Kernel Estimation of Bivariate Time-Varying Coefficient Model for Longitudinal Data with Terminal Event

Yue Wang, Bin Nan, and John D. Kalbfleisch

Department of Statistics, University of California, Irvine, Irvine, CA; Department of Biostatistics, University of Michigan, Ann Arbor, MI

ABSTRACT

We propose a nonparametric bivariate time-varying coefficient model for longitudinal measurements with the occurrence of a terminal event that is subject to right censoring. The time-varying coefficients capture the longitudinal trajectories of covariate effects along with both the followup time and the residual lifetime. The proposed model extends the parametric conditional approach given terminal event time in recent literature, and thus avoids potential model misspecification. We consider a kernel smoothing method for estimating regression coefficients in our model and use cross-validation for bandwidth selection, applying undersmoothing in the final analysis to eliminate the asymptotic bias of the kernel estimator. We show that the kernel estimates follow a finite-dimensional normal distribution asymptotically under mild regularity conditions, and provide an easily computed sandwich covariance matrix estimator. We conduct extensive simulations that show desirable performance of the proposed approach, and apply the method to analyzing the medical cost data for patients with end-stage renal disease. Supplementary materials for this article are available online.

1. Introduction

In longitudinal studies, it is often the case that the collection of repeated measurements is stopped by the occurrence of some terminal event, for example, death. There are two sets of widely used approaches for modeling longitudinal measures with a terminal event: the joint modeling approach using latent frailty and the marginal estimating equation approach using Inverse Probability Weighting (IPW). Under the joint modeling framework, the survival time and the longitudinal process are assumed independent conditional on some latent random effects. Thorough reviews of this type of approach can be found in Tsiatis and Davidian (2004) and Rizopoulos (2012). For the marginal estimating equation approach with IPW, readers can refer to Robins, Rotnitzky, and Zhao (1995). These ideas have also been applied to modeling recurrent events in the presence of a terminal event, see, for example, Kalbfleisch et al. (2013) and Ghosh and Lin (2002). They may fall short in certain situations, however. First, as pointed out by Kong et al. (2018), they do not explicitly model the association between the terminal event time and the longitudinally measured response variable, which is of primary interest in many applications. Second, in health studies where death is a terminal event, some approaches treat the occurrence of death as “dropout,” either informative or non-informative, which implicitly defines the underlying longitudinally measured stochastic processes of health status beyond death. In other words, death causes “missing data” in such a view, which is questionable since death itself is a fundamental characteristic of health.

For these reasons, several reverse-time models have been considered in the recent literature. Chan and Wang (2010) considered a nonparametric approach for the mean of a reverse-time process. Li, Tosteson, and Bakitas (2013) considered a likelihood-based approach for the reverse-time model with applications to palliative care, with extension to a semiparametric approach introduced in Li et al. (2017). Dempsey and McCullogh (2018) considered reverse alignment as a general technique for constructing models for survival processes and investigated several related statistical consequences. These methods model backward time with event time as the time origin, but lose the interpretation of chronological time effects that are of primary interest in conventional longitudinal studies. To keep the desired chronological time interpretation of regression coefficients and meanwhile to describe the effect of terminal event in longitudinal studies, Kong et al. (2018) proposed a parametric nonlinear regression model conditional on the terminal event time which builds the residual lifetime into covariate effects. They showed that the complete case analysis that only uses data with uncensored event times is a valid approach, and proposed a two-stage approach that improves the efficiency of parameter estimates of the complete case analysis. But a parametric model can be easily misspecified, as we observe in our data example, and their two-stage method cannot handle time-varying covariates that occur overwhelmingly often in longitudinal studies.

In this article, we propose a nonparametric extension of Kong et al. (2018). In particular, regression coefficients are bivariate functions of both chronological followup time $t$ and residual lifetime $T - t$ with unknown form, where $t$ denotes the...
followup time and $T$ denotes the terminal event time. Moreover, time-varying covariates are incorporated in our model. Such a modeling strategy allows us to assess the varying effect of certain covariate when patients approach death, which is of particular interest for the analysis of End-Stage Renal Disease (ESRD) medical cost data. We estimate the regression coefficients using kernel smoothing and establish the asymptotic normality of kernel estimates together with convergence rate that depends on the bandwidth size. We also provide a consistent sandwich variance estimator that helps construct pointwise confidence bands.

The rest of the article is organized as follows. In Section 2 we introduce the time-varying coefficient model and the kernel estimating method with bandwidth determined via undersmoothing after cross-validation. We outline the asymptotic properties in Section 3. We provide simulations in Section 4 and the analysis of ESRD medical cost data in Section 5. We give a few concluding remarks in Section 6, and provide detailed proofs and additional numerical results in the supplementary material.

2. Modeling Strategy and Estimating Method

2.1. Bivariate Time-Varying Coefficient Model

Let $Y(t)$ be a stochastic process denoting the response variable measured over time in a longitudinal study. Let $X(t) = (X_1(t), \ldots, X_p(t))$ be $p$ covariate processes. Note that we use bold letter to represent either a vector or a matrix in this article. Suppose the longitudinal cohort data consists of $n$ independent copies of $(Y(t), X(t))$, representing $n$ individuals’ observations in the study cohort, where the $i$th individual’s data $(Y_i(t), X_{i1}(t), \ldots, X_{ip}(t))$ are measured at random time points $\tau_{ij}, j = 1, \ldots, m_i$. Baseline covariates take constant values over time. We define $X_{it}(t) \equiv 1$ for any $i$ and $t$, which determines the intercept. Suppose each individual has $m$ visits, where $m$ is a finite number, but not all of them are observed because of early stopping due to terminal event or right censoring, which makes the number of actual visits varying among individuals. Specifically for subject $i$, denote the terminal event time as $T_i$ and the right censoring time as $C_i$, then the number of visits of subject $i$ is $m_i = \max\{j : j \leq m, \tau_{ij} \leq T_i \land C_i\}$, where $a \land b = \min\{a, b\}$. Denote the set of subjects whose terminal events are observed by $D = \{i : T_i \leq C_i\}$.

We consider the following model for the longitudinal response variable $Y_i$ observed at time $\tau_{ij}$:

$$Y_i(\tau_{ij}) = \sum_{k=1}^{p} X_{ik}(\tau_{ij}) \beta_k(\tau_{ij}, T_i - \tau_{ij}) + \epsilon_i(\tau_{ij}), \quad (2.1)$$

where each $\epsilon_i(t)$ is a zero-mean stochastic process with variance function $\sigma^2(t)$ and covariance function $\rho(t_1, t_2)$ for any $t_1 \neq t_2$. Assume all the quantities involved in this model are iid across individuals, which include $\{\tau_{ij}\}_{i=1}^{m_i}, T_i, C_i, \{X_{ik}(\cdot)\}_{k=1}^{p}$ and $\epsilon_i(\cdot)$. Here iid is defined for processes on any finite index set. Suppressing the subscript $i$ here without causing any confusion, we further assume that for each individual we have: (a) given $\tau_j = t$, $\epsilon(t)$ has the same distribution as $\epsilon(t)$ and is independent of $T$ and $C$ and $\{X_{ik}(\tau_j)\}_{k=1}^{p}$; (b) given $(\tau_{j1}, \tau_{j2}) = (t_1, t_2)$, $(\epsilon(\tau_{j1}), \epsilon(\tau_{j2}))$ has the same distribution as $(\epsilon(t_1), \epsilon(t_2))$ and is independent of $T$, $C$, $\{X_k(\tau_j)\}_{k=1}^{p}$ and $\{X_k(\tau_j)\}_{k=1}^{p}$.

In other words, data observed on a set of random times behave like observed on a set of constant times, which is commonly assumed in longitudinal data analysis. Because we do not estimate the survival function for the complete case analysis considered in this article, we do not need to assume $T$ and $C$ are independent or conditionally independent given covariates, which is a crucial assumption in traditional survival analysis.

Unlike the usual time-varying coefficient model for longitudinal data, a particularly important feature of model (2.1) is that the unknown coefficient $\beta_k(t, T - t)$ is allowed to be a bivariate function not only varying with time since entry, $t$ (the usual setup, see, for example, Hoover et al. 1998), but also varying with time from $t$ to the terminal event, $T - t$ (also referred to as residual lifetime if $T$ is death time). Unlike any conventional modeling strategy for longitudinal data with terminal event, allowing $\beta_k$ to depend on $T - t$ directly captures the way in which impending failure modifies the effect of $X_k(t)$. If none of the $\beta_k, k = 1, \ldots, p$, varies with $T - t$, then the above model (2.1) reduces to a standard time-varying coefficient model.

Model (2.1) extends Kong et al. (2018) from a parametric model to a nonparametric model, from an intercept varying with $T - t$ only to all regression coefficients varying with both $t$ and $T - t$, and from fixed baseline covariates only to time-varying covariates. The model also extends Lu et al. (2010), who only considered a nonparametric intercept varying with $T - t$ without pursuing the asymptotic properties of their spline based estimating method. Li et al. (2018) considered a bivariate mean model, included no covariates and did not provide asymptotic results for their spline estimating method.

It becomes clear that model (2.1) is well-defined when $T_i$ is observed, so all the observations collected from time at entry to $T_i$ are complete data, whereas observations collected from time at entry to $C_i$ before $T_i$ are incomplete. This is another major distinction between a model that is conditional on $T_i$ and conventional regression models for longitudinal data with terminal events. Since $T_i$ is subject to right censoring, the problem determined by model (2.1) becomes a regression problem with censored covariate, for which the complete case analysis is a valid approach. This is the method we consider in this article for estimating unknown bivariate coefficient functions $\beta_k, k = 1, \ldots, p$. Including observations for censored individuals faces multifaceted difficulties, which will be discussed in Section 6.

2.2. Bivariate Kernel Estimation

For any fixed point $(t_0, t_0)$, we apply bivariate kernel smoothing to estimate $\beta(t_0, t_0)$ by minimizing the following loss function with respect to $\beta_k, k = 1, \ldots, p$:

$$I_n(t_0, t_0) = \sum_{i \in D} \sum_{j=1}^{m} \left( Y_{ij} - \sum_{k=1}^{p} X_{ik}(\tau_{ij}) \beta_k \right)^2 K \left( \frac{t_j - t_0}{h}, \frac{T_i - t_j - t_0}{h} \right), \quad (2.2)$$

where $K : \mathbb{R}^2 \rightarrow \mathbb{R}$ is the kernel function, $h > 0$ is the bandwidth. There are two major distinctions between the resulting estimator from (2.2) and the estimator in Hoover et al. (1998): First, the estimator in (2.2) involves terminal event time
and is based on complete data. Second, since β_k’s are bivariate functions, a bivariate kernel function is used. Note that we use the same bandwidth for both time axes and ignore the off-diagonal element of the $2 \times 2$ bandwidth matrix in order to simplify the numerical implementation. Using multiple bandwidths requires multiple bandwidth selection procedures thus is more computationally cumbersome especially for large datasets. Later we show both theoretically and numerically that the kernel estimation using a common bandwidth performs satisfactorily.

Rewrite (2.2) into the following matrix form:

$$ L_n(t_0, s_0) = \sum_{i \in D} (Y_i - X_i b)^T K_i(t_0, s_0; h) (Y_i - X_i b), $$

where

$$ X_i = \begin{pmatrix} X_{i1}(\tau_{i1}) & \ldots & X_{ip}(\tau_{i1}) \\ \vdots & \ddots & \vdots \\ X_{i1}(\tau_{im}) & \ldots & X_{ip}(\tau_{im}) \end{pmatrix}, $$

$$ K_i(t_0, s_0; h) $$

is a diagonal matrix with jth element given by

$$ K((t_j - t_0)/h, (T_i - t_j - s_0)/h)^2, $$

and $Y_i = (Y_{i1}, \ldots, Y_{im})^T$. We estimate the time-varying coefficients $\beta_k(t, s)$ by minimizing $L_n(t_0, s_0)$ with respect to $b_k$, $k = 1, \ldots, p$, that is,

$$ \hat{\beta}(t_0, s_0; h) = \arg\min_{\beta} L_n(t_0, s_0), $$

which has a closed form solution given by

$$ \hat{\beta}(t_0, s_0; h) = \left( \sum_{i \in D} X_i^T K_i(t_0, s_0; h) X_i \right)^{-1} \left( \sum_{i \in D} X_i^T K_i(t_0, s_0; h) Y_i \right). $$

(2.3)

The estimator (2.3) ignores the within-subject correlation following the working independence assumption, which was shown by Lin and Carroll (2000) to be most efficient when a standard kernel is applied and the cluster size is finite. This counter-intuitive result was explained by Wang (2003) who also showed that higher efficiency could be achieved by using an alternative kernel method, which we do not pursue here because of both the numerical advantages of the working independence assumption and the efficiency result of Lin and Carroll (2000) for using a standard kernel in (2.3).

### 2.3. Automatic Bandwidth Selection and Undersmoothing

A typical approach for automatic bandwidth selection is through $K$-fold cross-validation (CV). To keep the independence between training set and validation set, we partition the longitudinal data at the subject level such that all repeatedly measured observations of each subject belong to only 1-fold. The criterion for selecting bandwidth is to minimize the average predictive squared errors across all validation sets. In particular, let $S_k$, $k = 1, \ldots, K$, be the index set of subjects in the kth fold, where $\cup_k S_k = D$ and $S_k \cap S_l = \emptyset$ for any $k \neq l$, then the average predictive squared error criterion is given by

$$ CV(h) = \frac{1}{\sum_{i \in D} m_i} \sum_{k=1}^K \sum_{i \in S_k} \sum_{j=1}^{m_i} [Y_{ij} - X_i \hat{\beta}^{(-k)}(\tau_{ij}, T_j - t_j; h)]^2, $$

(2.4)

where $\hat{\beta}^{(-k)}$ represents the kernel estimator calculated by leaving out all observations in the kth fold. Note that only the complete cases $D$ are partitioned into folds. In practice, the criterion is minimized on a preselected grid of $h$.

Standard approaches to constructing nonparametric confidence bands for functions are complicated by the impact of bias. According to Hall (1992), bias decreases as the amount of statistical smoothing is reduced, which can be clearly seen from the asymptotic distributional results of kernel estimates. Therefore, one way of alleviating bias is to smooth the curve estimator less than would be optimal for point estimation. We choose to undersmooth by multiplying $n^{-\gamma}$ to the selected bandwidth using cross-validation for some $\gamma > 0$. We will see in Section 3 that undersmoothing still leads to the desirable asymptotic result as long as the undersmoothed bandwidth falls into the range specified by Condition 2 in Appendix A.

### 2.4. A Special Case with Potentially Improved Efficiency

Under a special circumstance where only baseline covariates (or the so-called defined time-dependent covariates that are completely determined by baseline covariates and the time $t$) are of concern as in Kong et al. (2018), or in an even less likely situation where time-varying covariates are of interest but the terminal event time only depends on baseline covariates, one might assume Gaussian error in model (2.1) and consider the following locally weighted pseudo likelihood function under working independence which includes both complete and censored data:

$$ \prod_{i=1}^n \left\{ \prod_{j=1}^{m_i} \left[ \frac{1}{\sqrt{2\pi}\sigma_{ij}} e^{-\frac{1}{2\sigma_{ij}^2}(Y_{ij} - X_{ij}^T b)^2} \right] K\left(\frac{t_{ij} - t_0, T_i - t_j - s_0}{h}\right) f_1(T_i|Z_i) \right\} \Delta_j $$

$$ \times \left\{ \int_{C_i} \cdots \int_{C_i} \prod_{j=1}^{m_i} \left[ \frac{1}{\sqrt{2\pi}\sigma_{ij}} e^{-\frac{1}{2\sigma_{ij}^2}(Y_{ij} - X_{ij}^T b)^2} \right] K\left(\frac{t_{ij} - t_0, T_i - t_j - s_0}{h}\right) dP(T_i \leq u|Z_i) \right\}^{1-\Delta_i} f_2(X_i, \tau_i, Z_i). $$

Note that, similar to Kong et al. (2018), the assumption of conditional independent censoring is also needed to obtain the above locally weighted pseudo likelihood function. In contrast, no assumption on the censoring mechanism is needed for the complete case analysis. In the above, $Z_i$ is a vector of baseline covariates of subject $i$, $f_1(\cdot|Z_i)$ is the survival density given $Z_i$ and $f_2$ is the joint density of $X_i$, $\tau_i$, and $Z_i$. The nuisance parameter $\sigma_{ij}$ denotes $\sigma(\tau_{ij})$ and can be estimated by the kernel estimator $\hat{\sigma}_{ij}^2 = (nh)^{-1} \sum_{j'} \hat{\sigma}_{ij'}^2 K'(\tau_{ij'} - \tau_{ij})/h$, where $K'$ is some univariate kernel function, and another nuisance parameter $P(T_i \leq u|Z_i)$ can be estimated using a proper survival model. The estimation of $\hat{\beta}(t, s)$ can be obtained by maximizing the above locally weighted pseudo likelihood with respect to $b$. A simulation study, summarized in the supplementary material, shows that this method seems to yield valid results with improved efficiency over the complete cases analysis. Its theoretical justification, however, is beyond the scope of this work thus not pursued here, and as discussed in Section 6, several additional difficulties
preclude the application of this approach to the case with time-varying covariates that are commonly observed in longitudinal studies.

3. Asymptotic Properties

3.1. Asymptotic Normality of $\hat{\beta}$

We consider the asymptotic joint normality of $\hat{\beta}$ at a finite number of pairs of distinct time points $\{(t_1, s_1), \ldots, (t_d, s_d)\}$. Under several mild regularity conditions given in the Appendix A, we can show that the estimator (2.3) follows a multivariate normal distribution as $n$ approaches infinity. First we introduce some notation:

$$
\mu_0 = \int K^2(x, y)dx dy,
$$

$$
\mu_2 = \left(\int x^2 K(x, y)dx dy, \int x K(x, y)dx dy, \int y^2 K(x, y)dx dy\right),
$$

$$
\eta_j(t, s) = E[1(T \leq \tau_j)|X(\tau_j)]= t - \tau_j = s,
$$

$$
g_{j,k}(t, s) = \nabla \eta_j(t, s) f_j(t, s) \nabla \beta_k(t, s)^T + \frac{1}{2} \eta_j(t, s) \nabla^2 \beta_k(t, s)^T + \frac{1}{2} \eta_j(t, s) \nabla^2 \beta_k(t, s)^T,
$$

$$
\Gamma(t, s) = \sum_{j = 1}^{m} \sum_{k = 1}^{p} (\mu_2, g_{j,k}(t, s)),
$$

$$
\Omega(t, s) = \sum_{j = 1}^{m} \eta_j(t, s) f_j(t, s).
$$

In the above, $\eta_j(t, s)$ is the $(r, k)$-element of $p \times p$ matrix $\eta_j(t, s)$; $f_j$ denotes the joint density of $\tau_j$ and $T - \tau_j$; $\nabla \beta_k(t, s)$ is the $2 \times 1$ gradient of $\beta_k$ at $(t, s)$ and $\nabla^2 \beta_k(t, s)$ is the $2 \times 2$ Hessian matrix; $\langle A, B \rangle$ is the Frobenius inner product of matrices $A$ and $B$, that is, $\langle A, B \rangle = tr(A^T B)$. Note that the subscript $t$ is suppressed for random variables in above $\eta_j(t, s)$ to simplify the notation since these are all defined for a generic subject and observations are assumed iid.

Furthermore, the following quantities are defined at a set of finite number of distinct time points $\{(t_1, s_1), \ldots, (t_d, s_d)\}$. Let $t = (t_1, \ldots, t_d)^T$ and $s = (s_1, \ldots, s_d)^T$. Define

$$
\hat{\beta}(t, s; h) = \left(\begin{array}{c}
\hat{\beta}(t_1, s_1; h) \\
\vdots \\
\hat{\beta}(t_d, s_d; h)
\end{array}\right)_{dp \times 1},
$$

$$
\beta(t, s) = \left(\begin{array}{c}
\beta(t_1, s_1) \\
\vdots \\
\beta(t_d, s_d)
\end{array}\right)_{dp \times 1},
$$

$$
B(t, s) = \left(\begin{array}{c}
\Omega^{-1}(t_1, s_1) \Gamma(t_1, s_1) \\
\vdots \\
\Omega^{-1}(t_d, s_d) \Gamma(t_d, s_d)
\end{array}\right)_{dp \times 1},
$$

$$
V(t, s) = \mu_0 \left(\begin{array}{c}
\sigma^2(t_1) \Omega^{-1}(t_1, s_1) \\
\vdots \\
\sigma^2(t_d) \Omega^{-1}(t_d, s_d)
\end{array}\right)_{dp \times dp}.
$$

Theorem 3.1. For a finite integer $d$ and fixed vectors $t$ and $s$ satisfying $t_i, s_i > 0$ for $i = 1, \ldots, d$ and $t_i \neq t_{i'}, s_i \neq s_{i'}$ when $l \neq l'$, under regularity conditions 1, 2a, 2b, and 3–7 given in Appendix A, we have

$$
n^{1/2} h(\hat{\beta}(t, s; h) - \beta(t, s)) \to_d N(h_0^2 B(t, s), V(t, s))
$$

as $n \to \infty$, where $h_0$ is defined as the limit of $n^{-1/20}$, which shows satisfactory performance in simulations.

3.2. Sandwich Estimator and Pointwise Confidence Interval

With undersmoothing, the asymptotic bias in Theorem 3.1 disappears when $n$ goes to infinity. Hence, for any pair of time points in Theorem 3.1, denoted by $(t_0, s_0)$, we only need to estimate the variance $V(t_0, s_0)$ in order to construct the pointwise confidence interval. It turns out that the following sandwich estimator is a valid variance estimator:

$$
\hat{\nu}(t_0, s_0) = nh^2 \left(\sum_{i \in D} X_i^T K_{0i} X_i\right)^{-1} \left(\sum_{i \in D} X_i^T K_{0i} \tilde{e}_i \tilde{e}_i^T K_{0i} X_i\right)^{-1} \left(\sum_{i \in D} X_i^T K_{0i} X_i\right)^{-1},
$$

where $\tilde{e}_i$ is the residual vector for the $i$th subject and $K_{0i}$ is short for $K_i(t_0, s_0; h)$. The elements of $\tilde{e}_i$ are calculated by

$$
\hat{e}_{ij} = \hat{e}_i(t_j) = Y_i(t_j) - \sum_{k = 1}^{p} X_{ik}(t_j) \hat{\beta}_k(t_j, T_i - t_j), \quad 1 \leq j \leq m_i.
$$

The following theorem demonstrates the consistency of (3.4).

Theorem 3.2. For the covariance matrix in Theorem 3.1, under regularity conditions 1, 2a, 2c, and 3–7 given in Appendix A, we have

$$
\hat{\nu}(t_0, s_0) \to_p V(t_0, s_0)
$$

as $n \to \infty$, where $V(t_0, s_0)$ is the corresponding diagonal block matrix in $V(t, s)$ at $(t_0, s_0)$. 
With this theorem, an approximate $1 - \alpha$ pointwise confidence interval of $\beta_k(t_0, s_0)$ without bias correction can be constructed as

$$\hat{\beta}_k(t_0, s_0) \pm z_{\alpha/2} \left( \frac{1}{nh^2} \hat{v}(t_0, s_0)_{kk} \right)^{1/2}.$$ 

4. A Simulation Study

This section reports the numerical performance of the kernel estimator (2.3). For the simulation study, consider model (2.1) with $p = 3$ and $m = 20$, where the coefficients are the following functions:

$$\beta_1(x, y) = \frac{x}{4} \exp \left( -\frac{x^2 + y^2}{100} \right),$$
$$\beta_2(x, y) = \frac{1}{2} \left[ \sin \left( \frac{2x}{5} \right) - \sin \left( \frac{y}{2} \right) \right],$$
$$\beta_3(x, y) = \cos \left( \frac{x^2 + y^2}{100} \right).$$

These functions are similar to those used in Wu, Chiang, and Hoover (1998). We generate visiting times in the following way: for subject $i$, the first visit time $\tau_{i1}$ is generated uniformly on $[0, 1]$, then $\tau_{ij}, j > 1$, is generated independently from $\tau_{ij} - (j - 1) \sim \text{Beta}(\tau_{i1}/4\nu^2, (1 - \tau_{i1})/4\nu^2)$, where $\nu$ serves as an upper bound of standard deviation of the Beta distribution and is set to be 0.01. The generated interarrival time $\tau_{ij} - \tau_{ij-1}$ falls into $[0, 2]$ with mean 1 and a very small variance. Thus, the generated visiting schedule is approximately evenly spaced, mimicking a designed longitudinal study with annual visits. For covariates, $X_1$ is always 1, $X_2$ is generated from a standard normal distribution, and $X_3(t)$ is a mean-zero Gaussian process with covariance function $\text{cov}(X_3(t), X_3(s)) = \exp(-|t-s|^2)$. Moreover, $X_2$ and $X_3(t)$ are correlated with covariance $\text{cov}(X_2, X_3(t)) = 0.8 \exp(-t^2)$. Terminal event time $T$ is generated from an exponential distribution with rate $\exp(3X_{12} + X_{33}(0) - 5)$, then truncated at 15 and added 5. Thus, $T \geq 5$ with probability 1. Censoring time follows a uniform distribution between 5 and $2T - 5$. This leads to dependent censoring and yields 50% censoring rate. The error term $\varepsilon_i(\tau_{ij})$ is generated by a nonhomogeneous Ornstein-Uhlenbeck (NOU) process $U_i(t)$ plus a random error. The NOU process satisfies $\text{var}(U_i(t)) = \exp(1 - 0.1t)$ and $\text{corr}(U_i(t_1), U_i(t_2)) = 0.5|t_1 - t_2|$, and the random error follows a standard normal distribution.

With this design, we simulate 1000 independent replications, each with a sample size $n = 4000$ that is about 10% of the sample size of the ESRD data analyzed in the next section. The kernel function is the density of a standard bivariate normal distribution truncated by a circle around $(0,0)$ which contains probability 0.95. To save the computing cost, we first run 5-fold cross-validation on 10 independent datasets with a grid search on $\{1.2^i : i = -10, \ldots, 10\}$, which yield an average bandwidth of 1. We then undersmooth it to obtain a bandwidth of $1 \times n^{-1/20} \approx 0.66$ and fix it for all the 1000 simulation replications. To achieve a better visual effect of bivariate functions, we plot a few slices of estimated coefficients. Specifically, we plot $\hat{\beta}_1(t, T - t)$ varying with $t$ at $T = 8, 12,$ and 16, separately, which are the estimated covariate effects from the time of entry to the terminal event for individuals who died at time 8, 12, and 16. Among the $3 \times 3$ panels in Figure 1, each row represents a time-varying coefficient.
Figure 2. Coverage probabilities of 95% pointwise confidence intervals for $\beta_1$, $\beta_2$, and $\beta_3$. The dotted line shows the nominal level of 0.95.

and each column represents one chosen value of $T$. There are six curves in each panel: the true function $\beta_k$ (solid), the sample mean of estimators $\hat{\beta}_k$ (long dashed), upper and lower 95% confidence bands calculated by the sample mean $\pm 1.96$ times the sample standard deviation of the estimates (dashed), and the sample averages of upper and lower 95% confidence bands calculated using the sandwich variance estimates (dot-dashed). We can see that across all panels, the undersmoothing yields negligible biases, and the two types of confidence bands are nearly identical, indicating the validity of the proposed variance estimator.

To have a clear view of the performance of the 95% pointwise confidence intervals, in Figure 2 we further provide their coverage probabilities to the same coefficient curves depicted in Figure 1. From Figure 2 we see that the coverage probability is mostly around 95%, but can drop to near 85% on the boundaries or regions with large curvature of the coefficient due to relatively large biases of kernel smoothing in such regions.

At the request of an anonymous reviewer, we have implemented simulations with a much smaller sample size of $n = 400$, and provided results in Section D, Figure S2, of the supplementary material. The overall performance is very similar, with a slightly larger bias for $\beta_2$ at $T = 8$ and, unsurprisingly, wider pointwise confidence bands.

We further verify the performance of hypothesis testing based on the asymptotic multivariate normal distribution given in Theorem 3.1. Such a test allows for comparing coefficient values at any two different pairs of time points. In order to also systematically evaluate the size of the test, we modify the simulation setup slightly by setting $\beta_3 = 0.5$ while keeping other simulation parameters unchanged, which creates a true null hypothesis of $\Delta \beta_3 = 0$. In Tables 1 and 2 we summarize the empirical power, or size when the null hypothesis is true, of a two-sided $z$-test. The empirical power is the rejection frequency of the test among 1000 simulation replications. In Table 1 we consider cases with the same failure time $T$ but two different visit times $t_1$ and $t_2$. In Table 2 we consider the same visit time $t$ but two different failure times $T_1$ and $T_2$. We see from simulation results presented in both Tables 1 and 2 that empirical sizes of the tests are close to 0.05, showing the validity of the test. We also observe larger empirical powers for larger magnitudes of the coefficient differences.

| $t_1, t_2, T$ | $\Delta \beta_1$ (EP$_1$) | $\Delta \beta_2$ (EP$_2$) | $\Delta \beta_3$ (EP$_3$) |
|--------------|--------------------------|--------------------------|--------------------------|
| 2, 4, 8      | 0.391(0.980)             | -0.146(0.197)            | 0(0.056)                 |
| 2, 6, 8      | 0.670(1.000)             | -0.786(1.000)            | 0(0.054)                 |
| 4, 6, 8      | 0.279(0.905)             | -0.640(1.000)            | 0(0.053)                 |
| 2, 4, 12     | 0.273(0.654)             | 0.610(0.995)             | 0(0.054)                 |
| 2, 6, 12     | 0.553(0.998)             | 0.616(0.993)             | 0(0.055)                 |
| 4, 6, 12     | 0.281(0.737)             | 0.006(0.026)             | 0(0.053)                 |
| 2, 4, 16     | 0.134(0.141)             | 0.038(0.032)             | 0(0.051)                 |
| 2, 6, 16     | 0.317(0.706)             | 0.214(0.298)             | 0(0.066)                 |
| 4, 6, 16     | 0.183(0.295)             | 0.176(0.201)             | 0(0.057)                 |
Liu, Wolfe, and Kalbfleisch (2007) found an increasing pattern between observation time and time to death. The pattern of end-of-life Medicare cost response is the daily inpatient cost paid by Medicare and the stage renal disease (ESRD) from year 2007 to 2018 collected by for the hypothesis test $H_0 : \Delta \beta_k = 0$ versus $H_0 : \Delta \beta_k \neq 0, k = 1, 2, 3$.

| $T_1$, $T_2$, $t$ | $\Delta \beta_1$ (EP1) | $\Delta \beta_2$ (EP2) | $\Delta \beta_3$ (EP3) |
|------------------|------------------------|------------------------|------------------------|
| 8, 12, 2         | $0.158(0.272)$         | $0.637(0.990)$         | $0.060(0.000)$         |
| 8, 16, 2         | $0.267(0.564)$         | $0.872(1.000)$         | $0.058(0.000)$         |
| 12, 16, 2        | $0.109(0.139)$         | $0.235(0.322)$         | $0.063(0.000)$         |
| 8, 12, 4         | $0.277(0.685)$         | $0.119(0.218)$         | $0.058(0.000)$         |
| 8, 16, 4         | $0.524(0.990)$         | $0.668(1.000)$         | $0.055(0.000)$         |
| 12, 16, 4        | $0.247(0.524)$         | $0.807(1.000)$         | $0.048(0.000)$         |
| 8, 12, 6         | $0.275(0.749)$         | $0.765(1.000)$         | $0.039(0.000)$         |
| 8, 16, 6         | $0.620(1.000)$         | $0.128(0.210)$         | $0.064(0.000)$         |
| 12, 16, 6        | $0.345(0.849)$         | $0.637(0.995)$         | $0.059(0.000)$         |

### 5. The ESRD Medicare Data Analysis

We consider inpatient medical cost data of patients with end-stage renal disease (ESRD) from year 2007 to 2018 collected by the United States Renal Data System (USRDS). The longitudinal response is the daily inpatient cost paid by Medicare and the terminal event is death. The pattern of end-of-life Medicare cost has been identified in previous work. For example, Chan and Wang (2010) showed an increasing and then decreasing pattern in Medicare costs before death among ovarian cancer patients. Liu, Wolfe, and Kalbfleisch (2007) found an increasing pattern in monthly outpatient EPO costs starting from 6 months prior to death and an initial jump since entry time, followed by a linear drop. When it comes to inpatient cost among ESRD patients, Kong et al. (2018) established similar initial pattern as in Liu, Wolfe, and Kalbfleisch (2007) and an increasing then decreasing terminal pattern using a parametric model. Here we aim to investigate the patterns with a much larger sample size using our nonparametric modeling approach.

Following Kong et al. (2018), we only include black and white patients who started their ESRD services in 2007 and were at least 65 years old when they started. We exclude patients who received kidney transplant because they could potentially have very different trajectories of inpatient costs. We also exclude patients who never had any hospitalization nor filled out the CMS Medical Evidence Report during the follow-up. Instead of selecting a simple random sample of available ESRD patients for the analysis as in Kong et al. (2018), all eligible patients are included in our analysis. Additionally, we are able to take advantage of the most updated data from USRDS, for which the follow-up ended on June 30th, 2018. We end up with a much larger sample size of 42,253 patients who died before the end of follow-up, much longer follow-up with an average of 34.6 months, and a very low censoring rate of only 3.74%. In the original data, a total cost is given for each hospitalization period. The initial pattern is different too: in our analysis, we find that the initial pattern is increasing and then decreasing, followed by a linear drop. However, in Kong et al. (2018), the peak value, $\Delta \beta_1 + \Delta \beta_2$, corresponds to Medicare the secondary payer. The coefficient $\Delta \beta_1$ represents the Medicare payment difference between payer types, but plotting $\beta_1 + \beta_2$ provides a direct illustration of the medical cost pattern of Medicare payment as the secondary payer. Similar to how we display simulation results, we choose to only visualize $\beta_k(t-T-1)$ under several fixed values of $T$. Here we choose $T = 360, 900$, and 1440, corresponding to patients who died roughly 1 year, 2.5 years and 4 years after entry, respectively.

In Figure 3, we plot the estimated curves and their confidence bands for $\beta_1$ and $\beta_1 + \beta_2$, respectively, obtained using a selected bandwidth of 12 days, where $\beta_1$ represents the log-transformed daily Medicare payment trajectory among ESRD patients when Medicare is the primary payer, and $\beta_1 + \beta_2$ corresponds to Medicare the secondary payer. The coefficient $\beta_1$ provides a direct illustration of the medical cost pattern of Medicare payment as the secondary payer. Similar to how we display simulation results, we choose to only visualize $\beta_k(t-T-1)$ under several fixed values of $T$. Here we choose $T = 360, 900$, and 1440, corresponding to patients who died roughly 1 year, 2.5 years and 4 years after entry, respectively.

From Figure 3 we see that the Medicare cost as primary payer starts to escalate from roughly 150 days prior to death, similar to the pattern observed in Kong et al. (2018). The peak value, however, is at the time of death, which is different to Kong et al. (2018) where the peak was around three weeks before death. The initial pattern is different too: in our analysis, we find that the Medicare cost decreases drastically in the first two months after entry, then becomes stabilized overtime until close to death, whereas in Kong et al. (2018), and in Liu, Wolfe, and Kalbfleisch (2007) as well, it increases first then decreases. Such differences are likely due to the restrictive parametric assumptions imposed in Kong et al. (2018) and Liu, Wolfe, and Kalbfleisch (2007). We also find that, when Medicare is secondary payer, the pattern of inpatient costs is similar but the magnitude is much smaller and at most times very close to zero. This is anticipated because a large portion of medical costs was paid first by some other insurance.

We further conduct formal statistical tests of Medicare payment patterns toward the end of life to confirm a key difference between our findings and Kong et al. (2018). These tests are based on the finite-dimensional asymptotic independence and joint normality established in Theorem 3.1 for the intercept parameter $\beta_1$. We have showed in Section 4 that such a test has also been applied successfully to a large sample size of ESRD patients.
the correct power to detect the coefficient difference. In particular, for a given \( T \), we consider three different residual lifetimes: 90 days, 30 days and 15 days prior to death. The tests compare \( \beta_1 \) at 90 versus 30 days and at 30 versus 15 days, respectively. Table 3 summarizes each estimated difference \( \beta_1(t_2, T - t_2) - \beta_1(t_1, T - t_1) \) with its 95% confidence interval, and the \( p \)-value of a one-sided test, showing significant evidence of a continuously increasing pattern of the intercept parameter \( \beta_1 \) toward the end of life for each of three different values of \( T = 360, 900, 1440 \) days. In other words, these tests reject the pattern obtained in Kong et al. (2018) where the medical cost peaks and then declines during the last month before death.

It is well-known that kernel methods may yield larger biases at the boundaries. The boundary biases, however, diminish as the sample size increases (see Fan, Heckman, and Wand 1995) and can be further reduced by undersmoothing that is the approach we take in the ESRD data analysis. Although the choice of the undersmoothing factor \( n^{1/20} \) seems working well in our simulations, in practice it is hard to determine what choice of bandwidth would result in a reasonable undersmoothing. However, given the fact that the estimated values of \( \beta_1 \) are very large with confidence intervals far from zero at the boundaries, it is reasonable to believe that the medical spending is much higher at the time when an ESRD patient becomes Medicare eligible and at the end of a patient’s life.

It would be also of interest to test the medical spending pattern from a different angle, in other words, to compare \( \beta_1 \) with the same \( t \) but different \( T \). Several such comparisons are shown in Table 4, where we see higher daily medical cost among subjects with shorter residual lifetime at the same time since enrollment. Such difference diminishes among longer survivors, however.

### 6. Discussion

We propose a nonparametric bivariate time-varying coefficient model for longitudinal data and a kernel estimating method based on complete cases analysis, that is, using data when terminal events are available, which is shown to be a valid approach. Our main contributions are 2-fold: (a) The flexible nonparametric modeling of longitudinal associations that not only vary with chronological observational time, but also depend on residual lifetime, with support defined on a triangular region. This particular modeling strategy provides meaningful interpretations for longitudinal studies with study participants being followed during their lifespans, as for the motivating example of medical spending for ESRD patients considered in this article. We believe this is the first time that such a nonparametric conditional modeling strategy given a terminal event is considered in the literature. (b) Nontrivial theoretical development showing the desirable large sample behavior of the bivariate kernel smoothing method for our flexible nonparametric model under less restrictive conditions. Specifically, unlike any standard survival analysis method, the censoring mechanism to the terminal event does not have any impact to the validity of the estimating method.

---

**Table 3.** Estimated difference \( \beta_1(t_2, T - t_2) - \beta_1(t_1, T - t_1) \), its 95% confidence interval, and \( p \)-value for the monotonicity test: \( H_0 : \beta_1(t_1, T - t_1) = \beta_1(t_2, T - t_2) \) versus \( H_a : \beta_1(t_1, T - t_1) < \beta_1(t_2, T - t_2) \).

| \( T - t_1 \) | \( T - t_2 \) | \( T - t_1 = 90, T - t_2 = 30 \) | \( T - t_1 = 30, T - t_2 = 15 \) |
|----------------|----------------|----------------|----------------|
| \( T = 360 \)  | \( 7.56 \times 10^{-5} \) | 0.004, 0.118 | 2.61 \( \times 10^{-5} \) |
| \( T = 900 \)  | 3.78 \( \times 10^{-3} \) | 0.028, 0.113 | 5.31 \( \times 10^{-4} \) |
| \( T = 1440 \) | 1.11 \( \times 10^{-3} \) | 0.026, 0.132 | 1.79 \( \times 10^{-3} \) |

**Table 4.** Estimated difference \( \beta_1(t, T_2 - t) - \beta_1(t, T_1 - t) \), its 95% confidence interval, and \( p \)-value for the monotonicity test: \( H_0 : \beta_1(t, T_1 - t) = \beta_1(t, T_2 - t) \) versus \( H_a : \beta_1(t, T_1 - t) < \beta_1(t, T_2 - t) \).

| \( T_1 = 900, T_2 = 360 \) | \( t = 90 \) | 0.014, (−0.004, 0.033) | 6.56 \( \times 10^{-2} \) |
| \( T_1 = 1440, T_2 = 900 \) | \( t = 180 \) | 0.030, (0.012, 0.047) | 4.67 \( \times 10^{-4} \) |
|                          | \( t = 270 \) | 0.087, (0.065, 0.110) | 1.02 \( \times 10^{-14} \) |
In contrast, Kong et al. (2018) considered a parametric model for baseline covariates only. They proposed a two-stage likelihood based estimating method that uses all observations, including both censored and uncensored observations. Clearly parametric models suffer from model misspecification, but Kong et al. (2018) showed that, with correctly specified longitudinal and survival models, including censored observations improves efficiency over the complete case analysis. We pointed out in Section 2.4 that it may be possible to include censored observations and implement a pseudo likelihood estimating approach in our nonparametric framework, but under the undesirable assumption that the terminal event time only depends on baseline covariates, together with the normal error assumption that is not required by our proposed kernel smoothing method. Moreover, it is difficult to generalize the pseudo likelihood approach when a survival model involves time-varying covariates due to following two reasons: (a) It requires estimating the survival function \( P(T_i > t | \bar{X}_i(t)) \) beyond censoring time, that is, \( t > C_i \), if one wants to include censored individual \( i \), where \( \bar{X}_i(t) \) is the history of time-dependent covariates \( X_i \) up to time \( t \). This would require modeling of \( \bar{X}_i(t) \), which is otherwise unnecessary as in most regression problems, and even when it can be done, this approach would introduce measurement errors that are difficult to estimate. (b) Oftentimes in longitudinal studies, the time-varying covariates are internal covariates (Kalbfleisch and Prentice 2002) that make the conditional survival function undefined, leading to invalid likelihood based inference. The ESRD data analysis in previous section falls into this category where the indicator variable of Medicare as secondary payer is time-varying and clearly an internal covariate. Hence, we argue that the complete case analysis is most appropriate for longitudinal data with time-varying covariates, as in our analysis of the USRDS data. Additionally, efficiency loss should not be a concern in our analysis of the USRDS data because of the very low censoring rate. It is worth pointing out that even under scenarios where efficiency loss could be a serious concern, our proposed method can be used as an exploratory tool for finding an appropriate parametric model.

The kernel estimator (2.3) and its asymptotic properties are motivated by Wu, Chiang, and Hoover (1998) with extensions to bivariate time-varying coefficients. However, there are major differences between our setup and theirs. First, we assume fixed maximum number of observations \( m \) instead of letting \( m \) go to infinity as the sample size \( n \) approaches infinity. This is categorized as the sparse functional case by Li and Hsing (2010) and induces a simplified version of the asymptotic variance, where the correlation between observations vanishes. Second, each observation time \( \tau_j, 1 \leq j \leq m \), is allowed to have its own distribution rather than being iid. As a result, \( \tau_j \) can flexibly depend on the history up to \( \tau_{j-1} \), which reflects more practical settings of longitudinal studies.

Extension of the proposed modeling strategy to generalized linear models can be of interest. In particular, if regression coefficients are univariate functions of either one of the time components \( t \) and \( T - t \), then the local linear smoothing method of Sun et al. (2019) may be directly applicable for the complete case analysis. For bivariate coefficient functions considered in this article, however, local linear smoothing would need to estimate several additional unknown functions, and its implementation and theoretical justification for longitudinal data with a terminal event would be of future interest.

### Appendix A: Regularity Conditions

Denote by \( C^q \) the class of functions with \( q \)th order continuous derivatives. For the points \((t_l, s_j), l = 0, 1, \ldots, d \) in Theorems 3.1 and 3.2, we need the following regularity conditions.

1. \( K \) is a probability density of the form \( f(\|x\|_2) \), where \( f(\cdot) \) is of bounded variation on bounded support.
2. (a) \( n^{1/4}h \rightarrow h_0 < \infty \), where \( h_0 \geq 0 \).
   (b) \( nh^2 \rightarrow \infty \).
   (c) \( n^{3/4}h^2/ \log n \rightarrow \infty \).
3. (a) For any \( j, k, l, n_j \), \( f_j(x, y) \) is of class \( C^1 \) in a neighborhood of \((t_l, s_j)\).
   (b) For any \( j, k, l, n_j \), \( E [X_j (\tau_j)^8 | \tau_j = x, T - \tau_j = y] \) is bounded in a neighborhood of \((t_l, s_j)\).
   (c) For any \( j_1 \neq j_2 \) and \( k, E [X_j (\tau_j)^8 | \tau_j = x, \tau_{j_2} = y, T - \tau_{j_1} = \varepsilon] \) is bounded in a neighborhood of \((t_l, t_l, s_j)\).
4. (a) For any \( j, f_j \) is of class \( C^2 \) in a neighborhood of \((t_l, s_j)\).
   (b) For any \( j, f_j \) is of class \( C^2 \) in a neighborhood of \((t_l, t_l, s_j)\).
5. For any \( k, \phi_k \) is of class \( C^2 \) in a neighborhood of \((t_l, s_j)\).
6. (a) \( \sigma^2 \) is continuous at \( t_l \).
   (b) \( E(\cdot)^3 \) is bounded in a neighborhood of \( t_l \).
7. There exists \( j \) such that \( \eta_j(t_l, s_j) \) is positive definite and \( f_j(t_l, s_j) \) is positive.

Remark: Most of the regularity conditions are direct extensions of those in Wu, Chiang, and Hoover (1998) to the bivariate case. Specifically, Condition 1 ensures that \( K \) has a compact support on \( \mathbb{R}^2 \) and is symmetric, that is,

\[
\iint xK(x, y)dx dy = 0, \quad \iint yK(x, y)dx dy = 0.
\]

Conditions 2a and 2b, or 2a and 2c, together specify a range of feasible bandwidths, which justifies the use of undersmoothed bandwidth. Note that, pointed out by an anonymous reviewer, Condition 2c gives a more restrictive lower bound for bandwidth \( h \) and is only required for Theorem 3.2. For Theorem 3.1, more relaxed Condition 2b is sufficient. The exponent 3/4 in Condition 2c is a result of finite moments of \( X \) and \( \varepsilon \) in Conditions 3b and 6b. A similar condition is given in Einmahl and Mason (2005) for the Nadaraya-Watson estimator. In particular, when \( Y \) is assumed to have a finite 8th order moment, which is a result of our conditions 3b, 3c, and 6b, the bandwidth \( h \) in Einmahl and Mason (2005) needs to satisfy \( (n/ \log n)^{3/4}h \rightarrow \infty \) for a univariate kernel estimator. We would also like to point out that the assumptions of finite higher order moments for \( X \) and \( \varepsilon \) in Conditions 3b, 3c, and 6b automatically hold for sub-Gaussian or sub-exponential processes. Lastly, Condition 7 ensures that \( \sum_{j \in D} K_j X_j / n \) is invertible asymptotically. This is commonly assumed for regression models.

### Supplementary Materials

The supplementary material contains detailed proofs of main theorems and additional numerical results.

### Acknowledgments

The data reported here have been supplied by the United States Renal Data System (USRDS). The interpretation and reporting of these data are the responsibility of the authors and in no way should be seen as an official policy or interpretation of the U.S. government.
Disclosure Statement

The authors report there are no competing interests to declare.

Funding

This work was supported in part by NIH grants R01AG056764 and RF1AG075107, and by NSF grant DMS-1915711.

References

Chan, K. C. G., and Wang, M.-C. (2010), “Backward Estimation of Stochastic Processes with Failure Events as Time Origins,” *The Annals of Applied Statistics*, 4, 1602–1620. [1102,1108]

Dempsey, W., and McCullagh, P. (2018), “Survival Models and Health Sequences,” *Lifetime Data Analysis*, 24, 550–584. [1102]

Einmahl, U., and Mason, D. M. (2005), “Uniform in Bandwidth Consistency of Kernel-Type Function Estimators,” *The Annals of Statistics*, 33, 1380–1403. [1110]

Fan, J., Heckman, N. E., and Wand, M. P. (1995), “Local Polynomial Kernel Regression for Generalized Linear Models and Quasi-Likelihood Functions,” *Journal of the American Statistical Association*, 90, 141–150. [1109]

Ghosh, D., and Lin, D. Y. (2002), “Marginal Regression Models for Recurrent and Terminal Events,” *Statistica Sinica*, 12, 663–688. [1102]

Hall, P. (1992), “Effect of Bias Estimation on Coverage Accuracy of Bootstrap Confidence Intervals for a Probability Density,” *The Annals of Statistics*, 20, 675–694. [1104]

Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L.-P. (1998), “Nonparametric Smoothing Estimates of Time-varying Coefficient Models with Longitudinal Data,” *Biometrika*, 85, 809–822. [1103]

Kalbfleisch, J. D., and Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data* (2nd ed.), Hoboken, NJ: Wiley. [1110]

Kalbfleisch, J. D., Schaubel, D. E., Ye, Y., and Gong, Q. (2013), “An Estimating Function Approach to the Analysis of Recurrent and Terminal Events,” *Biometrics*, 69, 366–374. [1102]

Kong, S., Nan, B., Kalbfleisch, J. D., Saran, R., and Hirth, R. (2018), “Conditional Modeling of Longitudinal Data With Terminal Event,” *Journal of the American Statistical Association*, 113, 357–368. [1102,1103,1104,1108,1109,1110]

Li, L., Wu, C.-H., Ning, J., Huang, X., Shih, Y.-C. T., and Shen, Y. (2018), “Semiparametric Estimation of Longitudinal Medical Cost Trajectory,” *Journal of the American Statistical Association*, 113, 582–592. [1103]

Li, Y., and Hsing, T. (2010), “Uniform Convergence Rates for Nonparametric Regression and Principal Component Analysis in Functional/Longitudinal Data,” *The Annals of Statistics*, 38, 3321–3351. [1110]

Li, Z., Frost, H., Tosteson, T. D., Zhao, L., Liu, L., Lyons, K., Chen, H., Cole, R., Currow, D., and Bakitas, M. (2017), “A Semiparametric Joint Model for Terminal Trend of Quality of Life and Survival in Palliative Care Research,” *Statistics in Medicine*, 36, 4692–4704. [1102]

Li, Z., Tosteson, T. D., and Bakitas, M. A. (2013), “Joint Modeling Quality of Life and Survival Using a Terminal Decline Model in Palliative Care Studies,” *Statistics in Medicine*, 32, 1394–1406. [1102]

Lin, X., and Carroll, R. J. (2000), “Nonparametric Function Estimation for Clustered Data When the Predictor is Measured without/with Error,” *Journal of the American statistical Association*, 95, 520–534. [1104]

Liu, L., Wolfe, R. A., and Kalbfleisch, J. D. (2007), “A Shared Random Effects Model for Censored Medical Costs and Mortality,” *Statistics in Medicine*, 26, 139–155. [1108]

Liu, X., Nan, B., Song, P., and Sowers, M. (2010), “Longitudinal Data Analysis with Event Time as a Covariate,” *Statistics in Biosciences*, 2, 65–80. [1103]

Rizopoulos, D. (2012), *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*, Boca Raton, FL: CRC press. [1102]

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995), “Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data,” *Journal of the American Statistical Association*, 90, 106–121. [1102]

Sun, Y., Qi, L., Heng, F., and Gilbert, P. B. (2019), “Analysis of Generalized Semiparametric Mixed Varying-Coefficients Models for Longitudinal Data,” *Canadian Journal of Statistics*, 47, 352–373. [1110]

Tsatis, A. A., and Davidian, M. (2004), “Joint Modeling of Longitudinal and Time-to-Event Data: An Overview,” *Statistica Sinica*, 14, 809–834. [1102]

Wang, N. (2003), “Marginal Nonparametric Kernel Regression Accounting for Within-Subject Correlation,” *Biometrika*, 90, 43–52. [1104]

Wu, C. O., Chiang, C.-T., and Hoover, D. R. (1998), “Asymptotic Confidence Regions for Kernel Smoothing of a Varying-Coefficient Model with Longitudinal Data,” *Journal of the American statistical Association*, 93, 1388–1402. [1106,1108,1110]