Functional Individualized Treatment Regimes with Imaging Features

Xinyi Li[a] and Michael R. Kosorok[b]

[a]Clemson University, [b]University of North Carolina at Chapel Hill

Abstract: Precision medicine seeks to discover an optimal personalized treatment plan and thereby provide informed and principled decision support, based on the characteristics of individual patients. With recent advancements in medical imaging, it is crucial to incorporate patient-specific imaging features in the study of individualized treatment regimes. We propose a novel, data-driven method to construct interpretable image features which can be incorporated, along with other features, to guide optimal treatment regimes. The proposed method treats imaging information as a realization of a stochastic process, and employs smoothing techniques in estimation. We show that the proposed estimators are consistent under mild conditions. The proposed method is applied to a dataset provided by the Alzheimer's Disease Neuroimaging Initiative.

Key words and phrases: Precision medicine; Functional data; Bivariate spline; Imaging data; Treatment regime.

1. Introduction

Precision medicine is an approach to medical treatment that takes into account individual variability in various features, such as genomic characteristics, environment factors, medical history, lifestyle, and current health status. In contrast to traditional treatment strategies, which tend to be uniform and one-size-fits-all, precision medicine aims to provide optimal treatment that is tailored to the unique characteristics of each individual. To achieve this goal, individualized treatment rules are developed to formalize the decision-making process that translates patient information into recommended treatment. The identification and application of individualized treatment regimes (ITRs) are crucial areas of investigation in precision medicine research, particularly in the study of chronic diseases or disorders that necessitate diverse medical interventions. Notably, ITRs have been successfully applied in various clinical contexts, including but not limited to Alzheimer's Disease (Isaacson et al., 2019), diabetes (Luckett et al., 2020), cancers (Thall et al., 2007; Zhao et al., 2011), and attention deficit hyperactivity disorder (Pelham and Fabiano, 2008).

The field of medical imaging has experienced significant advancements that have had a notable impact on disease and health studies. In addition to characterizing abnormalities and providing morphologic disease information, imaging can also define the imaging phenotype of the disease, using an increasingly diverse set of modern imaging tools and imaging biomarkers (Herold et al., 2016). The

Address for correspondence: Xinyi Li, School of Mathematical and Statistical Sciences, Clemson University, Clemson, SC 29634, USA. Email: lixinyi@clemson.edu

integration of imaging features in precision medicine research has the potential to significantly enhance patient care by enabling effective early detection and diagnosis tools, as well as guiding optimal treatments and lifestyle interventions. For instance, Alzheimer's disease-related pathophysiologic changes can be visualized up to 15 years before the onset of clinical dementia (Bateman et al., 2012). Thus, biomedical imaging and imaging-guided interventions, which provide multiparametric morphologic and functional information, constitute crucial elements in the infrastructure of precision medicine. Consequently, it is essential to develop strategies that incorporate imaging data in conjunction with other abundant information in precision medicine research.

In this work, we propose a novel and flexible approach, Functional Individualized Treatment Regimes with Imaging features (FITRI), to adapt semiparametric learning and functional data analysis (FDA) frameworks to precision medicine with abundant features, including medical images, genetic features, environment factors, etc. Semiparametric modeling and FDA offer powerful tools to deal with image features, allowing for a flexible range of possible model specifications. Specifically, we apply FDA for low-dimensional representation of high-resolution and high-dimensional imaging features. In other words, the imaging features are represented as functional terms in our analysis. We first develop a multi-dimensional functional principal component (FPC) basis expansion tool to approximate the functional imaging features, and then conduct a Q-learning (Clifton and Laber, 2020; Kosorok and Laber, 2019) type analysis to estimate the optimal treatment regimes. We also investigate the theoretical properties of the proposed bases and estimators, and develop computationally efficient algorithms accordingly.

The proposed work presents a significant contribution to the field of precision medicine research in several ways. First and foremost, the FITRI approach bridges the gap in the current literature in precision medicine studies that incorporate imaging features, which is not a trivial task due to the large volume and complexity of image data. The development and analysis of FITRI require efficient statistical methods and decision-support algorithms, which will be discussed later. And the proposed estimates provide an efficient data-driven summary statistic for the information contained in the imaging features. Moreover, the proposed work expands the existing methods that use functional data to construct optimal ITRs. Prior studies, including McKeague and Qian (2014), Ciarleglio et al. (2015), Ciarleglio et al. (2016), Ciarleglio et al. (2018), Laber and Staicu (2018) and Park et al. (2021), have developed methods for various FDA models for dense or sparse functional data. All of these works focus on functional data with a one-dimensional index continuum. In contrast, the proposed basis tool demonstrates both theoretical superiority and computational efficiency in dealing with imaging data, which can be widely used in the FDA approach dealing with image data (Nathoo et al., 2019; Zhu et al., 2022). Such a smoothing tool is lacking in the current literature on functional data indexed by two or higher-dimensional continuums. Not only do the proposed multi-dimensional FPC bases work for index continuum dimensions greater than 1, they also provide practical, data-driven interpretation

for the dominant information in the imaging features.

The rest of the paper is organized as follows. Section 2 presents the construction methods for the proposed multi-dimensional FPC basis for multivariate functional data, with a particular focus on the two-dimensional (2D) case. In Section 3, we apply the proposed 2D-FPC basis to the functional model and describe the corresponding estimator. Section 4 provides a presentation of the theoretical findings, while Section 5 elaborates on the implementation aspects. In Section 6, we furnish empirical investigations to showcase the performance of the proposed estimator. Additionally, Section 7 provides an illustrative example of real-world data in the context of Alzheimer's Disease. The conclusion of the paper is presented in Section 8, which discusses open issues and future work. The supplemental materials provide theoretical details.

2. Bases approximation for imaging features

We assume the observed data are $\{\mathbf{X}_i, Z_i(\boldsymbol{s}), A_i, Y_i\}_{i=1}^n$, which comprise n independent, identically distributed (iid) copies of a trajectory $\{\mathbf{X}, Z(\boldsymbol{s}), A, Y\}$, where $\mathbf{X} \in \mathcal{X} \subseteq \mathbb{R}^q$ denotes patient information, $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{V}\}$ is a stochastic process indexed by $\mathcal{V} \subset \mathbb{R}^m$ with $m \geq 1$, $A \in \mathcal{A} = \{0, 1\}$ is a binary treatment, and Y is real-valued outcome for which higher values are more desirable. For the ith subject, the real-valued imaging measure $Z_i(\boldsymbol{s}_j)$ is observed at point $\boldsymbol{s}_j \in \mathcal{V}$ for $j = 1, \ldots, N_s$. For example, when $m = 2$, $\boldsymbol{s}_j = (s_{j1}, s_{j2})^\top$ could be the coordinates of a voxel center in a two-dimensional brain image. Let μ be a nonnegative finite measure on $\mathcal{V}$, and define $L_2(\mathcal{V}, \mu) = \{f : \|f\|_{\mu,2} < \infty\}$, where $\|f\|_{\mu,2}^2 = \int_{\mathcal{V}} |f(\boldsymbol{s})|^2 \mathrm{d}\mu(\boldsymbol{s})$. Note that $L_2(\mathcal{V}, \mu)$ is a closed linear space, specifically, a Hilbert space. We assume that $Z_1, \ldots, Z_n$ are iid with distribution f such that $\|Z\|_{\mu,2} < \infty$ almost surely.

We employ a Q-learning type analysis. Specifically, define the Q-function as

$$Q\{\mathbf{x}, z(\boldsymbol{s}), a\} = \mathrm{E}\{Y|\mathbf{X} = \mathbf{x}, Z(\boldsymbol{s}) = z(\boldsymbol{s}), A = a\}. \tag{2.1}$$

Assume that we want to parametrize all linear functionals of Z, which can be characterized as

$$\int_{\mathcal{V}} Z(\boldsymbol{s}) g(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}) \tag{2.2}$$

for some $g \in L_2(\mathcal{V}, \mu)$. To be specific, we assume a semiparametric working model:

$$\mathrm{E}\{Y|\mathbf{X} = \mathbf{x}, Z(\boldsymbol{s}) = z(\boldsymbol{s}), A = a\}$$
$$= \boldsymbol{\alpha}_1^\top \mathbf{X} + \int_{\mathcal{V}} \beta_1(\boldsymbol{s}) Z(\boldsymbol{s}) \mathrm{d}\boldsymbol{s} + A \left\{\boldsymbol{\alpha}_2^\top \mathbf{X} + \int_{\mathcal{V}} \beta_2(\boldsymbol{s}) Z(\boldsymbol{s}) \mathrm{d}\boldsymbol{s}\right\}, \tag{2.3}$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_2^\top)^\top$ are linear coefficients, and $\beta_1(\boldsymbol{s})$ and $\beta_2(\boldsymbol{s})$ are coefficient functions/maps.

Efficient tools are essential for addressing coefficient maps in estimating the Q-function and, subsequently, determining the optimal treatment regime. The analysis of imaging data presents challenges

such as high dimensionality and resolution, complex geometric structures, including complex boundaries, and spatial heterogeneity. A variety of nonparametric methods have been developed for analyzing images; see for example, tensor-product-based kernel smoothing (Zhu et al., 2014), thin plate spline smoothing (Ramsay, 2002), soap film smoothing (Wood et al., 2008), and multivariate spline over triangulation (MST; Lai and Schumaker, 2007). MST has demonstrated superiority in analyzing multi-dimensional (MD) imaging data, as evidenced by its application in bivariate spline analysis of 2D images (Lai and Wang, 2013) and trivariate spline analysis of 3D images (Li et al., 2023). The construction of MST involves two main steps: constructing triangulation to approximate the entire domain and constructing multivariate splines based on the triangulation. In dealing with MD data, the complex and irregular shape poses a primary challenge compared to 1D data, which conventional nonparametric methods struggle to estimate accurately along the boundary, commonly known as the "leakage" issue (Ramsay, 2002; Wang and Ranalli, 2007). However, the triangulation, a set of MD simplices whose union approximates the domain $\mathcal{V}$ well, provides an efficient tool to address the "leakage" issue. Based on the constructed triangulation, the multivariate spline can be constructed with explicit formulas, allowing efficient approximation of the MD images. One can refer to Lai and Schumaker (2007) for additional technical details.

The use of MST presents numerous advantages in the analysis of MD imaging data. However, when applied to the model (2.3), theoretical and computational limitations emerge. Specifically, to establish the asymptotic consistency of the proposed coefficients estimates and, consequently, the estimated optimal regime, the stability condition for the bases inside the integral is required. Unfortunately, such a condition is lacking in the literature for MST. Moreover, the construction of multivariate splines typically involves hundreds or thousands of bases, rendering the whole system underdetermined, given the limited sample size. To overcome these limitations, we propose an MD-FPC basis based on MST, in a parallel fashion to the construction of 1D FPC bases. By constructing a Reproducing Kernel Hilbert Space (RKHS) based on the multivariate spline space, the low-dimensional representation of MD-FPC can approximate the multivariate spline space, and thus provide a good approximation to the original space $L_2(\mathcal{V}, \mu)$. This approach offers theoretical and computational advantages and allows for the handling of the complicated MD image domain.

2.1. Construction of MD-FPC basis

In the following, while we develop the methodology for general L_2 spaces, including $\mathcal{V} \subset \mathbb{R}^m$, for $m \geq 1$; we also will specialize some of the results for $\mathcal{V} \subset \mathbb{R}^2$, that is, for the 2D-FPC basis.

We first need the following assumptions:

(A1) (*Space*) Suppose we have a Hilbert space $\mathcal{H}_0 \equiv L_2(\mathcal{V}, \mu)$, and assume $Z \in \mathcal{H}_0$ with probability one.

(A2) (*Covariance function*) Assume $\mathrm{E}\|Z\|_{\mu,2}^2 < \infty$. Let $V_0(\boldsymbol{s}, \boldsymbol{s}') = \mathrm{E}\{Z(\boldsymbol{s}) - \mathrm{E}Z(\boldsymbol{s})\}\{Z(\boldsymbol{s}') - \mathrm{E}Z(\boldsymbol{s}')\}$, and suppose $V_0(\boldsymbol{s}, \boldsymbol{s}') = \sum_{k=1}^{\infty} \lambda_k \phi_k(\boldsymbol{s})\phi_k(\boldsymbol{s}')$, where $\{\phi_k, k \geq 1\}$ is an orthonormal basis contained in $\mathcal{H}_0$, with corresponding eigenvalues $0 \leq \cdots < \lambda_2 < \lambda_1 < \infty$.

(A3) (*Eigenvalues*) $\sum_{k=1}^{\infty} \lambda_k < \infty$.

Remark 1. *By Assumptions (A1) and (A2), we have the Karhunen-Loéve expansion $Z(\boldsymbol{s}) = \mathrm{E}Z(\boldsymbol{s}) + \sum_{k=1}^{\infty} \xi_k \phi_k(\boldsymbol{s})$, where $\xi_k = \int_{\mathcal{V}} \phi_k(\boldsymbol{s})\{Z(\boldsymbol{s}) - \mathrm{E}Z(\boldsymbol{s})\}\mathrm{d}\mu(\boldsymbol{s})$, $\mathrm{E}\xi_k = 0$, $\mathrm{Var}(\xi_k) = \lambda_k$, and $\mathrm{E}(\xi_k \xi_{k'}) = 0$, for $1 \leq k < k'$. Let $\mathcal{H}_1$ be the closed linear span of the basis functions $\{\phi_k, k \geq 1\}$, then consequently $\mathcal{H}_1 \subset \mathcal{H}_0$ is also a Hilbert space, and Assumption (A3) will then imply that $\mathrm{Pr}(Z - \mathrm{E}Z \in \mathcal{H}_1) = 1$.*

Let $\mathcal{B}_1 = \{b(\boldsymbol{s}) \in \mathcal{H}_0 : \|b\|_{\mu,2} \leq 1\}$, $\mathcal{F}_1 = \{\int_{\mathcal{V}} b(\boldsymbol{s})\{Z(\boldsymbol{s}) - \mathrm{E}Z(\boldsymbol{s})\}\mathrm{d}\mu(\boldsymbol{s}) : b(\boldsymbol{s}) \in \mathcal{B}_1\}$, and $\mathcal{F}_2 = \mathcal{F}_1 \times \mathcal{F}_1 = \{\int_{\mathcal{V} \times \mathcal{V}} a(\boldsymbol{s})b(\boldsymbol{s}')\{Z(\boldsymbol{s}) - \mathrm{E}Z(\boldsymbol{s})\}\{Z(\boldsymbol{s}') - \mathrm{E}Z(\boldsymbol{s}')\}\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}') : a(\boldsymbol{s}), b(\boldsymbol{s}') \in \mathcal{B}_1\}$. Theorem A.1 in supplemental materials shows that both $\mathcal{F}_1$ and $\mathcal{F}_2$ are Glivenko-Cantelli classes of functions of Z. Let $\mathcal{S}$ be the closed linear space spanned by MST in $\mathcal{H}_0$. Let G_N be the operator that projects onto $\mathcal{S}$, where the index N is the number of spline bases that grow with the sample size n. We need the following additional assumption:

(A4) (*Convergence of spline space*) Assume that for any $g \in \mathcal{H}_1$ such that $\|g\|_{\mu,2} \leq 1$, we have that $\|G_N g - g\|_{\mu,2} \to 0$ as $n \to \infty$.

Remark 2. *Assumption (A4) assumes the convergence of the spline space, which is a regular conclusion in the MST literature; see, for example, Lai and Wang (2013) and Li et al. (2023).*

Denote the empirical variance function as $V_n(\boldsymbol{s}, \boldsymbol{s}') = n^{-1}\sum_{i=1}^{n}\{Z_i(\boldsymbol{s}) - \bar{Z}_n(\boldsymbol{s})\}\{Z_i(\boldsymbol{s}') - \bar{Z}_n(\boldsymbol{s}')\}$, where $\bar{Z}_n = n^{-1}\sum_{i=1}^{n} Z_i$. We first obtain the sequence of theoretical bases based on $V_n(\boldsymbol{s}, \boldsymbol{s}')$. Define

$$\widehat{\phi}_{n1}(\boldsymbol{s}) = \arg\max_{f \in \mathcal{B}_1 \cap \mathcal{S}} \int_{\mathcal{V} \times \mathcal{V}} f(\boldsymbol{s})f(\boldsymbol{s}')V_n(\boldsymbol{s}, \boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}'). \tag{2.4}$$

As we show in Theorem A.2 in Section A in supplemental materials, $\widehat{\phi}_{n1}$ converges to ϕ_1 up to sign almost surely, as $n \to \infty$. In addition, we define

$$\begin{aligned}
\widehat{\mathcal{H}}_{n1} &= \{\text{closed linear span of } \{\widehat{\phi}_{n1}(\boldsymbol{s})\} \text{ in } \mathcal{H}_0\}, \\
\widehat{\phi}_{n2}(\boldsymbol{s}) &= \arg\max_{f \in \mathcal{B}_1 \cap \mathcal{S} \cap \widehat{\mathcal{H}}_{n1}^{\perp}} \int_{\mathcal{V} \times \mathcal{V}} f(\boldsymbol{s})f(\boldsymbol{s}')V_n(\boldsymbol{s}, \boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}'),
\end{aligned}$$

where $\widehat{\mathcal{H}}_{n1}^{\perp}$ denotes the closed orthocomplements of $\widehat{\mathcal{H}}_{n1}$ in $L_2(\mathcal{V}, \mu)$. Similarly, $\widehat{\phi}_{n2}$ converges to ϕ_2

5

up to sign almost surely, as $n \to \infty$. Repeating the construction recursively, we have for $k = 1, \ldots, K$,

$$\widehat{\mathcal{H}}_{nk} = \{\text{closed linear span of } \{\widehat{\phi}_{n1}, \ldots, \widehat{\phi}_{nk}\} \text{ in } \mathcal{H}_0\},$$

$$\mathcal{H}_{0k} = \{\text{closed linear span of } \{\phi_1, \ldots, \phi_k\} \text{ in } \mathcal{H}_0\}, \tag{2.5}$$

$$\widehat{\phi}_{n(k+1)}(\boldsymbol{s}) = \underset{f \in \mathcal{B}_1 \cap \mathcal{S} \cap \widehat{\mathcal{H}}_{nk}^\perp}{\arg \max} \int_{\mathcal{V} \times \mathcal{V}} f(\boldsymbol{s}) f(\boldsymbol{s}') V_n(\boldsymbol{s}, \boldsymbol{s}') \mathrm{d}\mu(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}').$$

As shown in Theorem 1 in Section 4.1, $\{\widehat{\phi}_{n1}, \ldots, \widehat{\phi}_{nK}\}$ form an orthonormal system on $\mathcal{S} \cap \mathcal{H}_0$, and $\{\widehat{\phi}_{nk}\}_k$ converge to $\{\phi_k\}_k$ up to sign almost surely, as $n \to \infty$.

In practice, in the 2D-FPC setting, we obtain the measurements of the image data $\{Z_i(\boldsymbol{s_j})\}_{i=1,j=1}^{n,N_s}$ on a finite grid or random set of points, instead of the whole continuum. That means, $V_n(\boldsymbol{s}, \boldsymbol{s}')$ in (2.4) and (2.5) is not computable for practical data settings; consequently, $\{\widehat{\phi}_{nk}\}_k$ is not directly accessible. Therefore, we propose to employ bivariate splines over triangulation (BST) as an initial smoothing tool to construct the 2D-FPC basis function. In 2D-FPC, we are specializing to the setting where $\mathcal{V} \subset \mathbb{R}^2$.

Specifically, let $\mathbf{B}(\boldsymbol{s}) = \{B_J(\boldsymbol{s}) \in \mathcal{S}\}_{J \in \mathcal{J}}$ be the BST basis of $\mathcal{S}$, and let $\mathbf{H} = \int_{\mathcal{V}} \mathbf{B}(\boldsymbol{s}) \mathbf{B}^\top(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s})$ with dimension $|\mathcal{J}| \times |\mathcal{J}|$, where $|\mathcal{J}|$ is the cardinality of the bivariate basis index set $\mathcal{J}$. Define a "presmoothed" variance-covariance matrix by bivariate spline as

$$\mathbf{K}_n = \mathbf{H}^{-1/2} \left\{ \int_{\mathcal{V} \times \mathcal{V}} \mathbf{B}(\boldsymbol{s}) V_n(\boldsymbol{s}, \boldsymbol{s}') \mathbf{B}(\boldsymbol{s}')^\top \mathrm{d}\mu(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}') \right\} \mathbf{H}^{-1/2}.$$

Let $\{\widehat{\boldsymbol{\varphi}}_{nk}\}_{k=1}^{p_n}$ be the eigenvectors of $\mathbf{K}_n$, ordered from largest eigenvalue to smallest, where $p_n = |\mathcal{J}| \wedge n$. Then for $k \geq 1$, we define

$$\widehat{\phi}_{nk}^{\mathcal{S}}(\boldsymbol{s}) = \widehat{\boldsymbol{\varphi}}_{nk}^\top \mathbf{H}^{-1/2} \mathbf{B}(\boldsymbol{s}), \quad \widehat{\lambda}_{nk} = \int_{\mathcal{V} \times \mathcal{V}} \widehat{\phi}_{nk}(\boldsymbol{s}) \widehat{\phi}_{nk}(\boldsymbol{s}') V_n(\boldsymbol{s}, \boldsymbol{s}') \mathrm{d}\mu(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}'). \tag{2.6}$$

As shown in Theorem 1 (iii) in Section 4.1, $\{\widehat{\phi}_{nk}(\boldsymbol{s})\}_k$ and $\{\widehat{\phi}_{nk}^{\mathcal{S}}(\boldsymbol{s})\}_k$ are asymptotically equal up to sign almost surely, as $n \to \infty$. Also, by Theorem 1 (iv) in Section 4.1, we have that $\{\widehat{\lambda}_{nk}\}_k$ converge to $\{\lambda_k\}_k$ almost surely. Note that the $\widehat{\lambda}_{nk}$ values are also the eigenvalues of $\mathbf{K}_n$. Therefore, we obtain that the $\{\widehat{\phi}_{nk}^{\mathcal{S}}(\boldsymbol{s})\}_k$ functions are the constructed orthonormal bases.

3. Functional learning with imaging features for estimating optimal ITRs

With all the preparation in Section 2, we are ready to implement functional learning with imaging features for estimating the optimal ITRs.

An individualized treatment rule is a function $\pi : \mathcal{X} \times \mathcal{V} \to \mathcal{A}$ so that under π, a patient presenting with $\mathbf{X} = \mathbf{x}$ and $Z(\boldsymbol{s}) = z(\boldsymbol{s})$ will be assigned to treatment $\pi(\mathbf{x}, z(\boldsymbol{s}))$. We employ the potential outcome framework to define an optimal ITR. Let $Y^*(a)$ be the potential outcome under treatment $a \in \mathcal{A}$, and under any regime π, $Y^*(\pi) = \sum_{a \in \mathcal{A}} Y^*(a) 1\{\pi(\mathbf{X}, Z(\boldsymbol{s})) = a\}$. For a decision rule π,

6

let $V(\pi) = \mathrm{E}Y^*(\pi)$ be the value of π. An optimal regime for Y, denoted as π^{opt}, satisfies $V(\pi^{\mathrm{opt}}) \geq V(\pi)$ for any other regime π.

Throughout the paper, we use capital letters, like $\mathbf{X}$, to denote random variables, and lower cases like $\mathbf{x}$ for instances of corresponding random variables. To identify the optimal regimes in terms of the data-generating model, we make the following assumptions:

(A5) (Consistency) $Y = Y^*(A)$.

(A6) (Positivity) There exists some constant c such that $\Pr\{A = a|\mathbf{X}, Z(\boldsymbol{s})\} \geq c > 0$ for each $a \in \mathcal{A}$, $\mathbf{X} \in \mathcal{X}$ and $Z(\boldsymbol{s}) \in \mathcal{V}$ almost surely.

(A7) (Ignorability) $\{Y^*(-1), Y^*(1)\} \perp A|\mathbf{X}, Z(\boldsymbol{s})$.

Remark 3. *Assumptions (A5)–(A7) are standard in the context of precision medicine (Hernán and Robins, 2010; Robins, 2004). Assumption (A6) is verifiable in the study. Assumption (A7) implies there are no unmeasured confounders, which generally holds by construction in randomized studies but is not in general empirically verifiable in observational studies (Rosenbaum, 1984; Rosenbaum and Rubin, 1983).*

Under Assumptions (A5)–(A7), and recalling the definition of the Q-function in (2.1), the optimal ITR is $\pi^{\mathrm{opt}}\{\mathbf{x}, z(\boldsymbol{s})\} = \arg\max_{a \in \mathcal{A}} Q\{\mathbf{x}, z(\boldsymbol{s}), a\}$. Let $\widehat{Q}\{\mathbf{x}, z(\boldsymbol{s}), a\}$ be the functional-regression-based estimators fit using the data. Then the optimal ITR can be estimated through $\widehat{Q}\{\mathbf{x}, z(\boldsymbol{s}), a\}$ via

$$\widehat{\pi}^{\mathrm{Q}}\{\mathbf{x}, z(\boldsymbol{s})\} = \arg\max_{a \in \mathcal{A}} \widehat{Q}\{\mathbf{x}, z(\boldsymbol{s}), a\}.$$

Switch back now to functional regression modeling. Recall our model in (2.3), or equivalently, assume

$$Y = \boldsymbol{\alpha}_1^\top \mathbf{X} + A\boldsymbol{\alpha}_2^\top \mathbf{X} + \int_{\mathcal{V}} \beta_1(\boldsymbol{s})Z(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}) + A \int_{\mathcal{V}} \beta_2(\boldsymbol{s})Z(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}) + \varepsilon,$$

where $\mathbf{X}$ includes a $\{1\}$ term (intercept) and is in $\mathbb{R}^q$, $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \mathbb{R}^q$, A is binary treatment indicator (can assume $A \in \{0, 1\}$), $Z(\boldsymbol{s})$ is a stochastic process on $\mathcal{H}_0$, and $\beta_1(\boldsymbol{s}), \beta_2(\boldsymbol{s}) \in \mathcal{H}_0$. Assume $\mathrm{E}\|\mathbf{X}\|_2^2 < \infty$, $\mathrm{E}(\varepsilon|\mathbf{X}, A, Z) = 0$ almost surely, $\mathrm{E}\{\mathrm{Var}(\varepsilon|\mathbf{X}, A, Z)\} < \infty$ and $\mathrm{Var}(\varepsilon) = \sigma^2$.

Denote $\boldsymbol{\theta} = (\boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_2^\top, \beta_1, \beta_2)^\top$. Let the generic covariate vector $\mathbf{W} = (\mathbf{X}^\top, A\mathbf{X}^\top, Z, AZ)^\top$, and define $\boldsymbol{\theta}(\mathbf{W}) = \boldsymbol{\alpha}_1^\top \mathbf{X} + \boldsymbol{\alpha}_2^\top A\mathbf{X} + \int_{\mathcal{V}} \beta_1(\boldsymbol{s})Z(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}) + \int_{\mathcal{V}} \beta_2(\boldsymbol{s})AZ(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s})$. Also, for subject i, the observed covariate is $\mathbf{W}_i = (\mathbf{X}_i^\top, A_i\mathbf{X}_i^\top, Z_i, A_iZ_i)^\top$. Note that $\boldsymbol{\theta}$ is a linear operator, instead of a pure linear coefficient; and $\mathbf{W}$ and $\mathbf{W}_i$ are "functional matrices", namely a hybrid of both matrices and functional covariates. We need the following Assumption (A8) for the linear operator $\boldsymbol{\theta}$.

(A8) (*Linear operator*) $\boldsymbol{\theta}(\mathbf{W}_i) = 0$ a.s. $\Rightarrow \|\boldsymbol{\theta}\|_* = 0$, where $\|\boldsymbol{\theta}\|_* = \|\boldsymbol{\alpha}_1\|_2 + \|\boldsymbol{\alpha}_2\|_2 + \|\beta_1\|_{\mu,2} + \|\beta_2\|_{\mu,2}$.

Remark 4. *Assumption (A8) can be implied by* $\mathrm{E}\mathbf{W}_i\mathbf{W}_i^\top$ *being positive definite, since* $\mathrm{E}\{\boldsymbol{\theta}(\mathbf{W}_i)\}^2 = \boldsymbol{\theta}^\top\mathrm{E}\mathbf{W}_i\mathbf{W}_i^\top\boldsymbol{\theta} = 0$ *if and only if* $\boldsymbol{\theta} = \mathbf{0}$.

We employ the 2D-FPC basis proposed in Section 2 to approximate the covariates Z and AZ. To be specific, denote the 2D-FPC bases for $Z(\boldsymbol{s})$ and $AZ(\boldsymbol{s})$ as $\{\widehat{\phi}_{1,nk}(\boldsymbol{s})\}_k$ and $\{\widehat{\phi}_{2,nk}(\boldsymbol{s})\}_k$, respectively, then

$$Z(\boldsymbol{s}) \approx \sum_{k=1}^{K_1}\widehat{U}_{1,k}\widehat{\phi}_{1,nk}(\boldsymbol{s}), \quad AZ(\boldsymbol{s}) \approx \sum_{k=1}^{K_2}\widehat{U}_{2,k}\widehat{\phi}_{2,nk}(\boldsymbol{s}), \tag{3.1}$$

where $\widehat{U}_{1,k} = \int_{\mathcal{V}}\widehat{\phi}_{1,nk}(\boldsymbol{s})Z(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s})$, $\widehat{U}_{2,k} = \int_{\mathcal{V}}\widehat{\phi}_{2,nk}(\boldsymbol{s})AZ(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s})$, and K_1 and K_2 are basis numbers. Note that we exclude the expectation terms in Equation (3.1), which are absorbed into the intercept terms and have no effect on the estimation procedure. Consequently, $\forall k$, $\widehat{U}_{1,k}$ and $\widehat{U}_{2,k}$ does not necessarily have mean zero. Fix $1 \le K_1, K_2 < \infty$. Note that K_1 and K_2 can be zero, in which case the corresponding terms vanish in the functional regression model. To avoid such trivialities, we will require $K_1, K_2 \ge 1$. As for the selection of K_1 and K_2, we discuss this in detail in Section 5.1. Accordingly, the coefficient functions $\beta_1(\boldsymbol{s})$ and $\beta_2(\boldsymbol{s})$ have the representations $\beta_\ell(\boldsymbol{s}) = \sum_{k=1}^{K_\ell}\gamma_{\ell,k}\widehat{\phi}_{\ell,nk}(\boldsymbol{s})$ for $\ell = 1, 2$, where $\gamma_{\ell,k} = \int_{\mathcal{V}}\widehat{\phi}_{\ell,nk}(\boldsymbol{s})\beta_\ell(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s})$.

For $1 \le K_\ell \le p_n$, let $\widehat{\mathbf{U}}_\ell(K_\ell) = (\widehat{U}_{\ell,1}, \ldots, \widehat{U}_{\ell,K_\ell})^\top$ and $\widehat{\mathbf{U}}_{\ell,i}(K_\ell) = (\widehat{U}_{\ell,i1}, \ldots, \widehat{U}_{\ell,iK_\ell})^\top$, $\ell = 1, 2$, where for $1 \le i \le n$,

$$\widehat{U}_{1,ik} = \int_{\mathcal{V}}\widehat{\phi}_{1,nk}(\boldsymbol{s})Z_i(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}), \ 1 \le k \le K_1,$$

$$\widehat{U}_{2,ik} = \int_{\mathcal{V}}\widehat{\phi}_{2,nk}(\boldsymbol{s})A_iZ_i(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}), \ 1 \le k \le K_2.$$

Note that $\mathrm{Var}(\widehat{U}_{\ell,ik}) = \widehat{\lambda}_{\ell,nk}$, where $\widehat{\lambda}_{\ell,nk}$ are defined in (2.6) with $V_{\ell,n}(\boldsymbol{s}, \boldsymbol{s}')$, $\ell = 1, 2$, respectively, with

$$V_{1,n}(\boldsymbol{s}, \boldsymbol{s}') = \frac{1}{n}\sum_{i=1}^n\{Z_i(\boldsymbol{s}) - \bar{Z}_n(\boldsymbol{s})\}\{Z_i(\boldsymbol{s}') - \bar{Z}_n(\boldsymbol{s}')\},$$

$$V_{2,n}(\boldsymbol{s}, \boldsymbol{s}') = \frac{1}{n}\sum_{i=1}^n\{A_iZ_i(\boldsymbol{s}) - (\overline{AZ})_n(\boldsymbol{s})\}\{A_iZ_i(\boldsymbol{s}') - (\overline{AZ})_n(\boldsymbol{s}')\}. \tag{3.2}$$

In a parallel fashion to $\mathbf{W}$ and $\mathbf{W}_i$, for $1 \le i \le n$, let $\widehat{\mathbf{W}} = (\mathbf{X}^\top, A\mathbf{X}^\top, \widehat{\mathbf{U}}_1^\top(K_1), A\widehat{\mathbf{U}}_2^\top(K_2))^\top$ and $\widehat{\mathbf{W}}_i = (\mathbf{X}_i^\top, A_i\mathbf{X}_i^\top, \widehat{\mathbf{U}}_{1,i}^\top(K_1), A_i\widehat{\mathbf{U}}_{2,i}^\top(K_2))^\top$. Assume the full parameter space is $\boldsymbol{\Theta} = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \mathbb{R}^q, \beta_1, \beta_2 \in \mathcal{H}_0\}$. Recall the definitions of $\mathcal{H}_{0K}$ and $\widehat{\mathcal{H}}_{nK}$ in (2.5), which are the closed linear spaces spanned by bases $\{\widehat{\phi}_{nk}\}_{k=1}^K$ and $\{\phi_k\}_{k=1}^K$, respectively. Define the parameter spaces

$$\boldsymbol{\Theta}_{K_1,K_2} = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \mathbb{R}^q, \beta_1 \in \mathcal{H}_{0K_1}, \beta_2 \in \mathcal{H}_{0K_2}\}, \tag{3.3}$$

$$\widehat{\boldsymbol{\Theta}}_{K_1,K_2} = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \mathbb{R}^q, \beta_1 \in \widehat{\mathcal{H}}_{nK_1}, \beta_2 \in \widehat{\mathcal{H}}_{nK_2}\}.$$

8

As shown in Theorem 1 in Section 4.1, $\widehat{\Theta}_{K_1,K_2} \to \Theta_{K_1,K_2}$ almost surely, as $n \to \infty$. Since $\Theta_{K_1,K_2} \subset \Theta$, we also have $\widehat{\Theta}_{K_1,K_2} \subset \Theta$ almost surely, as $n \to \infty$. Let $\boldsymbol{\gamma}_1 = (\gamma_{1,1}, \ldots, \gamma_{1,K_1})^\top$ and $\boldsymbol{\gamma}_2 = (\gamma_{2,1}, \ldots, \gamma_{2,K_2})^\top$. With a slight abuse of notation, denote $\widetilde{\boldsymbol{\theta}} = (\boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_2^\top, \boldsymbol{\gamma}_1^\top, \boldsymbol{\gamma}_2^\top)^\top$, and the corresponding parameter space as $\widetilde{\Theta}_{K_1,K_2} = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \mathbb{R}^q, \boldsymbol{\gamma}_1 \in \mathbb{R}^{K_1}, \boldsymbol{\gamma}_2 \in \mathbb{R}^{K_2}\}$. Obviously, the closed linear span of $\widehat{\Theta}_{K_1,K_2}$ is equal to the closed linear span of $\widetilde{\Theta}_{K_1,K_2}$ almost surely, as $n \to \infty$; in the sense that, for each $\widehat{\boldsymbol{\theta}} \in \widehat{\Theta}_{K_1,K_2}$, there exists one and only one $\widetilde{\boldsymbol{\theta}} \in \widetilde{\Theta}_{K_1,K_2}$ such that $\widetilde{\boldsymbol{\theta}}(\widehat{\mathbf{W}}) = \widehat{\boldsymbol{\theta}}(\mathbf{W})$ almost surely. Actually, this is an isomorphism by Assumption (A8).

Combining the model (2.3) and the basis expansion with 2D-FPC basis in (3.1), we are able to obtain the estimates

$$\widehat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta} \in \widehat{\Theta}_{K_1,K_2}}{\arg\min} \; n^{-1} \sum_{i=1}^n \{Y_i - \widehat{\boldsymbol{\theta}}(\mathbf{W}_i)\}^2. \tag{3.4}$$

Consequently, the optimal decision rule is estimated as

$$\widehat{\pi} = \arg\max_A \widehat{Q}\{\widehat{\boldsymbol{\theta}}_n(\mathbf{W})\} \;\; = \;\; I\left(\left[\widehat{Q}\{\widehat{\boldsymbol{\theta}}_n(\mathbf{W})\}|(A=1) - \widehat{Q}\{\widehat{\boldsymbol{\theta}}_n(\mathbf{W})\}|(A=0)\right] > 0\right)$$

$$= \;\; I\left(\widehat{\alpha}_2 \mathbf{X} + \int_{\mathcal{V}} \widehat{\beta}_2(\boldsymbol{s}) Z(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}) > 0\right). \tag{3.5}$$

Combining all the aforementioned steps, the complete algorithm, which we refer to as the semiparametric functional learning algorithm, is given below in Algorithm 1.

Input   : $\{\mathbf{X}_i, Z_i(\boldsymbol{v}), A_i, Y_i\}_{i=1}^n$.
Output   : $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{\pi}$.

Step 1. Based on $\{Z_i(\boldsymbol{s}_j)\}_{i=1,j=1}^{n,N_s}$, construct 2D-FPC basis functions $\{\widehat{\phi}_{\ell,nk}^{\mathcal{S}}(\boldsymbol{s})\}_{k=1}^{K_\ell}$ by (2.6) with respect to covariance functions $V_{\ell,n}(\boldsymbol{s}, \boldsymbol{s}')$ defined in (3.2), $\ell = 1, 2$.
Step 2. Obtain the estimates $\widehat{\boldsymbol{\theta}}_n$ by (3.4).
Step 3. Obtain the estimated optimal regime $\widehat{\pi}$ through (3.5).

Algorithm 1: Semiparametric Functional Learning Algorithm.

---

4. Theoretical results

In this section, we establish the asymptotic properties of the bases $\{\widehat{\phi}_{nk}(\boldsymbol{s})\}_k$ and $\{\widehat{\phi}_{nk}^{\mathcal{S}}(\boldsymbol{s})\}_k$ proposed in Section 2, and the estimated $\widehat{\boldsymbol{\theta}}$ and estimated optimal regime $\widehat{\pi}$ proposed in Section 3. Our main results are summarized in Theorems 1 and 2.

4.1. Theoretical results for 2D-FPC basis

The main results are stated below. Theorem 1 states that as $n \to \infty$, the proposed bases $\{\widehat{\phi}_{nk}(\boldsymbol{s})\}_k$ converge to the theoretical ones up to sign almost surely, the projection of the closed linear span of the

proposed bases $\{\widehat{\phi}_{nk}(s)\}_k$ converge to the theoretical ones almost surely, the proposed "pre-smoothed" bases $\{\widehat{\phi}_{nk}^{\mathcal{S}}(s)\}_k$ are asymptotically equal to $\{\widehat{\phi}_{nk}(s)\}_k$ up to sign, and the eigenvalues $\{\widehat{\lambda}_{nk}\}_k$ converge to their theoretical values almost surely.

Theorem 1. *Under Assumption (A3), there exists a sequence $K_n \to \infty$ as $n \to \infty$, such that*

(i) *for each k, there exists a sign sequence $\{S_{nk}\}$ such that $\max_{1\leq k\leq K_n} \|S_{nk}\widehat{\phi}_{nk} - \phi_k\|_{\mu,2} \xrightarrow{\text{a.s.}} 0$;*

(ii) *for any $g_n \in \mathcal{H}_0$ with $\limsup_{n\to\infty} \|g_n\|_{\mu,2} < \infty$, $\|\widehat{\mathcal{H}}_{nK_n}g_n - \mathcal{H}_{0K_n}g_n\|_{\mu,2} \xrightarrow{\text{a.s.}} 0$ and $\|\widehat{\mathcal{H}}_{nK_n}^{\perp}g_n - \mathcal{H}_{0K_n}^{\perp}g_n\|_{\mu,2} \xrightarrow{\text{a.s.}} 0$;*

(iii) *$\{\widehat{\phi}_{nk}(s)\}_k$ and $\{\widehat{\phi}_{nk}^{\mathcal{S}}(s)\}_k$ are asymptotically equal up to sign;*

(iv) *$\max_k |\widehat{\lambda}_{nk} - \lambda_k| \xrightarrow{\text{a.s.}} 0$.*

Sketch Proof of Theorem 1. To show Theorem 1, we need to show the following conclusions in sequence:

(a) $\mathcal{F}_1$ and $\mathcal{F}_2$ are Glivenko-Cantelli classes;

(b) Recall that $\widehat{\phi}_{n1} = \arg\max_{f\in\mathcal{B}_1\cap\mathcal{S}} \int_{\mathcal{V}\times\mathcal{V}} f(s)f(s')V_n(s,s')\mathrm{d}\mu(s)\mathrm{d}\mu(s')$. Then $\exists$ a sign sequence $\{S_{n1} : \forall n, S_{n1} \in \{-1,1\}\}$ such that $\|S_{n1}\widehat{\phi}_{n1} - \phi_1\|_{\mu,2} \xrightarrow{\text{a.s.}} 0$, as $n \to \infty$.

(c) Suppose for some $1 \leq K < \infty$, $\exists$ sign sequences $\{S_{n1}, S_{n2}, \ldots, S_{nK} : S_{nk} \in \{-1,1\}, k = 1, \ldots, K\}_n$ such that $\max_{1\leq k\leq K} \|S_{nk}\widehat{\phi}_{nk} - \phi_k\|_{\mu,2} \xrightarrow{\text{a.s.}} 0$, as $n \to \infty$, where $\widehat{\phi}_{nk} \in \mathcal{B}_1 \cap \mathcal{S}$, $\forall 1 \leq k \leq K$ and $\{\widehat{\phi}_{n1}, \ldots, \widehat{\phi}_{nK}\}$ are orthogonal in $\mathcal{H}_0$. Let $\{g_n\} \in \mathcal{H}_0$ be a sequence satisfying $\limsup_{n\to\infty} \|g_n\|_{\mu,2} < \infty$, then both

$$\|\widehat{\mathcal{H}}_{nK}g_n - \mathcal{H}_{0K}g_n\|_{\mu,2} \xrightarrow{\text{a.s.}} 0 \quad \text{and} \quad \|\widehat{\mathcal{H}}_{nK}^{\perp}g_n - \mathcal{H}_{0K}^{\perp}g_n\|_{\mu,2} \xrightarrow{\text{a.s.}} 0, \quad \text{as } n \to \infty.$$

(d) Assume for some $1 \leq K < \infty$, that $\{\widehat{\phi}_{n1}, \ldots, \widehat{\phi}_{nK}\}$ form an orthonormal system on $\mathcal{S} \cap \mathcal{H}_0$ with $\max_{1\leq k\leq K} \|S_{nk}\widehat{\phi}_{nk} - \phi_k\|_{\mu,2} \xrightarrow{\text{a.s.}} 0$, as $n \to \infty$, for some sign sequences $\{S_{n1}, \ldots, S_{nK} : S_{nk} \in \{-1,1\}, k = 1, \ldots, K\}_n$. Let $\widehat{\phi}_{n(K+1)} = \arg\max_{f\in\mathcal{B}_1\cap\mathcal{S}\cap\widehat{\mathcal{H}}_{nK}^{\perp}} \int_{\mathcal{V}\times\mathcal{V}} f(s)f(s')V_n(s,s')\mathrm{d}\mu(s)\mathrm{d}\mu(s')$. Then $\exists$ a sign sequence $\{S_{n(K+1)} \in \{-1,1\}\}_n$ such that $\|S_{n(K+1)}\widehat{\phi}_{n(K+1)} - \phi_{(K+1)}\|_{\mu,2} \xrightarrow{\text{a.s.}} 0$, as $n \to \infty$.

The detailed proof and more technical details are given in Section A in supplemental materials. $\square$

4.2. Theoretical results for individualized treatment regimes estimation

Let $\theta_0 = (\alpha_{01}, \alpha_{02}, \beta_{01}, \beta_{02}) \in \Theta$ be the true parameter value. We require further the following Assumptions (A9)–(A10) to develop the consistency results of the estimates:

10

(A9) (Covariates) $\mathrm{E}\|\mathbf{X}\|^2 < \infty$. The eigenvalues of $\mathrm{E}(\mathbf{X}\mathbf{X}^\top)$ are bounded away from 0 and infinity.

(A10) (Coefficients) $\|\alpha_{01}\| < \infty$, $\|\alpha_{02}\| < \infty$, $\beta_{01}, \beta_{02} \in \mathcal{H}_0$, $\|\beta_{01}\|^2_{\mu,2} < \infty$ and $\|\beta_{02}\|^2_{\mu,2} < \infty$.

In what follows, P denotes taking the expectation over a single observation $(\mathbf{W}, Y)$. The following Theorem 2 (i) gives the consistency result for $\widehat{\boldsymbol{\theta}}_n$, and Theorem 2 (ii), which is similar to results found in Qian and Murphy (2011) but generalized to our setting, establishes asymptotic optimality of the treatment regime estimated from the forgoing regression model.

Theorem 2. *Under Assumptions (A5) – (A10), as $n \to \infty$,*

(i) $P\left\{\widehat{\boldsymbol{\theta}}_n(\mathbf{W}) - \boldsymbol{\theta}_0(\mathbf{W})\right\}^2 \xrightarrow{\text{a.s.}} 0$;

(ii) $V(\pi^{opt}) - V(\widehat{\pi}) \xrightarrow{\text{a.s.}} 0$.

Sketch Proof of Theorem 2. The conclusion in (i) follows from the facts of the convergence of bases, the negligibility of the basis approximation error caused by the basis cutoff, and the fact that $\limsup \|\widehat{\boldsymbol{\theta}}_n\|_*$ is bounded. The conclusion in (ii) now follows from Qian and Murphy (2011) directly. See the detailed proof and more technical details in Section B in supplemental materials. □

5. Implementation

This section outlines the practical implementation of the proposed procedure. We begin by discussing the selection criteria for bases numbers, followed by the implementation details of BST.

5.1. Selection Criteria for Bases Numbers

Selecting the leading principal components (PC) is a popular topic in FDA, particularly for functional data with univariate indexes. To this end, two methods have gained favor in the literature: ranking PC based on the percentage of variance explained (PVE; Kong et al., 2016), and the percentage of association–variation explained (PAVE; Su et al., 2017), due to their high testing power and computational efficiency. In this study, we adopt both PVE and PAVE criteria to select the leading PC bases, similar to the univariate functional linear model setting. Specifically, for a given threshold α, we define the eigenvalues $\widehat{\lambda}_{nk}$, $k = 1, \ldots, K_n$ as in (2.6). Then

- PVE selects the number of bases K_n such that

$$\frac{\sum_{k=1}^{K_n} \widehat{\lambda}_{nk}}{\sum_{k=1}^{\infty} \widehat{\lambda}_{nk}} \geq \alpha \quad \text{and} \quad \frac{\sum_{k=1}^{K_n-1} \widehat{\lambda}_{nk}}{\sum_{k=1}^{\infty} \widehat{\lambda}_{nk}} < \alpha;$$

- PAVE selects the number of bases K_n such that

$$\frac{\sum_{k=1}^{K_n} \widehat{\lambda}_{nk}\widehat{\gamma}_k^2}{\sum_{k=1}^{K_n^*} \widehat{\lambda}_{nk}\widehat{\gamma}_k^2} \geq \alpha \quad \text{and} \quad \frac{\sum_{k=1}^{K_n-1} \widehat{\lambda}_{nk}\widehat{\gamma}_k^2}{\sum_{k=1}^{K_n^*} \widehat{\lambda}_{nk}\widehat{\gamma}_k^2} < \alpha,$$

11

where K_n^* is selected by pre-fitting the truncated model with a high threshold of PVE, i.e., somewhere in the range $[0.95, 0.99]$, and where $\{\widehat{\gamma}_k\}_{k=1}^{K_n^*}$ are the corresponding PC basis coefficients.

The threshold α is usually chosen to be within $[0.95, 0.99]$ in practice. In our work, we set $\alpha = 0.99$ for PVE, PAVE, and the threshold of PVE in the pre-fitting step for PAVE.

In the following, we adopt both PVE and PAVE as selection criteria for bases numbers, denoting the results obtained by the selection criteria PVE and PAVE with the superscripts "PVE" and "PAVE", respectively.

5.2. Construction of bivariate spline bases over triangulation

The proposed 2D-FPC bases are constructed using the BST as the initial basis. The construction of BST involves the input of parameters for triangulation and spline bases. To investigate the impact of BST on the performance of the proposed 2D-FPC bases, we conducted the simulation study in various settings, such as coarse or fine grids for triangulation, and different smoothness conditions for the splines, with larger degree $d = 5$ or smaller degree $d = 3$. The simulation results indicate that the proposed method is robust regarding the initial bases. We briefly discuss the implementation of BST in this section.

Triangulation. Optimal triangulation involves determining both the number and shape of the triangles. According to the literature (Li et al., 2021; Mu et al., 2018; Yu et al., 2021), BST performs consistently well when an adequate number of triangles is used. In practice, it is suggested that various numbers of triangles with coarse or fine grids can be attempted to select the optimal number of triangles. Once the number of triangles is determined, triangulated meshes can be constructed using typical triangulation methods, such as Delaunay Triangulation (Bern and Eppstein, 1995), which is implemented in the R package Triangulation (Wang and Lai, 2019).

BST parameters. BST with a higher degree d is expected to provide a more accurate approximation and requires more computational power. Throughout, we use $(d, r) = (5, 1)$ for BST, as it attains full approximation power asymptotically (Lai and Schumaker, 2007). After the triangulation is constructed, BST can be generated using the R package BPST (Wang et al., 2019).

6. Empirical study

In this section, we investigate the finite sample performance of the proposed method. The experimental data is generated from the underlying model:

$$Y_i = \boldsymbol{\alpha}_1^\top \mathbf{X}_i + \int_{\mathcal{V}} \beta_1(\boldsymbol{s}) Z_i(\boldsymbol{s}) \mathrm{d}\boldsymbol{s} + A_i \left\{ \boldsymbol{\alpha}_2^\top \mathbf{X}_i + \int_{\mathcal{V}} \beta_2(\boldsymbol{s}) Z_i(\boldsymbol{s}) \mathrm{d}\boldsymbol{s} \right\} + \varepsilon_i, \ \ i = 1, \ldots, n.$$

For each subject i, we generate $\mathbf{X}_i \in \mathbb{R}^q$ independently from $\mathrm{MVN}(\mathbf{0}_q, \boldsymbol{\Omega}_q(r))$, where $\{\boldsymbol{\Omega}_q(r)\}_{\ell, \ell'} = r^{|\ell - \ell'|}$, and the dependence structure in $\mathbf{X}_i$ is indexed by the autocorrelation parameter r; the error

ε_i is independently generated from $N(0,1)$. To simulate the within-image dependence, we generate the imaging data $Z_i(s) = \sum_{k=1}^{2} \zeta_{ik} Z_k^{\zeta}(s)$ at a grid of $N_s = 40 \times 40$ pixels, where $(\zeta_{i1}, \zeta_{i2})^{\top} \sim$ MVN$(\mathbf{0}_2, \mathbf{I}_2)$, and $Z_1^{\zeta}(\cdot)$ and $Z_2^{\zeta}(\cdot)$ are quadratic and exponential functions, respectively, with forms $Z_1^{\zeta}(s) = 20\{(s_1 - 0.5)^2 + (s_2 - 0.5)^2\}$ and $Z_2^{\zeta}(s) = \exp[-15\{(s_1 - 0.5)^2 + (s_2 - 0.5)^2\}]$. The contour plots of $Z_1^{\zeta}(\cdot)$ and $Z_2^{\zeta}(\cdot)$ are illustrated in Figure 6.1, left panel. Throughout, we take $q = 5$,
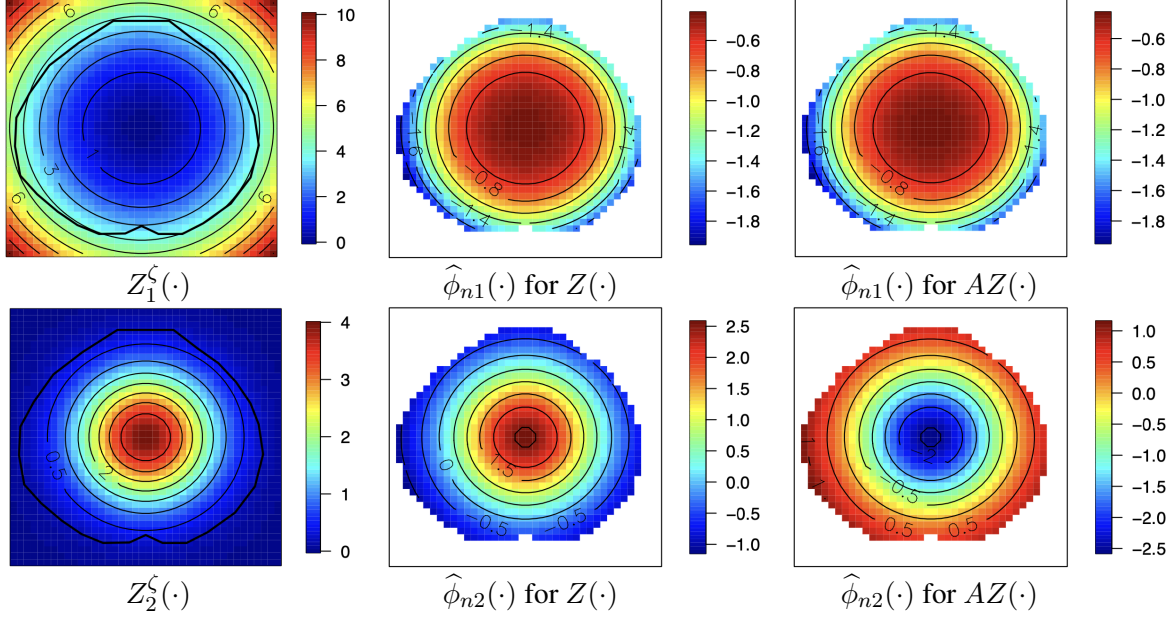
Figure 6.1: Contour plots of $Z_j^{\zeta}(\cdot)$ (left panel) and the corresponding bases functions $\widehat{\phi}_{nj}(\cdot)$ for $Z(\cdot)$ (middle panel) and $AZ(\cdot)$ (right panel), $j = 1, 2$.

$\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \mathbf{1}_5^{\top}$, and the coefficient maps $\beta_1(s) = \beta_2(s) = 1$, and we consider the autocorrelation $r = 0$ and $r = 0.5$, which corresponds to the independent and dependent covariate structures, respectively.

We evaluate the methods on the accuracy of the estimated ITR with respect to the average marginal mean outcome $V(\widehat{\pi}^{\mathrm{PVE}})$ and $V(\widehat{\pi}^{\mathrm{PAVE}})$, for PVE and PAVE criteria, respectively. For comparison, we also compute the average optimal marginal mean outcome $V(\pi^{opt})$, which works as a benchmark and is only computable in simulations. We also evaluate the methods on the accuracy of coefficient estimation by computing the mean squared errors (MSEs). All the results are based on 100 Monte Carlo replications.

As illustrated in Figure 6.1, the generated bases functions are the normalized true function up to a sign difference, which validates the conclusion in Theorem 1. Table 6.1 presents the marginal mean outcome for the proposed methods. In all settings, the average marginal mean outcomes of PVE and PAVE criteria perform similarly, and both are close to those with optimal treatment regime, π^{opt}. Table 6.2 shows the MSEs for linear coefficients $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$. With the increase of sample size, the MSEs of all estimates decrease significantly. And the estimates with bases selected by PVE and PAVE perform

similarly, regardless of the settings.

Table 6.1: Estimated mean outcome for optimal π and proposed method with PVE and PAVE criteria.

	$r = 0$			$r = 0.5$		
	$n = 100$	$n = 200$	$n = 500$	$n = 100$	$n = 200$	$n = 500$
$V(\widehat{\pi}^{opt})$	2.791	2.814	2.798	3.433	3.442	3.456
$V(\widehat{\pi}^{\mathrm{PVE}})$	2.763	2.808	2.804	3.431	3.455	3.459
$V(\widehat{\pi}^{\mathrm{PAVE}})$	2.763	2.808	2.804	3.431	3.455	3.459

Table 6.2: MSEs for linear coefficients $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$.

r	n	Criteria	MSE of $\boldsymbol{\alpha}_1$ ($\times 10^{-2}$)					MSE of $\boldsymbol{\alpha}_2$ ($\times 10^{-2}$)				
			α_{11}	α_{12}	α_{13}	α_{14}	α_{15}	α_{21}	α_{22}	α_{23}	α_{24}	α_{25}
0	100	PVE	2.53	1.29	1.30	1.19	1.42	6.85	1.08	1.35	1.31	1.47
		PAVE	2.53	1.29	1.30	1.19	1.42	6.85	1.08	1.35	1.31	1.47
	200	PVE	1.13	0.65	0.68	0.47	0.60	3.73	0.54	0.54	0.58	0.61
		PAVE	1.13	0.65	0.68	0.47	0.60	3.73	0.54	0.54	0.58	0.61
	500	PVE	0.46	0.19	0.20	0.22	0.22	1.48	0.23	0.21	0.17	0.20
		PAVE	0.46	0.19	0.20	0.22	0.22	1.48	0.23	0.21	0.17	0.20
0.5	100	PVE	2.53	2.01	2.56	1.89	1.27	6.85	1.71	2.01	1.79	1.89
		PAVE	2.53	2.01	2.56	1.89	1.27	6.85	1.70	2.01	1.79	1.89
	200	PVE	1.13	0.67	1.03	1.03	1.02	3.73	0.67	0.78	1.08	0.78
		PAVE	1.13	0.67	1.03	1.03	1.02	3.73	0.67	0.78	1.08	0.78
	500	PVE	0.46	0.29	0.31	0.34	0.25	1.48	0.28	0.32	0.39	0.27
		PAVE	0.46	0.29	0.31	0.34	0.25	1.48	0.28	0.32	0.39	0.27

7. Application to ADNI data

The data that drives our research comes from the large neuroimaging datasets in the Alzheimer's Disease Neuroimaging Initiative (ADNI, `http://adni.loni.usc.edu`). The longitudinal cohort study in ADNI, which has gone through three phases including ADNI1, ADNI GO, and ADNI2, is a comprehensive neuroimaging study that collected a variety of necessary phenotypic measures, including structural, functional, and molecular neuroimaging, biomarkers, clinical and neuropsychological variables, and genomic information (Petersen et al., 2010; Weiner and Veitch, 2015). These data provide unprecedented resources for statistical methods development and scientific discovery.

We now analyze the records from 441 participants through the ADNI1 and ADNI GO phases. The data contains the following variables:

- Mini-mental state examination (MMSE) scores: response variable, ranging from 15 to 30, where lower values indicate a more severe AD status.

- Fludeoxyglucose positron emission tomography (PET) scans: neuroimaging representing brain metabolism activity level and can be used to make early diagnoses of AD, with 79×95 pixels, with the measurements ranging from 0.013 to 2.149. The left panel in Figure 7.3 shows the PET images for four randomly selected subjects.

- Age: the participants' ages, ranging from 55 to 89 years.

- Education: the participants' educational status, ranging from 4 to 20 years.

- Gender: the participants' gender, with 169 female and 278 male. We created a dummy variable with value 1 representing female and 0 for male.

- Ethnicity: the participants' ethnic categories, with 12 Hispanic/Latino, 429 not Hispanic/Latino, and 6 unknown. We created a dummy variable with value 1 representing Hispanic/Latino and 0 for others.

- Race: the participants' racial categories, with 1 Indian/Alaskan, 7 Asians, 24 Blacks, 413 Whites, and 2 more than one category. We created a dummy variable with value 1 representing white and 0 for others.

- Marriage: the participants' marital status, with 35 divorced, 344 married, 12 never married, and 56 widowed. We created a dummy variable with value 1 representing married and 0 for others.

- Apolipoprotein (APOE) gene: the number of copies of APOE4 gene, the most prevalent genetic risk factor for AD (Ashford and Mortimer, 2002), ranging from 0 to 2. We created two dummy variables, APOE1 and APOE2, to denote those with one and two copies of APOE4 gene, respectively.

- Treatment: During the ADNI1 and ADNI GO study periods, the US FDA-approved therapies for AD symptoms included cholinesterase inhibitors and the NMDA-partial receptor antagonist memantine. Cholinesterase inhibitors, including donepezil, galantamine, and rivastigmine, are prescribed for mild-to-moderate-stage AD. Memantine is prescribed for the treatment of AD either as monotherapy or in combination with one of the cholinesterase inhibitors for moderate-to-severe stage AD (Schneider et al., 2011). We denote by $A = 1$ those participants taking one or more combinations of Donepezil (Aricept), Galantamine (Razadyne), Rivastigmine (Exelon), and Memantine (Namenda), while we use $A = 0$ for those wihout concurrent medical records, or taking some other treatments or supplements. The distribution of participants with $A = 1$ is illstruated in Figure 7.1.

We apply the proposed method to those data using both PVE and PAVE criteria. The estimated coefficient map for $\beta_1(\cdot)$ and $\beta_2(\cdot)$ are shown in the top line of Figure 7.2, in which PVE and PAVE
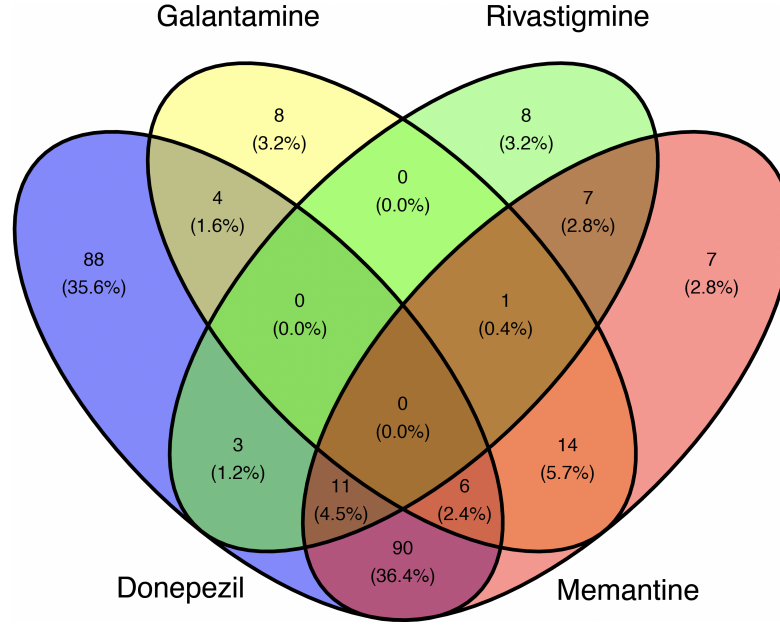
Figure 7.1: Distribution of patients by treatments (total sample size 441, with 247 $A = 1$).

criteria produce similar results, and the estimates of $\beta_2(\cdot)$ illustrate brain structures. We also display visually the three leading principal component basis maps in the middle and bottom lines of Figure 7.2. All these estimated basis maps illustrate brain structures.

Tables 7.1 presents the estimated coefficients for the nonfunctional predictors, along with the corresponding 95% bootstrap confidence intervals. The main effect of the treatment can improve the performance in MMSE by around $4 \sim 6$ units on average. With the increase of age, the general MMSE scores will decrease; and with the increase of education level, the MMSE scores also increase in general: both are supported by studies on cognitive reserve in aging and AD (Fratiglioni et al., 2007; Stern, 2012). As for the well-known risk genetic factor APOE gene, with more copies of epsilon 4 alleles in the APOE gene, the MMSE scores decrease, which means a higher risk for the onset of AD and agrees with other current studies (Schneider et al., 2011). The difference between females and males is not significant.

When one patient enters into the database with the imaging feature and other features, we can utilize the estimated coefficients to provide the optimal treatment regime based on Equation (3.5). To be specific, for the imaging feature, for Subjects 27, 48, 55, and 160, the process can be visualized as depicted in the following Figure 7.3. The estimated optimal ITRs for these four subjects are $A = 1$, $A = 0$, $A = 0$, and $A = 1$, respectively. The differences in the images are subtle, note, for example, that the blue and orange parts are a little darker for those assigned $A = 0$ versus those assigned $A = 1$.
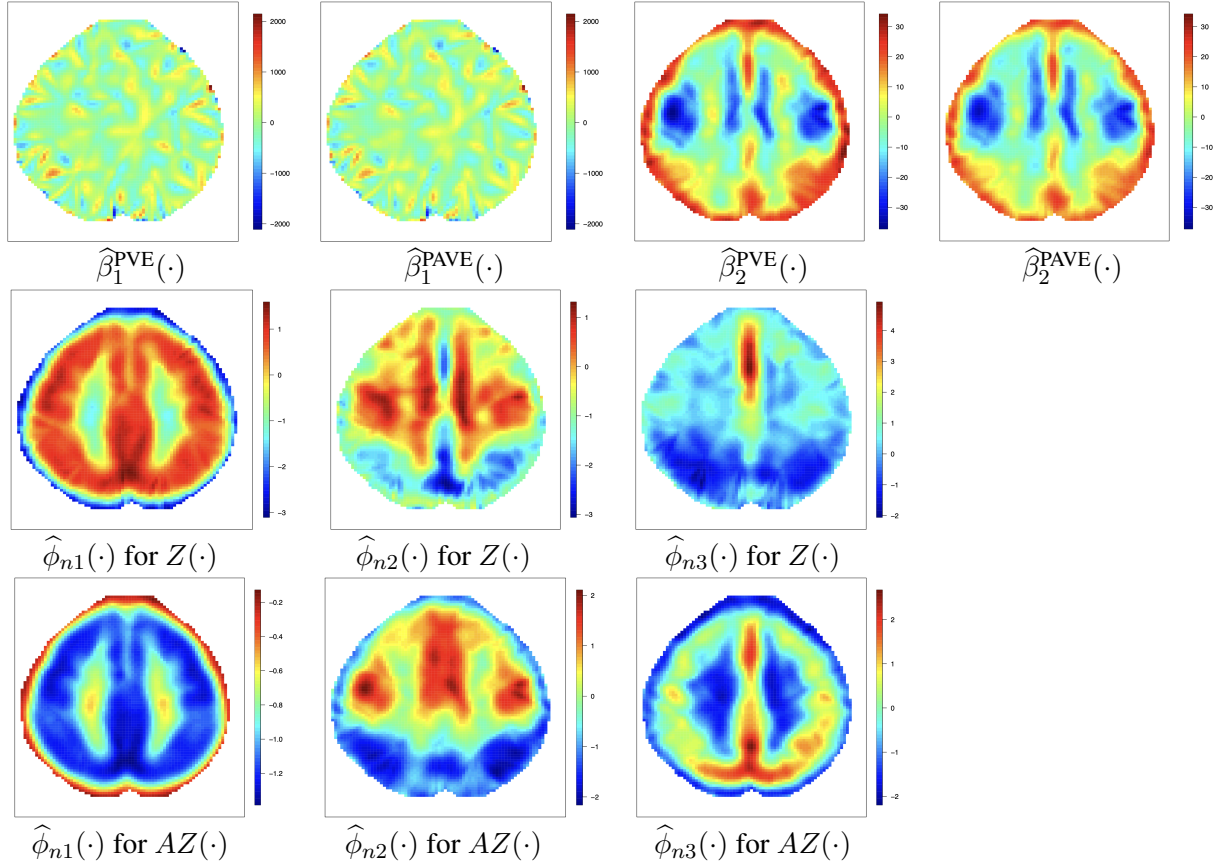
Figure 7.2: Top: Estimated coefficient maps for $\beta_1(\cdot)$ and $\beta_2(\cdot)$ using PVE and PAVE criteria, respectively. Middle: The three leading PC basis maps (top three ranked by both PVE and PAVE) for $Z(\cdot)$. Bottom: The three leading PC basis maps (top three ranked by both PVE and PAVE) for $AZ(\cdot)$.

Table 7.1: Estimated coefficients and 95% bootstrap confidence intervals for the linear covariates using PVE and PAVE criteria, respectively.

	PVE		PAVE	
Term	Estimate	95% Bootstrap CI	Estimate	95% Bootstrap CI
Intercept	20.0650	(8.655, 31.93)	19.8176	(8.663, 31.562)
Age	-0.0560	(-0.113, -0.004)	-0.0568	(-0.111, -0.005)
Education	0.2225	(0.137, 0.340)	0.2196	(0.137, 0.334)
Gender	-0.1998	(-0.909, 0.707)	-0.1811	(-0.896, 0.643)
APOE1	-0.3260	(-1.079, 0.318)	-0.3201	(-1.080, 0.272)
APOE2	-0.9185	(-2.059, 0.321)	-0.9003	(-2.022, 0.333)
Ethinicity	0.2363	(-1.513, 2.580)	0.2427	(-1.536, 2.451)
Race	1.2455	(0.256, 2.664)	1.3468	(0.312, 2.527)
Marriage	0.1760	(-0.711, 0.837)	0.1451	(-0.703, 0.801)
Treatment	4.1362	(-6.893, 15.867)	6.0930	(-6.759, 12.850)
Treatment×Age	-0.0594	(-0.107, -0.006)	-0.0643	(-0.107, -0.014)
Treatment×Education	0.0001	(-0.104, 0.076)	0.0056	(-0.102, 0.072)
Treatment×Gender	-0.0302	(-0.711, 0.704)	-0.0279	(-0.620, 0.710)
Treatment×APOE1	0.4237	(-0.312, 1.089)	0.4050	(-0.298, 1.036)
Treatment×APOE2	-0.7102	(-1.767, 0.576)	-0.8079	(-1.764, 0.508)
Treatment×Ethinicity	-0.0115	(-2.067, 2.216)	0.0661	(-1.957, 2.263)
Treatment×Race	1.3543	(0.368, 2.536)	1.3859	(0.381, 2.477)
Treatment×Marriage	-0.2662	(-0.998, 0.552)	-0.3088	(-0.963, 0.524)
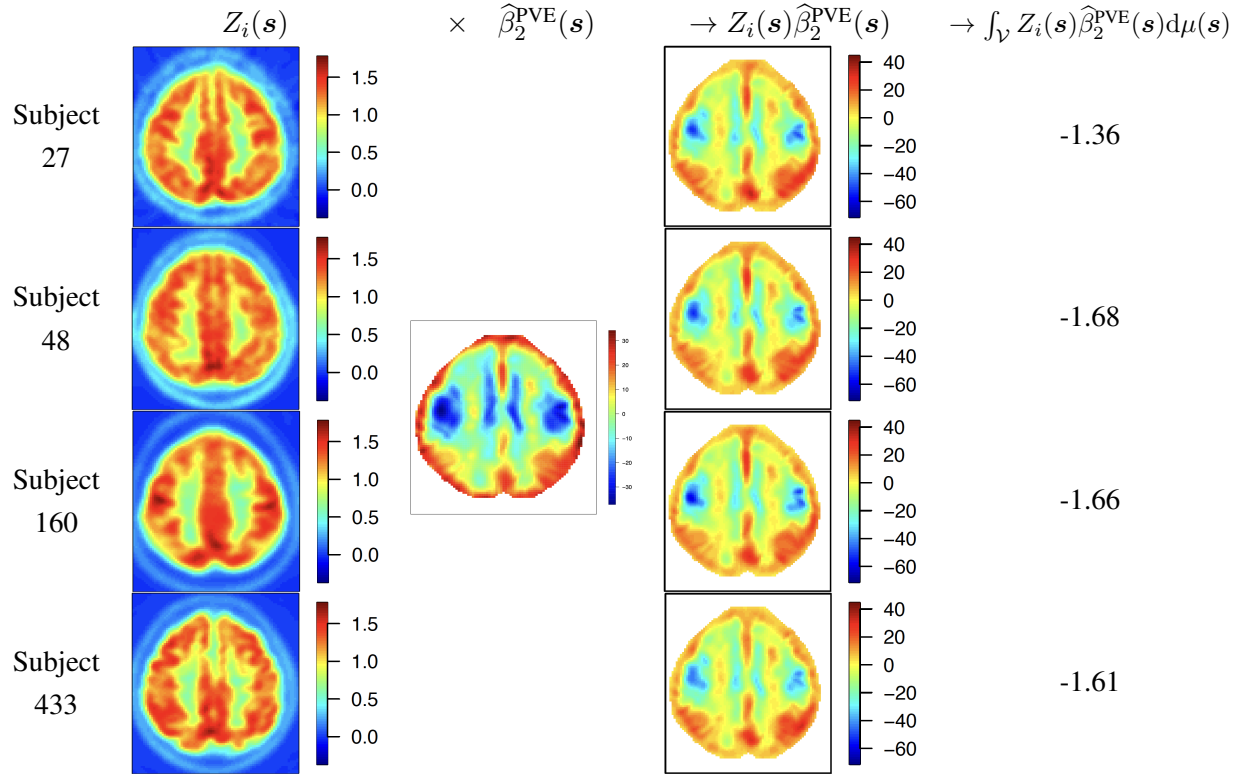
Figure 7.3: An illustration of procedures for dealing with imaging features to obtain ITRs. The left panel represents the imaging features of four randomly selected subjects, Subjects 27, 48, 160, and 433, with MMSE scores of 30, 27, 15, and 23, respectively. The multiplication of the imaging feature $Z_i(s)$ and the coefficient map $\widehat{\beta}_2^{\text{PVE}}(s)$, as shown in the third column, represents the data-driven summary of the imaging features, which differs in the activity level in the blue region and orange region in the bottom. The integral of the multiplication gives the image contribution of the interaction term; adding $\mathbf{X}\widehat{\alpha}$ and taking the positivity indicator will result in the optimal ITR.

8. Discussion

We proposed semiparametric functional learning with imaging features to estimate the optimal ITRs. Overall, the proposed approach can efficiently and precisely estimate the optimal treatment regime with abundant features. It can overcome the "leakage" problem in handling the complex domain of imaging data and is also computationally efficient. This approach provides an efficient and powerful tool to estimate optimal treatment regimes by incorporating abundant features into the precision medicine framework.

There are some natural extensions based on the proposed work. One extension is to extend the current regression model to a generalized regression model, in which the response is allowed to be discrete or categorical. For example, for the study of Alzheimer's disease, one could take the disease stage as the response. Another extension could be to generalize the current single-decision setting to the multi-stage decision setting, which is more common in chronic disease settings such as Alzheimer's disease.

Supplemental Materials for **"Functional Individualized Treatment Regimes with Imaging Features"**

In this document, we provide the technical details of the conclusions presented in the main paper. Specifically, we give detailed proofs of Theorems 1 and 2 in the paper, and present a number of technical lemmas and other supporting results used in the proofs.

A. Proof of 2D-FPC Bases Properties

We first develop the properties of the proposed 2D-FPC bases. We generally follow the sketch proof of Theorem 1 in Section 4.1 in the paper, and provide a detailed proof for each conclusion, with corresponding supporting lemmas.

Lemma A.1. *For any $b(\boldsymbol{s}) \in \mathcal{H}_0^{\perp}$, $\mathrm{Var}\{\int_{\mathcal{V}} b(\boldsymbol{s}) Z(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s})\} = 0$.*

Proof. For any $b(\boldsymbol{s}) \in \mathcal{H}_0^{\perp}$,

$$\mathrm{Var}\left\{ \int_{\mathcal{V}} b(\boldsymbol{s}) Z(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}) \right\} = \mathrm{E}\left[\int_{\mathcal{V}} b(\boldsymbol{s})\{Z(\boldsymbol{s}) - \mathrm{E}Z(\boldsymbol{s})\} \mathrm{d}\mu(\boldsymbol{s}) \right]^2$$

$$= \mathrm{E}\left\{ \sum_{k=1}^{\infty} \xi_k \int_{\mathcal{V}} b(\boldsymbol{s}) \phi_k(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}) \right\}^2 = 0.$$

$\square$

Remark A.1. *From Lemma A.1, for any linear functional $\int_{\mathcal{V}} b(\boldsymbol{s}) Z(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s})$, we can assume without loss of generality that $b \in \mathcal{H}_0$.*

In the following, we denote the theoretical and empirical expectation operators as P and $\mathbb{P}_n$, respectively. The following Theorem A.1 shows that both $\mathcal{F}_1$ and $\mathcal{F}_2$ are Glivenko-Cantelli classes, which can be further applied for consistency proofs.

Theorem A.1. *Suppose Assumption (A3) holds. Then the following two conclusions hold:*

(i) $\mathcal{F}_1$ is a Glivenko-Cantelli class with envelope

$$F_1(Z) = \left(\sum_{k=1}^{\infty} \left[\int_{\mathcal{V}} \phi_k(\boldsymbol{s}) \left\{ Z(\boldsymbol{s}) - \mathrm{E}Z(\boldsymbol{s}) \right\} \mathrm{d}\mu(\boldsymbol{s}) \right]^2 \right)^{1/2}, \tag{A.1}$$

such that $F_1 \leq \|Z - \mathrm{E}Z\|_{\mu,2}$ and $\mathrm{E}F_1^2 < \infty$.

(ii) $\mathcal{F}_2$ is also a P-Glivenko-Cantelli class with envelope F_1^2, where $\mathrm{E}F_1^2 < \infty$.

Proof. (i) By Assumption (A3), fix $\varepsilon > 0$, $\exists K < \infty$, such that $\sum_{k=K+1}^{\infty} \lambda_k \leq \varepsilon^2$. Since $\mathcal{B}_1 \subset \mathcal{H}_0$, $b \in \mathcal{B}_1$ has a unique Fourier representation $b(\boldsymbol{s}) = \sum_{k=1}^{\infty} a_k \phi_k(\boldsymbol{s})$, where

21

$\int_{\mathcal{V}} b^2(s)\mathrm{d}\mu(s) = \sum_{k=1}^{\infty} a_k^2 \leq 1$ by definition of $\mathcal{B}_1$ and projection of a basis. Then, letting $\ell_2 = \{a : \|a\|_2 \leq \infty\}$, we can write $\mathcal{F}_1$ as

$$\mathcal{F}_1 = \left\{ \sum_{k=1}^{\infty} a_k \int_{\mathcal{V}} \phi_k(s)\{Z(s) - \mathrm{E}Z(s)\}\mathrm{d}\mu(s) : a \in \ell_2 \text{ and } \|a\|_2 \leq 1 \right\}.$$

Let $\mathcal{A}_1 = \{a \in \ell_2 \text{ and } \|a\|_2 \leq 1\}$ and recall $\bar{Z}_n(s) = n^{-1}\sum_{i=1}^n Z_i(s)$. Thus, $\forall f \in \mathcal{F}_1$,

$$\sup_{f \in \mathcal{F}_1} \|\mathbb{P}_n f - Pf\| = \sup_{a \in \mathcal{A}_1} \left| \sum_{k=1}^{\infty} a_k \int_{\mathcal{V}} \phi_k(s)\{\bar{Z}_n(s) - \mathrm{E}Z(s)\}\mathrm{d}\mu(s) \right|$$

$$\leq \sup_{a \in \mathcal{A}_1} \left| \sum_{k=K+1}^{\infty} a_k \int_{\mathcal{V}} \phi_k(s)\{\bar{Z}_n(s) - \mathrm{E}Z(s)\}\mathrm{d}\mu(s) \right| + \sup_{a \in \mathcal{A}_1} \left| \sum_{k=1}^{K} a_k \int_{\mathcal{V}} \phi_k(s)\{\bar{Z}_n(s) - \mathrm{E}Z(s)\}\mathrm{d}\mu(s) \right|$$

$$\equiv A_n + B_n.$$

By the Cauchy-Schwarz inequality,

$$A_n = \sup_{a \in \mathcal{A}_1} \left| n^{-1}\sum_{i=1}^n \sum_{k=K+1}^{\infty} a_k \int_{\mathcal{V}} \phi_k(s)\{Z_i(s) - \mathrm{E}Z(s)\}\mathrm{d}\mu(s) \right|$$

$$\leq \sup_{a \in \mathcal{A}_1} \left| n^{-1}\sum_{i=1}^n \left(\sum_{k=K+1}^{\infty} a_k^2 \right)^{1/2} \left(\sum_{k=K+1}^{\infty} \left[\int_{\mathcal{V}} \phi_k(s)\{Z_i(s) - \mathrm{E}Z(s)\}\mathrm{d}\mu(s) \right]^2 \right)^{1/2} \right|$$

$$\leq n^{-1}\sum_{i=1}^n \left(\sum_{k=K+1}^{\infty} \left[\int_{\mathcal{V}} \phi_k(s)\{Z_i(s) - \mathrm{E}Z(s)\}\mathrm{d}\mu(s) \right]^2 \right)^{1/2}.$$

Let $U_i = (\sum_{k=K+1}^{\infty}[\int_{\mathcal{V}} \phi_k(s)\{Z_i(s) - \mathrm{E}Z(s)\}\mathrm{d}\mu(s)]^2)^{1/2}$. Note that $(\mathrm{E}U_i)^2 \leq \mathrm{E}U_i^2 = \sum_{k=K+1}^{\infty} \lambda_k$. Since this is an iid sum, we have that $\limsup_{n\to\infty} A_n = \limsup_{n\to\infty} n^{-1}\sum_{i=1}^n U_i \leq \varepsilon$ almost surely. Let $\mathcal{A}_2 = \{a \in \mathbb{R}^K \text{ and } \|a\|_2 \leq 1\}$. Then

$$B_n = \sup_{a \in \mathcal{A}_2} \left| \sum_{k=1}^{K} a_k \int_{\mathcal{V}} \phi_k(s)\{\bar{Z}_n(s) - \mathrm{E}Z(s)\}\mathrm{d}\mu(s) \right|.$$

Since $\mathcal{A}_2$ is a compact ball, $\exists$ finite subset $\mathcal{M}_2 \subset \mathcal{A}_2$, such that $\sup_{a \in \mathcal{A}_2} \inf_{\tilde{a} \in \mathcal{M}_2} \|a - \tilde{a}\| \leq \varepsilon$. Accordingly,

$$B_n \leq \max_{\tilde{a} \in \mathcal{M}_2} \left| \sum_{k=1}^{K} \tilde{a}_k \int_{\mathcal{V}} \phi_k(s)\{\bar{Z}_n(s) - \mathrm{E}Z(s)\}\mathrm{d}\mu(s) \right|$$

$$+ \sup_{a \in \mathcal{A}_2, \tilde{a} \in \mathcal{M}_2 : \|a - \tilde{a}\| \leq \varepsilon} \left| \sum_{k=1}^{K} (a_k - \tilde{a}_k) \int_{\mathcal{V}} \phi_k(s)\{\bar{Z}_n(s) - \mathrm{E}Z(s)\}\mathrm{d}\mu(s) \right|$$

$$\equiv B_{1n} + B_{2n}.$$

By the standard strong law of large numbers, $B_{1n} \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$. As for B_{2n}, by the Cauchy-Schwarz inequality,

$$\sup_{\boldsymbol{a} \in \mathcal{A}_2, \widetilde{\boldsymbol{a}} \in \mathcal{M}_2 : \|\boldsymbol{a} - \widetilde{\boldsymbol{a}}\| \leq \varepsilon} \left| \sum_{k=1}^{K} (a_k - \widetilde{a}_k) \int_{\mathcal{V}} \phi_k(\boldsymbol{s}) \{ \bar{Z}_n(\boldsymbol{s}) - \mathrm{E}Z(\boldsymbol{s}) \} \mathrm{d}\mu(\boldsymbol{s}) \right|$$

$$= \sup_{\boldsymbol{a} \in \mathcal{A}_2, \widetilde{\boldsymbol{a}} \in \mathcal{M}_2 : \|\boldsymbol{a} - \widetilde{\boldsymbol{a}}\| \leq \varepsilon} \left| n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} (a_k - \widetilde{a}_k) \int_{\mathcal{V}} \phi_k(\boldsymbol{s}) \{ Z_i(\boldsymbol{s}) - \mathrm{E}Z(\boldsymbol{s}) \} \mathrm{d}\mu(\boldsymbol{s}) \right|$$

$$\leq \varepsilon \left| n^{-1} \sum_{i=1}^{n} \left(\sum_{k=1}^{K} \left[\int_{\mathcal{V}} \phi_k(\boldsymbol{s}) \{ Z_i(\boldsymbol{s}) - \mathrm{E}Z(\boldsymbol{s}) \} \mathrm{d}\mu(\boldsymbol{s}) \right]^2 \right)^{1/2} \right|$$

$$\leq \varepsilon \left| n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{\infty} \left[\int_{\mathcal{V}} \phi_k(\boldsymbol{s}) \{ Z_i(\boldsymbol{s}) - \mathrm{E}Z(\boldsymbol{s}) \} \mathrm{d}\mu(\boldsymbol{s}) \right]^2 \right|^{1/2}.$$

Let $V_i = \sum_{k=1}^{\infty} \left[\int_{\mathcal{V}} \phi_k(\boldsymbol{s}) \{ Z_i(\boldsymbol{s}) - \mathrm{E}Z(\boldsymbol{s}) \} \mathrm{d}\mu(\boldsymbol{s}) \right]^2$. Note that V_i does not depend on ε, and

$$\mathrm{E}V_i = \sum_{k=1}^{\infty} \lambda_k < \infty \Rightarrow \limsup_{n \to \infty} B_{2n} \leq \varepsilon C_0 \ \text{ a.s.,}$$

where $C_0 = |\sum_{k=1}^{\infty} \lambda_k|^{1/2}$ does not depend on ε. Therefore, $\sup_{f \in \mathcal{F}_1} \|\mathbb{P}_n f - Pf\| \xrightarrow{\text{a.s.}} 0$, thus, $\mathcal{F}_1$ is a P-Glivenko-Cantelli class by definition. Also note that $\mathcal{F}_1$ is enveloped by

$$F_1(Z) = \left(\sum_{k=1}^{\infty} \left[\int_{\mathcal{V}} \phi_k(\boldsymbol{s}) \{ Z(\boldsymbol{s}) - \mathrm{E}Z(\boldsymbol{s}) \} \mathrm{d}\mu(\boldsymbol{s}) \right]^2 \right)^{1/2}$$

$$\leq \left(\sum_{k=1}^{\infty} \left\{ \int_{\mathcal{V}} \phi_k^2(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}) \right\} \left[\int_{\mathcal{V}} \{ Z(\boldsymbol{s}) - \mathrm{E}Z(\boldsymbol{s}) \}^2 \mathrm{d}\mu(\boldsymbol{s}) \right] \right)^{1/2} = \|Z - \mathrm{E}Z\|_{\mu,2},$$

and

$$\mathrm{E}F_1^2(Z) \leq \mathrm{E} \left(\sum_{k=1}^{\infty} \left[\int_{\mathcal{V}} \phi_k(\boldsymbol{s}) \{ Z_i(\boldsymbol{s}) - \mathrm{E}Z(\boldsymbol{s}) \} \mathrm{d}\mu(\boldsymbol{s}) \right]^2 \right) = \sum_{k=1}^{\infty} \lambda_k < \infty.$$

(ii) By (i) and Corollary 9.27 of Kosorok (2008), we can conclude that $\mathcal{F}_2$ is also a P-Glivenko-Cantelli class, with envelope F_1^2.

$\square$

Given the conclusions in Theorem A.1, we can therefore develop the convergence results for the bases $\{\widehat{\phi}_{nk}\}_k$ through the following Theorems A.2 – A.4.

Theorem A.2. *Let* $\widehat{\phi}_{n1} = \arg\max_{f \in \mathcal{B}_1 \cap \mathcal{S}} \int_{\mathcal{V} \times \mathcal{V}} f(\boldsymbol{s})f(\boldsymbol{s}')V_n(\boldsymbol{s}, \boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}')$. *Suppose Assumption (A3) holds. Then* $\exists$ *a sign sequence* $\{S_{n1} : \forall n, S_{n1} \in \{-1, 1\}\}$ *such that*

$$\|S_{n1}\widehat{\phi}_{n1} - \phi_1\|_{\mu,2} \xrightarrow{\text{a.s.}} 0, \quad as\ n \to \infty.$$

Proof of Theorem A.2. Let

$$B_{1n} = \sup_{f \in \mathcal{B}_1} \left| \int_{\mathcal{V} \times \mathcal{V}} f(\boldsymbol{s})f(\boldsymbol{s}')\{V_n(\boldsymbol{s}, \boldsymbol{s}') - V_0(\boldsymbol{s}, \boldsymbol{s}')\}\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}') \right|. \tag{A.2}$$

By Theorem A.1 (ii), $B_{1n} \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$. Then

$$\int_{\mathcal{V} \times \mathcal{V}} \widehat{\phi}_{n1}(\boldsymbol{s})\widehat{\phi}_{n1}(\boldsymbol{s}')V_n(\boldsymbol{s}, \boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}') \leq \int_{\mathcal{V} \times \mathcal{V}} \widehat{\phi}_{n1}(\boldsymbol{s})\widehat{\phi}_{n1}(\boldsymbol{s}')V_0(\boldsymbol{s}, \boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}') + B_{1n}$$

$$\leq \sup_{f \in \mathcal{B}_1} \int_{\mathcal{V} \times \mathcal{V}} f(\boldsymbol{s})f(\boldsymbol{s}')V_0(\boldsymbol{s}, \boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}') + B_{1n} \xrightarrow{\text{a.s.}} \lambda_1, \quad as\ n \to \infty.$$

Let $\widetilde{\phi}_{n1}(\boldsymbol{s})$ be the projection of $\phi_1(\boldsymbol{s})$ onto $\mathcal{B}_1 \cap \mathcal{S}$. Then on the other side, as $n \to \infty$,

$$\int_{\mathcal{V} \times \mathcal{V}} \widehat{\phi}_{n1}(\boldsymbol{s})\widehat{\phi}_{n1}(\boldsymbol{s}')V_n(\boldsymbol{s}, \boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}') \geq \int_{\mathcal{V} \times \mathcal{V}} \widetilde{\phi}_{n1}(\boldsymbol{s})\widetilde{\phi}_{n1}(\boldsymbol{s}')V_n(\boldsymbol{s}, \boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}')$$

$$\geq \int_{\mathcal{V} \times \mathcal{V}} \widetilde{\phi}_{n1}(\boldsymbol{s})\widetilde{\phi}_{n1}(\boldsymbol{s}')V_0(\boldsymbol{s}, \boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}') - B_{1n} \xrightarrow{\text{a.s.}} \int_{\mathcal{V} \times \mathcal{V}} \phi_1(\boldsymbol{s})\phi_1(\boldsymbol{s}')V_0(\boldsymbol{s}, \boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}') = \lambda_1.$$

Therefore, if we define $\widehat{a}_{nk} = \int_{\mathcal{V}} \widehat{\phi}_{n1}(\boldsymbol{s})\phi_k(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s})$, we can conclude that as $n \to \infty$,

$$\int_{\mathcal{V} \times \mathcal{V}} \widehat{\phi}_{n1}(\boldsymbol{s})\widehat{\phi}_{n1}(\boldsymbol{s}')V_0(\boldsymbol{s}, \boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}') = \int_{\mathcal{V} \times \mathcal{V}} \sum_{k=1}^{\infty} \widehat{\phi}_{n1}(\boldsymbol{s})\widehat{\phi}_{n1}(\boldsymbol{s}')\lambda_k\phi_k(\boldsymbol{s})\phi_k(\boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}')$$

$$= \sum_{k=1}^{\infty} \widehat{a}_{nk}^2\lambda_k \xrightarrow{\text{a.s.}} \lambda_1.$$

Note that $\sum_{k=1}^{\infty} \widehat{a}_{nk}^2 \leq 1$ since $\widehat{\phi}_{n1} \in \mathcal{B}_1$, and since $\infty > \lambda_1 > \lambda_2 > \ldots$: this now leads to $\widehat{a}_{n1}^2 \xrightarrow{\text{a.s.}} 1$ and $\sum_{j=2}^{\infty} \widehat{a}_{nk}^2 \xrightarrow{\text{a.s.}} 0$. Hence there exists a sign sequence $\{S_{n1} \in \{-1, 1\}\}_n$ such that $\|S_{n1}\widehat{\phi}_{n1} - \phi_1\|_{\mu,2} \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$. $\square$

Theorem A.3. *Suppose for some* $1 \leq K < \infty$, $\exists$ *sign sequences* $\{S_{n1}, S_{n2}, \ldots, S_{nK} : S_{nk} \in \{-1, 1\}, k = 1, \ldots, K\}_n$ *such that*

$$\max_{1 \leq k \leq K} \|S_{nk}\widehat{\phi}_{nk} - \phi_k\|_{\mu,2} \xrightarrow{\text{a.s.}} 0, \quad as\ n \to \infty,$$

where $\widehat{\phi}_{nk} \in \mathcal{B}_1 \cap \mathcal{S}, \forall 1 \leq k \leq K$ *and* $\{\widehat{\phi}_{n1}, \ldots, \widehat{\phi}_{nK}\}$ *are orthogonal in* $\mathcal{H}_0$. *Recall the definitions of* $\mathcal{H}_{0K}$ *and* $\widehat{\mathcal{H}}_{nK}$ *in (2.5), and let* $\widehat{\mathcal{H}}_{nK}^{\perp}$ *and* $\mathcal{H}_{0K}^{\perp}$ *denote the respective closed*

orthocomplements in $\mathcal{H}_0$. Let $\{g_n\} \in \mathcal{H}_0$ be a sequence satisfying $\limsup_{n\to\infty} \|g_n\|_{\mu,2} < \infty$.
Suppose Assumption (A3) holds, then both

$$\|\widehat{\mathcal{H}}_{nK}g_n - \mathcal{H}_{0K}g_n\|_{\mu,2} \xrightarrow{\text{a.s.}} 0 \quad \text{and} \quad \|\widehat{\mathcal{H}}_{nK}^{\perp}g_n - \mathcal{H}_{0K}^{\perp}g_n\|_{\mu,2} \xrightarrow{\text{a.s.}} 0, \quad \text{as } n \to \infty.$$

Proof of Theorem A.3.

$$\widehat{\mathcal{H}}_{nK}g_n(\boldsymbol{s}) = \sum_{k=1}^{K} \widehat{\phi}_{nk}(\boldsymbol{s}) \int_{\mathcal{V}} \widehat{\phi}_{nk}(\boldsymbol{s}')g_n(\boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s}') = \sum_{k=1}^{K} S_{nk}\widehat{\phi}_{nk}(\boldsymbol{s}) \int_{\mathcal{V}} S_{nk}\widehat{\phi}_{nk}(\boldsymbol{s}')g_n(\boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s}')$$

$$= H_{0K}g_n(\boldsymbol{s}) + \sum_{k=1}^{K}\{S_{nk}\widehat{\phi}_{nk}(\boldsymbol{s}) - \phi_k(\boldsymbol{s})\} \int_{\mathcal{V}} \phi_k(\boldsymbol{s}')g_n(\boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s}')$$

$$+ \sum_{k=1}^{K} S_{nk}\widehat{\phi}_{nk}(\boldsymbol{s}) \int_{\mathcal{V}} \{S_{nk}\widehat{\phi}_{nk}(\boldsymbol{s}') - \phi_k(\boldsymbol{s}')\}g_n(\boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s}')$$

$$= H_{0K}g_n(\boldsymbol{s}) + F_{1n}(\boldsymbol{s}) + F_{2n}(\boldsymbol{s}).$$

Note that as $n \to \infty$,

$$\|F_{1n}\|_{\mu,2} \leq \sum_{k=1}^{K} \|S_{nk}\widehat{\phi}_{nk} - \phi_k\|_{\mu,2}\|g_n\|_{\mu,2} \xrightarrow{\text{a.s.}} 0, \quad \|F_{2n}\|_{\mu,2} \leq \sum_{k=1}^{K} \|S_{nk}\widehat{\phi}_{nk} - \phi_k\|_{\mu,2}\|g_n\|_{\mu,2} \xrightarrow{\text{a.s.}} 0.$$

Thus, $\|\widehat{\mathcal{H}}_{nK}g_n - \mathcal{H}_{0K}g_n\|_{\mu,2} \xrightarrow{\text{a.s.}} 0$, as $n \to \infty$. Since

$$\|\widehat{\mathcal{H}}_{nK}^{\perp}g_n - \mathcal{H}_{0K}^{\perp}g_n\|_{\mu,2} = \|(\mathbb{I} - \widehat{\mathcal{H}}_{nK})g_n - (\mathbb{I} - \mathcal{H}_{0K})g_n\|_{\mu,2} = \|\widehat{\mathcal{H}}_{nK}g_n - \mathcal{H}_{0K}g_n\|_{\mu,2},$$

where $\mathbb{I}$ is the identity operator, we now have the second conclusion of the theorem. $\square$

Theorem A.4. *Suppose Assumption (A3) holds. Assume for some $1 \leq K < \infty$, that $\{\widehat{\phi}_{n1}, \ldots, \widehat{\phi}_{nK}\}$ form an orthonormal system on $\mathcal{S} \cap \mathcal{H}_0$ with*

$$\max_{1 \leq k \leq K} \|S_{nk}\widehat{\phi}_{nk} - \phi_k\|_{\mu,2} \xrightarrow{\text{a.s.}} 0, \quad \text{as } n \to \infty,$$

for some sign sequences $\{S_{n1}, \ldots, S_{nK} : S_{nk} \in \{-1, 1\}, k = 1, \ldots, K\}_n$. Let

$$\widehat{\phi}_{n(K+1)} = \underset{f \in \mathcal{B}_1 \cap \mathcal{S} \cap \widehat{\mathcal{H}}_{nK}^{\perp}}{\arg\max} \int_{\mathcal{V} \times \mathcal{V}} f(\boldsymbol{s})f(\boldsymbol{s}')V_n(\boldsymbol{s}, \boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}').$$

Then $\exists$ a sign sequence $\{S_{n(K+1)} \in \{-1, 1\}\}_n$ such that

$$\|S_{n(K+1)}\widehat{\phi}_{n(K+1)} - \phi_{(K+1)}\|_{\mu,2} \xrightarrow{\text{a.s.}} 0, \quad \text{as } n \to \infty.$$

25

Proof of Theorem A.4. Let $V_{0K}(s, s') = \sum_{k=K+1}^{\infty} \lambda_k \phi_k(s) \phi_k(s')$. Recall the definition of B_{1n} given in (A.2) and $B_{1n} \xrightarrow{\text{a.s.}} 0$ as shown in Theorem A.2. Again by Theorem A.1 (ii), we have as $n \to \infty$,

$$\int_{\mathcal{V} \times \mathcal{V}} \widehat{\phi}_{n(K+1)}(s) \widehat{\phi}_{n(K+1)}(s') V_n(s, s') \mathrm{d}\mu(s) \mathrm{d}\mu(s') \tag{A.3}$$

$$\leq \int_{\mathcal{V} \times \mathcal{V}} \widehat{\phi}_{n(K+1)}(s) \widehat{\phi}_{n(K+1)}(s') V_0(s, s') \mathrm{d}\mu(s) \mathrm{d}\mu(s') + B_{1n}$$

$$\leq \int_{\mathcal{V} \times \mathcal{V}} \widehat{\phi}_{n(K+1)}(s) \widehat{\phi}_{n(K+1)}(s') V_{0K}(s, s') \mathrm{d}\mu(s) \mathrm{d}\mu(s') + B_{1n} + B_{2n},$$

$$\leq \sup_{f \in \mathcal{B}_1} \int_{\mathcal{V} \times \mathcal{V}} f(s) f(s') V_{0K}(s, s') \mathrm{d}\mu(s) \mathrm{d}\mu(s') + B_{1n} + B_{2n} \xrightarrow{\text{a.s.}} \lambda_{K+1},$$

where $B_{2n} = |\sum_{k=1}^{K} \lambda_k \{\int_{\mathcal{V}} \widehat{\phi}_{n(K+1)}(s) \phi_k(s) \mathrm{d}\mu(s)\}^2| \xrightarrow{\text{a.s.}} 0$ by Theorem A.3. Let $\widetilde{\phi}_{n(K+1)} = \mathcal{B}_1 \mathcal{S} \phi_{K+1}$, then

$$(A.3) \geq \int_{\mathcal{V} \times \mathcal{V}} \widetilde{\phi}_{n(K+1)}(s) \widetilde{\phi}_{n(K+1)}(s') V_n(s, s') \mathrm{d}\mu(s) \mathrm{d}\mu(s')$$

$$\geq \int_{\mathcal{V} \times \mathcal{V}} \widetilde{\phi}_{n(K+1)}(s) \widetilde{\phi}_{n(K+1)}(s') V_0(s, s') \mathrm{d}\mu(s) \mathrm{d}\mu(s') - B_{1n} \xrightarrow{\text{a.s.}} \lambda_{K+1}, \quad \text{as } n \to \infty.$$

Thus,

$$\int_{\mathcal{V} \times \mathcal{V}} \widehat{\phi}_{n(K+1)}(s) \widehat{\phi}_{n(K+1)}(s') V_{0K}(s, s') \mathrm{d}\mu(s) \mathrm{d}\mu(s') \xrightarrow{\text{a.s.}} \lambda_{K+1}, \quad \text{as } n \to \infty. \tag{A.4}$$

Let $\widehat{a}_{nk} = \int_{\mathcal{V}} \widehat{\phi}_{n(K+1)}(s) \phi_k(s) \mathrm{d}\mu(s)$. Using arguments similar to those in the proof of Theorem A.2, since $\sum_{k=1}^{\infty} \widehat{a}_{nk}^2 \leq 1$ and by $(A.4)$, we have $\sum_{k=K+1}^{\infty} \widehat{a}_{nk}^2 \lambda_k \xrightarrow{\text{a.s.}} \lambda_{K+1}$, which forces $\widehat{a}_{n(K+1)}^2 \xrightarrow{\text{a.s.}} 1$ and $\sum_{k=1}^{K} \widehat{a}_{nk}^2 + \sum_{k=K+2}^{\infty} \widehat{a}_{nk}^2 \xrightarrow{\text{a.s.}} 0$. Then there exists a sign sequence $\{S_{n(K+1)} \in \{-1, 1\}\}_n$ such that $\|S_{n(K+1)} \widehat{\phi}_{n(K+1)} - \phi_{K+1}\|_{\mu,2} \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$. $\qquad \square$

Proof of Theorem 1. Based on these theorems and lemmas introduced above, we are ready to give the detailed proof of Theorem 1.

(i) The conclusion of (i) follows by Theorems A.2 – A.4 directly.

(ii) The conclusion of (ii) has been shown in Theorem A.3.

(iii) It is obvious that for $1 \leq k \leq K_n \wedge p_n$,

$$\int_{\mathcal{V} \times \mathcal{V}} \widehat{\phi}_{nk}(s) \widehat{\phi}_{nk}(s') V_n(s, s') \mathrm{d}\mu(s) \mathrm{d}\mu(s') \geq \int_{\mathcal{V} \times \mathcal{V}} \widehat{\phi}_{nk}^{\mathcal{S}}(s) \widehat{\phi}_{nk}^{\mathcal{S}}(s') V_n(s, s') \mathrm{d}\mu(s) \mathrm{d}\mu(s').$$

26

On the other side, for any $f \in \mathcal{B}_1 \cap \mathcal{S}$, $\exists a \in \mathbb{R}^{J_n} \cap \mathcal{A}_1$, such that $f(s) = a^\top \mathbf{H}^{-1/2} \mathbf{B}(s)$. Thus

$$
\int_{\mathcal{V} \times \mathcal{V}} f(s) f(s') V_n(s, s') \mathrm{d}\mu(s) \mathrm{d}\mu(s')
$$

$$
= a^\top \mathbf{H}^{-1/2} \int_{\mathcal{V} \times \mathcal{V}} \mathbf{B}(s) V_n(s, s') \mathbf{B}^\top(s') \mathrm{d}\mu(s) \mathrm{d}\mu(s') \mathbf{H}^{-1/2} a
$$

$$
= a^\top \mathbf{K}_n a \le \widehat{\varphi}_{n1}^\top \mathbf{K}_n \widehat{\varphi}_{n1} = \int_{\mathcal{V} \times \mathcal{V}} \widehat{\phi}_{n1}^{\mathcal{S}}(s) \widehat{\phi}_{n1}^{\mathcal{S}}(s') V_n(s, s') \mathrm{d}\mu(s) \mathrm{d}\mu(s').
$$

We therefore obtain $\widehat{\phi}_{n1}^{\mathcal{S}}(s) = \arg\max_{f \in \mathcal{B}_1 \cap \mathcal{S}} \int_{\mathcal{V} \times \mathcal{V}} f(s) f(s') V_n(s, s') \mathrm{d}\mu(s) \mathrm{d}\mu(s')$ by taking the supremum over $\mathcal{B}_1 \cap \mathcal{S}$. Thus, $\widehat{\phi}_{n1}(s)$ is equal to $\widehat{\phi}_{n1}^{\mathcal{S}}(s)$ up to sign. Suppose $\forall k < K_n \wedge p_n$, we have $\widehat{\phi}_{n(k-1)}(s) = \widehat{\phi}_{n(k-1)}^{\mathcal{S}}(s)$. Similarly, we can obtain

$$
\widehat{\phi}_{n_k}^{\mathcal{S}}(s) = \arg\max_{f \in \mathcal{B}_1 \cap \mathcal{S} \cap \widehat{\mathcal{H}}_{n(k-1)}^{\perp}} \int_{\mathcal{V} \times \mathcal{V}} f(s) f(s') V_n(s, s') \mathrm{d}\mu(s) \mathrm{d}\mu(s')
$$

by noticing that $\widehat{\varphi}_{nk} = \arg\max_{a \in \mathbb{R}^{J_n} \cap \mathcal{A}_1} a^\top \widehat{\mathbf{K}}_{nk} a$, where

$$
\widehat{\mathbf{K}}_{nk} = \mathbf{H}^{-1/2} \int_{\mathcal{V} \times \mathcal{V}} \mathbf{B}(s) \sum_{k' \ge k} \widehat{\lambda}_{nk'} \widehat{\phi}_{nk}(s) \widehat{\phi}_{nk}(s') \mathbf{B}^\top(s') \mathrm{d}\mu(s) \mathrm{d}\mu(s') \mathbf{H}^{-1/2}.
$$

Hence, for $1 \le k \le K_n \wedge p_n$, $\widehat{\phi}_{nk}(s)$ and $\widehat{\phi}_{nk}^{\mathcal{S}}(s)$ are equal up to sign.

(iv) By the definitions of $\widehat{\lambda}_{nk}$ and λ_k, for $1 \le k \le K_n \wedge p_n$, we have

$$
\left| \widehat{\lambda}_{nk} - \lambda_k \right| = \left| \int_{\mathcal{V} \times \mathcal{V}} \left\{ \widehat{\phi}_{nk}(s) \widehat{\phi}_{nk}(s') V_n(s, s') - \phi_k(s) \phi_k(s') V_0(s, s') \right\} \mathrm{d}\mu(s) \mathrm{d}\mu(s') \right|
$$

$$
\le \left| \int_{\mathcal{V} \times \mathcal{V}} \widehat{\phi}_{nk}(s) \widehat{\phi}_{nk}(s') \left\{ V_n(s, s') - V_0(s, s') \right\} \mathrm{d}\mu(s) \mathrm{d}\mu(s') \right|
$$

$$
+ \left| \int_{\mathcal{V} \times \mathcal{V}} \left\{ \widehat{\phi}_{nk}(s) \widehat{\phi}_{nk}(s') - \phi_k(s) \phi_k(s') \right\} V_0(s, s') \mathrm{d}\mu(s) \mathrm{d}\mu(s') \right|.
$$

From the proofs of Theorem A.1 and Theorem A.4, as $n \to \infty$,

$$
\left| \int_{\mathcal{V} \times \mathcal{V}} \widehat{\phi}_{nk}(s) \widehat{\phi}_{nk}(s') \left\{ V_n(s, s') - V_0(s, s') \right\} \mathrm{d}\mu(s) \mathrm{d}\mu(s') \right| \xrightarrow{\text{a.s.}} 0.
$$

Since $\sum_{k=1}^\infty \lambda_k < \infty$ and $\lambda_1 > \lambda_2 > \ldots \ge 0$, $\int_{\mathcal{V} \times \mathcal{V}} V_0^2(s, s') \mathrm{d}\mu(s) \mathrm{d}\mu(s') = \sum_{k=1}^\infty \lambda_k^2 <$

∞. Thus,

$$\left| \int_{\mathcal{V}\times\mathcal{V}} \left\{ S_{nk}\widehat{\phi}_{nk}(\boldsymbol{s})S_{nk}\widehat{\phi}_{nk}(\boldsymbol{s}') - \phi_k(\boldsymbol{s})\phi_k(\boldsymbol{s}') \right\} V_0(\boldsymbol{s},\boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}') \right|$$

$$\leq \left| \int_{\mathcal{V}\times\mathcal{V}} S_{nk}\widehat{\phi}_{nk}(\boldsymbol{s}) \left\{ S_{nk}\widehat{\phi}_{nk}(\boldsymbol{s}') - \phi_k(\boldsymbol{s}') \right\} V_0(\boldsymbol{s},\boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}') \right|$$

$$+ \left| \int_{\mathcal{V}\times\mathcal{V}} \phi_k(\boldsymbol{s}') \left\{ S_{nk}\widehat{\phi}_{nk}(\boldsymbol{s}) - \phi_k(\boldsymbol{s}) \right\} V_0(\boldsymbol{s},\boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}') \right|$$

$$\leq \|S_{nk}\widehat{\phi}_{nk}\|_{\mu,2}\|S_{nk}f_{nj} - \phi_k\|_{\mu,2} \left\{ \int_{\mathcal{V}\times\mathcal{V}} V_0^2(\boldsymbol{s},\boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}') \right\}^{1/2}$$

$$+ \|\phi_k\|_{\mu,2}\|S_{nk}f_{nj} - \phi_k\|_{\mu,2} \left\{ \int_{\mathcal{V}\times\mathcal{V}} V_0^2(\boldsymbol{s},\boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}') \right\}^{1/2}$$

$$\leq 2\|S_{nk}f_{nj} - \phi_k\|_{\mu,2} \left\{ \int_{\mathcal{V}\times\mathcal{V}} V_0^2(\boldsymbol{s},\boldsymbol{s}')\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}') \right\}^{1/2} \to 0$$

almost sure as $n \to \infty$. The conclusion follows.

$\square$

B. Proof of Convergence of Estimators

In this section, we develop the convergence results of the coefficient estimator based on the proposed 2D-FPC bases. Following the sketch proof of Theorem 2 in Section 4.2 in the paper, we provide the supporting lemmas and the detailed proof.

We first define those coefficients and operators that will be employed in the proof. Recall that $\widetilde{\boldsymbol{\theta}} = (\boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_2^\top, \boldsymbol{\gamma}_1^\top, \boldsymbol{\gamma}_2^\top)^\top \in \widetilde{\boldsymbol{\Theta}}_{K_1,K_2}$, where $\boldsymbol{\alpha}_\ell \in \mathbb{R}^q$, $\ell = 1, 2$, are the linear coefficients, and $\boldsymbol{\gamma}_\ell \in \mathbb{R}^{K_\ell}$, $\ell = 1, 2$, are the basis coefficients. Let $\widetilde{\boldsymbol{\theta}}_n = \arg\min_{\widetilde{\boldsymbol{\theta}}\in\widetilde{\boldsymbol{\Theta}}_{K_1,K_2}} n^{-1} \sum_{i=1}^n \{Y_i - \widetilde{\boldsymbol{\theta}}(\widehat{\mathbf{W}}_i)\}^2$. We write $\widetilde{\boldsymbol{\theta}}_n = (\widetilde{\boldsymbol{\alpha}}_{1n}, \widetilde{\boldsymbol{\alpha}}_{2n}, \widetilde{\boldsymbol{\gamma}}_1, \widetilde{\boldsymbol{\gamma}}_2)$, with $\widetilde{\boldsymbol{\gamma}}_1 = (\widetilde{\gamma}_{1,1}, \ldots, \widetilde{\gamma}_{1,K_1})^\top$ and $\widetilde{\boldsymbol{\gamma}}_2 = (\widetilde{\gamma}_{2,1}, \ldots, \widetilde{\gamma}_{2,K_2})^\top$, then

$$\widehat{\boldsymbol{\theta}}_n = \left(\widetilde{\boldsymbol{\alpha}}_{1n}, \widetilde{\boldsymbol{\alpha}}_{2n}, \sum_{k=1}^{K_1} \widetilde{\gamma}_{1,k}\widehat{\phi}_{1,nk}(\boldsymbol{s}), \sum_{k=1}^{K_2} \widetilde{\gamma}_{2,k}\widehat{\phi}_{2,nk}(\boldsymbol{s}) \right)^\top .$$

Recall that $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}_{01}, \boldsymbol{\alpha}_{02}, \beta_{01}, \beta_{02}) \in \boldsymbol{\Theta}$ is the true parameter value. Let $\gamma_{0\ell,k} = \int_{\mathcal{V}} \phi_{\ell,k}(\boldsymbol{s})\beta_{0\ell}(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s})$, and thus $\beta_{0\ell}(\boldsymbol{s}) = \sum_{k=1}^\infty \gamma_{0\ell,k}\phi_{\ell,k}(\boldsymbol{s})$, $\ell = 1, 2$, where $\{\phi_{1,k}(\boldsymbol{s})\}_k$ and $\{\phi_{2,k}(\boldsymbol{s})\}_k$ are theoretical orthonormal bases for $Z(\boldsymbol{s})$ and $AZ(\boldsymbol{s})$, respectively. Define

$$\boldsymbol{\theta}_{0n} = \left(\boldsymbol{\alpha}_{01}^\top, \boldsymbol{\alpha}_{02}^\top, \sum_{k=1}^{K_1} \gamma_{01,k}S_{1,nk}\phi_{1,k}(\boldsymbol{s}), \sum_{k=1}^{K_2} \gamma_{02,k}S_{2,nk}\phi_{2,k}(\boldsymbol{s}) \right)^\top ,$$

and

$$\widetilde{\boldsymbol{\theta}}_{0n} = \left(\boldsymbol{\alpha}_{01}^\top, \boldsymbol{\alpha}_{02}^\top, (\gamma_{01,1}, \ldots, \gamma_{01,K_1}), (\gamma_{02,1}, \ldots, \gamma_{02,K_2})\right)^\top.$$

We first explore the properties of the truncated terms. Let

$$\Delta_{0n}(\mathbf{W}_i) = \sum_{k=K_1+1}^{\infty} \gamma_{01,k} \int_{\mathcal{V}} S_{nk} \phi_k(\boldsymbol{s}) Z_i(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}) + \sum_{k=K_2+1}^{\infty} \gamma_{02,k} \int_{\mathcal{V}} S_{nk} \phi_k(\boldsymbol{s}) A_i Z_i(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}).$$

(B.1)

We use the following Lemma B.1 to state that the truncated terms are negligible.

Lemma B.1. *Under Assumptions (A3) − (A8), as $n \to \infty$, $n^{-1} \sum_{i=1}^{n} \Delta_{0n}^2(\mathbf{W}_i) \to 0$.*

Proof. Decompose $\Delta_{0n}(\mathbf{W}_i) = I_1 + I_2$, where

$$I_1 = \sum_{k=K_1+1}^{\infty} \gamma_{01,k} \int_{\mathcal{V}} S_{nk} \phi_k(\boldsymbol{s}) Z_i(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}), \qquad I_2 = \sum_{k=K_2+1}^{\infty} \gamma_{02,k} \int_{\mathcal{V}} S_{nk} \phi_k(\boldsymbol{s}) A_i Z_i(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}).$$

We have

$$\begin{aligned}
\mathrm{E} I_1^2 &= \sum_{k=K_1+1}^{\infty} \sum_{k'=K_1+1}^{\infty} \gamma_{01,k} \gamma_{01,k'} \int_{\mathcal{V} \times \mathcal{V}} S_{nk} S_{nk'} \phi_k(\boldsymbol{s}) \phi_{k'}(\boldsymbol{s}') \left\{ \mathrm{E} Z_i(\boldsymbol{s}) Z_i(\boldsymbol{s}') \right\} \mathrm{d}\mu(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}') \\
&= \sum_{k=K_1+1}^{\infty} \sum_{k'=K_1+1}^{\infty} \gamma_{01,k} \gamma_{01,k'} \int_{\mathcal{V} \times \mathcal{V}} S_{nk} S_{nk'} \phi_k(\boldsymbol{s}) \phi_{k'}(\boldsymbol{s}') \left\{ \sum_{k''=1}^{\infty} \lambda_{k''} \phi_{k''}(\boldsymbol{s}) \phi_{k''}(\boldsymbol{s}') \right\} \mathrm{d}\mu(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}') \\
&= \sum_{k=K_1+1}^{\infty} \sum_{k'=K_1+1}^{\infty} \sum_{k''=1}^{\infty} \gamma_{01,k} \gamma_{01,k'} \lambda_{k''} S_{nk} S_{nk'} \int_{\mathcal{V}} \phi_k(\boldsymbol{s}) \phi_{k''}(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}) \int_{\mathcal{V}} \phi_{k'}(\boldsymbol{s}') \phi_{k''}(\boldsymbol{s}') \mathrm{d}\mu(\boldsymbol{s}') \\
&= \sum_{k=K_1+1}^{\infty} \gamma_{01,k}^2 \lambda_k.
\end{aligned}$$

Similarly, $\mathrm{E} I_2^2 \leq \sum_{k=K_2+1}^{\infty} \gamma_{02,k}^2 \lambda_k$, and $\mathrm{E} I_1 I_2 \leq \sum_{k=K_1 \vee K_2+1}^{\infty} \gamma_{01,k} \gamma_{02,k} \lambda_k$. Then as $n \to \infty$,

$$\begin{aligned}
n^{-1} \sum_{i=1}^{n} \Delta_{0n}^2(\mathbf{W}_i) \to \mathrm{E} \Delta_{0n}^2(\mathbf{W}) &\leq 2 \left(\sum_{k=K_1+1}^{\infty} \gamma_{01,k}^2 \lambda_k + \sum_{k=K_2+1}^{\infty} \gamma_{02,k}^2 \lambda_k \right) \\
&\leq 2 \left(\sum_{k=K_1+1}^{\infty} \gamma_{01,k}^2 + \sum_{k=K_2+1}^{\infty} \gamma_{02,k}^2 \right) \left(\sum_{k=K_1 \wedge K_2+1}^{\infty} \lambda_k \right) \equiv \mathrm{E}_{1n}^*(K_1, K_2).
\end{aligned}$$

Thus, we have $\infty > \mathrm{E}_{1n}^*(K_1, K_2) \to 0$ as $K_1 \wedge K_2 \to \infty$. $\qquad\square$

Next, we use the following Lemma B.2 to state that the errors induced by the approximation of $\{\widehat{\phi}_{nk}\}$ to $\{\phi_k\}$ are negligible.

29

Lemma B.2. *Under Assumptions (A3) – (A8), as $n \to \infty$, $n^{-1} \sum_{i=1}^n \{\boldsymbol{\theta}_{0n}(\mathbf{W}_i) - \widetilde{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i)\}^2 \xrightarrow{\text{a.s.}} 0$.*

Proof. By the definitions of $\boldsymbol{\theta}_{0n}(\mathbf{W}_i)$ and $\widetilde{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i)$,

$$\boldsymbol{\theta}_{0n}(\mathbf{W}_i) - \widetilde{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i) = \sum_{k=1}^{K_1} \gamma_{01,k} \int_{\mathcal{V}} \left\{ S_{nk}\phi_k(\boldsymbol{s}) - \widehat{\phi}_{nk}(\boldsymbol{s}) \right\} Z_i(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s})$$

$$+ \sum_{k=1}^{K_2} \gamma_{02,k} \int_{\mathcal{V}} \left\{ S_{nk}\phi_k(\boldsymbol{s}) - \widehat{\phi}_{nk}(\boldsymbol{s}) \right\} A_i Z_i(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}).$$

So $|\boldsymbol{\theta}_{0n}(\mathbf{W}_i) - \widehat{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i)| \leq \sum_{k=1}^{K_1} \gamma_{01,k} \|S_{nk}\phi_k - \widehat{\phi}_{nk}\|_{\mu,2} \|Z_i\|_{\mu,2} + \sum_{k=1}^{K_2} \gamma_{02,k} \|S_{nk}\phi_k - \widehat{\phi}_{nk}\|_{\mu,2} \|Z_i\|_{\mu,2}$ since $|A_i| \leq 1$. By Theorem 1, there exist a sequence $\{r_n\}$, such that $r_n \|S_{nk}\phi_k - \widehat{\phi}_{nk}\|_{\mu,2} \to 0$. Take $K_1 \vee K_2 = r_n^{-1/2}$, as $n \to \infty$,

$$n^{-1} \sum_{i=1}^n \left\{ \boldsymbol{\theta}_{0n}(\mathbf{W}_i) - \widetilde{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i) \right\}^2$$

$$\leq \left(\sum_{j=1}^{K_1 \vee K_2} \gamma_{01,k}^2 + \gamma_{02,k}^2 \right) \left(2 \sum_{j=1}^{K_1 \vee K_2} \|S_{nk}\phi_k - \widehat{\phi}_{nk}\|_{\mu,2}^2 \right) n^{-1} \sum_{i=1}^n \|Z_i\|_{\mu,2}^2 \xrightarrow{\text{a.s.}} 0.$$

$\square$

Combining Lemmas B.1 and B.2, we use the following Lemma B.3 to illuminate the convergence rate of $\widetilde{\boldsymbol{\theta}}_{0n}$ to the observed Y_i.

Lemma B.3. *Under Assumptions (A3) – (A8), as $n \to \infty$, $n^{-1} \sum_{i=1}^n \{Y_i - \widetilde{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i)\}^2 \xrightarrow{\text{a.s.}} C_1$ for some constant C_1.*

Proof. By definition,

$$n^{-1} \sum_{i=1}^n \left\{ Y_i - \widetilde{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i) \right\}^2 = n^{-1} \sum_{i=1}^n \left\{ \boldsymbol{\theta}_0(\mathbf{W}_i) + \varepsilon_i - \widetilde{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i) \right\}^2$$

$$= n^{-1} \sum_{i=1}^n \left\{ \varepsilon_i + \Delta_{0n}(\mathbf{W}_i) + \boldsymbol{\theta}_{0n}(\mathbf{W}_i) - \widetilde{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i) \right\}^2 \quad \text{a.s.},$$

where $\Delta_{0n}(\mathbf{W}_i)$ is given in (B.1). Since $n^{-1} \sum_{i=1}^n \varepsilon_i^2 \to \mathrm{Var}(\varepsilon) = \sigma^2 < \infty$, combined with Lemmas B.1 and B.2, we have

$$n^{-1} \sum_{i=1}^n \left\{ Y_i - \widetilde{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i) \right\}^2 \xrightarrow{\text{a.s.}} C_1 < \infty \tag{B.2}$$

for some constant C_1.

$\square$

Note that $n^{-1} \sum_{i=1}^n \{Y_i - \widetilde{\boldsymbol{\theta}}_n(\widehat{\mathbf{W}}_i)\}^2 \leq n^{-1} \sum_{i=1}^n \{Y_i - \widetilde{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i)\}^2$. Consequently,

$$-\frac{2}{n} \sum_{i=1}^n \left\{ Y_i - \widetilde{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i) \right\} \left\{ \widetilde{\boldsymbol{\theta}}_n(\widehat{\mathbf{W}}_i) - \widetilde{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i) \right\} + \frac{1}{n} \sum_{i=1}^n \left\{ \widetilde{\boldsymbol{\theta}}_n(\widehat{\mathbf{W}}_i) - \widetilde{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i) \right\}^2 \leq 0.$$

(B.3)

Let $\widetilde{\boldsymbol{\delta}}_n = \widetilde{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}_{0n}$ and $\widetilde{\boldsymbol{r}}_n = \widetilde{\boldsymbol{\delta}}_n / \|\widetilde{\boldsymbol{\delta}}_n\|_2$. We use the following Lemma B.4 to build the boundedness of $\widetilde{\boldsymbol{\delta}}_n$.

Lemma B.4. *Under Assumptions (A3) – (A10),* $\limsup_{n \to \infty} \|\widetilde{\boldsymbol{\delta}}_n\|_2 \leq C_2 < \infty$ *almost surely for some constant* C_2.

Proof.

$$(B.3) \Rightarrow \frac{\|\widetilde{\boldsymbol{\delta}}_n\|_2^2}{1 + \|\widetilde{\boldsymbol{\delta}}_n\|_2} \frac{1}{n} \sum_{i=1}^n \{\widetilde{\boldsymbol{r}}_n(\widehat{\mathbf{W}}_i)\}^2 \leq \frac{2}{n} \sum_{i=1}^n \left\{ Y_i - \widetilde{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i) \right\} \widetilde{\boldsymbol{r}}_n(\widehat{\mathbf{W}}_i)$$

$$\leq 2 \left[\frac{1}{n} \sum_{i=1}^n \left\{ Y_i - \widetilde{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i) \right\}^2 \right]^{1/2} \left[\frac{1}{n} \sum_{i=1}^n \left\{ \widetilde{\boldsymbol{r}}_n(\widehat{\mathbf{W}}_i) \right\}^2 \right]^{1/2},$$

$$\Rightarrow \frac{\|\widetilde{\boldsymbol{\delta}}_n\|_2^2}{1 + \|\widetilde{\boldsymbol{\delta}}_n\|_2} \leq \frac{2 \left[\frac{1}{n} \sum_{i=1}^n \left\{ Y_i - \widetilde{\boldsymbol{\theta}}_{0n}(\widehat{\mathbf{W}}_i) \right\}^2 \right]^{1/2}}{\left[\frac{1}{n} \sum_{i=1}^n \left\{ \widetilde{\boldsymbol{r}}_n(\widehat{\mathbf{W}}_i) \right\}^2 \right]^{1/2}}.$$

(B.4)

By Lemma B.3, the numerator on the right-hand-side (RHS) of (B.4) $\xrightarrow{\text{a.s.}} 2C_1^{1/2}$ as $n \to \infty$.

As for the denominator on the RHS of $(B.4)$, we first define

$$\widehat{\mathbf{D}}_n = \frac{1}{n} \sum_{i=1}^n \widehat{\mathbf{W}}_i \widehat{\mathbf{W}}_i^\top = \begin{pmatrix} \widehat{\mathbf{D}}_{11} & \widehat{\mathbf{D}}_{12} & \widehat{\mathbf{D}}_{13} & \widehat{\mathbf{D}}_{14} \\ \widehat{\mathbf{D}}_{12}^\top & \widehat{\mathbf{D}}_{22} & \widehat{\mathbf{D}}_{23} & \widehat{\mathbf{D}}_{24} \\ \widehat{\mathbf{D}}_{13}^\top & \widehat{\mathbf{D}}_{23}^\top & \widehat{\mathbf{D}}_{33} & \widehat{\mathbf{D}}_{34} \\ \widehat{\mathbf{D}}_{14}^\top & \widehat{\mathbf{D}}_{24}^\top & \widehat{\mathbf{D}}_{34}^\top & \widehat{\mathbf{D}}_{44} \end{pmatrix},$$

where for each entry of $\widehat{\mathbf{D}}_n$,

$\widehat{\mathbf{D}}_{11} = n^{-1} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top,$ $\widehat{\mathbf{D}}_{12} = \widehat{\mathbf{D}}_{22} = n^{-1} \sum_{i=1}^n A_i \mathbf{X}_i \mathbf{X}_i^\top,$

$\widehat{\mathbf{D}}_{13} = n^{-1} \sum_{i=1}^n \mathbf{X}_i \widehat{\mathbf{U}}_{1,i}^\top(K_1),$ $\widehat{\mathbf{D}}_{14} = \widehat{\mathbf{D}}_{24} = n^{-1} \sum_{i=1}^n A_i \mathbf{X}_i \widehat{\mathbf{U}}_{2,i}^\top(K_2),$

$\widehat{\mathbf{D}}_{23} = n^{-1} \sum_{i=1}^n A_i \mathbf{X}_i \widehat{\mathbf{U}}_{1,i}^\top(K_1),$ $\widehat{\mathbf{D}}_{33} = n^{-1} \sum_{i=1}^n \widehat{\mathbf{U}}_{1,i}(K_1) \widehat{\mathbf{U}}_{1,i}^\top(K_1),$

$\widehat{\mathbf{D}}_{34} = n^{-1} \sum_{i=1}^n A_i \widehat{\mathbf{U}}_{1,i}(K_1) \widehat{\mathbf{U}}_{2,i}^\top(K_2),$ $\widehat{\mathbf{D}}_{44} = n^{-1} \sum_{i=1}^n A_i \widehat{\mathbf{U}}_{2,i}(K_2) \widehat{\mathbf{U}}_{2,i}^\top(K_2).$

Then we have $n^{-1} \sum_{i=1}^n \{\widetilde{\boldsymbol{r}}_n(\widehat{\mathbf{W}}_i)\}^2 = \widetilde{\boldsymbol{r}}_n^\top \widehat{\mathbf{D}}_n \widetilde{\boldsymbol{r}}_n.$

In a parallel fashion to $\widehat{\mathbf{W}}_i$ and $\widehat{\mathbf{D}}_n$, for $i = 1, \ldots, n$, let $\widetilde{\mathbf{W}}_i = (\mathbf{X}_i^\top, A_i \mathbf{X}_i^\top, \widetilde{\mathbf{U}}_{1,i}^\top(K_1), A_i \widetilde{\mathbf{U}}_{2,i}^\top(K_2))^\top$, where $\widetilde{\mathbf{U}}_{\ell,i}(K_\ell) = (\widetilde{U}_{\ell,i1}, \ldots, \widetilde{U}_{\ell,iK_\ell})^\top$, $\ell = 1, 2$, and

$$
\begin{aligned}
\widetilde{U}_{1,ik} &= \int_{\mathcal{V}} S_{1,nk} \phi_{1,k}(\boldsymbol{s}) Z_i(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}), \ \ 1 \le k \le K_1, \\
\widetilde{U}_{2,ik} &= \int_{\mathcal{V}} S_{2,nk} \phi_{2,k}(\boldsymbol{s}) A_i Z_i(\boldsymbol{s}) \mathrm{d}\mu(\boldsymbol{s}), \ \ 1 \le k \le K_2.
\end{aligned}
$$

Then define

$$
\mathbf{D} = \mathrm{E}\widetilde{\mathbf{W}}_i \widetilde{\mathbf{W}}_i^\top = \begin{pmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} & \mathbf{D}_{13} & \mathbf{D}_{14} \\ \mathbf{D}_{12}^\top & \mathbf{D}_{22} & \mathbf{D}_{23} & \mathbf{D}_{24} \\ \mathbf{D}_{13}^\top & \mathbf{D}_{23}^\top & \mathbf{D}_{33} & \mathbf{D}_{34} \\ \mathbf{D}_{14}^\top & \mathbf{D}_{24}^\top & \mathbf{D}_{34}^\top & \mathbf{D}_{44} \end{pmatrix},
$$

where for each entry of $\mathbf{D}$,

$$
\begin{aligned}
\mathbf{D}_{11} &= \mathrm{E}(\mathbf{X}_i \mathbf{X}_i^\top), \ \ \mathbf{D}_{12} = \mathbf{D}_{22} = \mathrm{E}(A_i \mathbf{X}_i \mathbf{X}_i^\top), \\
\mathbf{D}_{13} &= \mathrm{E}\{\mathbf{X}_i \widetilde{\mathbf{U}}_{1,i}^\top(K_1)\} = \int_{\mathcal{V}} \mathrm{E}\left\{ \mathbf{X}_i Z_i(\boldsymbol{s})\Big(S_{1,n1}\phi_{1,1}(\boldsymbol{s}), \ldots, S_{1,nK_1}\phi_{1,K_1}(\boldsymbol{s}) \Big) \right\} \mathrm{d}\mu(\boldsymbol{s}), \\
\mathbf{D}_{14} &= \mathbf{D}_{24} = \mathrm{E}\{A_i \mathbf{X}_i \widetilde{\mathbf{U}}_{2,i}^\top(K_2)\} \\
&= \int_{\mathcal{V}} \mathrm{E}\left[A_i \mathbf{X}_i \{A_i Z_i(\boldsymbol{s})\} \Big(S_{2,n1}\phi_{2,1}(\boldsymbol{s}), \ldots, S_{2,nK_2}\phi_{2,K_2}(\boldsymbol{s}) \Big) \right] \mathrm{d}\mu(\boldsymbol{s}),
\end{aligned}
$$

$$
\mathbf{D}_{23} = \mathrm{E}\{A_i \mathbf{X}_i \widetilde{\mathbf{U}}_{1,i}^\top(K_1)\} = \int_{\mathcal{V}} \mathrm{E}\left\{ A_i \mathbf{X}_i Z_i(\boldsymbol{s})\Big(S_{1,n1}\phi_{1,1}(\boldsymbol{s}), \ldots, S_{1,nK_1}\phi_{1,K_1}(\boldsymbol{s}) \Big) \right\} \mathrm{d}\mu(\boldsymbol{s}),
$$

$$
\begin{aligned}
\mathbf{D}_{33} &= \mathrm{E}\{\widetilde{\mathbf{U}}_{1,i}(K_1) \widetilde{\mathbf{U}}_{1,i}^\top(K_1)\} = \int_{\mathcal{V}\times\mathcal{V}} \mathrm{E}\Big\{ Z_i(\boldsymbol{s}) Z_i(\boldsymbol{s}') \times \\
& \qquad \begin{pmatrix} S_{1,n1}\phi_{1,1}(\boldsymbol{s}) \\ \vdots \\ S_{1,nK_1}\phi_{1,K_1}(\boldsymbol{s}) \end{pmatrix} \Big(S_{1,n1}\phi_{1,1}(\boldsymbol{s}'), \ldots, S_{1,nK_1}\phi_{1,K_1}(\boldsymbol{s}') \Big) \Big\} \mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}'),
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{D}_{34} &= \mathrm{E}\{A_i \widetilde{\mathbf{U}}_{1,i}(K_1) \widetilde{\mathbf{U}}_{2,i}^\top(K_2)\} = \int_{\mathcal{V}\times\mathcal{V}} \mathrm{E}\big[A_i \{Z_i(\boldsymbol{s})\} \{A_i Z_i(\boldsymbol{s}')\} \times \\
& \qquad \begin{pmatrix} S_{1,n1}\phi_{1,1}(\boldsymbol{s}) \\ \vdots \\ S_{1,nK_1}\phi_{1,K_1}(\boldsymbol{s}) \end{pmatrix} \Big(S_{2,n1}\phi_{2,1}(\boldsymbol{s}'), \ldots, S_{2,nK_2}\phi_{2,K_2}(\boldsymbol{s}') \Big) \big] \mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}'),
\end{aligned}
$$

$$\mathbf{D}_{44} = \mathrm{E}\{A_i\widetilde{\mathbf{U}}_{2,i}(K_2)\widetilde{\mathbf{U}}_{2,i}^\top(K_2)\} = \int_{\mathcal{V}\times\mathcal{V}} \mathrm{E}\left[A_i\left\{A_iZ_i(\boldsymbol{s})\right\}\left\{A_iZ_i(\boldsymbol{s}')\right\} \times\right.$$

$$\left.\begin{pmatrix} S_{2,n1}\phi_{2,1}(\boldsymbol{s}) \\ \vdots \\ S_{2,nK_2}\phi_{2,K_2}(\boldsymbol{s}) \end{pmatrix}\left(S_{2,n1}\phi_{2,1}(\boldsymbol{s}'),\dots,S_{2,nK_2}\phi_{2,K_2}(\boldsymbol{s}')\right)\right]\mathrm{d}\mu(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}').$$

By strong law of large numbers and Theorem A.1, we can recycle previous arguments to verify that $\|\widehat{\mathbf{D}}_n - \mathbf{D}\|_\infty \xrightarrow{\text{a.s.}} 0$. From Assumptions (A3) and (A9), we know that the minimum eigenvalue of $\mathbf{D} \geq c_1$ for some constant $c_1 > 0$. Hence

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\left\{\widetilde{\boldsymbol{r}}_n(\widehat{\mathbf{W}}_i)\right\}^2 \geq c_1 > 0, \quad \text{a.s.}$$

Let $C_2 = 2C_1^{1/2}c_1^{-1}$, then the $\limsup_{n\to\infty}$ of the RHS of (B.4) $\leq C_2 < \infty$.

Now, if $\frac{a^2}{1+a} \leq k$ for all $a \geq 0$ and some $k < \infty$, then $a \leq k + \sqrt{k}$ (follows from algebra). Thus,

$$\limsup_{n\to\infty}\|\widetilde{\boldsymbol{\delta}}_n\|_2 \leq C_2 + \sqrt{C_2} < \infty.$$

$\square$

The following Theorem B.1 gives the tightness of $\widehat{\boldsymbol{\theta}}_n$.

Theorem B.1. *Under Assumptions (A5) – (A10), for fixed $1 \leq K_1, K_2 < \infty$, $\exists\, C_\theta < \infty$, such that*

$$\limsup_{n\to\infty}\|\widehat{\boldsymbol{\theta}}_n\|_* \leq C_\theta \quad \text{a.s.} \tag{B.5}$$

Proof of Theorem B.1. The conclusion follows directly from Lemmas B.2 and B.4, and the isomorphism between $\widehat{\boldsymbol{\Theta}}_{K_1,K_2}$ and $\widetilde{\boldsymbol{\Theta}}_{K_1,K_2}$. $\square$

Lemma B.5. *Under Assumptions (A3) – (A10), as $n \to \infty$,*

$$\mathbb{P}_n\left\{Y - \widehat{\boldsymbol{\theta}}_n(\mathbf{W})\right\}^2 - P\left\{Y - \widehat{\boldsymbol{\theta}}_n(\mathbf{W})\right\}^2 \xrightarrow{\text{a.s.}} 0. \tag{B.6}$$

Proof. By the isomorphism between $\widehat{\boldsymbol{\Theta}}_{K_1,K_2}$ and $\widetilde{\boldsymbol{\Theta}}_{K_1,K_2}$, it is equivalent to show

$$\mathbb{P}_n\left\{Y - \widetilde{\boldsymbol{\theta}}_n(\widehat{\mathbf{W}})\right\}^2 - P\left\{Y - \widetilde{\boldsymbol{\theta}}_n(\widehat{\mathbf{W}})\right\}^2 \xrightarrow{\text{a.s.}} 0.$$

Note that

$$\widetilde{\boldsymbol{\theta}}_n(\widehat{\mathbf{W}}) = \widetilde{\boldsymbol{\alpha}}_{1n}^\top \mathbf{X} + \widetilde{\boldsymbol{\alpha}}_{2n}^\top A\mathbf{X} + \sum_{k=1}^{K_1} \widetilde{\gamma}_{1,k}\widehat{U}_{1,k} + \sum_{k=1}^{K_2} \widetilde{\gamma}_{2,k}\widehat{U}_{2,k}$$

$$= \widetilde{\boldsymbol{\alpha}}_{1n}^\top \mathbf{X} + \widetilde{\boldsymbol{\alpha}}_{2n}^\top A\mathbf{X} + \int_{\mathcal{V}} \sum_{k=1}^{K_1} \widetilde{\gamma}_{1,k}\widehat{\phi}_{1,nk}(\boldsymbol{s})Z(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s}) + \int_{\mathcal{V}} \sum_{k=1}^{K_2} \widetilde{\gamma}_{2,k}\widehat{\phi}_{2,nk}(\boldsymbol{s})AZ(\boldsymbol{s})\mathrm{d}\mu(\boldsymbol{s})$$

$$\triangleq \widehat{f}_n(\mathbf{W}).$$

By Theorem A.1 and Lemma B.4, $\widehat{f}_n$ is contained in a Glivenko-Cantelli (G-C) class with probability going to one as $n \to \infty$. By Glivenko-Cantelli Preservation, $Y - \widetilde{\boldsymbol{\theta}}_n(\widehat{\mathbf{W}})$ is also a G-C class, and by Theorem A.1, $\{Y - \widetilde{\boldsymbol{\theta}}_n(\widehat{\mathbf{W}})\}^2$ is also a G-C class, and thus the conclusion follows. $\qquad\square$

With all these preparations, we can show the detailed proof of Theorem 2.

Proof of Theorem 2. Let Ω be the set of $\boldsymbol{\omega}$ for which both (B.5) and (B.6) hold. Fix $\boldsymbol{\omega} \in \Omega$. By definition of Ω, $\exists$ a subsequence $\{n'\}$ such that along this subsequence

$$\widehat{\boldsymbol{\theta}}_{n'} \to \boldsymbol{\theta}_* \in \boldsymbol{\Theta}_{K_1,K_2},$$

for some limit $\boldsymbol{\theta}_*$, where $\boldsymbol{\Theta}_{K_1,K_2}$ is given in (3.3). Note that by Assumption (A8), $P\{Y - \boldsymbol{\theta}(\mathbf{W})\}^2$ has a unique minimizer over $\boldsymbol{\Theta}_{K_1,K_2}$, $\boldsymbol{\theta}_0^*$. Let

$$r_{1n} = \left| \mathbb{P}_n\{Y - \boldsymbol{\theta}_0^*(\mathbf{W})\}^2 - P\{Y - \boldsymbol{\theta}_0^*(\mathbf{W})\}^2 \right|,$$

$$r_{2n} = \left| \mathbb{P}_n\left\{Y - \widehat{\boldsymbol{\theta}}_n(\mathbf{W})\right\}^2 - P\left\{Y - \widehat{\boldsymbol{\theta}}_n(\mathbf{W})\right\}^2 \right|.$$

By Lemma B.5, $r_{2n} \to 0$ as $n \to \infty$. Thus,

$$P\{Y - \boldsymbol{\theta}_0^*(\mathbf{W})\}^2 \geq \mathbb{P}_n\{Y_i - \boldsymbol{\theta}_0^*(\mathbf{W})\}^2 - r_{1n} \geq \mathbb{P}_n\{Y_i - \widehat{\boldsymbol{\theta}}_n(\mathbf{W})\}^2 - r_{1n}$$

$$\geq P\{Y - \widehat{\boldsymbol{\theta}}_n(\mathbf{W})\}^2 - r_{1n} - r_{2n}$$

$$\to P\{Y - \boldsymbol{\theta}_*(\mathbf{W})\}^2 - \lim r_{1n} \geq P\{Y - \boldsymbol{\theta}_0^*(\mathbf{W})\}^2,$$

as $n \to \infty$, where the convergence is attained along the subsequence n'. Since the subsequence is arbitrary, and with the help of the strong law, we can conclude that $\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0^*$ and $r_{1n} \to 0$ as $n \to \infty$. Thus, for fixed $1 \leq K_1, K_2 < \infty$, integrates P over $\boldsymbol{\omega}$ in Ω, we have

$$P\left\{Y - \widehat{\boldsymbol{\theta}}_n(\mathbf{W})\right\}^2 - P\{Y - \boldsymbol{\theta}_0^*(\mathbf{W})\}^2 \xrightarrow{\text{a.s.}} 0. \tag{B.7}$$

However, this is also true if we allow $K_{1n}, K_{2n} \to \infty$ sufficiently slowly. Recycling previous arguments as in the proofs of Lemmas B.1 – B.4, we can now show that

$$P\{\boldsymbol{\theta}_0(\mathbf{W}) - \boldsymbol{\theta}_0^*(\mathbf{W})\}^2 \to 0, \quad \text{as } K_{1n}, K_{2n} \to \infty. \tag{B.8}$$

Then

$$
\begin{aligned}
&P\{Y - \widehat{\boldsymbol{\theta}}_n(\mathbf{W})\}^2 - P\{Y - \boldsymbol{\theta}_0^*(\mathbf{W})\}^2 \\
=&P\left[\{Y - \boldsymbol{\theta}_0^*(\mathbf{W})\} - \left\{\widehat{\boldsymbol{\theta}}_n(\mathbf{W}) - \boldsymbol{\theta}_0^*(\mathbf{W})\right\}\right]^2 - P\{Y - \boldsymbol{\theta}_0^*(\mathbf{W})\}^2 \\
=&P\left\{\widehat{\boldsymbol{\theta}}_n(\mathbf{W}) - \boldsymbol{\theta}_0^*(\mathbf{W})\right\}^2 - 2P\left[\{Y - \boldsymbol{\theta}_0^*(\mathbf{W})\}\left\{\widehat{\boldsymbol{\theta}}_n(\mathbf{W}) - \boldsymbol{\theta}_0^*(\mathbf{W})\right\}\right].
\end{aligned}
$$

Let $E_n = 2P[\{Y - \boldsymbol{\theta}_0^*(\mathbf{W})\}\{\widehat{\boldsymbol{\theta}}_n(\mathbf{W}) - \boldsymbol{\theta}_0^*(\mathbf{W})\}]$. By (B.8), we have

$$
\begin{aligned}
|E_n| \leq &2\left|P\left[\{\varepsilon + \boldsymbol{\theta}_0(\mathbf{W}) - \boldsymbol{\theta}_0^*(\mathbf{W})\}\left\{\widehat{\boldsymbol{\theta}}_n(\mathbf{W}) - \boldsymbol{\theta}_0^*(\mathbf{W})\right\}\right]\right| \\
=&2\left|P\left[\{\boldsymbol{\theta}_0(\mathbf{W}) - \boldsymbol{\theta}_0^*(\mathbf{W})\}\left\{\widehat{\boldsymbol{\theta}}_n(\mathbf{W}) - \boldsymbol{\theta}_0^*(\mathbf{W})\right\}\right]\right| \\
\leq&\left[P\{\boldsymbol{\theta}_0(\mathbf{W}) - \boldsymbol{\theta}_0^*(\mathbf{W})\}^2\right]^{1/2}\left[P\left\{\widehat{\boldsymbol{\theta}}_n(\mathbf{W}) - \boldsymbol{\theta}_0^*(\mathbf{W})\right\}^2\right]^{1/2} \xrightarrow{\text{a.s.}} 0,
\end{aligned}
$$

as $n \to \infty$. Combining with (B.7) above, we obtain that $P\{\widehat{\boldsymbol{\theta}}_n(\mathbf{W}) - \boldsymbol{\theta}_0^*(\mathbf{W})\}^2 \xrightarrow{\text{a.s.}} 0$. Thus combining with (B.8), as $n \to \infty$, the conclusion follows by

$$
\begin{aligned}
P\left\{\widehat{\boldsymbol{\theta}}_n(\mathbf{W}) - \boldsymbol{\theta}_0(\mathbf{W})\right\}^2 &= P\left\{\widehat{\boldsymbol{\theta}}_n(\mathbf{W}) - \boldsymbol{\theta}_0^*(\mathbf{W}) + \boldsymbol{\theta}_0^*(\mathbf{W}) - \boldsymbol{\theta}_0(\mathbf{W})\right\}^2 \\
&\xrightarrow{\text{a.s.}} P\left\{\widehat{\boldsymbol{\theta}}_n(\mathbf{W}) - \boldsymbol{\theta}_0^*(\mathbf{W})\right\}^2 + P\{\boldsymbol{\theta}_0^*(\mathbf{W}) - \boldsymbol{\theta}_0(\mathbf{W})\}^2 \xrightarrow{\text{a.s.}} 0.
\end{aligned}
$$

$\square$

References

Ashford, J. W. and Mortimer, J. A. (2002), "Non-familial Alzheimer's disease is mainly due to genetic factors," *Journal of Alzheimer's disease*, 4, 169–177.

Bateman, R. J., Xiong, C., Benzinger, T. L., Fagan, A. M., Goate, A., Fox, N. C., Marcus, D. S., Cairns, N. J., Xie, X., Blazey, T. M., Holtzman, D. M., Santacruz, A., Buckles, V., Oliver, A., Moulder, K., Aisen, P. S., Ghetti, B., Klunk, W. E., McDade, E., Martins, R. N., Masters, C. L., Mayeux, R., Ringman, J. M., Rossor, M. N., Schofield, P. R., Sperling, R. A., Salloway, S., and Morris, J. C. (2012), "Clinical and biomarker changes in dominantly inherited Alzheimer's disease," *New England Journal of Medicine*, 367, 795–804.

Bern, M. and Eppstein, D. (1995), "Mesh generation and optimal triangulation," in *Computing in Euclidean geometry*, World Scientific, pp. 47–123.

Ciarleglio, A., Petkova, E., Ogden, R. T., and Tarpey, T. (2015), "Treatment decisions based on scalar and functional baseline covariates," *Biometrics*, 71, 884–894.

Ciarleglio, A., Petkova, E., Ogden, T., and Tarpey, T. (2018), "Constructing treatment decision rules based on scalar and functional predictors when moderators of treatment effect are unknown," *Journal of the Royal Statistical Society Series C*, 67, 1331–1356.

Ciarleglio, A., Petkova, E., Tarpey, T., and Ogden, R. T. (2016), "Flexible functional regression methods for estimating individualized treatment rules," *Stat*, 5, 185–199.

Clifton, J. and Laber, E. (2020), "Q-learning: Theory and applications," *Annual Review of Statistics and Its Application*, 7, 279–301.

Fratiglioni, L., Winblad, B., and von Strauss, E. (2007), "Prevention of Alzheimer's disease and dementia. Major findings from the Kungsholmen Project," *Physiology & behavior*, 92, 98–104.

Hernán, M. A. and Robins, J. M. (2010), *Causal Inference*, CRC Boca Raton, FL;.

Herold, C. J., Lewin, J. S., Wibmer, A. G., Thrall, J. H., Krestin, G. P., Dixon, A. K., Schoenberg, S. O., Geckle, R. J., Muellner, A., and Hricak, H. (2016), "Imaging in the age of precision medicine: summary of the proceedings of the 10th Biannual Symposium of the International Society for Strategic Studies in Radiology," *Radiology*, 279, 226–238.

Isaacson, R. S., Hristov, H., Saif, N., Hackett, K., Hendrix, S., Melendez, J., Safdieh, J., Fink, M., Thambisetty, M., Sadek, G., et al. (2019), "Individualized clinical management of patients at risk for Alzheimer's dementia," *Alzheimer's & Dementia*, 15, 1588–1602.

Kong, D., Staicu, A.-M., and Maity, A. (2016), "Classical testing in functional linear models," *Journal of nonparametric statistics*, 28, 813–838.

Kosorok, M. R. (2008), *Introduction to Empirical Processes and Semiparametric Inference*, Springer: New York.

Kosorok, M. R. and Laber, E. B. (2019), "Precision medicine," *Annual review of statistics and its application*, 6, 263–286.

Laber, E. B. and Staicu, A.-M. (2018), "Functional feature construction for individualized treatment regimes," *Journal of the American Statistical Association*, 113, 1219–1227.

Lai, M.-J. and Schumaker, L. L. (2007), *Spline functions on triangulations.*, Cambridge University Press.

Lai, M.-J. and Wang, L. (2013), "Bivariate penalized splines for regression," *Statistica Sinica*, 23, 1399–1417.

Li, X., Wang, L., and Wang, H. J. (2021), "Sparse Learning and Structure Identification for Ultrahigh-Dimensional Image-on-Scalar Regression," *Journal of the American Statistical Association*, 116, 1994–2008.

Li, X., Yu, S., Wang, Y., Wang, G., Wang, L., and Lai, M.-J. (2023), "Nonparametric Regression for 3D Point Cloud Learning," `arXiv:2106.04255`.

Luckett, D. J., Laber, E. B., Kahkoska, A. R., Maahs, D. M., Mayer-Davis, E., and Kosorok, M. R. (2020), "Estimating dynamic treatment regimes in mobile health using V-learning," *Journal of the American Statistical Association*, 115, 692–706.

McKeague, I. W. and Qian, M. (2014), "Estimation of treatment policies based on functional predictors," *Statistica Sinica*, 24, 1461–1485.

Mu, J., Wang, G., and Wang, L. (2018), "Estimation and inference in spatially varying coefficient models," *Environmetrics*, 29, e2485.

Nathoo, F. S., Kong, L., Zhu, H., and Initiative, A. D. N. (2019), "A review of statistical methods in imaging genetics," *Canadian Journal of Statistics*, 47, 108–131.

Park, H., Petkova, E., Tarpey, T., and Ogden, R. T. (2021), "Functional additive models for optimizing individualized treatment rules," *Biometrics*, 1–14.

Pelham, W. E. and Fabiano, G. A. (2008), "Evidence-based psychosocial treatments for attention-deficit/hyperactivity disorder," *Journal of Clinical Child & Adolescent Psychology*, 37, 184–214.

Petersen, R. C., Aisen, P., Beckett, L. A., Donohue, M., Gamst, A., Harvey, D. J., Jack, C., Jagust, W., Shaw, L., Toga, A., and Trojanowski, J. (2010), "Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization," *Neurology*, 74, 201–209.

Qian, M. and Murphy, S. A. (2011), "Performance guarantees for individualized treatment rules," *Annals of statistics*, 39, 1180.

Ramsay, T. (2002), "Spline smoothing over difficult regions." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 307–319.

Robins, J. M. (2004), "Optimal Structural Nested Models for Optimal Sequential Decisions," in *Proceedings of the Second Seattle Symposium in Biostatistics*, Springer, pp. 189–326.

Rosenbaum, P. R. (1984), "From association to causation in observational studies: The role of tests of strongly ignorable treatment assignment," *Journal of the American Statistical Association*, 79, 41–48.

Rosenbaum, P. R. and Rubin, D. B. (1983), "Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome," *Journal of the Royal Statistical Society: Series B (Methodological)*, 45, 212–218.

Schneider, L. S., Insel, P. S., Weiner, M. W., Initiative, A. D. N., et al. (2011), "Treatment with cholinesterase inhibitors and memantine of patients in the Alzheimer's Disease Neuroimaging Initiative," *Archives of neurology*, 68, 58–66.

Stern, Y. (2012), "Cognitive reserve in ageing and Alzheimer's disease," *The Lancet Neurology*, 11, 1006–1012.

Su, Y.-R., Di, C.-Z., and Hsu, L. (2017), "Hypothesis testing in functional linear models," *Biometrics*, 73, 551–561.

Thall, P. F., Wooten, L. H., Logothetis, C. J., Millikan, R. E., and Tannir, N. M. (2007), "Bayesian and frequentist two-stage treatment strategies based on sequential failure times subject to interval censoring," *Statistics in medicine*, 26, 4687–4702.

Wang, G., Wang, L., Lai, M.-J., Kim, M., Li, X., Mu, J., Wang, Y., and Yu, S. (2019), "BPST: Bivariate Spline over Triangulation," R package version 1.0, `https://github.com/funstatpackages/BPST`.

Wang, H. and Ranalli, M. G. (2007), "Low-rank smoothing splines on complicated domains." *Biometrics*, 63, 209–217.

Wang, L. and Lai, M.-J. (2019), "Triangulation," R package version 1.0, `https://github.com/funstatpackages/Triangulation`.

Weiner, M. W. and Veitch, D. P. (2015), "Introduction to special issue: overview of Alzheimer's Disease Neuroimaging Initiative," *Alzheimer's & Dementia*, 11, 730–733.

Wood, N. S., Bravington, V. M., and Hedley, L. S. (2008), "Soap film smoothing." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70, 931–955.

Yu, S., Wang, G., Wang, L., and Yang, L. (2021), "Multivariate spline estimation and inference for image-on-scalar regression," *Statistica Sinica*, 31, 1463–1487.

Zhao, Y., Zeng, D., Socinski, M. A., and Kosorok, M. R. (2011), "Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer," *Biometrics*, 67, 1422–1433.

Zhu, H., Fan, J., and Kong, L. (2014), "Spatially varying coefficient model for neuroimaging data with jump discontinuities," *Journal of the American Statistical Association*, 109, 1084–1098.

Zhu, H., Li, T., and Zhao, B. (2022), "Statistical learning methods for neuroimaging data analysis with applications," .