DIMENSION REDUCTION OF CONDITIONAL QUANTILES FOR FUNCTIONAL DATA WITH ITS EXTENSION

by

Shanshan Wang

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2026

Approved by:

Dr. Eliana Christou

Dr. Hee Cheol Chung

Dr. Wenyu Gao

Dr. Monica Johar

ABSTRACT

SHANSHAN WANG. Dimension Reduction of Conditional Quantiles for Functional Data with Its Extension. (Under the direction of DR. ELIANA CHRISTOU)

Functional data analysis (FDA) has become increasingly important in modern applications where observations are recorded as functions over a continuum, such as neuroimaging signals, electrocardiograms, and other biomedical measurements. While most existing FDA methodologies focus on modeling the conditional mean, relatively little attention has been given to modeling conditional quantiles. Moreover, the infinite-dimensional nature of functional predictors necessitates effective dimension reduction techniques. This dissertation develops a unified framework for dimension reduction in the conditional quantile analysis of functional data.

First, we propose a dimension reduction methodology for the conditional quantiles of functional data with functional predictors. The proposed approach replaces the infinite-dimensional functional predictors with a small number of sufficient predictors while preserving the information relevant to the conditional quantiles. We establish the convergence rates of the proposed estimators and evaluate their finite-sample performance through simulation studies and an application to a functional magnetic resonance imaging (fMRI) dataset.

Second, we extend this framework to accommodate both functional and categorical predictors. Existing approaches for quantile-based dimension reduction primarily focus on quantitative predictors. To address this limitation, we introduce a partial dimension reduction method that incorporates categorical variables together with functional predictors. We present an efficient estimation algorithm, derive the convergence rates of the estimators, and demonstrate the performance of the proposed method through simulations and an application to fMRI data.

Third, we further generalize the proposed framework to nonlinear dimension reduc-

tion for conditional quantiles of functional data. While the previous approaches rely on linear dimension reduction structures, the proposed nonlinear framework captures more complex relationships between the response and functional predictors. The method constructs two nested functional spaces: a Hilbert space representing the functional predictors and a reproducing kernel Hilbert space that captures nonlinear structure. Based on this framework, we propose a new estimator, called the τth functional generalized central quantile subspace (τ-fGCQS), and establish its convergence rate. The performance of the method is illustrated through simulations and applications to health-related datasets, including studies of ADHD, Parkinson's disease, and body mass index (BMI).

Overall, this dissertation advances the methodology of functional data analysis by developing a series of dimension reduction techniques for conditional quantiles, progressively extending from linear models with functional predictors to mixed predictors and nonlinear structures.

DEDICATION

To my parents, Xiujuan Wang and Yongan Wang, my beloved Su Xu, and my grandmother Mingqin Li; and in loving memory of my grandparents, Shuqin Wang, Lianhe Wang, and Xianfa Wang.

ACKNOWLEDGEMENTS

First, I would like to express my deepest gratitude to my Ph.D. advisor, Dr. Eliana Christou, for all of her support, guidance, and encouragement over the past five years. I have always felt that one of the luckiest parts of my Ph.D. journey was having her as my advisor. She treated me not only as a student, but also with the warmth and care of a friend. Whenever I felt frustrated, stressed, or uncertain, I could always count on her encouragement and support. I am truly grateful for everything she has done for me, both academically and personally.

I would also like to sincerely thank my committee members, Dr. Wenyu Gao, Dr. Hee Cheol Chung, and Dr. Monica Johar, for their time, support, and valuable feedback on my dissertation. I am also grateful to the faculty members in the department who have helped me throughout my Ph.D. journey, especially Dr. Shaozhong Deng, Dr. Shaoyu Li, and Dr. Duan Chen. Their support, encouragement, and guidance have meant a great deal to me.

I am also grateful for the financial support that made my doctoral studies possible. I sincerely thank the Graduate School for its continued support, including the Graduate School Summer Fellowship (GSSF). I also acknowledge support from the National Science Foundation through grant DMS-2213140.

Finally, I would like to thank my family and friends for their love, support, and constant presence throughout this journey. I would not be where I am today without the unconditional love and support of my parents. I am especially grateful to my boyfriend, Su Xu, for being by my side for the past eight years, through almost my entire academic journey. I also want to thank my friend Suxian Zhou for her companionship throughout my Ph.D., and my friend Mengyuan Yang for her friendship and support over so many years. Last but not least, I am grateful for my dear dogs, PP and Bobby, who brought me so much comfort and joy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADHD	Attention Deficit Hyperactivity Disorder
BMI	Body Mass Index
FCPQS	Functional Central Partial Quantile Subspace
FCQS	Functional Central Quantile Subspace
FDA	Functional Data Analysis
fMRI	Functional Magnetic Resonance Imaging
FSIR	Functional Sliced Inverse Regression
MCMR	Multiple Correlation of Multivariate Rank
NHANES	National Health and Nutrition Examination Survey
QR	Quantile Regression
RMSE	Root Mean Squared Error
SDR	Sufficient Dimension Reduction

CHAPTER 1: INTRODUCTION

1.1 Functional Data Analysis (FDA)

Functional Data Analysis (FDA) has received increasing attention in recent years and affects various fields such as finance [1], natural language processing (NLP) [2], electric grid stabilization [3, 4], and notably the medical field [5, 6]. For example, neuroimaging studies on disorders such as attention deficit hyperactivity disorder (ADHD) or Alzheimer's disease aim to understand neural development in neuropsychiatric and neurodegenerative conditions, substance use, and brain structure and function. These studies often collect functional data such as functional magnetic resonance imaging (fMRI) and electroencephalogram (EEG) signals. While such data have traditionally been analyzed using conventional methods, FDA offers a more precise representation over a continuum.

FDA was first introduced by Ramsay [7] and further developed by Ramsay and Dalzell [8]. Its growing popularity is reflected in the adaptation of numerous classical statistical methods to the functional data setting. For example, functional linear mean regression has been extensively investigated in the works of [9, 10, 11, 12, 13, 14, 15, 16, 17], among others. In addition, nonparametric models for functional data have been applied in a variety of fields, including chemometrics, meteorology, speech recognition, and medicine; see [18, 19, 20, 21, 22]. Other extensions include models with multiple functional predictors [23, 24], as well as generalized linear models with functional predictors [25, 26, 27, 28].

However, one of the main challenges in FDA arises from the fact that functional data live in infinite-dimensional spaces, which necessitates the use of *dimension reduction techniques*. Traditional approaches developed for finite-dimensional settings

must therefore be adapted to accommodate the complexities of functional predictors. Several methods have been proposed to extend classical sufficient dimension reduction techniques to the functional setting. For instance, Ferré and Yao [29] extended sliced inverse regression (SIR) [30] to functional sliced inverse regression (FSIR), and later introduced functional inverse regression (FIR) using kernel smoothing [31]. Similarly, Lian and Li [32] extended sliced average variance estimation (SAVE; [33]) to functional data, while Wang, Lin and Zhang [34] extended contour regression [35] to functional contour regression (FCR). More recently, Li and Song [36] proposed functional dimension reduction techniques for situations where both the response and predictors are random functions, and Solea et al. [37] introduced a robust version of FSIR for multivariate elliptical functional data. A brief overview of dimension reduction techniques is provided in Section 1.3.

While mean regression is widely used in FDA, it provides only a limited view of the data distribution and may fail to capture variability across different quantiles. In contrast, *quantile regression (QR)* offers a more comprehensive perspective by modeling different parts of the conditional distribution, which is particularly useful when focusing on tails, rare events, or heteroscedastic models. However, QR for functional data remains relatively underexplored. For functional linear QR, Cardot et al. [38] used smoothing splines, while Kato [39] applied functional principal component analysis (FPCA). Nonparametric approaches via inversion of the conditional distribution were considered by [40] and [41]. Other related works include functional partially linear QR [42, 43, 44], functional quadratic QR [45], functional linear semiparametric models [46], and generalized regression quantiles [47]. A brief overview of quantile regression is provided in Section 1.4.

1.2 Basics on Hilbert Spaces

In this section, we briefly review the Hilbert space concepts and notation that will be used throughout the dissertation. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $\mathcal{H}$ a

separable Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and induced norm $\|\cdot\|_{\mathcal{H}}$.

Single functional predictor. If $U : \Omega \to \mathcal{H}$ is a measurable function with respect to the Borel σ-field $\mathcal{B}$, then U is called a *random element* in $\mathcal{H}$. In this dissertation, $\mathcal{H}$ will typically be a Hilbert space of functions of time, and U is viewed as a random function. If $E\|U\|_{\mathcal{H}} < \infty$, the linear functional $\mathcal{H} \to \mathbb{R}$, $f \to E\langle f, U \rangle_{\mathcal{H}}$ is bounded. By Riesz's representation theorem, there exists a unique *mean element* $\mu_U = E(U) \in \mathcal{H}$, such that

$$\langle \mu_U, f \rangle_{\mathcal{H}} = E\langle f, U \rangle_{\mathcal{H}}, \text{ for all } f \in \mathcal{H}. \tag{1.1}$$

If $E\|U\|_{\mathcal{H}}^2 < \infty$, the covariance operator of U is defined by

$$\Sigma_{UU} = E[\{U - E(U)\} \otimes \{U - E(U)\}] = E(U \otimes U) - \{E(U)\} \otimes \{E(U)\}$$

where $\otimes$ denotes the tensor product on $\mathcal{H}$. For $f, g \in \mathcal{H}$, the tensor operator associates to any $h \in \mathcal{H}$ the element $(f \otimes g)(h) = f\langle g, h \rangle_{\mathcal{H}}$. Under the assumption $E\|U\|_{\mathcal{H}}^2 < \infty$, Σ_{UU} is a trace-class operator that satisfies the spectral decomposition $\sum_{r=1}^{\infty} \delta_r \phi_r \otimes \phi_r$, where $\{\delta_r\}_{r \geq 1}$ are the eigenvalues satisfying $\delta_1 \geq \delta_2 \geq \cdots \geq 0$, and $\{\phi_r\}_{r \geq 1}$ are the eigenfunctions forming an orthonormal basis in $\mathcal{H}$. Then, the Karhunen-Loève expansion of $U - \mu_U$ is

$$U - \mu_U = \sum_{r=1}^{\infty} \delta_r^{1/2} \xi_r \phi_r, \tag{1.2}$$

where $\xi_r = \delta_r^{-1/2} \langle U - \mu_U, \phi_r \rangle_{\mathcal{H}}$, $r = 1, 2, \ldots$ are zero mean, unit variance, and uncorrelated random variables, called the scores [48].

A special case of a Hilbert space is $L_2(T)$, the space of square integrable real functions which comprises of a collection of Lebesgue measurable real-valued functions

f such that $\int_T f^2(t)dt < \infty$, where T is a bounded closed interval in $\mathbb{R}$. The inner product is then defined by $\langle f, g\rangle_{\mathcal{H}} = \int_T f(t)g(t)dt$ with a norm $\|f\|_{\mathcal{H}} = \sqrt{\langle f, f\rangle} = \{\int_T f^2(t)dt\}^{1/2}$. Moreover, for a random element $U \in L_2(T)$, the mean value is $E(U)(t) = E\{U(t)\}$ and the covariance operator is an integral operator with kernel $k_{UU}(s,t) = Cov\{U(s), U(t)\} = E[\{U(s) - EU(s)\}\{U(t) - EU(t)\}]$.

Multivariate functional predictors. For multivariate functional predictors, we extend the above formulation to the case of vector-valued random functions. Let $\mathscr{H} = \bigoplus_{i=1}^p \mathcal{H}_i$ be the direct sum of $\mathcal{H}_1, \ldots, \mathcal{H}_p$, i.e., the Cartesian product $\mathcal{H}_1 \times \cdots \times \mathcal{H}_p$. A member $f \in \bigoplus_{i=1}^p \mathcal{H}_i$ is $f = (f_1, \ldots, f_p)$, where $f_i \in \mathcal{H}_i$. Then, for $f, g \in \bigoplus_{i=1}^p \mathcal{H}_i$, the inner product is $\langle f, g\rangle_{\oplus\mathcal{H}} = \sum_{i=1}^p \langle f_i, g_i\rangle_{\mathcal{H}_i}$.

Let $\mathbb{U} = (U^1, \ldots, U^p)$ be a random element in $\bigoplus_{i=1}^p \mathcal{H}_i$, where U^i, $i = 1, \ldots, p$, is the ith component of $\mathbb{U}$. The mean of $\mathbb{U}$ is $\mu_{\mathbb{U}} = (\mu_{U^1}, \ldots, \mu_{U^p})$, where μ_{U^i} is the mean of U^i, as defined in (1.1). Moreover, for $i, j = 1, \ldots, p$, the covariance operator between U^i and U^j is $\Sigma_{U^iU^j} = E\{(U^i - \mu_{U^i}) \otimes (U^j - \mu_{U^j})\}$. Note that, $\Sigma_{U^iU^j} \in \mathscr{B}(\mathcal{H}_j, \mathcal{H}_i)$, where $\mathscr{B}(\mathcal{H}_j, \mathcal{H}_i)$ denotes the set of all bounded operators from $\mathcal{H}_j$ to $\mathcal{H}_i$. Then, $\Sigma_{\mathbb{U}\mathbb{U}}$ is defined as the $p \times p$ matrix whose (i,j)th entry is $\Sigma_{U^iU^j}$ and is linear, self-adjoint, positive semi-definite and compact. Its decomposition is $\Sigma_{\mathbb{U}\mathbb{U}} = \sum_{r=1}^\infty \gamma_r \psi_r \otimes \psi_r$, where $\{\gamma_r\}_{r\geq 1}$ are the eigenvalues satisfying $\gamma_1 \geq \gamma_2 \geq \cdots \geq 0$, and $\{\psi_r\}_{r\geq 1}$ are the eigenfunctions forming an orthonormal basis in $\bigoplus_{i=1}^p \mathcal{H}_i$. The Karhunen-Loève expansion of $\mathbb{U} - \mu_{\mathbb{U}}$ is

$$\mathbb{U} - \mu_{\mathbb{U}} = \sum_{r=1}^{\infty} \gamma_r^{1/2} \rho_r \psi_r, \tag{1.3}$$

where $\rho_r = \gamma_r^{-1/2}\langle \mathbb{U} - \mu_{\mathbb{U}}, \psi_r\rangle_{\oplus\mathcal{H}}$, $r = 1, 2, \ldots$, are zero mean, unit variance, and uncorrelated random variables [24].

We conclude this section with several additional notations and terminologies that will be used throughout the dissertation. For two Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, denote by $\mathscr{B}(\mathcal{H},\mathcal{K})$ the set of all bounded operators from $\mathcal{H}$ to $\mathcal{K}$; if $\mathcal{H}=\mathcal{K}$, we abbreviate $\mathscr{B}(\mathcal{H},\mathcal{H})$ by $\mathscr{B}(\mathcal{H})$. Then, for a linear operator $A\in\mathscr{B}(\mathcal{H},\mathcal{K})$, define A^* to be the adjoint operator of A, $\ker(A)=\{h\in\mathcal{H}: A(h)=0\}$ to be the kernel of A, $\mathrm{ran}(A)=\{A(h): h\in\mathcal{H}\}$ to be the range of A, and $\overline{\mathrm{ran}}(A)$ to be the closure of $\mathrm{ran}(A)$. Moreover, for a self-adjoint operator $A\in\mathscr{B}(\mathcal{H})$, let $A|_{\ker(A)^\perp}:\ker(A)^\perp\to\mathcal{H}$ be the restriction of A to $\ker(A)^\perp$. Then, the operator $A^\dagger:\mathrm{ran}(A)\to\ker(A)^\perp$ that maps each $g\in\mathrm{ran}(A)$ to the unique element $f\in\ker(A)^\perp$ such that $Af=g$ is called the Moore–Penrose inverse of A (see [49, Definition 3.5.7]). If, in addition, A is positive semi-definite, then, for any $\alpha>0$, we define $A^{\dagger\alpha}=(A^\alpha)^\dagger$.

1.3 Sufficient Dimension Reduction (SDR)

This section provides a brief overview of sufficient dimension reduction (SDR) in both linear and nonlinear settings. SDR is a regression-based approach for reducing the dimension of the predictor space without losing information about the relationship between the response and the predictors. In each case, we consider both scalar-on-scalar and scalar-on-function problems.

1.3.1 Linear Sufficient Dimension Reduction

We begin with the linear SDR framework in the classical scalar-on-scalar setting and then extend the discussion to the scalar-on-function case.

Scalar-on-Scalar. Let the response variable $Y:\Omega\to\mathbb{R}$ and the $p\times 1$ predictor vector $\mathbf{X}:\Omega\to\mathbb{R}^p$ are scalar-valued. Assume that there exists a $p\times d$ matrix $\mathbf{A}$, with $d\le p$, such that

$$Y \perp\!\!\!\perp \mathbf{X} \mid \mathbf{A}^\top\mathbf{X}. \tag{1.4}$$

That is, Y and $\mathbf{X}$ are independent given $\mathbf{A}^\top\mathbf{X}$. This implies that the original predictor $\mathbf{X}$ can be replaced by the reduced $d \times 1$ vector $\mathbf{A}^\top\mathbf{X}$ without loss of information about the regression of Y on $\mathbf{X}$. The space spanned by $\mathbf{A}$ is called the *dimension reduction subspace* for the regression of Y on $\mathbf{X}$, and the greatest dimension reduction is achieved using the smallest dimension reduction subspace, called the *central subspace*, which is denoted by $\mathcal{S}_{Y|\mathbf{X}}$. Classical methods for estimating the central subspace including sliced inverse regression (SIR, [30]), sliced average variance estimation (SAVE; [33]), coutour regression ([35]), among others.

Scalar-on-Function. The concept of linear SDR was extended to the functional setting by Li and Song [36]. Specifically, for $i = 1, \ldots, p$, let $\mathcal{H}_i$ be a separable Hilbert space of real-valued functions on T, where T is a bounded closed interval in $\mathbb{R}$. Let $Y : \Omega \to \mathbb{R}$ be a univariate response and $X = (X^1, \ldots, X^p) : \Omega \to \bigoplus_{i=1}^p \mathcal{H}_i$ be a random element. Assume a finite rank linear operator $L \in \mathscr{B}(\bigoplus_{i=1}^p \mathcal{H}_i, \mathbb{R}^d)$, $d \in \mathbb{N}$, such that

$$Y \perp\!\!\!\perp X | L(X). \tag{1.5}$$

Then, $\overline{\text{ran}}(L^*)$ is called the *functional dimension reduction subspace*. Li and Song [36] noted that (1.5) is consistent with the formulation of Ferré and Yao [29]. By Riesz's representation theorem, there exists $f_1, \ldots, f_d \in \bigoplus_{i=1}^p \mathcal{H}_i$ such that

$$L(X) = (L_1(X), \ldots, L_d(X)) = (\langle f_1, X\rangle_{\oplus\mathcal{H}}, \ldots, \langle f_d, X\rangle_{\oplus\mathcal{H}})$$

Hence, for any $v \in \mathbb{R}^d$, $v^\top L(X) = \sum_{j=1}^d v_j \langle f_j, X\rangle_{\oplus\mathcal{H}} = \langle \sum_{j=1}^d v_j f_j, X\rangle_{\oplus\mathcal{H}}$. Then, $L^*(v) = \sum_{j=1}^d v_j f_j$ and $\overline{\text{ran}}(L^*) = \text{span}(f_1, \ldots, f_d)$.

Define $\cap\{\overline{\text{ran}}(L^*) : L \text{ satisfies (1.5)}\}$, which we assume also satisfies (1.5). The assumption that this intersection is a functional dimension reduction subspace can

be proved similarly to Proposition 6.4 of [50] for the classical setting and [51] for the functional data setting. The condition is mild and taken for granted without further development. We call this intersection the *functional central subspace* (FCS), denoted by $\mathscr{S}_{Y|X}$. We assume $\mathscr{S}_{Y|X}$ exists and satisfies (1.5), and that $L \in \mathscr{B}(\bigoplus_{i=1}^{p} \mathcal{H}_i, \mathbb{R}^d)$ denotes an operator such that $\mathscr{S}_{Y|X} = \overline{\text{ran}}(L^*)$.

Ferré and Yao [29] proposed FSIR and proved that, under model (1.5) and Assumption 1.1 given below, $E(X|Y) - E(X)$ belongs to $\Sigma_{XX}\mathscr{S}_{Y|X}$, where Σ_{XX} is the covariance operator of X.

Assumption 1.1. There is a bounded linear operator $\Lambda : \text{ran}(L) \to \bigoplus_{i=1}^{p} \mathcal{H}_i$ such that $E\{X|L(X)\} = \Lambda L(X)$.

1.3.2 Nonlinear Sufficient Dimension Reduction

More recently, sufficient dimension reduction has been generalized to nonlinear settings. In contrast to linear SDR, which seeks low-dimensional linear combinations of the predictors, nonlinear SDR relaxes the linearity assumption and aims to recover more general predictive structures. This broader framework is better suited for capturing complex relationships between the response and the predictors. We briefly review the nonlinear formulation in both scalar-on-scalar and scalar-on-function settings.

Scalar-on-Scalar. A general theory for *nonlinear sufficient dimension reduction* was developed by Li et al. [52] and Lee et al. [53]. Specifically, let $Y \in \mathbb{R}$ and $X \in \mathbb{R}^p$. Let $\mathcal{F}_X$ be the Borel σ-field generated by open sets in $\mathbb{R}^p$, and let $\sigma(X)$ represent the sub σ-field in $\mathcal{F}$ generated by X, that is $\sigma(X) = X^{-1}(\mathcal{F}_X)$.

A sub σ-field $\mathcal{G} \subseteq \sigma(X)$ is called an SDR σ-field for Y versus X if and only if

$$Y \perp\!\!\!\perp X|\mathcal{G}$$

Among all such sub σ-fields, the intersection is called the *central σ-field*. This general-

izes the notion of the central subspace from linear SDR and provides the foundation for nonlinear dimension reduction. Based on this formulation, Lee, Li, and Chiaromonte [53] further introduced the corresponding *central class*, which serves as the nonlinear analogue of the central subspace.

Scalar-on-function. Li and Song [54] extended the nonlinear SDR framework of Lee et al. [53] to the functional setting. Specifically, let $Y \in \mathbb{R}$ be a scalar response, and let $X = (X^1, \ldots, X^p) : \Omega \to \bigoplus_{i=1}^{p} \mathcal{H}_i$ is a random element, where each $\mathcal{H}_i$ is a Hilbert space of real-valued functions defined on an interval $T \subseteq \mathbb{R}$. The goal is to identify a set of nonlinear functions $f_1, \ldots, f_d$ of X such that

$$Y \perp\!\!\!\perp X | f_1(X), \ldots, f_d(X)$$

To accommodate this framework, they considered two nested functional spaces. First, the predictor X takes values in a Hilbert space of functions defined on T. Second, the nonlinear sufficient predictors $f_1, \ldots, f_d$ belong to another functional space whose domain is the first predictor space. In this sense, the nonlinear reduction is defined not directly on the original time domain, but through functions acting on the functional predictor itself.

To formalize this idea, Li and Song [54] introduced a reproducing kernel Hilbert space associated with the predictor space and used it to characterize the nonlinear reduction structure. This framework provides a natural extension of nonlinear SDR from multivariate predictors to functional data and leads to a functional analogue of the central class.

1.4 Quantile Regression (QR)

Quantile Regression (QR) was first introduced by Koenker and Bassett [55] and has since received a lot of attention. For a univariate response Y and a $p \times 1$ predictor vector $\mathbf{X}$, and for $\tau \in (0, 1)$, let $Q_\tau(Y|\mathbf{x}) = \inf\{y : P(Y \leq y|\mathbf{X} = \mathbf{x}) \geq \tau\}$ denote the

τth conditional quantile of Y given $\mathbf{X} = \mathbf{x}$, which satisfies

$$Q_\tau(Y|\mathbf{x}) = \arg\min_q E\{\rho_\tau(Y-q)|\mathbf{X}=\mathbf{x}\}$$

where $\rho_\tau(u) = \{\tau - I(u<0)\}u$ is the loss function, commonly referred to as the *check function*.

For heteroscedastic data, interest often lies in modeling different parts of the conditional distribution rather than only the conditional mean. In such cases, conditional quantile regression provides a useful alternative. In analogy with classical sufficient dimension reduction, we seek a $p \times d_\tau$ matrix $\mathbf{B}_\tau$, with $d_\tau \leq p$, such that

$$Y \perp\!\!\!\perp Q_\tau(Y|\mathbf{X})|\mathbf{B}_\tau^\top \mathbf{X} \tag{1.6}$$

The space spanned by $\mathbf{B}_\tau$ is called the *τth quantile dimension reduction subspace* for the regression of Y on $\mathbf{X}$, and the smallest such space is called the *τth central quantile subspace* (τ-CQS), which is obtained by taking the intersection of all the τ-CQS get above. Methods for estimating the τ-CQS include [56], [57], [58], [59].

The setting above is referred to as linear quantile dimension reduction because the reduced predictor is represented through the linear combination $\mathbf{B}_\tau^\top \mathbf{X}$. However, nonlinear dimension reduction has the potential to capture more general predictive structures and may further reduce dimensionality. To this end, Christou et al. [60] extended the above framework to nonlinear functions in a Hilbert space $\mathcal{M}$, such that

$$Y \perp\!\!\!\perp Q_\tau(Y|\mathbf{X})|f_{\tau,1}(\mathbf{X}), \ldots, f_{\tau,\widetilde{d}_\tau}(\mathbf{X}).$$

This formulation generalizes the linear quantile dimension reduction framework and provides a more flexible approach for modeling complex relationships between the response and the predictors.

1.5 Overview of the Chapter

The discussion above highlights two main challenges and gaps in dimension reduction techniques for functional data. Most existing methods focus primarily on mean regression and therefore provide only a limited view of the conditional distribution. In addition, relatively little attention has been given to quantile-based dimension reduction methods for functional data, especially in more complex settings involving mixed-type predictors or nonlinear dependence structures. Motivated by these gaps, this dissertation studies dimension reduction for the conditional quantiles of functional data and its extension, beginning with the linear setting for functional predictors, then moving to settings involving both functional and categorical predictors, and finally considering a nonlinear dimension reduction framework for functional data.

Chapter 2 develops a dimension reduction methodology for the conditional quantiles of functional data and introduces the *τth functional central quantile subspace* (τ-FCQS). The chapter studies the estimation of the τ-FCQS, establishes the convergence rates of the proposed estimators, and investigates their finite-sample performance through simulation studies and an application to a functional magnetic resonance imaging (fMRI) dataset.

Chapter 3 extends the framework of Chapter 2 to the setting where both functional and categorical predictors are present and introduces the *τth functional central partial quantile subspace* (τ-FCPQS). The resulting methodology provides a partial dimension reduction approach for conditional quantiles with mixed-type predictors. The chapter presents an estimation algorithm, derives the convergence rates of the resulting estimators, and demonstrates the performance of the proposed method through simulation studies and a real-data application.

Chapter 4 further generalizes the previous developments to a nonlinear setting under a single-index model, thereby relaxing the linear assumption used in Chapters

2 and 3. The proposed methodology is formulated through a Hilbert space together with a reproducing kernel Hilbert space, allowing more complex relationships between the response and the functional predictors to be captured. The chapter introduces a new estimator, called the *τth functional generalized central quantile subspace* (τ-fGCQS), establishes its convergence rate, and investigates its performance through simulation studies and real-data applications involving several health indicators, including ADHD, Parkinson's disease, and body mass index (BMI).

CHAPTER 2: The τth FUNCTIONAL CENTRAL QUANTILE SUBSPACE (τ-FCQS)

2.1 The Methodology of τ-FCQS

For each $i = 1, \ldots, p$, $\mathcal{H}_i$ is a separable Hilbert space of real-valued functions on T, which is a bounded closed interval in $\mathbb{R}$. Let $Y : \Omega \to \mathbb{R}$ be a univariate response and $X = (X^1, \ldots, X^p) : \Omega \to \bigoplus_{i=1}^p \mathcal{H}_i$ be a random element, such that

Assumption 2.1. $E \|X\|_{\mathscr{H}}^2 < \infty$,

where, from now on, $\mathscr{H}$ denotes $\bigoplus_{i=1}^p \mathcal{H}_i$, i.e., $\mathscr{H} = \bigoplus_{i=1}^p \mathcal{H}_i$.

Definition 2.1. For a finite rank linear operator $L_\tau \in \mathscr{B}(\mathscr{H}, \mathbb{R}^{d_\tau})$, $d_\tau \in \mathbb{N}$, such that

$$Y \perp\!\!\!\perp Q_\tau(Y|X)|L_\tau(X), \tag{2.1}$$

$\overline{\text{ran}}(L_\tau^*)$ is called the *τth functional quantile dimension reduction subspace*.

This formulation is equivalent to assuming that there exist functional parameters $\beta_{\tau,1}, \ldots, \beta_{\tau,d_\tau} \in \mathscr{H}$, such that $Y \perp\!\!\!\perp Q_\tau(Y|X)|\langle \beta_{\tau,1}, X\rangle_{\mathscr{H}}, \ldots, \langle \beta_{\tau,d_\tau}, X\rangle_{\mathscr{H}}$.

Definition 2.2. Under the assumption that $\cap\{\overline{\text{ran}}(L_\tau^*) : L_\tau \text{ satisfies (2.1)}\}$ satisfies (2.1), we call it the *τth functional central quantile subspace* (τ-FCQS) and denote it by $\mathscr{S}_{Q_\tau(Y|X)}$.

Remark 2.1. The assumption that the intersection of all τth functional quantile dimension reduction subspaces is itself a functional quantile dimension reduction subspace is mild and can be proven similarly to the functional dimension reduction subspace case. Henceforth, we assume that $\mathscr{S}_{Q_\tau(Y|X)}$ exists, satisfies (2.1), and that $L_\tau \in \mathscr{B}(\mathscr{H}, \mathbb{R}^{d_\tau})$ is an operator such that $\mathscr{S}_{Q_\tau(Y|X)} = \overline{\text{ran}}(L_\tau^*)$.

Remark 2.2. It is evident that $\mathscr{S}_{Q_\tau(Y|X)} \subseteq \mathscr{S}_{Y|X}$, for any τ. Thus, the τ-FCQS can offer additional dimension reduction when these subspaces differ. For example, Model I from Section 2.5.1 is a case where $\mathscr{S}_{Q_\tau(Y|X)} = \mathscr{S}_{Y|X} = \text{span}\{\beta_1\}$, for all τ. However, in Model IV, $\mathscr{S}_{Q_\tau(Y|X)} = \text{span}\{\beta_1 + Q_\tau(\varepsilon)\beta_2\}$ for any τ, while $\mathscr{S}_{Y|X} = \text{span}\{\beta_1, \beta_2\}$; see Section 2.5.1 for details.

Assumption 1.1 is necessary for the FCS. However, when focusing on a specific conditional quantile, the assumption must hold for each quantile level τ under consideration. This leads to the following modification.

Assumption 2.2. For a given τ, there is a bounded linear operator $\Lambda_\tau : \text{ran}(L_\tau) \to \mathscr{H}$ such that $E\{X|L_\tau(X)\} = \Lambda_\tau L_\tau(X)$. That is, for any function $b_\tau \in \mathscr{H}$ there exists constants $c_{\tau 0}, c_{\tau 1}, \ldots, c_{\tau d_\tau}$, such that

$$E(\langle b_\tau, X\rangle_{\mathscr{H}} | \langle \beta_{\tau 1}, X\rangle_{\mathscr{H}}, \ldots, \langle \beta_{\tau d_\tau}, X\rangle_{\mathscr{H}}) = c_{\tau 0} + c_{\tau 1}\langle \beta_{\tau 1}, X\rangle_{\mathscr{H}} + \cdots + c_{\tau d_\tau}\langle \beta_{\tau d_\tau}, X\rangle_{\mathscr{H}}.$$

The following theorem is analogous to a known result in the classical setting (see, e.g. [50], page 57) and is necessary to ensure that the functional parameters derived in Theorems 2.2 and 2.3 belong to $\mathscr{S}_{Q_\tau(Y|X)}$. Essentially, Theorem 2.1 states that $E\{X|L_\tau(X)\}$ is equal to the projection of X onto the subspace spanned by $L_\tau(X)$.

Theorem 2.1. *Under Assumptions 2.1, 2.2, and $E(X) = 0$, we have $E\{X|L_\tau(X)\} = \Sigma_{XX} L_\tau^*(L_\tau \Sigma_{XX} L_\tau^*)^\dagger L_\tau(X)$.*

Proof. See Appendix A.4.1 □

We now focus on retrieving $\beta_{\tau,1}, \ldots, \beta_{\tau,d_\tau}$, such that $\mathscr{S}_{Q_\tau(Y|X)} = \text{span}\{\beta_{\tau,1}, \ldots, \beta_{\tau,d_\tau}\}$. Our methodology relies on two main theorems. Theorem 2.3 shows that given one functional parameter b_τ in $\mathscr{S}_{Q_\tau(Y|X)}$, we can construct another one using $E(\langle b_\tau, X\rangle_{\mathscr{H}} X)$. Since the τ-FCQS is spanned by d_τ functional parameters, we can use Theorem 2.3 to construct as many as needed. However, Theorem 2.3 requires an initial b_τ to start the

iterative process. Theorem 2.2 provides a starting point by showing that one functional parameter can be extracted through minimizing $\arg\min_{(a_\tau,b_\tau)} E\{Q_\tau(Y|X) - a_\tau - \langle b_\tau, X\rangle_{\mathscr{H}}\}^2$, for $a_\tau \in \mathbb{R}$ and $b_\tau \in \mathscr{H}$. Note that, in practice, this minimization problem requires estimating $Q_\tau(Y|X)$, which is challenging due to the infinite-dimensional nature of X. Therefore, since $Y \perp\!\!\!\perp X|L(X)$, we perform an initial dimension reduction by replacing X with $L(X)$.

Theorem 2.2. *For a given $\tau \in (0,1)$, assume that $Y \perp\!\!\!\perp Q_\tau(Y|X)|L_\tau(X)$, where $L_\tau \in \mathscr{B}(\mathscr{H}, \mathbb{R}^{d_\tau})$ is such that $\mathscr{S}_{Q_\tau(Y|X)} = \overline{\text{ran}}(L_\tau^*)$. Under Assumptions 2.1 and 2.2, and if*

$$(\alpha_\tau^*, \beta_\tau^*) = \arg\min_{(a_\tau,b_\tau)} E[Q_\tau\{Y|L(X)\} - a_\tau - \langle b_\tau, X\rangle_{\mathscr{H}}]^2, \tag{2.2}$$

where $L \in \mathscr{B}(\mathscr{H}, \mathbb{R}^d)$ is such that $\mathscr{S}_{Y|X} = \overline{\text{ran}}(L^*)$, then $\beta_\tau^* \in \mathscr{S}_{Q_\tau(Y|X)}$.

Proof. See Appendix A.4.2 □

Theorem 2.2 retrieves one functional parameter such that $Y \perp\!\!\!\perp Q_\tau(Y|X)|\langle\beta_\tau^*, X\rangle_{\mathscr{H}}$. If a single-index functional QR model is assumed ($d_\tau = 1$), then $\beta_\tau^* = \beta_{\tau,1}$, and the process ends. However, if $d_\tau > 1$, Theorem 2.2 is insufficient, and more functional parameters are required, provided by the next theorem. We note that the idea for constructing these additional parameters originates from [61]; see their Theorem 3.

Theorem 2.3. *For a given $\tau \in (0,1)$, assume that $Y \perp\!\!\!\perp Q_\tau(Y|X)|L_\tau(X)$, where $L_\tau \in \mathscr{B}(\mathscr{H}, \mathbb{R}^{d_\tau})$ is such that $\mathscr{S}_{Q_\tau(Y|X)} = \overline{\text{ran}}(L_\tau^*)$. Under Assumptions 2.1, 2.2, and the assumption that $b_\tau \in \mathscr{S}_{Q_\tau(Y|X)}$, then*

$$E(\langle b_\tau, X\rangle_{\mathscr{H}} X) \in \Sigma_{XX}\mathscr{S}_{Q_\tau(Y|X)}.$$

Proof. See Appendix A.4.3 □

2.2 Estimation of τ-FCQS

2.2.1 Population Level

Theorem 2.3 suggests a method for constructing the τ-FCQS. Specifically, if $b_{\tau,0} \in \mathscr{S}_{Q_\tau(Y|X)}$, then, for $j = 1, \ldots,$ we can construct additional functional parameters in $\mathscr{S}_{Q_\tau(Y|X)}$ using

$$E(\langle b_{\tau,j-1}, X\rangle_{\mathscr{H}} X). \tag{2.3}$$

Therefore, it is enough to find an initial $b_{\tau 0}$. For that, we can use $b_{\tau 0} = \beta^*_\tau$. Then, the procedure will be as follows:

1. Set $b_{\tau 0} = \beta^*_\tau$, where β^*_τ is given in (2.2).
2. Choose an integer m and, for $j = 1, \ldots, m$, sequentially form $b_{\tau j} = E(\langle b_{\tau j-1}, X\rangle_{\mathscr{H}} X)$.
3. Let $B_\tau = \sum_{j=0}^{m} b_{\tau j} \otimes b_{\tau j}$.
4. Let $v_{\tau 1}, \ldots, v_{\tau d_\tau}$ be the set of solutions to the generalized eigenvalue problem $B_\tau v_{\tau j} = \lambda \Sigma_{XX} v_{\tau j}$, $j = 1, \ldots, d_\tau$, then $v_{\tau j} \in \mathscr{S}_{Q_\tau(Y|X)}$. This is because, for $j = 1, \ldots, d_\tau$, $v_{\tau j} = \lambda^{-1} \Sigma^\dagger_{XX} B_\tau v_{\tau j} \in \Sigma^\dagger_{XX} \text{span}\{b_{\tau 0}, \ldots, b_{\tau m}\} \subseteq \mathscr{S}_{Q_\tau(Y|X)}$.

Note that, an eigenvector v of a generalized eigenvalue problem $Av = \lambda Bv$ implies that $v = B^{-1/2}u$, where u is an eigenvector of $B^{-1/2}AB^{-1/2}$. Thus, the eigenvectors $v_{\tau,j}$, $j = 1, \ldots, d_\tau$, of Step 4 that satisfy

$$\begin{aligned} &\text{argmax} \quad \langle v_\tau, B_\tau v_\tau\rangle_{\mathscr{H}} \\ &\text{subject to} \quad v_\tau \in \mathscr{H}, \langle v_\tau, \Sigma_{XX} v_\tau\rangle_{\mathscr{H}} = 1, \langle v_\tau, \Sigma_{XX} v_{\tau,j}\rangle_{\mathscr{H}} = 0, j = 1, \ldots, d_\tau - 1, \end{aligned}$$

can be expressed as $v_{\tau,j} = \Sigma_{XX}^{\dagger 1/2}\eta_{\tau,j}$, leading to the following problem

$$\begin{aligned} &\text{argmax} \quad \langle \eta_\tau, \Sigma_{XX}^{\dagger 1/2} B_\tau \Sigma_{XX}^{\dagger 1/2} \eta_\tau \rangle_{\mathscr{H}} \\ &\text{subject to} \quad \eta_\tau \in \mathscr{H}, \langle \eta_\tau, \eta_\tau \rangle_{\mathscr{H}} = 1, \langle \eta_\tau, \eta_{\tau,j} \rangle_{\mathscr{H}} = 0, j = 1, \ldots, d_\tau - 1. \end{aligned} \tag{2.4}$$

Remark 2.3. Step 2 of the above algorithm requires to choose an integer m. We tried different values during the simulation studies and concluded that the algorithm is not sensitive to the choice of m. For this work, we decided to use $m = pq - 1$, where p is the number of predictor variables and q is the number of basis functions used to approximate the functional predictors.

Remark 2.4. In this part, we assume the structural dimension d_τ is known. However, in practice, it is unknown and must be estimated. One approach is the cross-validation Bayesian information criterion (CVBIC) from [54]; details and simulations are provided in Supplementary of [62].

2.2.2 Sample Level

We now derive the sample estimates of the expressions (2.2) and (2.3) when the functions are fully observed. Specifically, for $u = 1, \ldots, n$, let Y_u be an independent and identically distributed (iid) sample from Y, and $X_1, \ldots, X_n$ be an independent sample from the random element $X = (X^1, \ldots, X^p)$, with $X_u = (X_u^1, \ldots, X_u^p)^\top$.

To achieve (4.5), we , we need to use a standard dimension reduction technique to replace X with $L(X)$. For that, we apply the FSIR method from [29] and replace X with the new d-dimensional predictor vector $\widehat{L}(X)$; see [29] for details. Then, we use the data $\{Y_u, X_u\}_{u=1}^n$ to to estimate β_τ^* using

$$(\widehat{\alpha}_\tau, \widehat{\beta}_\tau) = \arg\min_{(a_\tau, b_\tau)} \sum_{u=1}^{n} [\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\} - a_\tau - \langle b_\tau, X_u \rangle_{\mathscr{H}}]^2, \tag{2.5}$$

where $\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}$ is a nonparametric estimate of $Q_\tau\{Y|\widehat{L}(X_u)\}$. For that we use

the local linear conditional quantile estimation method of [63], where $\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\} = \widehat{q}_\tau\{\widehat{L}(X_u)\}$ and

$$(\widehat{q}_\tau\{\widehat{L}(X_u)\}, \widehat{\mathbf{s}}_\tau\{\widehat{L}(X_u)\}) = \\ \arg\min_{(q_\tau, \mathbf{s}_\tau)} \sum_{k=1}^{n} \rho_\tau\Big[Y_k - q_\tau - \mathbf{s}_\tau^\top\{\widehat{L}(X_k) - \widehat{L}(X_u)\}\Big] K\left\{\frac{\widehat{L}(X_k) - \widehat{L}(X_u)}{h}\right\} \quad (2.6)$$

Note that $K(\cdot)$ is a d-dimensional kernel function, and $h > 0$ is the bandwidth. We use a Gaussian kernel and choose the bandwidth h according to the rule-of-thumb in [64]. Specifically, $h = h_m[\tau(1-\tau)/[\phi\{\Phi^{-1}(\tau)\}]^2]^{1/5}$, where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the probability density and cumulative distribution functions of the standard normal distribution, respectively, and h_m is the optimal bandwidth for local mean regression.

Next, to achieve (2.3), we set $\widehat{\beta}_{\tau,0} = \widehat{\beta}_\tau$, and, for $j = 1, \ldots, m$, we form

$$\widehat{\beta}_{\tau,j} = n^{-1} \sum_{u=1}^{n} \langle \widehat{\beta}_{\tau,j-1}, X_u \rangle_{\mathscr{H}} X_u. \quad (2.7)$$

After obtaining $\widehat{\beta}_{\tau,0}, \widehat{\beta}_{\tau,1}, \ldots, \widehat{\beta}_{\tau,m}$, set $\widehat{B}_\tau = \sum_{j=0}^{m} \widehat{\beta}_{\tau,j} \otimes \widehat{\beta}_{\tau,j}$ and focus on the eigenvalue problem (2.4), which implies solving

$$\Sigma_{XX}^{\dagger 1/2} B_\tau \Sigma_{XX}^{\dagger 1/2} \eta_\tau = \lambda \eta_\tau, \quad v_\tau = \Sigma_{XX}^{\dagger 1/2} \eta_\tau.$$

At the sample level, we estimate Σ_{XX} using the $p \times p$ matrix $\widehat{\Sigma}_{XX}$, where $\widehat{\Sigma}_{X^i X^j} = E_n[\{X^i - E_n(X^i)\} \otimes \{X^j - E_n(X^j)\}]$, $i, j = 1, \ldots, p$, and $\Sigma_{XX}^{\dagger 1/2}$ using the regularized inverse $(\widehat{\Sigma}_{XX} + \epsilon_n I_p)^{-1/2}$, where $\{\epsilon_n\}_{n \geq 1}$ is a sequence of positive numbers such that $\epsilon_n \to 0$ as $n \to \infty$, and I_p is the $p \times p$ identity matrix. Therefore, we want to compute the first d_τ eigenfunctions $\widehat{\eta}_{\tau 1}, \ldots, \widehat{\eta}_{\tau d_\tau}$ of

$$\widehat{M}_\tau = (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} \widehat{B}_\tau (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} \quad (2.8)$$

and transform back to the eigenfunctions $\widehat{v}_{\tau,j} = (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}\widehat{\eta}_{\tau,j}$ and the sufficient predictors $\langle \widehat{v}_{\tau,j}, X\rangle_{\mathscr{H}}$, $j = 1, \ldots, d_\tau$.

2.3 Asymptotic Theory

We will start by deriving the consistency and convergence rate of $\widehat{M}_\tau$, where the population counterparts M_τ and $\eta_{\tau 1}, \ldots, \eta_{\tau d_\tau}$ are defined as $M_\tau = \Sigma_{XX}^{\dagger 1/2} B_\tau \Sigma_{XX}^{\dagger 1/2}$ and $\eta_{\tau 1}, \ldots, \eta_{\tau d_\tau}$ as the first d_τ eigenvectors of M_τ. Before proceeding with the main results, we need to make one more assumption, commonly used in the literature; see, for example, Assumption 7 of [36].

Assumption 2.3. $\mathscr{S}_{Q_\tau(Y|X)} \subseteq \overline{\text{ran}}(\Sigma_{XX})$.

This assumption is not restrictive and suggests that the τ-FCQS is identifiable up to the range of Σ_{XX}. This is because if $\beta_\tau \perp \overline{\text{ran}}(\Sigma_{XX})$, then $Var(\langle \beta_\tau, X\rangle_{\mathscr{H}}) = 0$, implying β_τ is orthogonal to the support of $X - E(X)$.

Theorem 2.4. *Let Assumptions 2.1, 2.2, 3.2, and S1-S6 from the Appendix A.1 hold. Then, if $E(X) = 0$, $\widehat{L}(X)$ is consistent of the directions of the FCS, and $n^{-1/4} \prec \epsilon_n \prec 1$, for a given $\tau \in (0,1)$, $\widehat{M}_\tau$, given in (2.8), is a consistent estimate of M_τ, and*

$$\left\| \widehat{M}_\tau - M_\tau \right\| = O_p(n^{-1/2}\epsilon_n^{-2}) + O(\epsilon_n^{1/2}),$$

where $M_\tau = \Sigma_{XX}^{\dagger 1/2} B_\tau \Sigma_{XX}^{\dagger 1/2}$, and $\|\cdot\|$ is the operator norm.

Proof. See Appendix A.4.4 □

Next, the first d_τ eigenvalues and eigenfunctions of $\widehat{M}_\tau$ converge to those of M_τ at the same rate, as in Corollary 2 of [36].

Corollary 2.1. *Under the assumptions of Theorem 2.4, and for a given $\tau \in (0,1)$*

and $j = 1, \ldots, d_\tau$,

$$\widehat{\lambda}_j - \lambda_j = O_p(n^{-1/2}\epsilon_n^{-2}) + O(\epsilon_n^{1/2}),$$
$$\|\widehat{\eta}_{\tau,j} - \eta_{\tau,j}\|_{\mathscr{H}} = O_p(n^{-1/2}\epsilon_n^{-2}) + O(\epsilon_n^{1/2}),$$

where $\widehat{\lambda}_j$ *and* $\widehat{\eta}_{\tau,j}$ *are the eigenvalues and eigenfunctions of* $\widehat{M}_\tau$*, respectively, and* λ_j *and* $\eta_{\tau,j}$ *are the eigenvalues and eigenfunctions of* M_τ*, respectively.*

Proof. The proof uses similar steps as those of Corollary 2 of [36] and thus, it is omitted. □

The eigenfunctions of interest are $\widehat{v}_{\tau,j} = (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}\widehat{\eta}_{\tau,j}$, and the next theorem gives the convergence rates of $\widehat{v}_{\tau,j}$ and of the predictors $\langle \widehat{v}_{\tau,j}, X\rangle_{\mathscr{H}}$, $j = 1, \ldots, d_\tau$.

Theorem 2.5. *Let Assumptions 2.1, 2.2, 3.2, and S1-S6 from the Appendix A.1 hold. Then, if* $E(X) = 0$*,* $\widehat{L}(X)$ *is consistent of the directions of the FCS, and* $n^{-1/5} \prec \epsilon_n \prec 1$*, for a given* $\tau \in (0, 1)$ *and for* $j = 1, \ldots, d_\tau$*,*

$$\|\widehat{v}_{\tau,j} - v_{\tau,j}\|_{\mathscr{H}} = O_p(n^{-1/2}\epsilon_n^{-5/2}) + O(\epsilon_n^{1/2}),$$
$$\langle \widehat{v}_{\tau,j}, X\rangle_{\mathscr{H}} - \langle v_{\tau,j}, X\rangle_{\mathscr{H}} = O_p(n^{-1/2}\epsilon_n^{-5/2}) + O(\epsilon_n^{1/2}),$$

where $\widehat{v}_{\tau,j} = (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}\widehat{\eta}_{\tau,j}$ *and* $v_{\tau,j} = \Sigma_{XX}^{\dagger 1/2}\eta_{\tau,j}$.

2.4 Implementation

In practice, the functions $X_u(t)$, $u = 1, \ldots, n$, are observed at a finite set of points, $t_{u1}, \ldots, t_{uN_u}$, and need to be estimated using the observed data $\{(t, X_u(t)) : t = t_{u1}, \ldots, t_{uN_u}\}$. Commonly used methods for estimating $X_u(t)$ are based on the coordinate representation with a finite number of basis functions, such as B-splines. This will convert a function to a vector representation, a linear operator to a matrix representation, and an eigenfuction problem to an eigenvector problem. Preliminaries

on coordinate representation are provided in Section A.5. We omit indices like ${}_{\mathcal{G}_j}[\cdot]_{\mathcal{G}_i}$ and $[\cdot]_{\mathcal{G}_i}$, as the bases for the coordinates can be identified by the operators' domain and range. In this section, square brackets are reserved for coordinates.

For $i = 1, \ldots, p$, assume that $\mathcal{H}_i$ is spanned by a finite set of basis functions $\mathcal{G}_i = \{g_1^i, \ldots, g_{k_n}^i\}$, such that each X_u^i, $u = 1, \ldots, n$, can be approximated by $[X_u^i]^\top g_{1:k_n}^i$. Although $\widehat{X}_u^i$ could be used for clarity, we omit this distinction for simplicity. Let $Q_n = I_n - n^{-1} 1_n 1_n^\top$ be the projection onto the orthogonal complement of the subspace spanned by 1_n, where I_n is the $n \times n$ identity matrix and 1_n is the n-dimensional vector of ones, and let $G = \mathrm{diag}(G_i : i = 1, \ldots, p) \in \mathbb{R}^{pk_n \times pk_n}$, where G_i is the Gram matrix of $\mathcal{H}_i$. Then, $[X_u] = ([X_u^1]^\top, \ldots, [X_u^p]^\top)^\top \in \mathbb{R}^{pk_n}$ is the coordinate representation of X_u, $u = 1, \ldots, n$, i.e., the ith block is the coordinate representation of X_u^i. Therefore, the matrix $[X_{1:n}]$ is the $pk_n \times n$ matrix $([X_1], \ldots, [X_n])$.

The coordinate representations of problems (2.5) and (2.7) with respect to $\mathcal{G} = \bigoplus_{i=1}^p \mathcal{G}_i$ are given by

$$[(\widehat{\alpha}_\tau, \widehat{\beta}_\tau)] = (\widehat{\alpha}_\tau, [\widehat{\beta}_\tau]) = \arg\min_{(a_\tau, b_\tau)} \sum_{u=1}^n [\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\} - a_\tau - [b_\tau]^\top G[X_u]]^2, \tag{2.9}$$

where $\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\} = \widehat{q}_\tau\{\widehat{L}(X_u)\}$ and

$$\begin{aligned}
&(\widehat{q}_\tau\{\widehat{L}(X_u)\}, \widehat{\mathbf{s}}_\tau\{\widehat{L}(X_u)\}) \\
&= \arg\min_{(q_\tau, \mathbf{s}_\tau)} \sum_{k=1}^n \rho_\tau\Big[Y_k - q_\tau - \mathbf{s}_\tau^\top\{\widehat{L}(X_k) - \widehat{L}(X_u)\}\Big] K\left\{\frac{\widehat{L}(X_k) - \widehat{L}(X_u)}{h}\right\}
\end{aligned} \tag{2.10}$$

and, for $j = 1, \ldots, m$,

$$[\widehat{\beta}_{\tau,j}] = n^{-1} \sum_{u=1}^n [\widehat{\beta}_{\tau,j-1}]^\top G[X_u][X_u], \tag{2.11}$$

where $\widehat{\beta}_{\tau,0} = \widehat{\beta}_\tau$ from (2.9). Then, for $\widehat{B}_\tau = \sum_{j=0}^m \widehat{\beta}_{\tau,j} \otimes \widehat{\beta}_{\tau,j}$, we have

$$[\widehat{B}_\tau] = \sum_{j=0}^m [\widehat{\beta}_{\tau j} \otimes \widehat{\beta}_{\tau j}] = \sum_{j=0}^m [\widehat{\beta}_{\tau j}][\widehat{\beta}_{\tau j}]^\top G = [\widehat{\beta}_{\tau 0:m}][\widehat{\beta}_{\tau 0:m}]^\top G$$

and the eigenvalue problem (2.4) becomes

$$\begin{aligned} &\text{argmax} \quad [\eta_\tau]^\top G[\widehat{\Sigma}_{XX}^{\dagger 1/2}][\widehat{\beta}_{\tau,0:m}][\widehat{\beta}_{\tau,0:m}]^\top G[\widehat{\Sigma}_{XX}^{\dagger 1/2}][\eta_\tau] \\ &\text{subject to} \quad [\eta_\tau]^\top G[\eta_\tau] = 1, [\eta_\tau]^\top G[\eta_{\tau,j}] = 0, j = 1, \ldots, d_{\tau-1}. \end{aligned}$$

Let $\omega_\tau = G^{1/2}[\eta_\tau]$, then we get the eigenvector problem

$$\begin{aligned} &\text{argmax} \quad \omega_\tau^\top G^{1/2}[\widehat{\Sigma}_{XX}^{\dagger 1/2}][\widehat{\beta}_{\tau,0:m}][\widehat{\beta}_{\tau,0:m}]^\top G[\widehat{\Sigma}_{XX}^{\dagger 1/2}] G^{\dagger 1/2}\omega_\tau \\ &\text{subject to} \quad \omega_\tau^\top \omega_\tau = 1, \omega_\tau^\top \omega_{\tau,j} = 0, j = 1, \ldots, d_{\tau-1}. \end{aligned}$$

Using $[\widehat{\Sigma}_{XX}^{\dagger 1/2}] = G^{\dagger 1/2}\{n^{-1}G^{1/2}([X_{1:n}]Q_n[X_{1:n}]^\top)G^{1/2}\}^{\dagger 1/2}G^{1/2}$, proven in Appendix A.4.6, and setting $A = \{n^{-1}G^{1/2}([X_{1:n}]Q_n[X_{1:n}]^\top)G^{1/2}\}^{\dagger 1/2}$, we get

$$\begin{aligned} &\text{argmax} \quad \omega_\tau^\top A G^{1/2}[\widehat{\beta}_{\tau,0:m}][\widehat{\beta}_{\tau,0:m}]^\top G^{1/2} A\omega_\tau \\ &\text{subject to} \quad \omega_\tau^\top \omega_\tau = 1, \omega_\tau^\top \omega_{\tau,j} = 0, j = 1, \ldots, d_{\tau-1}. \end{aligned} \tag{2.12}$$

Then, $[\widehat{v}_\tau] = [\widehat{\Sigma}_{XX}^{\dagger 1/2}]G^{\dagger 1/2}\omega_\tau = G^{\dagger 1/2}A\omega_\tau$ and $\langle X_u, \widehat{v}_\tau\rangle_{\mathscr{H}} = [X_u]^\top G[\widehat{v}_\tau] = [X_u]^\top G^{1/2}A\omega_\tau$, $u = 1, \ldots, n$, and the algorithm can be summarized as follows:

Algorithm. Let $\{Y_u, (t, X_u(t)) : t = t_{u1}, \ldots, t_{uN_u}\}_{u=1}^n$ iid sample.

(i) For each $u = 1, \ldots, n$, obtain the coordinates of X_u relative to the basis $\mathcal{G} = \bigoplus_{i=1}^p \mathcal{G}_i$ of $\mathscr{H}$ and derive the gram matrix G.

(ii) Use FSIR of [29] to compute the d-dimensional predictors $\widehat{L}(X_u)$, $u = 1, \ldots, n$.

(iii) For each $u = 1, \ldots, n$, estimate $Q_\tau\{Y|\widehat{L}(X_u)\}$ using the local linear condi-

tional quantile estimation method of [63]. Specifically, take $\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\} = \widehat{q}_\tau\{\widehat{L}(X_u)\}$, where $\widehat{q}_\tau\{\widehat{L}(X_u)\}$ is obtained by (2.10).

(iv) Compute $[\widehat{\beta}_\tau]$ according to (2.9) and set $[\widehat{\beta}_{\tau 0}] = [\widehat{\beta}_\tau]$.

(v) For $j = 1, \ldots, m$, compute $[\widehat{\beta}_{\tau j}]$ according to (2.11).

(vi) Compute the matrix $A = \{n^{-1}G^{1/2}([X_{1:n}]Q_n[X_{1:n}]^\top)G^{1/2}\}^{\dagger 1/2}$.

(vii) Solve the eigenvalue problem (2.12) and obtain the d_τ solutions $\omega_{\tau j}$, $j = 1, \ldots, d_\tau$.

(viii) Obtain $[\widehat{v}_{\tau j}] = G^{\dagger 1/2}A\omega_{\tau j}$ and the sufficient predictors $\langle X_u, \widehat{v}_{\tau j}\rangle_{\mathscr{H}} = [X_u]^\top G^{1/2}A\omega_{\tau j}$, $j = 1, \ldots, d_\tau$, $u = 1, \ldots, n$.

2.5 Simulation Studies

2.5.1 Computational Remarks

Algorithm and parameters. Step 1 of the algorithm requires to obtain the coordinates of X_u, $u = 1, \ldots, n$. For that, we use a B-spline basis with $q = 4$ bases. Step 2 of the algorithm uses the FSIR method of [29] for a preliminary dimension reduction. For that, we use $H = 10$ number of slices. Step 3 of the algorithm performs a local linear conditional quantile estimation, where the kernel and bandwidth are chosen as described in Section 2.2.2. Finally, Step 5 of the algorithm uses m to create additional directions, which was discussed in Remark 2.3, and is chosen as $pq - 1$.

Simulation Setting. The simulation setting is an extension of [65] to multivariate data. First, consider the time interval $[0, 1]$ and create a grid of points by 0.01, resulting in 101 equally spaced time points. Then, for $i = 1, \ldots, p$ and $u = 1, \ldots, n$, we generate $X_u^i(t)$ according to the univariate Karhunen-Loève expansion (1.2). Specifically, we assume $\mu_{X^i} = 0$ and use $X_u^i(t) = \sum_{r=1}^4 \xi_{ur}^i \phi_r^i(t)$, $t \in [0, 1]$, where $\phi_1^i(t) = \sqrt{2}\sin(2\pi t)$, $\phi_2^i(t) = \sqrt{2}\cos(2\pi t)$, $\phi_3^i(t) = \sqrt{2}\sin(4\pi t)$, and $\phi_4^i(t) = \sqrt{2}\cos(4\pi t)$, and ξ_{ur}^i are mutually independent random variables with zero mean and variance $var(\xi_{ur}^i) = \lambda_r$, $r = 1, \ldots, 4$, with $\lambda_1 = 2, \lambda_2 = 1, \lambda_3 = 1/2, \lambda_4 = 1/4$.

To extend the simulation process to the multivariate functional data $X_u(t) = (X_u^1(t), \ldots, X_u^p(t))$ using the Karhunen-Loève expansion (1.3), we follow Proposition 5 of [24] and use the multivariate FPCA eigenfunctions through an orthogonalization of the univariate eigenfunctions. Specifically, for $i = 1, \ldots, p$ and $u = 1, \ldots, n$, let $\boldsymbol{\xi}_u^i = (\xi_{u1}^i, \ldots, \xi_{u4}^i)^\top$ and $\boldsymbol{\xi}_u = ((\boldsymbol{\xi}_u^1)^\top, \ldots, (\boldsymbol{\xi}_u^p)^\top)^\top$. Moreover, define $\mathbb{Z} \in \mathbb{R}^{4p \times 4p}$ to be the covariance matrix of the univariate FPCA scores $\boldsymbol{\xi}_u$ with (j, j')th entry the matrix $\mathbb{Z}^{jj'} = cov(\boldsymbol{\xi}_u^j, \boldsymbol{\xi}_u^{j'}) \in \mathbb{R}^{4\times 4}$. Then, the kth eigenfunction $\boldsymbol{\psi}_k(t) = (\psi_k^1(t), \ldots, \psi_k^p(t))^\top$ of Σ_{XX} is defined by $\psi_k^i(t) = \boldsymbol{\phi}^i(t)^\top \boldsymbol{z}_k^i$, $k = 1, \ldots, 4p$, where $\boldsymbol{\phi}^i(t) = (\phi_1^i(t), \ldots, \phi_4^i(t))^\top$ and $\boldsymbol{z}_k^i = (z_{k1}^i, \ldots, z_{k4}^i)^\top$ denotes the ith block of the eigenvector $\boldsymbol{z}_k$ of $\mathbb{Z}$. Finally, the scores are $\rho_{uk} = \sum_{i=1}^p \sum_{r=1}^4 z_{kr}^i \xi_{ur}^i$, $k = 1, \ldots, 4p$ and $u = 1, \ldots, n$, where the coordinate-wise scores ξ_{ur}^i are standard normal random variables.

For the estimation accuracy, we use the multiple correlation between the true and estimated predictors; see [36]. Specifically, for U and V random vectors of the same dimension d, define C_{UU}, C_{UV}, and C_{VV} the sample covariance matrices. Then, the multiple correlation between U and V is defined by

$$mcorr(U, V) = tr(C_{VV}^{-1/2} C_{VU} C_{UU}^{-1} C_{UV} C_{VV}^{-1/2}),$$

and it results in a number between 0 and d. A number closer to d indicates better performance.

All simulation results are based on $N = 100$ iterations. Unless otherwise stated, the sample size is chosen to be $n = 400$, the number of functional predictors is chosen to be $p = 5$, the quantiles under consideration are $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$, and the dimension d_τ is assumed to be known.

Models under consideration. For the simulation studies we consider the following

single- and multi-index models.

$$\text{M-I:} \quad Y = \frac{1}{0.5 + (\langle \beta_1, X \rangle_{\mathscr{H}} + 1)^2} + 0.2\varepsilon, \text{ where } \mathscr{S}_{Q_\tau(Y|X)} = \text{span}\{\beta_1\}$$

$$\text{M-II:} \quad Y = \arctan(\pi \langle \beta_1, X \rangle_{\mathscr{H}}/2) + \varepsilon, \text{ where } \mathscr{S}_{Q_\tau(Y|X)} = \text{span}\{\beta_1\}$$

$$\text{M-III:} \quad Y = \exp(\langle \beta_1, X \rangle_{\mathscr{H}})\varepsilon, \text{ where } \mathscr{S}_{Q_\tau(Y|X)} = \text{span}\{\beta_1\}$$

$$\text{M-IV:} \quad Y = \langle \beta_1, X \rangle_{\mathscr{H}} + \langle \beta_2, X \rangle_{\mathscr{H}}\varepsilon, \text{ where } \mathscr{S}_{Q_\tau(Y|X)} = \text{span}\{\beta_1 + Q_\tau(\varepsilon)\beta_2\}$$

$$\text{M-V:} \quad Y = \arctan(\pi \langle \beta_1, X \rangle_{\mathscr{H}}) + 0.5 \sin(\pi \langle \beta_2, X \rangle_{\mathscr{H}}/6) + 0.1\varepsilon, \text{ where } \mathscr{S}_{Q_\tau(Y|X)} = \text{span}\{\beta_1, \beta_2\}$$

$$\text{M-VI:} \quad Y = \langle \beta_1, X \rangle_{\mathscr{H}} + \langle \beta_2, X \rangle_{\mathscr{H}} + \sqrt{0.5 + \langle \beta_1, X \rangle^2_{\mathscr{H}} + \langle \beta_2, X \rangle^2_{\mathscr{H}}}\varepsilon,$$
$$\text{where } \mathscr{S}_{Q_\tau(Y|X)} = \text{span}\{\beta_1, \beta_2\}$$

$$\text{M-VII:} \quad Y = \langle \beta_1, X \rangle^3_{\mathscr{H}} + \exp(\langle \beta_2, X \rangle_{\mathscr{H}}) + \langle \beta_3, X \rangle_{\mathscr{H}}\varepsilon, \text{ where } \mathscr{S}_{Q_\tau(Y|X)} = \text{span}\{\beta_1, \beta_2, \beta_3\}$$

$$\text{M-VIII:} \quad Y = \langle \beta_1, X \rangle^3_{\mathscr{H}} + \langle \beta_2, X \rangle_{\mathscr{H}} + \langle \beta_3, X \rangle_{\mathscr{H}}\varepsilon, \text{ where } \mathscr{S}_{Q_\tau(Y|X)} = \text{span}\{\beta_1, \beta_2 + Q_\tau(\varepsilon)\beta_3\},$$

where $\beta_1(t) = \psi_1(t)$, $\beta_2(t) = \psi_2(t)$, and $\beta_3(t) = \psi_3(t)$ are the first three eigenfunctions of Σ_{XX}, X is simulated as described above, and the error ε is generated according to (1) standard normal ($\mathcal{N}$), (2) chi-square with three degrees of freedom ($\mathcal{X}^2_3$), and (3) exponential with rate of 0.5 (Exp(0.5)) distribution. We note that, all examples were considered during the simulation runs and the results are available. However, to save space, and since all models exhibit similar patterns, we only report selective ones in the next subsection.

2.5.2 Results

Example 1 - Effect of n and p. We begin by considering the performance of the algorithm for different choices of n and p. The data are generated according to Models I-VIII; however, we report the results for Model I. The sample size is given by $n = 200, 400$ or 1000, and the number of predictors is $p = 5, 10, 20$, or 40. Tables 2.1 - 2.3 report the observed, over the 100 simulation runs, means and standard deviations (in parenthesis) of the multiple correlation for various τ and the different

error distributions. As expected, the efficiency of the proposed methodology increases with n and decreases with p. Moreover, observe that the performance is robust to the quantile level and the error distribution.

Table 2.1: *Mean (and standard deviation) of multiple correlation for Model I, when* $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$ *and the error follows a standard normal distribution.*

n	p	0.1	0.25	0.5	0.75	0.9
200	5	0.9957 (0.0039)	0.9956 (0.0042)	0.9957 (0.0040)	0.9957 (0.0039)	0.9957 (0.0038)
	10	0.9912 (0.0105)	0.9911 (0.0108)	0.9913 (0.0106)	0.9913 (0.0105)	0.9910 (0.0105)
	20	0.9848 (0.0542)	0.9849 (0.0565)	0.9911 (0.0146)	0.9907 (0.0173)	0.9908 (0.0155)
	40	0.9692 (0.0827)	0.9680 (0.0825)	0.9722 (0.0657)	0.9726 (0.0698)	0.9538 (0.1210)
400	5	0.9964 (0.0045)	0.9965 (0.0039)	0.9965 (0.0035)	0.9966 (0.0034)	0.9966 (0.0032)
	10	0.9948 (0.0048)	0.9948 (0.0047)	0.9949 (0.0047)	0.9949 (0.0046)	0.9949 (0.0047)
	20	0.9926 (0.0055)	0.9927 (0.0053)	0.9926 (0.0054)	0.9926 (0.0054)	0.9927 (0.0054)
	40	0.9898 (0.0092)	0.9897 (0.0095)	0.9892 (0.0137)	0.9898 (0.0092)	0.9897 (0.0097)
1000	5	0.9974 (0.0018)	0.9974 (0.0017)	0.9974 (0.0017)	0.9974 (0.0017)	0.9974 (0.0016)
	10	0.9966 (0.0027)	0.9966 (0.0027)	0.9966 (0.0026)	0.9966 (0.0026)	0.9966 (0.0026)
	20	0.9952 (0.0035)	0.9950 (0.0035)	0.9951 (0.0035)	0.9950 (0.0034)	0.9950 (0.0034)
	40	0.9935 (0.0045)	0.9936 (0.0044)	0.9935 (0.0044)	0.9935 (0.0045)	0.9934 (0.0046)

Table 2.2: *Mean (and standard deviation) of multiple correlation for Model I, when* $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$ *and the error follows a chi-square distribution.*

n	p	0.1	0.25	0.5	0.75	0.9
200	5	0.9945 (0.0061)	0.9944 (0.0064)	0.9945 (0.0061)	0.9945 (0.0061)	0.9944 (0.0062)
	10	0.9932 (0.0136)	0.9916 (0.0215)	0.9909 (0.0274)	0.9910 (0.0266)	0.9905 (0.0309)
	20	0.9732 (0.0804)	0.9805 (0.0386)	0.9837 (0.0233)	0.9801 (0.0590)	0.9782 (0.0705)
	40	0.9185 (0.2049)	0.9373 (0.1546)	0.9249 (0.1824)	0.9124 (0.2315)	0.9180 (0.1956)
400	5	0.9965 (0.0045)	0.9965 (0.0050)	0.9965 (0.0047)	0.9966 (0.0044)	0.9966 (0.0043)
	10	0.9950 (0.0060)	0.9950 (0.0061)	0.9950 (0.0059)	0.9951 (0.0056)	0.9949 (0.0060)
	20	0.9924 (0.0075)	0.9924 (0.0074)	0.9833 (0.0923)	0.9927 (0.0068)	0.9926 (0.0070)
	40	0.9867 (0.0169)	0.9847 (0.0246)	0.9855 (0.0270)	0.9849 (0.0318)	0.9844 (0.0366)
1000	5	0.9973 (0.0018)	0.9973 (0.0019)	0.9973 (0.0018)	0.9973 (0.0018)	0.9974 (0.0017)
	10	0.9959 (0.0032)	0.9960 (0.0032)	0.9960 (0.0031)	0.9960 (0.00310)	0.9960 (0.0031)
	20	0.9959 (0.0030)	0.9959 (0.0030)	0.9959 (0.0030)	0.9959 (0.0030)	0.9959 (0.0030)
	40	0.9936 (0.0038)	0.9936 (0.0037)	0.9936 (0.0037)	0.9936 (0.0037)	0.9936 (0.0037)

Example 2 - Performance of the algorithm. We now demonstrate the performance of the algorithm for all models under consideration. Table 2.4 reports the observed, over the 100 simulation runs, means and standard deviations (in parenthesis) of the multiple correlation for various τ and the different error distributions. We can see that the mean multiple correlation is really close to d_τ, which is 1 for Models I-IV, 2 for Models V, VI, and VIII, and 3 for Model VII.

Table 2.3: *Mean (and standard deviation) of multiple correlation for Model I, when* $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$ *and the error follows an exponential distribution.*

n	p	0.1	0.25	0.5	0.75	0.9
200	5	0.9953 (0.0046)	0.9955 (0.0043)	0.9955 (0.0044)	0.9955 (0.0047)	0.9953 (0.0055)
	10	0.9896 (0.0302)	0.9893 (0.0337)	0.9904 (0.0236)	0.9923 (0.0098)	0.9899 (0.0275)
	20	0.9790 (0.0597)	0.9819 (0.0445)	0.9789 (0.0684)	0.9765 (0.0722)	0.9746 (0.0825)
	40	0.9704 (0.0804)	0.9702 (0.0607)	0.9717 (0.0460)	0.9688 (0.0776)	0.9600 (0.1258)
400	5	0.9966 (0.0053)	0.9967 (0.0047)	0.9967 (0.0047)	0.9968 (0.0044)	0.9968 (0.0041)
	10	0.9943 (0.0078)	0.9943 (0.0084)	0.9943 (0.0082)	0.9942 (0.0091)	0.9943 (0.0083)
	20	0.9913 (0.0095)	0.9913 (0.0095)	0.9913 (0.0094)	0.9913 (0.0094)	0.9913 (0.0095)
	40	0.9887 (0.0110)	0.9892 (0.0098)	0.9890 (0.0099)	0.9881 (0.0131)	0.9888 (0.0110)
1000	5	0.9970 (0.0020)	0.9970 (0.0021)	0.9970 (0.0022)	0.9970 (0.0021)	0.9970 (0.0023)
	10	0.9961 (0.0034)	0.9961 (0.0034)	0.9961 (0.0032)	0.9961 (0.0032)	0.9962 (0.0032)
	20	0.9952 (0.0037)	0.9952 (0.0036)	0.9952 (0.0036)	0.9952 (0.0036)	0.9952 (0.0036)
	40	0.9931 (0.0043)	0.9931 (0.0044)	0.9931 (0.0044)	0.9931 (0.0044)	0.9931 (0.0045)

Table 2.4: *Mean (and standard deviation) of multiple correlation for Models I-VIII, when* $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$.

M	error	0.1	0.25	0.5	0.75	0.9
I	$\mathcal{N}$	0.9967 (0.0033)	0.9967 (0.0035)	0.9967 (0.0035)	0.9967 (0.0035)	0.9968 (0.0031)
	$\mathcal{X}_2^3$	0.9952 (0.0056	0.9953 (0.0054)	0.9952 (0.0053)	0.9953 (0.0053)	0.9954 (0.0052)
	Exp	0.9965 (0.0031)	0.9966 (0.0030)	0.9966 (0.0031)	0.9966 (0.0031)	0.9966 (0.0031)
II	$\mathcal{N}$	0.9975 (0.0016)	0.9975 (0.0016)	0.9975 (0.0017)	0.9975 (0.0016)	0.9975 (0.0016)
	$\mathcal{X}_2^3$	0.9961 (0.0048)	0.9961 (0.0047)	0.9961 (0.0045)	0.9961 (0.0046)	0.9961 (0.0043)
	Exp	0.9967 (0.0032)	0.9967 (0.0032)	0.9967 (0.0032)	0.9967 (0.0032)	0.9965 (0.0037)
III	$\mathcal{N}$	0.9975 (0.0020)	0.9973 (0.0031)	0.9955 (0.0082)	0.9971 (0.0038)	0.9975 (0.0018)
	$\mathcal{X}_2^3$	0.9970 (0.0024)	0.9970 (0.0022)	0.9971 (0.0020)	0.9971 (0.0020)	0.9971 (0.0021)
	Exp	0.9969 (0.0023)	0.9969 (0.0022)	0.9969 (0.0022)	0.9969 (0.0023)	0.9969 (0.0023)
IV	$\mathcal{N}$	0.6305 (0.0393)	0.8380 (0.0270)	0.9969 (0.0035)	0.8373 (0.0292)	0.6297 (0.0413)
	$\mathcal{X}_2^3$	0.9426 (0.0660)	0.9176 (0.0828)	0.7953 (0.1576)	0.7018 (0.1972)	0.6628 (0.2151)
	Exp	0.9340 (0.0991)	0.9604 (0.0346)	0.8205 (0.1053)	0.6547 (0.1619)	0.5882 (0.1979)
V	$\mathcal{N}$	1.9692 (0.0852)	1.9591 (0.1186)	1.9538 (0.1353)	1.9637 (0.0971)	1.9690 (0.0881)
	$\mathcal{X}_2^3$	1.9492 (0.1383)	1.9448 (0.1537)	1.9352 (0.1763)	1.9423 (0.1489)	1.9425 (0.1561)
	Exp	1.9679 (0.0869)	1.9629 (0.0967)	1.9625 (0.0829)	1.9619 (0.0931)	1.9688 (0.0813)
VI	$\mathcal{N}$	1.9802 (0.0425)	1.9808 (0.0404)	1.9805 (0.0410)	1.9793 (0.0438)	1.9751 (0.0496)
	$\mathcal{X}_2^3$	1.9549 (0.1063)	1.9569 (0.1006)	1.9549 (0.0884)	1.9269 (0.1210)	1.8588 (0.2145)
	Exp	1.9676 (0.1039)	1.9680 (0.1034)	1.9659 (0.1053)	1.9571 (0.1197)	1.9054 (0.1754)
VII	$\mathcal{N}$	2.3663 (0.3282)	2.4085 (0.3093)	2.4119 (0.3117)	2.4280 (0.3098)	2.4434 (0.3042)
	$\mathcal{X}_2^3$	2.8988 (0.1798)	2.9270 (0.1569)	2.9400 (0.1270)	2.9380 (0.1144)	2.9234 (0.1185)
	Exp	2.8644 (0.2028)	2.8925 (0.1689)	2.9173 (0.1231)	2.9084 (0.1369)	2.8836 (0.1693)
VIII	$\mathcal{N}$	1.3976 (0.1730)	1.6780 (0.1561)	1.9568 (0.0816)	1.6614 (0.1481)	1.3793 (0.1805)
	$\mathcal{X}_2^3$	1.604 (0.2456)	1.8709 (0.1626)	1.9040 (0.1342)	1.8619 (0.1563)	1.8475 (0.1684)
	Exp	1.4798 (0.2692)	1.7974 (0.1893)	1.9124 (0.1116)	1.8273 (0.1653)	1.7864 (0.2003)

Example 3 - Methods to compare. We now compare the performance of the proposed methodology with that of FSIR [29] and robust functional sliced inverse regression (R-FSIR) [37]. Since FSIR and R-FSIR focus on the functional central

subspace, whereas our method targets the functional central quantile subspace, we restrict our comparison to models for which the two subspaces coincide in order to ensure meaningful comparisons.

In these situations, the choice of τ is not critical. Although the simulations are conducted for all values of τ, we report the results only for $\tau = 0.5$. The data are generated according to Models I–III and V–VII.

Table 2.5 reports the means and standard deviations (in parentheses) of the multiple correlation, computed over 100 simulation runs, for the proposed method as well as for FSIR and R-FSIR. We observe that the proposed method performs best across all models and error distributions considered, except for Model III, where FSIR yields a slightly larger average multiple correlation when the error distribution is normal.

Furthermore, the difference between the three methods is particularly pronounced for multi-index models, where the proposed method outperforms the competing approaches by a substantial margin.

2.6 Application

In this section, we apply the proposed methodology to the ADHD-200 fMRI data containing resting-state fMRI (rs-fMRI) signals obtained from the New York University Child Study Center from 116 patients with ADHD defined as predominantly hyperactivity/impulsive ($n = 2$), predominantly inattentive ($n = 44$) and combined (ADHD-C; $n = 77$), and 99 controls. The data are publicly available from the ADHD-200 Consortium (`http://fcon_1000.projects.nitrc.org/indi/adhd200/index.html`). For our analysis, we use the subjects in the ADHD-C group, after removing five subjects due to significant amount of missing observations, resulting in $n = 72$. Pre-processing of the fMRI data was done by the Neuro Bureau organization using the Athena pipeline (`http://www.theneurobureau.org/`). The 116 regions of interest were constructed for the pre-processed rs-fMRI based on the automated anatomical labeling atlas (AAL) developed by Craddock et al. (2012). Then, for each

Table 2.5: *Mean (and standard deviation) of multiple correlation for Models I-VII, for FSIR, R-FSIR, and* 0.5*-FCQS and various error distributions.*

M	error	FSIR	R-FSIR	0.5-FCQS
I	$\mathcal{N}$	0.9652 (0.0138)	0.9623 (0.0156)	**0.9968** (0.0033)
	$\mathcal{X}_3^2$	0.9414 (0.0263)	0.9151 (0.0458)	**0.9964** (0.0027)
	Exp	0.9532 (0.0207)	0.9373 (0.0315)	**0.9961** (0.0059)
II	$\mathcal{N}$	0.9821 (0.0071)	0.9829 (0.0070)	**0.9972** (0.0019)
	$\mathcal{X}_3^2$	0.9519 (0.0207)	0.8963 (0.0803)	**0.9967** (0.0030)
	Exp	0.9703 (0.0124)	0.9568 (0.0217)	**0.9968** (0.0032)
III	$\mathcal{N}$	**0.9900** (0.0039)	0.9898 (0.0036)	0.9872 (0.0555)
	$\mathcal{X}_3^2$	0.9926 (0.0028)	0.9926 (0.0031)	**0.9972** (0.0020)
	Exp	0.9887 (0.0046)	0.9874 (0.0057)	**0.9969** (0.0032)
V	$\mathcal{N}$	1.9287 (0.0235)	1.9300 (0.0242)	**1.9666** (0.0606)
	$\mathcal{X}_3^2$	1.9009 (0.0414)	1.8945 (0.0428)	**1.9476** (0.1281)
	Exp	1.9141 (0.0298)	1.9122 (0.0336)	**1.9626** (0.0685)
VI	$\mathcal{N}$	1.0634 (0.1199)	1.0213 (0.0735)	**1.9856** (0.0243)
	$\mathcal{X}_3^2$	1.0116 (0.1265)	0.9403 (0.1256)	**1.9563** (0.0778)
	Exp	1.0603 (0.1446)	1.0003 (0.0959)	**1.9713** (0.0765)
VII	$\mathcal{N}$	1.9413 (0.1792)	1.7187 (0.2106)	**2.5043** (0.3515)
	$\mathcal{X}_3^2$	2.1001 (0.2947)	1.7029 (0.2785)	**2.9454** (0.0856)
	Exp	2.1188 (0.2686)	1.7358 (0.2709)	**2.9289** (0.1420)

of the 116 regions, fMRI time series were extracted by averaging all voxels time series within each region at each time point, resulting in 116 different regional fMRI time series, observed at 172 time points. The AAL atlas and the regional fMRI time series are publicly available at NITRC (`www.nitrc.org`).

The goal of this analysis is to apply the proposed methodology to explore whether brain activity as measured by rs-fMRI can be associated with the ADHD index for various quantile levels. Specifically, for each subject $u = 1, \ldots, 72$, $X_u(t) = (X_u^1(t), \ldots, X_u^{116}(t))$ represents the 116 regional rs-fMRI data, where Y_u represents the ADHD score that measures the severity of the disorder; the higher the number of this measure, the more severe the ADHD disease. The histogram of the response observations, Y_u, in Figure 2.1 suggests a right-skewed response distribution with extreme values of ADHD score, justifying the use of QR and the application of our method.

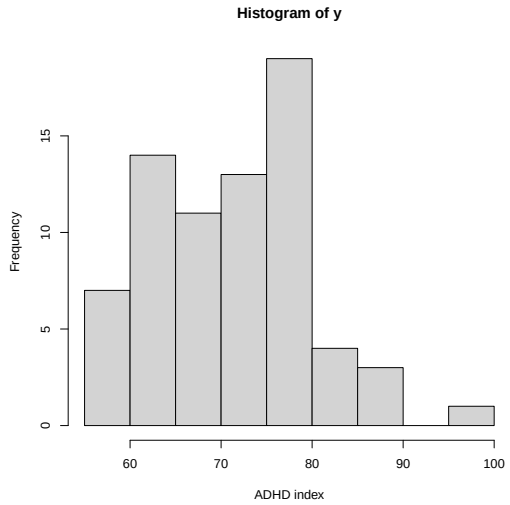

Figure 2.1: Histogram of the response variables (ADHD scores).

We apply our method to derive the first d_τ sufficient predictors $\langle \widehat{v}_{\tau 1}, X\rangle_{\mathscr{H}}, \ldots, \langle \widehat{v}_{\tau d_\tau}, X\rangle_{\mathscr{H}}$ for $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$. For all τ, we choose $d_\tau = 5$ and estimate the fMRI data using 15 B-splines basis functions of order 4. Figure 2.2 shows scatterplots of the first two sufficient predictors by quantile, with different colors indicating different ADHD scores. The distinct grouping of lighter and darker colors demonstrate a clear separation of subject based on their ADHD scores. These groupings effectively highlight the separation of subjects with low, moderate, and high severity ADHD scores.

To compare our method with FSIR, we split the data into training (80%) and test (20%) sets and fit the local linear QR [63] for various quantiles ($\tau = 0.1, 0.25, 0.5, 0.75, 0.9$). We use the first five sufficient predictors from FSIR and τ-FCQS, aligning τ with the quantile of the local linear model. Table 2.6 reports the average mean square error, indicting that our methodology outperforms FSIR, except for $\tau = 0.25$.

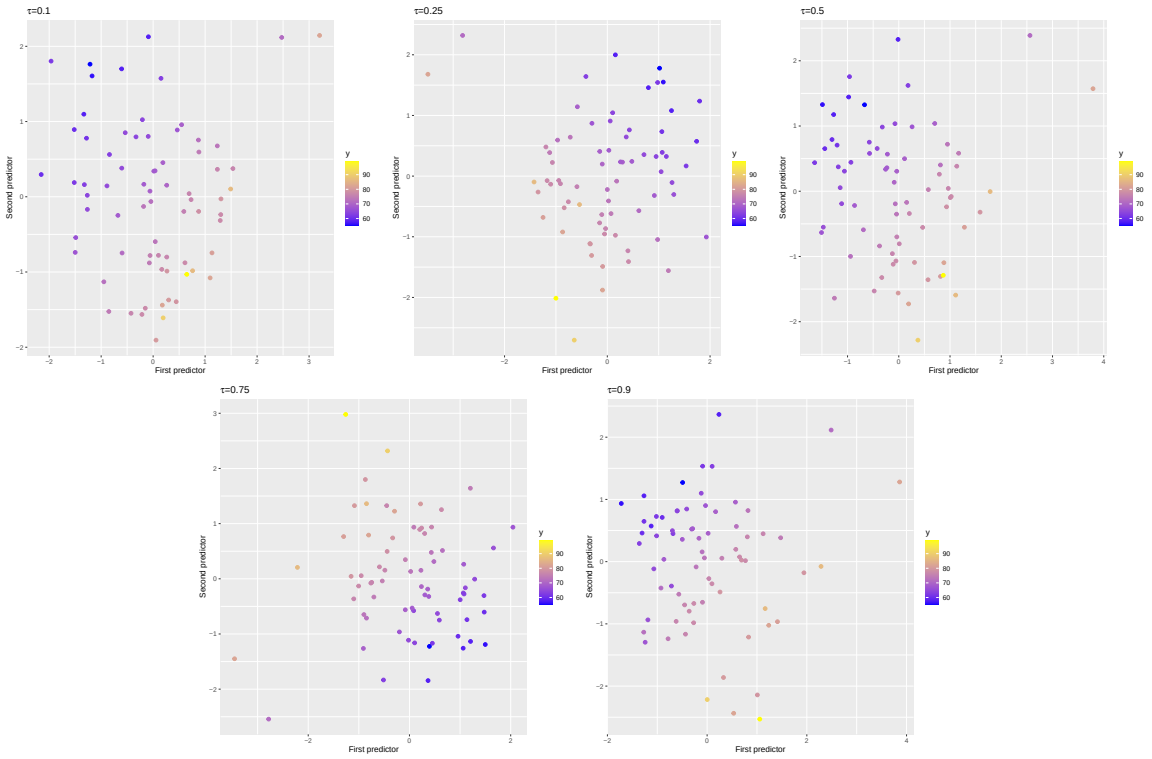

Figure 2.2: Scatterplots of the first two sufficient predictors for the fMRI data set across quantile levels $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$. Each point represents an observation, with colors indicating ADHD scores. Lighter colors correspond to higher ADHD scores. The axes represent the values of the first and second sufficient predictors.

Table 2.6: *Average mean square error of local linear QR model using the first d_τ sufficient predictors constructed by FSIR and τ-FCQS.*

Method	0.1	0.25	0.5	0.75	0.9
FSIR	5.18	4.75	4.53	5.89	7.95
τ-FCQS	4.26	5.21	3.84	5.42	6.93

CHAPTER 3: τth FUNCTIONAL CENTRAL PARTIAL QUANTILE SUBSPACE (τ-FCPQS)

3.1 The Methodology of τ-FCPQS

Let $Y : \Omega \to \mathbb{R}$ denote a univariate response, $X = (X^1, \ldots, X^p) : \Omega \to \mathscr{H}$ denote a random element, and W denote a categorical predictor or a combination of categorical predictors that partition the population into c subpopulations.

Definition 3.1. If there is a finite rank linear operator $M_\tau \in \mathscr{B}(\mathscr{H}, \mathbb{R}^{d_\tau^{(W)}})$, $d_\tau^{(W)} \in \mathbb{N}$, such that

$$Y \perp\!\!\!\perp Q_\tau(Y|X, W)|(M_\tau(X), W), \tag{3.1}$$

then $\overline{\text{ran}}(M_\tau^*)$ is called the *τth functional partial quantile dimension reduction subspace.*

An equivalent formulation of (3.1) that will be used in the Appendix B, is to assume that there is a subspace $\mathcal{S}_\tau$ such that $Y \perp\!\!\!\perp Q_\tau(Y|X, W)|(P_{\mathcal{S}_\tau} X, W)$, where $P_{\mathcal{S}_\tau}$ is the projection of X onto the space $\mathcal{S}_\tau$.

Definition 3.2. Under the assumption that $\cap\{\overline{\text{ran}}(M_\tau^*) : M_\tau \text{ satisfies (3.1)}\}$ satisfies (3.1), we call it the *τth functional central partial quantile subspace* (τ-FCPQS) and denote it by $\mathscr{S}^{(W)}_{Q_\tau(Y|X)}$.

Note that, using the equivalent formulation introduced above, $\mathscr{S}^{(W)}_{Q_\tau(Y|X)}$ is such that $Y \perp\!\!\!\perp Q_\tau(Y|X, W)|\ (P_{\mathscr{S}^{(W)}_{Q_\tau(Y|X)}} X, W)$.

Analogous to the ideas of [66] for the central subspace and [67] for the τth central quantile subspace, estimating the τ-FCPQS reduces to performing dimension

reduction within each subpopulation defined by the categorical variable. Specifically, let (Y_w, X_w) denote the observations corresponding to a specific category w, that is, $(Y, X) \mid (W = w)$. Let $\mathscr{S}_{Q_\tau(Y_w|X_w)}$ denote the τ-FCQS obtained from (Y_w, X_w) within each subpopulation. Then, $\mathscr{S}^{(W)}_{Q_\tau(Y|X)} = \bigoplus_{w=1}^{c} \mathscr{S}_{Q_\tau(Y_w|X_w)}$. Formally,

Proposition 3.1. *Let $\mathscr{S}_{Q_\tau(Y_w|X_w)}$ denote the τ-FCQS for the regression of $Y|W = w$ to $X|W = w$. Then,*

$$\mathscr{S}^{(W)}_{Q_\tau(Y|X)} = \bigoplus_{w=1}^{c} \mathscr{S}_{Q_\tau(Y_w|X_w)},$$

where $\bigoplus$ represents the direct sum between the subspaces.

Proof. See Appendix B.1. □

The above implies that the τ-FCPQS can be determined by combining the τ-FCQS within subpopulations. Note that, no dimension reduction is performed on W, rather it is used to separate the data into subpopulations to which to perform dimension reduction. Moreover, although the subspaces $\mathscr{S}_{Q_\tau(Y_w|X_w)}$, $w = 1, \ldots, c$ can overlap, their direct sum coincide with $\mathscr{S}^{(W)}_{Q_\tau(Y|X)}$.

The following proposition, established by Cook and Dritchley [68], further explains the relationship between $\mathscr{S}_{Q_\tau(Y|X)}$ and $\mathscr{S}^{(W)}_{Q_\tau(Y|X)}$ and suggests that the difference between the two subspaces is due to the joining regression of W on X. Moreover, it implies that if $X \perp\!\!\!\perp W$, then $\mathscr{S}_{Q_\tau(Y|X)} \subseteq \mathscr{S}^{(W)}_{Q_\tau(Y|X)}$.

Proposition 3.2. *Let $\mathscr{S}_{W|X}$ denote the FCS for the regression of W on X. Then,*

$$\mathscr{S}_{Q_\tau(Y|X)} \subseteq \mathscr{S}_{W|X} \bigoplus \mathscr{S}^{(W)}_{Q_\tau(Y|X)}.$$

Proof. See Appendix B.2. □

Finally, the next proposition explains the relationship between $\mathscr{S}_{Q_\tau(Y|X)}$ and $\mathscr{S}^{(W)}_{Q_\tau(Y|X)}$.

The two subspaces might overlap and, depending on the situation, we can have $\mathscr{S}_{Q_\tau(Y|X)} \subseteq \mathscr{S}^{(W)}_{Q_\tau(Y|X)}$ and/or $\mathscr{S}^{(W)}_{Q_\tau(Y|X)} \subseteq \mathscr{S}_{Q_\tau(Y|X)}$.

Proposition 3.3. *The following holds:*

(i) If $W \perp\!\!\!\perp Y|X$ *and* $W \perp\!\!\!\perp Q_\tau(Y|X)|P_{\mathscr{S}^{(W)}_{Q_\tau(Y|X)}}X$ *or if* $W \perp\!\!\!\perp Y|X$ *and* $W \perp\!\!\!\perp Y|P_{\mathscr{S}^{(W)}_{Q_\tau(Y|X)}}X$, *then* $\mathscr{S}_{Q_\tau(Y|X)} \subseteq \mathscr{S}^{(W)}_{Q_\tau(Y|X)}$.

(ii) If $W \perp\!\!\!\perp Y|X$ *and* $W \perp\!\!\!\perp Y|P_{\mathscr{S}_{Q_\tau(Y|X)}}X$, *then* $\mathscr{S}^{(W)}_{Q_\tau(Y|X)} \subseteq \mathscr{S}_{Q_\tau(Y|X)}$

Proof. See Appendix B.3. □

3.2 Estimation of τ-FCPQS

3.2.1 Population Level

Proposition 3.1 suggests a straightforward approach for constructing the τ-FCPQS. Specifically,

(i) Within each subpopulation $w = 1, \ldots, c$, apply methodology of τ-FCQS, described in Chapter 2, to obtain $v^{(w)}_{\tau,1}, \ldots, v^{(w)}_{\tau,d^{(w)}_\tau}$, such that $v^{(w)}_{\tau,j} \in \mathscr{S}_{Q_\tau(Y_w|X_w)}$, $j = 1, \ldots, d^{(w)}_\tau$, where $d^{(w)}_\tau$ denotes the dimension of each $\mathscr{S}_{Q_\tau(Y_w|X_w)}$.

(ii) Let $V^{(W)}_\tau = \sum_{w=1}^{c}\sum_{j=1}^{d^{(w)}_\tau} v^{(w)}_{\tau,j} \otimes v^{(w)}_{\tau,j}$. Then, for $\eta_{\tau,1}, \ldots, \eta_{\tau,d^{(W)}_\tau}$ being the set of solutions to the generalized eigenvalue problem $V^{(W)}_\tau \eta_{\tau,j} = \lambda\eta_{\tau,j}$, we have that $\eta_{\tau,j} \in \mathscr{S}^{(W)}_{Q_\tau(Y|X)}$, $j = 1, \ldots, d^{(W)}_\tau$.

The eigenvalue problem at the population level is described as follows

$$\begin{aligned} &\text{argmax} \quad \langle \eta_\tau, V^{(W)}_\tau \eta_\tau \rangle_{\mathscr{H}} \\ &\text{subject to} \quad \eta_\tau \in \mathscr{H}, \langle \eta_\tau, \eta_\tau \rangle_{\mathscr{H}} = 1, \langle \eta_\tau, \eta_{\tau j} \rangle_{\mathscr{H}} = 0, j = 1, \ldots, d^{(W)}_\tau - 1. \end{aligned} \tag{3.2}$$

Remark 3.1. The methodology of τ-FCQS relies on Assumption 2.2. However, since we perform their method within each subpopulation w, we assume that the assumption holds for each $w = 1, \ldots, c$, as follows.

Assumption 3.1. For a given τ and a given w, there is a bounded linear operator $\Lambda_\tau^{(w)} : \mathrm{ran}(L_\tau^{(w)}) \to \mathscr{H}$ such that $\mathrm{E}\{X_w | L_\tau^{(w)}(X_w)\} = \Lambda_\tau^{(w)} L_\tau^{(w)}(X_w)$, where $L_\tau^{(w)} \in \mathscr{B}(\mathscr{H}, \mathbb{R}^{d_\tau^{(w)}})$ and $\mathscr{S}_{Q_\tau(Y_w|X_w)} = \overline{\mathrm{ran}}(L_\tau^{(w)*})$.

3.3 Implementation

In this section, we present the algorithm for the proposed methodology under the common assumption that the functions $X_u(t)$, $u = 1, \ldots, n$, are not fully observed. Instead, they are observed at a finite set of points $t_{u1}, \ldots, t_{uN_u}$. In such situations, $X_u(t)$ needs to be estimated using the observed data $\{(t, X_u(t)) : t = t_{u1}, \ldots, t_{uN_u}\}$. A common approach is to express $X_u(t)$ through a coordinate representation using a finite number of basis functions; see Section A.5 for details.

3.3.1 Algorithm for the τ-FCPQS

For $i = 1, \ldots, p$, let $\mathcal{H}_i$ spanned by a finite set of basis functions $\mathcal{B}_i = \{b_1^i, \cdots, b_{k_n}^i\}$. Then, each X_u^i, $u = 1, \ldots, n$, of the observed data $\{(t, X_u(t)) : t = t_{u1}, \ldots, t_{uN_u}\}$, can be approximated by $[X_u^i]^\top b_{1:k_n}^i$. Moreover, $[X_u] = ([X_u^1]^\top, \ldots, [X_u^p]^\top)^\top \in \mathbb{R}^{pk_n}$ is the coordinate representation of X_u, $u = 1, \ldots, n$, and $[X_{1:n}]$ is the $pk_n \times n$ matrix $([X_1], \ldots, [X_n])$. Note that, we omit indices such as $[\cdot]_{\mathcal{B}_i}$, as the bases for the coordinates can be identified by the domain and range of the operators involved.

Let $\mathcal{G} = \mathrm{diag}(\mathcal{G}_i : i = 1, \ldots, p)$, where each $\mathcal{G}_i$ represents the Gram matrix of $\mathcal{H}_i$. Then, the coordinate representation of the sample level version $\widehat{V}_\tau^{(W)} = \sum_{w=1}^c \sum_{j=1}^{d_\tau^{(w)}} \widehat{v}_{\tau,j}^{(w)} \otimes \widehat{v}_{\tau,j}^{(w)}$ is given by

$$\begin{aligned}
[\widehat{V}_\tau^{(W)}] &= \sum_{w=1}^{c} \sum_{j=1}^{d_\tau^{(w)}} [\widehat{v}_{\tau,j}^{(w)} \otimes \widehat{v}_{\tau,j}^{(w)}] = \sum_{w=1}^{c} \sum_{j=1}^{d_\tau^{(w)}} [\widehat{v}_{\tau,j}^{(w)}][\widehat{v}_{\tau,j}^{(w)}]^\top \mathcal{G} \\
&= ([\widehat{v}_{\tau,1:d_\tau^{(1)}}^{(1)}]^\top, \ldots, [\widehat{v}_{\tau,1:d_\tau^{(c)}}^{(c)}]^\top)^\top ([\widehat{v}_{\tau,1:d_\tau^{(1)}}^{(1)}]^\top, \ldots, [\widehat{v}_{\tau,1:d_\tau^{(c)}}^{(c)}]^\top) \mathcal{G}.
\end{aligned}$$

The generalized eigenvalue problem (3.2) can be expressed as

$$\begin{aligned}&\text{argmax} \quad [\eta_\tau]^\top \mathcal{G}([\widehat{v}^{(1)}_{\tau,1:d_\tau^{(1)}}]^\top, \ldots, [\widehat{v}^{(c)}_{\tau,1:d_\tau^{(c)}}]^\top)^\top([\widehat{v}^{(1)}_{\tau,1:d_\tau^{(1)}}]^\top, \ldots, [\widehat{v}^{(c)}_{\tau,1:d_\tau^{(c)}}]^\top)\mathcal{G}[\eta_\tau] \\ &\text{subject to} \quad [\eta_\tau]^\top \mathcal{G}[\eta_\tau] = 1, [\eta_\tau]^\top \mathcal{G}[\eta_{\tau,j}] = 0, j = 1, \ldots, d_\tau^{(W)} - 1.\end{aligned}$$

Let $\omega_\tau = \mathcal{G}^{1/2}[\eta_\tau]$ and solve

$$\begin{aligned}&\text{argmax} \quad \omega_\tau{}^\top \mathcal{G}^{\dagger 1/2}\mathcal{G}([\widehat{v}^{(1)}_{\tau,1:d_\tau^{(1)}}]^\top, \ldots, [\widehat{v}^{(c)}_{\tau,1:d_\tau^{(c)}}]^\top)^\top([\widehat{v}^{(1)}_{\tau,1:d_\tau^{(1)}}]^\top, \ldots, [\widehat{v}^{(c)}_{\tau,1:d_\tau^{(c)}}]^\top)\mathcal{G}\mathcal{G}^{\dagger 1/2}\omega_\tau \\ &\text{subject to} \quad \omega_\tau^\top \omega_\tau = 1, \omega_\tau^\top \omega_{\tau,j} = 0, j = 1, \ldots, d_\tau^{(W)} - 1.\end{aligned} \tag{3.3}$$

That is, $\omega_{\tau,1}, \ldots, \omega_{\tau,d_\tau^{(W)}}$ are the eigenvectors of the matrix

$$\mathcal{G}^{\dagger 1/2}\mathcal{G}([\widehat{v}^{(1)}_{\tau,1:d_\tau^{(1)}}]^\top, \ldots, [\widehat{v}^{(c)}_{\tau,1:d_\tau^{(c)}}]^\top)^\top([\widehat{v}^{(1)}_{\tau,1:d_\tau^{(1)}}]^\top, \ldots, [\widehat{v}^{(c)}_{\tau,1:d_\tau^{(c)}}]^\top)\mathcal{G}\mathcal{G}^{\dagger 1/2}$$

that correspond to the $d_\tau^{(W)}$ largest eigenvalues.

Algorithm. Let $\{Y_u, (t, X_u(t)) : t = t_{u1}, \ldots, t_{uN_u}\}_{u=1}^n$ be an iid sample.

(i) For each $u = 1, \ldots, n$, obtain the coordinates of X_u relative to the basis $\mathcal{B} = \bigoplus_{i=1}^p \mathcal{B}_i$ of $\mathscr{H}$ and derive the Gram matrix $\mathcal{G}$.

(ii) Within each subpopulation $w = 1, \ldots, c$, use Christou et al. (2025) to obtain $[\widehat{v}^{(w)}_{\tau,1}], \ldots, [\widehat{v}^{(w)}_{\tau,d_\tau^{(w)}}]$ and form the vector $([\widehat{v}^{(1)}_{\tau,1}], \ldots, [\widehat{v}^{(1)}_{\tau,d_\tau^{(1)}}], \ldots, [\widehat{v}^{(c)}_{\tau,1}], \ldots, [\widehat{v}^{(c)}_{\tau,d_\tau^{(c)}}])$. This vector will be denoted by $([\widehat{v}^{(1)}_{\tau,1:d_\tau^{(1)}}]^\top, \ldots, [\widehat{v}^{(c)}_{\tau,1:d_\tau^{(c)}}]^\top)$.

(iii) Form $[\widehat{V}_\tau^{(W)}] = ([\widehat{v}^{(1)}_{\tau,1:d_\tau^{(1)}}]^\top, \ldots, [\widehat{v}^{(c)}_{\tau,1:d_\tau^{(c)}}]^\top)^\top([\widehat{v}^{(1)}_{\tau,1:d_\tau^{(1)}}]^\top, \ldots, [\widehat{v}^{(c)}_{\tau,1:d_\tau^{(c)}}]^\top)\mathcal{G}$.

(iv) Solve the eigenvalue problem (3.3) and obtain the $d_\tau^{(W)}$ solutions $\omega_{\tau,j}$, $j = 1, \ldots, d_\tau^{(W)}$, associated with the $d_\tau^{(W)}$ largest eigenvalues.

(v) Obtain $[\widehat{\eta}_{\tau,j}] = \mathcal{G}^{\dagger 1/2}\omega_{\tau,j}$ and the sufficient predictors $\langle X_u, \widehat{\eta}_{\tau,j}\rangle_{\mathscr{H}} = [X_u]^\top \mathcal{G}^{1/2}\omega_{\tau,j}$, $j = 1, \ldots, d_\tau^{(W)}$, $u = 1, \ldots, n$.

Remark 3.2. The dimension $d_\tau^{(W)}$ of the τ-FCPQS, as well as the dimension $d_\tau^{(w)}$ of the τ-FCQS $\mathscr{S}_{Q_\tau(Y_w|X_w)}$, for $w = 1, \ldots, c$, are unknown in practice and need to be estimated. For that, we can use the cross-validation Bayesian information criterion (CVBIC) ([52]). However, for this Chapter, we assume that the dimensions are known.

3.4 Asymptotic Theory

We now present the consistency and convergence rate of $\widehat{V}_\tau^{(W)} = \sum_{w=1}^{c}\sum_{j=1}^{d_\tau^{(w)}} \widehat{v}_{\tau,j}^{(w)} \otimes \widehat{v}_{\tau,j}^{(w)}$, where $\widehat{v}_{\tau,j}^{(w)}$, $j = 1, \ldots, d_\tau^{(w)}$, $w = 1, \ldots, c$, are the estimated eigenfunctions resulting from the algorithm of τ-FCQS. Before proceeding to the results, we need to introduce some additional notation. Let n_w denote the sample size for each category w, i.e., the number of observations within each subpopulation, n_* denote the smallest sample size, i.e., $n_* = \min\{n_w : w = 1, \ldots, c\}$, and $\Sigma_{XX}^{(w)}$ denote the covariance operator of $X|W = w$, for $w = 1, \ldots, c$. Also, for two positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \prec b_n$ if $a_n/b_n \to 0$.

An additional assumption necessary for the theory is the following:

Assumption 3.2. For a given w, $\mathscr{S}_{Q_\tau(Y_w|X_w)} \subseteq \overline{\text{ran}}(\Sigma_{XX}^{(w)})$.

This assumption was used in [36] and it is less restrictive than assuming that the covariance operator of the predictors is the same across subpopulations, an assumption used, for example, in [66] and [69].

Theorem 3.1. *If assumptions 2.1, 3.1, 3.2, $E(X_w) = 0$, $\widehat{L}(X_w)$ is consistent of the directions of the FCS for all $w = 1, \ldots, c$, and S1-S6 from Appendix A.1 hold, the number of categories c is fixed, and $n_*^{-1/5} \prec \epsilon_n \prec 1$, then, for a given $\tau \in (0,1)$, $\widehat{V}_\tau^{(W)}$ is a consistent estimate of V_τ, and*

$$\left\|\widehat{V}_\tau^{(W)} - V_\tau^{(W)}\right\| = O_p(n_*^{-1/2}\epsilon_n^{-5/2}) + O(\epsilon_n^{1/2}),$$

where $\|\cdot\|$ *denotes the operator norm.*

Proof. See Appendix B.4. □

Corollary 3.1. *Under the assumptions of Theorem 3.1, and for a given* $\tau \in (0,1)$ *and* $j = 1, \ldots, d_\tau^{(W)}$,

$$\|\widehat{\eta}_{\tau,j} - \eta_{\tau,j}\|_{\mathscr{H}} = O_p(n_*^{-1/2}\epsilon_n^{-5/2}) + O(\epsilon_n^{1/2}),$$
$$\langle \widehat{\eta}_{\tau,j}, X\rangle_{\mathscr{H}} - \langle \eta_{\tau,j}, X\rangle_{\mathscr{H}} = O_p(n_*^{-1/2}\epsilon_n^{-5/2}) + O(\epsilon_n^{1/2}).$$

Proof. The proof follows the same approach as Corollary 2 in [36] and it is omitted. □

Theorem 3.1 assumes that the number of categories c is fixed. However, there are situations where c can increase with the sample size n, i.e., $c_n \to \infty$ as $n \to \infty$. The following corollary shows how the convergence rate will be affected and it can be used to provide an optimal number c_n.

Corollary 3.2. *If assumptions 2.1, 3.1, 3.2,* $E(X_w) = 0$, $\widehat{L}(X_w)$ *is consistent of the directions of the FCS for all* $w = 1, \ldots, c$, *and S1-S6 from Appendix A.1 hold, the number of categories* c_n *is such that* $n_*^{-1/5} \prec \epsilon_n \prec c_n^{-2} \prec 1$, *then, for a given* $\tau \in (0,1)$, $\widehat{V}_\tau^{(W)}$ *is a consistent estimate of* V_τ, *and*

$$\left\|\widehat{V}_\tau^{(W)} - V_\tau^{(W)}\right\| = O_p(n_*^{-1/2}\epsilon_n^{-5/2}) + O(c_n\epsilon_n^{1/2}),$$

where $n_* = \min\{n_w : w = 1, \ldots, c_n\}$.

Proof. The proof follows from Theorem 3.1, and thus it is omitted. □

3.5 Simulation Studies

Simulation Setting. The simulation setting is the same as that used for the τ-FCQS method in Section 2.5.1. The categorical predictor W is generated from a

Binomial$(1, 0.5)$ distribution. Over the time interval $[0, 1]$, we create 101 equally spaced time points with a step size of 0.01. All simulation results are based on $N = 100$ Monte Carlo replications. Unless otherwise stated, the sample size is set to $n = 400$, the number of functional predictors is $p = 5$, the quantile levels considered are $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$, and the dimension $d_\tau^{(W)}$ is assumed to be known.

Algorithm and related parameters. Step 1 of the algorithm applies the τ-FCQS method within each subpopulation W. The parameters are chosen as described in Section 2.5.1. Moreover, to obtain the coordinate representation of $X_u, u = 1, \ldots, n$, we approximate the functional predictors using B-splines with four basis functions.

Models under consideration. The models that we consider throughout the simulation studies are:

$$\text{M-I:} \quad Y = \sin(\pi\langle\beta_1, X\rangle_{\mathscr{H}}/4) + W + 0.5\varepsilon, \text{ where } \mathscr{S}^{(W)}_{Q_\tau(Y|X)} = \text{span}\{\beta_1\}$$

$$\text{M-II:} \quad Y = \frac{1}{0.5 + (\langle\beta_1, X\rangle_{\mathscr{H}} + 1)^2} + W + 0.2\varepsilon, \text{ where } \mathscr{S}^{(W)}_{Q_\tau(Y|X)} = \text{span}\{\beta_1\}$$

$$\text{M-III:} \quad Y = (1 + W)\frac{\exp(\langle\beta_1, X\rangle_{\mathscr{H}})}{1 + \exp(\langle\beta_1, X\rangle_{\mathscr{H}})} + 2W + 0.1\varepsilon, \text{ where } \mathscr{S}^{(W)}_{Q_\tau(Y|X)} = \text{span}\{\beta_1\}$$

$$\text{M-IV:} \quad Y = \langle\beta_1, X\rangle_{\mathscr{H}} + \langle\beta_2, X\rangle_{\mathscr{H}}\varepsilon + W, \text{ where } \mathscr{S}^{(W)}_{Q_\tau(Y|X)} = \text{span}\{\beta_1 + Q_\tau(\varepsilon)\beta_2\}$$

$$\text{M-V:} \quad Y = (1 + W)\langle\beta_1, X\rangle^3_{\mathscr{H}} + \langle\beta_2, X\rangle_{\mathscr{H}} + 2W + \varepsilon, \text{ where } \mathscr{S}^{(W)}_{Q_\tau(Y|X)} = \text{span}\{\beta_1, \beta_2\}$$

$$\text{M-VI:} \quad Y = \langle\beta_1, X\rangle_{\mathscr{H}} + \langle\beta_2, X\rangle_{\mathscr{H}} + W + \sqrt{0.5 + \langle\beta_1, X\rangle^2_{\mathscr{H}} + \langle\beta_2, X\rangle^2_{\mathscr{H}}}\varepsilon, \text{ where } \mathscr{S}^{(W)}_{Q_\tau(Y|X)} = \text{span}\{\beta_1, \beta_2\}$$

$$\text{M-VII:} \quad Y = \langle\beta_1, X\rangle^3_{\mathscr{H}} + \exp(\langle\beta_2, X\rangle_{\mathscr{H}}) + \langle\beta_3, X\rangle_{\mathscr{H}}\varepsilon + W, \text{ where } \mathscr{S}^{(W)}_{Q_\tau(Y|X)} = \text{span}\{\beta_1, \beta_2, \beta_3\} \text{ when } Q_\tau(\varepsilon) \neq 0 \text{ and } \text{span}\{\beta_1, \beta_2\} \text{ when } Q_\tau(\varepsilon) = 0$$

$$\text{M-VIII:} \quad Y = W\langle\beta_1, X\rangle_{\mathscr{H}} + (1 + W)\langle\beta_2, X\rangle_{\mathscr{H}} + \varepsilon, \text{ where } \mathscr{S}^{(W)}_{Q_\tau(Y|X)} = \text{span}\{\beta_1, \beta_2\},$$

and $\beta_1(t) = \psi_1(t)$, $\beta_2(t) = \psi_2(t)$, and $\beta_3(t) = \psi_3(t)$ are the first, second, and third

eigenfunctions of Σ_{XX}, respectively, X is simulated as described above, and the error ε is generated according to (1) standard normal ($\mathcal{N}$), (2) chi-square with three degrees of freedom ($\mathcal{X}^2_3$), and (3) exponential with rate 0.5 (Exp(0.5)). All models were considered during the simulation runs but, due to space constraints, we only report selective ones.

Example 1 - Effect of n, p, and error distribution. This example demonstrates the performance of the proposed algorithm for different choices of n, p, and error distribution. We consider different sample sizes $n = \{200, 400, 1000\}$, number of predictors $p = \{5, 10, 20\}$, and error distributions as described above. Tables 3.1–3.3 present the means and standard deviations of the multiple correlations for Model I for various values of τ. The proposed methodology shows an expected increase in efficiency with larger sample sizes n, and a decrease with an increasing number of predictors p. Additionally, its performance remains robust across various quantile levels and error distributions.

Table 3.1: *Mean (and standard deviation) of multiple correlation for Model I, when* $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$ *and the error follows a standard normal distribution.*

n	p	0.1	0.25	0.5	0.75	0.9
200	5	0.9311 (0.1121)	0.9333 (0.1066)	0.9363 (0.0945)	0.9398 (0.0760)	0.9343 (0.0791)
	10	0.8569 (0.1285)	0.8565 (0.1279)	0.8575 (0.1262)	0.8605 (0.1246)	0.8601 (0.1263)
	20	0.7842 (0.1945)	0.7775 (0.2023)	0.7725 (0.2128)	0.7678 (0.2168)	0.7562 (0.2270)
400	5	0.9633 (0.0458)	0.9637 (0.0452)	0.9627 (0.0473)	0.9615 (0.0489)	0.9596 (0.0531)
	10	0.8679 (0.1393)	0.8686 (0.1382)	0.8663 (0.1491)	0.8635 (0.1567)	0.8635 (0.1495)
	20	0.8376 (0.1431)	0.8421 (0.1282)	0.8492 (0.1161)	0.8470 (0.1274)	0.8447 (0.1381)
1000	5	0.9792 (0.0221)	0.9793 (0.0225)	0.9793 (0.0230)	0.9792 (0.0237)	0.9790 (0.0244)
	10	0.8903 (0.1083)	0.8902 (0.1089)	0.8910 (0.1093)	0.8889 (0.1119)	0.8877 (0.1145)
	20	0.8696 (0.1132)	0.8696 (0.1118)	0.8703 (0.1095)	0.8685 (0.1096)	0.8633 (0.1226)

Example 2 - Performance of the algorithm. This example demonstrates the performance of the algorithm across all examined models. Table 3.4 presents the means and standard deviations of the multiple correlations for various τ and error distributions. The anticipated outcome is that the multiple correlation approximates the true dimension $d_\tau^{(W)}$ closely. Our observations confirm this expectation: the

Table 3.2: *Mean (and standard deviation) of multiple correlations for Model I, when* $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$ *and the error follows a chi-square distribution.*

n	p	0.1	0.25	0.5	0.75	0.9
200	5	0.9168 (0.1116)	0.9167 (0.0967)	0.9142 (0.0980)	0.8961 (0.1216)	0.8569 (0.2098)
	10	0.8403 (0.1386)	0.8343 (0.1517)	0.8396 (0.1469)	0.8165 (0.1834)	0.8240 (0.1732)
	20	0.7437 (0.2391)	0.7391 (0.2427)	0.7354 (0.2338)	0.7180 (0.2513)	0.7228 (0.2532
400	5	0.9658 (0.0354)	0.9659 (0.0343)	0.9655 (0.0345)	0.9628 (0.0376)	0.9442 (0.1018)
	10	0.8817 (0.1035)	0.8814 (0.1054)	0.8805 (0.1094)	0.8725 (0.1343)	0.8607 (0.1425)
	20	0.8228 (0.1541)	0.8188 (0.1525)	0.8036 (0.1756)	0.8016 (0.1763)	0.7853 (0.1967)
1000	5	0.9750 (0.0265)	0.9748 (0.0262)	0.9744 (0.0258)	0.9727 (0.0255)	0.9638 (0.0494)
	10	0.9152 (0.0699)	0.9147 (0.0711)	0.9143 (0.0725)	0.9121 (0.0719)	0.9049 (0.0924)
	20	0.8755 (0.0968)	0.8741 (0.0990)	0.8731 (0.1019)	0.8721 (0.1076)	0.8590 (0.1261)

Table 3.3: *Mean (and standard deviation) of multiple correlations for Model I, when* $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$ *and the error follows an exponential distribution.*

n	p	0.1	0.25	0.5	0.75	0.9
200	5	0.9395 (0.0700)	0.9400 (0.0690)	0.9382 (0.0704)	0.9345 (0.0792)	0.9088 (0.1365)
	10	0.8018 (0.2220)	0.7937 (0.2261)	0.7833 (0.2367)	0.7864 (0.2220)	0.7829 (0.2247)
	20	0.7377 (0.2319)	0.7323 (0.2463)	0.7326 (0.2266)	0.7297 (0.2187)	0.7225 (0.2200)
400	5	0.9563 (0.0409)	0.9566 (0.0403)	0.9569 (0.0397)	0.9555 (0.0403)	0.9495 (0.0512)
	10	0.8922 (0.0928)	0.8943 (0.0872)	0.8956 (0.0862)	0.8880 (0.1141)	0.8794 (0.1246)
	20	0.8149 (0.1565)	0.8140 (0.1591)	0.8142 (0.1606)	0.8112 (0.1614)	0.8055 (0.1635)
1000	5	0.9763 (0.0235)	0.9762 (0.0236)	0.9756 (0.0239)	0.9743 (0.0253)	0.9721 (0.0271)
	10	0.9164 (0.0735)	0.9148 (0.0794)	0.9123 (0.0865)	0.9087 (0.0994)	0.9058 (0.1124)
	20	0.8651 (0.1219)	0.8654 (0.1223)	0.8648 (0.1226)	0.8625 (0.1238)	0.8588 (0.1250)

multiple correlation is close to 1 for Models I-IV, to 2 for Models V, VI, and VIII, and to 3 for Model VII, demonstrating the algorithm's accuracy. Note that for Model VII, where the error term $\varepsilon \sim N(0, 1)$, the conditional median of ε is 0. As a result, the subspace $\mathscr{S}^{(W)}_{Q_\tau(Y|X)}$ has a dimension of 2. Therefore, the multiple correlation for Model VII closely approximates 2 under standard normal error at $\tau = 0.5$, aligning with the expected outcome.

Example 3 - Effect of categorical predictor W. This example evaluates the performance of the proposed methodology under different scenarios of the categorical predictor W. Specifically, W is generated according to three different scenarios:

- Scenario 1 (Unbalanced case) - W follows a Binomial(1, 0.1) distribution.
- Scenario 2 (Multiclass case) - W follows a multinomial distribution with three classes, having probabilities of 0.2, 0.3, and 0.5, respectively.

Table 3.4: *Mean (and standard deviation) of multiple correlations for Models I-VIII, when* $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$.

M	error	0.1	0.25	0.5	0.75	0.9
I	$\mathcal{N}$	0.9681 (0.0331)	0.9679 (0.0324)	0.9682 (0.0315)	0.9677 (0.0321)	0.9669 (0.0329)
	$\mathcal{X}_3^2$	0.9555 (0.0477)	0.9551 (0.0480)	0.9532 (0.0522)	0.9411 (0.0977)	0.9356 (0.0911)
	Exp	0.9579 (0.0454)	0.9575 (0.0466)	0.9569 (0.0481)	0.9553 (0.0516)	0.9295 (0.1274)
II	$\mathcal{N}$	0.9498 (0.0822)	0.9474 (0.0934)	0.9573 (0.0516)	0.9610 (0.0419)	0.9612 (0.0427)
	$\mathcal{X}_3^2$	0.9027 (0.1565)	0.9393 (0.0763)	0.9420 (0.0718)	0.9440 (0.0698)	0.9383 (0.0746)
	Exp	0.9351 (0.0751)	0.9418 (0.0782)	0.9467 (0.0664)	0.9511 (0.0497)	0.9476 (0.0537)
III	$\mathcal{N}$	0.9622 (0.0421)	0.9610 (0.0443)	0.9602 (0.0459)	0.9586 (0.0503)	0.9574 (0.0542)
	$\mathcal{X}_3^2$	0.9517 (0.0577)	0.9515 (0.0583)	0.9504 (0.0601)	0.9485 (0.0646)	0.9470 (0.0667)
	Exp	0.9525 (0.0558)	0.9523 (0.0558)	0.9517 (0.0572)	0.9495 (0.0616)	0.9463 (0.0690)
IV	$\mathcal{N}$	0.7251 (0.0387)	0.7241 (0.0377)	0.7246 (0.0367)	0.7243 (0.0371)	0.7237 (0.0373)
	$\mathcal{X}_3^2$	0.6140 (0.2061)	0.6125 (0.2068)	0.6116 (0.2066)	0.6077 (0.2087)	0.6132 (0.2047)
	Exp	0.6104 (0.1739)	0.6034 (0.1792)	0.6014 (0.1778)	0.6188 (0.1634)	0.6141 (0.1716)
V	$\mathcal{N}$	1.6157 (0.3191)	1.6447 (0.3161)	1.6302 (0.3318)	1.6493 (0.3078)	1.6438 (0.3086)
	$\mathcal{X}_3^2$	1.6480 (0.2983)	1.6798 (0.2929)	1.6681 (0.3014)	1.6577 (0.3002)	1.6544 (0.2878)
	Exp	1.5788 (0.3278)	1.5714 (0.3504)	1.5613 (0.3372)	1.5777 (0.3320)	1.5886 (0.3220)
VI	$\mathcal{N}$	1.7924 (0.2384)	1.8012 (0.2129)	1.8031 (0.2060)	1.7993 (0.2117)	1.7789 (0.2308)
	$\mathcal{X}_3^2$	1.7541 (0.2358)	1.7289 (0.2636)	1.7015 (0.2760)	1.6585 (0.3001)	1.6307 (0.3046)
	Exp	1.7919 (0.2071)	1.7854 (0.2184)	1.7863 (0.2084)	1.7768 (0.2281)	1.6828 (0.2904)
VII	$\mathcal{N}$	2.4373 (0.3176)	2.4460 (0.3301)	1.7081 (0.2788)	2.4974 (0.3318)	2.5162 (0.3268)
	$\mathcal{X}_3^2$	2.6392 (0.2961)	2.6577 (0.3063)	2.6442 (0.3321)	2.6515 (0.3038)	2.6320 (0.3195)
	Exp	2.5359 (0.3413)	2.5377 (0.3375)	2.5636 (0.3204)	2.5432 (0.3251)	2.5244 (0.3414)
VIII	$\mathcal{N}$	1.7157 (0.2836)	1.7213 (0.2767)	1.7255 (0.2684)	1.7155 (0.2733)	1.7234 (0.2673)
	$\mathcal{X}_3^2$	1.7060 (0.2769)	1.7096 (0.2820)	1.7188 (0.2756)	1.7406 (0.2370)	1.7286 (0.2437)
	Exp	1.7007 (0.2685)	1.7153 (0.2487)	1.7131 (0.2422)	1.6828 (0.2772)	1.7046 (0.2538)

- Scenario 3 (Dependent case) - W is generated based on the predictor X. Specifically, W follows a Binomial(1, p), where

$$p = \frac{\exp(\langle \beta_1, X \rangle_{\mathscr{H}})}{1 + \exp(\langle \beta_1, X \rangle_{\mathscr{H}})}.$$

Since the effect of the error distribution was examined in previous examples, we present results only for the standard normal error distribution. Table 3.5 presents the means and standard deviations of the multiple correlations for Model III (single-index model) and Model VI (multi-index model). Comparing these results with those in Table 3.4, we observe that the estimation accuracy remains largely unchanged for Scenarios 1 and 3. In the multiclass case (Scenario 2), the multiple correlations are slightly lower, but the difference is marginal. Overall, the performance of the

algorithm appears to be robust across different scenarios of W.

Table 3.5: *Mean (and standard deviation) of multiple correlations for Models III and VI under three scenarios for W at* $\tau \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$

M	W	0.1	0.25	0.5	0.75	0.9
III	Scenario 1	0.9676 (0.0223)	0.9670 (0.0255)	0.9679 (0.0223)	0.9664 (0.0258)	0.9669 (0.0262)
	Scenario 2	0.9151 (0.1407)	0.9146 (0.1430)	0.9131 (0.1441)	0.9093 (0.1501)	0.9003 (0.1650)
	Scenario 3	0.9502 (0.0634)	0.9508 (0.0623)	0.9498 (0.0656)	0.9489 (0.0687)	0.9473 (0.0755)
VI	Scenario 1	1.8355 (0.1824)	1.8283 (0.1975)	1.8355 (0.1887)	1.8363 (0.1693)	1.8352 (0.1740)
	Scenario 2	1.6873 (0.3066)	1.7197 (0.2620)	1.7313 (0.2526)	1.7316 (0.2443)	1.7205 (0.2409)
	Scenario 3	1.7734 (0.2566)	1.7810 (0.2413)	1.7893 (0.2349)	1.8084 (0.2059)	1.7990 (0.2218)

Example 4 - Comparisons. To the best of our knowledge, τ-FCQS is the first work to consider dimension reduction techniques for the conditional quantiles of functional data, and therefore, the proposed methodology is the first extension to categorical predictors. Since there are no quantile regression methods to compare with, we consider existing dimension reduction techniques for functional data that focus on the entire conditional distribution, such as FSIR of [29]. The extension of FSIR to categorical predictors follows the same idea as in [66] and it was developed by [69]; we denote this method with FPSIR. Note that our method focuses on FCPQS, while FPSIR focuses on the entire conditional distribution. For meaningful comparisons, we only consider models where the two subspaces coincide, i.e., homoscedastic models. For these situations, the choice of τ does not matter and, although we run the simulations for all τ, we only report the results for $\tau = 0.5$. Table 3.6 reports the mean values and standard deviations of the multiple correlations for all models except Model IV and Model VII. The results suggest that, generally, our method performs comparable with FPSIR for the single-index Models I-III, while it outperforms FPSIR for the multi-index Models V, VI and VIII.

3.6 Applications

In this section, we apply our methodology to the fMRI dataset from the ADHD-200 Consortium, which is the same dataset analyzed in Chapter 2 and introduced in Section 2.6. Here, we investigate whether the ADHD index can be better predicted

Table 3.6: *Mean (and standard deviation) of multiple correlations for Models I-III, and V, VI, VIII, for FPSIR and the proposed methodology τ-FCPQS for $\tau = 0.5$ and various error distributions. The value closest to the true dimension is in bold.*

Model	error	FPSIR	0.5-FCPQS
I	$\mathcal{N}$	**0.9824** (0.0067)	0.9570 (0.0500)
	$\mathcal{X}_3^2$	**0.9522** (0.0240)	0.9416 (0.1156)
	Exp	**0.9683**(0.0152)	0.9648 (0.0915)
II	$\mathcal{N}$	**0.9574** (0.0203)	0.9488 (0.0634)
	$\mathcal{X}_3^2$	0.9099 (0.0831)	**0.9428** (0.1056)
	Exp	0.9332 (0.0417)	**0.9675** (0.0582)
III	$\mathcal{N}$	**0.9960** (0.0012)	0.9561 (0.0501)
	$\mathcal{X}_3^2$	**0.9912** (0.0035)	0.9403 (0.1116)
	Exp	**0.9938** (0.0020)	0.9660 (0.0705)
V	$\mathcal{N}$	**1.6863** (0.2380)	1.6029 (0.3162)
	$\mathcal{X}_3^2$	1.3157 (0.3910)	**1.6934** (0.2851)
	Exp	1.5209 (0.3213)	**1.7219** (0.2820)
VI	$\mathcal{N}$	0.6948 (0.3183)	**1.7145** (0.2795)
	$\mathcal{X}_3^2$	0.6825 (0.2591)	**1.7339** (0.2627)
	Exp	0.6978 (0.2749)	**1.7948** (0.2231)
VIII	$\mathcal{N}$	0.5071 (0.2912)	**1.7666** (0.2468)
	$\mathcal{X}_3^2$	0.5338 (0.2928)	**1.7454** (0.2658)
	Exp	0.5332 (0.3191)	**1.7622** (0.2502)

by incorporating gender information in addition to rs-fMRI data. We also compare the resulting findings with those obtained in Chapter 2.

For the analysis, we estimate the fMRI functional data using 15 B-splines basis functions of order four and use the proposed methodology to extract the sufficient predictors $\langle \widehat{\eta}_{\tau 1}, X\rangle_{\mathscr{H}}, \ldots, \langle \widehat{\eta}_{\tau d_\tau^{(W)}}, X\rangle_{\mathscr{H}}$ for $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$ and $d_\tau^{(W)} = 2$. The left part of Figure 3.1 presents scatterplots of the first two sufficient predictors, where observations are color-coded to reflect various ADHD score levels. Additionally, distinct shapes represent gender levels: circles for females and triangles for males. We note a distinct separation between male and female subjects at all quantile levels. Providing a focused enlargement on the male subjects for further analysis, we observe a clear stratification of the subjects based on their ADHD scores within each gender category. This separation is evident as subjects with low, moderate, and high ADHD severity cluster distinctly, thus demonstrating the efficacy of the method in

distinguishing severity levels among the ADHD-C subjects. For further comparison, the right part of Figure 3.1 presents the first two sufficient predictors resulting from τ-FCQS, i.e., without considering the categorical predictor 'gender'. We observe that, although there is a good separation of the subjects according to their ADHD score, there is an obvious overlap between male and female subjects.

For further comparisons, we evaluate the performance of our proposed methodology and that of FPSIR. For that, we use sufficient predictors resulting from each method and fit a local linear conditional quantile model [63]. Table 3.7 displays the root mean squared error (RMSE) for each method across five distinct quantiles, demonstrating that our method consistently surpasses FPSIR in terms of performance at quantiles $\tau = 0.1, 0.25, 0.5$, and 0.9. Additionally, at $\tau = 0.75$, our proposed method and FPSIR exhibit comparable performance.

Table 3.7: *RMSE of the local linear quantile regression model using the first two sufficient predictors constructed by FPSIR and τ-FCPQS.*

Method	0.1	0.25	0.5	0.75	0.9
FPSIR	7.06	4.34	3.81	**4.31**	7.53
τ-FCPQS	**5.62**	**3.73**	**3.19**	4.35	**5.53**

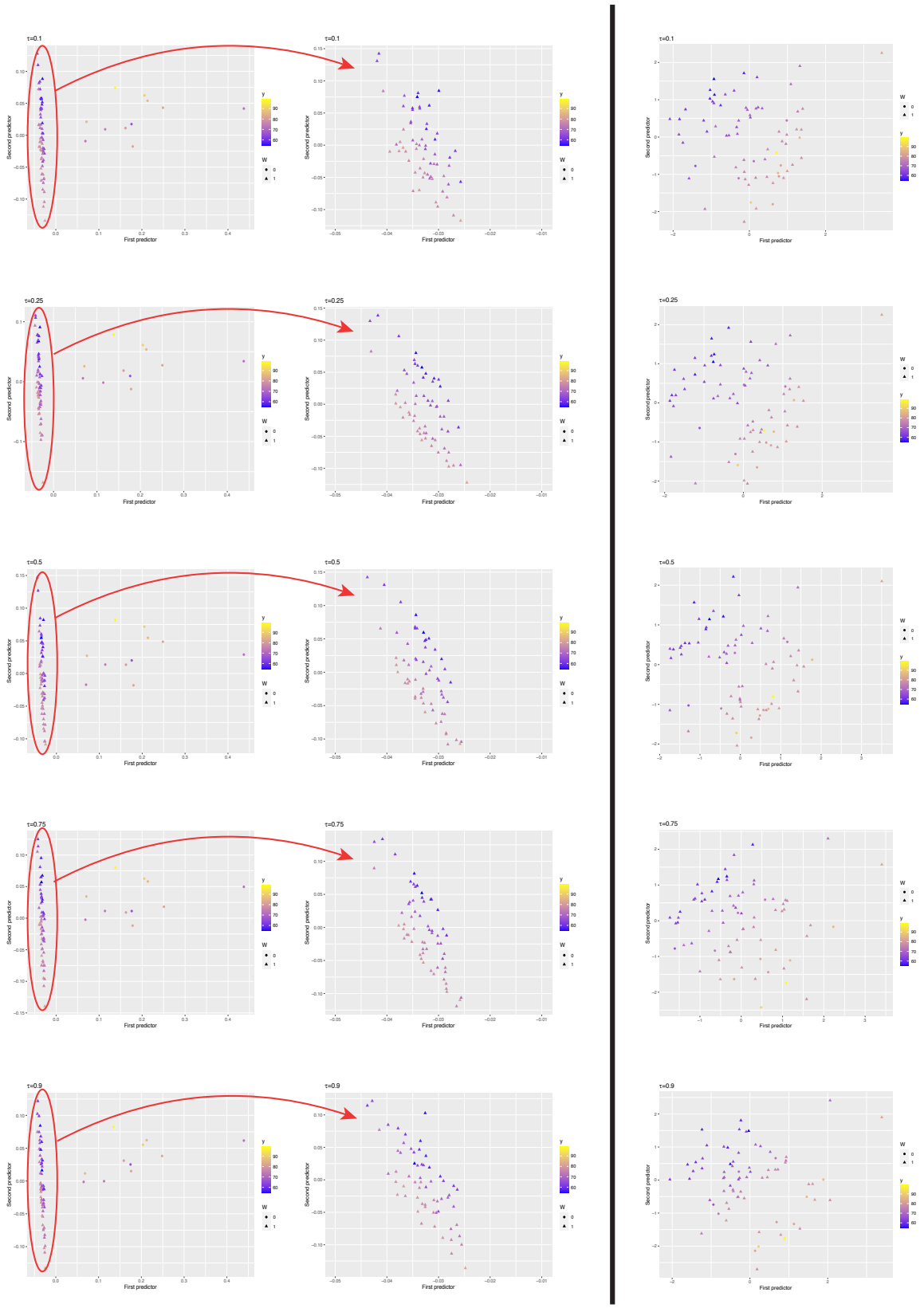

Figure 3.1: Left panel: Scatterplots of the first two sufficient predictors from the τ-FCPQS method (Chapter 3); Right panel: Scatterplots of the first two sufficient predictors from the τ-FCQS method (Chapter 2).

CHAPTER 4: The τth FUNCTIONAL GENERALIZED CENTRAL QUANTILE SUBSPACE (τ-fGCQS)

4.1 Functional Nonlinear Sufficient Dimension Reduction

Let $(\Omega, \mathcal{F}, P)$ be a probability space and for $i = 1, \ldots, p$, let $\mathcal{H}_i$ be a separable Hilbert space of real-valued functions on a bounded closed interval T, equipped with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_i}$ and an induced norm $\|\cdot\|_{\mathcal{H}_i}$. The space $\mathscr{H} = \bigoplus_{i=1}^{p} \mathcal{H}_i$ represents the direct sum of these Hilbert spaces, which is itself a Hilbert space, with an inner product $\langle \phi, \psi \rangle_{\mathscr{H}} = \langle \phi_1, \psi_1 \rangle_{\mathcal{H}_1} + \cdots + \langle \phi_p, \psi_p \rangle_{\mathcal{H}_p}$, for any $\phi = (\phi_1, \ldots, \phi_p) \in \mathscr{H}$ and $\psi = (\psi_1, \ldots, \psi_p) \in \mathscr{H}$, and a corresponding norm $\|\cdot\|_{\mathscr{H}}$. For $Y : \Omega \to \mathbb{R}$ a univariate response and $X = (X^1, \ldots, X^p) : \Omega \to \mathscr{H}$ a random element, let $\mathcal{F}_X$ be the Borel σ-field generated by open sets in $\mathscr{H}$, and P_X be the distribution of X, i.e., the induced measure $P \circ X^{-1}$ on $(\mathscr{H}, \mathcal{F}_X)$. Moreover, let $\sigma(X)$ represent the sub σ-field in $\mathcal{F}$ generated by X, that is $\sigma(X) = X^{-1}(\mathcal{F}_X)$, and $P_{Q_\tau(Y|X)|Y} : \mathcal{F}_X \times \mathbb{R} \to \mathbb{R}$ be the conditional distribution of $Q_\tau(Y|X)$ given Y.

Definition 4.1. A sub σ-field $\mathcal{G}_\tau$ of $\sigma(X)$ is a τth quantile dimension reduction σ-field for Y versus $Q_\tau(Y|X)$ if and only if

$$Y \perp\!\!\!\perp Q_\tau(Y|X) | \mathcal{G}_\tau. \tag{4.1}$$

This definition is general and includes all the situations mentioned in Section **??**. For example, if $\mathscr{H} = \mathbb{R}^p$, i.e., $\mathbf{X} : \Omega \to \mathbb{R}^p$, and $\mathcal{F}_X$ is the Borel σ-field generated by the open sets in $\mathbb{R}^p$, then by taking $\mathcal{G}_\tau = \sigma(\mathbf{B}_\tau^\top \mathbf{X})$, (4.1) reduces to the classical form (1.6). Similarly, when $X : \Omega \to \mathscr{H}$, taking $\mathcal{G}_\tau$ to be the σ-field generated by $\langle \beta_{\tau,1}, X \rangle_{\mathscr{H}}, \ldots, \langle \beta_{\tau,q_\tau}, X \rangle_{\mathscr{H}}$, for $\beta_{\tau,1}, \ldots, \beta_{\tau,q_\tau} \in \mathscr{H}$, leads to the functional case

(2.1). Thus, formulation (4.1) provides a natural extension, allowing us to generalize sufficient dimension reduction to the nonlinear case. It does so by considering an arbitrary sub σ-field $\mathcal{G}_\tau$, which does not need to be generated by linear functionals, enabling a more flexible approach to capturing complex, nonlinear relationships.

Obviously, there are many sub σ-fields of $\sigma(X)$ that satisfy (4.1). For instance, $\sigma(X)$ itself satisfies (4.1), but this choice provides no dimension reduction. Thus, the objective is to achieve the maximal dimension reduction by seeking the smallest τth quantile dimension reduction σ-field, as introduced in the next definition.

Definition 4.2. The intersection of all τth quantile dimension reduction σ-fields is called the τth central quantile σ-field and is denoted by $\mathcal{G}_{Q_\tau(Y|X)}$.

The τth central quantile σ-field $\mathcal{G}_{Q_\tau(Y|X)}$, defined as an intersection, does not necessarily satisfy (4.1). However, extending the argument in Theorem 1 of Lee et al. (2013), we can show that, if $\{P_{Q_\tau(Y|X)|Y}(\cdot|y) : y \in \mathbb{R}\}$ is dominated by a σ-finite measure, then $\mathcal{G}_{Q_\tau(Y|X)}$ also satisfies (4.1). The following theorem guarantees the existence and uniqueness of this minimal τth quantile dimension reduction σ-field.

Theorem 4.1. *Suppose that the family of probability measures* $\{P_{Q_\tau(Y|X)|Y}(\cdot|y) : y \in \mathbb{R}\}$ *is dominated by a* σ*-finite measure. Then, the* τ*th central quantile* σ*-field* $\mathcal{G}_{Q_\tau(Y|X)}$ *is such that*

(i) $Y \perp\!\!\!\perp Q_\tau(Y|X)|\mathcal{G}_{Q_\tau(Y|X)}$

(ii) if $\mathcal{G}_\tau$ *is a sub* σ*-field of* $\sigma(X)$ *such that* $Y \perp\!\!\!\perp Q_\tau(Y|X)|\mathcal{G}_\tau$*, then* $\mathcal{G}_{Q_\tau(Y|X)} \subseteq \mathcal{G}_\tau$.

Proof. See Appendix C.1. □

4.2 Nested Hilbert Space

The functional and nonlinear nature of the problem leads to the construction of two nested functional spaces: one to represent the functional data and the other to capture nonlinear functions whose domains lie in the first space. Accordingly, let $\mathscr{H}$ denote

the Hilbert space defining the range of X. To establish the second-level space $\mathfrak{M}_X$, we employ a kernel function mapping $\mathscr{H} \times \mathscr{H}$ to $\mathbb{R}$. Li and Song [54] provide such a kernel in their Definition 2. Specifically,

Definition 4.3 (Definition 2 in Li and Song [54])**.** We say that a positive definite kernel $\kappa : \mathscr{H} \times \mathscr{H} \to \mathbb{R}$ is induced by $\langle \cdot, \cdot \rangle_{\mathscr{H}}$ if there is a function $\rho : \mathbb{R}^3 \to \mathbb{R}^+$ such that, for any $\phi, \psi \in \mathscr{H}$,

$$\kappa(\phi, \psi) = \rho(\langle \phi, \phi \rangle_{\mathscr{H}}, \langle \phi, \psi \rangle_{\mathscr{H}}, \langle \psi, \psi \rangle_{\mathscr{H}}).$$

Examples of such kernels are the Gaussian radial basis function given by $\kappa(\phi, \psi) = \exp(-\gamma \left\| \phi - \psi \right\|_{\mathscr{H}}^2)$ and the polynomial kernel $\kappa(\phi, \psi) = (\alpha + \langle \phi, \psi \rangle_{\mathscr{H}})^k$, where $\gamma > 0$ is a tuning parameter, $\alpha \in \mathbb{R}$, and $k \geq 1$ is an integer.

In this work, we refer to $\mathscr{H}$ as the first-level Hilbert space and define the second-level space $\mathfrak{M}_X$ as the reproducing kernel Hilbert space (RKHS) generated by κ, i.e.,

$$\mathfrak{M}_X = \overline{\text{span}}\{\kappa(\cdot, \phi) : \phi \in \mathscr{H}\}, \quad \langle \kappa(\cdot, \phi), \kappa(\cdot, \psi) \rangle_{\mathfrak{M}_X} = \kappa(\phi, \psi).$$

Thus, determining $\mathscr{H}$ and $\mathfrak{M}_X$ ultimately depends on the choice of $\mathscr{H}$ and ρ; see Section 3 of [54] for specific examples. Throughout the rest of the paper, we use symbols such as ϕ and ψ to denote functions in the first-level Hilbert space, and f and g to represent functions in the second-level Hilbert space. Note that, while $\mathscr{H}$ is assumed to be any separable Hilbert space, $\mathfrak{M}_X$ is specifically required to be an RKHS.

To transition our focus to estimating functions, we reframe the problem of estimating the τth central quantile σ-field, a rather abstract object to estimate, into the task of estimating a subspace, specifically a subset of $\mathfrak{M}_X$. To achieve this, we consider $L_2(P_X)$, the class of all square-integrable functions of X with respect to P_X,

i.e., $Ef^2(X) < \infty$ under P_X. Note that, since constants are irrelevant for dimension reduction, we assume throughout that all functions in $L_2(P_X)$ have mean 0; see also Section 4.3. Now, suppose the following assumption holds:

Assumption 4.1. $\mathfrak{M}_X$ is a dense subset of $L_2(P_X)$ modulo constants; that is, for any $f \in L_2(P_X)$, there is a sequence $\{f_n\} \subseteq \mathfrak{M}_X$ such that $\text{var}\{f_n(X) - f(X)\} \to 0$.

This assumption guarantees that the second-level space $\mathfrak{M}_X$, which is generated by the kernel κ, characterizes the τth central quantile σ-field. Note that the kernel κ is required to be a universal kernel with respect to the $L_2(P_X)$-norm, which means that $\mathfrak{M}_X$ is rich enough to approximate any $L_2(P_X)$ function arbitrarily closely. For example, this condition is satisfied by the Gaussian radial basis function kernel, but not by the polynomial kernel. This assumption is standard and commonly used in RKHS; see [54], [70], [71]. In many practical settings, the assumption is mild. For example, if $\mathfrak{M}_X$ contains all continuous functions on a compact domain or all functions in a universal RKHS, the density property holds automatically. We have added a short illustrative example.

Example 4.1. Let ϕ, ψ be random functions taking values in a compact subset $\mathcal{X}$ of L_2, and let $\kappa(\phi, \psi) = \exp(-\gamma \|\phi - \psi\|_{L_2}^2)$ be a Gaussian radial basis functional kernel defined on $\mathcal{X}$. On a compact metric space, the Gaussian kernel is universal, meaning its RKHS is dense in $C(\mathcal{X})$ under the sup norm, and therefore, dense in $L_2(P_X)$. It follows that $\mathfrak{M}_X$ is dense in $L_2(P_X)$ modulo constants, and Assumption 4.1 holds.

Under Assumption 4.1, the τth central quantile σ-field is generated by a finite set of functions $f_{\tau,1}, \ldots, f_{\tau,\widetilde{q}_\tau}$ in $\mathfrak{M}_X$, satisfying

$$Y \perp\!\!\!\perp Q_\tau(Y|X) | f_{\tau,1}(X), \ldots, f_{\tau,\widetilde{q}_\tau}(X). \tag{4.2}$$

That is, equation (4.2) defines the τth central quantile σ-field as $\mathcal{G}_{Q_\tau(Y|X)} = \sigma[f_{\tau,1}(X), \ldots, f_{\tau,\widetilde{q}_\tau}(X)]$. This alternative representation gives the following definition.

Definition 4.4. Under Assumption 4.1, the class of functions in $\mathfrak{M}_X$ that are $\mathcal{G}_\tau$-measurable is called the τth quantile dimension reduction class, and the class of functions in $\mathfrak{M}_X$ that are $\mathcal{G}_{Q_\tau(Y|X)}$-measurable is called the τth central quantile class, denoted by $\mathfrak{G}_{Q_\tau(Y|X)}$. That is,

$$\mathfrak{G}_{Q_\tau(Y|X)} = \overline{\text{span}}\{f_\tau \in \mathfrak{M}_X : f_\tau \text{ is measurable } \mathcal{G}_{Q_\tau(Y|X)}\}.$$

Our goal is to recover $\mathfrak{G}_{Q_\tau(Y|X)}$ from a random sample of (X, Y). If a subspace $\mathfrak{G}_\tau$ of $\mathfrak{M}_X$ is contained in $\mathfrak{G}_{Q_\tau(Y|X)}$, then we say it is unbiased. If it is equal to $\mathfrak{G}_{Q_\tau(Y|X)}$, then we say it is exhaustive.

Equation (4.2) implies that the conditional quantile $Q_\tau(Y|X)$ is a function of $f_{\tau,1}(X), \ldots, f_{\tau,\widetilde{q}_\tau}(X)$, i.e., $Q_\tau(Y|X) = g_\tau(f_{\tau,1}(X), \ldots, f_{\tau,\widetilde{q}_\tau}(X)) + \epsilon$, for some function g_τ. However, the nonlinear nature of the problem allows us to capture the information in one function and express $g_\tau(f_{\tau,1}(X), \ldots, f_{\tau,\widetilde{q}_\tau}(X))$ as $g_\tau^*(f_\tau^*(X))$. Therefore, in this work, we focus on the *single-index* model

$$Y \perp\!\!\!\perp Q_\tau(Y|X)|f_\tau^*(X). \tag{4.3}$$

Remark 4.1. The reproducing nature of the RKHS $\mathfrak{M}_X$ implies that equation (4.2) can be stated as

$$Y \perp\!\!\!\perp Q_\tau(Y|X)|\langle f_{\tau,1}, \kappa(\cdot, X)\rangle_{\mathfrak{M}_X}, \ldots, \langle f_{\tau,\widetilde{q}_\tau}, \kappa(\cdot, X)\rangle_{\mathfrak{M}_X},$$

which resembles the linear sufficient dimension reduction problem (1.6).

4.3 Functional generalized central quantile subspace and its population-level properties

Consider two generic Hilbert spaces, $\mathcal{H}$ and $\mathcal{K}$. Let $\mathcal{B}(\mathcal{H}, \mathcal{K})$ represent the class of bounded linear operators from $\mathcal{H}$ to $\mathcal{K}$. In the special case where $\mathcal{H} = \mathcal{K}$, we use

the shorthand notation $\mathcal{B}(\mathcal{H})$ instead of $\mathcal{B}(\mathcal{H},\mathcal{H})$. For any operator $A \in \mathcal{B}(\mathcal{H},\mathcal{K})$, we denote its adjoint operator by A^*, its kernel by $\ker(A)$, its range by $\mathrm{ran}(A)$, and the closure of its range by $\overline{\mathrm{ran}}(A)$. Importantly, if A is a self-adjoint operator A, then $\overline{\mathrm{ran}}(A)$ is the orthogonal complement of $\ker(A)$, i.e., $\ker(A)^\perp = \overline{\mathrm{ran}}(A)$. Finally, if $\mathcal{S}$ is a subset of $\mathrm{dom}(A)$, then we use $A|\mathcal{S}$ to denote the operator restricted on $\mathcal{S}$, i.e., $(A|\mathcal{S}) : \mathcal{S} \to \mathcal{K}$. For a linear subspace $\mathcal{S}$ of $\mathcal{H}$, the projection operator onto $\mathcal{S}$ will be written as $P_{\mathcal{S}}$.

Assumption 4.2. There is a constant $C_1 > 0$ such that, for any $f \in \mathfrak{M}_X$, $\mathrm{var}\{f(X)\} \leq C_1 \|f\|^2_{\mathfrak{M}_X}$.

Define $L_2^{(c)}(P_X) = \{f - Ef(X) : f \in L_2(P_X)\}$ as the centered $L_2(P_X)$ space. Assumption 4.2 ensures that the mapping $\mathscr{H} \to L_2^{(c)}(P_X)$, $f \to f$ is continuous and guarantees that the bilinear form $\mathfrak{M}_X \times \mathfrak{M}_X \to \mathbb{R}$, $(f,g) \to cov\{f(X), g(X)\}$ is bounded. As a result, it induces an operator $\Sigma_{XX} \in \mathcal{B}(\mathfrak{M}_X)$ such that $\langle f, \Sigma_{XX} g\rangle_{\mathfrak{M}_X} = cov\{f(X), g(X)\}$. By definition, Σ_{XX} is a bounded, self-adjoint, and positive semi-definite operator, known as the variance operator of X. Moreover, for any $f \in \mathfrak{M}_X$, we have $\langle f, \Sigma_{XX} f\rangle_{\mathfrak{M}_X} = cov\{f(X), f(X)\} = var\{f(X)\}$.

Although Σ_{XX} is defined on $\mathfrak{M}_X$, its effective domain is the space $\overline{\mathrm{ran}}(\Sigma_{XX})$, since members of $\ker(\Sigma_{XX})$ are constants almost surely. Under Assumptions 4.1 and 4.2, the covariance operator $\Sigma_{XX} : \overline{\mathrm{ran}}(\Sigma_{XX}) \to \overline{\mathrm{ran}}(\Sigma_{XX})$ is defined through the relation

$$\langle f, \Sigma_{XX} g\rangle_{\overline{\mathrm{ran}}(\Sigma_{XX})} = \langle f, g\rangle_{L_2(P_X)},$$

for $f, g \in L_2^{(c)}(P_X)$. The subspace $\overline{\mathrm{ran}}(\Sigma_{XX})$ has an explicit expression, as described in Lemma 1 of [54]. Specifically, under Assumption 4.2, we have $\overline{\mathrm{ran}}(\Sigma_{XX}) = \mathfrak{M}_X^0$, where

$$\mathfrak{M}_X^0 = \overline{\mathrm{span}}\{\kappa(\cdot, x) - \mu_X : x \in \mathscr{H}\} \tag{4.4}$$

and μ_X is the linear functional $f \to E\{f(X)\}$ from $\mathfrak{M}_X$ to $\mathbb{R}$. It is important to note that $E\{f(X)\}$ is bounded since, under Assumption 4.2, there exists a constant $C_2 > 0$ such that $E|f(X)| \leq C_2 \|f\|_{\mathfrak{M}_X}$ for any $f \in \mathfrak{M}_X$.

Expression (4.4) gives a way to construct the second-level Hilbert space using the kernel κ. Specifically, we define

$$\mathfrak{M}_X = \overline{\text{span}}\{\kappa(\cdot, \phi) : \phi \in \mathscr{H}\}, \quad \langle \kappa(\cdot, \phi), \kappa(\cdot, \psi) \rangle_{\mathfrak{M}_X} = \kappa(\phi, \psi),$$

and $\mathfrak{M}_X^0$ as defined in (4.4). Then, the pair $(\mathscr{H}, \mathfrak{M}_X)$ is a pair of nested Hilbert spaces; see Definition 1 of [72].

Since any $f \in \text{ran}(\Sigma_{XX})$ can be uniquely decomposed as $\Sigma_{XX}(g_1 + g_2)$, where $g_1 \in \ker(\Sigma_{XX})$ and $g_2 \in \overline{\text{ran}}(\Sigma_{XX}) = \mathfrak{M}_X^0$, the mapping $f \to g_2$ from $\text{ran}(\Sigma_{XX})$ to $\mathfrak{M}_X^0$ is well-defined. This mapping is referred to as the Moore-Penrose inverse of Σ_{XX}, denoted by $\Sigma_{XX}^\dagger$. Therefore, $\Sigma_{XX}^\dagger$ is an operator from $\text{ran}(\Sigma_{XX})$ to $\overline{\text{ran}}(\Sigma_{XX})$ such that, for any $f \in \text{ran}(\Sigma_{XX})$, $\Sigma_{XX}^\dagger f$ is the unique member of $\overline{\text{ran}}(\Sigma_{XX})$ such that $\Sigma_{XX}(\Sigma_{XX}^\dagger f) = f$. Note that, $\Sigma_{XX}^\dagger$ is typically an unbounded operator and makes its estimation difficult. However, $\Sigma_{XX}^\dagger$ often appears in combination with another operator, and the composite operator is typically bounded.

We now present the main results on how to estimate $\mathfrak{G}_{Q_\tau(Y|X)}$ for the single-index model (4.3).

Theorem 4.2. *Under Assumptions 4.1 and 4.2, we have that* $\widetilde{f}_\tau \in \mathfrak{G}_{Q_\tau(Y|X)}$*, where*

$$\begin{aligned}(\widetilde{\alpha}_\tau, \widetilde{f}_\tau) &= \arg \min_{(a_\tau, f_\tau) \in \mathbb{R} \times \mathfrak{M}_X} E[Q_\tau\{Y|T^{CS}(X)\} - a_\tau - \langle f_\tau, \kappa(\cdot, X) \rangle_{\mathfrak{M}_X}]^2 \\ &= \arg \min_{(a_\tau, f_\tau) \in \mathbb{R} \times \mathfrak{M}_X} E[Q_\tau\{Y|T^{CS}(X)\} - a_\tau - f_\tau(X)]^2 \qquad (4.5)\end{aligned}$$

and T^{CS} *is a set of functions in the central class* $\mathfrak{G}_{Y|X}$*; see [53] and [54].*

Proof. See Appendix C.2.

For what follows, we will call the proposed methodology the τth functional generalized central quantile subspace (τ-fGCQS).

4.4 Implementation

We now present the implementation of the τ-fGCQS at the sample level, which will be achieved by replacing characteristics of a distribution with those of the empirical distribution and expressing relevant operators as matrices using a coordinate representation system; see details in Appendix A.5.

4.4.1 Construction of the Nested Hilbert Spaces

We now turn our attention to the construction of $\mathscr{H}$ and $\mathfrak{M}_X$ using an independent and identically distributed (iid) sample of (X, Y). Note that, the functions $X_u(t)$, $u = 1, \ldots, n$, are not fully observed; instead, they are observed on a finite set of points $t_{u1}, \ldots, t_{uN_u}$. Therefore, $X_u(t)$ need to be estimated using the observed data $\{(t, X_u(t)) : t = t_{u1}, \ldots, t_{uN_u}\}$. For simplicity, assume that all subjects are observed at the same time points, i.e., $t = t_1, \ldots, t_N$. For the case where the sets of observed time points can vary from subject to subject, the following development can be modified using similar arguments as [54].

First-level Hilbert space. We begin by constructing the first-level space $\mathscr{H}$. Since X is p-dimensional, with p possibly greater than 1, we first construct $\mathcal{H}_i$, for $i = 1, \ldots, p$. Let K_T be the $N \times N$ Gram matrix $\{\kappa_T(t_k, t_l) : k, l : 1, \ldots, N\}$ and, using the coordinate representation given in (A.13), we define the inner product in $\mathcal{H}_i$ as

$$\langle \phi_i, \psi_i \rangle_{\mathcal{H}_i} = \Big\langle \sum_{k=1}^{N} ([\phi_i])_k \kappa_T(\cdot, t_k), \sum_{l=1}^{N} ([\psi_i])_l \kappa_T(\cdot, t_l) \Big\rangle_{\mathcal{H}_i} = [\phi_i]^\top K_T [\psi_i].$$

Then, for each $i = 1, \ldots, p$, the function X_u^i can be expressed as $X_u^i = \sum_{k=1}^{N} ([X_u^i])_k \kappa_T(\cdot, t_k)$,

and its evaluation at time point t_l as

$$X_u^i(t_l) = \langle X_u^i, \kappa_T(\cdot, t_l)\rangle_{\mathcal{H}_i} = \sum_{k=1}^{N}([X_u^i])_k \kappa_T(t_l, t_k). \tag{4.6}$$

Therefore, for $T^* = (t_1, \ldots, t_N)^\top$ and $X_i = X_u^i(T^*) = \big(X_u^i(t_1), \ldots, X_u^i(t_N)\big)^\top$, we have that $X_u^i(T^*) = K_T[X_u^i]$ and the coordinate representation of X_u^i is $[X_u^i] = K_T^{-1} X_u^i(T^*)$. To avoid overfitting, we apply the Tychonoff regularization, which adds a multiple of the identity matrix to the sample covariance matrix before inverting it, to get

$$[X_u^i] = (K_T + \varepsilon_T I_N)^{-1} X_u^i(T^*),$$

where the regularization parameter ϵ_T is discussed in Section 4.4.3. Therefore, we define $\mathcal{H}_i$ as the space spanned by $\{\sum_{l=1}^{N}([X_u^i])_l \kappa_T(\cdot, t_l) : u = 1, \ldots, n\}$, with the inner product determined by $\langle \kappa_T(\cdot, s), \kappa_T(\cdot, t)\rangle_{\mathcal{H}_i} = \kappa_T(s, t)$. Thus,

$$\begin{aligned} \langle X_u^i, X_{u'}^i\rangle_{\mathcal{H}_i} &= [X_u^i]^\top K_T [X_{u'}^i] \\ &= X_u^{i\top}(T^*)(K_T + \varepsilon_T I_N)^{-1} K_T (K_T + \varepsilon_T I_N)^{-1} X_{u'}^i(T^*). \end{aligned} \tag{4.7}$$

Finally, we define $\mathscr{H} = \bigoplus_{i=1}^{p} \mathcal{H}_i$, and the inner product between two functions $\phi, \psi \in \mathscr{H}$ is given by $\langle \phi, \psi\rangle_{\mathscr{H}} = \sum_{i=1}^{p} \langle \phi_i, \psi_i\rangle_{\mathcal{H}_i}$. Therefore, equation (4.7) leads to relation $\langle X_u, X_{u'}\rangle_{\mathscr{H}} = \sum_{i=1}^{p} \langle X_u^i, X_{u'}^i\rangle_{\mathcal{H}_i}$.

Second-level Hilbert Space. We now focus on the construction of the second-level space $\mathfrak{M}_X$. First, note that, for $u, u' = 1, \ldots, n$

$$\begin{aligned} \kappa_X(X_u, X_{u'}) &= \rho(\langle X_u - X_{u'}, X_u - X_{u'}\rangle_{\mathscr{H}}) \\ &= \rho(\langle X_u, X_u\rangle_{\mathscr{H}} - 2\langle X_u, X_{u'}\rangle_{\mathscr{H}} + \langle X_{u'}, X_{u'}\rangle_{\mathscr{H}}), \end{aligned}$$

where $\langle X_u, X_{u'} \rangle_{\mathscr{H}}$ is defined in (4.7) and ρ is introduced in Definition 4.3. Then, $\mathfrak{M}_X$ is the space spanned by $\{\kappa_X(\cdot, X_u) : u = 1, \ldots, n\}$ and K_X is the $n \times n$ Gram matrix $\{\kappa_X(X_u, X_{u'}) : u, u' = 1, \ldots, n\}$. For $f, g \in \mathfrak{M}_X$, the inner product is given by

$$\langle f, g \rangle_{\mathfrak{M}_X} = [f]^\top K_X [g]. \tag{4.8}$$

4.4.2 Implementation of the τ-fGCQS

As mentioned earlier, the relevant subspace for dimension reduction is $\mathfrak{M}_X^0$ because functions in its orthogonal complement are constant almost surely P_X. Let μ_X be the Riesz representation of the linear functional $f \to E_n f(X)$ for $f \in \mathscr{H}$, and let $\mathfrak{M}_X^0 = \text{span}\{\kappa_X(\cdot, X_u) - \mu_X : u = 1, \ldots, n\}$, where μ_X can be expressed as $n^{-1} \sum_{u=1}^n \kappa_X(\cdot, X_u)$; see Proposition 2 of [54].

The sample level version of (4.5) is given by

$$(\widehat{\widetilde{\alpha}}_\tau, \widehat{\widetilde{f}}_\tau) = \arg \min_{(a_\tau, f_\tau) \in \mathbb{R} \times \mathfrak{M}_X} \sum_{u=1}^n \left[\widehat{Q}_\tau \{Y | \widehat{T}^{CS}(X_u)\} - a_\tau - f_\tau(X_u) \right]^2, \tag{4.9}$$

where $\widehat{T}^{CS}(X)$ results from the generalized functional SIR of [54] and $\widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}$ is a nonparametric estimate of $Q_\tau\{Y|\widehat{T}^{CS}(X_u)\}$. For that, we use the local linear conditional quantile estimation method of [63]. Then, using the derivations of Section

4.4.1, we have that the coordinate representation of (4.9) is

$$\begin{aligned}
[(\widehat{\widetilde{\alpha}}_\tau, \widehat{\widetilde{f}}_\tau)] &= (\widehat{\widetilde{\alpha}}_\tau, [\widehat{\widetilde{f}}_\tau]) \\
&= \arg \min_{(a_\tau, f_\tau) \in \mathbb{R} \times \mathfrak{M}_X} \sum_{u=1}^{n} \left\{ \widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\} - a_\tau - [f_\tau(X_u)] \right\}^2 \\
&= \arg \min_{(a_\tau, f_\tau) \in \mathbb{R} \times \mathfrak{M}_X} \sum_{u=1}^{n} \left\{ \widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\} - a_\tau - [\langle f_\tau, \kappa_X(\cdot, X_u)\rangle_{\mathfrak{M}_X}] \right\}^2 \\
&= \arg \min_{(a_\tau, f_\tau) \in \mathbb{R} \times \mathfrak{M}_X} \sum_{u=1}^{n} \left\{ \widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\} - a_\tau - [f_\tau]^\top K_X [\kappa_X(\cdot, X_u)] \right\}^2 \\
&= \arg \min_{(a_\tau, f_\tau) \in \mathbb{R} \times \mathfrak{M}_X} \sum_{u=1}^{n} \left\{ \widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\} - a_\tau - [f_\tau]^\top K_X e_u \right\}^2 \\
&= \arg \min_{(a_\tau, f_\tau) \in \mathbb{R} \times \mathfrak{M}_X} \sum_{u=1}^{n} \left\{ \widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\} - a_\tau - [f_\tau]^\top K_X^{(u)} \right\}^2,
\end{aligned}$$

where $K_X^{(u)}$ denotes the u-th column of the K_X matrix, and e_u is the u-th basis vector. The fourth equality follows from the representation in (4.8), while the fifth equality uses the fact that the coordinate representation of $\kappa_X(\cdot, X_u)$ corresponds to the u-th basis vector.

4.4.3 Tuning parameters

This section focuses on the selection of tuning parameters, which are organized into two levels. For the first-level Hilbert space $\mathscr{H}$, we need to select the Tychonoff regularization constant ε_T, and, when using the Gauss radial basis (GRB) function, the constant γ_T. For the second-level space $\mathfrak{M}_X$, we select the parameters γ_X and ε_X. For the purpose of this section, we assume the use of GRB kernel, though the same ideas can be applied to other kernels.

Tuning first-level functions. We begin with the first-level Hilbert space. Recall that K_T is the $N \times N$ Gram matrix $\{\kappa_T(t_k, , t_l) : k, l : 1, \ldots, N\}$, where the GRB is given by $\kappa_T(t_k, t_l) = \exp(-\gamma_T |t_k - t_l|^2)$. In practice, K_T^{-1} is replaced by its regularized

inverse $(K_T + \varepsilon_T I_N)^{-1}$ to avoid overfitting, so we focus on tuning the parameters $(\gamma_T, \varepsilon_T)$. To emphasize the dependence of K_T on γ_T, we use the notation $K_T(\gamma_T)$ throughout this section.

Following the recommendations of [54], we choose the initial values $\gamma_T^0 = 1/(2\rho)$, where $\rho = \binom{N}{2}^{-1} \sum_{k<l}(t_k - t_l)^2$, and $\varepsilon_T^0 = 0.05\widehat{\lambda}_1(K_T)$, where $\widehat{\lambda}_1(K_T)$ is the largest eigenvalue of K_T. These default choices are intuitive, γ_T^0 reflects the typical spacing of the design points, and ε_T^0 provides a small amount of regularization relative to the scale of K_T.

One approach is to use the initial values as the tuning parameters. However, we will instead use the generalized cross validation (GCV) criterion to tune the parameters around the default values. The idea of GCV is to provide a way to select smoothing parameters by balancing goodness of fit with model complexity. Specifically, let

$$\mathrm{gcv}(\gamma_T, \varepsilon_T) = \frac{N}{\mathrm{tr}\left\{I_N - \widehat{H}(\gamma_T, \varepsilon_T)\right\}^2} \sum_{u=1}^{n} \left\| \left\{I_N - \widehat{H}(\gamma_T, \varepsilon_T)\right\} X_u^i(T^*) \right\|^2,$$

where $\widehat{H}(\gamma_T, \varepsilon_T)$ is the matrix $K_T(\gamma_T)\left(K_T(\gamma_T) + \varepsilon_T I_N\right)^{-1}$, a smoothed projector that projects the random function X_u^i onto the important eigenfunctions of the kernel K_T.

Then, we minimize $\mathrm{gcv}(\gamma_T, \varepsilon_T)$ over a two-dimensional grid of γ_T in $[\gamma_T^0/20, 20\gamma_T^0]$ and a grid of ε_T in $[\varepsilon_T^0/50, 50\varepsilon_T^0]$ using the Gauss-Seidel method. Each grid consists of 20 points, equally spaced in log scale.

Tuning second-level functions. We now turn to the tuning parameters for second-level functions. Recall that K_X is the $n \times n$ Gram matrix $\{\kappa_X(X_u, X_{u'}) : u, u' = 1, \ldots, n\}$, where the GRB is given by $\kappa_X(X_u, X_{u'}) = \exp(-\gamma_X \left\|X_u - X_{u'}\right\|_{\mathscr{H}}^2)$, and K_X^{-1} is replaced by its regularized inverse $(K_X + \varepsilon_X I_n)^{-1}$. Therefore, we focus on tuning the parameters $(\gamma_X, \varepsilon_X)$.

Following the recommendations of [54], we choose the initial values γ_X^0 and ε_X^0 the

same as γ_T^0 and ε_T^0, but with K_T replaced by K_X, N replaced with n, and $(t_k - t_l)^2$ replaced by $\|X_u - X_{u'}\|^2$. Then, we use the leave-one-out cross validation (LOOCV) criterion to tune the parameters around the default values. LOOCV evaluates predictive performance by repeatedly leaving out one subject and predicting its response from the remaining sample. Specifically, let LOOCV defined by

$$\sum_{u=1}^{n} \{Y_u - \widehat{E}^{(-u)}(Y_u|X_u)\}^2,$$

where $\widehat{E}^{(-u)}(Y|X)$ is some nonparametric estimate of the conditional expectation $E(Y|X)$ based on the sample with the uth subject removed. This criterion balances fit with generalizability by penalizing parameter choices that overfit the training data but perform poorly for new observations.

Then, we minimize $\text{loocv}(\gamma_X, \varepsilon_X)$ over a two-dimensional grid of γ_X in $[\gamma_X^0/20, 20\gamma_X^0]$ and a grid of ε_X in $[\varepsilon_X^0/50, 50\varepsilon_X^0]$ using the Gauss-Seidel method. Each grid consists of 20 points, equally spaced in log scale. This procedure ensures that the smoothing and regularization levels are chosen adaptively according to the data.

4.5 Asymptotic Results

Theorem 4.3. *If assumptions 4.1, 4.2, and S1-S6 from Appendix A hold, then, for a given $\tau \in (0, 1)$,*

$$\left\| \widehat{\widetilde{f}}_\tau - \widetilde{f}_\tau \right\|_{\mathfrak{M}_X} = O_p(n^{-1/2}),$$

where $\widetilde{f}_\tau$ is given in (4.5), and $\widehat{\widetilde{f}}_\tau$ is given in (4.9).

Proof. See Appendix C.3.

This result establishes the consistency and convergence rate of the estimated functions $\widehat{\widetilde{f}}_\tau$. Specifically, under the stated assumptions, $\widehat{\widetilde{f}}_\tau$ converges to $\widetilde{f}_\tau$ at a rate of $n^{-1/2}$

as the sample size n increases.

4.6 Simulation Results

4.6.1 Computational Remarks

Algorithm and parameters. Equation (4.9) requires the estimation of $\widehat{T}^{CS}(X_u)$, which is obtained using the f-GSIR of [54], and the estimation of $\widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}$, which is obtained using the local linear conditional quantile estimation method of [63]. For that, we use a Gaussian kernel and a bandwidth based on the rule-of-thumb in Yu and Jones [64].

Simulation Setup. The simulation setting is the same as that used for the τ-FCQS method in Section 2.5.1.

Evaluation and simulation parameters. For the estimation accuracy, we will use the Multiple Correlation of Multivariate Rank (MCMR), a multivariate generalization of Spearman's rank correlation introduced by [54]. This measure extends the idea of [53], who employed Spearman's correlation to quantify the dependence between estimated and true predictors in scalar-on-scalar settings. Like Spearmanâs correlation, MCMR is computed on ranks and is therefore invariant under monotonic transformations of each coordinate, making it well suited for performance assessment. MCMR generalizes this idea to multivariate predictors, providing a robust and flexible measure of dependence between the estimated and true subspaces. Specifically, for two samples of random vectors $U_1, \ldots, U_n \in \mathbb{R}^r$ and $V_1, \ldots, V_n \in \mathbb{R}^s$, we define their multivariate ranks $\widetilde{U}$ and $\widetilde{V}$ as

$$\widetilde{U}_i = n^{-1} \sum_{l=1}^{n} (U_l - U_i) / \left\| U_l - U_i \right\|, \quad \widetilde{V}_i = n^{-1} \sum_{l=1}^{n} (V_l - V_i) / \left\| V_l - V_i \right\|.$$

Then, the MCMR between $\{U_1, \ldots, U_n\}$ and $\{V_1, \ldots, V_n\}$ is the multiple correlation

between the multivariate ranks of the two samples:

$$mcmr_n(U,V) = \left[tr[\{var_n(\widetilde{V})\}^{-1/2} cov_n(\widetilde{V},\widetilde{U})\{var_n(\widetilde{U})\}^{-1} cov_n(\widetilde{U},\widetilde{V})\{var_n(\widetilde{V})\}^{-1/2}]\right]^{1/2}.$$

MCMR ranges from 0 to d, the dimension of the vectors, with values near d indicating better performance.

Unless otherwise stated, simulations are based on 100 iterations, $n = 200$, $p = 10$, and $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$.

Models under consideration. We consider the following models

$$\begin{aligned}
\text{M-I:} \quad & Y = \frac{1}{0.5 + (\langle \beta_1, X\rangle_{\mathscr{H}} + 1)^2} + 0.2\epsilon, \\
\text{M-II:} \quad & Y = 0.5 + (\langle \beta_1, X\rangle_{\mathscr{H}} + 1.5)^2 + \epsilon, \\
\text{M-III:} \quad & Y = \sin(\pi\langle \beta_1, X\rangle_{\mathscr{H}}/4) + 0.5\epsilon, \\
\text{M-IV:} \quad & Y = \frac{e^{\langle \beta_1, X\rangle_{\mathscr{H}}}}{1 + e^{\langle \beta_1, X\rangle_{\mathscr{H}}}} + 0.1\epsilon, \\
\text{M-V:} \quad & Y = \arctan(\pi\langle \beta_1, X\rangle_{\mathscr{H}}) + 0.5\sin(\pi\langle \beta_2, X\rangle_{\mathscr{H}}/6) + 0.1\epsilon, \\
\text{M-VI:} \quad & Y = \langle \beta_1, X\rangle_{\mathscr{H}} + \langle \beta_2, X\rangle_{\mathscr{H}} + \epsilon, \\
\text{M-VII:} \quad & Y = \frac{\langle \beta_1, X\rangle_{\mathscr{H}}}{1 + e^{\langle \beta_2, X\rangle_{\mathscr{H}}}} + \epsilon, \\
\text{M-VIII:} \quad & Y = \frac{\langle \beta_1, X\rangle_{\mathscr{H}}}{1 + e^{\langle \beta_2, X\rangle_{\mathscr{H}}}} + \langle \beta_3, X\rangle_{\mathscr{H}} + \epsilon,
\end{aligned}$$

where $\beta_1(t) = \psi_1(t)$, $\beta_2(t) = \psi_2(t)$, and $\beta_3(t) = \psi_3(t)$ are the first three eigenfunctions of Σ_{XX}, X is simulated as described above, and the error ϵ is generated from standard normal distribution. Results are available for all models, but we report only a selection for brevity, as they exhibit similar patterns.

4.6.2 Results

Example 1 - Effect of choice of kernel. We examine how the choice of kernel influences the algorithm's performance, focusing on two options for computing K_T:

the GRB kernel, defined as $\kappa_T(t_k, t_l) = \exp(-\gamma_T|t_k - t_l|^2)$, and the Brownian kernel, given by $\kappa_T(t_k, t_l) = \min(t_k, t_l)$. Table 4.1 reports the means and standard deviations of the MCMR across Models I-VIII for various τ values. While the overall performance is comparable between the two kernels, the GRB kernel generally shows a slight edge. Notably, when the GRB kernel outperforms the Brownian kernel, the improvement is more pronounced. Based on these results, we adopt the GRB kernel for subsequent simulations.

Table 4.1: *Mean (and standard deviation) of MCMR for Models I-VIII using different kernels for computing K_T.*

Model	Kernel	0.1	0.25	0.5	0.75	0.9
I	GRB	0.8472 (0.0613)	0.8480 (0.0610)	0.8523 (0.0616)	0.8544 (0.0620)	0.8551 (0.0617)
	Brownian	0.7385 (0.0726)	0.7403 (0.0701)	0.7399 (0.0708)	0.7402 (0.0703)	0.7404 (0.0703)
II	GRB	0.7665 (0.0807)	0.7666 (0.0808)	0.7666 (0.0808)	0.7666 (0.0808)	0.7666 (0.0808)
	Brownian	0.8146 (0.0868)	0.8147 (0.0868)	0.8146 (0.0868)	0.8146 (0.0868)	0.8146 (0.0868)
III	GRB	0.9206 (0.0249)	0.9209 (0.0246)	0.9210 (0.0245)	0.9208 (0.0247)	0.9205 (0.0251)
	Brownian	0.8602 (0.0363)	0.8601 (0.0362)	0.8603 (0.0361)	0.8605 (0.0363)	0.8602 (0.0365)
IV	GRB	0.9719 (0.0081)	0.9719 (0.0082)	0.9719 (0.0082)	0.9719 (0.0082)	0.9719 (0.0082)
	Brownian	0.9674 (0.0071)	0.9674 (0.0070)	0.9674 (0.0071)	0.9674 (0.0071)	0.9674 (0.0071)
V	GRB	0.8883 (0.0204)	0.8881 (0.0205)	0.8879 (0.0206)	0.8881 (0.0206)	0.8881 (0.0206)
	Brownian	0.8905 (0.0194)	0.8904 (0.0195)	0.8903 (0.0195)	0.8903 (0.0195)	0.8904 (0.0195)
VI	GRB	0.9150 (0.0951)	0.9150 (0.0951)	0.9150 (0.0951)	0.9150 (0.0951)	0.9150 (0.0951)
	Brownian	0.9324 (0.0287)	0.9324 (0.0287)	0.9324 (0.0287)	0.9324 (0.0287)	0.9324 (0.0287)
VII	GRB	0.8055 (0.0329)	0.8055 (0.0329)	0.8054 (0.0329)	0.8054 (0.0328)	0.8053 (0.0329)
	Brownian	0.7325 (0.0465)	0.7325 (0.0465)	0.7326 (0.0466)	0.7327 (0.0464)	0.7329 (0.0463)
VIII	GRB	0.9036 (0.0233)	0.9036 (0.0233)	0.9036 (0.0233)	0.9036 (0.0233)	0.9036 (0.0233)
	Brownian	0.8934 (0.0225)	0.8934 (0.0225)	0.8934 (0.0225)	0.8935 (0.0225)	0.8934 (0.0225)

Example 2 - Effect of n and p. We evaluate the algorithm's performance across sample sizes $n = 100, 200$ and varying number of predictors $p = 5, 10, 20, 40$. Table 4.2 reports the means and standard deviations of the MCMR for Models III and VII at different τ values. The results indicate that the algorithm's efficiency improves with larger sample sizes (n) and declines as the number of predictors (p) increases.

Example 3 - Compare performance with other methods. To evaluate the performance of the algorithm, we compare it with existing methods. However, to the best of our knowledge, no other nonlinear dimension reduction technique specifically targets conditional quantiles. For illustration purposes, we compare our approach with the f-GSIR and f-GSAVE methods ([54]). Since our method focuses on condi-

Table 4.2: *Mean (and standard deviation) of MCMR for Model III and VII for various n and p.*

Model	n	p	0.1	0.25	0.5	0.75	0.9
III	100	5	0.9325 (0.0357)	0.9336 (0.0353)	0.9338 (0.0351)	0.9336 (0.0351)	0.9332 (0.0354)
		10	0.9021 (0.0422)	0.9037 (0.0428)	0.9040 (0.0425)	0.9037 (0.0421)	0.9029 (0.0424)
		20	0.8420 (0.0636)	0.8423 (0.0639)	0.8431 (0.0637)	0.8425 (0.0640)	0.8428 (0.0635)
		40	0.6413 (0.1320)	0.6415 (0.1322)	0.6415 (0.1322)	0.6414 (0.1322)	0.6412 (0.1320)
	200	5	0.9501 (0.0204)	0.9506 (0.0196)	0.9507 (0.0195)	0.9508 (0.0193)	0.9509 (0.0191)
		10	0.9241 (0.0351)	0.9246 (0.0353)	0.9245 (0.0354)	0.9245 (0.0353)	0.9243 (0.0357)
		20	0.8864 (0.0319)	0.8867 (0.0322)	0.8867 (0.0321)	0.8868 (0.0319)	0.8864 (0.0323)
		40	0.8197 (0.0475)	0.8199 (0.0477)	0.8200 (0.0477)	0.8198 (0.0477)	0.8196 (0.0477)
VII	100	5	0.8114 (0.0530)	0.8117 (0.0528)	0.8114 (0.0529)	0.8112 (0.0530)	0.8115 (0.0527)
		10	0.7898 (0.0414)	0.7897 (0.0411)	0.7901 (0.0413)	0.7898 (0.0413)	0.7901 (0.0409)
		20	0.7774 (0.0420)	0.7782 (0.0417)	0.7782 (0.0417)	0.7782 (0.0419)	0.7781 (0.0417)
		40	0.7543 (0.0536)	0.7544 (0.0537)	0.7544 (0.0537)	0.7544 (0.0537)	0.7544 (0.0537)
	200	5	0.8168 (0.0352)	0.8167 (0.0353)	0.8167 (0.0352)	0.8166 (0.0351)	0.8167 (0.0352)
		10	0.7994 (0.0322)	0.7993 (0.0324)	0.7994 (0.0323)	0.7995 (0.0322)	0.7993 (0.0321)
		20	0.7907 (0.0345)	0.7908 (0.0344)	0.7908 (0.0345)	0.7907 (0.0344)	0.7907 (0.0344)
		40	0.7660 (0.0352)	0.7662 (0.0353)	0.7663 (0.0352)	0.7662 (0.0352)	0.7662 (0.0353)

tional quantiles, we limit the comparison to homoscedastic models and report results for $\tau = 0.5$.

Table 4.3 reports the means and standard deviations of the MCMR. Overall, all methods demonstrate competitive performance, with τ-fGCQS and f-GSIR yielding similar results. There are situations where f-GSAVE outperforms the other methods, and cases where τ-fGCQS and f-GSIR perform better. However, these differences are generally small, except for Model VIII, where τ-fGCQS and f-GSIR significantly outperform f-GSAVE.

It is very important to note that, while this comparison provides useful insights, the proposed methodology is explicitly designed for conditional quantiles, whereas f-GSIR and f-GSAVE are not.

Example 4 - Linear vs. Nonlinear dimension reduction. As highlighted in the Introduction, linear dimension reduction techniques may be insufficient for capturing complex nonlinear structures. Furthermore, these linear approaches often require more directions to adequately represent the information on the conditional quantiles. This example illustrates scenarios where the τ-FCQS method of Chapter 2 selects a

Table 4.3: *Mean (and standard deviation) of MCMR for τ-fGCQS, f-GSIR, and f-GSAVE for Models I-VIII.*

Model	τ-fGCQS	f-GSIR	f-GSAVE
I	0.8523 (0.0616)	0.8556 (0.0618)	0.8785 (0.1459)
II	0.7666 (0.0808)	0.7666 (0.0808)	0.9676 (0.0112)
III	0.9210 (0.0245)	0.9213 (0.0246)	0.8991 (0.1161)
IV	0.9719 (0.0082)	0.9719 (0.0082)	0.9696 (0.0163)
V	0.8879 (0.0206)	0.8888 (0.0205)	0.8566 (0.0327)
VI	0.9150 (0.0951)	0.9150 (0.0951)	0.9652 (0.0276)
VII	0.8054 (0.0329)	0.8056 (0.0330)	0.7711 (0.1130)
VIII	0.9036 (0.0233)	0.9036 (0.0233)	0.6927 (0.2438)

structural dimension q_τ greater than one, whereas τ-fGCQS successfully captures all relevant information within a single-index model.

Table 4.4 reports the means and standard deviations of the MCMR for Models V, VII, and VIII, comparing linear and nonlinear dimension reduction methods. This MCMR is intended to be close to the structural dimension of each method, and therefore, it is not directly comparable across methods when the dimensions differ. Notably, the structural dimension q_τ is 2 for Models V and VII, and 3 for Model VIII. This highlights a key advantage of nonlinear dimension reduction methods: the ability to capture all relevant information in fewer dimensions.

Table 4.4: *Mean (and standard deviation) of MCMR for τ-fGCQS and τ-fCQS for Models V, VII, and VIII.*

Model	Method	0.1	0.25	0.5	0.75	0.9
V	τ-fGCQS	0.8890 (0.0208)	0.8887 (0.0208)	0.8886 (0.0208)	0.8887 (0.0208)	0.8889 (0.0207)
	τ-fCQS	1.9711 (0.0308)	1.9721 (0.0246)	1.9702 (0.0351)	1.9618 (0.0975)	1.9713 (0.0381)
VII	τ-fGCQS	0.8103 (0.0391)	0.8101 (0.0392)	0.8102 (0.0391)	0.8104 (0.0388)	0.8102 (0.0381)
	τ-fCQS	1.8576 (0.2511)	1.8720 (0.2364)	1.8646 (0.2377)	1.8745 (0.2429)	1.8651 (0.2421)
VIII	τ-fGCQS	0.9037 (0.0275)	0.9037 (0.0274)	0.9037 (0.0275)	0.9037 (0.0275)	0.9037 (0.0275)
	τ-fCQS	2.7145 (0.3987)	2.7187 (0.4047)	2.7381 (0.3893)	2.7101 (0.4095)	2.7206 (0.3978)

4.7 Applications

4.7.1 Data Sets

We utilize several real-world datasets to apply our proposed methodology and compare its performance with existing methods. These datasets focus on health studies, examining various disorders such as ADHD and Parkinson's disease, along with key health metrics, such as BMI. Below is a brief description of each dataset considered.

ADHD-200 fMRI Data. We consider the same ADHD-200 resting-state fMRI (rs-fMRI) dataset introduced in Section 2.6 and used in Chapters 2 and 3. In contrast to the linear analyses in the earlier chapters, the present chapter revisits this dataset under the nonlinear framework developed here. The skewed distribution of the ADHD scores, with skewness equal to 0.3556, together with the presence of extreme observations, continues to support the use of quantile regression methods; see Figure 4.1. For comparison, we also relate the resulting findings to those reported in the earlier chapters.

Parkinson's Disease Telemonitoring Data. The Parkinson's Disease Telemonitoring Dataset, originally introduced by [73], consists of biomedical voice measurements collected from 42 individuals with early-stage Parkinson's disease who participated in a six-month trial of a telemonitoring device for remote symptom progression tracking. The recordings were automatically captured in the participants' homes and are publicly available at `https://archive.ics.uci.edu/dataset/189/parkinsons+telemonitoring`.

The dataset contains 5,875 voice recordings, with each row representing an individual recording. It includes subject information such as subject number, age, gender, and time interval from the baseline recruitment date, as well as motor and total Unified Parkinson's Disease Rating Scale (UPDRS) scores, which serve as key clinical measures of Parkinson's disease severity. Additionally, the dataset provides 16 biomedical voice features. The primary goal of the dataset is to predict motor and

total UPDRS scores from these voice measurements.

To ensure a consistent number of time points per patient, we refined the dataset by retaining only the first 100 voice recordings for each individual, resulting in a final dataset of 4200 recordings (100 per patient).

In this study, we apply the proposed methodology to investigate the relationship between voice measurements and motor UPDRS scores. When selecting the response variable, we considered both motor UPDRS and total UPDRS. Since motor UPDRS specifically measures motor symptoms of Parkinsonâs disease, such as tremors, rigidity, and bradykinesia, and given that the predictors are biomedical voice measurements - primarily associated with motor dysfunction (e.g., dysphonia, voice tremors, speech rate) - we chose motor UPDRS as the response variable. Consequently, we use the average motor UPDRS score for each patient as the response and the 16 voice features as the functional predictors.

NHANES 2011–2014 Accelerometry Data. The National Health and Nutrition Examination Survey (NHANES) is a large, ongoing cross-sectional study conducted by the Centers for Disease Control and Prevention (CDC), designed to assess the health and nutritional status of the non-institutionalized U.S. population. The study employs a multi-stage stratified sampling approach and collects extensive demographic, socioeconomic, lifestyle, and medical data, with variations in collected variables across different survey years.

A notable component of NHANES is the wrist-worn accelerometry data, collected during the 2011–2012 and 2013–2014 waves and publicly released in December 2020. Each participant recruited in these waves was asked to wear a wrist-worn physical activity monitor (ActiGraph GT3X+) for seven consecutive days. The raw 80 Hz tri-axial acceleration data were extracted from the device, processed, and released on the NHANES website in different units for the 2011–2012 and 2013–2014 cohorts. The accelerometry dataset is particularly valuable due to its linkage to the National

Death Index (NDI), its use of an open-source algorithm to process movement data into monitor-independent movement summary (MIMS) units, and its 24-hour continuous wear protocol, which captures physical activity and sleep patterns over multiple days.

The dataset initially included 14,693 participants who consented to wear an accelerometer. However, quality control measures led to the exclusion of 2,083 participants who had fewer than three “good days" of data (defined as at least 95% valid wear time). The final dataset comprises 12,610 participants, with an average age of 36.90 years, and a nearly equal gender distribution (51.23% female).

For this study, we utilize the single-level version of the NHANES accelerometry dataset, which was processed and organized into an analyzable format following a pipeline similar to that described in the `rnhanesdata` package. The data cleaning steps were adapted for the NHANES 2011-2014 cohorts. The single-level dataset compresses multiple days of minute-level accelerometry data by averaging the activity intensity at each minute across available days, resulting in 1,440 observations per participant. The data set is available at `https://functionaldataanalysis.org/` associated with the book by [74].

Some preprocessing steps were applied to the data. First, participants with missing BMI values were removed. Given the large sample size (12,610 participants), we further restricted the dataset to individuals aged 25 to 60 years. Finally, we limited the sample to participants who were followed for 72 months, resulting in a final dataset of 1,544 observations. As shown in Figure 4.1, BMI is right-skewed (with a skewness of 1.3746), highlighting the need for QR techniques.

This dataset enables the exploration of associations between physical activity patterns and key health outcomes. Our analysis focuses on modeling the relationship between the functional predictor MIMS and the response variable BMI.

This dataset enables the exploration of associations between physical activity patterns and key health outcomes. Our analysis focuses on modeling the relationship

between the functional predictor MIMS and the response variable BMI.

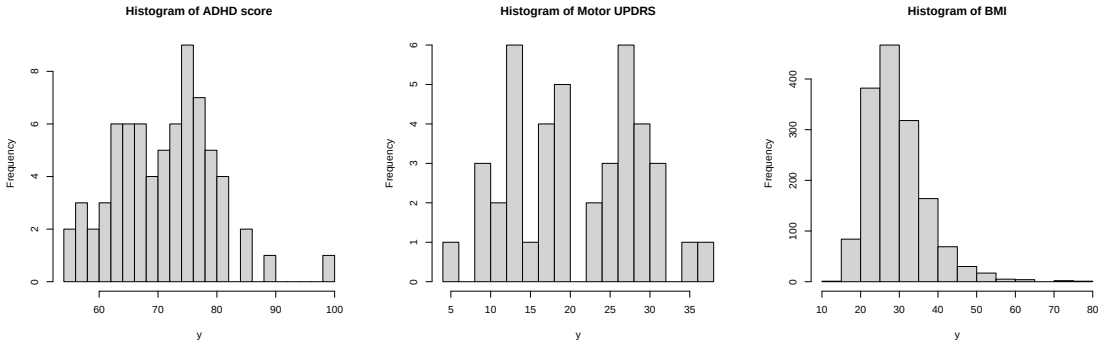

Figure 4.1: Histogram of the response variables for each data set.

4.7.2 Results

We split the data into training (80%) and test (20%) sets and use the training set to estimate the function $\widehat{\widetilde{f}}_\tau$, given in (4.9). To compare our method with f-GSIR and f-GSAVE of [54], we also extract the first nonlinear function using these two methods. We then fit the local linear QR model ([63]) on the test set for various quantiles ($\tau = 0.1, 0.25, 0.5, 0.75, 0.9$), using the extracted nonlinear functions as sufficient predictors to obtain an estimate of the conditional quantile function $Q_\tau\{Y|\widehat{\widetilde{f}}_\tau(X)\}$. We then compute the root mean squared error (RMSE) for each quantile level τ, defined as

$$RMSE(\tau) = \sqrt{\frac{1}{n_{test}} \sum_{u=1}^{n_{test}} \left(Y_u^{test} - \widehat{Q}_\tau\{Y|\widehat{\widetilde{f}}_\tau(X_u^{test})\} \right)^2}$$

where Y_u^{test} denotes the observed response in the test set, and $\widehat{Q}_\tau\{Y|\widehat{\widetilde{f}}_\tau(X_u^{test})\}$ is the estimated conditional quantile function from the local linear QR model. RMSE is calculated separately for each quantile level to assess the estimation accuracy, with smaller RMSE values indicating better performance.

In the ADHD dataset, Table 4.5 shows that our proposed method, τ-fGCQS, consistently achieves the lowest or tied-for-lowest RMSE across all quantiles, outperforming f-GSAVE and performing comparably with f-GSIR. These results demonstrate

the effectiveness of τ-fGCQS in reducing prediction error for ADHD severity scores. Notably, at higher quantiles of ADHD severity, the relationship between brain activity and ADHD scores becomes more complex and nonlinear. Individuals with more severe symptoms may exhibit atypical or irregular neurobiological patterns that linear methods like f-GSIR and f-GSAVE fail to capture effectively. This highlights the importance of modeling both nonlinear and quantile-specific relationships in neuroimaging studies to reflect the heterogeneity present in clinical populations.

For the Parkinson's disease dataset, τ-fGCQS achieves the lowest RMSE in four out of five quantiles, with f-GSIR outperforming it only at $\tau = 0.75$. In contrast, f-GSAVE consistently exhibits the highest RMSE across all quantiles, suggesting that it is less effective for capturing the nonlinear structure relevant to quantile regression. Notably, the performance gain of τ-fGCQS is especially prominent at the lower quantiles ($\tau = 0.1$ and 0.25), indicating that the nonlinear relationship between voice features and UPDRS scores is more detectable among patients with milder symptoms. Since vocal impairment develops gradually, flexible methods like τ-fGCQS are particularly valuable in identifying subtle patterns in the early stages of disease progression. These findings underscore the importance of modeling nonlinear and quantile-dependent relationships to understand the spectrum of Parkinson's disease symptoms.

The NHANES 2011-2014 dataset presents a more competitive scenario, with τ-fGCQS and f-GSIR producing comparable results across most quantiles. These results suggest that the relationship between physical activity and BMI may be moderately nonlinear, but not as strongly as in other datasets. Nonetheless, the ability of τ-fGCQS to achieve the lowest RMSE at $\tau = 0.1$ and $\tau = 0.25$ indicates that it can still capture subtle nonlinear deviations in the lower range of BMI. Its consistent performance across all quantiles highlights the robustness and flexibility of the proposed method, even when the underlying structure is not strongly nonlinear.

Table 4.5: *Root mean square error (RMSE) of local linear QR model using the first nonlinear sufficient predictor constructed by τ-fGCQS, f-GSIR and f-GSAVE. For each τ, the smallest value is bolded. In the case of ties, all relevant values are bolded.*

Dataset	Method	0.1	0.25	0.5	0.75	0.9
ADHD	τ-fGCQS	**8.3918**	**4.8254**	**4.9304**	**5.8281**	**6.5196**
	f-GSIR	**8.3918**	4.9994	**4.9304**	**5.8281**	6.7374
	f-GSAVE	9.6670	6.2290	5.0439	6.0731	8.4063
Parkinson	τ-fGCQS	**9.0026**	**7.0970**	**4.7880**	7.9622	**8.1822**
	f-GSIR	19.1994	10.6580	6.9821	**7.3021**	11.2321
	f-GSAVE	16.5418	9.3717	7.6600	8.7057	12.9980
NHANES	τ-fGCQS	**11.1124**	**8.9584**	7.3836	8.0711	12.1852
	f-GSIR	11.1366	8.9643	**7.3831**	**8.0340**	12.3182
	f-GSAVE	11.2203	8.9843	7.4383	8.1417	**12.0900**

CHAPTER 5: CONCLUSIONS

This dissertation studies dimension reduction for the conditional quantiles of functional data. The main goal is to move beyond traditional mean-based analysis and develop methods that provide a more complete understanding of the conditional distribution in the presence of functional predictors. Across the dissertation, this goal is pursued through a sequence of increasingly flexible frameworks, beginning with linear dimension reduction for functional predictors, extending to settings with both functional and categorical predictors, and culminating in a nonlinear dimension reduction framework.

Taken together, the results of this dissertation show that conditional quantile dimension reduction can be successfully developed in functional settings and can be adapted to increasingly complex data structures. The proposed methods provide new tools for studying functional data when the primary interest lies not only in the conditional mean, but also in other parts of the conditional distribution. In addition to their methodological contributions, the proposed approaches are supported by theoretical results, including convergence rates of the estimators, and by empirical studies demonstrating promising performance in simulations and real-data applications.

At the same time, several important directions remain for future research. First, the nonlinear framework developed in this dissertation is restricted to a single-index model. While this provides a useful starting point, such a structure may not be sufficiently flexible to capture more complex relationships involving multiple functional predictors or richer nonlinear dependence patterns. Extending the methodology to more general multi-index or fully nonlinear models would therefore be of substantial interest.

Second, the proposed methods rely on several user-specified parameters, and their practical performance may depend on these choices. This suggests the need for more data-driven procedures for parameter selection, as well as alternative estimation strategies that are less sensitive to tuning choices. Developing such methods would improve both the robustness and the accessibility of the proposed framework in practice.

Third, although the dissertation focuses on scalar responses and functional predictors, many modern applications involve more complicated data structures, such as multiple functional predictors observed on irregular grids, mixed responses, or other structured covariates. Extending conditional quantile dimension reduction to these settings would broaden the applicability of the methodology and raise new theoretical and computational questions.

Overall, this dissertation contributes new methodology and theory for conditional quantile analysis of functional data and highlights several promising avenues for future work. It is hoped that these results will provide a useful foundation for continued research on flexible and interpretable dimension reduction methods for complex functional data.

REFERENCES

[1] H.-G. Müller, R. Sen, and U. Stadmüller, "Functional data analysis for volatility," *Journal of Econometrics*, vol. 165, no. 2, pp. 233–245, 2011.

[2] M. Gubian, F. Torreira, and L. Boves, "Using functional data analysis for investigating multidimensional dynamic phonetic contrasts," *Journal of Phonetics*, vol. 49, pp. 16–40, 2015.

[3] A. Elías, R. Jiménez, and H. L. Shang, "On projection methods for functional time series forecasting," *Journal of Multivariate Analysis*, vol. 189, p. 104890, 2022.

[4] M. Fontana, M. Tavoni, and S. Vantini, "Functional data analysis of high-frequency load curves reveals drivers of residential electricity consumption," *PLOS ONE*, vol. 14, no. 6, 2019.

[5] K. Li and S. Luo, "Functional joint model for longitudinal and time-to-event data: An application to alzheimer's disease," *Statistics in Medicine*, vol. 36, no. 22, pp. 3560–3572, 2017.

[6] J. Pratt, W. Su, D. J. Hayes, J. P. Clancy, and R. D. Szczesniak, "An animated functional data analysis interface to cluster rapid lung function decline and enhance center-level care in cystic fibrosis," *Journal of Healthcare Engineering*, 2021.

[7] J. O. Ramsay, "When the data are functions," *Psychometrika*, vol. 47, no. 4, pp. 379–396, 1982.

[8] J. O. Ramsay and C. Dalzell, "Some tools for functional data analysis," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 53, no. 3, pp. 539–561, 1991.

[9] H. Cardot, F. Ferraty, and P. Sarda, "Functional linear model," *Statistics & Probability Letters*, vol. 45, no. 1, pp. 11–22, 1999.

[10] H. Cardot, F. Ferraty, and P. Sarda, "Spline estimators for the functional linear model," *Statistica Sinica*, pp. 571–591, 2003.

[11] F. Yao, H.-G. Müller, and J.-L. Wang, "Functional linear regression analysis for longitudinal data," 2005.

[12] T. T. Cai and P. Hall, "Prediction in functional linear regression," 2006.

[13] P. Hall and J. L. Horowitz, "Methodology and convergence rates for functional linear regression," 2007.

[14] C. Crambes, A. Kneip, and P. Sarda, "Smoothing splines estimators for functional linear regression," 2009.

[15] H. Shin, "Partial functional linear regression," *Journal of Statistical Planning and Inference*, vol. 139, no. 10, pp. 3405–3418, 2009.

[16] G. M. James, J. Wang, and J. Zhu, "Functional linear regression thatâs interpretable," 2009.

[17] M. Yuan and T. T. Cai, "A reproducing kernel hilbert space approach to functional linear regression," 2010.

[18] P. C. Besse, H. Cardot, and D. B. Stephenson, "Autoregressive forecasting of some functional climatic variations," *Scandinavian Journal of Statistics*, vol. 27, no. 4, pp. 673–687, 2000.

[19] F. Ferraty and P. Vieu, "The functional nonparametric model and application to spectrometric data," *Computational Statistics*, vol. 17, no. 4, pp. 545–564, 2002.

[20] F. Ferraty and P. Vieu, "Curves discrimination: a nonparametric functional approach," *Computational Statistics & Data Analysis*, vol. 44, no. 1-2, pp. 161–173, 2003.

[21] F. Ferraty and P. Vieu, "Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination," *Nonparametric Statistics*, vol. 16, no. 1-2, pp. 111–125, 2004.

[22] G. A. Pérez, H. Cardot, G. E. Pérez, and P. Vieu, "Maximum ozone forecasting by functional nonparametric approach," *Environmetrics*, vol. 15, pp. 675–685, 2004.

[23] J.-M. Chiou, Y.-F. Yang, and Y.-T. Chen, "Multivariate functional linear regression and prediction," *Journal of Multivariate Analysis*, vol. 146, pp. 301–312, 2016.

[24] C. Happ and S. Greven, "Multivariate functional principal component analysis for data observed on different (dimensional) domains," *Journal of the American Statistical Association*, vol. 113, no. 522, pp. 649–659, 2018.

[25] B. D. Marx and P. H. Eilers, "Generalized linear regression on sampled signals and curves: a p-spline approach," *Technometrics*, vol. 41, no. 1, pp. 1–13, 1999.

[26] G. M. James, "Generalized linear models with functional predictors," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 64, no. 3, pp. 411–432, 2002.

[27] H.-G. Müller and U. Stadtmüller, "Generalized functional linear models," 2005.

[28] G. M. James and B. W. Silverman, "Functional adaptive model estimation," *Journal of the American Statistical Association*, vol. 100, no. 470, pp. 565–576, 2005.

[29] L. Ferré and A.-F. Yao, "Functional sliced inverse regression analysis," *Statistics*, vol. 37, no. 6, pp. 475–488, 2003.

[30] K.-C. Li, "Sliced inverse regression for dimension reduction," *Journal of the American Statistical Association*, vol. 86, no. 414, pp. 316–327, 1991.

[31] L. Ferré and A.-F. Yao, "Smoothed functional inverse regression," *Statistica Sinica*, pp. 665–683, 2005.

[32] H. Lian and G. Li, "Series expansion for functional sufficient dimension reduction," *Journal of Multivariate Analysis*, vol. 124, pp. 150–165, 2014.

[33] R. D. Cook and S. Weisberg, "Sliced inverse regression for dimension reduction: Comment," *Journal of the American Statistical Association*, vol. 86, no. 414, pp. 328–332, 1991.

[34] G. Wang, N. Lin, and B. Zhang, "Functional contour regression," *Journal of Multivariate Analysis*, vol. 116, pp. 1–13, 2013.

[35] B. Li, H. Zha, and F. Chiaromonte, "Contour regression: a general approach to dimension reduction," 2005.

[36] B. Li and J. Song, "Dimension reduction for functional data based on weak conditional moments," *The Annals of Statistics*, vol. 50, no. 1, pp. 107–128, 2022.

[37] E. Solea, E. Christou, and J. Song, "Robust inverse regression for multivariate elliptical functional data," *Statistica Sinica*, 2024.

[38] H. Cardot, C. Crambes, and P. Sarda, "Quantile regression when the covariates are functions," *Nonparametric Statistics*, vol. 17, no. 7, pp. 841–856, 2005.

[39] K. Kato, "Estimation in functional linear quantile regression," 2012.

[40] F. Ferraty, A. Rabhi, and P. Vieu, "Conditional quantiles for dependent functional data with application to the climatic" el niño" phenomenon," *Sankhyā: The Indian Journal of Statistics*, pp. 378–398, 2005.

[41] K. Chen and H.-G. Müller, "Conditional quantile analysis when covariates are functions, with application to growth data," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 74, no. 1, pp. 67–89, 2012.

[42] Y. Lu, J. Du, and Z. Sun, "Functional partially linear quantile regression model," *Metrika*, vol. 77, no. 2, pp. 317–332, 2014.

[43] F. Yao, S. Sue-Chee, and F. Wang, "Regularized partially functional quantile regression," *Journal of Multivariate Analysis*, vol. 156, pp. 39–56, 2017.

[44] H. Ma, T. Li, H. Zhu, and Z. Zhu, "Quantile regression for functional partially linear model in ultra-high dimensions," *Computational Statistics & Data Analysis*, vol. 129, pp. 135–147, 2019.

[45] G. Shi, T. Xie, and Z. Zhang, "Statistical inference for the functional quadratic quantile regression model," *Metrika*, vol. 83, no. 8, pp. 937–960, 2020.

[46] T. Qingguo and L. Kong, "Quantile regression in functional linear semiparametric model," *Statistics*, vol. 51, no. 6, pp. 1342–1358, 2017.

[47] M. Guo, L. Zhou, J. Z. Huang, and W. K. Härdle, "Functional data analysis of generalized regression quantiles," *Statistics and Computing*, vol. 25, no. 2, pp. 189–202, 2015.

[48] D. Bosq, *Linear Processes in Function Spaces: Theory and Applications*, vol. 149 of *Lecture Notes in Statistics*. New York: Springer Science & Business Media, 2000.

[49] T. Hsing and R. L. Eubank, *Theoretical foundations of functional data analysis, with an introduction to linear operators*, vol. 997. Wiley Online Library, 2015.

[50] R. D. Cook, *Regression graphics: Ideas for studying regressions through graphics*. John Wiley & Sons, 2009.

[51] K.-Y. Lee and L. Li, "Functional sufficient dimension reduction through average fréchet derivatives," *Annals of statistics*, vol. 50, no. 2, p. 904, 2022.

[52] B. Li, A. Artemiou, and L. Li, "Principal support vector machines for linear and nonlinear sufficient dimension reduction," 2011.

[53] K.-Y. Lee, B. Li, and F. Chiaromonte, "A general theory for nonlinear sufficient dimension reduction: Formulation and estimation," *The Annals of Statistics*, pp. 221–249, 2013.

[54] B. Li and J. Song, "Nonlinear sufficient dimension reduction for functional data," 2017.

[55] R. Koenker and G. Bassett Jr, "Regression quantiles," *Econometrica: journal of the Econometric Society*, pp. 33–50, 1978.

[56] W. Luo, B. Li, and X. Yin, "On efficient dimension reduction with respect to a statistical functional of interest," 2014.

[57] E. Kong and Y. Xia, "An adaptive composite quantile approach to dimension reduction," 2014.

[58] E. Christou, "Central quantile subspace," *Statistics and Computing*, vol. 30, no. 3, pp. 677–695, 2020.

[59] C. E. Lee and H. Hilafu, "Quantile martingale difference divergence for dimension reduction," *Statistica Sinica*, vol. 32, no. 1, pp. 65–87, 2022.

[60] E. Christou, A. Settle, and A. Artemiou, "Nonlinear dimension reduction for conditional quantiles," *Advances in Data Analysis and Classification*, vol. 15, pp. 937–956, 2021.

[61] R. D. Cook and B. Li, "Dimension reduction for conditional mean in regression," *The Annals of Statistics*, vol. 30, no. 2, pp. 455–474, 2002.

[62] E. Christou, E. Solea, S. Wang, and J. Song, "Sufficient dimension reduction for the conditional quantiles of functional data," *Statistica Sinica*, 2025.

[63] E. Guerre and C. Sabbah, "Uniform bias study and bahadurrepresentation for local polynomial estimators ofthe conditional quantile function," *Econometric Theory*, vol. 28, no. 1, pp. 87–129, 2012.

[64] K. Yu and M. Jones, "Local linear quantile regression," *Journal of the American statistical Association*, vol. 93, no. 441, pp. 228–237, 1998.

[65] G. Wang, S. Liu, F. Han, and C.-Z. Di, "Robust functional principal component analysis via a functional pairwise spatial sign operator," *Biometrics*, vol. 79, no. 2, pp. 1239–1253, 2023.

[66] F. Chiaromonte, R. D. Cook, and B. Li, "Sufficient dimension reduction in regressions with categorical predictors," *Annals of Statistics*, pp. 475–497, 2002.

[67] E. Christou, "Sufficient dimension reduction for conditional quantiles with alternative types of data," *Journal of Statistical Computation and Simulation*, vol. 92, no. 2, pp. 300–317, 2022.

[68] R. D. Cook and F. Critchley, "Identifying regression outliers and mixtures graphically," *Journal of the American Statistical Association*, vol. 95, no. 451, pp. 781–794, 2000.

[69] G. Wang, "Dimension reduction in functional regression with categorical predictor," *Computational Statistics*, vol. 32, no. 2, pp. 585–609, 2017.

[70] B. Li and K. Kim, "On sufficient graphical models," *Journal of Machine Learning Research*, vol. 25, no. 17, pp. 1–64, 2024.

[71] S. Bhattacharjee, B. Li, and L. Xue, "Nonlinear global fréchet regression for random objects via weak conditional expectation," *Annals of statistics*, vol. 53, no. 1, p. 117, 2025.

[72] B. Li, "Linear operator-based statistical analysis: A useful paradigm for big data," *Canadian Journal of Statistics*, vol. 46, no. 1, pp. 79–103, 2018.

[73] A. Tsanas, M. Little, P. McSharry, and L. Ramig, "Accurate telemonitoring of parkinsonâs disease progression by non-invasive speech tests," *Nature Precedings*, pp. 1–1, 2009.

[74] C. M. Crainiceanu, J. Goldsmith, A. Leroux, and E. Cui, *Functional data analysis with R*. Chapman and Hall/CRC, 2024.

[75] B. Li and E. Solea, "A nonparametric graphical model for functional data with application to brain networks based on fmri," *Journal of the American Statistical Association*, vol. 113, no. 524, pp. 1637–1655, 2018.

[76] K. Fukumizu, F. R. Bach, and A. Gretton, "Statistical consistency of kernel canonical correlation analysis.," *Journal of Machine Learning Research*, vol. 8, no. 2, 2007.

[77] Q. Wang and X. Yin, "Sufficient dimension reduction and variable selection for regression mean function with two types of predictors," *Statistics & probability letters*, vol. 78, no. 16, pp. 2798–2803, 2008.

[78] R. R. Bahadur, "Sufficiency and statistical decision functions," *The annals of mathematical Statistics*, pp. 423–462, 1954.

[79] B. Li, *Sufficient dimension reduction: Methods and applications with R*. Chapman and Hall/CRC, 2018.

APPENDIX A: TECHNICAL DETAILS AND PROOFS FOR CHAPTER 2

The Appendix A includes the assumptions (Section A.1), some preliminary results and notation (Section A.2), useful lemmas needed for the proof of the main results (Section A.3), the proofs of Theorems 2.1, 2.2, 2.3, 4.3, and **??** (Section A.4), and additional simulation results (Section **??**).

A.1 Assumptions

S1. $E[Q_\tau^2\{Y|L(X)\}] < \infty$ for a given $\tau \in (0,1)$, where $L(\cdot)$ is defined in (1.5).

S2. $E\left\|X\right\|_{\mathscr{H}}^4 < \infty$.

S3. The nonnegative kernel function $K(\cdot)$, used in (2.6), is Lipschitz over $\mathbb{R}^d$, has a compact support, and satisfies $\int K(z)dz = 1$. For some $\underline{K} > 0$, $K(z) \geq \underline{K} I(z \in \mathcal{S}(0,1))$, where $\mathcal{S}(0,1)$ is the closed unit ball. The associated bandwidth h is in $[\underline{h}_n, \overline{h}_n]$ with $0 < \underline{h}_n \leq \overline{h}_n < \infty$, $lim_{n\to\infty}\overline{h}_n = 0$, and $\lim_{n\to\infty}(logn)/(n\overline{h}_n^d) = 0$.

S4. The distribution of $L(X)$ has a probability density function $f_L(\cdot)$ with respect to the Lebesgue measure, which is strictly positive and continuously differentiable over the compact support $\mathcal{L}(X)$ of $L(X)$.

S5. The cumulative distribution function $F_L(\cdot|\cdot)$ of Y given $L(X)$ has a continuous probability density function $f_L\{Y|L(X)\}$ with respect to the Lebesgue measure, which is strictly positive for Y in $\mathbb{R}$ and $L(X)$ in $\mathcal{L}(X)$. The partial derivative $\partial F_L\{Y|L(X)\}/\partial L(X)$ is continuous over $\mathbb{R} \times \mathcal{L}(X)$. There is an $N_0 > 0$ such that

$$|f_L\{Y|L(X)\} - f_L\{Y'|L'(X)\}| \leq N_0 \left\|(L(X), Y) - (L'(X), Y')\right\|$$

for all $(L(X), Y), (L'(X), Y)$ of $\mathcal{L}(X) \times \mathbb{R}$.

S6. Assume that $Q_\tau\{Y|L(X)\}$ is in $\mathcal{C}(N^*, s)$ for some $[s] \leq 1$, where $[s]$ denotes the lowest integer part of s. We say that $Q_\tau\{Y|L(X)\}$ is in $\mathcal{C}(N^*, s)$, $N^*, s > 0$, if

(a) for all $\tau \in [\underline{\tau}, \overline{\tau}]$, a subset of $(0, 1)$, $Q_\tau\{Y|L(X)\}$ is $[s]$th continuous differentiable over the support $\mathcal{L}(X)$;

(b) there exists a constant $N^* > 0$, such that for all $u = (u_1, \ldots, u_d)^\top$ with $|u| = u_1 + \cdots + u_d = [s]$, all τ in the interval $[\underline{\tau}, \overline{\tau}]$ and all $L(X), L'(X)$ in $\mathcal{L}(X)$,

$$|D^u Q_\tau\{Y|L(X)\} - D^u Q_\tau\{Y|L'(X)\}| \leq N^* \|L(X) - L'(X)\|^{s-[s]},$$

where $D^u Q_\tau\{Y|L(X)\}$ denotes the partial derivative $\partial^{|u|} Q_\tau\{Y|L(X)\}/\partial L_1^{u_1}(X) \ldots L_d^{u_d}(X)$ and $\|\cdot\|$ denotes the Euclidean norm.

Assumption S2 is necessary for the $\sqrt{n}$-consistency of $\widehat{\Sigma}_{XX}$. Assumptions S3-S6 come from the work of [63] and are needed for the consistency of $\widehat{Q}_\tau\{Y|\widehat{L}(X)\}$.

A.2 Preliminaries

For a bounded linear operator A, let $\|A\|$ denote its operator norm. For a Hilbert-Schmidt operator A, let $\|A\|_{HS}$ denote its Hilbert-Schmidt norm of A.

Let a_n be a sequence of positive numbers that converges to 0. If A is a bounded linear operator and $\widehat{A}_n$ is a sequence of bounded linear random operators such that $\left\|\widehat{A}_n - A\right\| = O_p(a_n)$, then we write $\widehat{A}_n = A + O_p(a_n)$. Similarly, if f is a function in a Hilbert space $\mathcal{H}$ and $\widehat{f}_n$ is a sequence of random functions in $\mathcal{H}$ such that $\left\|\widehat{f}_n - f\right\|_{\mathcal{H}} = O_p(a_n)$, then we write $\widehat{f}_n = f + O_p(a_n)$.

For two positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \prec b_n$ if $a_n/b_n \to 0$. For example, $a_n \prec 1$ means $a_n \to 0$. We write $a_n \succ b_n$ if $a_n/b_n \to \infty$. For example, $n^{-1/2} \prec a_n \prec 1$ means a_n goes to 0 at a slower rate than $n^{-1/2}$.

Useful facts that will be used throughout the proofs are:

- ([75], Lemma 6): Suppose A is a bounded and self-adjoint operator, $\widehat{A}_n$ is a sequence of bounded and self-adjoint random linear operators such that $\widehat{A}_n = O_p(1)$, ϵ_n is a positive sequence that converges to 0, and a, b are positive constants, then

$$\begin{aligned} \left\|(A+\epsilon_n I)^{-b} A^a\right\| &= O(\epsilon_n^{min\{0,a-b\}}) \\ \left\|(\widehat{A}_n+\epsilon_n I)^{-b} \widehat{A}_n^a\right\| &= O_p(\epsilon_n^{min\{0,a-b\}}). \end{aligned} \tag{A.1}$$

- ([49], Theorem 8.1.2): If Assumption S2 holds, then Σ_{XX} is a Hilbert Schmidt operator and

$$\left\|\widehat{\Sigma}_{XX} - \Sigma_{XX}\right\|_{HS} = O_p(n^{-1/2}). \tag{A.2}$$

- ([76], Section 5.2): If A and B are self adjoint and invertible linear operators, then

$$A^{-1/2} - B^{-1/2} = \{A^{-1/2}(B^{3/2} - A^{3/2}) + A - B\}B^{-3/2}. \tag{A.3}$$

- ([76], Lemma 8): Suppose A and B are self-adjoint operators on a Hilbert space such that $0 \leq A \leq cI$ and $0 \leq B \leq cI$ hold for a positive constant c. Then,

$$\left\|A^{3/2} - B^{3/2}\right\| \leq 3c^{1/2} \left\|A - B\right\|. \tag{A.4}$$

A.3 Useful Lemmas

Lemma A.1. *Under Assumptions S3-S6 and the assumption that $\widehat{L}(X)$ is consistent of the directions of the FCS then,*

$$\sup_{x\in\mathscr{H}} |\widehat{Q}_\tau\{Y|\widehat{L}(x)\} - Q_\tau\{Y|L(x)\}| = o_p(1),$$

where $\widehat{Q}_\tau\{Y|\widehat{L}(x)\}$ is as defined in (2.6).

Proof. Observe that

$$\begin{aligned}\sup_{x\in\mathscr{H}} |\widehat{Q}_\tau\{Y|\widehat{L}(x)\} - Q_\tau\{Y|L(x)\}| &\leq \sup_{x\in\mathscr{H}} |\widehat{Q}_\tau\{Y|\widehat{L}(x)\} - \widehat{Q}_\tau\{Y|L(x)\}| \\ &\quad + \sup_{x\in\mathscr{H}} |\widehat{Q}_\tau\{Y|L(x)\} - Q_\tau\{Y|L(x)\}| \\ &= o_p(1),\end{aligned}$$

where the first term follows from the Bahadur representation of $\widehat{Q}_\tau\{Y|\widehat{L}(x)\}-\widehat{Q}_\tau\{Y|L(x)\}$ (see [63]) and the consistency of $\widehat{L}(x)$, and the second term follows from Corollary 1 (ii) of [63]. □

Lemma A.2. *Under Assumptions 2.1, 2.2, 3.2, and the assumption that $E(X) = 0$, we can prove that*

(a) M_τ is bounded, and

(b) $B_\tau = \Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2} B_\tau \Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2}$,

where $M_\tau = \Sigma_{XX}^{\dagger 1/2} B_\tau \Sigma_{XX}^{\dagger 1/2}$ and $B_\tau = \sum_{j=0}^{m} b_{\tau,j}\otimes b_{\tau,j}$, for $b_{\tau,0} = \beta_\tau^*$ given in (4.5) and $b_{\tau,j} = E(\langle b_{\tau,j-1}, X\rangle_{\mathscr{H}} X)$, for $j = 1,\ldots,m$.

Proof. To prove the two results, we will first derive an alternative expression for B_τ.

By the definition of the tensor product,

$$\begin{aligned} b_{\tau,j} &= E(\langle b_{\tau,j-1}, X\rangle_{\mathscr{H}} X) = E(X\langle b_{\tau,j-1}, X\rangle_{\mathscr{H}}) = E(X\langle X, b_{\tau,j-1}\rangle_{\mathscr{H}}) \\ &= E\{(X\otimes X)b_{\tau,j-1}\} = E(X\otimes X)b_{\tau,j-1} = \Sigma_{XX}b_{\tau,j-1}. \end{aligned}$$

Using the above, we can express $b_{\tau,j}$, $j = 1,\ldots,m$, as

$$b_{\tau,1} = \Sigma_{XX}b_{\tau,0}, \quad b_{\tau,2} = \Sigma_{XX}^2 b_{\tau,0}, \ldots, \quad b_{\tau,m} = \Sigma_{XX}^m b_{\tau,0}. \tag{A.5}$$

Moreover, since $b_{\tau,0} \in \mathscr{S}_{Q_\tau(Y|X)}$ and, under Assumption 3.2, $\mathscr{S}_{Q_\tau(Y|X)} \subseteq \overline{\text{ran}}(\Sigma_{XX})$, then $b_{\tau,0}$ can be expressed as $b_{\tau,0} = \Sigma_{XX}\omega_\tau$, for $\omega_\tau \in \mathscr{H}$. Using this representation of $b_{\tau,0}$, we can rewrite $b_{\tau,j}$ in expression (A.5) as

$$b_{\tau,1} = \Sigma_{XX}^2\omega_\tau, \quad b_{\tau,2} = \Sigma_{XX}^3\omega_\tau, \ldots, \quad b_{\tau,m} = \Sigma_{XX}^{m+1}\omega_\tau.$$

Finally, B_τ can be expressed as

$$\begin{aligned} B_\tau &= \sum_{j=0}^{m} b_{\tau,j}\otimes b_{\tau,j} \\ &= \Sigma_{XX}\omega_\tau\otimes\Sigma_{XX}\omega_\tau + \Sigma_{XX}^2\omega_\tau\otimes\Sigma_{XX}^2\omega_\tau + \cdots + \\ &\quad \Sigma_{XX}^{m+1}\omega_\tau\otimes\Sigma_{XX}^{m+1}\omega_\tau. \end{aligned} \tag{A.6}$$

An alternative formulation is given by

$$\begin{aligned} B_\tau &= \Sigma_{XX}(\omega_\tau\otimes\omega_\tau)\Sigma_{XX} + \Sigma_{XX}^2(\omega_\tau\otimes\omega_\tau)\Sigma_{XX}^2 + \cdots + \\ &\quad \Sigma_{XX}^{m+1}(\omega_\tau\otimes\omega_\tau)\Sigma_{XX}^{m+1}. \end{aligned} \tag{A.7}$$

This is easy to see because, in general, for every $h \in \mathscr{H}$,

$$\begin{aligned}(\Sigma_{XX}\omega_\tau \otimes \Sigma_{XX}\omega_\tau)h &= \Sigma_{XX}\omega_\tau\langle\Sigma_{XX}\omega_\tau, h\rangle_{\mathscr{H}} = \Sigma_{XX}\omega_\tau\langle\omega_\tau, \Sigma_{XX}h\rangle_{\mathscr{H}} \\ &= \Sigma_{XX}(\omega_\tau \otimes \omega_\tau)\Sigma_{XX}h.\end{aligned}$$

We are now ready to prove parts (a) and (b) of the lemma. For part (a), we will show that for every $h \in \mathscr{H}$, $M_\tau h$ is bounded, where $M_\tau = \Sigma_{XX}^{\dagger 1/2} B_\tau \Sigma_{XX}^{\dagger 1/2}$. Observe that,

$$\begin{aligned}M_\tau h &= \Sigma_{XX}^{\dagger 1/2} B_\tau \Sigma_{XX}^{\dagger 1/2} h \\ &= \Sigma_{XX}^{\dagger 1/2}(\Sigma_{XX}\omega_\tau \otimes \Sigma_{XX}\omega_\tau + \Sigma_{XX}^2\omega_\tau \otimes \Sigma_{XX}^2\omega_\tau + \cdots + \Sigma_{XX}^{m+1}\omega_\tau \otimes \Sigma_{XX}^{m+1}\omega_\tau)\Sigma_{XX}^{\dagger 1/2}h \\ &= \Sigma_{XX}^{\dagger 1/2}(\Sigma_{XX}\omega_\tau \otimes \Sigma_{XX}\omega_\tau)\Sigma_{XX}^{\dagger 1/2}h + \Sigma_{XX}^{\dagger 1/2}(\Sigma_{XX}^2\omega_\tau \otimes \Sigma_{XX}^2\omega_\tau)\Sigma_{XX}^{\dagger 1/2}h + \cdots + \\ &\quad \Sigma_{XX}^{\dagger 1/2}(\Sigma_{XX}^{m+1}\omega_\tau \otimes \Sigma_{XX}^{m+1}\omega_\tau)\Sigma_{XX}^{\dagger 1/2}h \\ &= \Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}\omega_\tau\langle\Sigma_{XX}\omega_\tau, \Sigma_{XX}^{\dagger 1/2}h\rangle_{\mathscr{H}} + \Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^2\omega_\tau\langle\Sigma_{XX}^2\omega_\tau, \Sigma_{XX}^{\dagger 1/2}h\rangle_{\mathscr{H}} + \cdots + \\ &\quad \Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{m+1}\omega_\tau\langle\Sigma_{XX}^{m+1}\omega_\tau, \Sigma_{XX}^{\dagger 1/2}h\rangle_{\mathscr{H}} \\ &= \Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}\omega_\tau\langle\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}\omega_\tau, h\rangle_{\mathscr{H}} + \Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^2\omega_\tau\langle\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^2\omega_\tau, h\rangle_{\mathscr{H}} + \cdots + \\ &\quad \Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{m+1}\omega_\tau\langle\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{m+1}\omega_\tau, h\rangle_{\mathscr{H}},\end{aligned}$$

where the second equality follows from replacing B_τ with expression (A.6).

Then,

$$\begin{aligned}
\|M_\tau h\| &\leq \left\|\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}\right\| \|\omega_\tau\|_{\mathscr{H}} |\langle \Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}\omega_\tau, h\rangle_{\mathscr{H}}| + \\
&\quad \left\|\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{2}\right\| \|\omega_\tau\|_{\mathscr{H}} |\langle \Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{2}\omega_\tau, h\rangle_{\mathscr{H}}| + \cdots + \\
&\quad \left\|\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{m+1}\right\| \|\omega_\tau\|_{\mathscr{H}} |\langle \Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{m+1}\omega_\tau, h\rangle_{\mathscr{H}}| \\
&\leq \left\|\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}\right\| \|\omega_\tau\|_{\mathscr{H}} \left\|\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}\right\| \|\omega_\tau\|_{\mathscr{H}} \|h\|_{\mathscr{H}} + \\
&\quad \left\|\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{2}\right\| \|\omega_\tau\|_{\mathscr{H}} \left\|\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{2}\right\| \|\omega_\tau\|_{\mathscr{H}} \|h\|_{\mathscr{H}} + \cdots + \\
&\quad \left\|\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{m+1}\right\| \|\omega_\tau\|_{\mathscr{H}} \left\|\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{m+1}\right\| \|\omega_\tau\|_{\mathscr{H}} \|h\|_{\mathscr{H}},
\end{aligned}$$

where the second inequality follows from the Cauchy–Schwarz inequality. It is easy to see that all of the above terms are bounded and therefore, M_τ is bounded.

For part (b), we need to show that $B_\tau h = \Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2} B_\tau \Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2} h$, for every $h \in \mathscr{H}$. Observe that,

$$\begin{aligned}
&\Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2} B_\tau \Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2} h \\
= \;& \Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}(\Sigma_{XX}\omega_\tau \otimes \Sigma_{XX}\omega_\tau + \Sigma_{XX}^2\omega_\tau \otimes \Sigma_{XX}^2\omega_\tau + \cdots + \Sigma_{XX}^{m+1}\omega_\tau \otimes \Sigma_{XX}^{m+1}\omega_\tau)\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2} h \\
= \;& \Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}(\Sigma_{XX}\omega_\tau \otimes \Sigma_{XX}\omega_\tau)\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2} h + \Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}(\Sigma_{XX}^2\omega_\tau \otimes \Sigma_{XX}^2\omega_\tau)\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2} h + \cdots + \\
&\Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}(\Sigma_{XX}^{m+1}\omega_\tau \otimes \Sigma_{XX}^{m+1}\omega_\tau)\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2} h \\
= \;& \Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}\omega_\tau \langle \Sigma_{XX}\omega_\tau, \Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2} h\rangle_{\mathscr{H}} + \Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^2\omega_\tau \langle \Sigma_{XX}^2\omega_\tau, \Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2} h\rangle_{\mathscr{H}} + \cdots + \\
&\Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{m+1}\omega_\tau \langle \Sigma_{XX}^{m+1}\omega_\tau, \Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2} h\rangle_{\mathscr{H}} \\
= \;& \Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}\omega_\tau \langle \Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}\omega_\tau, h\rangle_{\mathscr{H}} + \Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^2\omega_\tau \langle \Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^2\omega_\tau, h\rangle_{\mathscr{H}} + \cdots + \\
&\Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{m+1}\omega_\tau \langle \Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{m+1}\omega_\tau, h\rangle_{\mathscr{H}}
\end{aligned}$$

$$
\begin{aligned}
&= \Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2}\Sigma_{XX}^{1/2}\omega_\tau\langle\Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2}\Sigma_{XX}^{1/2}\omega_\tau, h\rangle_{\mathscr{H}} \\
&\quad +\Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2}\Sigma_{XX}^{3/2}\omega_\tau\langle\Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2}\Sigma_{XX}^{3/2}\omega_\tau, h\rangle_{\mathscr{H}} + \cdots + \\
&\quad \Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2}\Sigma_{XX}^{m+1/2}\omega_\tau\langle\Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2}\Sigma_{XX}^{m+1/2}\omega_\tau, h\rangle_{\mathscr{H}} \\
&= \Sigma_{XX}^{1/2}\Sigma_{XX}^{1/2}\omega_\tau\langle\Sigma_{XX}^{1/2}\Sigma_{XX}^{1/2}\omega_\tau, h\rangle_{\mathscr{H}} + \Sigma_{XX}^{1/2}\Sigma_{XX}^{3/2}\omega_\tau\langle\Sigma_{XX}^{1/2}\Sigma_{XX}^{3/2}\omega_\tau, h\rangle_{\mathscr{H}} + \cdots + \\
&\quad \Sigma_{XX}^{1/2}\Sigma_{XX}^{m+1/2}\omega_\tau\langle\Sigma_{XX}^{1/2}\Sigma_{XX}^{m+1/2}\omega_\tau, h\rangle_{\mathscr{H}} \\
&= \Sigma_{XX}\omega_\tau\langle\Sigma_{XX}\omega_\tau, h\rangle_{\mathscr{H}} + \Sigma_{XX}^2\omega_\tau\langle\Sigma_{XX}^2\omega_\tau, h\rangle_{\mathscr{H}} + \cdots + \Sigma_{XX}^{m+1}\omega_\tau\langle\Sigma_{XX}^{m+1}\omega_\tau, h\rangle_{\mathscr{H}} \\
&= (\Sigma_{XX}\omega_\tau \otimes \Sigma_{XX}\omega_\tau)h + (\Sigma_{XX}^2\omega_\tau \otimes \Sigma_{XX}^2\omega_\tau)h + \cdots + (\Sigma_{XX}^{m+1}\omega_\tau \otimes \Sigma_{XX}^{m+1}\omega_\tau)h \\
&= B_\tau h,
\end{aligned}
$$

where the first equality follows from replacing B_τ with expression (A.6) and the sixth equality follows from the fact that $\Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2} = \Sigma_{XX}^{1/2}$; see Theorem 3.5.8 of [49]. □

Lemma A.3. *Under Assumptions 2.1, 2.2, 3.2, the assumption that $E(X) = 0$, the assumption that $\widehat{L}(X)$ is consistent of the directions of the FCS, and assumptions S1-S6, we have for $j = 0, \ldots, m$ that*

$$\left\|\widehat{\beta}_{\tau,j} - b_{\tau,j}\right\|_{\mathscr{H}} = O_p(n^{-1/2}),$$

*where $\widehat{\beta}_{\tau,0} = \widehat{\beta}_\tau$ is given in (**??**), and, for $j = 1, \ldots, m$, $\widehat{\beta}_{\tau,j} = n^{-1}\sum_{u=1}^n \langle\widehat{\beta}_{\tau,j-1}, X_u\rangle_{\mathscr{H}} X_u$.*

Proof. The proof will be completed in two steps.

- Step 1: Prove that $\widehat{\beta}_{\tau,0} = b_{\tau,0} + O_p(n^{-1/2})$, where $b_{\tau,0} = \beta_\tau^*$ given in (4.5).
- Step 2: Prove that $\widehat{\beta}_{\tau,j} = b_{\tau,j} + O_p(n^{-1/2})$, for $j = 1, \ldots, m$, where $b_{\tau,j} = E(\langle b_{\tau,j-1}, X\rangle_{\mathscr{H}} X)$.

For Step 1: We know that $\widehat{\beta}_{\tau,0} = \widehat{\beta}_\tau$, where, according to (**??**),

$$(\widehat{\alpha}_\tau, \widehat{\beta}_\tau) = \arg\min_{(a_\tau, b_\tau)} \sum_{u=1}^n [\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\} - a_\tau - \langle b_\tau, X_u\rangle_{\mathscr{H}}]^2.$$

Without loss of generality, we assume that $\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}$ and X_u are centered. Then,

$$\widehat{\beta}_\tau = \arg\min_{b_\tau} \sum_{u=1}^{n} [\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\} - \langle b_\tau, X_u\rangle_{\mathscr{H}}]^2.$$

Minimizing the above objective function with respect to b_τ is equivalent with minimizing

$$\widehat{R}_n(b_\tau) = \frac{1}{2}\sum_{u=1}^{n} [\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\} - \langle b_\tau, X_u\rangle_{\mathscr{H}}]^2 - \frac{1}{2}\sum_{u=1}^{n} [\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}]^2,$$

with respect to b_τ. Expanding the square, $\widehat{R}_n(b_\tau)$ can be expressed as

$$\begin{aligned}
\widehat{R}_n(b_\tau) &= \frac{1}{2}\sum_{u=1}^{n} [\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}]^2 - \sum_{u=1}^{n} \widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}\langle b_\tau, X_u\rangle_{\mathscr{H}} \\
&\quad + \frac{1}{2}\sum_{u=1}^{n} \langle b_\tau, X_u\rangle_{\mathscr{H}}^2 - \frac{1}{2}\sum_{u=1}^{n} [\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}]^2 \\
&= -\sum_{u=1}^{n} \widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}\langle b_\tau, X_u\rangle_{\mathscr{H}} + \frac{1}{2}\sum_{u=1}^{n} \langle b_\tau, X_u\rangle_{\mathscr{H}}^2 \\
&= -\sum_{u=1}^{n} \widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}\langle b_\tau, X_u\rangle_{\mathscr{H}} + \frac{n}{2}\sum_{u=1}^{n} \langle b_\tau, \widehat{\Sigma}_{XX} b_\tau\rangle_{\mathscr{H}},
\end{aligned}$$

where the last equality follows from the fact that $\langle b_\tau, X_u\rangle_{\mathscr{H}}^2 = \langle b_\tau, X_u\rangle_{\mathscr{H}}\langle b_\tau, X_u\rangle_{\mathscr{H}} = \langle b_\tau, X_u\langle X_u, b_\tau\rangle_{\mathscr{H}}\rangle_{\mathscr{H}} = \langle b_\tau, (X_u \otimes X_u) b_\tau\rangle_{\mathscr{H}}$ and $\sum_{u=1}^{n}\langle b_\tau, X_u\rangle_{\mathscr{H}}^2 = \sum_{u=1}^{n}\langle b_\tau, (X_u \otimes X_u) b_\tau\rangle_{\mathscr{H}} = n\langle b_\tau, \widehat{\Sigma}_{XX} b_\tau\rangle_{\mathscr{H}}$.

Let $\gamma_\tau = \sqrt{n}(b_\tau - \beta_\tau^*)$, which implies that b_τ can be written as $b_\tau = \gamma_\tau/\sqrt{n} + \beta_\tau^*$. Then, $\widehat{R}_n(b_\tau)$ can be expressed as

$$\begin{aligned}
\widehat{R}_n(b_\tau) &= \widehat{R}_n(\gamma_\tau/\sqrt{n} + \beta_\tau^*) \\
&= -\frac{1}{\sqrt{n}}\sum_{u=1}^{n} \widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}\langle \gamma_\tau, X_u\rangle_{\mathscr{H}} - \sum_{u=1}^{n} \widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}\langle \beta_\tau^*, X_u\rangle_{\mathscr{H}} \\
&\quad + \frac{1}{2n}\sum_{u=1}^{n} \langle \gamma_\tau, X_u\rangle_{\mathscr{H}}^2 + \frac{1}{\sqrt{n}}\sum_{u=1}^{n} \langle \gamma_\tau, X_u\rangle_{\mathscr{H}}\langle \beta_\tau^*, X_u\rangle_{\mathscr{H}} + \frac{1}{2}\sum_{u=1}^{n} \langle \beta_\tau^*, X_u\rangle_{\mathscr{H}}^2.
\end{aligned}$$

To prove that $\widehat{\beta}_\tau$ is $\sqrt{n}$-consistent estimate of β^*_τ, we will prove that for any given $\delta_\tau > 0$, there exists a constant C_τ such that

$$P\left\{\inf_{\|\gamma_\tau\|_{\mathscr{H}} \geq C_\tau} \widehat{R}_n(\gamma_\tau/\sqrt{n} + \beta^*_\tau) > \widehat{R}_n(\beta^*_\tau)\right\} \geq 1 - \delta_\tau,$$

which implies that with probability at least $1 - \delta_\tau$ there exists a local minimum in the ball $\{\gamma_\tau/\sqrt{n} + \beta^*_\tau : \|\gamma_\tau\|_{\mathscr{H}} \leq C_\tau\}$. Since the objective function is strictly convex, the local minimum is a global minimum and $\left\|\widehat{\beta}_\tau - \beta^*_\tau\right\|_{\mathscr{H}} = O_p(n^{-1/2})$.

Enough to show that $\widehat{R}_n(\gamma_\tau/\sqrt{n} + \beta^*_\tau) - \widehat{R}_n(\beta^*_\tau) > 0$ with probability tending to one. Observe that

$$\begin{aligned}
& \widehat{R}_n(\gamma_\tau/\sqrt{n} + \beta^*_\tau) - \widehat{R}_n(\beta^*_\tau) \\
= \; & -\frac{1}{\sqrt{n}} \sum_{u=1}^{n} \widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}\langle \gamma_\tau, X_u\rangle_{\mathscr{H}} - \sum_{u=1}^{n} \widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}\langle \beta^*_\tau, X_u\rangle_{\mathscr{H}} \\
& + \frac{1}{2n} \sum_{u=1}^{n} \langle \gamma_\tau, X_u\rangle^2_{\mathscr{H}} + \frac{1}{\sqrt{n}} \sum_{u=1}^{n} \langle \gamma_\tau, X_u\rangle_{\mathscr{H}} \langle \beta^*_\tau, X_u\rangle_{\mathscr{H}} + \frac{1}{2} \sum_{u=1}^{n} \langle \beta^*_\tau, X_u\rangle^2_{\mathscr{H}} \\
& + \sum_{u=1}^{n} \widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}\langle \beta^*_\tau, X_u\rangle_{\mathscr{H}} - \frac{1}{2} \sum_{u=1}^{n} \langle \beta^*_\tau, X_u\rangle^2_{\mathscr{H}} \\
= \; & \frac{1}{2n} \sum_{u=1}^{n} \langle \gamma_\tau, X_u\rangle^2_{\mathscr{H}} + \frac{1}{\sqrt{n}} \sum_{u=1}^{n} \langle \gamma_\tau, X_u\rangle_{\mathscr{H}} \langle \beta^*_\tau, X_u\rangle_{\mathscr{H}} \\
& - \frac{1}{\sqrt{n}} \sum_{u=1}^{n} \widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}\langle \gamma_\tau, X_u\rangle_{\mathscr{H}} \\
= \; & H_1 + H_2 + H_3.
\end{aligned}$$

It is enough to show that $\widehat{R}_n(\gamma_\tau/\sqrt{n} + \beta^*_\tau) - \widehat{R}_n(\beta^*_\tau)$ is dominated by the first term H_1, which is positive, and that $H_2 + H_3$ is bounded.

For H_1,

$$H_1 = \frac{1}{2n}\sum_{u=1}^{n}\langle\gamma_\tau, X_u\rangle_{\mathscr{H}}^2 = \frac{1}{2}\langle\gamma_\tau, \widehat{\Sigma}_{XX}\gamma_\tau\rangle_{\mathscr{H}},$$

which is positive and bounded.

For $H_2 + H_3$,

$$\begin{aligned} H_2 + H_3 &= \frac{1}{\sqrt{n}}\sum_{u=1}^{n}\langle\gamma_\tau, X_u\rangle_{\mathscr{H}}\langle\beta_\tau^*, X_u\rangle_{\mathscr{H}} - \frac{1}{\sqrt{n}}\sum_{u=1}^{n}\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}\langle\gamma_\tau, X_u\rangle_{\mathscr{H}} \\ &= \frac{1}{\sqrt{n}}\sum_{u=1}^{n}[\langle\beta_\tau^*, X_u\rangle_{\mathscr{H}} - \widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}]\langle\gamma_\tau, X_u\rangle_{\mathscr{H}} \\ &= \frac{1}{\sqrt{n}}\sum_{u=1}^{n}[\langle\beta_\tau^*, X_u\rangle_{\mathscr{H}} - \widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}]X_u^*(\gamma_\tau), \end{aligned}$$

where $X_u^* : \mathscr{H} \to \mathbb{R}$ is a linear operator such that $X_u^*(\gamma_\tau) = \langle\gamma_\tau, X_u\rangle_{\mathscr{H}}$. Since $E\|X\|_{\mathscr{H}} < \infty$, then $E\|X^*\| < \infty$ and $X^* = O_p(1)$. Then, it is enough to show that $n^{-1/2}\sum_{u=1}^{n}[\langle\beta_\tau^*, X_u\rangle_{\mathscr{H}} - \widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}] = O_p(1)$ by showing that the second moment is bounded. Since the quantity involves $\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}$, which is data dependent and not deterministic, we define $n^{-1/2}\sum_{u=1}^{n}[\langle\beta_\tau^*, X_u\rangle_{\mathscr{H}} - \phi_\tau\{Y|L(X_u)\}]$, where $\phi_\tau : \mathbb{R}^{d+1} \to \mathbb{R}$ is a function in the class Φ_τ, whose value at $(Y, L(X_u)) \in \mathbb{R}^{d+1}$ can be written as $\phi_\tau\{Y|L(X)\}$ and satisfying $E[\phi_\tau^2\{Y|L(X_u)\}] < \infty$. Since Φ_τ includes $Q_\tau\{Y|L(X_u)\}$ and, according to Lemma A.1, includes $\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}$ for n large enough, almost surely, we will prove that the second moment of $n^{-1/2}\sum_{u=1}^{n}[\langle\beta_\tau^*, X_u\rangle_{\mathscr{H}} - \phi_\tau\{Y|L(X_u)\}]$ is bounded, uniformly on $\phi_\tau \in \Phi_\tau$.

Then,

$$\begin{aligned}
& \sup_{\phi_\tau \in \Phi_\tau} E\left[\frac{1}{\sqrt{n}} \sum_{u=1}^{n} [\langle \beta_\tau^*, X_u \rangle_{\mathscr{H}} - \phi_\tau\{Y|L(X_u)\}]\right]^2 \\
\leq & \sup_{\phi_\tau \in \Phi_\tau} \left[\frac{1}{n} \sum_{u=1}^{n} E\left[\langle \beta_\tau^*, X_u \rangle_{\mathscr{H}} - \phi_\tau\{Y|L(X_u)\}\right]^2\right] \\
\leq & \sup_{\phi_\tau \in \Phi_\tau} \left[\frac{1}{n} \sum_{u=1}^{n} \left[E(\langle \beta_\tau^*, X_u \rangle_{\mathscr{H}}^2) + E[\phi_\tau^2\{Y|L(X_u)\}]\right]\right] \\
= & \sup_{\phi_\tau \in \Phi_\tau} \left[E[\langle \beta_\tau^*, \widehat{\Sigma}_{XX} \beta_\tau^* \rangle_{\mathscr{H}}] + \frac{1}{n} \sum_{u=1}^{n} E[\phi_\tau^2\{Y|L(X_u)\}\right] \\
= & O(1).
\end{aligned}$$

Since the result holds uniformly on ϕ_τ and since the class Φ_τ includes $\widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}$ for n large enough, almost surely, then $n^{-1/2}\sum_{u=1}^{n}[\langle \beta_\tau^*, X_u \rangle_{\mathscr{H}} - \widehat{Q}_\tau\{Y|\widehat{L}(X_u)\}] = O_p(1)$ and the proof is complete.

For Step 2: We already proved in Lemma A.2 that, for $j = 1, \ldots, m$, $b_{\tau,j} = \Sigma_{XX}^j b_{\tau,0}$. Similarly, we can prove that, for $j = 1, \ldots, m$, $\widehat{\beta}_{\tau,j} = \widehat{\Sigma}_{XX}^j \widehat{\beta}_{\tau,0}$. Then,

$$\begin{aligned}
\widehat{\beta}_{\tau,j} - b_{\tau,j} &= \widehat{\Sigma}_{XX}^j \widehat{\beta}_{\tau,0} - \Sigma_{XX}^j b_{\tau,0} = (\widehat{\Sigma}_{XX}^j \pm \Sigma_{XX}^j)\widehat{\beta}_{\tau,0} - \Sigma_{XX}^j b_{\tau,0} \\
&= (\widehat{\Sigma}_{XX}^j - \Sigma_{XX}^j)\widehat{\beta}_{\tau,0} + \Sigma_{XX}^j(\widehat{\beta}_{\tau,0} - b_{\tau,0}) \\
&= (\widehat{\Sigma}_{XX}^j - \Sigma_{XX}^j)(\widehat{\beta}_{\tau,0} \pm b_{\tau,0}) + \Sigma_{XX}^j(\widehat{\beta}_{\tau,0} - b_{\tau,0}) \\
&= (\widehat{\Sigma}_{XX}^j - \Sigma_{XX}^j)(\widehat{\beta}_{\tau,0} - b_{\tau,0}) + (\widehat{\Sigma}_{XX}^j - \Sigma_{XX}^j)b_{\tau,0} + \Sigma_{XX}^j(\widehat{\beta}_{\tau,0} - b_{\tau,0}) \\
&= O_p(n^{-1/2})O_p(n^{-1/2}) + O_p(n^{-1/2}) + O_p(n^{-1/2}) \\
&= O_p(n^{-1}) + O_p(n^{-1/2}) = O_p(n^{-1/2}).
\end{aligned}$$

□

Lemma A.4. *Under the assumptions of Lemma A.3,* $\widehat{B}_\tau$ *is a consistent estimate of*

B_τ, and

$$\left\| \widehat{B}_\tau - B_\tau \right\| = O_p(n^{-1/2}),$$

where $\widehat{B}_\tau = \sum_{j=0}^{m} \widehat{\beta}_{\tau,j} \otimes \widehat{\beta}_{\tau,j}$ *and* $B_\tau = \sum_{j=0}^{m} b_{\tau,j} \otimes b_{\tau,j}$.

Proof. Recall that $\widehat{B}_\tau = \sum_{j=0}^{m} \widehat{\beta}_{\tau,j} \otimes \widehat{\beta}_{\tau,j}$, where $\widehat{\beta}_{\tau,0} = \widehat{\beta}_\tau$ given in (**??**), and, for $j = 1, \ldots, m$, $\widehat{\beta}_{\tau,j} = n^{-1} \sum_{u=1}^{n} \langle \widehat{\beta}_{\tau,j-1}, X_u \rangle_{\mathscr{H}} X_u$. Similarly, $B_\tau = \sum_{j=0}^{m} b_{\tau,j} \otimes b_{\tau,j}$, where $b_{\tau,0} = \beta_\tau^*$ given in (4.5), and, for $j = 1, \ldots, m$, $b_{\tau,j} = E(\langle b_{\tau,j-1}, X \rangle_{\mathscr{H}} X)$.

We have that

$$\begin{aligned}
\widehat{B}_\tau &= \sum_{j=0}^{m} \widehat{\beta}_{\tau,j} \otimes \widehat{\beta}_{\tau,j} = \sum_{j=0}^{m} (\widehat{\beta}_{\tau,j} \pm b_{\tau,j}) \otimes (\widehat{\beta}_{\tau,j} \pm b_{\tau,j}) \\
&= \sum_{j=0}^{m} (\widehat{\beta}_{\tau,j} - b_{\tau,j}) \otimes (\widehat{\beta}_{\tau,j} - b_{\tau,j}) + \sum_{j=0}^{m} (\widehat{\beta}_{\tau,j} - b_{\tau,j}) \otimes b_{\tau,j} \\
&\quad + \sum_{j=0}^{m} b_{\tau,j} \otimes (\widehat{\beta}_{\tau,j} - b_{\tau,j}) + \sum_{j=0}^{m} b_{\tau,j} \otimes b_{\tau,j} \\
&= O_p(n^{-1}) + O_p(n^{-1/2}) + O_p(n^{-1/2}) + B_\tau \\
&= O_p(n^{-1/2}) + B_\tau,
\end{aligned}$$

where the third equality follows from Lemma A.3.

□

A.4 Proofs of Main Results

A.4.1 Proof of Theorem 2.1

Assumption 2.2 states that, for a given τ, there is a bounded linear operator $\Lambda_\tau : \mathrm{ran}(L_\tau) \to \mathscr{H}$ such that $E\{X | L_\tau(X)\} = \Lambda_\tau L_\tau(X)$. We take the tensor product with $L_\tau(X)$ on both sides to get $E\{X | L_\tau(X)\} \otimes L_\tau(X) = \Lambda_\tau L_\tau(X) \otimes L_\tau(X)$. Then, we

take the expectation on both sides to get

$$\begin{aligned} & E[E\{X|L_\tau(X)\} \otimes L_\tau(X)] = E\{\Lambda_\tau L_\tau(X) \otimes L_\tau(X)\} \\ \Rightarrow \quad & E[E\{X \otimes L_\tau(X)|L_\tau(X)\}] = E\{\Lambda_\tau L_\tau(X) \otimes L_\tau(X)\}. \end{aligned}$$

The left hand side of the above expression is equal to $E\{X \otimes L_\tau(X)\}$ and therefore, we get that

$$E\{X \otimes L_\tau(X)\} = E\{\Lambda_\tau L_\tau(X) \otimes L_\tau(X)\}. \tag{A.8}$$

Next, proving that

(a) $X \otimes L_\tau(X) = (X \otimes X)L_\tau^*$; and

(b) $\Lambda_\tau L_\tau(X) \otimes L_\tau(X) = \Lambda_\tau L_\tau(X \otimes X)L_\tau^*$,

we can replace the left hand side of expression (A.8) with (a) and the right hand side of expression (A.8) with (b), i.e.,

$$\begin{aligned} & E\{X \otimes L_\tau(X)\} = E\{\Lambda_\tau L_\tau(X) \otimes L_\tau(X)\} \\ \Rightarrow \quad & E\{(X \otimes X)L_\tau^*\} = E\{\Lambda_\tau L_\tau(X \otimes X)L_\tau^*\} \\ \Rightarrow \quad & E(X \otimes X)L_\tau^* = \Lambda_\tau L_\tau E(X \otimes X)L_\tau^* \\ \Rightarrow \quad & \Sigma_{XX}L_\tau^* = \Lambda_\tau L_\tau \Sigma_{XX}L_\tau^*. \end{aligned}$$

Then, multiply both sides of the equation from the right by $(L_\tau \Sigma_{XX} L_\tau^*)^\dagger$ and use the fact $AA^\dagger = I_{\text{ran}(A)}$ to get $\Lambda_\tau = \Sigma_{XX}L_\tau^*(L_\tau \Sigma_{XX} L_\tau^*)^\dagger$. Finally, plugging Λ_τ to Assumption 2.2, we get $E\{X|L_\tau(X)\} = \Lambda_\tau L_\tau(X) = \Sigma_{XX}L_\tau^*(L_\tau \Sigma_{XX} L_\tau^*)^\dagger L_\tau(X)$.

Remains to prove relations (a) and (b). For (a) we have that for any $h \in \mathscr{H}$,

$$\{X \otimes L_\tau(X)\}h = X\langle L_\tau(X), h\rangle_{\mathscr{H}} = X\langle X, L_\tau^* h\rangle_{\mathscr{H}} = (X \otimes X)L_\tau^* h,$$

and therefore, $X \otimes L_\tau(X) = (X \otimes X)L_\tau^*$. Similarly for (b), we have that for any $h \in \mathscr{H}$,

$$\begin{aligned}\Lambda_\tau L_\tau(X) \otimes L_\tau(X)h &= \Lambda_\tau L_\tau(X)\langle L_\tau(X), h\rangle_{\mathscr{H}} = \Lambda_\tau L_\tau(X)\langle X, L_\tau^* h\rangle_{\mathscr{H}} \\ &= \Lambda_\tau L_\tau(X \otimes X)L_\tau^* h,\end{aligned}$$

and therefore, $\Lambda_\tau L_\tau(X) \otimes L_\tau(X) = \Lambda_\tau L_\tau(X \otimes X)L_\tau^*$. The proof is complete.

A.4.2 Proof of Theorem 2.2

Without loss of generality, assume that $E(X) = 0$. Define $P_{L_\tau}(\Sigma_{XX}) = L_\tau^*(L_\tau \Sigma_{XX} L_\tau^*)^\dagger L_\tau \Sigma_{XX}$ such that $E\{X|L_\tau(X)\} = P_{L_\tau}^*(\Sigma_{XX})X$. Then,

$$\begin{aligned}R(a_\tau, b_\tau) &= E[Q_\tau\{Y|L(X)\} - a_\tau - \langle b_\tau, X\rangle_{\mathscr{H}}]^2 \\ &= E[E[[Q_\tau\{Y|L(X)\} - a_\tau - \langle b_\tau, X\rangle_{\mathscr{H}}]^2|L_\tau(X)]] \\ &\geq E[[E[Q_\tau\{Y|L(X)\} - a_\tau - \langle b_\tau, X\rangle_{\mathscr{H}}|L_\tau(X)]]^2] \\ &= E[[E[Q_\tau\{Y|L(X)\}|L_\tau(X)] - a_\tau - E[\langle b_\tau, X\rangle_{\mathscr{H}}|L_\tau(X)]]^2] \\ &= E[[E[Q_\tau\{Y|L(X)\}|L_\tau(X)] - a_\tau - \langle b_\tau, E\{X|L_\tau(X)\}\rangle_{\mathscr{H}}]^2] \\ &= E[[E[Q_\tau\{Y|L(X)\}|L_\tau(X)] - a_\tau - \langle b_\tau, P_{L_\tau}^*(\Sigma_{XX})X\rangle_{\mathscr{H}}]^2] \\ &= E[[E[Q_\tau\{Y|L(X)\}|L_\tau(X)] - a_\tau - \langle P_{L_\tau}(\Sigma_{XX})b_\tau, X\rangle_{\mathscr{H}}]^2] \\ &= E[Q_\tau\{Y|L(X)\} - a_\tau - \langle P_{L_\tau}(\Sigma_{XX})b_\tau, X\rangle_{\mathscr{H}}]^2 = R(a_\tau, P_{L_\tau}(\Sigma_{XX})b_\tau),\end{aligned}$$

where the last line follows from the fact that $Q_\tau(Y|X) = Q_\tau\{Y|L(X)\} = Q_\tau\{Y|L_\tau(X)\}$ and therefore,

$$\begin{aligned}E[Q_\tau\{Y|L(X)\}|L_\tau(X)] &= E[Q_\tau\{Y|L_\tau(X)\}|L_\tau(X)] \\ &= Q_\tau\{Y|L_\tau(X)\} = Q_\tau\{Y|L(X)\}.\end{aligned}$$

Finally, since $R(a_\tau, b_\tau)$ is convex and has a unique minimum $(\alpha_\tau^*, \beta_\tau^*)$, then $\beta_\tau^* = P_{L_\tau}(\Sigma_{XX})\beta_\tau^*$ and $\beta_\tau^* \in \mathscr{S}_{Q_\tau(Y|X)}$.

A.4.3 Proof of Theorem 2.3

Without loss of generality, assume that $E(X) = 0$. Define $P_{L_\tau}(\Sigma_{XX}) = L_\tau^*(L_\tau \Sigma_{XX} L_\tau^*)^\dagger L_\tau \Sigma_{XX}$ such that $E\{X|L_\tau(X)\} = P_{L_\tau}^*(\Sigma_{XX})X$. To prove that $E(\langle b_\tau, X\rangle_{\mathscr{H}} X) \in \Sigma_{XX}\mathscr{S}_{Q_\tau(Y|X)}$ it is enough to prove that $E(\langle b_\tau, X\rangle_{\mathscr{H}} X) = P_{L_\tau}^*(\Sigma_{XX})E(\langle b_\tau, X\rangle_{\mathscr{H}} X)$. Since $b_\tau \in \mathscr{S}_{Q_\tau(Y|X)}$, then $\langle b_\tau, X\rangle_{\mathscr{H}}$ is a measurable function of $L_\tau(X)$. Then

$$\begin{aligned} E(\langle b_\tau, X\rangle_{\mathscr{H}} X) &= E[E\{\langle b_\tau, X\rangle_{\mathscr{H}} X|L_\tau(X)\}] = E[\langle b_\tau, X\rangle_{\mathscr{H}} E\{X|L_\tau(X)\}] \\ &= E\{\langle b_\tau, X\rangle_{\mathscr{H}} P_{L_\tau}^*(\Sigma_{XX})X\} = P_{L_\tau}^*(\Sigma_{XX})E(\langle b_\tau, X\rangle_{\mathscr{H}} X). \end{aligned}$$

The reason $P_{L_\tau}^*(\Sigma_{XX})E(\langle b_\tau, X\rangle_{\mathscr{H}} X) \in \Sigma_{XX}\mathscr{S}_{Q_\tau(Y|X)}$ is because if $A \in \mathscr{B}(\mathscr{H})$, then $\Sigma_{XX}L_\tau^* A \in \Sigma_{XX}\mathscr{S}_{Q_\tau(Y|X)}$. To prove this, note that for $A \in \mathscr{B}(\mathscr{H})$, $\overline{\text{ran}(L_\tau^* A)} \subseteq \overline{\text{ran}(L_\tau^*)}$. Then, if $x \in \mathscr{H}$,

$$\Sigma_{XX}L_\tau^* A(x) \in \Sigma_{XX}\overline{\text{ran}(L_\tau^* A)} \subseteq \Sigma_{XX}\overline{\text{ran}(L_\tau^*)}.$$

Therefore,

$$E(\langle b_\tau, X\rangle_{\mathscr{H}} X) \in \overline{\text{ran}(\Sigma_{XX}L_\tau^*)} \subseteq \Sigma_{XX}\mathscr{S}_{Q_\tau(Y|X)}.$$

A.4.4 Proof of Theorem 2.4

We want to prove that $\widehat{M}_\tau = (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}\widehat{B}_\tau(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}$ is a consistent estimate of $M_\tau = \Sigma_{XX}^{\dagger 1/2} B_\tau \Sigma_{XX}^{\dagger 1/2}$.

Note that,

$$\begin{aligned}
\widehat{M}_\tau &= (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} \widehat{B}_\tau (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} \\
&= (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} (\widehat{B}_\tau \pm B_\tau)(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} \\
&= (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} (\widehat{B}_\tau - B_\tau)(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} \\
&\quad + (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} B_\tau (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} \\
&= I_1 + I_2.
\end{aligned}$$

The first expression I_1 can be written as

$$\begin{aligned}
I_1 &= (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} (\widehat{B}_\tau - B_\tau)(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} \\
&= O_p(\epsilon_n^{-1/2}) O_p(n^{-1/2}) O_p(\epsilon_n^{-1/2}) = O_p(n^{-1/2}\epsilon_n^{-1}),
\end{aligned}$$

where the second equality follows from (A.2), the fact that $n^{-1/2} \prec n^{-1/4} \prec \epsilon_n \prec 1$, and Lemma A.4.

The second expression I_2 can be written as

$$\begin{aligned}
I_2 &= (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} B_\tau (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} \\
&= \{(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} \pm (\Sigma_{XX} + \epsilon_n I)^{-1/2}\} B_\tau \{(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} \pm (\Sigma_{XX} + \epsilon_n I)^{-1/2}\} \\
&= \{(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} - (\Sigma_{XX} + \epsilon_n I)^{-1/2}\} B_\tau \{(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} - (\Sigma_{XX} + \epsilon_n I)^{-1/2}\} \\
&\quad + \{(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} - (\Sigma_{XX} + \epsilon_n I)^{-1/2}\} B_\tau (\Sigma_{XX} + \epsilon_n I)^{-1/2} \\
&\quad + (\Sigma_{XX} + \epsilon_n I)^{-1/2} B_\tau \{(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} - (\Sigma_{XX} + \epsilon_n I)^{-1/2}\} \\
&\quad + (\Sigma_{XX} + \epsilon_n I)^{-1/2} B_\tau (\Sigma_{XX} + \epsilon_n I)^{-1/2} \\
&= J_1 + J_2 + J_3 + J_4. \qquad \text{(A.9)}
\end{aligned}$$

For J_1, we use relationship (A.3) to expand $(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} - (\Sigma_{XX} + \epsilon_n I)^{-1/2}$ and

relationship (A.7) to express B_τ. This will give us

$$
\begin{aligned}
J_1 &= \{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}-(\Sigma_{XX}+\epsilon_n I)^{-1/2}\}B_\tau\{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}-(\Sigma_{XX}+\epsilon_n I)^{-1/2}\} \\
&= [(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\{(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}\}+\widehat{\Sigma}_{XX}-\Sigma_{XX}](\Sigma_{XX}+\epsilon_n I)^{-3/2} \\
&\quad \{\Sigma_{XX}(\omega_\tau\otimes\omega_\tau)\Sigma_{XX}+\Sigma_{XX}^2(\omega_\tau\otimes\omega_\tau)\Sigma_{XX}^2+\cdots+\Sigma_{XX}^{m+1}(\omega_\tau\otimes\omega_\tau)\Sigma_{XX}^{m+1}\} \\
&\quad [(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\{(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}\}+\widehat{\Sigma}_{XX}-\Sigma_{XX}](\Sigma_{XX}+\epsilon_n I)^{-3/2} \\
&= [(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\{(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}\}+\widehat{\Sigma}_{XX}-\Sigma_{XX}](\Sigma_{XX}+\epsilon_n I)^{-3/2}\Sigma_{XX} \\
&\quad (\omega_\tau\otimes\omega_\tau)\Sigma_{XX}[(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\{(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}\}+\widehat{\Sigma}_{XX}-\Sigma_{XX}](\Sigma_{XX}+\epsilon_n I)^{-3/2} \\
&\quad +[(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\{(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}\}+\widehat{\Sigma}_{XX}-\Sigma_{XX}](\Sigma_{XX}+\epsilon_n I)^{-3/2}\Sigma_{XX}^2 \\
&\quad (\omega_\tau\otimes\omega_\tau)\Sigma_{XX}^2[(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\{(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}\}+\widehat{\Sigma}_{XX}-\Sigma_{XX}](\Sigma_{XX}+\epsilon_n I)^{-3/2} \\
&\quad +\ldots \\
&\quad +[(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\{(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}\}+\widehat{\Sigma}_{XX}-\Sigma_{XX}](\Sigma_{XX}+\epsilon_n I)^{-3/2}\Sigma_{XX}^{m+1} \\
&\quad (\omega_\tau\otimes\omega_\tau)\Sigma_{XX}^{m+1}[(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\{(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}\}+\widehat{\Sigma}_{XX}-\Sigma_{XX}](\Sigma_{XX}+\epsilon_n I)^{-3/2}.
\end{aligned}
$$

Note that, since

- $\widehat{\Sigma}_{XX}=\Sigma_{XX}+O_p(n^{-1/2})$ and $n^{-1/2}\prec\epsilon_n\prec 1$, then $(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}=O_p(\epsilon_n^{-1/2})$,
- relationship (A.4) implies that $(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}=O_p(n^{-1/2})$,
- relationship (A.1) implies that $(\Sigma_{XX}+\epsilon_n I)^{-3/2}\Sigma_{XX}^a=O(\epsilon_n^{min\{0,a-3/2\}})$, for $a=1,\ldots,m+1$,
- $(\omega_\tau\otimes\omega_\tau)\Sigma_{XX}^a=O(1)$, for every $a=1,\ldots,m+1$,
- $(\Sigma_{XX}+\epsilon_n I)^{-3/2}=O(\epsilon_n^{-3/2})$,

then J_1 gives us

$$
\begin{aligned}
J_1 &= \{O_p(\epsilon_n^{-1/2})O_p(n^{-1/2}) + O_p(n^{-1/2})\}O(\epsilon_n^{min\{0,1-3/2\}})O(1)\{O_p(\epsilon_n^{-1/2})O_p(n^{-1/2}) + O_p(n^{-1/2})\}O(\epsilon_n^{-3/2}) \\
&\quad +\{O_p(\epsilon_n^{-1/2})O_p(n^{-1/2}) + O_p(n^{-1/2})\}O(\epsilon_n^{min\{0,2-3/2\}})O(1)\{O_p(\epsilon_n^{-1/2})O_p(n^{-1/2}) + O_p(n^{-1/2})\}O(\epsilon_n^{-3/2}) \\
&\quad +\ldots \\
&\quad +\{O_p(\epsilon_n^{-1/2})O_p(n^{-1/2}) + O_p(n^{-1/2})\}O(\epsilon_n^{min\{0,m+1-3/2\}})O(1)\{O_p(\epsilon_n^{-1/2})O_p(n^{-1/2}) + O_p(n^{-1/2})\}O(\epsilon_n^{-3/2}) \\
&= O_p(n^{-1/2}\epsilon_n^{-1/2})O(\epsilon_n^{-1/2})O(1)O_p(n^{-1/2}\epsilon_n^{-1/2})O(\epsilon_n^{-3/2}) \\
&\quad +O_p(n^{-1/2}\epsilon_n^{-1/2})O(1)O(1)O_p(n^{-1/2}\epsilon_n^{-1/2})O(\epsilon_n^{-3/2}) + \ldots \\
&\quad +O_p(n^{-1/2}\epsilon_n^{-1/2})O(1)O(1)O_p(n^{-1/2}\epsilon_n^{-1/2})O(\epsilon_n^{-3/2}) \\
&= O_p(n^{-1}\epsilon_n^{-3}) + O_p(n^{-1}\epsilon_n^{-5/2}) + \cdots + O_p(n^{-1}\epsilon_n^{-5/2}) \\
&= O_p(n^{-1}\epsilon_n^{-3}).
\end{aligned}
$$

Using similar arguments and steps for J_2 we have

$$
\begin{aligned}
J_2 &= \{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}-(\Sigma_{XX}+\epsilon_n I)^{-1/2}\}B_\tau(\Sigma_{XX}+\epsilon_n I)^{-1/2} \\
&= [(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\{(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}\}+\widehat{\Sigma}_{XX}-\Sigma_{XX}](\Sigma_{XX}+\epsilon_n I)^{-3/2} \\
&\quad \{\Sigma_{XX}(\omega_\tau\otimes\omega_\tau)\Sigma_{XX}+\Sigma_{XX}^2(\omega_\tau\otimes\omega_\tau)\Sigma_{XX}^2+\cdots+\Sigma_{XX}^{m+1}(\omega_\tau\otimes\omega_\tau)\Sigma_{XX}^{m+1}\}(\Sigma_{XX}+\epsilon_n I)^{-1/2} \\
&= [(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\{(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}\}+\widehat{\Sigma}_{XX}-\Sigma_{XX}](\Sigma_{XX}+\epsilon_n I)^{-3/2}\Sigma_{XX} \\
&\quad (\omega_\tau\otimes\omega_\tau)\Sigma_{XX}(\Sigma_{XX}+\epsilon_n I)^{-1/2} \\
&\quad +[(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\{(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}\}+\widehat{\Sigma}_{XX}-\Sigma_{XX}](\Sigma_{XX}+\epsilon_n I)^{-3/2}\Sigma_{XX}^2 \\
&\quad (\omega_\tau\otimes\omega_\tau)\Sigma_{XX}^2(\Sigma_{XX}+\epsilon_n I)^{-1/2} \\
&\quad +\ldots \\
&\quad +[(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\{(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}\}+\widehat{\Sigma}_{XX}-\Sigma_{XX}](\Sigma_{XX}+\epsilon_n I)^{-3/2}\Sigma_{XX}^{m+1} \\
&\quad (\omega_\tau\otimes\omega_\tau)\Sigma_{XX}^{m+1}(\Sigma_{XX}+\epsilon_n I)^{-1/2} \\
&= [(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\{(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}\}+\widehat{\Sigma}_{XX}-\Sigma_{XX}](\Sigma_{XX}+\epsilon_n I)^{-3/2}\Sigma_{XX} \\
&\quad (\omega_\tau\otimes\omega_\tau)(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX} \\
&\quad +[(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\{(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}\}+\widehat{\Sigma}_{XX}-\Sigma_{XX}](\Sigma_{XX}+\epsilon_n I)^{-3/2}\Sigma_{XX}^2 \\
&\quad (\omega_\tau\otimes\omega_\tau)(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^2 \\
&\quad +\ldots \\
&\quad +[(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\{(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}\}+\widehat{\Sigma}_{XX}-\Sigma_{XX}](\Sigma_{XX}+\epsilon_n I)^{-3/2}\Sigma_{XX}^{m+1} \\
&\quad (\omega_\tau\otimes\omega_\tau)(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{m+1},
\end{aligned}
$$

where the last equality follows from the fact that Σ_{XX}^a commutes with $(\Sigma_{XX}+\epsilon_n I)^{-3/2}$, for $a=1,2,\ldots,m+1$. Then, using similar rates of convergence as before,

we have

$$\begin{aligned}
J_2 &= \{O_p(\epsilon_n^{-1/2})O_p(n^{-1/2}) + O_p(n^{-1/2})\}O(\epsilon_n^{min\{0,1-3/2\}})O(1)O(\epsilon_n^{min\{0,1-1/2\}}) + \{O_p(\epsilon_n^{-1/2})O_p(n^{-1/2}) \\
&\quad +O_p(n^{-1/2})\}O(\epsilon_n^{min\{0,2-3/2\}})O(1)O(\epsilon_n^{min\{0,2-1/2\}}) + \dots \\
&\quad +\{O_p(\epsilon_n^{-1/2})O_p(n^{-1/2}) + O_p(n^{-1/2})\}O(\epsilon_n^{min\{0,m+1-3/2\}})O(1)O(\epsilon_n^{min\{0,m+1-1/2\}}) \\
&= O_p(n^{-1/2}\epsilon_n^{-1}) + O_p(n^{-1/2}\epsilon_n^{-1/2}) + \cdots + O_p(n^{-1/2}\epsilon_n^{-1/2}) \\
&= O_p(n^{-1/2}\epsilon_n^{-1}).
\end{aligned}$$

For J_3, we use similar steps to obtain

$$\begin{aligned}
J_3 &= (\Sigma_{XX} + \epsilon_n I)^{-1/2} B_\tau \{(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} - (\Sigma_{XX} + \epsilon_n I)^{-1/2}\} \\
&= (\Sigma_{XX} + \epsilon_n I)^{-1/2} \{\Sigma_{XX}(\omega_\tau \otimes \omega_\tau)\Sigma_{XX} + \Sigma_{XX}^2(\omega_\tau \otimes \omega_\tau)\Sigma_{XX}^2 + \cdots + \Sigma_{XX}^{m+1}(\omega_\tau \otimes \omega_\tau)\Sigma_{XX}^{m+1}\} \\
&\quad [(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}\{(\Sigma_{XX} + \epsilon_n I)^{3/2} - (\widehat{\Sigma}_{XX} + \epsilon_n I)^{3/2}\} + \widehat{\Sigma}_{XX} - \Sigma_{XX}](\Sigma_{XX} + \epsilon_n I)^{-3/2} \\
&= (\Sigma_{XX} + \epsilon_n I)^{-1/2}\Sigma_{XX}(\omega_\tau \otimes \omega_\tau)\Sigma_{XX} \\
&\quad [(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}\{(\Sigma_{XX} + \epsilon_n I)^{3/2} - (\widehat{\Sigma}_{XX} + \epsilon_n I)^{3/2}\} + \widehat{\Sigma}_{XX} - \Sigma_{XX}](\Sigma_{XX} + \epsilon_n I)^{-3/2} \\
&\quad +(\Sigma_{XX} + \epsilon_n I)^{-1/2}\Sigma_{XX}^2(\omega_\tau \otimes \omega_\tau)\Sigma_{XX}^2 \\
&\quad [(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}\{(\Sigma_{XX} + \epsilon_n I)^{3/2} - (\widehat{\Sigma}_{XX} + \epsilon_n I)^{3/2}\} + \widehat{\Sigma}_{XX} - \Sigma_{XX}](\Sigma_{XX} + \epsilon_n I)^{-3/2} \\
&\quad +\dots \\
&\quad +(\Sigma_{XX} + \epsilon_n I)^{-1/2}\Sigma_{XX}^{m+1}(\omega_\tau \otimes \omega_\tau)\Sigma_{XX}^{m+1} \\
&\quad [(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}\{(\Sigma_{XX} + \epsilon_n I)^{3/2} - (\widehat{\Sigma}_{XX} + \epsilon_n I)^{3/2}\} + \widehat{\Sigma}_{XX} - \Sigma_{XX}](\Sigma_{XX} + \epsilon_n I)^{-3/2} \\
&= O(\epsilon_n^{min\{0,1-1/2\}})O(1)\{O_p(\epsilon_n^{-1/2})O_p(n^{-1/2}) + O_p(n^{-1/2})\}O(\epsilon_n^{-3/2}) \\
&\quad +O(\epsilon_n^{min\{0,2-1/2\}})O(1)\{O_p(\epsilon_n^{-1/2})O_p(n^{-1/2}) + O_p(n^{-1/2})\}O(\epsilon_n^{-3/2}) \\
&\quad +\dots \\
&\quad +O(\epsilon_n^{min\{0,m+1-1/2\}})O(1)\{O_p(\epsilon_n^{-1/2})O_p(n^{-1/2}) + O_p(n^{-1/2})\}O(\epsilon_n^{-3/2}) \\
&= O_p(n^{-1/2}\epsilon_n^{-2}).
\end{aligned}$$

Finally, for J_4, observe that

$$\begin{aligned}
J_4 &= (\Sigma_{XX}+\epsilon_n I)^{-1/2} B_\tau (\Sigma_{XX}+\epsilon_n I)^{-1/2} \\
&= (\Sigma_{XX}+\epsilon_n I)^{-1/2} \Sigma_{XX}^{1/2} \Sigma_{XX}^{\dagger 1/2} B_\tau \Sigma_{XX}^{\dagger 1/2} \Sigma_{XX}^{1/2} (\Sigma_{XX}+\epsilon_n I)^{-1/2} \\
&= (\Sigma_{XX}+\epsilon_n I)^{-1/2} \Sigma_{XX}^{1/2} M_\tau \Sigma_{XX}^{1/2} (\Sigma_{XX}+\epsilon_n I)^{-1/2} \\
&= \{(\Sigma_{XX}+\epsilon_n I)^{-1/2} \Sigma_{XX}^{1/2} \pm I\} M_\tau \{\Sigma_{XX}^{1/2} (\Sigma_{XX}+\epsilon_n I)^{-1/2} \pm I\} \\
&= \{(\Sigma_{XX}+\epsilon_n I)^{-1/2} \Sigma_{XX}^{1/2} - I\} M_\tau \{\Sigma_{XX}^{1/2} (\Sigma_{XX}+\epsilon_n I)^{-1/2} - I\} + M_\tau \{\Sigma_{XX}^{1/2} (\Sigma_{XX}+\epsilon_n I)^{-1/2} - I\} \\
&\quad + \{(\Sigma_{XX}+\epsilon_n I)^{-1/2} \Sigma_{XX}^{1/2} - I\} M_\tau + M_\tau \\
&= \{(\Sigma_{XX}+\epsilon_n I)^{-1/2} \Sigma_{XX}^{1/2} - I\} M_\tau \{(\Sigma_{XX}+\epsilon_n I)^{-1/2} \Sigma_{XX}^{1/2} - I\} + M_\tau \{(\Sigma_{XX}+\epsilon_n I)^{-1/2} \Sigma_{XX}^{1/2} - I\} \\
&\quad + \{(\Sigma_{XX}+\epsilon_n I)^{-1/2} \Sigma_{XX}^{1/2} - I\} M_\tau + M_\tau,
\end{aligned}$$

where the second equality follows from part (b) of Lemma A.2 and the last equality follows from the fact that $\Sigma_{XX}^{1/2}$ and $(\Sigma_{XX}+\epsilon_n I)^{-1/2}$ are commutative. Therefore, the main term to investigate is $(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2} - I$.

Observe that

$$(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2} - I = (\Sigma_{XX}+\epsilon_n I)^{-1/2}\{\Sigma_{XX}^{1/2} - (\Sigma_{XX}+\epsilon_n I)^{1/2}\}.$$

If γ_1 denotes the largest eigenvalue of Σ_{XX}, then $\left\|\Sigma_{XX}^{1/2} - (\Sigma_{XX}+\epsilon_n I)^{1/2}\right\| = \gamma_1^{1/2} - (\gamma_1+\epsilon_n)^{1/2} = O(\epsilon_n)$. Therefore,

$$\begin{aligned}
(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2} - I &= (\Sigma_{XX}+\epsilon_n I)^{-1/2}\{\Sigma_{XX}^{1/2} - (\Sigma_{XX}+\epsilon_n I)^{1/2}\} \\
&= O(\epsilon_n^{-1/2})O(\epsilon_n) = O(\epsilon_n^{1/2}). \qquad \text{(A.10)}
\end{aligned}$$

Finally, going back to J_4, we get

$$\begin{aligned} J_4 &= \{(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2} - I\}M_\tau\{(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2} - I\} \\ &\quad + M_\tau\{(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2} - I\} + \{(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2} - I\}M_\tau + M_\tau \\ &= O(\epsilon_n^{1/2})O(1)O(\epsilon_n^{1/2}) + O(1)O(\epsilon_n^{1/2}) + O(\epsilon_n^{1/2})O(1) + M_\tau \\ &= O(\epsilon_n^{1/2}) + M_\tau. \end{aligned}$$

Combining everything,

$$\begin{aligned} I_2 &= J_1 + J_2 + J_3 + J_4 \\ &= O_p(n^{-1}\epsilon_n^{-3}) + O_p(n^{-1/2}\epsilon_n^{-1}) + O_p(n^{-1/2}\epsilon_n^{-2}) + O(\epsilon_n^{1/2}) + M_\tau \\ &= O_p(n^{-1}\epsilon_n^{-3} + n^{-1/2}\epsilon_n^{-2}) + O(\epsilon_n^{1/2}) + M_\tau \\ &= O_p(n^{-1/2}\epsilon_n^{-2}) + O(\epsilon_n^{1/2}) + M_\tau, \end{aligned}$$

where the third equality follows from the fact that $n^{-1/2}\epsilon_n^{-1} \prec n^{-1/2}\epsilon_n^{-2} \prec 1$ and the fourth equality follows from the fact that $n^{-1}\epsilon_n^{-3} \prec n^{-1/2}\epsilon_n^{-2} \prec 1$.

Finally,

$$\begin{aligned} \widehat{M_\tau} &= I_1 + I_2 = O_p(n^{-1/2}\epsilon_n^{-1}) + O_p(n^{-1/2}\epsilon_n^{-2}) + O(\epsilon_n^{1/2}) + M_\tau \\ &= O_p(n^{-1/2}\epsilon_n^{-2}) + O(\epsilon_n^{1/2}) + M_\tau. \end{aligned}$$

A.4.5 Proof of Theorem 2.5

We know that, for any $j = 1, \ldots, d_\tau$,

$$\widehat{v}_{\tau,j} = (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}\widehat{\eta}_{\tau,j} \quad \text{and} \quad v_{\tau,j} = \Sigma_{XX}^{\dagger 1/2}\eta_{\tau,j}.$$

We will focus on $(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}\widehat{M_\tau}$ and show that is consistent for $\Sigma_{XX}^{\dagger 1/2}M_\tau$.

We use similar decomposition of $\widehat{M_\tau}$ as in Section A.4.4 to get

$$\begin{aligned}(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\widehat{M_\tau} &= (\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\widehat{B}_\tau \pm B_\tau)(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2} \\ &= (\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1}(\widehat{B}_\tau - B_\tau)(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2} \\ &\quad +(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1}B_\tau(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2} \\ &= I_1^* + I_2^*,\end{aligned}$$

where

$$\begin{aligned}I_1^* &= (\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1}(\widehat{B}_\tau - B_\tau)(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2} = O_p(\epsilon_n^{-1})O_p(n^{-1/2})O_p(\epsilon_n^{-1/2}) \\ &= O_p(n^{-1/2}\epsilon_n^{-3/2}).\end{aligned}$$

As far as I_2^*, we use the decomposition (A.9) to get,

$$I_2^* = (\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2} I_2 = (\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(J_1+J_2+J_3+J_4),$$

where

$$\begin{aligned}(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}J_1 &= O_p(\epsilon_n^{-1/2})O_p(n^{-1}\epsilon_n^{-3}) = O_p(n^{-1}\epsilon_n^{-7/2}) \\ (\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}J_2 &= O_p(\epsilon_n^{-1/2})O_p(n^{-1/2}\epsilon_n^{-1}) = O_p(n^{-1/2}\epsilon_n^{-3/2}) \\ (\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}J_3 &= O_p(\epsilon_n^{-1/2})O_p(n^{-1/2}\epsilon_n^{-2}) = O_p(n^{-1/2}\epsilon_n^{-5/2}).\end{aligned}$$

As far as $(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}J_4$, we use the following decomposition:

$$\begin{aligned}&(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}J_4 \\ =\ &(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}B_\tau(\Sigma_{XX}+\epsilon_n I)^{-1/2} \\ =\ &(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}B_\tau\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2} \\ =\ &(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}M_\tau\Sigma_{XX}^{1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2},\end{aligned}$$

where the second equality follows from part (b) of Lemma A.2 and the third equality follows from the definition of M_τ.

Then,

$$
\begin{aligned}
& (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2} J_4 \\
= \; & \{(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}(\Sigma_{XX} + \epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2} \pm \Sigma_{XX}^{\dagger 1/2}\} M_\tau \{\Sigma_{XX}^{1/2}(\Sigma_{XX} + \epsilon_n I)^{-1/2} \pm I\} \\
= \; & \{(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}(\Sigma_{XX} + \epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2} - \Sigma_{XX}^{\dagger 1/2}\} M_\tau \{\Sigma_{XX}^{1/2}(\Sigma_{XX} + \epsilon_n I)^{-1/2} - I\} \\
& + \{(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}(\Sigma_{XX} + \epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2} - \Sigma_{XX}^{\dagger 1/2}\} M_\tau \\
& + \Sigma_{XX}^{\dagger 1/2} M_\tau \{\Sigma_{XX}^{1/2}(\Sigma_{XX} + \epsilon_n I)^{-1/2} - I\} + \Sigma_{XX}^{\dagger 1/2} M_\tau \\
= \; & A_1 + A_2 + A_3 + A_4,
\end{aligned}
$$

where

$$
\begin{aligned}
A_1 &= \{(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}(\Sigma_{XX} + \epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2} - \Sigma_{XX}^{\dagger 1/2}\} M_\tau \{\Sigma_{XX}^{1/2}(\Sigma_{XX} + \epsilon_n I)^{-1/2} - I\} \\
&= \{O_p(n^{-1/2}\epsilon_n^{-1/2}) + O(\epsilon_n^{1/2})\} O(\epsilon_n^{1/2}) = O_p(n^{-1/2}) + O(\epsilon_n),
\end{aligned}
$$

$$
\begin{aligned}
A_2 &= \{(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}(\Sigma_{XX} + \epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2} - \Sigma_{XX}^{\dagger 1/2}\} M_\tau \\
&= O_p(n^{-1/2}\epsilon_n^{-1/2}) + O(\epsilon_n^{1/2}), \\
A_3 &= \Sigma_{XX}^{\dagger 1/2} M_\tau \{\Sigma_{XX}^{1/2}(\Sigma_{XX} + \epsilon_n I)^{-1/2} - I\} \\
&= O(1) O(\epsilon_n^{1/2}) = O(\epsilon_n^{1/2}), \\
A_4 &= \Sigma_{XX}^{\dagger 1/2} M_\tau.
\end{aligned}
$$

The above follow from few facts:

1. $\{(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}(\Sigma_{XX} + \epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2} - \Sigma_{XX}^{\dagger 1/2}\} M_\tau = O_p(n^{-1/2}\epsilon_n^{-1/2}) + O(\epsilon_n^{1/2})$, which is proven at the end of this section,

2. $\{\Sigma_{XX}^{1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}-I\}=O(\epsilon_n^{1/2})$, which follows from the facts that $\Sigma_{XX}^{1/2}$ and $(\Sigma_{XX}+\epsilon_n I)^{-1/2}$ commute and the relation (A.10),

3. $\Sigma_{XX}^{\dagger 1/2}M_\tau=O(1)$, which follows from the facts that Σ_{XX} is compact and M_τ is bounded (see part (b) of Lemma A.2).

Combining everything we have,

$$\begin{aligned}(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}J_4 &= A_1+A_2+A_3+A_4\\ &= O_p(n^{-1/2})+O(\epsilon_n)+O_p(n^{-1/2}\epsilon_n^{-1/2})+O(\epsilon_n^{1/2})\\ &\quad +O(\epsilon_n^{1/2})+\Sigma_{XX}^{\dagger 1/2}M_\tau\\ &= O_p(n^{-1/2}\epsilon_n^{-1/2})+O(\epsilon_n^{1/2})+\Sigma_{XX}^{\dagger 1/2}M_\tau.\end{aligned}$$

Then,

$$\begin{aligned}I_2^* &= (\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(J_1+J_2+J_3+J_4)\\ &= O_p(n^{-1}\epsilon_n^{-7/2})+O_p(n^{-1/2}\epsilon_n^{-3/2})+O_p(n^{-1/2}\epsilon_n^{-5/2})+O_p(n^{-1/2}\epsilon_n^{-1/2})\\ &\quad +O(\epsilon_n^{1/2})+\Sigma_{XX}^{\dagger 1/2}M_\tau\\ &= O_p(n^{-1}\epsilon_n^{-7/2}+n^{-1/2}\epsilon_n^{-5/2})+O(\epsilon_n^{1/2})+\Sigma_{XX}^{\dagger 1/2}M_\tau\\ &= O_p(n^{-1/2}\epsilon_n^{-5/2})+O(\epsilon_n^{1/2})+\Sigma_{XX}^{\dagger 1/2}M_\tau,\end{aligned}$$

where the last equality follows from the fact that $n^{-1}\epsilon_n^{-7/2}\prec n^{-1/2}\epsilon_n^{-5/2}\prec 1$. Finally,

$$\begin{aligned}(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\widehat{M}_\tau &= I_1^*+I_2^*\\ &= O_p(n^{-1/2}\epsilon_n^{-3/2})+O_p(n^{-1/2}\epsilon_n^{-5/2})+O(\epsilon_n^{1/2})+\Sigma_{XX}^{\dagger 1/2}M_\tau\\ &= O_p(n^{-1/2}\epsilon_n^{-5/2})+O(\epsilon_n^{1/2})+\Sigma_{XX}^{\dagger 1/2}M_\tau.\end{aligned}$$

We have shown that

$$(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}\widehat{M}_\tau - \Sigma_{XX}^{\dagger 1/2} M_\tau = O_p(n^{-1/2}\epsilon_n^{-5/2}) + O(\epsilon_n^{1/2}).$$

Then, since for $j = 1, \ldots, d_\tau$, $\widehat{\eta}_{\tau,j} = \widehat{\lambda}_j^{-1}\widehat{M}_\tau\widehat{\eta}_{\tau,j}$ and $\eta_{\tau,j} = \lambda_j^{-1}M_\tau\eta_{\tau,j}$, we have

$$\begin{aligned}
& (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}\widehat{\eta}_{\tau,j} - \Sigma_{XX}^{\dagger 1/2}\eta_{\tau,j} \\
= \; & (\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}\widehat{\lambda}_j^{-1}\widehat{M}_\tau\widehat{\eta}_{\tau,j} - \Sigma_{XX}^{\dagger 1/2}\lambda_j^{-1}M_\tau\eta_{\tau,j} \\
= \; & \{\Sigma_{XX}^{\dagger 1/2}M_\tau + O_p(n^{-1/2}\epsilon_n^{-5/2}) + O(\epsilon_n^{1/2})\}\widehat{\lambda}_j^{-1}\widehat{\eta}_{\tau,j} - \Sigma_{XX}^{\dagger 1/2}\lambda_j^{-1}M_\tau\eta_{\tau,j} \\
= \; & \Sigma_{XX}^{\dagger 1/2}M_\tau(\widehat{\lambda}_j^{-1}\widehat{\eta}_{\tau,j} - \lambda_j^{-1}\eta_{\tau,j}) + O_p(n^{-1/2}\epsilon_n^{-5/2}) + O(\epsilon_n^{1/2}).
\end{aligned}$$

From Corollary 2.1,

$$\widehat{\lambda}_j^{-1}\widehat{\eta}_{\tau,j} - \lambda_j^{-1}\eta_{\tau,j} = O_p(n^{-1/2}\epsilon_n^{-2}) + O(\epsilon_n^{1/2})$$

and therefore,

$$\begin{aligned}
(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}\widehat{\eta}_{\tau,j} - \Sigma_{XX}^{\dagger 1/2}\eta_{\tau,j} &= O_p(n^{-1/2}\epsilon_n^{-5/2}) + O(\epsilon_n^{1/2}) + O_p(n^{-1/2}\epsilon_n^{-2}) + O(\epsilon_n^{1/2}) \\
&= O_p(n^{-1/2}\epsilon_n^{-5/2}) + O(\epsilon_n^{1/2}).
\end{aligned}$$

Proof of $\{(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}(\Sigma_{XX} + \epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2} - \Sigma_{XX}^{\dagger 1/2}\}M_\tau = O_p(n^{-1/2}\epsilon_n^{-1/2}) + O(\epsilon_n^{1/2})$:

Let's first consider the term $\Sigma_{XX}^{1/2}\{(\widehat{\Sigma}_{XX} + \epsilon_n I)^{-1/2}(\Sigma_{XX} + \epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2} - \Sigma_{XX}^{\dagger 1/2}\}\Sigma_{XX}^{1/2}$

and observe that

$$
\begin{aligned}
& \Sigma_{XX}^{1/2}\{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-\Sigma_{XX}^{\dagger 1/2}\}\Sigma_{XX}^{1/2} \\
= \;& \Sigma_{XX}^{1/2}(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}-\Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2} \\
= \;& \Sigma_{XX}^{1/2}(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}-\Sigma_{XX}^{1/2} \\
= \;& \Sigma_{XX}^{1/2}\{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}-I\} \\
= \;& \Sigma_{XX}^{1/2}\{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-I\} \\
= \;& \Sigma_{XX}^{1/2}[\{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}\pm I\}\{(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}\pm I\}-I] \\
= \;& \Sigma_{XX}^{1/2}[\{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-I\}\{(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-I\} \\
& +\{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-I\}+\{(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-I\}+I-I] \\
= \;& \Sigma_{XX}^{1/2}[\{O_p(n^{-1/2}\epsilon_n^{-1/2})+O(\epsilon_n^{1/2})\}O(\epsilon_n^{1/2})+O_p(n^{-1/2}\epsilon_n^{-1/2})+O(\epsilon_n^{1/2})+O(\epsilon_n^{1/2})] \\
= \;& \Sigma_{XX}^{1/2}\{O_p(n^{-1/2}\epsilon_n^{-1/2})+O(\epsilon_n^{1/2})\},
\end{aligned}
$$

where the second equality follows from the fact that $\Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2}=\Sigma_{XX}^{1/2}$, the fourth equality follows from the fact that $\Sigma_{XX}^{1/2}$ and $(\Sigma_{XX}+\epsilon_n I)^{-1/2}$ commute, and the seventh equality follows from the fact that

$$
\begin{aligned}
& \{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-I\} \\
= \;& \{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}-(\Sigma_{XX}+\epsilon_n I)^{-1/2}\}\Sigma_{XX}^{1/2}+(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-I \\
= \;& [(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}\{(\Sigma_{XX}+\epsilon_n I)^{3/2}-(\widehat{\Sigma}_{XX}+\epsilon_n I)^{3/2}\}+\widehat{\Sigma}_{XX}-\Sigma_{XX}]\Sigma_{XX}^{1/2} \\
& +(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-I \\
= \;& \{O_p(\epsilon_n^{-1/2})O_p(n^{-1/2})+O_p(n^{-1/2})\}O(1)+O(\epsilon_n^{1/2}) \\
= \;& O_p(n^{-1/2}\epsilon_n^{-1/2})+O(\epsilon_n^{1/2}),
\end{aligned}
$$

and relation (A.10).

Therefore, we proved that

$$\begin{aligned} & \Sigma_{XX}^{1/2}\{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-\Sigma_{XX}^{\dagger 1/2}\}\Sigma_{XX}^{1/2} \\ = \; & \Sigma_{XX}^{1/2}\{O_p(n^{-1/2}\epsilon_n^{-1/2})+O(\epsilon_n^{1/2})\}, \end{aligned}$$

which implies that

$$\begin{aligned} & \{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-\Sigma_{XX}^{\dagger 1/2}\}\Sigma_{XX}^{1/2} \\ = \; & O_p(n^{-1/2}\epsilon_n^{-1/2})+O(\epsilon_n^{1/2}). \end{aligned} \tag{A.11}$$

Now, let's consider the original expression $\{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-\Sigma_{XX}^{\dagger 1/2}\}M_\tau$ and observe that

$$\begin{aligned} & \{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-\Sigma_{XX}^{\dagger 1/2}\}M_\tau \\ = \; & \{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-\Sigma_{XX}^{\dagger 1/2}\}\Sigma_{XX}^{\dagger 1/2}B_\tau\Sigma_{XX}^{\dagger 1/2} \\ = \; & \{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-\Sigma_{XX}^{\dagger 1/2}\}\Sigma_{XX}^{\dagger 1/2}\{\Sigma_{XX}(\omega_\tau\otimes\omega_\tau)\Sigma_{XX}+\dots \\ & +\Sigma_{XX}^{m+1}(\omega_\tau\otimes\omega_\tau)\Sigma_{XX}^{m+1}\}\Sigma_{XX}^{\dagger 1/2} \\ = \; & \{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-\Sigma_{XX}^{\dagger 1/2}\}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2}\Sigma_{XX}^{1/2}(\omega_\tau\otimes\omega_\tau)\Sigma_{XX}^{1/2}\Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}+\dots \\ & +\{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-\Sigma_{XX}^{\dagger 1/2}\}\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2}\Sigma_{XX}^{m+1/2}(\omega_\tau\otimes\omega_\tau)\Sigma_{XX}^{m+1/2}\Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2} \\ = \; & \{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-\Sigma_{XX}^{\dagger 1/2}\}\Sigma_{XX}^{1/2}(\omega_\tau\otimes\omega_\tau)\Sigma_{XX}^{1/2}+\dots \\ & +\{(\widehat{\Sigma}_{XX}+\epsilon_n I)^{-1/2}(\Sigma_{XX}+\epsilon_n I)^{-1/2}\Sigma_{XX}^{1/2}-\Sigma_{XX}^{\dagger 1/2}\}\Sigma_{XX}^{m+1/2}(\omega_\tau\otimes\omega_\tau)\Sigma_{XX}^{m+1/2} \\ = \; & \{O_p(n^{-1/2}\epsilon_n^{-1/2})+O(\epsilon_n^{1/2})\}O(1), \end{aligned}$$

where the second equality follows from relation (A.7), the fourth equality follows from the facts that $\Sigma_{XX}^{\dagger 1/2}\Sigma_{XX}^{1/2}$ is equal to the projection matrix to $\Sigma_{XX}^{1/2}$ and $\Sigma_{XX}^{1/2}\Sigma_{XX}^{\dagger 1/2}$ is the identity mapping from $\mathrm{ran}(\Sigma_{XX}^{1/2})$ onto $\mathrm{ran}(\Sigma_{XX}^{1/2})$, and the last equality follows from (A.11).

A.4.6 Proof of $[\widehat{\Sigma}_{XX}^{\dagger 1/2}] = G^{\dagger 1/2}\{n^{-1}G^{1/2}([X_{1:n}]Q_n[X_{1:n}]^\top)G^{1/2}\}^{\dagger 1/2}G^{1/2}$

First, we will prove that $[\widehat{\Sigma}_{X^iX^i}] = n^{-1}[X^i_{1:n}]Q_n[X^i_{1:n}]^\top G_1$. Note that,

$$\begin{aligned}\widehat{\Sigma}_{X^iX^i} &= E_n\{(X^i - E_n(X^i)) \otimes (X^i - E_n(X^i))\} \\ &= n^{-1}\sum_{u=1}^{n}(X^i_u - \bar{X}^i) \otimes (X^i_u - \bar{X}^i).\end{aligned}$$

Then,

$$\begin{aligned}[\widehat{\Sigma}_{X^iX^i}] &= n^{-1}\sum_{u=1}^{n}[X^i_u \otimes X^i_u - \bar{X}^i \otimes \bar{X}^i] \\ &= n^{-1}\sum_{u=1}^{n}[X^i_u][X^i_u]^\top G_1 - [\bar{X}^i][\bar{X}^i]^\top G_1 \\ &= n^{-1}[X^i_{1:n}][X^i_{1:n}]^\top G_1 - n^{-1}[X^i_{1:n}](1_n 1_n^\top/n)[X^i_{1:n}]^\top G_1 \\ &= n^{-1}[X^i_{1:n}]Q_n[X^i_{1:n}]^\top G_1.\end{aligned}$$

Using similar arguments, we can show that $[\widehat{\Sigma}_{XX}] = n^{-1}[X_{1:n}]Q_n[X_{1:n}]^\top G$. Then, using (A.15),

$$\begin{aligned}[\widehat{\Sigma}_{XX}^{\dagger 1/2}] &= G^{\dagger 1/2}(G^{1/2}[\widehat{\Sigma}_{XX}]G^{\dagger 1/2})^{\dagger 1/2}G^{1/2} \\ &= G^{\dagger 1/2}(G^{1/2}n^{-1}[X_{1:n}]Q_n[X_{1:n}]^\top G^{1/2})^{\dagger 1/2}G^{1/2}.\end{aligned}$$

A.5 Basics on Coordinate Representation

Let $\mathcal{H}_i$ be finite-dimensional Hilbert spaces with spanning systems $\mathcal{B}_i = \{b^i_1, \cdots, b^i_{k_i}\}$, respectively, and define the Gram matrix of $\mathcal{H}_i$, denoted by $\mathcal{G}_i$, as the matrix of inner products $\{\langle b^i_j, b^i_k\rangle_{\mathcal{H}_i} : j,k = 1,\ldots,k_i\}$, $i = 1,2$.

(i) **Coordinate representation of a function.** A function $f \in \mathcal{H}_1$ can be

written as a linear combination of the spanning system $\mathcal{B}_1$:

$$f = c_1^1 b_1^1 + \cdots + c_{k_1}^1 b_{k_1}^1$$

where $c_1^1, \cdots, c_{k_1}^1$ are constants. The vector $(c_1^1, \cdots, c_{k_1}^1)^\top$ is called the coordinate of f relative to $\mathcal{B}_1$, which can be represented by $[f]_{\mathcal{B}_1}$. That is,

$$f = \sum_{j=1}^{k_1} ([f]_{\mathcal{B}_1})_j b_j^1 = [f]_{\mathcal{B}_1}^\top b_{1:k_1}^1, \tag{A.12}$$

where $([f]_{\mathcal{B}_1})_j$ is the jth component of the vector $[f]_{\mathcal{B}_1}$ and $b_{1:k_1}^1 = (b_1^1, \cdots, b_{k_1}^1)^\top$.

(ii) **Coordinate representation of an inner product.** For $f, h \in \mathcal{H}_1$, the coordinate representation of the inner product can be expressed as

$$\begin{aligned} \langle f, h \rangle_{\mathcal{H}_1} &= \left\langle [f]_{\mathcal{B}_1}^\top b_{1:k_1}^1, [h]_{\mathcal{B}_1}^\top b_{1:k_1}^1 \right\rangle_{\mathcal{H}_1} = \left\langle \sum_{j=1}^{k_1} ([f]_{\mathcal{B}_1})_j b_j^1, \sum_{k=1}^{k_1} ([h]_{\mathcal{B}_1})_k b_k^1 \right\rangle_{\mathcal{H}_1} \\ &= \sum_{j=1}^{k_1} \sum_{k=1}^{k_1} ([f]_{\mathcal{B}_1})_j ([h]_{\mathcal{B}_1})_k \langle b_j^1, b_k^1 \rangle_{\mathcal{H}_1} = [f]_{\mathcal{B}_1}^\top \mathcal{G}_1 [h]_{\mathcal{B}_1}. \end{aligned} \tag{A.13}$$

(iii) **Coordinate representation of a linear operator.** Let $A : \mathcal{H}_1 \to \mathcal{H}_2$ be a linear operator and f a member of $\mathcal{H}_1$. Using equation (A.12), we have

$$Af = A\left(\sum_{j=1}^{k_1} ([f]_{\mathcal{B}_1})_j b_j^1 \right) = \sum_{j=1}^{k_1} ([f]_{\mathcal{B}_1})_j A b_j^1 = \sum_{j=1}^{k_1} ([f]_{\mathcal{B}_1})_j \sum_{k=1}^{k_2} ([A b_j^1]_{\mathcal{B}_2})_k b_k^2.$$

Then, the coordinate of Af relative to $\mathcal{B}_2$ is the vector $\{\sum_{j=1}^{k_1} ([f]_{\mathcal{B}_1})_j ([A b_j^1]_{\mathcal{B}_2})_k, k = 1, \ldots, k_2\}$, and the coordinates of A with respect to $\mathcal{B}_1$ and $\mathcal{B}_2$, denoted by ${}_{\mathcal{B}_2}[A]_{\mathcal{B}_1}$, is the $k_2 \times k_1$ matrix ${}_{\mathcal{B}_2}[A]_{\mathcal{B}_1} = ([A b_1^1]_{\mathcal{B}_2}, \ldots, [A b_{k_1}^1]_{\mathcal{B}_2})$. Then, we can simply write $[Af]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}[A]_{\mathcal{B}_1} [f]_{\mathcal{B}_1}$.

For more than one operator, let $A : \mathcal{H}_1 \to \mathcal{H}_2$ and $B : \mathcal{H}_2 \to \mathcal{H}_3$ be two linear operators, and let f be a member of $\mathcal{H}_1$. Then, using the same arguments as above,

$$_{\mathcal{B}_3}[BAf]_{\mathcal{B}_1} = {}_{\mathcal{B}_3}[B(Af)]_{\mathcal{B}_1} = {}_{\mathcal{B}_3}[B]_{\mathcal{B}_2}[Af]_{\mathcal{B}_1} = ({}_{\mathcal{B}_3}[B]_{\mathcal{B}_2})\,({}_{\mathcal{B}_2}[A]_{\mathcal{B}_1})\,[f]_{\mathcal{B}_1}$$

Therefore, ${}_{\mathcal{B}_3}[BA]_{\mathcal{B}_1} = ({}_{\mathcal{B}_3}[B]_{\mathcal{B}_2})\,({}_{\mathcal{B}_2}[A]_{\mathcal{B}_1})$.

(iv) **Coordinate representation of a tensor product.** For $f, g, h \in \mathcal{H}_1$ is given by

$$[(f \otimes g)h]_{\mathcal{B}_1} = [f\langle g, h\rangle_{\mathcal{H}_1}]_{\mathcal{B}_1} = [f]_{\mathcal{B}_1}[g]_{\mathcal{B}_1}^{\top}\mathcal{G}_1[h]_{\mathcal{B}_1},$$

where the last equality follows from the coordinate representation of an inner product (A.13).

Finally, a useful property that will be utilized in the following Chapter. For $V : \mathcal{H}_1 \to \mathcal{H}_1$ and $\alpha > 0$, we have ${}_{\mathcal{G}_1}[V^\alpha]_{\mathcal{G}_1} = G_1^{\dagger 1/2}(G_1^{1/2}{}_{\mathcal{G}_1}[V]_{\mathcal{G}_1}G_1^{\dagger 1/2})^\alpha G_1^{1/2}$. To show this, suppose V has spectral decomposition $\sum_{j=1}^{k_1}\delta_j(\zeta_j \otimes \zeta_j)$. Then, $V^\alpha = \sum_{j=1}^{k_1}\delta_j^\alpha(\zeta_j \otimes \zeta_j)$. By the coordinate representation of the tensor product derived above, we have

$$_{\mathcal{G}_1}[V^\alpha]_{\mathcal{G}_1} = \sum_{j=1}^{k_1}\delta_j^\alpha[\zeta_j \otimes \zeta_j]_{\mathcal{G}_1} = \left(\sum_{j=1}^{k_1}\delta_j^\alpha[\zeta_j]_{\mathcal{G}_1}[\zeta_j]_{\mathcal{G}_1}^{\top}\right)G_1.$$

Because $G_1G_1^\dagger = G_1^\dagger G_1$ and $[\zeta_j]_{\mathcal{G}_1} = G_1G_1^\dagger[\zeta_j]_{\mathcal{G}_1} = G_1^\dagger G_1[\zeta_j]_{\mathcal{G}_1}$, $j = 1, \ldots, r$, we have $[\zeta_j \otimes \zeta_j]_{\mathcal{G}_1} = [\zeta_j]_{\mathcal{G}_1}[\zeta_j]_{\mathcal{G}_1}^{\top}G_1 = G_1^{\dagger 1/2}\left(G_1^{1/2}[\zeta_j]_{\mathcal{G}_1}\right)\left(G_1^{1/2}[\zeta_j]_{\mathcal{G}_1}\right)^{\top}G_1^{1/2}$. Therefore,

$$_{\mathcal{G}_1}[V^\alpha]_{\mathcal{G}_1} = G_1^{\dagger 1/2}\left\{\sum_{j=1}^{k_1}\delta_j^\alpha\left(G_1^{1/2}[\zeta_j]_{\mathcal{G}_1}\right)\left(G_1^{1/2}[\zeta_j]_{\mathcal{G}_1}\right)^{\top}\right\}G_1^{1/2}. \tag{A.14}$$

Lastly, note that if (δ, ζ) is an eigenvalue-eigenvector pair of the linear operator

V, then $(\delta, G_1^{1/2}[\zeta]_{\mathcal{G}_1})$ is the corresponding eigenvalue-eigenvector pair of the matrix $G_1^{1/2}({}_{\mathcal{G}_1}[V]_{\mathcal{G}_1})G_1^{\dagger 1/2}$, which is symmetric because V is self-adjoint. Using this, the right hand side of (A.14) is equal to $G_1^{\dagger 1/2}(G_1^{1/2}{}_{\mathcal{G}_1}[V]_{\mathcal{G}_1}G_1^{\dagger 1/2})^{\alpha}G_1^{1/2}$. This gives us the desired result, ${}_{\mathcal{G}_1}[V^{\alpha}]_{\mathcal{G}_1} = G_1^{\dagger 1/2}(G_1^{1/2}{}_{\mathcal{G}_1}[V]_{\mathcal{G}_1}G_1^{\dagger 1/2})^{\alpha}G_1^{1/2}$. Similarly, we can prove

$$ {}_{\mathcal{G}_1}[V^{\dagger\alpha}]_{\mathcal{G}_1} = G_1^{\dagger 1/2}(G_1^{1/2}{}_{\mathcal{G}_1}[V]_{\mathcal{G}_1}G_1^{\dagger 1/2})^{\dagger\alpha}G_1^{1/2}. \tag{A.15} $$

APPENDIX B: TECHNICAL DETAILS AND PROOFS FOR CHAPTER 3

The following equivalences come from Proposition 4.6 of [50]. They were used in [66] and in subsequent papers that deal with dimension reduction under the presence of categorical predictors ([77], [67], among others). [69] also used these results when the predictor variable is functional and included the case where V_1, V_2, V_3, V_4 are random curves. For clarification purposes, the independence between a random variable V_1 and a random function $V_2 \in \mathcal{H}$ is defined as $V_1 \perp\!\!\!\perp V_2(t)$ for any $t \in T$.

Lemma B.1. *For generic random variables or random functions V_1, V_2, V_3, V_4, the following equivalences hold:*

$$\begin{aligned} & V_1 \perp\!\!\!\perp V_2|(V_3, V_4) \quad \textit{and} \quad V_1 \perp\!\!\!\perp V_4|V_3 \\ \Leftrightarrow\ & V_1 \perp\!\!\!\perp V_4|(V_2, V_3) \quad \textit{and} \quad V_1 \perp\!\!\!\perp V_2|V_3 \\ \Leftrightarrow\ & V_1 \perp\!\!\!\perp (V_2, V_4)|V_3 \end{aligned}$$

B.1 Proof of Proposition 3.1

The proof follows similar steps as the proof of Proposition 3.3 of [66] for the central subspace and of Proposition 2.1 of [67] for the τth central quantile subspace. For a generic subspace $\mathcal{S}_\tau$, the following holds

$$Y \perp\!\!\!\perp Q_\tau(Y|X, W)|(P_{\mathcal{S}_\tau}X, W) \Leftrightarrow Y \perp\!\!\!\perp Q_\tau(Y|X, W = w)|(P_{\mathcal{S}_\tau}X, W = w), \quad \forall w = 1, \ldots, c. \quad \text{(B.1)}$$

By definition, $\mathscr{S}^{(W)}_{Q_\tau(Y|X)}$ satisfies the left-hand side of (B.1) and, therefore, it also satisfies $Y \perp\!\!\!\perp Q_\tau(Y|X, W = w)|(P_{\mathscr{S}^{(W)}_{Q_\tau(Y|X)}}, W = w)$ for every $w = 1, \ldots, c$. However, since $\mathscr{S}_{Q_\tau(Y_w|X_w)}$ is the τ-FCQS for each subpopulation w, it is the minimal subspace satisfying $Y \perp\!\!\!\perp Q_\tau(Y|X, W = w)|(P_{\mathscr{S}_{Q_\tau(Y_w|X_w)}}X, W = w)$, implying that $\mathscr{S}_{Q_\tau(Y_w|X_w)} \subseteq \mathscr{S}^{(W)}_{Q_\tau(Y|X)}$ for every $w = 1, \ldots, c$. Since $\mathscr{S}^{(W)}_{Q_\tau(Y|X)}$ contains $\mathscr{S}_{Q_\tau(Y_w|X_w)}$

for every $w = 1, \ldots, c$, it also contains their direct sum, i.e.,

$$\bigoplus_{w=1}^{c} \mathscr{S}_{Q_\tau(Y_w|X_w)} \subseteq \mathscr{S}^{(W)}_{Q_\tau(Y|X)}. \tag{B.2}$$

On the other hand, since $\mathscr{S}_{Q_\tau(Y_w|X_w)} \subseteq \bigoplus_{w=1}^{c} \mathscr{S}_{Q_\tau(Y_w|X_w)}$, $w = 1, \ldots, c$, then $\bigoplus_{w=1}^{c} \mathscr{S}_{Q_\tau(Y_w|X_w)}$ satisfies the right-hand side of (B.1) and, therefore, it also satisfies $Y \perp\!\!\!\perp Q_\tau(Y|X,W)|(P_{\bigoplus_{w=1}^{c} \mathscr{S}_{Q_\tau(Y_w|X_w)}} X, W)$. Since $\mathscr{S}^{(W)}_{Q_\tau(Y|X)}$ is the τ-FCPQS, it is the minimal subspace satisfying $Y \perp\!\!\!\perp Q_\tau(Y|X,W)|(P_{\mathscr{S}^{(W)}_{Q_\tau(Y|X)}} X, W)$, implying that

$$\mathscr{S}^{(W)}_{Q_\tau(Y|X)} \subseteq \bigoplus_{w=1}^{c} \mathscr{S}_{Q_\tau(Y_w|X_w)}. \tag{B.3}$$

Therefore, (B.2) and (B.3) conclude that $\mathscr{S}^{(W)}_{Q_\tau(Y|X)} = \bigoplus_{w=1}^{c} \mathscr{S}_{Q_\tau(Y_w|X_w)}$.

B.2 Proof of Proposition 3.2

The proof follows similar steps as the proof of Proposition 1 of [77] for the central mean subspace and Proposition 2.2 of [67]) for the τth central quantile subspace. Define $\mathscr{S}_\tau = \mathscr{S}_{W|X} \bigoplus \mathscr{S}^{(W)}_{Q_\tau(Y|X)}$. By construction, $W \perp\!\!\!\perp X|P_{\mathscr{S}_\tau} X$ and $Y \perp\!\!\!\perp Q_\tau(Y|X,W)|(P_{\mathscr{S}_\tau} X, W)$.

The first conditional independence $W \perp\!\!\!\perp X|P_{\mathscr{S}_\tau} X$ implies that $W \perp\!\!\!\perp Q_\tau(Y|X)|P_{\mathscr{S}_\tau} X$, since $Q_\tau(Y|X)$ is a function of X.

The second conditional independence $Y \perp\!\!\!\perp Q_\tau(Y|X,W)|(P_{\mathscr{S}_\tau} X, W)$ implies that $Y \perp\!\!\!\perp Q_\tau(Y|X)|\ (P_{\mathscr{S}_\tau} X, W)$ since $Q_\tau(Y|X) = \sum_{w=1}^{c} Q_\tau(Y|X, W = w)P(W = w|X)$.

Use Lemma B.1 and take $V_1 = Q_\tau(Y|X)$, $V_2 = Y$, $V_3 = P_{\mathscr{S}_\tau} X$, and $V_4 = W$. Then, $V_4 \perp\!\!\!\perp V_1|V_3$ and $V_2 \perp\!\!\!\perp V_1|(V_3, V_4)$ is equivalent to $V_1 \perp\!\!\!\perp V_2|V_3$, which translates to $Y \perp\!\!\!\perp Q_\tau(Y|X)|P_{\mathscr{S}_\tau} X$. Since $\mathscr{S}_{Q_\tau(Y|X)}$ is the minimal subspace satisfying $Y \perp\!\!\!\perp Q_\tau(Y|X)|P_{\mathscr{S}_{Q_\tau(Y|X)}} X$, then $\mathscr{S}_{Q_\tau(Y|X)} \subseteq \mathscr{S}_{W|X} \bigoplus \mathscr{S}^{(W)}_{Q_\tau(Y|X)}$.

B.3 Proof of Proposition 3.3

The proof uses similar steps as the proof of Propositions 2 and 3 of [77]) for the central mean subspace and Proposition 2.3 of [67] for the τth central quantile subspace.

(i) To prove that $\mathscr{S}_{Q_\tau(Y|X)} \subseteq \mathscr{S}^{(W)}_{Q_\tau(Y|X)}$ we need to show that $Y \perp\!\!\!\perp Q_\tau(Y|X)|P_{\mathscr{S}^{(W)}_{Q_\tau(Y|X)}}X$. By the definition of $\mathscr{S}^{(W)}_{Q_\tau(Y|X)}$, we have that $Y \perp\!\!\!\perp Q_\tau(Y|X,W)|(P_{\mathscr{S}^{(W)}_{Q_\tau(Y|X)}}, W)$. Moreover, using the fact that $W \perp\!\!\!\perp Y|X$, the previous conditional independence can be re-written as

$$Y \perp\!\!\!\perp Q_\tau(Y|X)|(P_{\mathscr{S}^{(W)}_{Q_\tau(Y|X)}}, W). \tag{B.4}$$

Assume

$$W \perp\!\!\!\perp Q_\tau(Y|X)|P_{\mathscr{S}^{(W)}_{Q_\tau(Y|X)}}X. \tag{B.5}$$

Then, use Lemma B.1 with $V_1 = Q_\tau(Y|X)$, $V_2 = Y$, $V_3 = P_{\mathscr{S}^{(W)}_{Q_\tau(Y|X)}}$, and $V_4 = W$, to conclude that (B.4) and (B.5) imply $V_1 \perp\!\!\!\perp V_2|V_3$, which translates to $Y \perp\!\!\!\perp Q_\tau(Y|X)|P_{\mathscr{S}^{(W)}_{Q_\tau(Y|X)}}$.

Assume

$$W \perp\!\!\!\perp Y|P_{\mathscr{S}^{(W)}_{Q_\tau(Y|X)}}X. \tag{B.6}$$

Then, use Lemma B.1 with $V_1 = Q_\tau(Y|X)$, $V_2 = Y$, $V_3 = P_{\mathscr{S}^{(W)}_{Q_\tau(Y|X)}}X$, and $V_4 = W$, to conclude that (B.4) and (B.6) imply $V_1 \perp\!\!\!\perp V_2|V_3$, which translates to $Y \perp\!\!\!\perp Q_\tau(Y|X)|P_{\mathscr{S}^{(W)}_{Q_\tau(Y|X)}}X$.

(ii) To prove that $\mathscr{S}^{(W)}_{Q_\tau(Y|X)} \subseteq \mathscr{S}_{Q_\tau(Y|X)}$ we need to show that $Y \perp\!\!\!\perp Q_\tau(Y|X,W)|(P_{\mathscr{S}_{Q_\tau(Y|X)}}X, W)$.

By the definition of $\mathscr{S}_{Q_\tau(Y|X)}$, we have that

$$Y \perp\!\!\!\perp Q_\tau(Y|X)|P_{\mathscr{S}_{Q_\tau(Y|X)}}X. \tag{B.7}$$

Using the fact that $Q_\tau(Y|X) = Q_\tau(Y|P_{\mathscr{S}_{Q_\tau(Y|X)}}X)$ and the assumption $W \perp\!\!\!\perp Y|P_{\mathscr{S}_{Q_\tau(Y|X)}}X$, we get

$$W \perp\!\!\!\perp Y|(P_{\mathscr{S}_{Q_\tau(Y|X)}}X, Q_\tau(Y|X)). \tag{B.8}$$

Using Lemma B.1 with $V_1 = Y$, $V_2 = Q_\tau(Y|X)$, $V_3 = P_{\mathscr{S}_{Q_\tau(Y|X)}}X$, and $V_4 = W$, relations (B.7) and (B.8) imply that $V_1 \perp\!\!\!\perp V_2|(V_3, V_4)$, which translates to $Y \perp\!\!\!\perp Q_\tau(Y|X)|(P_{\mathscr{S}_{Q_\tau(Y|X)}}X, W)$. Finally, since $W \perp\!\!\!\perp Y|X$, then $Y \perp\!\!\!\perp Q_\tau(Y|X, W)|(P_{\mathscr{S}_{Q_\tau(Y|X)}}X, W)$.

B.4 Proof of Theorem 3.1

We want to prove that $\widehat{V}_\tau^{(W)} = \sum_{w=1}^{c} \sum_{j=1}^{d_\tau^{(w)}} \widehat{v}_{\tau j}^{(w)} \otimes \widehat{v}_{\tau j}^{(w)}$ is a consistent estimate of $V_\tau^{(W)} = \sum_{w=1}^{c} \sum_{j=1}^{d_\tau^{(w)}} v_{\tau j}^{(w)} \otimes v_{\tau j}^{(w)}$. Then,

$$
\begin{aligned}
\widehat{V}_\tau^{(W)} &= \sum_{w=1}^{c} \sum_{j=1}^{d_\tau^{(w)}} \widehat{v}_{\tau j}^{(w)} \otimes \widehat{v}_{\tau j}^{(w)} \\
&= \sum_{w=1}^{c} \sum_{j=1}^{d_\tau^{(w)}} (\widehat{v}_{\tau j}^{(w)} \pm v_{\tau j}^{(w)}) \otimes (\widehat{v}_{\tau j}^{(w)} \pm v_{\tau j}^{(w)}) \\
&= \sum_{w=1}^{c} \sum_{j=1}^{d_\tau^{(w)}} (\widehat{v}_{\tau j}^{(w)} - v_{\tau j}^{(w)}) \otimes (\widehat{v}_{\tau j}^{(w)} - v_{\tau j}^{(w)}) + \sum_{w=1}^{c} \sum_{j=1}^{d_\tau^{(w)}} (\widehat{v}_{\tau j}^{(w)} - v_{\tau j}^{(w)}) \otimes v_{\tau j}^{(w)} \\
&\quad + \sum_{w=1}^{c} \sum_{j=1}^{d_\tau^{(w)}} v_{\tau j}^{(w)} \otimes (\widehat{v}_{\tau j}^{(w)} - v_{\tau j}^{(w)}) + \sum_{w=1}^{c} \sum_{j=1}^{d_\tau^{(w)}} v_{\tau j}^{(w)} \otimes v_{\tau j}^{(w)} \\
&= \sum_{w=1}^{c} \{O_p(n_w^{-1/2} \epsilon_n^{-5/2}) + O(\epsilon_n^{1/2})\} \{O_p(n_w^{-1/2} \epsilon_n^{-5/2}) + O(\epsilon_n^{1/2})\} \\
&\quad + \sum_{w=1}^{c} \{O_p(n_w^{-1/2} \epsilon_n^{-5/2}) + O(\epsilon_n^{1/2})\} O(1) + \sum_{w=1}^{c} O(1) \{O_p(n_w^{-1/2} \epsilon_n^{-5/2}) + O(\epsilon_n^{1/2})\} \\
&\quad + V_\tau^{(W)} \\
&= \sum_{w=1}^{c} \{O_p(n_w^{-1} \epsilon_n^{-5}) + O(\epsilon_n) + O_p(n_w^{-1/2} \epsilon_n^{-5/2}) + O(\epsilon_n^{1/2}) + O_p(n_w^{-1/2} \epsilon_n^{-5/2}) + O(\epsilon_n^{1/2})\} \\
&\quad + V_\tau^{(W)} \\
&= \sum_{w=1}^{c} \{O_p(n_w^{-1/2} \epsilon_n^{-5/2}) + O(\epsilon_n^{1/2})\} + V_\tau^{(W)} \\
&= O_p(n^{*-1/2} \epsilon_n^{-5/2}) + O(\epsilon_n^{1/2}) + V_\tau^{(W)},
\end{aligned}
$$

where the fourth equality results from the convergence rate of $\widehat{v}_{\tau j}^{(w)}$, $j = 1, \ldots, d_\tau^{(w)}$, $w = 1, \ldots, c$, derived in Theorem **??** in Chapter 2. Finally, assumption $n^{*-1/5} \prec \epsilon_n \prec 1$ implies that $n_w^{-1/5} \prec \epsilon_n \prec 1$ for every $w = 1, \ldots, c$. Then, the sixth equality results from the fact that $n_w^{-1/5} \prec \epsilon_n \prec 1$ and the fact that $\epsilon_n^{1/2} \prec \epsilon_n \prec 1$, for $w = 1, \ldots, c$.

APPENDIX C: TECHNICAL DETAILS AND PROOFS FOR CHAPTER 4

C.1 Proof of Theorem 4.1

The proof builds upon the foundational work of [78] on minimal sufficient σ-fields in the classical setting and follows a similar approach to [53]. Define $\Pi_y = P_{Q_\tau(Y|X)|Y}(\cdot|y)$ and let $\mathbb{P} = \{\Pi_y : y \in \mathbb{R}\}$. Given that $\mathbb{P}$ is dominated by a σ-finite measure, there exists a countable subset $\mathbb{Q} = \{Q_k : k = 1, 2, \dots\} \subseteq \mathbb{P}$ such that $\mathbb{Q} \equiv \mathbb{P}$, where $\equiv$ means two families of measures dominating each other. Let Q_0 be the mixture of Q_k with positive mixing probabilities, i.e., $Q_0 = \sum_{k=1}^{\infty} c_k Q_k$, where $\{c_k : k = 1, 2, \dots\}$ is a sequence of positive numbers that sum to 1. Then, $\{Q_0\}$ dominates and is dominated by $\mathbb{Q}$, i.e., $\{Q_0\} \equiv \mathbb{Q} \equiv \mathbb{P}$. Define $\pi_y = d\Pi_y/dQ_0$ and let $\mathcal{G}_\tau$ be a sub σ-field of $\sigma(X)$. The two following statements are equivalent:

1. $Y \perp\!\!\!\perp Q_\tau(Y|X)|\mathcal{G}_\tau$

2. π_y is essentially measurable with respect to $\mathcal{G}_\tau$ for all $y \in \mathbb{R}$ modulo Q_0. In other words, for each $y \in \mathbb{R}$, there exists a $\mathcal{G}_\tau$-measurable function f_y such that $\pi_y = f_y$ almost everywhere with respect to Q_0. This property implies that the likelihood π_y, and hence the conditional distribution of $Q_\tau(Y|X)$ given Y, is fully determined by $\mathcal{G}_\tau$, saying that Y and $Q_\tau(Y|X)$ are conditionally independent given $\mathcal{G}_\tau$.

Proof of $1 \Rightarrow 2$: First, note that, for any integrable function $h(X)$,

$$E_{Q_0}\{h(X)\pi_y(X)\} = \int h(x)\pi_y(x)dQ_0 = \int h(x)d\Pi_y = E_{\Pi_y}\{h(X)\}.$$

Then, for $B \in \mathcal{F}_X$, it follows that

$$\begin{aligned} E_{Q_0}\{\pi_y(X)I_B(X)\} &= E_{\Pi_y}\{I_B(X)\} = E_{\Pi_y}[E_{\Pi_y}\{I_B(X)|\mathcal{G}_\tau\}] \\ &= E_{Q_0}[E_{\Pi_y}\{I_B(X)|\mathcal{G}_\tau\}\pi_y(X)] = E_{Q_0}\{\Pi_y(B|\mathcal{G}_\tau)\pi_y(X)\} \end{aligned} \quad \text{(C.1)}$$

Since $Y \perp\!\!\!\perp Q_\tau(Y|X)|\mathcal{G}_\tau$, $\Pi_y(B|\mathcal{G}_\tau)$ is the same for all $y \in \mathbb{R}$. That is, $\Pi_y(B|\mathcal{G}_\tau) = \Pi_{y'}(B|\mathcal{G}_\tau)$ for all $y' \in \mathbb{R}$. Therefore, $\Pi_y(B|\mathcal{G}_\tau) = Q_k(B|\mathcal{G}_\tau)$ for all k, which implies $\Pi_y(B|\mathcal{G}_\tau) = Q_0(B|\mathcal{G}_\tau)$. Therefore, the right-hand side of equation (C.1) can be written as

$$\begin{aligned} E_{Q_0}\{\Pi_y(B|\mathcal{G}_\tau)\pi_y(X)\} &= E_{Q_0}\{Q_0(B|\mathcal{G}_\tau)\pi_y(X)\} = E_{Q_0}[E_{Q_0}\{I_B(X)|\mathcal{G}_\tau\}\pi_y(X)] \\ &= E_{Q_0}[I_B(X)E_{Q_0}\{\pi_y(X)|\mathcal{G}_\tau\}], \end{aligned} \tag{C.2}$$

where the last equality follows from Proposition 2.2 of [79]. Combining (C.1) and (C.2), we have that, for all $B \in \mathcal{F}_X$,

$$E_{Q_0}\{\pi_y(X)I_B(X)\} = E_{Q_0}[I_B(X)E_{Q_0}\{\pi_y(X)|\mathcal{G}_\tau\}],$$

which implies that $\pi_y(X) = E_{Q_0}\{\pi_y(X)|\mathcal{G}_\tau\}$ a.s. Q_0.

Proof of $2 \Rightarrow 1$: For $A \in \mathcal{G}_\tau$, we have

$$\begin{aligned} E_{\Pi_y}[E_{Q_0}\{I_B(X)|\mathcal{G}_\tau\}I_A(X)] &= E_{Q_0}[E_{Q_0}\{I_B(X)|\mathcal{G}_\tau\}I_A(X)\pi_y(X)] \\ &= E_{Q_0}[I_B(X)I_A(X)E_{Q_0}\{\pi_y(X)|\mathcal{G}_\tau\}], \end{aligned} \tag{C.3}$$

where the last equality follows from Proposition 2.2. of [79]. By assertion 2, $E_{Q_0}\{\pi_y(X)|\mathcal{G}_\tau\} = \pi_y(X)$ and the right-hand side of (C.3) becomes

$$\begin{aligned} E_{Q_0}[I_B(X)I_A(X)E_{Q_0}\{\pi_y(X)|\mathcal{G}_\tau\}] &= E_{Q_0}[I_B(X)I_A(X)\pi_y(X)] = E_{\Pi_y}\{I_B(X)I_A(X)\} \\ &= \Pi_y(X \in A \cap B). \end{aligned}$$

Then, $E_{Q_0}\{I_B(X)|\mathcal{G}_\tau\} = Q_0(B|\mathcal{G}_\tau)$ is the conditional probability $\Pi_y(B|\mathcal{G}_\tau)$, which means that $\Pi_y(B|\mathcal{G}_\tau)$ does not depend on y. That is, 1 holds.

Finally, $\mathcal{G}_{Q_\tau(Y|X)}$ is a σ-field. Since π_y is essentially measurable with respect to all

τth quantile dimension reduction σ-fields for all $y \in \mathbb{R}$, it is also essentially measurable with respect to $\mathcal{G}_{Q_\tau(Y|X)}$ for all $y \in \mathbb{R}$. Consequently, $\mathcal{G}_{Q_\tau(Y|X)}$ is itself a τth quantile dimension reduction σ-field, which implies that it is also the smallest τth quantile dimension reduction σ-field. If $\mathcal{G}_\tau^*$ is another smallest τth quantile dimension reduction σ-field, then $\mathcal{G}_{Q_\tau(Y|X)} \subseteq \mathcal{G}_\tau^*$ and $\mathcal{G}_\tau^* \subseteq \mathcal{G}_{Q_\tau(Y|X)}$. □

C.2 Proof of Theorem 4.2

Under the single-index model (4.3), let $T_\tau(X) = f_\tau^*(X)$, where $\mathcal{G}_{Q_\tau(Y|X)} = \sigma[f_\tau^*(X)]$. Then,

$$\begin{aligned}
R(a_\tau, f_\tau) &= E[Q_\tau\{Y|T^{CS}(X)\} - a_\tau - f_\tau(X)]^2 \\
&= E[E[[Q_\tau\{Y|T^{CS}(X)\} - a_\tau - f_\tau(X)]^2|T_\tau(X)]] \\
&\geq E[[E[Q_\tau\{Y|T^{CS}(X)\} - a_\tau - f_\tau(X)|T_\tau(X)]]^2] \\
&= E[[E[Q_\tau\{Y|T^{CS}(X)\}|T_\tau(X)] - a_\tau - E\{f_\tau(X)|T_\tau(X)\}]^2] \\
&= E[Q_\tau\{Y|T^{CS}(X)\} - a_\tau - E\{f_\tau(X)|T_\tau(X)\}]^2,
\end{aligned}$$

where the last equality is based on the fact that $Q_\tau(Y|X) = Q_\tau\{Y|T^{CS}(X)\} = Q_\tau\{Y|T_\tau(X)\}$. This is because $\mathfrak{G}_{Q_\tau(Y|X)} \subseteq \mathfrak{G}_{Y|X}$ for every τ. Consequently,

$$E[Q_\tau\{Y|T^{CS}(X)\}|T_\tau(X)] = E[Q_\tau\{Y|T_\tau(X)\}|T_\tau(X)] = Q_\tau\{Y|T_\tau(X)\} = Q_\tau\{Y|T^{CS}(X)\}.$$

Therefore, for any $\alpha_\tau \in \mathbb{R}$, we have that $R(a_\tau, f_\tau) \geq R(a_\tau, M(f_\tau))$, where $M(f_\tau) = E\{f_\tau(X)|T_\tau(X)\}$. If there exists an $f_\tau \in \mathfrak{M}_X$ that is measurable with respect to $\sigma(T_\tau(X))$, then

$$E\{f_\tau(X)|T_\tau(X)\} = f_\tau(X),$$

and therefore, $R(a_\tau, f_\tau) = R(a_\tau, M(f_\tau))$.

Assuming $f'_\tau \in \mathfrak{M}_X$ is not measurable with respect to $\sigma(T_\tau(X))$, then $R(a_\tau, f'_\tau) > R(a_\tau, M(f_\tau))$. Since $\mathfrak{M}_X$ is a subset of $L_2(P_X)$, it implies that $M(f_\tau) \in L_2(P_X)$ for all $f_\tau \in \mathfrak{M}_X$. Therefore, for any $\alpha_\tau \in \mathbb{R}$ and since the mapping $f_\tau \to R(\alpha_\tau, f_\tau)$ is continuous in f_τ with respect to $L_2(P_X)$ norm, we can select $f''_\tau \in \mathfrak{M}_X$ such that

$$R(\alpha_\tau, f'_\tau) > R(\alpha_\tau, f''_\tau) > R(\alpha_\tau, M(f_\tau)).$$

Therefore, f'_τ cannot be the minimizer. Finally, since $R(a_\tau, f_\tau)$ is convex and has a unique minimum, then $\widetilde{f}_\tau$ is measurable with respect to $\sigma(T_\tau(X))$ and $\widetilde{f}_\tau \in \mathfrak{G}_{Q_\tau(Y|X)}$.

□

C.3 Proof of Theorem 4.3

Without loss of generality, assume that $\widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}$ is centered. Then, (4.9) reduces to

$$\widehat{\widetilde{f}}_\tau = \arg\min_{f_\tau \in \mathfrak{M}_X} \sum_{u=1}^{n} [\widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\} - f_\tau(X_u)]^2.$$

Minimizing the above objective function with respect to f_τ is equivalent with minimizing

$$\widehat{R}_n(f_\tau) = \frac{1}{2}\sum_{u=1}^{n} [\widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\} - f_\tau(X_u)]^2 - \frac{1}{2}\sum_{u=1}^{n} [\widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}]^2,$$

with respect to f_τ. Expanding the square, $\widehat{R}_n(f_\tau)$ can be expressed as

$$\begin{aligned}
\widehat{R}_n(f_\tau) &= \frac{1}{2}\sum_{u=1}^{n}[\widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}]^2 - \sum_{u=1}^{n}\widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}f_\tau(X_u) \\
&+ \frac{1}{2}\sum_{u=1}^{n} f_\tau(X_u)^2 - \frac{1}{2}\sum_{u=1}^{n}[\widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}]^2 \\
&= -\sum_{u=1}^{n}\widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}f_\tau(X_u) + \frac{1}{2}\sum_{u=1}^{n} f_\tau(X_u)^2 \\
&= -\sum_{u=1}^{n}\widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}\langle f_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X} + \frac{1}{2}\sum_{u=1}^{n}\langle f_\tau, \kappa(\cdot, X_u)\rangle^2_{\mathfrak{M}_X} \\
&= -\sum_{u=1}^{n}\widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}\langle f_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X} + \frac{n}{2}\langle f_\tau, \widehat{\Sigma}_{XX} f_\tau\rangle_{\mathfrak{M}_X},
\end{aligned}$$

where the last equality follows from the fact that

$$\begin{aligned}
\langle f_\tau, \kappa(\cdot, X_u)\rangle^2_{\mathfrak{M}_X} &= \langle f_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X}\langle f_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X} = \langle f_\tau, \kappa(\cdot, X_u)\langle\kappa(\cdot, X_u), f_\tau\rangle_{\mathfrak{M}_X}\rangle_{\mathfrak{M}_X} \\
&= \langle f_\tau, \{\kappa(\cdot, X_u)\otimes\kappa(\cdot, X_u)\}f_\tau\rangle_{\mathfrak{M}_X}
\end{aligned}$$

and $\sum_{u=1}^{n}\langle f_\tau, \{\kappa(\cdot, X_u)\otimes\kappa(\cdot, X_u)\}f_\tau\rangle_{\mathfrak{M}_X} = n\langle f_\tau, \widehat{\Sigma}_{XX} f_\tau\rangle_{\mathfrak{M}_X}$, for $\widehat{\Sigma}_{XX} = E_n\{\kappa(\cdot, X_u)\otimes \kappa(\cdot, X_u)\} = n^{-1}\sum_{u=1}^{n}\kappa(\cdot, X_u)\otimes\kappa(\cdot, X_u)$.

Let $\gamma_\tau = \sqrt{n}(f_\tau - \widetilde{f}_\tau)$, which implies that f_τ can be written as $f_\tau = \gamma_\tau/\sqrt{n} + \widetilde{f}_\tau$. Then $\widehat{R}_n(f_\tau)$ can be expressed as

$$\begin{aligned}
\widehat{R}_n(f_\tau) &= \widehat{R}_n(\gamma_\tau/\sqrt{n} + \widetilde{f}_\tau) \\
&= -\sum_{u=1}^{n}\widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}\langle\gamma_\tau/\sqrt{n} + \widetilde{f}_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X} + \frac{1}{2}\sum_{u=1}^{n}\langle\gamma_\tau/\sqrt{n} + \widetilde{f}_\tau, \kappa(\cdot, X_u)\rangle^2_{\mathfrak{M}_X} \\
&= -\frac{1}{\sqrt{n}}\sum_{u=1}^{n}\widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}\langle\gamma_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X} - \sum_{u=1}^{n}\widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}\langle\widetilde{f}_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X} \\
&+ \frac{1}{2n}\sum_{u=1}^{n}\langle\gamma_\tau, \kappa(\cdot, X_u)\rangle^2_{\mathfrak{M}_X} + \frac{1}{\sqrt{n}}\sum_{u=1}^{n}\langle\gamma_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X}\langle\widetilde{f}_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X} + \frac{1}{2}\sum_{u=1}^{n}\langle\widetilde{f}_\tau, \kappa(\cdot, X_u)\rangle^2
\end{aligned}$$

To prove that $\widehat{\widetilde{f}}_\tau$ is $\sqrt{n}$-consistent estimate of $\widetilde{f}_\tau$, we will prove that for any given

$\delta_\tau > 0$, there exists a constant C_τ such that

$$P\left\{\inf_{\|\gamma_\tau\|_{\mathfrak{M}_X} \geq C_\tau} \widehat{R}_n(\gamma_\tau/\sqrt{n} + \widetilde{f}_\tau) > \widehat{R}_n(\widetilde{f}_\tau)\right\} \geq 1 - \delta_\tau,$$

which implies that with probability at least $1 - \delta_\tau$ there exists a local minimum in the ball $\{\gamma_\tau/\sqrt{n} + \widetilde{f}_\tau : \|\gamma_\tau\|_{\mathfrak{M}_X} \leq C_\tau\}$. Since the objective function is strictly convex, the local minimum is a global minimum and $\left\|\widehat{\widetilde{f}}_\tau - \widetilde{f}_\tau\right\|_{\mathfrak{M}_X} = O_p(n^{-1/2})$.

Enough to show that $\widehat{R}_n(\gamma_\tau/\sqrt{n} + \widetilde{f}_\tau) - \widehat{R}_n(\widetilde{f}_\tau) > 0$ with probability tending to one. Observe that

$$\begin{aligned}
& \widehat{R}_n(\gamma_\tau/\sqrt{n} + \widetilde{f}_\tau) - \widehat{R}_n(\widetilde{f}_\tau) \\
= & -\frac{1}{\sqrt{n}}\sum_{u=1}^{n} \widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}\langle\gamma_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X} - \sum_{u=1}^{n} \widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}\langle\widetilde{f}_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X} \\
+ & \frac{1}{2n}\sum_{u=1}^{n}\langle\gamma_\tau, \kappa(\cdot, X_u)\rangle^2_{\mathfrak{M}_X} + \frac{1}{\sqrt{n}}\sum_{u=1}^{n}\langle\gamma_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X}\langle\widetilde{f}_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X} \\
+ & \frac{1}{2}\sum_{u=1}^{n}\langle\widetilde{f}_\tau, \kappa(\cdot, X_u)\rangle^2_{\mathfrak{M}_X} + \sum_{u=1}^{n} \widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}\langle\widetilde{f}_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X} - \frac{1}{2}\sum_{u=1}^{n}\langle\widetilde{f}_\tau, \kappa(\cdot, X_u)\rangle^2_{\mathfrak{M}_X} \\
= & \frac{1}{2n}\sum_{u=1}^{n}\langle\gamma_\tau, \kappa(\cdot, X_u)\rangle^2_{\mathfrak{M}_X} + \frac{1}{\sqrt{n}}\sum_{u=1}^{n}\langle\gamma_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X}\langle\widetilde{f}_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X} \\
- & \frac{1}{\sqrt{n}}\sum_{u=1}^{n} \widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}\langle\gamma_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X} \\
= & A_1 + A_2 + A_3
\end{aligned}$$

It is enough to show that $\widehat{R}_n(\gamma_\tau/\sqrt{n} + \widetilde{f}_\tau) - \widehat{R}_n(\widetilde{f}_\tau)$ is dominated by the first term A_1, which is positive, and that $A_2 + A_3$ is bounded. For A_1,

$$A_1 = \frac{1}{2n}\sum_{u=1}^{n}\langle\gamma_\tau, \kappa(\cdot, X_u)\rangle^2_{\mathfrak{M}_X} = \frac{1}{2}\langle\gamma_\tau, \widehat{\Sigma}_{XX}\gamma_\tau\rangle_{\mathfrak{M}_X},$$

which is positive and bounded.

For $A_2 + A_3$,

$$\begin{aligned} A_2 + A_3 &= \frac{1}{\sqrt{n}} \sum_{u=1}^{n} \langle \gamma_\tau, \kappa(\cdot, X_u) \rangle_{\mathfrak{M}_X} \langle \widetilde{f}_\tau, \kappa(\cdot, X_u) \rangle_{\mathfrak{M}_X} - \frac{1}{\sqrt{n}} \sum_{u=1}^{n} \widehat{Q}_\tau \{ Y | \widehat{T}^{CS}(X_u) \} \langle \gamma_\tau, \kappa(\cdot, X_u) \rangle_{\mathfrak{M}_X} \\ &= \frac{1}{\sqrt{n}} \sum_{u=1}^{n} [\langle \widetilde{f}_\tau, \kappa(\cdot, X_u) \rangle_{\mathfrak{M}_X} - \widehat{Q}_\tau \{ Y | \widehat{T}^{CS}(X_u) \}] \langle \gamma_\tau, \kappa(\cdot, X_u) \rangle_{\mathfrak{M}_X} \\ &= \frac{1}{\sqrt{n}} \sum_{u=1}^{n} [\langle \widetilde{f}_\tau, \kappa(\cdot, X_u) \rangle_{\mathfrak{M}_X} - \widehat{Q}_\tau \{ Y | \widehat{T}^{CS}(X_u) \}] \kappa^*(\cdot, X_u)(\gamma_\tau), \end{aligned}$$

where $\kappa^*(\cdot, X_u) : \mathfrak{M}_X \to \mathbb{R}$ is a linear operator such that $\kappa^*(\cdot, X_u)(\gamma_\tau) = \langle \gamma_\tau, \kappa(\cdot, X_u) \rangle_{\mathfrak{M}_X}$. Since $E \left\| \kappa(\cdot, X) \right\|_{\mathfrak{M}_X} < \infty$, it follows that $E \left\| \kappa^*(\cdot, X) \right\| < \infty$ and $\kappa^*(\cdot, X) = O_p(1)$. Consequently, it suffices to show that

$$\frac{1}{\sqrt{n}} \sum_{u=1}^{n} \left[\langle \widetilde{f}_\tau, \kappa(\cdot, X_u) \rangle_{\mathfrak{M}_X} - \widehat{Q}_\tau \{ Y | \widehat{T}^{CS}(X_u) \} \right] = O_p(1)$$

by showing that the second moment of this quantity is bounded.

Given that $\widehat{Q}_\tau \{ Y | \widehat{T}^{CS}(X_u) \}$ is data-dependent and non-deterministic, we define

$$\frac{1}{\sqrt{n}} \sum_{u=1}^{n} \left[\langle \widetilde{f}_\tau, \kappa(\cdot, X_u) \rangle_{\mathfrak{M}_X} - \phi_\tau \{ Y | T^{CS}(X_u) \} \right],$$

where ϕ_τ is a function in the class Φ_τ, whose value at $(Y, T^{CS}(X_u))$ can be written as $\phi_\tau \{ Y | T^{CS}(X) \}$ and satisfying $E[\phi_\tau^2 \{ Y | T^{CS}(X_u) \}] < \infty$. Since the class Φ_τ includes $Q_\tau \{ Y | T^{CS}(X_u) \}$ and also includes $\widehat{Q}_\tau \{ Y | \widehat{T}^{CS}(X_u) \}$ for sufficiently large n, almost surely, we will prove that the second moment of $n^{-1/2} \sum_{u=1}^{n} \left[\langle \widetilde{f}_\tau, \kappa(\cdot, X_u) \rangle_{\mathfrak{M}_X} - \phi_\tau \{ Y | T^{CS}(X_u) \} \right]$

is bounded, uniformly over $\phi_\tau \in \Phi_\tau$. Then,

$$\begin{aligned}
& \sup_{\phi_\tau \in \Phi_\tau} E\left[\frac{1}{\sqrt{n}}\sum_{u=1}^{n}[\langle \widetilde{f}_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X} - \phi_\tau\{Y|T^{CS}(X_u)\}]\right]^2 \\
\leq & \sup_{\phi_\tau \in \Phi_\tau}\left[\frac{1}{n}\sum_{u=1}^{n} E\left[\langle \widetilde{f}_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X} - \phi_\tau\{Y|T^{CS}(X_u)\}\right]^2\right] \\
\leq & \sup_{\phi_\tau \in \Phi_\tau}\left[\frac{1}{n}\sum_{u=1}^{n}\left[E\{\langle \widetilde{f}_\tau, \kappa(\cdot, X_u)\rangle^2_{\mathfrak{M}_X}\} + E[\phi^2_\tau\{Y|T^{CS}(X_u)\}]\right]\right] \\
= & \sup_{\phi_\tau \in \Phi_\tau}\left[E\{\langle \widetilde{f}_\tau, \widehat{\Sigma}_{XX}\widetilde{f}_\tau\rangle_{\mathfrak{M}_X}\} + \frac{1}{n}\sum_{u=1}^{n} E[\phi^2_\tau\{Y|T^{CS}(X_u)\}]\right] \\
= & \; O(1).
\end{aligned}$$

Since the result holds uniformly on ϕ_τ and since the class Φ_τ includes $\widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}$ for n large enough, almost surely, then $n^{-1/2}\sum_{u=1}^{n}[\langle \widetilde{f}_\tau, \kappa(\cdot, X_u)\rangle_{\mathfrak{M}_X} - \widehat{Q}_\tau\{Y|\widehat{T}^{CS}(X_u)\}] = O_p(1)$ and the proof is complete. □