axioms

MDPI

Article

A Joint Model for Longitudinal Data with Terminal Event: Integrating Time-Varying Coefficients, Residual Life Effects, and Censored Data

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

[1] College of Big Data and Software Engineering, Zhejiang Wanli University, Ningbo 315100, China; lisimeng@zwu.edu.cn
[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, a terminal event such as death typically halts data collection, and patients may exhibit marked changes as they approach the event. Existing methods face three related challenges: interpreting the effect of residual life on the longitudinal response, accommodating time-varying covariates and coefficients, and retaining information from censored individuals. We propose a joint model that addresses these challenges in a unified likelihood framework. The longitudinal submodel includes an explicit residual-life effect $g(T_i - t, \boldsymbol{\xi})$, time-varying covariates $\mathbf{X}_i(t)$ with time-varying coefficients $\boldsymbol{\beta}(t)$, and shared random effects. Exponential-decay and Gaussian-kernel specifications are considered for $g(\cdot)$. The survival submodel depends on the latent state $m_i(t) = \mathbf{X}_i(t)^\top \boldsymbol{\beta}(t) + \mathbf{Z}_i(t)^\top \mathbf{b}_i$. The coefficient functions are approximated by B-splines, and the full observed-data likelihood is maximized numerically with random-effects integrals evaluated by Gauss–Hermite quadrature. Under the stated regularity conditions, we establish consistency, root-n asymptotic normality of the finite-dimensional parameters, and the sieve convergence rate of the time-varying coefficient functions. Simulation studies demonstrate accurate recovery of the coefficient and residual-life functions. In the MADIT application, the fitted hazard ratio for ICD implantation is 0.433, and the residual-life effect indicates an increase in medical costs approximately three weeks before death.

Keywords: joint modeling; longitudinal data; terminal event; time-varying coefficients; residual life effects; censored data

MSC: 62F12; 62F10; 62N02; 62J05

Check for updates

Academic Editor: Xuejun Wang

Received: 16 July 2026
Revised: 18 August 2026
Accepted: 25 August 2026
Published: 29 August 2026

Copyright: © 2026 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.

1. Introduction

In biomedical research, epidemiological investigations, and clinical follow-up studies, the simultaneous collection of longitudinal data and survival data has become commonplace. For example, in long-term follow-up studies of chronic diseases such as kidney disease, cardiovascular disease, and cancer, researchers not only regularly measure patients' biomarkers (e.g., CD4 counts, blood pressure, renal function indicators) but also record patients' time of death or disease recurrence. How to scientifically analyze such complex data and accurately reveal the joint effects of covariates on longitudinal outcomes and survival times is an important problem in biostatistics [1,2].

Existing research methods mainly fall into three categories: joint models, inverse probability weighting, and conditional models. Joint models connect the longitudinal process and survival time through shared random effects, effectively handling non-ignorable dropout and internal covariate issues [1–3]. However, existing joint models have a notable limitation: they cannot intuitively explain the direct effect of residual life on the longitudinal response variable.

In clinical practice, patients often exhibit significant physiological changes as they approach death. For instance, Liu et al. [4] found that medical costs for patients with end-stage renal disease show a significant upward trend in the six months before death. Chan and Wang [5] revealed a pattern of medical costs first increasing and then decreasing before death among ovarian cancer patients. Sowers et al. [6] discovered that the degree of functional limitation significantly worsens before death in elderly populations. Harlow et al. [7] showed that menstrual cycles become longer and more irregular as women approach menopause. This "residual life effect" holds significant clinical importance and research value–itreflects the patterns of physiological changes in the terminal stages of disease and is crucial for end-of-life care, medical resource planning, and patient prognosis assessment. However, existing joint models fail to explicitly separate and quantify this effect, instead mixing it into random effects or error terms, thereby limiting model interpretability.

The conditional model proposed by Kong et al. [8] introduces a residual life effect term $g(T_i - t, \boldsymbol{\xi})$ directly into the longitudinal model, providing an intuitive interpretation of how impending death affects the response variable. The innovation of this approach lies in treating residual life as an explicit covariate in the model. When the observation time is far from the terminal event, the residual life effect approaches zero, and the model reduces to its standard form; when the observation time is close to the terminal event, the model flexibly captures the impact of residual life. However, this method has two limitations.

First, there is a restriction on covariate types. The method can only handle baseline covariates and cannot incorporate time-varying covariates. In real clinical studies, patients' treatment status, comorbidity conditions, and medication regimens often change over time. For example, in ESRD patient studies, whether Medicare serves as a secondary payer can change over time; in HIV research, patients' antiviral treatment regimens may be adjusted due to drug resistance; in cardiovascular disease studies, patients' antihypertensive medication doses may vary with blood pressure control. Ignoring these time-varying covariates can lead to severe estimation bias and even yield opposite conclusions.

Second, there is a limitation in censored data processing. This approach to handling censored data requires two-stage estimation: first estimating the survival distribution, then integrating over possible death times for censored individuals. In the case of time-varying covariates, this approach requires predicting the future covariate trajectories of censored individuals, which is extremely difficult both theoretically and practically. This is because the future values of time-varying covariates depend not only on current values but also on whether the individual survives–thisis the classic "internal covariate" problem in survival analysis. Kalbfleisch and Prentice [9] pointed out that when covariates are internal, the conditional survival function cannot be mathematically well-defined, because observing the covariate history itself implies the individual is alive. This fundamental difficulty makes it challenging to extend Kong et al.'s method to time-varying covariate scenarios.

Wang et al. [10] extended the conditional model to a nonparametric bivariate time-varying coefficient model that flexibly handles time-varying covariates. Their study allows regression coefficients to vary simultaneously with follow-up time t and residual life $T - t$, offering great flexibility. However, this study employs a complete-case analysis, discarding all censored data, which results in substantial efficiency loss when the censoring rate is high. Using ESRD data as an example, although the censoring rate in that study was only 3.74%,

making the efficiency loss negligible, in many practical studies the censoring rate can reach 30–50%, where complete-case analysis would lose a considerable amount of information and could even lead to selection bias–ifcensoring is related to health status, analyzing only complete cases may yield biased conclusions.

Deng [11] proposed a cumulative mean function estimation method based on joint models and inverse probability weighting, capable of handling time-varying covariates and right-censored data, providing an important tool for analyzing medical cost data. However, in this method, covariate effects are assumed to be time-invariant constants, which may not adequately capture nonlinear relationships in the data. Li et al. [12] employed polynomial regression to estimate the nonparametric component of joint models, enabling the description of nonlinear relationships between longitudinal variables and time points. Li et al. [13] further proposed a symmetric kernel smoothing estimation method for time-varying coefficient models of medical cost functions. Wang et al. [14] studied kernel quantile regression estimation for semiparametric partially linear time-varying coefficient models based on a history process of longitudinal data, developed a three-stage estimation procedure using inverse probability of censoring weighting, and established the asymptotic normality of the estimators and the oracle property of the adaptive LASSO penalized estimators.

This paper integrates three components that have not previously been available in a single likelihood-based framework. First, it extends the residual-life modelof Kong et al. [8] by allowing time-varying covariates to have time-varying coefficients. Second, unlike the complete-case analysisof Wang et al. [10], it retains censored subjects through their observed longitudinal histories and survival contributions. Third, shared random effectslink the longitudinal and survival processes while the residual-life term is kept separate from the latent state entering the hazard. The individual ingredients build on existing work; the contribution is their unified formulation, likelihood-based estimation, and accompanying sieve theory under the assumptions stated in Section 4.

The remainder of this paper is organized as follows. Section 2 presents the model specification. Section 3 describes constrained numerical maximization of the full likelihood, using B-splines and Gauss–Hermitequadrature. Section 4 establishes asymptotic properties of the sieve maximum likelihood estimators under a known baseline hazard. Section 5 presents simulation studies, Section 6 analyzes theMADIT medical cost data, and Section 7 gives conclusions and future directions. Proofs are provided in the Appendixes A and B.

2. Description of Model

In this study, we consider longitudinal history data subject to a terminal event and censoring. The modeling framework is formulated as follows:

Let T denote the terminal event time (e.g., time to death). The stochastic process $\{Y(t) : t \geq 0\}$ represents the patient's health state history, and is assumed to satisfy $Y(t) = 0$ for all $t \geq T$, i.e., the state function is zero after the occurrence of the terminal event.

For each individual, observation of the state process $Y(t)$ is not continuous; rather, it ceases at the earlier of two competing events: the terminal event T and a censoring time C (e.g., loss to follow-up). An assumption of this work is that, conditional on relevant covariates, the censoring time C is independent of both the terminal event time T and the latent state process $\{Y(t)\}$; that is, the censoring mechanism is non-informative.

Moreover, the state process $Y(t)$ and the event time T do not evolve independently. They are not only correlated with each other but are also jointly influenced by a set of covariates, including time-dependent covariates $X(t)$ and $Z(t)$, and a time-independent covariate W.

Regarding the observed data, for the ith subject, let T_i be the event time and C_i the censoring time. The observed follow-up time is then $V_i = \min(T_i, C_i)$. We define an event indicator $\Delta_i = I(T_i \leq C_i)$, where $\Delta_i = 1$ indicates that the terminal event was observed, and $\Delta_i = 0$ indicates that the observation was censored.

For clarity, n denotes the number of subjects and n_i the number of longitudinal measurements for subject i. The dimensions of the time-varying fixed-effect, random-effect, and baseline-covariate vectors are p, q, and r, respectively. The vector $\boldsymbol{\xi}$ contains the finite-dimensional parameters of the residual-life function, D is the $q \times q$ random-effects covariance matrix, σ^2 is the measurement-error variance, and K_n denotes the B-spline sieve dimension used below.

In practice, continuous recording of the state process is not feasible. Instead, for subject i observations are obtained at discrete time points t_{ij}, subject to measurement error, up to the observed follow-up time V_i. The observed longitudinal data thus consist of $\{(\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i, w_i, V_i, \Delta_i); i = 1, 2, \ldots, n\}$, where $\boldsymbol{y}_i = \{y_i(t_{ij}) : j = 1, 2, \ldots, n_i\}$ are the observed values of the state process at times t_{ij}, and $\boldsymbol{x}_i = \{x_i(t_{ij}) : j = 1, \ldots, n_i\}$ and $\boldsymbol{z}_i = \{z_i(t_{ij}) : j = 1, \ldots, n_i\}$ are the corresponding values (or measurements) of the time-dependent covariates, and w_i is the time-independent covariate.

In summary, the data structure under consideration exhibits the following complexity: each subject contributes longitudinal measurements at multiple discrete time points, with follow-up terminated by a potentially censored survival event, and the entire process is jointly influenced by both time-invariant and time-varying covariates.

2.1. Longitudinal Submodel

For subject $i = 1, \ldots, n$, at observation times $t_{ij}, j = 1, \ldots, n_i$, we observe a longitudinal response $Y_i(t_{ij})$ and a p-dimensional vector of time-varying covariates $\mathbf{X}_i(t_{ij})$. We propose the following longitudinal submodel:

$$Y_i(t) = \mathbf{X}_i(t)^T \beta(t) + g(T_i - t, \boldsymbol{\xi}) + \mathbf{Z}_i(t)^T \boldsymbol{b}_i + \epsilon_i(t), \tag{1}$$

where $\beta(t)$ is a $p \times 1$ vector of time-varying coefficient, $g(T_i - t, \boldsymbol{\xi})$ is the residual life effect, with $g(u, \boldsymbol{\xi}) \to 0$ as $u \to \infty$, $\mathbf{Z}_i(t)$ is the $q \times 1$ design vector for random effects, $\boldsymbol{b}_i$ is the $q \times 1$ random effects vector, with $\boldsymbol{b}_i \sim N(0, D)$, and $\epsilon_i(t)$ is the measurement error, where $\epsilon_i(t) \sim N(0, \sigma^2)$ independent of $\boldsymbol{b}_i$.

We should notice that when T_i is not observed (censored individual), T_i in $g(T_i - t, \boldsymbol{\xi})$ represents the potential, unobserved time of death, which will be handled through integration in the likelihood function.

2.2. Survival Submodel

We define the latent true state directly by

$$m_i(t) = \boldsymbol{X}_i(t)^T \boldsymbol{\beta}(t) + \boldsymbol{Z}_i(t)^T \boldsymbol{b}_i. \tag{2}$$

The longitudinal response can therefore be written as $Y_i(t) = m_i(t) + g(T_i - t, \boldsymbol{\xi}) + \epsilon_i(t)$. This ordering makes explicit the modeling assumption that $m_i(t)$ enters the hazard, whereas $g(\cdot)$ describes an additional terminal-proximity component of the longitudinal response and does not affect survival through another pathway. Moreover, $\boldsymbol{X}_i(t)$ and $\boldsymbol{Z}_i(t)$ contain only current or predictable covariate information available up to time t; no future measurements are used. We assume that the hazard depends on $m_i(t)$:

$$\lambda_i(t|\boldsymbol{b}_i) = \lambda_0(t) \exp\Big(\boldsymbol{\gamma}^T \boldsymbol{W}_i + \alpha m_i(t)\Big), \tag{3}$$

where $\lambda_0(t)$ is a known positive baseline hazard in the present theoretical and computational development, $\boldsymbol{W}_i$ is the $r \times 1$ vector of baseline covariates, γ is the $r \times 1$ regression coefficient vector, and α is an association parameter. The corresponding survival function is

$$S_T(t|\boldsymbol{b}_i) = \exp\Big(-\int_0^t \lambda_0(s)\exp\Big(\boldsymbol{\gamma}^T\boldsymbol{W}_i + \alpha m_i(s)\Big)ds\Big). \tag{4}$$

3. Parameter Estimation

3.1. Likelihood Function

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}(\cdot), \boldsymbol{\xi}, \boldsymbol{\gamma}, \alpha, \mathbf{D}, \sigma^2)$ denote the vector of all unknown parameters in the model, where $\boldsymbol{\beta}(\cdot)$ is the infinite-dimensional time-varying coefficient function, and $(\boldsymbol{\xi}, \boldsymbol{\gamma}, \alpha, \mathbf{D}, \sigma^2)$ are the finite-dimensional parameters. Based on models (1)–(4), the full likelihood is given by

$$\begin{aligned} L(\boldsymbol{\theta}) = \prod_{i=1}^{n}\int_{\boldsymbol{b}_i}\Bigg[\prod_{j=1}^{n_i} f_Y(Y_{ij} \mid \boldsymbol{b}_i, T_i, \boldsymbol{X}_i(t_{ij}), \boldsymbol{\beta}(t_{ij}), \boldsymbol{\xi})\Bigg] \\ \times [\lambda_i(T_i \mid \boldsymbol{b}_i)]^{\Delta_i} \times [S_T(C_i \mid \boldsymbol{b}_i)]^{1-\Delta_i} \times f_{\boldsymbol{b}}(\boldsymbol{b}_i)\, d\boldsymbol{b}_i, \end{aligned} \tag{5}$$

where

$$\begin{aligned} f_Y(Y_{ij}|\dots) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(Y_{ij} - \boldsymbol{X}_i(t_{ij})^T\boldsymbol{\beta}(t_{ij}) - g(T_i - t_{ij}, \boldsymbol{\xi}) - \boldsymbol{Z}_i(t_{ij})^T\boldsymbol{b}_i)^2}{2\sigma^2}\right) \\ f_{\boldsymbol{b}}(\boldsymbol{b}_i) &= (2\pi)^{-q/2}|\boldsymbol{D}|^{-1/2}\exp\left(-\frac{1}{2}\boldsymbol{b}_i^T\boldsymbol{D}^{-1}\boldsymbol{b}_i\right) \\ \lambda_i(T_i|\boldsymbol{b}_i) &= \lambda_0(T_i)\exp\Big(\gamma^T\boldsymbol{W}_i + \alpha m_i(T_i)\Big) \\ S_T(C_i|\boldsymbol{b}_i) &= \exp\Big(-\int_0^{C_i}\lambda_0(s)\exp\Big(\gamma^T\boldsymbol{W}_i + \alpha m_i(s)\Big)ds\Big) \\ m_i(s) &= \boldsymbol{X}_i(s)^T\boldsymbol{\beta}(s) + \boldsymbol{Z}_i(s)^T\boldsymbol{b}_i. \end{aligned}$$

To accommodate the time-varying nature of $\boldsymbol{\beta}(t)$, we use a nonparametric approach. Specifically, we approximate $\boldsymbol{\beta}(t)$ using a B-spline basis expansion of the form

$$\boldsymbol{\beta}(t) = \boldsymbol{B}(t)^T\boldsymbol{\eta}, \tag{6}$$

where $\boldsymbol{B}(t)$ is a $K \times p$ matrix of B-spline basis functions evaluated at time t, and η is a corresponding vector of spline coefficients to be estimated. This transforms the nonparametric component into a finite-dimensional optimization problem.

We next specify the function $g(T_i - t, \boldsymbol{\xi})$. Following Kong et al. [8], we consider the following parametric forms:

Gaussian kernel:

$$g(u, \boldsymbol{\xi}) = \xi_1\exp\Big(-\xi_2(u - \xi_3)^2\Big), \quad \boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3) \tag{7}$$

Exponential decay model:

$$g(u, \boldsymbol{\xi}) = \xi_1\exp(-\xi_2 u), \quad u \geq 0 \tag{8}$$

The condition $g(u, \boldsymbol{\xi}) \to 0$ as $u \to \infty$ ensures that the model reduces to a standard form when the observation time is far from the terminal event. The exponential specification is a parsimonious model for an effect that is strongest near the terminal event and decreases monotonically with residual life. The Gaussian specification allows the effect to peak before the terminal event and then decrease. These are working parametric models rather than

universally appropriate forms; their suitability should be assessed by scientific context and sensitivity analysis.

3.2. Constrained Full-Likelihood Optimization

After the B-spline representation, all unknown quantities are collected in the finite-dimensional vector $\vartheta = (\boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{\gamma}, \alpha, \text{vech}(D), \sigma^2)$. We maximize the observed-data log-likelihood corresponding to (5) directly, subject to parameter bounds and positive-definiteness of D. The nonlinear residual-life parameters $\boldsymbol{\xi}$ are optimized jointly with the other parameters; they are not treated as a linear design component and do not have aone-step weighted least-squares update.

For each subject, the integral over the bivariate random effects is evaluated using product Gauss–Hermite quadrature. Computations are performed on the log scale using a log-sum-exp calculation to reduce numerical underflow. The numerical optimizer is the bound-constrained L-BFGS-B routine in R. The implementation uses cubic B-splines with dimensions 3, 4, and 10 for the three coefficient functions and permits at most 1000 optimization iterations. The optimizer control parameter is `factr` $= 10^7$, corresponding to a relative objective-function reduction tolerance of approximately $10^7 \epsilon_{\text{machine}} = 2.22 \times 10^{-9}$ in double precision; no separate projected-gradient tolerance is specified. In the simulation study, the spline-coefficient starting values are obtained by least-squares projection of the true coefficient functions onto their respective B-spline bases. The residual-life parameters, γ, α, and the elements of D are initialized at their data-generating values ($\gamma = 1$ and $\alpha = 0.25$). In the baseline simulations, the exponential and Gaussian residual-life scenarios use 3 and 5 quadrature nodes per random-effect dimension, respectively. An optimization is classified as successfully converged only when the optimizer returns convergence code 0. Simulation summaries are calculated using successfully converged replications only, with failed or unfinished runs excluded. The optimization procedure achieved a practically acceptable convergence rate across the two scenarios, and the mean computing time was approximately 5 min per replication.

For a censored subject ($\Delta_i = 0$), all longitudinal measurements observed up to C_i and the survival probability $S_T(C_i \mid \boldsymbol{b}_i)$ contribute to the likelihood. Thus, censored subjects are retained rather than discarded in a complete-case analysis. The current theoretical and computational development treats $\lambda_0(t)$ as known; in the simulations it is fixed at $1/3000$. Joint estimation of an unknown baseline hazard is beyond the scope of the present theory.

All computations were performed using R version 4.3.1 (16 June 2023 ucrt) on a computer running Microsoft Windows 11 Home, version 24H2 (OS build 26100.8875), with an Intel Core i7-14700 processor and 16 GB DDR5 memory. The confirmed package versions were `MASS` 7.3-60, `splines` 4.3.1, and `survival` 3.5-5.

4. Asymptotic Properties

Before presenting the asymptotic results, we define the following spaces and functions:

- Θ_f: the space of all finite-dimensional parameters $(\boldsymbol{\xi}, \gamma, \alpha, \boldsymbol{D}, \sigma^2)$
- $\mathcal{B}_n$: the B-spline sieve space defined as

$$\mathcal{B}_n = \left\{ \beta(t) = \sum_{k=1}^{K_n} \eta_k B_k(t) : \|\boldsymbol{\eta}\| \leq M_\eta \right\}$$

  where $\{B_k(t)\}$ are B-spline basis functions of degree d and M_η is a positive constant.
- K_n: the number of B-spline basis functions, satisfying $K_n \to \infty$ and $K_n^3/n \to 0$ as $n \to \infty$
- $\Theta_n = \Theta_f \times \mathcal{B}_n$: the sieve parameter space.

- $M(\boldsymbol{\theta}) = \mathbb{E}[\log L_i(\boldsymbol{\theta})]$: the expected log-likelihood.

Based on n independent and identically distributed observations $\mathcal{O}_1, \ldots, \mathcal{O}_n$, define the loglikelihood function as

$$\ell_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \log L_i(\boldsymbol{\theta}),$$

where $L_i(\boldsymbol{\theta})$ is the likelihood contribution for individual i given in (5). Let

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta_n} \ell_n(\boldsymbol{\theta})$$

be the sieve maximum likelihood estimator.

We require the following conditions for establishing consistency.

4.1. Regularity Conditions

A1 (Compactness of Parameter Space). Θ_f is a compact subset of $\mathbb{R}^{d_f}$. The sieve space B_n has bounded coefficients with $\|\boldsymbol{\eta}\| \leq M_\eta$. Consequently, $\Theta_n = \Theta_f \times \mathcal{B}_n$ is compact. The number of basis functions satisfies $K_n \to \infty$ and $K_n^3/n \to 0$.

A2 (Sieve Approximation). The true function $\beta_0(t)$ is twice continuously differentiable on $[0, \tau]$. There exists a sequence $\beta_n \in \mathcal{B}_n$ such that

$$\|\beta_n - \beta_0\|_\infty \to 0,$$

where $\|\cdot\|_\infty$ denotes the supremum norm on $[0, \tau]$.

A3 (Bounded Covariates). There exists a constant $M_X < \infty$ such that for all $t \in [0, \tau]$

$$\|\boldsymbol{X}_i(t)\|_\infty \leq M_X, \quad \|\boldsymbol{Z}_i(t)\|_\infty \leq M_X, \quad \|\boldsymbol{W}_i\|_\infty \leq M_X \quad \text{a.s.}$$

Moreover, the expectation in $\mathbb{E}[X_i(t)X_i(t)^T]$ is unconditional under the joint data-generating distribution, and this matrix ispositive definite with its minimum eigenvaluebounded away from zero uniformly for $t \in [0, \tau]$.

A4 (Known Baseline Hazard). For the theoretical results established in this section, the baseline hazard function $\lambda_0(t)$ is known and satisfies

$$0 < \underline{\lambda} \leq \lambda_0(t) \leq \overline{\lambda} < \infty, \quad \forall t \in [0, \tau].$$

where $\underline{\lambda}$ and $\overline{\lambda}$ are positive constants.

A5 (Exponential Moment of Random Effects). Let $M_\alpha = \sup_{\theta \in \Theta_n} |\alpha|$ and let q be the dimension of $\mathbf{b}_i$.There exists a constant $c_0 > c_* := M_\alpha \sqrt{q} M_X$ such that

$$\mathbb{E}[\exp(c_0 \|\mathbf{b}_i\|)] < \infty.$$

For $\mathbf{b}_i \sim N(0, D)$, this condition holds for every finite $c_0 > 0$. By compactness of the covariance-parameter space in A1, there exist constants $0 < \underline{d} \leq \overline{d} < \infty$ such that all eigenvalues of D lie in $[\underline{d}, \overline{d}]$ uniformly over Θ_n.

A6 (Non-informative Censoring). The censoring time C_i is independent of (T_i, Y_i) given covariates (X_i, Z_i, W_i). Moreover, there exists $\tau > 0$ such that $P(C_i \geq \tau) > 0$.

A7 (Identifiability). The joint support of follow-up time t and residual life $T_i - t$ contains sufficient independent variation. The family $\{g(\cdot, \boldsymbol{\xi})\}$ is identifiable, satisfies $g(u, \boldsymbol{\xi}) \to 0$ as $u \to \infty$, and no nonzero difference $g(\cdot, \boldsymbol{\xi}_1) - g(\cdot, \boldsymbol{\xi}_2)$ can be represented by the follow-up-time B-spline intercept over that joint support. The fixed- and random-effect design matrices have full column rank, and the finite-dimensional information matrix is

nonsingular. Under these conditions, distinct parameter values induce distinct observed-data distributions.

Condition A7 separates the follow-up-time intercept $\beta_0(t)$ from the residual-life effect, which varies with $T_i - t$. In practical terms, subjects observed at similar follow-up times must exhibit variation in residual life, and vice versa; without this variation the two components cannot be estimated separately.

Theorem 1. *Under Regularity Conditions A1–A7, the sieve maximum likelihood estimator $\hat{\theta}_n$ converges in probability to the true value θ_0:*

$$\hat{\theta}_n \xrightarrow{P} \theta_0 \quad \textit{in the norm} \|\cdot\|_{\Theta_n},$$

where $\|\theta\|_{\Theta_n} = \|\theta_f\| + \|\beta\|_\infty$.

The proof can be found in Appendix A.

4.2. Additional Conditions for Asymptotic Normality

For establishing the asymptotic normality of the sieve maximum likelihood estimator, we require the following additional conditions.

B1 (Sieve Dimension Condition). The number of B-spline basis functions K_n satisfies

$$K_n \to \infty, \qquad \frac{K_n^4}{n} \to 0, \qquad \sqrt{n} K_n^{-4} \to 0$$

as $n \to \infty$. Thus, for example, $n^{1/8} \ll K_n \ll n^{1/4}$. The lower bound makes the second-order sieve approximation bias negligible on the root-n scale, whereas the upper bound controls the increasing-dimensional Taylor remainder.

B2 (Score Function Regularity). There exists a neighborhood $\mathcal{U}$ of $\boldsymbol{\theta}_0$ such that

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\mathcal{U}} \left\| \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^2\right] < \infty,$$

and

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\mathcal{U}} \left\| \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right\|\right] < \infty,$$

where $\ell_i(\boldsymbol{\theta}) = \log L_i(\boldsymbol{\theta})$ is the individual log-likelihood contribution. In addition, writing $d_n = d_f + pK_n$ and $\nabla^3 \ell_i(\boldsymbol{\theta})$ for the third derivative as a trilinear form, there are nonnegative random variables G_i with $E(G_i) < \infty$ such that

$$\sup_{\boldsymbol{\theta}\in\mathcal{U}_n} \sup_{\|a\|=\|b\|=\|c\|=1} |\nabla^3 \ell_i(\boldsymbol{\theta})[a,b,c]| \le K_n G_i.$$

This bound is uniform over a shrinking neighborhood $\mathcal{U}_n$ of the sieve projection $\boldsymbol{\theta}_{0n}$. It allows for the growth caused by the $O(K_n)$ nuisance coordinates while requiring the covariate processes and the first three derivatives of the model components to have integrable envelopes.

B3 (Positive Definiteness of Fisher Information). The sieve Fisher information matrix

$$\mathbf{I}_n(\boldsymbol{\theta}_0) = \mathbb{E}\left[\frac{\partial \ell_i(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{\partial \ell_i(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}^\top\right]$$

is positive definite with minimum eigenvalue bounded away from zero uniformly in n:

$$\lambda_{\min}(\mathbf{I}_n(\boldsymbol{\theta}_0)) \ge c_I > 0, \quad \forall n,$$

for some constant $c_I > 0$.

B4 (Differentiability under the Integral). For the random effects density $f_{\mathbf{b}}(\cdot)$,

$$\int \|\mathbf{b}\|^2 f_{\mathbf{b}}(\mathbf{b})\, d\mathbf{b} < \infty,$$

and the score function satisfies

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{\mathbf{b}} h(\mathbf{b}, \boldsymbol{\theta})\, d\mathbf{b} = \int_{\mathbf{b}} \frac{\partial h(\mathbf{b}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\, d\mathbf{b}$$

for $h(\mathbf{b}, \boldsymbol{\theta}) = f_Y(\cdot) f_{\mathbf{b}}(\mathbf{b}) \lambda_i(\cdot)^{\Delta_i} S_T(\cdot)^{1-\Delta_i}$, i.e., differentiation and integration are interchangeable.

B5 (Profile Orthogonality and Sieve Rate). Let $S_{f,i}$ and $S_{\eta,i}$ be the finite-dimensional and nuisance scores at $\boldsymbol{\theta}_{0n}$, and define the efficient profile score

$$S_{f,i}^{\text{eff}} = S_{f,i} - I_{f\eta,n} I_{\eta\eta,n}^{-1} S_{\eta,i}.$$

Its pathwise derivative in every nuisance direction is zero at the truth (Neyman orthogonality). The estimator satisfies

$$\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}\| = O_p\Big(\sqrt{K_n/n}\Big), \qquad \|\boldsymbol{\beta}_{0n} - \boldsymbol{\beta}_0\|_\infty = O(K_n^{-2}),$$

and the efficient information $I_{\text{eff,n}} = E(S_{f,i}^{\text{eff}} S_{f,i}^{\text{eff}\top})$ converges to a positive-definite matrix I_{eff}.

Theorem 2. *Under Regularity Conditions A1–A7 and B1–B5, the sieve maximum likelihood estimator* $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\theta}}_f^\top, \hat{\boldsymbol{\beta}}_n)^\top$ *satisfies the following asymptotic properties.*

(a) ***Asymptotic normality of finite-dimensional parameters.***

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_f - \boldsymbol{\theta}_{f0}) \xrightarrow{D} N\Big(0, \mathbf{I}(\boldsymbol{\theta}_0)^{-1}\Big),$$

where $\boldsymbol{\theta}_{f0}$ *denotes the true value of the finite-dimensional parameters* $(\boldsymbol{\xi}_0, \gamma_0, \alpha_0, \mathbf{D}_0, \sigma_0^2)$, *and*

$$\mathbf{I}(\boldsymbol{\theta}_0) = \lim_{n\to\infty} \{\mathbf{I}_{ff,n} - \mathbf{I}_{f\eta,n} \mathbf{I}_{\eta\eta,n}^{-1} \mathbf{I}_{\eta f,n}\} = \mathbf{I}_{\text{eff}}$$

is the efficient Fisher information matrix for $\boldsymbol{\theta}_f$, *with* $\mathbf{I}_n(\boldsymbol{\theta}_0)$ *partitioned as*

$$\mathbf{I}_n(\boldsymbol{\theta}_0) = \begin{pmatrix} \mathbf{I}_{ff}(\boldsymbol{\theta}_0) & \mathbf{I}_{f\eta}(\boldsymbol{\theta}_0) \\ \mathbf{I}_{\eta f}(\boldsymbol{\theta}_0) & \mathbf{I}_{\eta\eta}(\boldsymbol{\theta}_0) \end{pmatrix}.$$

(b) ***Rate forthe nonparametric coefficient function.*** *For the time-varying coefficient function,*

$$\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| = O_p\Big(\sqrt{K_n/n} + K_n^{-2}\Big).$$

Accordingly, no root-n pointwise normal limit is asserted for $\hat{\boldsymbol{\beta}}_n(t)$ *when* $K_n \to \infty$; *such a limit generally requires a* $\sqrt{n/K_n}$ *normalization and additional local-basis conditions.*

The proof can be found in Appendix B.

5. Simulation Study

In this section, we conduct simulation studies to evaluate the finite-sample performance of the proposed joint model. We consider two different specifications of the residual

life effect function to assess the performance of the proposed estimation method under various scenarios.

5.1. Data Generation and Model Specification

The joint model used in our simulation study is given by

$$\begin{aligned} Y_i(t) &= \mathbf{X}_i(t)^\top \boldsymbol{\beta}(t) + g(T_i - t, \boldsymbol{\xi}) + \mathbf{Z}_i(t)^\top \mathbf{b}_i + \varepsilon_i(t), \\ \lambda_i(t \mid \mathbf{b}_i) &= \lambda_0(t) \exp\Big(\gamma^\top \mathbf{W}_i + \alpha m_i(t)\Big), \end{aligned} \tag{9}$$

where $m_i(t) = \mathbf{X}_i(t)^\top \boldsymbol{\beta}(t) + \mathbf{Z}_i(t)^\top \mathbf{b}_i$ is the true state process.

We consider two specifications of the residual life effect function:
Scenario 1 (Exponential Decay Model):

$$g(u, \boldsymbol{\xi}) = \xi_1 \exp(-\xi_2 u), \quad u \geq 0, \tag{10}$$

where $\xi_1 > 0$ controls the magnitude of the effect and $\xi_2 > 0$ controls the decay rate. This model assumes that the residual life effect decreases monotonically as the time to death increases.
Scenario 2 (Gaussian Kernel Model):

$$g(u, \boldsymbol{\xi}) = \xi_1 \exp\{-\xi_2 (u - \xi_3)^2\}, \quad u \geq 0, \tag{11}$$

where $\xi_1 > 0$ controls the magnitude, $\xi_2 > 0$ controls the dispersion, and $\xi_3 > 0$ controls the location of the peak. This model allows the effect to reach its maximum at a specific time point before death.

Under both scenarios, the true time-varying coefficient functions are set to

$$\beta_0(t) = 1.5t^{0.8}, \quad \beta_1(t) = 1.2t^{0.5}, \quad \beta_2(t) = 1.2t^{0.2}. \tag{12}$$

Baseline covariates are generated as $X_{2i} \sim N(1, 2)$ and $W_i \sim \text{Bernoulli}(0.5)$. The time-varying covariate is generated as $X_{1i}(t) = r_i \sin(t)$, where $r_i \sim \text{Bernoulli}(0.5)$.

The random effects $\mathbf{b}_i = (b_{0i}, b_{1i})^\top$ follow a bivariate normal distribution with mean zero and covariance matrix

$$\mathbf{D} = \begin{pmatrix} 0.3808 & -0.0033 \\ -0.0033 & 0.00024 \end{pmatrix}, \tag{13}$$

with corresponding standard deviations $\sigma_{b0} = 0.617$, $\sigma_{b1} = 0.0155$, and correlation $\rho = -0.34$.

Survival times are generated via the inverse transform method with constant baseline hazard $\lambda_0(t) = 1/3000$, corresponding to the baseline survival function $S_0(t) = \exp(-t/3000)$. The treatment effect parameter is set to $\gamma = 1.0$ and the association parameter is set to $\alpha = 0.25$. The censoring rate is set to approximately 50% to evaluate the proposed method under moderate censoring. Longitudinal observations are generated at discrete time points $t = 1, 2, \ldots, T_i^*$ with measurement error $\varepsilon_i(t) \sim N(0, 1)$.

The residual life effect parameters under the two scenarios are specified as follows:

- **Scenario 1 (Exponential Decay):** $\xi_1 = 10.0$, $\xi_2 = 0.09$. Then $g(0) = 10$, and the effect reaches its maximum at the time of death and decays exponentially as u increases.
- **Scenario 2 (Gaussian Kernel):** $\xi_1 = 10.0$, $\xi_2 = 0.09$, $\xi_3 = 2.6$. The effect reaches its peak at $u = 2.6$ time units before death.

Except for the specification of $g(u)$, all other data-generatingsettings are identical between the two scenarios, including the time-varying coefficient functions, the random effects covariance matrix, the baseline hazard function, the treatment effect, the association

parameter, the sample size, and the B-spline basis dimensions. This design ensures that differences in estimation performance can primarily be attributedto the choice of the residual life effect function.

The baselinesample size is $n = 500$. The numbers of B-spline basis functions are $K_0 = 3$, $K_1 = 4$, and $K_2 = 10$. For both scenarios, only successfully converged replications were retained for the parameter and function summaries; failed or unfinished runs were excluded.

For each replication, the data are generated as follows:

1. Generate covariates and draw the bivariate random effects from $N(0, D)$.
2. Generate the terminal-event time by numerical inversion of the cumulative hazard determined by $m_i(t)$.
3. Independently generate censoring and construct $V_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$.
4. At each $t_{ij} \leq V_i$, generate $Y_i(t_{ij}) = m_i(t_{ij}) + g(T_i - t_{ij}, \boldsymbol{\xi}) + \epsilon_i(t_{ij})$.
5. Fit the model and retain the convergence code, runtime, parameter estimates, and estimated functions.

We use the following metrics to evaluate estimation performance:

1. Bias, standard deviation (SD), and root mean square error (RMSE) for finite-dimensional parameters;
2. The **relative root mean squared error (RRMSE)** for atime-varying coefficient, defined over the grid $\{t_g\}_{g=1}^{N_{\text{grid}}}$:

$$\text{RRMSE}(\hat{\beta}_k) = \left\{ \frac{\sum_{g=1}^{N_{\text{grid}}} \left(\hat{\beta}_k(t_g) - \beta_k(t_g)\right)^2}{\sum_{g=1}^{N_{\text{grid}}} \beta_k(t_g)^2} \right\}^{1/2}; \tag{14}$$

3. The **RRMSE** for the residual-lifeeffect function, defined over the u-grid:

$$\text{RRMSE}(\hat{g}) = \left\{ \frac{\sum_{\ell=1}^{N_u} (\hat{g}(u_\ell) - g(u_\ell))^2}{\sum_{\ell=1}^{N_u} g(u_\ell)^2} \right\}^{1/2}. \tag{15}$$

RRMSE standardizes the root integrated squared error by the signal magnitude, facilitating comparison across functions with different scales. It is calculated within each successfully converged replication and then summarized across replications.

5.2. Scenario 1: Exponential Decay Model

Table 1 presents the estimation results for the finite-dimensional parameters under Scenario 1. For the residual life effect parameters, the bias of $\hat{\xi}_1$ is 0.058 (RMSE = 0.082) and the bias of $\hat{\xi}_2$ is 0.0002 (RMSE = 0.003), indicating that both parameters of the exponential decay model can be accurately estimated. The treatment effect parameter γ has a bias of -0.022 (RMSE = 0.042), and the association parameter α has a bias of -0.015 (RMSE = 0.021), both within reasonable ranges.

The estimates of the random effects covariance parameters perform well. The estimate of D_{11} (intercept variance) has a mean of 0.3802 (true value 0.3808) with a bias of only -0.0006; the estimate of D_{12} (covariance) has a mean of -0.0020 (true value -0.0033) with a bias of 0.0013; and the estimate of D_{22} (slope variance) has a mean of 0.00012 (true value 0.00024) with a bias of -0.00012. All random effect parameter estimates are close to their true values, validating the model's ability to capture individual heterogeneity.

Table 1. Finite-dimensional parameter estimation results for Scenario 1 (Exponential Decay).

Parameter	True Value	Mean	Bias	SD	RMSE
ξ_1	10.000	10.058	0.058	0.057	0.082
ξ_2	0.090	0.090	0.000	0.003	0.003
γ	1.000	0.978	−0.022	0.035	0.042
α	0.250	0.235	−0.015	0.015	0.021
D_{11}	0.3808	0.3802	−0.0006	0.0008	0.0010
D_{12}	−0.0033	−0.0020	0.0013	0.0027	0.0029
D_{22}	0.00024	0.00012	−0.00012	0.00009	0.00015

Table 2 reports the RRMSE summaries for the coefficient and residual-life functionsunder Scenario 1. The mean RRMSEs are small for all four functions, ranging from 0.0137 for $g(u)$ to 0.0521 for $\beta_0(t)$, indicating accurate recovery of the coefficient and residual-life functions. These summaries are calculated from the successfully converged replication-level curve estimates.

Table 2. RRMSEsfor function estimation under Scenario 1 (Exponential Decay).

Function	MSE	Mean RRMSE	SD of RRMSE
$\beta_0(t)$	0.487	0.052097	0.000348
$\beta_1(t)$	0.028	0.037149	0.005421
$\beta_2(t)$	0.010	0.048773	0.013321
$g(u)$	0.014	0.013666	0.010465

Figure 1 displays the estimated time-varying coefficient functions (blue solid lines) along with the true functions (red dashed lines) and the corresponding 95% pointwise Monte Carlo percentile bands (light-blueshaded areas) under Scenario 1. The three coefficient panels use linear y-axes and retain their zero-valued endpoints at $t = 0$. Following the reviewer's suggestion, the residual-life panel displays $\log\{g(u)\}$ to improve the visibility of relative differences as the effect decays. In Figures 1 and 2, log denotes the base-10 logarithm.Both $\hat{\beta}_1(t)$ and $\hat{\beta}_2(t)$ closely match their true curves over most of the time domain, with narrow percentile bandsreflecting high estimation precision. The overall trend of $\hat{\beta}_0(t)$ is correctly captured, with wider percentile bandsat the boundaries, consistent with the boundary effects commonly observed in B-spline estimation.

Figure 1d shows the estimated residual life effect function $\hat{g}(u)$. The estimated curve closely matches the true function across the entire u domain, with narrow confidence intervals, particularly over the main support $u \in [0, 8]$. When $u > 10$, the effect has decayed to near zero, and the confidence intervals widen slightly but remain limited in magnitude. This result verifies the effective characterization of the residual life effect by the exponential decay model.

5.3. Scenario 2: Gaussian Kernel Model

Table 3 presents the estimation results for the finite-dimensional parameters under Scenario 2. For the Gaussian kernel parameters, $\hat{\xi}_1$ has a bias of 0.159 (RMSE = 0.161), $\hat{\xi}_2$ has a bias of −0.012 (RMSE = 0.016), and $\hat{\xi}_3$ has a bias of 0.061 (RMSE = 0.061). The treatment effect parameter γ has a bias of −0.055 (RMSE = 0.081), and the association parameter α has a bias of −0.041 (RMSE = 0.041). The random effect covariance parameters are also well estimated, with D_{11} showing a bias of only 0.0005 (RMSE = 0.0024).

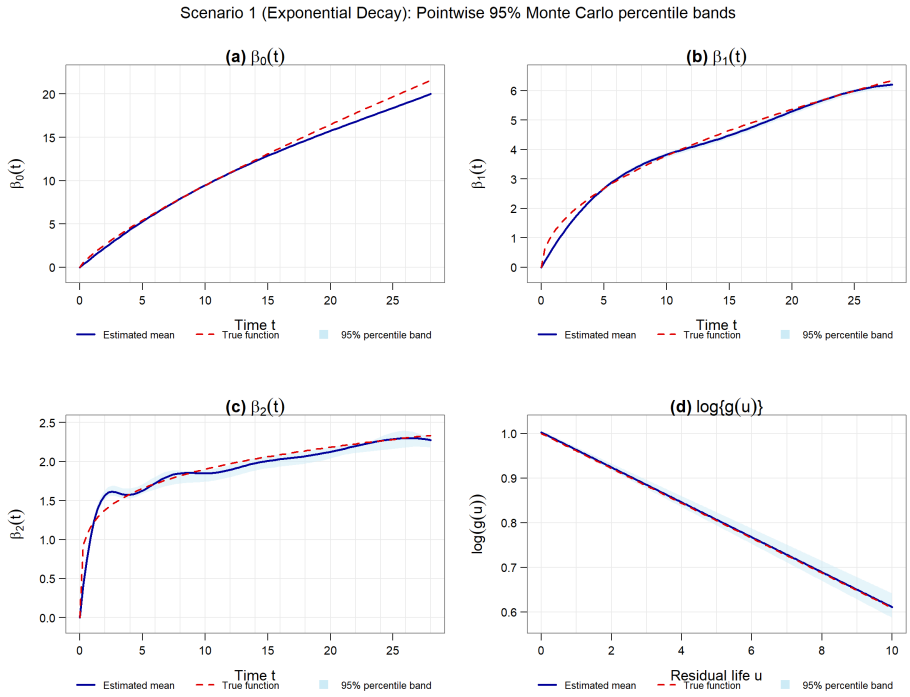


Figure 1. Estimation results under Scenario 1 (Exponential Decay): (**a**) $\hat{\beta}_0(t)$, (**b**) $\hat{\beta}_1(t)$, (**c**) $\hat{\beta}_2(t)$, and (**d**) $\log\{\hat{g}(u)\}$. The coefficient panels use linear y-axes and retain the values at $t = 0$; only the residual-life panel is displayed on the base-10 logarithmic scale. Blue solid lines represent the estimated means, light-blueshaded areas represent 95% pointwise Monte Carlo percentile bands, and red dashed lines represent the true functions.

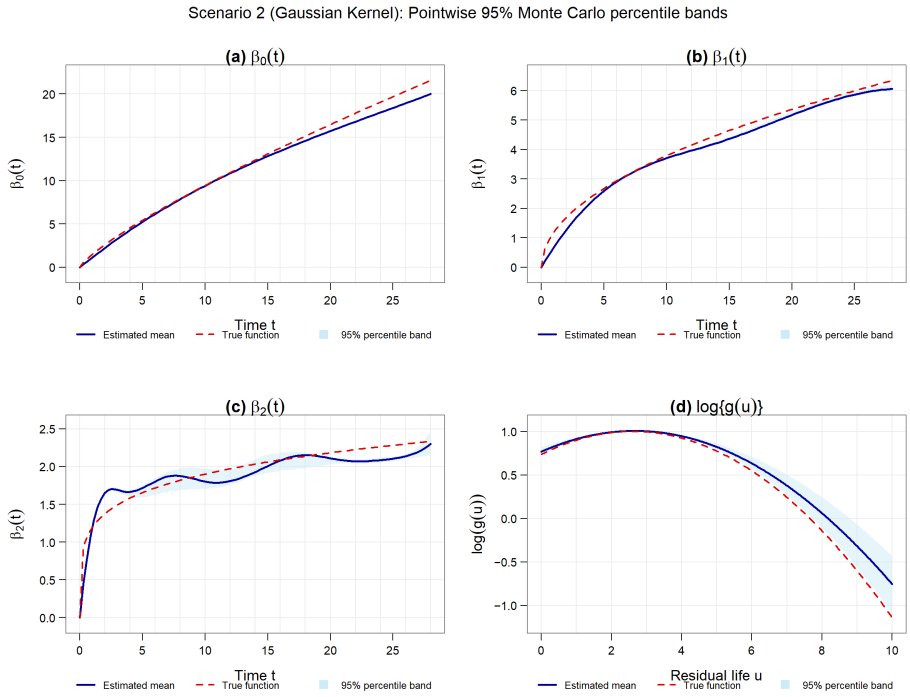


Figure 2. Estimation results under Scenario 2 (Gaussian Kernel): (**a**) $\hat{\beta}_0(t)$, (**b**) $\hat{\beta}_1(t)$, (**c**) $\hat{\beta}_2(t)$, and (**d**) $\log\{\hat{g}(u)\}$. The coefficient panels use linear y-axes and retain the values at $t = 0$; only the residual-life panel is displayed on the base-10 logarithmic scale. Blue solid lines represent the estimated means, light-blueshaded areas represent 95% pointwise Monte Carlo percentile bands, and red dashed lines represent the true functions.

Table 3. Finite-dimensional parameter estimation results for Scenario 2 (Gaussian Kernel).

Parameter	True Value	Mean	Bias	SD	RMSE
ξ_1	10.000	10.159	0.159	0.023	0.161
ξ_2	0.090	0.078	−0.012	0.010	0.016
ξ_3	2.600	2.661	0.061	0.009	0.061
γ	1.000	0.945	−0.055	0.060	0.081
α	0.250	0.209	−0.041	0.006	0.041
D_{11}	0.3807	0.3812	0.0005	0.0024	0.0024
D_{12}	−0.0032	−0.0027	0.0005	0.0013	0.0014
D_{22}	0.00024	0.00014	−0.00010	0.00009	0.00013

Table 4 reports the corresponding RRMSE summaries under Scenario 2, calculated separately for each successfully converged replication. The mean RRMSE ranges from 0.0502 for $\beta_1(t)$ to 0.0749 for $g(u)$. The larger between-replication variability for $g(u)$ reflects the additional difficulty of estimating both the magnitude and location of the Gaussian peak.

Table 4. RRMSEsfor function estimation under Scenario 2 (Gaussian Kernel).

Function	MSE	Mean RRMSE	SD of RRMSE
$\beta_0(t)$	0.497959	0.052674	0.000340
$\beta_1(t)$	0.051212	0.050237	0.004071
$\beta_2(t)$	0.020022	0.071511	0.006158
$g(u)$	0.293910	0.074946	0.043318

Figure 2 displays the estimated functions along with the true functions and 95% pointwise Monte Carlo percentile bandsunder Scenario 2. The coefficient panels again use linear y-axes and retain the values at $t = 0$, whereas the residual-life panel displays $\log\{g(u)\}$. The estimates closely track the true functions across the displayeddomain. For $\hat{g}(u)$, the estimated curve successfully captures the peak at approximately $u = 2.6$, consistent with the true value of ξ_3, and the percentile band isnarrow around the peak region, reflecting precise estimation of the residual life effect.

5.4. Comparison of the Two Scenarios

Both scenarios demonstrate that the proposed estimation procedure performs well in finite samples. The biases for most parameters are small, and the time-varying coefficient functions are consistently recovered. Under Scenario 1 (Exponential Decay), the estimation of ξ_2 is particularly precise with a bias of only 0.0002, while under Scenario 2 (Gaussian Kernel), the additional parameter ξ_3 is estimated with a bias of 0.061 and RMSE of 0.061, reflecting the inherent difficulty in jointly identifying the peak location alongside the flexible time-varying intercept $\beta_0(t)$.

The RRMSE results show that recovery is more difficult under the Gaussian-kernel specification. The mean RRMSE for $\beta_0(t)$ is similar inthe two scenarios (0.052674 versus 0.052097), whereas the Gaussian-kernel values are higher for $\beta_1(t)$ (0.050237 versus 0.037149), $\beta_2(t)$ (0.071511 versus 0.048773), and especially $g(u)$ (0.074946 versus 0.013666). The standard deviation of the RRMSE for $g(u)$ also increases from 0.010465 to 0.043318. These differences are consistent with the greater flexibility of the Gaussian form and the need to estimate its peak location in addition to its magnitude and scale.

Overall, the simulation results provide strong evidence for the practical utility of the proposed joint modeling approach. The method successfully accommodates time-varying

covariates, utilizes censored data through the joint modeling framework, and captures the residual life effect while separating it from the true state to avoid endogeneity.

5.5. Scope of the Simulation Study

The simulation study is deliberately restricted to two correctly specified residual-life scenarios with a common sample size and censoring design. A broader factorial study varying sample size, censoring proportion, and residual-life misspecification would require many additional high-dimensional likelihood optimizations. We therefore do not interpret the present results as establishing robustness under all sample sizes, censoring mechanisms, or misspecified forms. These limitations and the corresponding extensions are discussed in the Future Work discussion in Section 7. The R code for the simulation study can be found in the Supplementary Material.

6. Application: MADIT Medical Cost Data

6.1. Data and Model Specification

In this section, we apply the proposed joint model to analyze data from the Multicenter Automatic Defibrillator Implantation Trial (MADIT). This randomized clinical trial was designed to evaluate the effectiveness of implantable cardioverter-defibrillators (ICD) in preventing sudden death in high-risk patients. A total of 181 patients from 36 centers in the United States were enrolled, with 89 assigned to the ICD group and 92 assigned to the conventional therapy group.

Medical costs were recorded daily and aggregated into 12-day time units, resulting in 11,310 longitudinal observations. Six types of costs were considered: 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), and Type 6 (medications). Due to the highly skewed distribution of medical costs, we applied a log transformation, i.e., $Y(t) = \log(\text{cost}/1000 + 1)$, as the response variable. The censoring rate was 74.03%, meaning that approximately 26% of patients died during follow-up.

Based on the data characteristics, we specify the following joint model. The longitudinal submodel takes the form

$$\begin{aligned} Y_i(t) = {} & \beta_0(t) + \beta_1(t)\text{Type1}_i(t) + \beta_2(t)\text{Type2}_i(t) + \beta_3(t)\text{Type3}_i(t) \\ & + \beta_4(t)\text{Type4}_i(t) + \beta_5(t)\text{Type5}_i(t) + \beta_6(t)\text{Type6}_i(t) \\ & + g(T_i - t, \boldsymbol{\xi}) + b_{0i} + b_{1i}t + \epsilon_i(t), \end{aligned} \tag{16}$$

where $Y_i(t)$ is the log-transformed medical cost, $\beta_0(t)$ is the time-varying intercept, $\beta_1(t), \ldots, \beta_6(t)$ are time-varying coefficients for the six cost types, $g(u, \boldsymbol{\xi}) = \xi_1 \exp\{-\xi_2(u - \xi_3)^2\}$ is the residual life effect function, $(b_{0i}, b_{1i})^\top$ are random effects, and $\epsilon_i(t)$ is the measurement error. The survival submodel is given by

$$\lambda_i(t \mid \mathbf{b}_i) = \lambda_0(t) \exp(\gamma \cdot \text{Treatment}_i + \alpha \cdot m_i(t)),$$

where $m_i(t) = \beta_0(t) + \beta_1(t)\text{Type1}_i(t) + \cdots + \beta_6(t)\text{Type6}_i(t) + b_{0i} + b_{1i}t$ is the true state, Treatment $i = 1$ indicates the ICD group, and $\lambda_0(t)$ is the baseline hazard function.

In the supplied MADIT implementation, $\lambda_0(t) = h_0$ is modeled as a positive constant and estimated jointly on the log scale. The time-varying intercept uses three cubic B-spline basis functions, while each of the six time-varying cost coefficients uses five cubic B-spline basis functions. A second-difference penalty with tuning parameter 10 is applied to the intercept curve. The bivariate random-effects integral is evaluated by product Gauss–Hermite quadrature with 10 nodes per dimension. Estimation uses the bound-constrained

L-BFGS-B optimizer with at most 500 iterations and `factr` $= 10^7$. The initial values are zero for all spline coefficients, $(5, 0.1, 2)$ for $\boldsymbol{\xi}$, -1.3869 for γ, 0.1 for α, $\log(0.001)$ for $\log h_0$, and $(0.5, 0, 0.1)$ for (D_{11}, D_{12}, D_{22}). The main fit returned convergence code 0 with the message `REL_REDUCTION_OF_F <= FACTR*EPSMCH`; it required approximately 497 s, all fitted parameters were finite, and none was within 10^{-6} of its imposed bound.

Table 5 presents the parameter estimates from the proposed method. Figure 3 displays the estimated time-varying coefficients $\hat{\beta}_j(t)$ for the six cost types.

Table 5. Parameter estimates for the MADIT data.

Parameter	Interpretation	Estimate
γ	Treatment effect (log HR)	-0.837
$\exp(\gamma)$	Hazard ratio	0.433
α	Association parameter	0.338
ξ_1	g magnitude	4.386
ξ_2	g decay rate	0.338
ξ_3	g peak location	1.812
D_{11}	Intercept variance	1.064
D_{12}	Covariance	0.063
D_{22}	Slope variance	0.088

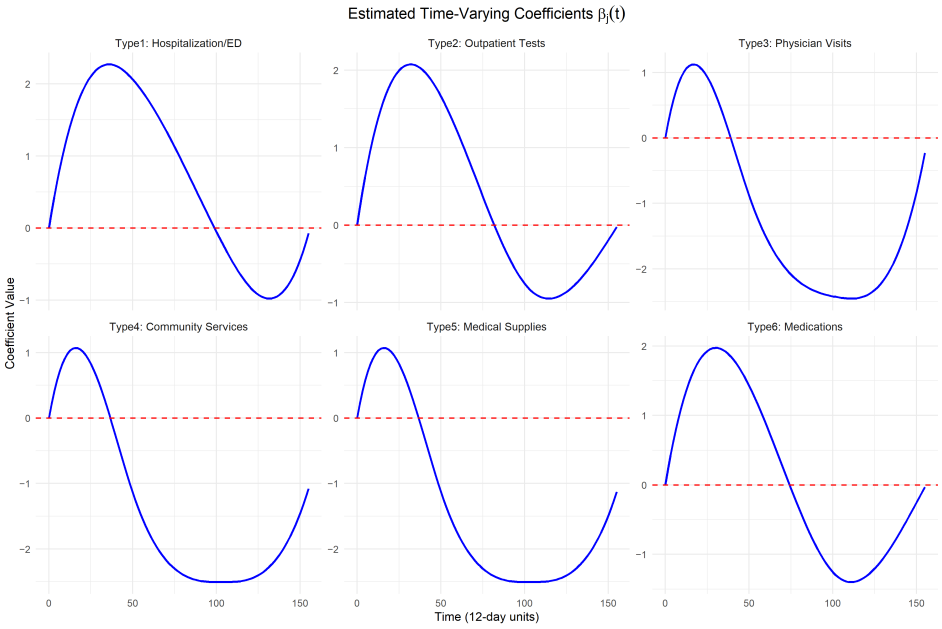


Figure 3. Estimated time-varying coefficient functions for the six cost types in the MADIT data.

6.2. Interpretation and Comparisonof Results

The fittedhazard ratio is 0.433, corresponding to a 56.7% lower fitted mortality hazard for the ICD group under this model. This is an adjusted model-based association and should not be interpreted as an unconditional causal effect. Deng [11], Li et al. [12], and Li et al. [13] reported hazard ratios of 0.250, 0.301, and 0.295 under different specifications. The more moderate estimate here may reflect the separation oftime-varying cost patterns and terminal-proximity effects, although differences across models should be interpreted cautiously.

The association estimate $\hat{\alpha} = 0.338$ corresponds to an estimated hazard ratio of $e^{0.338} \approx 1.40$ fora one-unit increase in the latent state. This is a point estimate, and no claim of statistical significance is made.

The estimated peak location is $\hat{\xi}_3 = 1.812$, which translates to approximately 22 days (since each time unit corresponds to 12 days), indicating that the fitted residual-life effect peaksabout three weeks before death. Notably, Kong et al. [8] analyzed end-stage renal disease (ESRD) patient data using a conditional model with the same Gaussian kernel and reported a peak location of approximately 23 days before death. The similarity between our pointestimate (22 days) and theirs (23 days) suggests a potentially comparableend-of-life cost pattern, although this comparison should be interpreted cautiously without uncertainty estimates. In contrast, Li et al. [12] introduced a nonparametric function $g(t) = \eta_0 + \eta_1 t + \eta_2 t^2 + \eta_3 t^3$ to capture the absolute time trend, which depends only on t and cannot distinguish between natural disease progression and the approach of death.

Figure 3 reveals that all six curves exhibit highly similar patterns: they start at 0 at time 0, rise to a common peak of approximately 2.5 at around time 10, and then gradually decline to approximately -1.8 by time 60. This "co-movement" pattern stands in sharp contrast to Figure 9 of Li et al. [13], where the six curves show distinctly different trajectories. This difference can be explained by two key features of our model. First, the inclusion of random effects (b_{0i}, b_{1i}) absorbs individual heterogeneity, allowing the fixed-effect time-varying coefficients to reflect the pattern for an "average" individual. Second, the residual life effect term $g(T_i - t, \xi)$ separates end-of-life surges, so that $\beta_j(t)$ capture the effects during routine follow-up.

The estimated random effect variances are $\hat{D}_{11} = 1.064$ (intercept) and $\hat{D}_{22} = 0.087$ (slope), both positive point estimates. The intercept variance is an order of magnitude larger than the slope variance, indicating greater fitted heterogeneityin baseline costs thanin cost trajectories. Compared to Deng [11], who reported $D_{11} = 0.381$ and $D_{22} = 0.00024$, our fitted variances are larger under the present specification. Li et al. [13] did not include random effects and thus could not assess individual heterogeneity.

This application of the proposed joint model to the MADIT medical cost data demonstrates its practical utility. The estimated hazard ratio of 0.433 is more moderate than existing estimates, reflecting the model's ability to separate indirect treatment effects. The association parameter of 0.338 is approximately 34% larger than comparable estimates, highlighting the importance of explicitly modeling end-of-life cost surges. The residual life effect peaks at approximately 22 days before death, consistent with findings from ESRD patients. The time-varying coefficients exhibit a striking "co-movement" pattern across all six cost types, contrasting with existing results and suggesting a common latent driver. These findings collectively demonstrate the value of integrating time-varying coefficients, residual life effects, and joint modeling for analyzing complex longitudinal-survival data in medical cost research.

7. Conclusions

In this paper, we have proposed a novel joint model for analyzing longitudinal data with terminal events. This model effectively addresses the challenges faced by existing methods through the integration of three important directions. First, we introduce a residual life effect term $g(T_i - t, \boldsymbol{\xi})$ that directly captures the impact of impending death on the longitudinal response variable. This specification allows the model to reduce to its standard form when the observation time is far from the terminal event, while flexibly capturing the effect of residual life when the observation time is close to the terminal event, thereby providing an intuitive clinical interpretation. Second, we allow covariate effects to vary over time by estimating time-varying coefficients $\boldsymbol{\beta}(t)$ using B-splines, overcoming the limitation of traditional fixed-coefficient models that cannot capture dynamic relationships. Third, we adopt a joint modeling framework that links the longitudinal process and the

survival process through shared random effects, naturally handling non-ignorable dropout and utilizing information from censored individuals.

The baseline simulations under exponential and Gaussian residual-life specifications show smallfinite-dimensional estimation errors and close recovery of the coefficient functions. The results support finite-sample performanceunder the two correctly specified settings considered here; they should not be read as a general robustness claim.

In the MADIT application, the fitted ICD hazard ratio is 0.433 and the estimated residual-life effect peaksapproximately 22 days before death. These are model-based associations that may help characterizeend-of-life resource use; they do not by themselves define a clinical decision rule or establish an unconditional causal effect.

Future Work

Several limitations motivatefuture research. First, the exponential and Gaussian residual-life functions are parametric working models;nonparametric or semiparametric extensions and formal specification diagnostics are desirable. The present simulations use one sample size and censoring design under correct specification, so additional small-sample, high-censoring, and misspecification analyses are needed to assess robustness. Second, a systematic comparison with modern joint-modeling software would be valuable, although no directly comparable implementation currently supports all features of the proposed model. Third, the current theory assumes a known baseline hazard, and joint sieve estimation of an unknown baseline hazard remains to be established. Fourth, multivariate longitudinal outcomes and competing risks would broaden the range of applications. Finally, constrained full-likelihood optimization with Gauss–Hermite quadrature and resampling-based uncertainty estimatescan be computationally intensive. Parallel computation, improved covariance parameterizations, adaptive quadrature, and variational approximations are possible directions for improving efficiency.

Despite these limitations, we believe that the proposed modeling framework provides a flexible and powerful tool for analyzing complex longitudinal-survival data and will play an important role in biomedical and public health research. The ability to simultaneously capture time-varying covariate effects, residual life effects, and individual heterogeneity, while properly accounting for censoring, makes this approach particularly well-suited for applications in medical cost analysis, chronic disease management, and end-of-life care planning.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/axioms15090647/s1. The complete R code used for data generation, constrained full-likelihood estimation and simulation post-processing will be uploaded as Supplementary Material upon publication of the article.

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

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: DeepSeek-V3 and ChatGPT (GPT-4o, OpenAI) were used in a limited, supportive role during manuscript preparation, specifically for language polishing, code comment generation, and formatting of tables and figures. All scientific content, methodology, results, and interpretations were developed and verified by the authors, who take full responsibility for the final manuscript.

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

Appendix A. Proof of Theorem 1 (Consistency)

Proof. We verify the five sufficient conditions for sieve M-estimation consistency [15–17]. Recall that $M(\boldsymbol{\theta}) = \mathbb{E}[\log L_i(\boldsymbol{\theta})]$ and $\ell_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n} \log L_i(\boldsymbol{\theta})$.

Step 1: Compactness (C1). By Condition A1, Θ_f is a compact subset of $\mathbb{R}^{d_f}$. The sieve space is defined as

$$\mathcal{B}_n = \Big\{\boldsymbol{\beta}(t) = \mathbf{B}(t)^\top \boldsymbol{\eta} : \|\boldsymbol{\eta}\| \le M_\eta\Big\},$$

which is closed and bounded in $\mathbb{R}^{K_n \times p}$, hence compact. Therefore, $\Theta_n = \Theta_f \times \mathcal{B}_n$ is compact.

Step 2: Continuity (C2). For fixed observed data $\mathcal{O}_i = (Y_i, T_i, \Delta_i, \mathbf{X}_i, \mathbf{Z}_i, \mathbf{W}_i)$, the individual likelihood contribution is

$$L_i(\boldsymbol{\theta}) = \int_{\mathbf{b}} \left[\prod_{j=1}^{n_i} f_Y(Y_{ij} \mid \mathbf{b}, T_i, \mathbf{X}_i(t_{ij}), \boldsymbol{\beta}(t_{ij}), \boldsymbol{\xi})\right] \cdot [\lambda_i(T_i \mid \mathbf{b})]^{\Delta_i} \cdot [S_T(C_i \mid \mathbf{b})]^{1-\Delta_i} \cdot f_{\mathbf{b}}(\mathbf{b})\, d\mathbf{b}.$$

The component functions are continuous in $\boldsymbol{\theta}$:

- $f_Y(\cdot)$ is continuous in $(\boldsymbol{\beta}, \boldsymbol{\xi})$ because the normal density is continuous in its mean parameters;
- $\lambda_i(T_i \mid \mathbf{b}) = \lambda_0(T_i)\exp\{\boldsymbol{\gamma}^\top \mathbf{W}_i + \alpha[\mathbf{X}_i(T_i)^\top \boldsymbol{\beta}(T_i) + \mathbf{Z}_i(T_i)^\top \mathbf{b}]\}$ is continuous in $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha)$, since the exponential function is continuous and $\boldsymbol{\beta}$ is continuous by the B-spline representation;
- $S_T(C_i \mid \mathbf{b}) = \exp\{-\int_0^{C_i} \lambda_0(s)\exp\{\boldsymbol{\gamma}^\top \mathbf{W}_i + \alpha[\mathbf{X}_i(s)^\top \boldsymbol{\beta}(s) + \mathbf{Z}_i(s)^\top \mathbf{b}]\}\, ds\}$ is continuous in $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha)$ by the dominated convergence theorem, as the integrand is bounded by Condition A4 (bounded baseline hazard) and Condition A3 (bounded covariates);
- $f_{\mathbf{b}}(\mathbf{b})$ is continuous in $\mathbf{D}$ on the space of positive definite matrices.

We now give an explicit dominating function for the outer random-effects integral. By A1, the measurement-error variance is bounded below by some $\underline{\sigma}^2 > 0$, the finite-dimensional parameters are uniformly bounded, and the B-spline coefficients are bounded. Together with A3, these facts imply that the product of the longitudinal normal densities is bounded above by

$$C_{Yi} = (2\pi\underline{\sigma}^2)^{-n_i/2}.$$

Let $M_\gamma = \sup_{\theta\in\Theta_n} \|\boldsymbol{\gamma}\|$ and $M_\beta = \sup_{\theta\in\Theta_n, t\in[0,\tau]} \|\boldsymbol{\beta}(t)\|$. For the event contribution, A3–A4 give

$$[\lambda_i(T_i \mid \mathbf{b})]^{\Delta_i} \le \max(1, \overline{\lambda})\exp\big\{M_\gamma\sqrt{r}M_X + M_\alpha\sqrt{p}M_X M_\beta + c_*\|\mathbf{b}\|\big\},$$

where $c_* = M_\alpha\sqrt{q}M_X$. The survival factor is at most one. Moreover, the uniform eigenvalue bounds in A5 imply

$$f_{\mathbf{b}}(\mathbf{b}; D) \le (2\pi\underline{d})^{-q/2}\exp\left(-\frac{\|\mathbf{b}\|^2}{2\overline{d}}\right).$$

Consequently, for a finite constant C_i independent of $\theta \in \Theta_n$, the complete integrand is bounded by

$$F_i(\mathbf{b}) = C_i\exp\left\{c_*\|\mathbf{b}\| - \frac{\|\mathbf{b}\|^2}{2\overline{d}}\right\}.$$

This function is integrable on $\mathbb{R}^q$: using $c_* x \le x^2/(4\overline{d}) + c_*^2\overline{d}$ for $x \ge 0$,

$$F_i(\mathbf{b}) \le C_i\exp(c_*^2\overline{d})\exp\left(-\frac{\|\mathbf{b}\|^2}{4\overline{d}}\right),$$

whose integral is finite. Equivalently, A5 ensures the required exponential moment, and the displayed Gaussian envelope verifies it explicitly for the assumed normal random effects. The same bound $\overline{\lambda}\exp\{M_\gamma\sqrt{r}M_X + M_\alpha\sqrt{p}M_X M_\beta + c_*\|\mathbf{b}\|\}$ dominates the cumulative-hazard integrand over $s \in [0,\tau]$. The Dominated Convergence Theorem therefore applies first to the cumulative-hazard integral and then to the outer random-effects integral, proving continuity of $L_i(\boldsymbol{\theta})$ and, since $L_i(\boldsymbol{\theta}) > 0$, of $\log L_i(\boldsymbol{\theta})$.

Step 3: Uniform Law of Large Numbers (C3). By Conditions A3–A5, $|\log L_i(\boldsymbol{\theta})|$ is uniformly bounded. Moreover, $\log L_i(\boldsymbol{\theta})$ is Lipschitz in $\boldsymbol{\theta}$ with an integrable Lipschitz constant:

$$|\log L_i(\boldsymbol{\theta}_1) - \log L_i(\boldsymbol{\theta}_2)| \le \Psi_i \cdot \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|,$$

where $\mathbb{E}[\Psi_i] < \infty$ (guaranteed by Condition A3 for bounded covariates and Condition A5 for the exponential moment of random effects). Therefore, the function class $\{\log L_i(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_n\}$ is Glivenko-Cantelli [17] (Theorem 19.4). Consequently,

$$\sup_{\boldsymbol{\theta}\in\Theta_n} |\ell_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta})| \xrightarrow{p} 0.$$

Step 4: Identifiability (C4). By Condition A7 and the Kullback–Leibler information inequality,

$$M(\boldsymbol{\theta}_0) - M(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}_0}\left[\log \frac{L_i(\boldsymbol{\theta}_0)}{L_i(\boldsymbol{\theta})}\right] \ge 0,$$

with equality if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Thus $\boldsymbol{\theta}_0$ is the unique maximizer of $M(\boldsymbol{\theta})$. By compactness and continuity, for any $\delta > 0$,

$$\sup_{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\ge\delta} M(\boldsymbol{\theta}) \le M(\boldsymbol{\theta}_0) - \rho(\delta),$$

for some $\rho(\delta) > 0$.

Step 5: Sieve Approximation (C5). By Condition A2, the true function $\boldsymbol{\beta}_0(t)$ is twice continuously differentiable on $[0,\tau]$. By the approximation property of B-splines [18], there exists $\boldsymbol{\beta}_n \in \mathcal{B}_n$ such that

$$\|\boldsymbol{\beta}_n - \boldsymbol{\beta}_0\|_\infty = O(K_n^{-2}) \to 0.$$

Let $\boldsymbol{\theta}_{f0}$ denote the true finite-dimensional parameters and define $\boldsymbol{\theta}_n = (\boldsymbol{\theta}_{f0}, \boldsymbol{\beta}_n) \in \Theta_n$. By continuity of M, we have $M(\boldsymbol{\theta}_n) \to M(\boldsymbol{\theta}_0)$.

Step 6: Conclusion. Let $\delta > 0$ be arbitrary. Define the event $\mathcal{E}_n = \{\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_{\Theta_n} \ge \delta\}$. On $\mathcal{E}_n$, by Step 4,

$$M(\hat{\boldsymbol{\theta}}_n) \le M(\boldsymbol{\theta}_0) - \rho(\delta).$$

By the definition of $\hat{\boldsymbol{\theta}}_n$ (maximizing ℓ_n) and Step 5,

$$\ell_n(\hat{\boldsymbol{\theta}}_n) \ge \ell_n(\boldsymbol{\theta}_n).$$

Thus,

$$M(\hat{\boldsymbol{\theta}}_n) - M(\boldsymbol{\theta}_0) = [M(\hat{\boldsymbol{\theta}}_n) - \ell_n(\hat{\boldsymbol{\theta}}_n)] + [\ell_n(\hat{\boldsymbol{\theta}}_n) - \ell_n(\boldsymbol{\theta}_n)] + [\ell_n(\boldsymbol{\theta}_n) - M(\boldsymbol{\theta}_0)].$$

The first term is bounded below by $-\sup_{\boldsymbol{\theta}} |\ell_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta})| \xrightarrow{p} 0$ by Step 3; the second term is ≥ 0 by maximality of $\hat{\boldsymbol{\theta}}_n$; the third term $\xrightarrow{p} 0$ by Step 5. Hence on $\mathcal{E}_n$,

$$-\rho(\delta) \ge M(\hat{\boldsymbol{\theta}}_n) - M(\boldsymbol{\theta}_0) \ge -o_P(1),$$

which implies $P(\mathcal{E}_n) \to 0$, since $\rho(\delta) > 0$ is a fixed constant. Therefore,

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0.$$

This completes the proof of Theorem 1. □

Appendix B. Proof of Theorem 2 (Asymptotic Normality)

Proof. The sieve and penalization argument follows the general framework of Shen [19].

Let $\ell_n(\theta) = n^{-1}\sum_{i=1}^n \ell_i(\theta)$ and let $\theta_{0n} = (\theta_{f0}^\top, \eta_{0n}^\top)^\top$ be the Kullback–Leibler projection of the true parameter onto the sieve. Its dimension is $d_n = d_f + pK_n = O(K_n)$, and Condition A2 gives $\|\beta_{0n} - \beta_0\|_\infty = O(K_n^{-2})$.

Step 1: Uniform Taylor remainder. Put $S_n(\theta) = \nabla\ell_n(\theta)$ and $H_n(\theta) = \nabla^2\ell_n(\theta)$. Taylor's theorem in integral form gives

$$0 = S_n(\theta_{0n}) + H_n(\theta_{0n})(\hat{\theta}_n - \theta_{0n}) + R_n,$$

where, for every unit vector a,

$$|a^\top R_n| \le \frac{1}{2}\sup_{\theta\in\mathcal{U}_n}\|\nabla^3\ell_n(\theta)\|_{\mathrm{op}}\|\hat{\theta}_n - \theta_{0n}\|^2.$$

By Condition B2 and the law of large numbers,

$$\sup_{\theta\in\mathcal{U}_n}\|\nabla^3\ell_n(\theta)\|_{\mathrm{op}} \le K_n\, n^{-1}\sum_{i=1}^n G_i = O_p(K_n).$$

Condition B5 therefore yields the explicit order

$$\|R_n\| = O_p\left(K_n\frac{K_n}{n}\right) = O_p(K_n^2/n).$$

Since $K_n^4/n \to 0$,

$$\sqrt{n}\,\|R_n\| = O_p(K_n^2/\sqrt{n}) = o_p(1), \qquad \text{or equivalently} \qquad R_n = o_p(n^{-1/2}).$$

Thus the Taylor remainder remains negligible even though the sieve dimension increases.

Step 2: Profile score expansion. Partition the score and information according to (θ_f, η) and use the efficient score

$$S_{f,n}^{\mathrm{eff}} = S_{f,n} - I_{f\eta,n}I_{\eta\eta,n}^{-1}S_{\eta,n}.$$

Eliminating the nuisance increment from the two score equations gives

$$0 = S_{f,n}^{\mathrm{eff}}(\theta_{0n}) - I_{\mathrm{eff,n}}(\hat{\theta}_f - \theta_{f0}) + r_{1n} + r_{2n},$$

where r_{1n} is the projected Taylor remainder and hence $r_{1n} = O_p(K_n^2/n) = o_p(n^{-1/2})$ by Step 1. Condition B5 makes the profile score orthogonal to first-order nuisance perturbations. Consequently, the effect of replacing β_0 by its sieve projection is quadratic:

$$\|r_{2n}\| = O(\|\beta_{0n} - \beta_0\|_\infty^2) = O(K_n^{-4}).$$

The third requirement in B1 gives $\sqrt{n}\|r_{2n}\| = O(\sqrt{n}K_n^{-4}) = o(1)$. Notice that the former, incorrect requirement $\sqrt{n}K_n^{-2} \to 0$ is neither used nor implied by $K_n^4/n \to 0$.

https://doi.org/10.3390/axioms15090647

Step 3: Central limit theorem and conclusion. For the fixed-dimensional efficient score, Conditions B2–B5 imply the Lindeberg condition and the information identity. Hence

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_{f,i}^{\text{eff}}(\theta_{0n}) \xrightarrow{D} N(0, I_{\text{eff}}), \qquad I_{\text{eff,n}} \to I_{\text{eff}}.$$

Multiplying the profile expansion by $\sqrt{n}$, applying the preceding two remainder bounds, and using Slutsky's theorem yields

$$\sqrt{n}(\hat{\theta}_f - \theta_{f0}) = I_{\text{eff}}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_{f,i}^{\text{eff}}(\theta_{0n}) + o_p(1) \xrightarrow{D} N(0, I_{\text{eff}}^{-1}).$$

Finally, the stochastic coefficient error is $O_p(\sqrt{K_n/n})$ and the spline approximation error is $O(K_n^{-2})$; therefore

$$\|\hat{\beta}_n - \beta_0\| = O_p\Big(\sqrt{K_n/n} + K_n^{-2}\Big).$$

This proves the two assertions of Theorem 2. □

References

1. Tsiatis, A.A.; Davidian, M. Joint modeling of longitudinal and time-to-event data: An overview. *Stat. Sin.* **2004**, *14*, 809–834.
2. Rizopoulos, D. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*; CRC Press: Boca Raton, FL, USA, 2012.
3. Hsieh, F.; Tseng, Y.-K.; Wang, J.-L. Joint modeling of survival and longitudinal data: Likelihood approach revisited. *Biometrics* **2006**, *62*, 1037–1043. [CrossRef]
4. Liu, L.; Wolfe, R.A.; Kalbfleisch, J.D. A shared random effects model for censored medical costs and mortality. *Stat. Med.* **2007**, *26*, 139–155. [CrossRef]
5. Chan, K.C.G.; Wang, M.-C. Backward estimation of stochastic processes with failure events as time origins. *Ann. Appl. Stat.* **2010**, *4*, 1602–1620. [CrossRef]
6. Sowers, M.; Tomey, K.; Jannausch, M.; Eyvazzadeh, A.; Nan, B.; Randolph, J. Physical functioning and menopause states. *Obstet. Gynecol.* **2007**, *110*, 1290–1296. [CrossRef]
7. Harlow, S.D.; Mitchell, E.S.; Crawford, S.; Nan, B.; Little, R.; Taffe, J. The restage collaboration: Defining optimal bleeding criteria for onset of early menopausal transition. *Fertil. Steril.* **2008**, *89*, 129–140. [CrossRef]
8. Kong, S.; Nan, B.; Kalbfleisch, J.D.; Saran, R.; Hirth, R. Conditional modeling of longitudinal data with terminal event. *J. Am. Stat. Assoc.* **2018**, *113*, 357–368. [CrossRef]
9. Kalbfleisch, J.D.; Prentice, R.L. *The Statistical Analysis of Failure Time Data*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2002.
10. Wang, Y.; Nan, B.; Kalbfleisch, J.D. Kernel estimation of bivariate time-varying coefficient model for longitudinal data with terminal event. *J. Am. Stat. Assoc.* **2024**, *119*, 1102–1111. [CrossRef]
11. Deng, D. Estimating the cumulative mean function for history process with time-dependent covariates and censoring mechanism. *Stat. Med.* **2016**, *35*, 4624–4636. [CrossRef]
12. 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]
13. Li, S.; Deng, D.; Han, Y. A symmetric kernel smoothing estimation of the time-varying coefficient for medical costs. *Symmetry* **2024**, *16*, 389. [CrossRef]
14. Wang, W.; Liu, X.; Deng, D. Kernel quantile regression for semiparametric partially linear time-varying-coefficient model based on a history process of longitudinal data. *J. Multivar. Anal.* **2026**, *214*, 105629. [CrossRef]
15. Chen, X. Large sample sieve estimation of semi-nonparametric models. In *Handbook of Econometrics*; Heckman, J.J., Leamer, E.E., Eds.; Elsevier: Amsterdam, The Netherlands, 2007; Volume 6B, Chapter 76, pp. 5549–5632.
16. Shen, X.; Wong, W.H. Convergence rate of sieve estimates. *Ann. Stat.* **1994**, *22*, 580–615. [CrossRef]
17. van der Vaart, A.W. *Asymptotic Statistics*; Cambridge University Press: Cambridge, UK, 1998.
18. de Boor, C. *A Practical Guide to Splines*, revised ed.; Springer: New York, NY, USA, 2001.
19. Shen, X. On methods of sieves and penalization. *Ann. Stat.* **1997**, *25*, 2555–2591. [CrossRef]

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.