \frac{1}{m}\sum_{i=1}^{m}\frac{I\{T_i \geqslant t\}}{S_T(t)}(\hat{Y}_i(t) - Y_i(t))\\ &+ \frac{1}{m}\sum_{i=1}^{m}\frac{I\{T_i \geqslant t\}}{S_T(t)}Y_i(t) - EY(t)\\ =& (I) + (II) + (III).\end{aligned}$$

Under the assumption that $\epsilon(t)$ is independent of T and C conditional on $X(t)$ and W, from the law of large numbers,

$$(III) = \frac{1}{m}\sum_{i=1}^{m}\frac{I\{T_i \geqslant t\}}{S_T(t)}Y_i(t) - EY(t) \to 0 \quad \text{a.s. as } m \to \infty.$$

For the second term (II), we have

$$\begin{aligned}(II) =& \frac{1}{m}\sum_{i=1}^{m}\frac{I\{T_i \geqslant t\}}{S_T(t)}(\hat{Y}_i(t) - Y_i(t))\\ =& \frac{1}{m}\sum_{i=1}^{m}\frac{I\{T_i \geqslant t\}}{S_T(t)}\boldsymbol{x}_i(t)(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t))\\ &+ \frac{1}{m}\sum_{i=1}^{m}\frac{I\{T_i \geqslant t\}}{S_T(t)}(-\epsilon_i(t))\\ \equiv& (IV) + (V).\end{aligned}$$

By Assumption (6), $\| \boldsymbol{x}_i(t) \| \leqslant C_0$ for some positive constants C_0. By Theorem 3.1 in Zeng and Cai [17] and Assumption (11), we have that

$$(IV) \leqslant \frac{1}{m}\sum_{i=1}^{m}\frac{I\{T_i \geqslant t\}}{S_T(t)} \| \boldsymbol{x}_i(t) \| \| \hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t) \| \leqslant \frac{C_0}{a^2} \| \hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t) \| \quad \text{a.s. as } m \to \infty.$$

By Theorem 1 in Wu et al. [12] and Assumptions (1)–(5), we have that

$$(IV) \leqslant \frac{C_0}{a^2} \| \hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t) \| \to 0 \quad \text{a.s. as } m \to \infty.$$

Also, from the law of large numbers,

$$\begin{aligned}(V) =& \frac{1}{m}\sum_{i=1}^{m}\frac{I\{T_i \geqslant t\}}{S_T(t)}(-\epsilon_i(t))\\ =& E\left(\frac{I\{T \geqslant t\}}{S_T(t)}(-\epsilon(t))\right) = \frac{E(I\{T \geqslant t\})}{S_T(t)}E(-\epsilon(t)) = 0 \quad \text{a.s. as } m \to \infty\end{aligned}$$

Similarly, by Assumption (9), Theorem 3.1 in Zeng and Cai [17], and Theorem 2 in Phadia and Ryzin [21], we have that

$$\begin{aligned}(I) &= \frac{1}{m}\sum_{i=1}^{m}\frac{I\{T_i \geqslant t\}}{\hat{S}_T(t)S_T(t)}\boldsymbol{x}_i(t)\hat{\boldsymbol{\beta}}(t)[S_T(t)-\hat{S}_T(t)] \\ &\leqslant \frac{1}{m}\sum_{i=1}^{m}\frac{I\{T_i \geqslant t\}}{\hat{S}_T(t)S_T(t)} \parallel \boldsymbol{x}_i(t) \parallel\parallel \hat{\boldsymbol{\beta}}(t) \parallel\mid S_T(t)-\hat{S}_T(t) \mid \\ &\leqslant \frac{M_0C_0^2}{a^4} \mid S_T(t)-\hat{S}_T(t) \mid \to 0 \quad \text{a.s. as } m \to \infty.\end{aligned}$$

Now, $\hat{\nu}(\tau)$ converges to $\nu(\tau)$ almost uniformly in $\tau \in [0, T]$.
Then, based on Egoroff Theorem in Bartle [22],$\hat{\mu}(\tau)$ converges to $\mu(\tau)$ in probability. □

This completes the proof of Theorem 3.

References

1. Bang, H.; Tsiatis, A. Estimating medical costs with censored data. *Biometrika* **2000**, *87*, 329–343. [CrossRef]
2. Lin, D.Y.; Feuer, E.J.; Etzioni, R.; Wax, Y. Estimating medical costs from incomplete follow-Up data. *Biometrics* **1997**, *53*, 419–434. [CrossRef] [PubMed]
3. Huang, Y.; Lovato, L. Tests for lifetime utility or cost via calibrating survival time. *Stat. Sin.* **2002**, *12*, 707–723.
4. Rizopoulos, D. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*; CRC Press: Boca Raton, FL, USA, 2012.
5. Deng, D. Estimating the cumulative mean function for history process with time-dependent covariates and censoring mechanism: Estimating the cumulative mean function for history process. *Stat. Med.* **2016**, *35*, 4624–4636. [CrossRef] [PubMed]
6. Korn, E.L. On estimating the distribution function for quality of life in cancer clinical trials. *Biometrika* **1993**, *80*, 535–542. [CrossRef]
7. Hoover, D.R.; Rice, J.A.; Wu, C.O.; Yang, L. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **1998**, *85*, 809–822. [CrossRef]
8. Zhao, X.; Tong, X.; Sun, L. Joint analysis of longitudinal data with dependent observation times. *Stat. Sin.* **2012**, *22*, 317–336. [CrossRef]
9. Li, C.; Xiao, L.; Luo, S. Joint model for survival and multivariate sparse functional data with application to a study of Alzheimer's Disease. *Biometrics* **2022**, *78*, 435–447. [CrossRef] [PubMed]
10. Do, H.; Nandi, S.; Putzel, P.; Smyth, P.; Zhong, J. A joint fairness model with applications to risk predictions for under-represented populations. *Biometrics* **2022**, *79*, 826–840. [CrossRef] [PubMed]
11. Eubank, R.L.;Speckman, P.L. Confidence bands in nonparametric regression. *J. Am. Stat. Assoc.* **1993**, *88*, 1287–1301.
12. Wu, C.O.; Chiang, C.T.; Hoover, D.R. Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Am. Stat. Assoc.* **1998**, *93*, 1388–1402. [CrossRef]
13. You, L.; Qiu, P. Joint modeling of multivariate nonparametric longitudinal data and survival data: A local smoothing approach. *Stat. Med.* **2021**, *40*, 6689–6706. [CrossRef] [PubMed]
14. Silverman, B.W. *Density Estimation: For Statistics and Data Analysis*; Chapman & Hall: London, UK, 2018.
15. Rice, J.A.; Silverman, B.W. Estimating the Mean and Covariance Structure Nonparametrically When the Data are Curves. *J. R. Stat. Soc. Ser. Methodol.* **1991**, *53*, 233–243. [CrossRef]
16. Kaplan, E.L.; Meier, P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **1958**, *53*, 457–481. [CrossRef]
17. Zeng, D.; Cai, J. Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *Ann. Stat.* **2005**, *33*, 2132–2163. [CrossRef]
18. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023.
19. Li, S.; Deng, D.; Han, Y.; Zhang, D. Joint model for estimating the asymmetric distribution of medical costs based on a history process. *Symmetry* **2023**, *15*, 2130. [CrossRef]
20. Kenyon, J.R. Analysis of Multivariate Survival Data. *Technometrics* **2002**, *44*, 86–87. [CrossRef]
21. Phadia, E.G.; Ryzin, J.V. A note on convergence rates for the product limit estimator. *Ann. Stat.* **1980**, *8*, 673–678. [CrossRef]
22. Bartle, R.G. *The Elements of Integration and Lebesgue Measure*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2011.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.