

Functional varying-coefficient Cox model and its application

Statistical Methods in Medical Research

1–12

© The Author(s) 2025



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09622802251406527

journals.sagepub.com/home/smm

Fansheng Kong¹, Maozai Tian^{1,2} , Zhihao Wang² and Man-lai Tang³ 

Abstract

When data become increasingly complex, desirable models are required to be more flexible for analyzing survival data. Building upon the existing functional Cox model, we introduce a novel functional varying-coefficient Cox model and the corresponding estimation algorithms are proposed in this article. The proposed model can simultaneously handle survival data with varying-coefficient covariates and functional covariates, thereby significantly enhancing the adaptability of survival models. The model performance is evaluated by simulation studies, and a real application using Alzheimer's disease neuroimaging initiative (ADNI) data is used to illustrate the practicality of the proposed model.

Keywords

Functional data, varying-coefficients, Cox model

1 Introduction

With the development of the economy, an increasing amount of time-to-event data is being observed and recorded, and the analysis of such data is widely applied in various fields, including medicine research, financial studies, and social science investigation. In biomedical research, time-to-event data are also known as survival data, and the analysis of time-to-event data is referred to as survival analysis. Survival analysis investigates the relationship between certain characteristics of study subjects (e.g. patients) and the time until an event occurs, such as death of the study subject, failure time of certain experiments or recovery. A common model for analyzing such data in statistical research is the hazard model, which associates the risk of an event (usually the onset of a disease or death) with one or more covariates related to that event. The Cox proportional hazards model is a very common type of hazard model and a popular model for analyzing time-to-event data with censored data and covariates.¹

An important assumption of the Cox proportional hazards model is that the hazard remains constant over time. For example, taking a specific drug may halve a person's possibility of getting a disease, or changing the specific material of certain manufacturing components may triple the possibility of their failure. However, in real-life application, this assumption of constant hazard may not be realistic. In clinical medicine, such non-constant hazard phenomena are quite common. For example, when studying the occurrence of mild cognitive impairment (MCI) related to time to death, the presence of MCI among participants may change over time. One explanation is the existence of covariates that change over time, which are referred to as time-varying covariates. Leemis et al.² proposed an algorithm for generating lifetimes when survival models include time-varying covariates, primarily based on inverting the cumulative hazard function to generate survival times. Austin³ studied the data generation process for the Cox proportional hazards model with time-varying covariates under exponential, Weibull, or Gompertz distributions. Yan and Huang⁴ proposed an adaptive lasso method for

¹Center for Applied Statistics, School of Statistics, Renmin University of China, PR China

²School of Statistics and Data Science, Xinjiang University of Finance and Economics, PR China

³Department of Physics, Astronomy and Mathematics, University of Hertfordshire, UK

*The first two authors contributed equally to this paper.

Corresponding author:

Man-lai Tang, Department of Physics, Astronomy and Mathematics, University of Hertfordshire, London, UK.

Email: m.l.tang@herts.ac.uk

variable selection in Cox models with time-varying covariates. Ngwa et al.⁵ used the Lambert W function to generate survival time data for the Cox proportional hazards model with time-varying covariates that follow exponential or Weibull distributions, and demonstrated that reliable and robust estimation can be performed through simulation studies. Cygu et al.⁶ studied the penalized Cox proportional hazards model in the presence of a large number of time-varying covariates and proposed a variant of the gradient descent algorithm to fit the penalized Cox model.

Another extension of the Cox proportional model is the expansion of coefficients, transitioning from originally fixed coefficients to time-varying coefficients. In this case, it is generally assumed that the varying coefficients depend on certain covariates, such as scores for disease severity. Typically, the variation of coefficients is related to the variation of time t , known as time-varying coefficients. In such cases, it is referred to as the Cox regression model with time-dependent or time-varying coefficients. Zucker and Karr⁷ and Hastie and Tibshirani⁸ conducted initial explorations of varying coefficient models. Cai et al.⁹ developed a local partial likelihood method to estimate time-dependent coefficients in the Cox regression model and obtained the asymptotic properties of the estimators, where the time-varying functions were expanded using first-order Taylor functions. Kim et al.¹⁰ studied the issue of missing data in varying coefficient proportional hazards models and developed a reverse probability-weighted estimator based on the local partial likelihood method. Song et al.¹¹ considered the Bayesian analysis of proportional hazards models with time-varying coefficients under two prior scenarios. Kim et al.¹² and Yang et al.¹³ considered the estimation problem of time-varying coefficient models with left-truncated data and latent variables, respectively.

Functional data analysis (FDA) has become increasingly popular in many fields, such as medical research, magnetic resonance imaging (MRI), meteorology and public health. The most common method for modeling functional data is the functional linear model (FLM). There are many articles about the extension of functional linear models. Shin¹⁴ introduced the partially functional linear regression model (PFLRM). Peng et al.¹⁵ extended the partially functional linear model and proposed the varying coefficient partially functional linear regression model (VCPFLM) based on Shin¹⁴. Wang et al.¹⁶ proposed the functional partially varying coefficient zero-inflated model (FPVCZIM). When modeling Cox hazard models, the covariates related to the interesting event include various types, commonly binary and numerical types, which are relatively simple to model. As data complexity increases, functional data is also incorporated into survival analysis models. Modeling functional data is more complex than modeling general numerical data. Some scholars have combined functional data with Cox models for modeling. Lee et al.¹⁷ proposed a Bayesian functional linear Cox regression model (BFLCRM) with functional and scalar covariates. Kong et al.¹⁸ proposed a functional linear Cox regression model to describe the association between event occurrence time data and a set of functional and scalar predictor variables, and developed an algorithm to calculate the maximum approximate partial likelihood estimation of unknown finite and infinite-dimensional parameters. In the final real data analysis, they showed that high-dimensional hippocampal surface data may be an important marker for predicting the time to conversion to Alzheimer's disease.

According to our knowledge, there are few articles that combined Cox hazard models with functional data, and those articles only considered modeling with functional covariates and fixed coefficients. However, in real life applications, varying coefficients are more common than fixed coefficients. In this article, we therefore propose the so-called functional varying-coefficient Cox (FVC-Cox) model, which extends the analysis of Cox hazard models to a framework that simultaneously incorporates both varying and fixed coefficients, as well as functional data. This extension can be better adapted to practical applications, and is more interpretive and flexible. In this model, functional principal components (FPCs) are used to approximate the functional covariates, B-spline methods are used to approximate the varying coefficient covariates. The performance of the proposed method is evaluated through simulation studies. A real application using Alzheimer's disease neuroimaging initiative (ADNI) data is utilized to illustrate the practicality of the proposed model.

2 Functional varying-coefficient Cox (FVC-Cox) model and estimation

2.1 Functional varying-coefficient Cox (FVC-Cox) model

Assume that $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})^T$ is a p -dimensional covariate vector for individual i . The general Cox proportional hazards model proposed at time t is defined as follows:

$$h_i(t | X_i) = h_0(t) \exp(X_i^T \beta) = h_0(t) \exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}),$$

where $h_0(t)$ is the baseline hazard function, and the vector X_i remains fixed over time within the same individual. The parameter $\beta = (\beta_1, \dots, \beta_p)$ represents the coefficients, which also remain constant.

For functional data, we assume that the response variable Y is a real-valued random variable which defined on the probability space (Ω, \mathbb{B}, P) , $\{X(t) : t \in T\}$ is a second-order stochastic process defined on the same probability space

with $\{E|X(t)|^2 < \infty, \forall t \in T\}$, and $X \in L^2(T)$ with $L^2(T)$ being the Hilbert space of square-integrable functions defined on a non-degenerate compact interval T . Without loss of generality, we let T be the unit interval $[0, 1]$. Therefore, the relationship between $X(t)$ and Y can be expressed as:

$$Y = \int_T X(t)\beta(t)dt + \varepsilon. \quad (1)$$

Model (1) is commonly referred as the functional linear model (FLM), where $X(t)$ is the functional predictor variable and $\beta(t)$ is the unknown functional coefficient.

2.2 Functional varying-coefficient Cox model (FVC-Cox)

Kong et al. (2018) proposed the functional linear Cox regression model (FLCRM) which introduced functional data into the Cox regression model to describe the relationship between survival outcome and a set of finite-dimensional and infinite-dimensional predictors. The corresponding hazard model is defined as follows:

$$h(t | X, Z) = h_0(t) \exp \left(Z\gamma + \int_T X(t)\beta(t)dt \right).$$

In this model, the hazard function is related to the functional covariate $X(t)$ and the general covariate Z with the covariate coefficient γ being fixed. This model incorporates functional covariates into the general Cox model, increasing the flexibility of the model. In practice, the covariate coefficient γ is sometimes not completely fixed. Therefore, we consider FVC-Cox that could simultaneously considers functional covariates and varying coefficients. Specifically, for the i -th individual, the hazard function for the FVC-Cox has the following form

$$h_i(t | X_i, Z_i, V_i) = h_0(t) \exp \left(V_i^T \theta + Z_i^T \alpha(U) + \int_T X_i(t)\beta(t)dt \right). \quad (2)$$

where covariates V_i , Z_i and $X_i(t)$ for individual i can be observed, and parameter θ is a fixed coefficient. Without loss of generality, we assume that U is a one-dimensional variable, $\alpha(U) = (\alpha_1(U), \alpha_2(U), \dots, \alpha_p(U))^T$ is a p -dimensional vector of unknown varying coefficients, and $\beta(t)$ is the functional coefficient. Specifically, when the parameter $\alpha(U) = 0$, Model (2) reduces to the FLCRM (i.e. Model (1)).

3 Estimation process

As $\beta(t)$ is an infinite dimensional parameter, we need to transform the problem of infinite parameters into that of finite parameters. For this purpose, we define the covariance function of the functional predictor variable $X(t)$ as $K(s, t) = Cov(X(s), X(t))$. Assume that $K(s, t)$ is continuous on $[0, 1]^2$.

According to Mercer's Theorem, we have $K(s, t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s)\phi_j(t)$, where λ_j are the eigenvalues ($\lambda_1 > \lambda_2 > \dots > 0$) and $\phi_j(s), j = 1, 2, \dots$ are the continuous orthonormal eigenfunctions corresponding to the covariance operator of $K(s, t)$. Define the inner product on space T as $\langle f(t), g(t) \rangle = \int_T f(t)g(t)dt$. Hence, the covariance operator can be defined as: $\langle K, f \rangle (u) = \int K(u, v)f(v)dv$, where $f \in L^2(T)$. The empirical covariance function of X is defined as $\hat{K}(s, t) = \frac{1}{n} \sum_{i=1}^n X_i(s)X_i(t)$. If $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq 0$ is the ordered sequence of eigenvalues of \hat{K} , the spectral decomposition of \hat{K} can be written as:

$$\hat{K}(s, t) = \sum_{j=1}^{\infty} \hat{\lambda}_j \hat{\phi}_j(s)\hat{\phi}_j(t).$$

It is noteworthy that the sequence $\phi_j(s)$'s also form an orthonormal basis on $[0, 1]^2$. According to the Karhunen–Loève Theorem, we have: $X(t) = \sum_{i=1}^{\infty} \xi_i \phi_i(t)$, where ξ_i 's are uncorrelated random variables with zero mean and variance λ_i , and $\xi_i = \langle X, \phi_i \rangle$.¹⁹

Similarly, we can obtain the decomposition:

$$X_i(t) = \sum_{j=1}^{\infty} \hat{\xi}_{ij} \hat{\phi}_j(t), \beta(t) = \sum_{j=1}^{\infty} \eta_j \hat{\phi}_j(t),$$

where $\hat{\xi}_{ij} = \langle X_i, \hat{\phi}_j \rangle$ and $\eta_j = \langle \beta, \hat{\phi}_j \rangle$. Substituting these decompositions into the functional part, we have:

$$\begin{aligned} \int_0^1 X_i(t)\beta(t)dt &= \int_0^1 \sum_{j=1}^{\infty} \hat{\xi}_{ij} \hat{\phi}_j(t) \sum_{k=1}^{\infty} \eta_k \hat{\phi}_k(t) dt \\ &= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \hat{\xi}_{ij} \eta_k \int_0^1 \hat{\phi}_j(t) \hat{\phi}_k(t) dt \\ &= \sum_{j=1}^{\infty} \hat{\xi}_{ij} \eta_j. \end{aligned} \quad (3)$$

Substituting Equation (3) into Equation (2) yields:

$$h_i(t | X_i, Z_i, V_i) = h_0(t) \exp \left(V_i^T \theta + Z_i^T \alpha(U) + \sum_{j=1}^{\infty} \hat{\xi}_{ij} \eta_j \right). \quad (4)$$

The term $\sum_{j=1}^{\infty} \hat{\xi}_{ij} \eta_j$ in Equation (4) is infinite-dimensional. Here, we approximate it by using the first m terms. That is,

$$h_i(t | X_i, Z_i, V_i) = h_0(t) \exp \left(V_i^T \theta + Z_i^T \alpha(U) + \sum_{j=1}^m \hat{\xi}_{ij} \eta_j \right). \quad (5)$$

The effectiveness of this approximation depends on whether the first m terms can closely approximate the original term. The choice of m has relatively significant impact on the estimation accuracy. In general, a smaller m leads to larger bias while a larger m may result in larger variance. The method for selecting m will be discussed later.

For the varying coefficient part, local polynomials or B-splines are commonly used for approximation. Since B-spline approximation is global and computationally efficient and fewer parameters need to be estimated, we adopt B-spline approximation for the varying coefficient part in this paper. Let k_n be the number of uniform internal knots on $[0, 1]$, $h + 1$ be the order, and the corresponding B-spline basis be $\pi(u) = (B_1(u), B_2(u), \dots, B_{k_n+h+1}(u))^T$. Thus, the varying coefficient $\alpha(U)$ can be approximated by

$$\alpha_i(U) \approx \sum_{j=1}^{k_n+h+1} B_j(u) \gamma_{i,j} = \pi^T(u) \gamma_i, \quad i = 1, 2, \dots, p, \quad (6)$$

where $\gamma_i = (\gamma_{i,1}, \dots, \gamma_{i,k_n+h+1})^T$ is the coefficient vector of the spline basis. In this paper, we choose $h = 3$, that is, cubic splines. Substituting Equation (6) into Equation (5), we have

$$h_i(t | X_i, Z_i, V_i) = h_0(t) \exp \left(V_i^T \theta + \sum_{k=1}^p \sum_{j=1}^{k_n+h+1} Z_{ik} B_j(u) \gamma_{kj} + \sum_{g=1}^m \hat{\xi}_{ig} \eta_g \right). \quad (7)$$

We outline the estimation procedure for the parameters in the model as follows:

Step 1: Using B-splines to smooth the function $\alpha(U)$ for individual i . This process yields a smooth estimate of $\alpha(U)$. The values of $\alpha(U)$ within interval T can be calculated and expressed as a combination of spline basis functions.

Step 2: Using FPC Analysis to decompose the functional sequence $\hat{X}_i(t)$ into the product of principal component scores with principal components. To achieve a better fitting result, a reasonable method to select m is necessary. This allows $\hat{X}_i(t)$ to be approximated as a linear combination.

For selecting the value of m , there are several methods. One method is to set a threshold for the first m FPC variances, denoted as $PV(m) = \sum_{i=1}^m \lambda_i / \sum_{i=1}^{\infty} \lambda_i$. For example, one might set $PV \geq 95\%$ or $PV \geq 90\%$, we can then get the corresponding value m . Another method is similar to variable selection. Since changing the value of m can significantly affect the accuracy of the estimation. Generally, a smaller m results in larger bias while a larger m may lead to larger variance. Therefore, the AIC criterion can be used to select the value of m . Here, AIC is defined as: $AIC(m) = 2m - 2\ln\{L\}$, where L is the likelihood function.

Step 3: Substituting the approximated results from Steps 1 and 2 into the proposed FVC-Cox Model. The parameter $\hat{\delta} = (\hat{\theta}^T, \hat{\alpha}^T, \hat{\eta}^T)^T$ can be estimated by maximizing the following log partial likelihood equation:

$$Q(\delta) = \sum_{i=1}^n \int_T \hat{w}_i^T \delta dN_i(t) - \int_T \log \left\{ \sum_{i=1}^n Y_i(t) \exp(\hat{w}_i^T \delta) \right\} d\bar{N}(t),$$

where $N_i(t) = I(\bar{T}_i \leq t, \delta_i = 1)$, $\bar{N}(t) = \sum_{i=1}^n N_i(t)$, and $R(t) = \{j : \bar{T}_j \geq t\}$ represents the set of subjects at risk and uncensored before time t . Define $Y_i(t) = I(\bar{T}_i \geq t) = I(i \in R(t))$ and $\hat{w}_i = (\hat{\theta}_{i1}, \dots, \hat{\theta}_{ip_1}, \alpha_{i1}, \dots, \alpha_{ip_2}, \eta_{i1}, \dots, \eta_{ip_m})^T$ for $i = 1, \dots, n$. In this paper, it is assumed that the observed failure times are distinct.

4 Large sample properties

In this section, we will study some properties of the estimators. Given (θ, α, β) , let $F_\varepsilon(\cdot)$ and $f_\varepsilon(\cdot)$ be the conditional distribution function and conditional density function of ε , respectively. When $U = 1$, the varying coefficients $\alpha(U)$ can be converted to the parameter θ . Thus, the properties of the parameter θ can be incorporated into the parameter $\alpha(U)$. Let K represent a positive constant that may take different values in different contexts. $\alpha_0(\cdot)$ and $\beta_0(\cdot) = (\beta_{01}(\cdot), \dots, \beta_{0p}(\cdot))^T$ denote the true values of $\alpha(\cdot)$ and $\beta(\cdot)$, respectively. $a_n \asymp b_n$ indicates that $\frac{a_n}{b_n}$ has a lower bound greater than 0 and a finite upper bound.

We consider the following conditions (i.e. (A1)–(A5)) on the considered survival data. Here, Conditions (A1)–(A4) can be regarded as a direct extension of some standard conditions in the literature.

(A1) $\int_0^\tau h_0(t)dt < \infty$.

(A2) Let $w_{iR} = (\xi_{i1R}, \dots, \xi_{ip_1R}, z_{i1}, \dots, z_{ip})^T$. For $d = 0, 1, 2$, we define

$$S^{(d)}(\eta_R, t) = n^{-1} \sum_{i=1}^n (w_{iR})^{(d)} Y_i(t) \exp(\eta_R^T w_{iR} + e_i),$$

where $(w_{iR})^{(0)} = 1$, $(w_{iR})^{(1)} = w_{iR}$ and $(w_{iR})^{(2)} = (w_{iR})^{\otimes 2}$. Moreover, there exists a neighborhood B of the true value of η_R , denoted as η_{R0} , and a scalar, a vector or a matrix continuous function $s^{(d)}(\eta_R, t) = ES^{(d)}(\eta_R, t)$ defined on $B \times [0; \tau]$ such that

$$\sup_{t \in [0, \tau], \eta_R \in B} \left\| S^{(d)}(\eta_R, t) - s^{(d)}(\eta_R, t) \right\| \xrightarrow{p} 0 \quad \text{for } d = 0, 1, 2.$$

(A3) The functions $s^{(d)}$ for $d = 0, 1, 2$ are bounded on $B \times [0; \tau]$ and $s^{(d)}(\cdot; t)$ are continuous in $\eta_R \in B$ uniformly in $t \in [0; \tau]$. Moreover, $s^{(0)}$ is bounded away from 0 on $B \times [0; \tau]$.

(A4) The matrix $\Sigma(\eta_{R0}) = \int_0^\tau v(\eta_{R0}, t) s^{(0)}(\eta_{R0}, t) h_0(t) dt$ is positive definite, where $v(\eta_R, t) = \{s^{(0)}\}^{-1} s^{(2)} - \{s^{(0)}\}^{-2} \{s^{(1)}\}^{\otimes 2}$.

(A5) For any $1 \leq k \leq p$, z_k is subgaussian.

(A6) $\lambda_j - \lambda_{j+1} \geq K j^{-a-1}$ for $j \geq 1$ with $a > 1$.

(A7) $|\eta_{j0}| \leq CK^{-b}$ for $j > 1$.

(A8) For all $j \geq 1$, $E|X(\cdot)|^4 \leq K$, and $E[\xi_j^4] \leq K \lambda_j^2$.

(A9) For $a > 1$, $K^{-1} j^{-a} \leq \lambda_j \leq K j^{-a}$ and for all $j \geq 1$, the eigenvalues v_j^s satisfy $v_j - v_{j+1} \geq K^{-1} j^{-a-1}$.

(A10) The truncation number m satisfies $m \asymp n^{\frac{1}{a+2b}}$.

(A11) For all $u \in [0, 1]$, matrix $E(\mathbf{Z}\mathbf{Z}^T | U = u)$ is positive definite matrix, with eigenvalues greater than 0 and uniformly bounded.

(A12) Varying coefficient $\alpha_k(u)$ ($k = 1, \dots, p$) has continuous q -th order derivatives $\alpha_k^{(q)}(u)$ ($k = 1, \dots, p$), and satisfies

$$\left\| \alpha_k^{(q)}(u) - \alpha_k^{(q)}(u') \right\| \leq K |u - u'|^v, \quad 0 \leq u, u' \leq 1, \quad k = 1, \dots, p,$$

where $K > 0$, $0 < v \leq 1$, and $r = q + v$, $r \geq 1$.

(A13) The number of knots k_n in the space \mathcal{S}_n composed of spline functions satisfies $k_n \asymp n^{\frac{1}{1+2r}}$.

(A14) For $l = 1, \dots, p$, $\alpha_{0l}(\cdot)$ is r -th ($r \geq 2$) continuously differentiable on $[0, 1]$.

Here, (A6)–(A10) are very common in the functional linear regression models.^{20,21} (A6) prevents the spacing between eigenvalues from being too small to identify $\beta(t)$ and (A7) ensures that $\beta(t)$ is sufficiently smooth relative to the covariance function. (A11) ensures that the varying coefficient function is identifiable. (A12) to (A13) guarantee that the parameter $\alpha_k(u)$ is sufficiently smooth and can be approximated by spline functions, thereby ensuring that its estimation can achieve the optimal convergence rate. (A14) represents the smoothness condition of the varying coefficient function, describing the requirements for the optimal convergence rate that can be achieved in estimating the varying coefficient function.²²

Theorem 1: Under conditions A(1)–A(14), $m \sim n^{1/(a+2b)}$ and $k_n \sim n^{1/(1+2r)}$, we have

$$\begin{aligned} \|\hat{\beta}(\cdot) - \beta_0(\cdot)\|^2 &= O_p\left(n^{-\frac{2b-1}{a+2b}}\right) + O_p\left(n^{-\frac{2b}{a+2b} + \frac{1}{2r+1}}\right), \text{ and} \\ \|\hat{\alpha}(\cdot) - \alpha_0(\cdot)\|^2 &= O_p\left(n^{-\frac{a+2b-1}{a+2b}}\right) + O_p\left(n^{-2r/(2r+1)}\right). \end{aligned}$$

5 Simulation studies

5.1 Model setting

In this section, simulation studies are conducted to study the performance of the FVC-Cox model. First, we assume that the data are coming from the following model:

$$h(t | X, Z, V) = h_0(t) \exp\left(\mathbf{v}\boldsymbol{\theta} + \mathbf{Z}\boldsymbol{\alpha}(U) + \int_0^1 X(t)\boldsymbol{\beta}(t)dt\right),$$

where $h_0(t) = 1$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T = (1, 0.15, 0.35)^T$, and $\mathbf{Z}\boldsymbol{\alpha}(U) = \alpha_1(U)Z_1 + \alpha_2(U)Z_2$ with $\alpha_1(U) = 2\cos(2\pi U)$ and $\alpha_2(U) = -2 + (3 - U)^3/5$. For the functional data, let $\boldsymbol{\beta}(t) = \sqrt{2}\sin(\pi t/2) + 3\sqrt{2}\sin(3\pi t/2)$. For the random function, we set $X(t) = \sum_{i=1}^{\infty} \xi_i \phi_i(t)$ with ξ_i following $N(0, \lambda_i)$, $\lambda_i = ((i - 0.5)\pi)^{-2}$ and $\phi_j(t) = \sqrt{2}\sin((j - 0.5)\pi t)$. According to (3), it can be seen that when performing parameter estimation on the functional part, it is hard to directly obtain the result of the unknown $\boldsymbol{\beta}(t)$ through parameter estimation. However, after decomposing $\boldsymbol{\beta}(t)$, since the coefficient $(1, 3)^T$ in the expression $\boldsymbol{\beta}(t) = \sqrt{2}\sin(\pi t/2) + 3\sqrt{2}\sin(3\pi t/2)$ can be estimated, the subsequent numerical comparison of functional parameter estimation in the simulation will be mainly focused on the parameters $(1, 3)^T$.

5.2 Analysis of simulation results

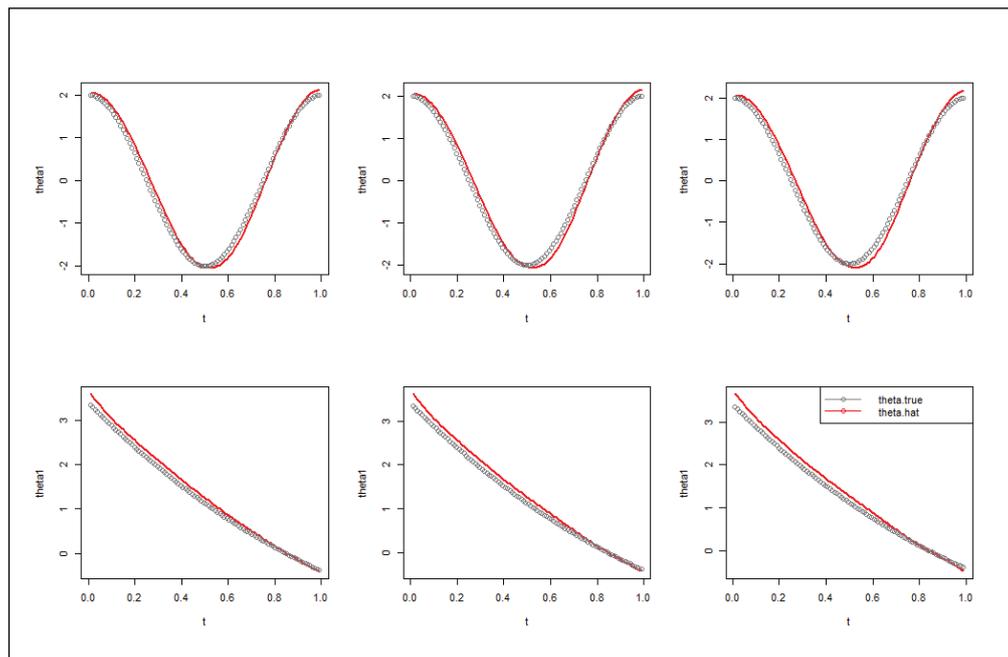
In this section, we consider scenarios with sample sizes of 500, 1000, and 2000, and simultaneously evaluate the proposed model under censoring rates of 0.1, 0.3, and 0.5. For all nine scenarios, we assess the performance of the model and estimators based on the root of average squared error (RASE) and integrated bias (IBIAS), which are defined as follows:

$$\begin{aligned} \text{RASE}(\hat{\beta}) &= \left(\frac{1}{n} \sum_{i=1}^n (\hat{\theta} - \theta)^2\right)^{1/2}, \\ \text{IBIAS}(\hat{\alpha}(\cdot)) &= \int_0^1 (E(\hat{\alpha}(u)) - \alpha(u))^2 du, \text{ and} \\ \text{IBIAS}(\hat{\beta}(\cdot)) &= \int_0^1 (E(\hat{\beta}(t)) - \beta(t))^2 dt. \end{aligned}$$

Table 1 shows the corresponding simulation results for the nine scenarios. As expected, all RASEs and IBIASs increase with censoring rate and decrease with sample size. Figure 1 plots the true and estimated values of the varying coefficient part at three different quantiles when the sample size is 500. According to Figure 1, there are insignificant discrepancies between the true values and the estimated values of the varying coefficients obtained from our proposed method. These results suggest that our model and estimation method are highly effective for estimating Cox models with varying coefficients.

Table 1. RASE and IBIAS for sample sizes of 500, 1000, and 2000 at censoring rates of 0.1, 0.3, and 0.5 (SZ: sample size; CR: censoring rate).

SZ	CR	θ_1	θ_2	θ_3	β_1	β_2	α_1	α_2
500	0.1	0.036	0.004	0.010	0.037	0.092	0.028	0.025
	0.3	0.040	0.000	0.010	0.040	0.094	0.030	0.028
	0.5	0.048	0.000	0.011	0.053	0.111	0.033	0.031
1000	0.1	0.014	0.003	0.008	0.018	0.043	0.005	0.013
	0.3	0.018	0.003	0.010	0.021	0.056	0.006	0.014
	0.5	0.022	0.005	0.010	0.026	0.063	0.006	0.014
2000	0.1	0.008	0.002	0.002	0.008	0.017	0.007	0.012
	0.3	0.009	0.001	0.002	0.007	0.027	0.007	0.012
	0.5	0.009	0.000	0.002	0.009	0.033	0.006	0.012

**Figure 1.** The true and estimated values of the varying coefficient part at three different quantiles, sample size is 500.

5.3 Comparison between varying coefficient and fixed coefficient functional Cox models

When the data is under the same model setting, a comparison can be made between the results of simulations with and without varying coefficients. To save space, only the results under fixed coefficient settings with censoring rates of 0.1, 0.3, and 0.5, and sample sizes of 500, 1000, and 2000 are presented in Table 2.

According to Table 2, it can be seen that when the varying coefficient part of the model is replaced with fixed coefficients, there is a significant difference between the simulated results and the true results. Compared to Table 1, the RASEs and IBIASs in Table 2 become much larger. These results clearly suggest that for varying coefficient functional data, the original functional Cox model cannot provide a good fit and estimation of the data. The FVC-Cox model proposed in this paper, however, can not only encompass both functional covariate effects and varying coefficient effects but also perform parameter estimation more effectively than the functional Cox model.

6 Real data analysis

The data used in this paper were obtained from the ADNI database (<https://adni.loni.usc.edu/>). The primary objective of ADNI is to test whether a combination of serial magnetic resonance imaging (MRI) and other biomarkers can be used to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD) since early and more

Table 2. RASE and IBIAS for fixed coefficient settings at censoring rates of 0.1, 0.3, and 0.5, and sample sizes of 500, 1000, and 2000 (SZ: sample size; CR: censoring rate).

SZ	CR	θ_1	θ_2	θ_3	β_1	β_2	α_1	α_2
500	0.1	0.551	0.084	0.190	0.549	1.667	2.929	2.711
	0.3	0.537	0.080	0.185	1.021	3.000	2.920	2.693
	0.5	0.536	0.078	0.184	1.020	3.000	2.874	2.758
1000	0.1	0.560	0.084	0.193	0.562	1.686	2.802	2.745
	0.3	0.545	0.081	0.189	0.549	1.636	2.831	2.711
	0.5	0.545	0.081	0.189	0.550	1.637	2.822	2.804
2000	0.1	0.560	0.081	0.200	0.559	1.692	2.801	2.758
	0.3	0.544	0.078	0.195	0.545	1.648	2.824	2.758
	0.5	0.544	0.079	0.195	0.545	1.644	2.800	2.751

Table 3. Estimated values of the linear component θ .

	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
Estimate	0.294	-0.127	-0.210	-0.153	0.139	0.015	-0.016
Standard error	0.187	0.327	0.302	0.365	0.765	0.027	0.216
	β_{11}	θ_{12}	θ_{13}	θ_{14}			
Estimate	-0.005	0.023	-0.077	-0.647			
Standard error	0.013	0.417	0.460	0.193			

accurate diagnosis of AD is considered an important therapeutic measure. This is the goal of researchers because therapeutic interventions are more likely to be beneficial during the early stages of the disease. At the same time, identifying sensitive and specific markers of early AD progression aims to help researchers and clinicians develop new treatments, monitor their effectiveness, and reduce the time and cost of clinical trials.

The hippocampus is one of the key brain regions affected by AD. The data set contains clinical and imaging measurements from 373 MCI individuals in ADNI, using them to predict the time from MCI conversion to AD. Among the 373 MCI individuals, 161 MCI individuals progressed to AD before the study was completed, while the remaining 212 individuals did not convert to AD by the end of the study. Therefore, the time from MCI to AD conversion can be regarded as time-to-event data, and the outcomes are interval censored.

In this paper, we use the proposed FVC-Cox model to fit the ADNI dataset. Here, the covariates include gender (1 = male; 0 = female), handedness (1 = right-handed; 2 = left-handed), whether widowed (0 = no; 1 = yes), whether divorced (0 = no; 1 = yes), whether married (0 = no; 1 = yes), years of education, whether retired (1 = yes; 0 = no), age, whether the first allele is genotype 3 (0 = no; 1 = yes), whether the first allele is genotype 4 (0 = no; 1 = yes), whether the second allele is genotype 3 (0 = no; 1 = yes), and ADAS-Cog score. This results in a design matrix Z with dimensions $(n, p) = (373, 12)$. For the functional predictor variables, we used the hippocampal radial distances at 2000 surface points on the left and right hippocampal surfaces as functional data. The radial distance, defined as the distance between the medial core of the hippocampus and the corresponding vertex, is a summary statistic of hippocampal shape and size.

Here, the interaction between ADAS-Cog score and age was considered, meaning that the coefficient of ADAS-Cog score is a function of age, denoted as $\alpha(U) = \alpha(\text{age})$. First, principal components are estimated by applying FPCA to the hippocampal radial distances. Next, the varying coefficient part is expanded using B-spline methods. Finally, parameters are estimated using FVC-Cox model and the proposed parameter estimation method. Based on the AIC criterion, the first three principal components are selected and these three principal components explain a total of 76.4% of the total variance. By utilizing the proposed FVC-Cox model and its estimation method, we can obtain the estimated values for the linear part θ (see Table 3) and the estimated values for the varying coefficient part α (age) (see Figure 2). For the functional coefficient $\beta(t)$, the graph of this coefficient can be derived using the formula $\hat{\beta}(t) = \sum_{j=1}^{\infty} \eta_j \hat{\phi}_j(t)$, as shown in Figure 3.

From Figure 2, the analysis results show that the coefficient of the ADAS-Cog score is a non-linear curve. For data with varying coefficients, using only fixed coefficients Cox analysis can lead to biased coefficient estimates and other issues. Therefore, the FVC-Cox model proposed in this paper is appropriate. Table 3 reports the parameter estimates and their corresponding standard errors. It can be seen that ADAS-Cog score is a significant factor, which is consistent with the conclusions from Kong et al.¹⁸ This result suggests that the ADAS-Cog score provides a good predictive power for the conversion from MCI to AD.

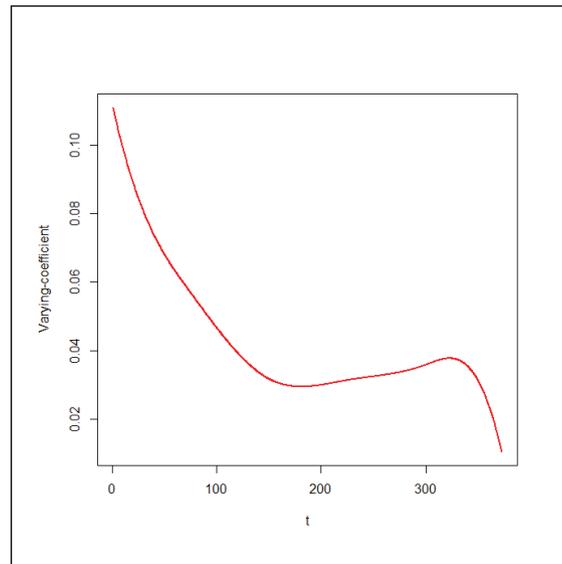


Figure 2. The variation of the varying coefficients $\alpha(\text{age})$ with age.

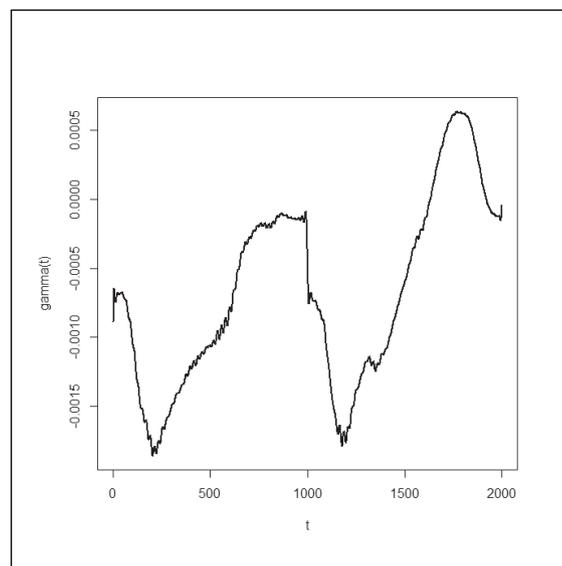


Figure 3. Functional coefficients $\beta(t)$ in ADNI data.

7 Conclusion

As data become more complex, the demand for model flexibility becomes higher. The functional Cox model incorporates functional data into the Cox model, increasing model flexibility and expanding the scope of application for functional data. Building on the functional Cox model, we further incorporate varying coefficients, which further enhances the flexibility of the survival model. Analysis of simulated data shows that this model performs better in scenarios where both functional data and varying coefficients exist at the same time than the functional Cox model. Our proposed model and estimation method also perform well in the real data analysis.

ORCID iDs

Maozai Tian  <https://orcid.org/0000-0002-0515-4477>

Man-lai Tang  <https://orcid.org/0009-0004-4247-6637>

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Professor Maozai Tian's work was partially supported by the Beijing Natural Science Foundation (No. 1242005), the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (25XNN015). Zhihao Wang's work was sponsored by Natural Science Foundation of Xinjiang Uygur Autonomous Region (No.2023D01A74).

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Cox DR. Regression models and life-tables. *J R Stat Soc: Ser B (Methodol)* 1972; **34**: 187–220.
2. Leemis LM, Shih LH and Reynertson K. Variate generation for accelerated life and proportional hazards models with time dependent covariates. *Stat Probab Lett* 1990; **10**: 335–339.
3. Austin PC. Generating survival times to simulate Cox proportional hazards models with time-varying covariates. *Stat Med* 2012; **31**: 3946–3958.
4. Yan J and Huang J. Model selection for Cox models with time-varying coefficients. *Biometrics* 2012; **68**: 419–428.
5. Ngwa JS, Cabral HJ, Cheng DM, et al. Generating survival times with time-varying covariates using the Lambert W function. *Commun Stat—Simul Comput* 2022; **51**:1: 135–153.
6. Cygu S, Dushoff J and Bolker BM. pCoxtime: penalized Cox proportional hazard model for time-dependent covariates. arXiv:2102.02297, 2021.
7. Zucker DM and Karr AF. Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *Ann Stat* 1990; **18**: 329–353.
8. Hastie T and Tibshirani R. Varying-coefficient models. *J R Stat Soc: Ser B (Methodol)* 1993; **55**: 757–796.
9. Cai Z and Sun Y. Local linear estimation for time-dependent coefficients in Cox's regression models. *Scandina J Stat* 2003; **30**: 93–111.
10. Song X and Wang CY. Time-varying coefficient proportional hazards model with missing covariates. *Stat Med* 2013; **32**: 2013–2030.
11. Kim G, Kim Y and Choi T. Bayesian analysis of the proportional hazards model with time-varying coefficients. *Scandina J Stat* 2017; **44**: 524–544.
12. Kim M, Paik MC, Jang J, et al. Cox proportional hazards models with left truncation and time-varying coefficient: application of age at event as outcome in cohort studies. *Biomet J* 2017; **59**: 405–419.
13. Yang Q, He H and Song X. Time-varying coefficient additive hazards model with latent variables. *Stat Methods Med Res* 2022; **31**: 928–946.
14. Shin H. Partial functional linear regression. *J Stat Plan Inference* 2009; **139**: 3405–3418.
15. Peng QY, Zhou JJ and Tang NS. Varying coefficient partially functional linear regression models. *Stat Papers* 2016; **57**: 827–841.
16. Wang Z, Liu Y, Tian M, et al. Functional partially varying coefficient model for zero-inflated count data. *Stat Res* 2021; **38**: 127–139.
17. Lee E, Zhu H, Kong D, et al. BFLCRM: a Bayesian functional linear Cox regression model for predicting time to conversion to Alzheimer's disease. *Ann Appl Stat* 2015; **9**: 2153–2178.
18. Kong D, Ibrahim JG, Lee E, et al. FLCRM: functional linear Cox regression model. *Biometrics* 2018; **74**: 109–117.
19. Horváth L and Kokoszka P. *Inference for functional data with applications*. 1st ed. Springer Series in Statistics, New York, 2012.
20. Cai TT and Hall P. Prediction in functional linear regression. *Ann Stat* 2006; **34**: 2159–2179.
21. Hall P and Horowitz JL. Methodology and convergence rates for functional linear regression. *Ann Stat* 2007; **35**: 70–91.
22. Wang K and Lin L. Robust structure identification and variable selection in partial linear varying coefficient models. *J Stat Plan Inference* 2016; **174**: 153–168.

Appendix

Proof of Theorem 1. For convenience, we define $V = (V_1, \dots, V_n)^T$ and $\omega = (\omega_1, \dots, \omega_n)^T$ are quasidesign matrix for slope function and varying-coefficient function component, respectively. Let $\eta_0 = (\eta_{0,1}, \dots, \eta_{0,m})^T$ be coefficient vector for true slope function $\beta_0(t)$. $\gamma_0 = (\gamma_{01}^T, \dots, \gamma_{0p}^T)^T$ be the true coefficients for the spline basis, $B = \text{diag}(f_1(0), \dots, f_n(0))$, $P = \omega(\omega^T B \omega)^{-1} \omega^T B$, $V^* = (I - P)V = (V_1^*, \dots, V_n^*)^T$, $H_n^2 = k_n \omega^T B \omega$ and $\Lambda_n = V^{*T} B V^*$.

We also let

$$\theta(\boldsymbol{\eta}) = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix} = \begin{pmatrix} \Lambda_n^{\frac{1}{2}} (\boldsymbol{\eta} - \boldsymbol{\eta}_0) \\ k_n^{-\frac{1}{2}} \mathbf{H}_n (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0) + \sqrt{k_n} \mathbf{H}_n^{-1} \boldsymbol{\omega}^T \mathbf{B} \mathbf{V} (\boldsymbol{\eta} - \boldsymbol{\eta}_0) \end{pmatrix}$$

and

$$\hat{\boldsymbol{\theta}} = \theta(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\gamma}})^T = (\hat{\boldsymbol{\theta}}_1^T, \hat{\boldsymbol{\theta}}_2^T)^T$$

Observe that

$$\begin{aligned} \|\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\|^2 &= \left\| \sum_{j=1}^m \hat{\eta}_j \hat{\phi}_j(t) - \sum_{j=1}^{\infty} \eta_{0j} \phi_j(t) \right\|^2 \\ &\leq 2 \left\| \sum_{j=1}^m \hat{\eta}_j \hat{\phi}_j(t) - \sum_{j=1}^m \eta_{0j} \phi_j(t) \right\|^2 + 2 \left\| \sum_{j=m+1}^{\infty} \eta_{0j} \phi_j(t) \right\|^2 \\ &\leq 4 \left\| \sum_{j=1}^m (\hat{\eta}_j - \eta_{0j}) \hat{\phi}_j(t) \right\|^2 + 4 \left\| \sum_{j=1}^m \eta_{0j} (\hat{\phi}_j(t) - \phi_j(t)) \right\|^2 \\ &\quad + 2 \sum_{j=m+1}^{\infty} \eta_{0j}^2 \\ &= 4A_1 + 4A_2 + 2A_3 \end{aligned}$$

where,

$$\begin{aligned} A_1 &= \left\| \sum_{j=1}^m (\hat{\eta}_j - \eta_{0j}) \hat{\phi}_j(t) \right\|^2 \\ A_2 &= \left\| \sum_{j=1}^m \eta_{0j} (\hat{\phi}_j(t) - \phi_j(t)) \right\|^2 \\ A_3 &= \sum_{j=m+1}^{\infty} \eta_{0j}^2 \end{aligned}$$

Invoking the orthogonality of $\hat{\phi}_j$, we have:

$$\begin{aligned} A_1 &= \left\| \sum_{j=1}^m (\hat{\eta}_j - \eta_{0j}) \hat{\phi}_j(t) \right\|^2 \leq \sum_{j=1}^m |\hat{\eta}_j - \eta_{0j}|^2 = \|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\|^2 \\ &= O_p(a_n^2) = O_p\left(n^{-\frac{2b-1}{a+2b}}\right) + O_p\left(k_n n^{-\frac{2b}{a+2b}}\right) \\ A_2 &= \left\| \sum_{j=1}^m \eta_{0j} (\hat{\phi}_j - \phi_j) \right\|^2 \leq m \sum_{j=1}^m \|\hat{\phi}_j - \phi_j\|^2 \eta_{0j}^2 \leq \frac{m}{n} \sum_{j=1}^m j^2 \eta_{0j}^2 \\ &= O_p\left(n^{-1} m \sum_{j=1}^m j^{2-2b}\right) = O_p(n^{-1} m) = o_p\left(n^{-\frac{2b-1}{a+2b}}\right) \\ A_3 &= \sum_{j=m+1}^{\infty} \eta_{0j}^2 \leq K \sum_{j=m+1}^{\infty} j^{-2b} = o\left(n^{-\frac{2b-1}{a+2b}}\right) \end{aligned}$$

Then, we can have:

$$\|\hat{\boldsymbol{\beta}}(\cdot) - \boldsymbol{\beta}_0(\cdot)\|^2 = 4A_1 + 4A_2 + 2A_3 = O_p\left(n^{-\frac{2b-1}{a+2b}}\right) + O_p\left(n^{-\frac{2b}{a+2b} + \frac{1}{2r+1}}\right)$$

By and Cauchy–Schwarz inequality and the definition of $\hat{\theta}$, we have:

$$\begin{aligned}
\|\hat{\gamma} - \gamma_0\|^2 &= \left\| \sqrt{k_n} \mathbf{H}_n^{-1} \hat{\theta}_2 - k_n \mathbf{H}_n^{-2} \boldsymbol{\omega}^T \mathbf{B} \mathbf{V} (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \right\|^2 \\
&\leq \left\| \sqrt{k_n} \mathbf{H}_n^{-1} \hat{\theta}_2 \right\|^2 + \left\| k_n \mathbf{H}_n^{-2} \boldsymbol{\omega}^T \mathbf{B} \mathbf{V} (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \right\|^2 \\
&= k_n \hat{\theta}_2^T \mathbf{H}_n^{-2} \hat{\theta}_2 + k_n^2 (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)^T \mathbf{V}^T \mathbf{B} \boldsymbol{\omega} \mathbf{H}_n^{-4} \boldsymbol{\omega}^T \mathbf{B} \mathbf{V} (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \\
&= k_n \hat{\theta}_2^T \mathbf{H}_n^{-2} \hat{\theta}_2 + k_n^2 (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)^T \mathbf{V}^T \mathbf{B} \boldsymbol{\omega} \mathbf{H}_n^{-2} (k_n \boldsymbol{\omega}^T \mathbf{B} \boldsymbol{\omega})^{-1} \boldsymbol{\omega}^T \mathbf{B} \mathbf{V} (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \\
&= O_p \left(\frac{k_n (k_n + m)}{n} \right) \\
&\quad + k_n \left(O_p \left(n^{-\frac{2b-1}{a+2b}} \right) + O_p \left(k_n n^{-\frac{2b}{a+2b}} \right) \right) O_p \left(\frac{1}{n} \right) O_p (n \hat{\lambda}_m) \\
&= O_p \left(\frac{k_n (k_n + m)}{n} \right).
\end{aligned}$$

Therefore, by Hölder inequality and the above equation, for $l = 1, \dots, p$, we have

$$\begin{aligned}
\|\hat{\alpha}_l(u) - \alpha_{0l}(u)\|^2 &= \left\| \boldsymbol{\omega}^T(u) \hat{\gamma}_l - \alpha_{0l}(u) \right\|^2 \\
&\leq 2 \left\| \boldsymbol{\omega}^T(u) (\hat{\gamma}_l - \gamma_{0l}) \right\|^2 + 2 \left\| \alpha_{0l}(u) - \boldsymbol{\omega}^T(u) \gamma_{0l} \right\|^2 \\
&= O \left(\frac{1}{k_n} \right) \|\hat{\gamma}_l - \gamma_{0l}\|^2 + O_p(k_n^{-2r}) \\
&= O \left(\frac{1}{k_n} \right) O_p \left(\frac{k_n (k_n + m)}{n} \right) + O_p(n^{-2r/(2r+1)}) \\
&= O_p(n^{-(a+2b-1)/(a+2b)}) + O_p(n^{-2r/(2r+1)})
\end{aligned}$$

Thus, this completes the proof of Theorem 1. □