QUANTILE REGRESSION METHODS IN FUNCTIONAL DATA ANALYSIS

by

Dengdeng Yu

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Statistics

Department of Mathematical and Statistical Sciences
University of Alberta

Abstract

In this thesis, we study the partial quantile regression methods in functional data analysis. In the first part, we propose a prediction procedure for the functional linear quantile regression model by using partial quantile covariance techniques and develop a simple partial quantile regression (SIMPQR) algorithm to efficiently extract partial quantile regression (PQR) basis for estimating functional coefficients. In the second part, we propose and implement an alternative formulation of partial quantile regression (APQR) for functional linear model by using block relaxation ideas and finite smoothing techniques. Such reformulation leads to insightful results and motivates new theory, demonstrating consistency and establishing convergence rates by applying advanced techniques from empirical process theory. In the third part, we propose and implement the generalization of PQR procedure to multidimensional functional linear model using tensor decomposition techniques. We also establish and demonstrate the corresponding asymptotic properties. In all three parts, extensive simulations and real data are investigated to show the superiority of our proposed methods, while the advantages of our proposed PQR basis are well demonstrated in various settings for functional linear quantile regression model.

Preface

Some of the research conducted for this thesis forms part of a collaboration with Dr. Linglong Kong and Dr. Ivan Mizera at the University of Alberta. The model formulating and data analysis in chapter 2, 3 and 4 are my original work, as well as the literature review in chapter 1 and concluding analysis in chapter 5.

Chapter 2 of this thesis has been published as D.Yu, L. Kong, and I. Mizera, "Partial functional linear quantile regression for neuroimaging data analysis," *Neurocomputing*, vol.195, pp.74-87, 2016. I was responsible for the model formulating and data analysis as well as the manuscript composition. Dr. Linglong Kong was the corresponding author, assisted with the data processing and was involved with concept formation and manuscript composition. Dr. Ivan Mizera assisted with the data analysis and contributed to manuscript edits.

Chapter 3 of this thesis has been submitted as D.Yu, L. Kong, and I. Mizera, "An Alternative Approach to Functional Linear Partial Quantile Regression". I was responsible for the model formulating and data analysis as well as the manuscript composition. Dr. Linglong Kong was the corresponding author and was involved with concept formation and manuscript composition. Dr. Ivan Mizera contributed to manuscript edits and assisted with the data analysis.

Chapter 4 of this thesis is currently under preparation to be submitted as D.Yu, L. Kong, and I. Mizera, "Partial quantile regression for multidimensional functional linear model". I will be responsible for modelling and data analysis as well as the manuscript composition. Dr. Linglong Kong will be the corresponding author, involved with data analysis and manuscript composition. Dr. Ivan Mizera will contribute to manuscript edits and assist with the data analysis.

Acknowledgements

I would like to thank all of those who have supported and helped me throughout my Ph.D. studies.

I am deeply indebted to the inspirational instructions, tremendous supports and invaluable guidance from my supervisors Dr. Linglong Kong and Dr. Ivan Mizera. They introduced me to functional data analysis and quantile regression, and helped me found a concrete background. This dissertation has benefited from their insights and intellectual acumen. I am privileged to have worked under the supervision of such two extraordinary academic mentors. My sincere thanks are extended to other members of my examining committee: Dr. Rohana Karunamuni, Dr. Hai Jiang and Dr. Lan Wang for their thorough review, insightful comments, and helpful suggestions. I also would like to thank Ms. Tara Schuetz for facilitating my doctoral study. My profound thankfulness and love to my fiancee, Shujing Gu, for her companionship, understanding and encouragement while I was working to complete this dissertation. I am also grateful for the support from my parents, Shicheng Yu and Suping Deng, whose generous and selfless support is crucial on my first step of pursuing the Ph.D. studies. Last but not the least, I am extremely thankful to my grandparents: Peijuan Qiu, Deling Deng, Aizhu He and Chengshao Yu. Their trust, encouragement and guidance are invaluable for my life.

This work and the conference travels during my study are supported by Dr. Linglong Kong, Dr. Ivan Mizera, Department of Mathematical and Statistical Department, Statistical Society of Canada, and Faculty of Graduate Studies and Research, for which I am grateful.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction and Overview of the Thesis

Functional data analysis (FDA) is about the analysis of information on curves, images, functions or more general objects where the primary object of observation can be viewed as a function [61]. While it has become a major branch of nonparametric statistics, it is still fast evolving as more data of larger scale and more complex structure emerge. As a popular tool, functional linear regression is often considered useful by statisticians to deal with such data [16, 29, 44, 80].

In imaging data analysis, massive functional data are observed/calculated at the same design points, such as time for functional images (e.g., PET and fMRI) and arclength for structure imaging (e.g. DTI). As an illustration, we present a smoothed functional data that we encounter in neuroimaging studies. We plot one diffusion property, called fractional anisotropy (FA), measured at 83 grid points along the midsagittal corpus callosum (CC) skeleton (Figure 1.1) from 30 randomly selected infants from the NIH Alzheimer's Disease Neuroimaging Initiative (ADNI) study. The corpus callosum (CC) is the largest fiber tract in the human brain and is a to-

pographically organized structure. It is responsible for much of the communication between the two hemispheres and connects homologous areas in the two cerebral hemispheres. Scientists are particularly interested in delineating the structure of the variability of these functional FA data and their prediction ability on mini-mental state examination (MMSE) with a set of covariates of interest, such as genetic information. MMSE is one of the most widely used screening tests on Alzheimer's Disease to provide brief and objective measures of cognitive functioning [70].

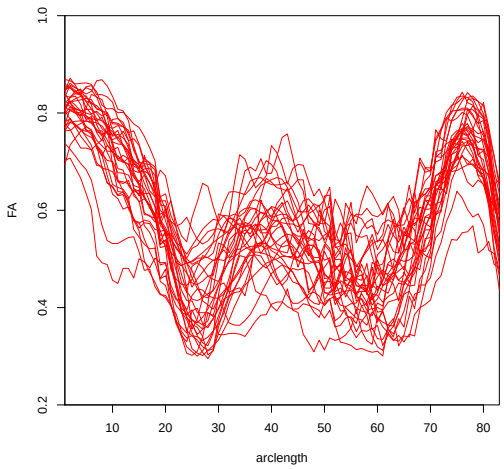


Figure 1.1: Representative functional neuroimaging data: Fractional anisotropy (FA) along the midsagittal corpus callosum (CC) skeleton from 30 randomly selected subjects from the NIH Alzheimer's Disease Neuroimaging Initiative (ADNI) study.

In order to predict the MMSE states, we can use the functional linear model:

$$y = \alpha + \mathbf{x}^T\boldsymbol{\beta} + \int_I \mathbf{z}^T(t)\boldsymbol{\gamma}(t)dt + \varepsilon, \tag{1.1}$$

where y is a scalar response, and $\mathbf{x}$ and $\mathbf{z}(t)$ are scalar and functional covariates. In the ADNI study discussed above, y is the MMSE states, $\mathbf{x}$ is the demographic information of each patients, and $\mathbf{z}(t)$ is the fractional anisotropy (FA) measured at 83

grid points along the midsagittal corpus callosum (CC) skeleton (Figure 1.1). The contributions of $\mathbf{z}(t)$ towards the variation of y is characterized by the functional coefficients $\boldsymbol{\gamma}(t)$ and change by t. For simplicity, we often assume that $t \in I = [0, 1]$. Model (1.1) is a generalization of the classical linear regression model corresponding to the case $\boldsymbol{\gamma}(t)$ is a constant. If it is not constant, the contributions of $\boldsymbol{z}(t)$ characterized by $\boldsymbol{\gamma}(t)$ change in terms of t. The model has been well studied and applied in many fields including neuroimaging data analysis [3, 36, 47, 66]. To facilitate the estimation of $\boldsymbol{\gamma}(t)$, we usually require that it satisfies certain smoothness conditions and restrict it onto a functional space. For example, we may require that its second derivative exists and that the square of $\boldsymbol{\gamma}(t)$ is integrable, that is, $\boldsymbol{\gamma}(t) \in L_2[0, 1]$. Even in such a case, the estimation is still an infinite-dimensional problem.

The common practice is to project $\boldsymbol{\gamma}(t)$ into a functional space with a finite functional basis. There are three major methods to choose the functional basis: general basis, functional principal component basis (fPC), and partial least square basis (PLS). There are various options on the selection of general basis, for example B-spline basis [8, 12], wavelet basis [79] and so on. In order to provide a good approximation of the functional coefficients, a large number of basis should be chosen. However, this may cause overfitting of the model and to remedy that various penalty methods have been proposed [16, 80]. The fPC method has been extensively studied [29, 44] where the fPC of $\mathbf{z}(t)$ serve as the basis. Its generalization to the reproducing kernel Hilbert space (RKHS) was proposed by Cai and Yuan [9] and Yuan and Cai [78] who also studied its minimax rates. Although fPC basis are more data-adapted than the general basis as they use the information of functional covariates and the formed space can explain most of the variation of $\boldsymbol{z}(t)$, it is not necessary all the fPC basis will contribute to the variation of the responses. Therefore, another appealing choice is the PLS basis which utilize both the information

of functional covariates and the responses and use the linear projects of $\mathbf{z}(t)$ which best predict the responses Delaigle and Hall [20].

An alternative to model (1.1) is the functional linear quantile regression where the conditional quantiles of the responses are modeled by a set of scalar covariates and functional covariates. In recent years, quantile regression, which was introduced by the seminal work of Koenker and Bassett [41], has been well developed and recognized in functional linear regression [37, 76]. In this thesis, we consider the functional linear quantile regression model:

$$Q_\tau(y|\mathbf{x}, \mathbf{z}(t)) = \alpha_\tau + \mathbf{x}^T\boldsymbol{\beta}_\tau + \int_0^1 \mathbf{z}^T(t)\boldsymbol{\gamma}_\tau(t)dt, \tag{1.2}$$

where $Q_\tau(y|\boldsymbol{x}, \boldsymbol{z}(t))$ is the τ-th conditional quantile of response y given scalar covariates $\boldsymbol{x}$ and functional covariate $\boldsymbol{z}(t)$ for a fixed quantile level $\tau \in (0, 1)$. As an alternative to least squares regression, the quantile regression method is more efficient and robust when the responses are non-normal, errors are heavy tailed, or outliers are present. It is also capable of dealing with the heteroscedasticity issues and providing a more complete picture of the response [38].

In the existing literature, model (1.2) has been well studied and various methods have been proposed. As in functional linear regression, to estimate functional coefficients $\boldsymbol{\gamma}_\tau(t)$ it is convenient to restrict it onto a functional space with a finite basis. Similarly, general basis like B-spline can be used to approximate the coefficient functions [11, 68]. fPC basis have also been throughly investigated with and without scalar covariate $\mathbf{x}$ while only one functional covariate presents [37, 52, 69]. However, there is no analogue to the PLS method in functional linear regression model. Therefore, none of the existing methods for model (1.2) is able to provide more efficient prediction by extracting information from the responses which

motivates us to propose a new method, partial quantile regression (PQR), to study functional linear quantile regression problem. The existing and proposed methods are summarized in Table 1.1.

	Least Squares Regression	Quantile Regression
General	Fourier, wavelet or B-spline [12, 30]	Smoothing splines [11]
Unsupervised	Functional PC [8, 29]	Functional PC [37, 52, 69]
Supervised	Partial Least Square [20, 60]	Partial Quantile Regression [proposed]

Table 1.1: The summary of existing and proposed methods in functional linear regression.

In Chapter 2, we propose a prediction procedure for the functional linear quantile regression model (1.2) by using partial quantile covariance techniques and develop an algorithm inspired by simple partial linear regression, SIMPLS [17], to efficiently extract partial quantile regression (PQR) basis for estimating functional coefficients. We further extend our partial quantile covariance techniques to functional composite quantile regression (CQR) [87] by defining partial composite quantile covariance. The major contributions of this chapter can be summarized as follows. We first define partial quantile covariance between two scalar variables through linear quantile regression. Motivated by extracting PLS basis in functional linear regression, we found PQR basis by sequentially maximizing the partial quantile covariance between the response and projections of functional covariates. In order to efficiently extract PQR basis, we develop a simple partial quantile regression (SIMPQR) algorithm analogue to SIMPLS. Under the homoscedasticity assumption, we

extend our techniques to partial composite quantile covariance and use it to find the partial composite quantile regression (PCQR) basis. The SIMPQR algorithm is then modified to obtain the SIMPCQR algorithm.

However, although the functional PQR approach appears to be an enticing choice alternative to the principal component method, there are certain limitations to uncovering the asymptotic properties mainly because of its iterative nature and the non-differentiability of the quantile loss functions. The difficulty due to the iterative formulation used to exist for functional PLS too. To deal with it, an alternative but equivalent PLS formulation (APLS) was proposed by Delaigle and Hall [20]. Based on the fact that there exists an equivalence between fPC space and functional APLS space, it is then possible to verify the consistency and convergence rates. Unfortunately, such equivalence does not exist between PQR and fPC spaces because of the non-additivity of conditional quantiles. In addition, the non-differentiability of the quantile loss function $\rho_\tau(\cdot)$ can prevent many methods and properties being directly applied in quantile regression context [75, 81].

To address these problems, in Chapter 3, we firstly propose a smoothing approximation for the quantile loss function by applying the finite smoothing techniques [13, 54]. The approximation function should uniformly converge towards the quantile loss function so that the minimizer of the former converges to the minimizer of the latter in a compact set. Replacing ρ_τ by such smoothing approximation, the quantile objective function becomes differentiable. Then for a given fixed K, namely the number of PQR basis, the original PQR formulation can be modified according to the block relaxation ideas [19] which updates and obtains the basis as a "block" instead of one by one sequentially. The value of K can be chosen using BIC or cross validation (CV) as in choosing the number of fPC basis adapted by Kato and other authors [37, 52, 69]. Such modification provides an alternative to

the original formulation for PQR (APQR) basis which leads to insightful results and motivates new theory. In particular, we can demonstrate consistency and establish convergence rates by applying advanced techniques from empirical processes theory (for more background of empirical processes, see van der Vaart [72]).

The functional covariate $\mathbf{z}$ can also be taken as a multidimensional function, i.e. its input is made up of multiple variables. In particular, without loss of generality, we let $\mathbf{t} \in [0,1]^q$ where $q \geq 2$ and $z(\mathbf{t}) \in \mathbf{R}$. Suppose that the multidimensional function covariates and coefficients, namely $z(\mathbf{t})$ and $\gamma(\mathbf{t}) \in \mathbf{R}$ with $\mathbf{t} \in [0,1]^q$. In particular, for a scalar response, a q-dimensional functional linear regression (qD-FLR) model is of the form:

$$y = \alpha + \boldsymbol{x}^T\boldsymbol{\beta} + \int_0^1 \cdots \int_0^1 z(t_1, \ldots, t_q)\gamma(t_1, \ldots, t_q)dt_q \cdots dt_1 + \varepsilon, \tag{1.3}$$

where α is the intercept, $\boldsymbol{\beta}$ is a p-dimensional vector of scalar coefficients, $\gamma(t_1, \ldots, t_q)$ is a q-dimensional functional coefficient, and ε is an error term usually with zero mean and finite variance.

Under the context of quantile regression, analogue to model (4.3), and for a given quantile level $\tau \in (0,1)$, the q-dimensional functional linear quantile regression (qD-FLQR) model is defined as:

$$Q_{y|x,z}(\tau|x,z) = \alpha_\tau + \mathbf{x}^T\boldsymbol{\beta}_\tau + \int_0^1 \cdots \int_0^1 z(t_1, \ldots, t_q)\gamma_\tau(t_1, \ldots, t_q)dt_q \cdots dt_1, \tag{1.4}$$

where $Q_{y|\mathbf{x},z}(\tau|x,z)$ is τ-th conditional quantile of y given scalar covariates $\mathbf{x}$ and multidimensional functional z.

Having $z(\mathbf{t})$ properly discretized, we can use $\mathbf{Z} \in \mathbf{R}^{I_1 \times \cdots \times I_q}$, a q-dimensional array, to represent $z(\mathbf{t})$, where the j-th element of $\mathbf{t}$ is observed at $0 = t_1 < \cdots <$

$t_{I_j} = 1$. Such data of q-dimensional array, also known as order-q tensor, is quite common in medical imaging. A notable example is magnetic resonance imaging (MRI) data where the anatomical MRI images can be observed as a matrix (order-2 tensor) of size 256 by 256. To predict certain clinical outcomes using tensor, a naive attempt to re-arrange $\mathbf{Z}$ into a vector then perform regression. However such practice is evidently unsatisfactory. First, the re-arranged (vectorized) vector of image covariates is of size $256^2 = 65,536$, implicitly requiring large number of regression parameters. Both computational cost and theoretical properties can be severely compromised due to such ultra-high dimensional setting. Furthermore, the vectorized $\mathbf{Z}$ loses information about the original data structure, so the regression model based on it would be lack of efficiency and hard to interpret.

For an order-q random tensor $\mathbf{Z} \in \mathbf{R}^{I_1 \times \cdots \times I_q}$, in order to achieve a similar decomposition as PC decomposition for vectors, namely order-1 tensors, Lu et al. [51] proposed a framework to conduct multilinear PC (MPC) decomposition. Its intention was to project the order-q tensor into a lower-dimensional space spanned by the product of a few feature vectors. In general, there are two ways to decompose a tensor: CANDECOMP/PARAFAC (CP) decomposition and Tucker decomposition [42].

In Chapter 4, we propose and implement the generalization of PQR procedure to multidimensional functional linear model (1.3) using tensor decomposition techniques. We also establish and demonstrate the corresponding asymptotic properties by applying advanced techniques from empirical process theory [72].

Chapter 2

Partial Functional Linear Quantile Regression for Neuroimaging Data Analysis

In this chapter, we propose a prediction procedure for the functional linear quantile regression model by using partial quantile covariance techniques and develop a simple partial quantile regression (SIMPQR) algorithm to efficiently extract partial quantile regression (PQR) basis for estimating functional coefficients. We further extend our partial quantile covariance techniques to functional composite quantile regression (CQR) defining partial composite quantile covariance. There are three major contributions. (1) We define partial quantile covariance between two scalar variables through linear quantile regression. We compute PQR basis by sequentially maximizing the partial quantile covariance between the response and projections of functional covariates. (2) In order to efficiently extract PQR basis, we develop a SIMPQR algorithm analogous to simple partial least squares (SIMPLS). (3) Under the homoscedasticity assumption, we extend our techniques to partial compos-

ite quantile covariance and use it to find the partial composite quantile regression (PCQR) basis. The SIMPQR algorithm is then modified to obtain the SIMPCQR algorithm. Two simulation studies show the superiority of our proposed methods. Two real data from ADHD-200 sample and ADNI are analyzed using our proposed methods.

2.1 Introduction

Nowadays, there is great need in the analysis of complex neuroimaging data obtained from various cross-sectional and clustered neuroimaging studies. These neuroimaging studies are essential to advancing our understanding of the neural development of neuropsychiatric and neurodegenerative disorders, substance use disorders, the normal brain and the interactive effects of environmental and genetic factors on brain structure and function. Such large imaging studies include the ADNI (Alzheimer's Disease Neuroimaging Initiative), the longitudinal magnetic resonance imaging (MRI) study of schizophrenia, autism, and attention deficit hyperactivity disorder (ADHD), the NIH human connectome project, among many others. Neuroimaging studies usually collect structural, neurochemical, and functional images over both time and space [23, 25, 58]. These structural, neurochemical, and functional imaging modalities include computed axial tomography (CT), diffusion tensor imaging (DTI), functional magnetic resonance imaging (fMRI), magnetic resonance imaging (MRI), magnetic resonance spectroscopy (MRS), positron emission tomography (PET), single photon emission tomography (SPECT), electroencephalography (EEG), and magnetoencephalography (MEG), among many others. For instance, by using anatomical MRI, various measures of the morphology of the cortical and subcortical structures (e.g., hippocampus) are extracted to

understand neuroanatomical differences in brain structure across different populations [22, 65]. In DTI, various diffusion properties and fiber tracts are extracted for quantitative assessment of anatomical connectivity across different populations [5, 84–86]. Functional images, such as resting-state functional MRI (rsfMRI), have been widely used in behavioral and cognitive neuroscience to understand functional segregation and integration of different brain regions across different populations [34, 59].

A common feature of many imaging techniques is that massive functional data are observed/calculated at the same design points, such as time for functional images (e.g., PET and fMRI) and arclength for structure imaging (e.g. DTI). As an illustration, we present two smoothed functional data that we encounter in neuroimaging studies. First, we consider the BOLD rsfMRI signal, which is based on hemodynamic responses secondary to resting-state. We plot the estimated hemodynamic response functions (HRF) with 172 time courses from 20 randomly selected children at a selected region of interest (ROI) of Anatomical Automatic Labeling (AAL) atlas [71] from the New York University (NYU) Child Study Center from the ADHD-200 Sample Initiative Project. Although the canonical form of the HRF is often used, when applying rsfMRI in a clinical population with possibly altered hemodynamic responses (Figure 2.1 (a)), using the subject's own HRF in rsfMRI data analysis may be advantageous because HRF variability is greater across subjects than across brain regions within a subject [2, 50]. We are particularly interested in delineating the structure of the variability of the HRF and their capacity of predicting ADHD index with a set of covariates of interest, such as diagnostic group [49]. Secondly, we plot one diffusion property, called fractional anisotropy (FA), measured at 83 grid points along the midsagittal corpus callosum (CC) skeleton (Figure 2.1 (b)) from 30 randomly selected infants from the NIH Alzheimer's

Disease Neuroimaging Initiative (ADNI) study. The corpus callosum (CC) is the largest fiber tract in the human brain and is a topographically organized structure. It is responsible for much of the communication between the two hemispheres and connects homologous areas in the two cerebral hemispheres. Scientists are particularly interested in delineating the structure of the variability of these functional FA data and their prediction ability on mini-mental state examination (MMSE) with a set of covariates of interest, such as genetic information. MMSE is one of the most widely used screening tests on Alzheimer's Disease to provide brief and objective measures of cognitive functioning [70]. We will systematically investigate these two prediction problems using functional imaging data over time or space in Section 2.7 after we develop our methodology.

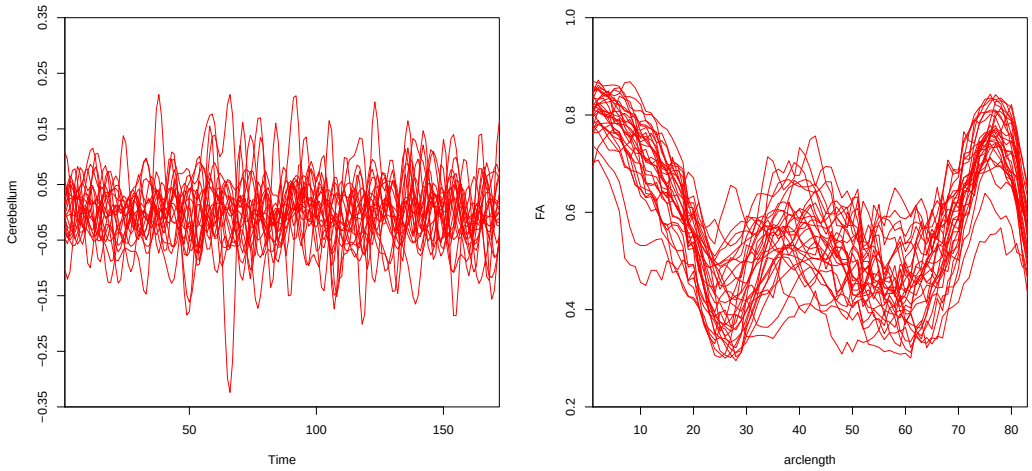


Figure 2.1: Representative functional neuroimaging data: (Left) the estimated hemodynamic response functions (HRF) corresponding to resting-state from 20 children at NYU from the ADHD-200 Sample Initiative Project and (Right) fractional anisotropy (FA) along the midsagittal corpus callosum (CC) skeleton from 30 randomly selected subjects from the NIH Alzheimer's Disease Neuroimaging Initiative (ADNI) study.

A functional linear regression model, where the responses such as the neuro-

logical or clinical outcomes (e.g. ADHD index or MMSE) are modeled by a set of scalar covariates and functional covariates of interest (e.g. HRF along time courses or FA along arclength), is a powerful statistical tool for addressing these scientific questions [26, 27, 84, 85]. In particular, denoting the neurological or clinical outcome of the i-th subject by y_i, $i = 1, \ldots, n$, the functional linear regression model is of the form

$$y_i = \alpha + \mathbf{x}_i^T \boldsymbol{\beta} + \int_0^1 \mathbf{z}_i^T(t)\boldsymbol{\gamma}(t)dt + \epsilon_i, \tag{2.1}$$

where α is the intercept, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^T$ is a $p \times 1$ vector of coefficients, $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})^T$ is a $p \times 1$ vector of scalar covariates of interest, $\boldsymbol{\gamma}(t) = (\gamma_1(t), \cdots, \gamma_q(t))^T$ is a $q \times 1$ vector of coefficient functions of t, $\mathbf{z}_i(t) = (z_{i1}(t), \cdots, z_{iq}(t))^T$ is a $q \times 1$ vector of functional covariates, and ϵ_i is a random error. It is usually assumed that ϵ_i is independent and identical copy of normal distribution with zero mean and variance σ^2. For simplicity, we let $t \in [0, 1]$. Model (2.1) is a generalization of the classical linear regression model corresponding to the case $\boldsymbol{\gamma}(t)$ is a constant. If it is not constant, the contributions of $\mathbf{z}_i(t)$ characterized by $\boldsymbol{\gamma}(t)$ change in terms of t. The model has been well studied and applied in many fields including neuroimaging data analysis [3, 36, 47, 66]. To facilitate the estimation of $\boldsymbol{\gamma}(t)$, we usually require that it satisfies certain smoothness conditions and restrict it onto a functional space. For example, we may require that its second derivative exists and that the square of $\boldsymbol{\gamma}(t)$ is integrable, that is, $\boldsymbol{\gamma}(t) \in L_2[0, 1]$. Even in such a case, the estimation is still an infinite-dimensional problem.

The common practice is to project $\boldsymbol{\gamma}(t)$ into a functional space with a finite functional basis. There are three major methods to choose the functional basis: general basis, functional principal component basis (fPC), and partial least square basis (PLS). There are various options on the selection of general basis, for exam-

ple B-spline basis [8, 12], wavelet basis [79] and so on. In order to provide a good approximation of the functional coefficients, a large number of basis should be chosen. However, this may cause overfitting of the model and to remedy that various penalty methods have been proposed [16, 80]. The fPC method has been extensively studied [29, 44] where the fPC of $\mathbf{z}_i(t)$ serve as the basis. Its generalization to the reproducing kernel Hilbert space (RKHS) was proposed by Cai and Yuan [9] and Yuan and Cai [78] who also studied its minimax rates. Although fPC basis are more data-adapted than the general basis as they use the information of functional covariates and the formed space can explain most of the variation of $\mathbf{z}_i(t)$, it is not necessary all the fPC basis will contribute to the variation of the responses. Therefore, another appealing choice is the PLS basis which use both the information of functional covariates and the responses. The PLS basis use the linear projects of $\mathbf{z}_i(t)$ which best predict the responses [20].

An alternative to model (2.1) is the functional linear quantile regression where the conditional quantiles of the responses are modeled by a set of scalar covariates and functional covariates. There are at least three advantages to use conditional quantiles instead of conditional means. First, quantile regression, in particular median regression, provides an alternative and complement to mean regression while being resistant to outliers in responses. It is more efficient than mean regression when the errors follow a distribution with heavy tails. Second, quantile regression is capable of dealing with heteroscedasticity, the situation when variances depend on some covariates. More importantly, quantile regression can give a more complete picture on how the responses are affected by covariates: for example, some tail behaviors of the responses conditional on covariates. For more background on quantile regression, see the monograph of Koenker [38]. In our case, we consider

functional linear quantile regression: for given $\tau \in (0, 1)$,

$$Q_\tau\left(y_i|\mathbf{x}_i, \mathbf{z}_i(t)\right) = \alpha_\tau + \mathbf{x}_i^T\boldsymbol{\beta}_\tau + \int_0^1 \mathbf{z}_i^T(t)\boldsymbol{\gamma}_\tau(t)dt, \tag{2.2}$$

where $Q_\tau\left(y_i|\mathbf{x}_i, \mathbf{z}_i(t)\right)$ is the τ-th conditional quantile of y_i given covariates $\mathbf{x}_i$ and $\mathbf{z}_i(t)$, α_τ is the intercept, $\boldsymbol{\beta}_\tau = (\beta_{1\tau}, \cdots, \beta_{p\tau})^T$ is a $p \times 1$ vector of coefficients and $\boldsymbol{\gamma}_\tau(t) = (\gamma_{1\tau}(t), \cdots, \gamma_{q\tau}(t))^T$ is a $q \times 1$ vector of coefficient functions. In the existing literature, model (2.2) has been well studied and various methods have been proposed. As in functional linear regression, to estimate functional coefficients $\gamma_\tau(t)$ it is convenient to restrict it in a functional space with a finite basis. Similarly, general basis like B-spline can be used to approximate the coefficient functions [11, 68]. fPC basis have also been thoroughly investigated with and without scalar covariate $\mathbf{x}_i$ while only one functional covariate presents [37, 52, 69]. However, there is no analogue to the PLS basis method in functional linear regression model. Therefore, none of the existing methods for model (2.2) is able to provide more efficient prediction by extracting information from the responses.

In this chapter, we propose a prediction procedure for the functional linear quantile regression model (2.2) by using partial quantile covariance techniques and develop an algorithm inspired by simple partial linear regression, SIMPLS [17], to efficiently extract partial quantile regression (PQR) basis for estimating functional coefficients. We further extend our partial quantile covariance techniques to functional composite quantile regression (CQR) [87] by defining partial composite quantile covariance. The major contributions of this chapter can be summarized as follows. We first define partial quantile covariance between two scalar variables through linear quantile regression. Motivated by extracting PLS basis in functional linear regression, we found PQR basis by sequentially maximizing the partial quan-

tile covariance between the response and projections of functional covariates. In order to efficiently extract PQR basis, we develop a simple partial quantile regression (SIMPQR) algorithm analogue to SIMPLS. Under the homoscedasticity assumption, we extend our techniques to partial composite quantile covariance and use it to find the partial composite quantile regression (PCQR) basis. The SIMPQR algorithm is then modified to obtain the SIMPCQR algorithm.

The rest of this chapter is organized as follows. In Section 2.2, we define partial quantile covariance and describe how to use it to extract PQR basis in functional linear quantile regression model. In Section 2.3, we develop the SIMPQR algorithm and discuss its properties. We discuss how to calculate the PCQR basis by using partial composite quantile covariance and propose the SIMPCQR algorithm in Section 2.4. Two sets of simulation studies are presented in Section 2.5 with the known ground truth to examine the finite sample performance of our proposed methodology. In Section 2.6, we use PQR and PCQR to predict ADHD index and MMSE using data from NYU site from ADHD-200 sample and ADNI, respectively. Some discussions and future research directions are given in Section 2.7.

2.2 Partial Functional Linear Quantile Regression

In model (2.2), we assume without loss of generality that $t \in [0, 1]$ and restrict the functional coefficients $\boldsymbol{\gamma}_\tau(t) \in L_2[0, 1]$. For simplicity, we assume $q = 1$, that is, we only consider one functional covariate. The extension of our methodology to more functional covariates is straightforward. The estimation $\boldsymbol{\gamma}_\tau(t)$ is in general a difficult question as it lies in an infinite-dimensional space. However, if it can be well approximated in a finite element space, say, $H[0, 1]$, the solution for the model (2.2) can be found. Let $b_{k\tau}(t)$, $k = 1, \ldots, K$ be a basis of $H[0, 1]$ and $\boldsymbol{\gamma}_\tau(t) =$

$\sum_{k=1}^{K} \boldsymbol{\gamma}_{k\tau} b_{k\tau}(t)$. Model (2.2) can be then rewritten as

$$Q_\tau\left(y_i|\mathbf{x}_i, \mathbf{z}_i(t)\right) = \alpha_\tau + \mathbf{x}_i^T \boldsymbol{\beta}_\tau + \sum_{k=1}^{K} \mathbf{z}_{ki} \boldsymbol{\gamma}_{k\tau}, \tag{2.3}$$

where $\mathbf{z}_{ki} = \int_0^1 \mathbf{z}_i(t) b_{k\tau}(t) dt$. Model (2.3) is simply a linear quantile regression problem, which is essentially a linear programing problem; its solutions can be obtained by many algorithms—for example, the simplex method [6], the interior point method [38], the MM algorithm [35] and many others, already implemented in various statistical softwares like `quantreg` in `R` [39].

In the literature, there are many methods devoted to find the crucial basis functions in model (2.3). The general basis B-spline was proposed and studied by Cardot and others [11, 68]. In various models, fPC basis has also been studied [37, 52, 69]. However, neither basis does use information of the responses and hence they are less efficient to do prediction. In this section, as the motivation of our proposal, we first review the PLS basis in model (2.1) where both information of the functional covariates and the responses are used to choose the basis functions. Then we propose our methodology to choose basis for model (2.2), namely, partial quantile regression (PQR) basis.

In functional linear regression model (2.1), the first PLS basis is chosen to be

$$b(t) = \arg_{b(t)} \min_{\alpha,\beta,b(t)} \sum_{i=1}^{n} \left(y_i - \alpha - \mathbf{x}_i^T \boldsymbol{\beta} - \int_0^1 \mathbf{z}_i(t) b(t) dt \right)^2, \tag{2.4}$$

which is the analogue to the partial least square regression in multivariate analysis. The subsequent basis is chosen by iteratively using (2.4) after taking account of and subtracting the information from previous basis. For more details, see Delaigle and Hall [20]. The essential idea of criteria (2.4) is to find a direction $b(t)$ so that the

projection of $\mathbf{z}(t)$ on it explains as much as possible the variation of the response after adjusting some covariates. Therefore, as shown in Delaigle and Hall [20], it is equivalent to find a basis $b(t)$ such that the partial covariance

$$COV\left(Y - \alpha - \mathbf{X}\boldsymbol{\beta}, \int_0^1 \mathbf{Z}(t)b(t)dt\right) \tag{2.5}$$

is maximized, where $Y = (y_1, \ldots, y_n)^T$, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ and $\mathbf{Z}(t) = (\mathbf{z}_1(t), \ldots, \mathbf{z}_n(t))^T$. Based on this equation, Delaigle and Hall [20] found an equivalent space with the same dimension as the PLS space and established the corresponding estimation and precision consistency for PLS methods.

The parameters in model (2.2) are estimated by solving

$$\min_{\alpha,\beta,b(t)} \sum_{i=1}^{n} \rho_\tau\left(y_i - \alpha - \mathbf{x}_i^T\boldsymbol{\beta} - \int_0^1 \mathbf{z}_i(t)b(t)dt\right), \tag{2.6}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the quantile loss function [38] with I as the indicator function. When $\tau = 0.5$, the loss is $\rho_\tau(u) = |u|/2$ and the results is then the median, or least absolute deviation (LAD) regression. To adapt to the idea of PLS basis, that is, to find a direction $b_\tau(t)$ so that the projection of $\mathbf{z}(t)$ on it contributes as much as possible to predict the quantile of the response after adjusting some covariates, we first propose the concepts of quantile covariance (QC) and partial quantile covariance (PQC). For given $\tau \in (0, 1)$ and a random variable X, the partial quantile covariance $COV_{qr}(Y, Z)$ between two random variables Y and Z is of the form

$$COV_{qr}(Y, Z) = \arg_{\gamma_\tau} \inf_{\alpha,\beta_\tau,\gamma_\tau} E\left(\rho_\tau\left(Y - \alpha - \beta_\tau X - \gamma_\tau Z\right)\right), \tag{2.7}$$

where we first normalize Z to have mean zero and variance one. If there is no X, then $COV_{qr}(Y, Z)$ is quantile covariance between Y and Z. The quantile covariance

measures the contribution of Z to the τ-th quantile of Y. It was first proposed and studied by Dodge and Whittaker [21] in the context of partial quantile regression. Li et al. [45] proposed a similar concept of quantile correlation and used it to study quantile autoregressive model.

To find the partial quantile regression basis (PQR), similar to that of PLS to maximize the covariance we propose to compute the $b_\tau(t)$ by maximizing

$$COV_{qr}\left(Y - \alpha_\tau - \mathbf{X}\boldsymbol{\beta}_\tau, \int_0^1 \mathbf{Z}(t)b_\tau(t)dt\right). \tag{2.8}$$

The subsequent basis is computed by iteratively maximizing (2.8) after taking account of and subtracting the information of the previous basis. Let $\mathbf{Z}_k = \int_0^1 \mathbf{Z}(t)b_{k\tau}(t)dt$, where $b_{k\tau}(t)$ is the k-th PQR basis. Denote $\mathbf{Z}_{(k+1)}(t)$ as $\mathbf{Z}(t)$ after subtracting the information from the first k basis. Then the $(k+1)$-th basis $b_{(k+1)\tau}(t)$ is obtained by maximizing the partial quantile covariance

$$COV_{qr}\left(Y - \alpha_\tau - \mathbf{X}\boldsymbol{\beta}_\tau - \sum_{j=1}^{k} \mathbf{Z}_j\boldsymbol{\gamma}_{j\tau}, \int_0^1 \mathbf{Z}_{(k+1)}(t)b_\tau(t)dt\right). \tag{2.9}$$

We will discuss detailed algorithms in the next section. Once we find an adequate number, K of functional basis elements, we have the approximation model (2.3), where the parameters are obtained by minimizing

$$\sum_{i=1}^{n} \rho_\tau\left(y_i - \alpha_\tau - \mathbf{x}_i^T\boldsymbol{\beta}_\tau - \sum_{k=1}^{K} \mathbf{z}_{ki}\boldsymbol{\gamma}_{k\tau}\right). \tag{2.10}$$

The number of PQR basis can be chosen using BIC or cross validation (CV) as in choosing the number of fPC basis adapted by Kato and other authors [37, 52, 69].

2.3 SIMPQR Algorithm

In this section, we propose a simple partial quantile regression (SIMPQR) algorithm to iteratively extract the PQR basis from the functional covariates $\mathbf{z}(t)$. Similar algorithm has been studied by Dodge and Whittaker [21] in partial quantile regression with multiple covariates. It is parallel to the SIMPLS for partial least square regression [17]. The motivation is to subsequently maximize (2.8) after accounting and subtracting the information of the previous basis. To simplify the description of the SIMPQR algorithm, we will drop the scalar covariates $\mathbf{x}$ in model (2.9) in this section. Let $0 < t_1 < \cdots < t_m < 1$ denote the discretized sample points for the functional covariates and we assume they are equally spaced. Recall that we set $q = 1$ and we focus on only one functional covariate $\mathbf{z}(t)$. The SIMPQR algorithm is described as follows.

1. *Initialization:* Normalize $\mathbf{z}_i(t_j)$ for each j so that it has mean zero and variance one.

2. *Repeat:*

   (a) Compute a functional basis $b_\tau(t_j) = COV_{qr}(Y, \mathbf{Z}(t_j))$ for $j = 1, \ldots, m$ and rescale it to have $\sum b_\tau^2(t_j) = 1$.

   (b) Project $\mathbf{z}_i(t_j)$ onto the basis $(b_\tau(t_1), \ldots, b_\tau(t_m))^T$ to obtain $\mathbf{z}_i = \sum \mathbf{z}_i(t_j) b_\tau(t_j)$. Denote $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)^T$ as the projections for each subjects.

   (c) Predict $\mathbf{z}_i(j)$ by using simple linear regression with the projection $\mathbf{z}_i$ as the covariate, denoting the result by $\hat{\mathbf{z}}_i(j)$.

   (d) Subtract the information from the projection $\mathbf{z}_i$ by replacing $\mathbf{z}_i(j)$ by their residuals $\mathbf{z}_i(j) - \hat{\mathbf{z}}_i(j)$.

3. *Stop:* Check stopping criterion and retain the projections $\mathbf{Z}_1, \ldots, \mathbf{Z}_K$.

4. *Model:* Fit the model (2.3) by minimizing equation (2.10).

Algorithm 1: Simple Partial Quantile Regression (SIMPQR)

The SIMPQR algorithm follows the same line of SIMPLS with the covariance being replaced by quantile covariance. The nature of our proposed quantile covariance implies that it is not necessary to adjust the response Y each time after a new basis is obtained. The resulting functional basis is orthogonal to each other due to the prediction in step 2.3. However, it is worth noting that due to the nonadditivity of conditional quantiles, we need to fit model (2.3) after all the basis elements are picked out, instead of estimating the coefficients of each basis projections once they are chosen.

2.4 Partial Functional Linear Composite Quantile Regression

Despite the success of quantile regression (QR), its relative efficiency to the least square regression can be arbitrarily small [38, 87]. Composite quantile regression (CQR) proposed by Zou and Yuan [87] inherits some good properties of QR and is capable of providing more efficient estimators under certain conditions. Given two random variables X and Y and quantile level set $0 < \tau_1 < \cdots < \tau_L < 1$, the CQR parameters $(\alpha_{\tau_1}, \ldots, \alpha_{\tau_L}, \beta)$ are defined as

$$\inf_{\alpha_{\tau_1},\ldots,\alpha_{\tau_L},\beta} E \sum_{l=1}^{L} \left(\rho_{\tau_l}\left(Y - \alpha_{\tau_l} - \beta X\right)\right), \tag{2.11}$$

where ρ_{τ_l} is the τ_l-th quantile loss function and E us the expectation with respect to random variables of X and Y. Under the homoscedasticity assumption, that is, the model errors do not depend on covariates, all conditional regression quantiles are parallel and they have the same slope β but different intercepts. The CQR is equivalent to fit QR at different quantile levels. However, CQR estimators are more efficient.

For given $0 < \tau_1 < \cdots < \tau_L < 1$ and a random variable X similar to (2.7), the partial composite quantile covariance (PCQC) $COV_{cqr}(Y, Z)$ between two random variable Y and Z is of the form

$$COV_{cqr}(Y, Z) = \arg_{\gamma} \inf_{\alpha_{\tau_1},\ldots,\alpha_{\tau_L},\beta,\gamma} E \sum_{l=1}^{L} \left(\rho_{\tau_l}\left(Y - \alpha_{\tau_l} - \beta X - \gamma Z\right)\right), \tag{2.12}$$

where we first normalize Z to have mean zero and variance one. If there is no X, then $COV_{cqr}(Y, Z)$ is composite quantile covariance (CQC) between Y and Z. The

composite quantile covariance measures the contribution of Z to the quantiles of Y at levels $0 < \tau_1 < \cdots < \tau_L < 1$. There are some connections between composite quantile covariance and covariance; however, these are beyond the scope of this chapter and we plan to discuss them elsewhere.

With the definition of PCQC, we can obtain the PCQR basis for functional linear composite quantile regression by maximizing

$$COV_{cqr}\left(Y - \mathbf{X}\boldsymbol{\beta}, \int_0^1 \mathbf{Z}(t)b(t)dt\right), \tag{2.13}$$

for a given quantile level set $0 < \tau_1 < \cdots < \tau_L < 1$. The subsequent basis is computed by iteratively maximizing (2.13) after accounting and subtracting the information of the previous basis. Once the PCQR basis is found, the functional linear composite quantile regression can be easily fitted by a linear program, for example `quantreg` in `R` [39]. The algorithm to compute the PCQR basis follows the same line of SIMPQR in the last section; we only need to replace COV_{qr} by COV_{cqr} in step 2.1 and keep the rest unchanged.

2.5 Simulation Studies

In this section, we investigate the finite sample performance of our proposed prediction methods, namely partial quantile regression (PQR) basis and partial composite quantile regression (PCQR) basis methods. We compare them with the fPC basis method in functional linear quantile regression (QRfPC) and functional linear composite quantile regression (CQRfPC) models, where fPC method has served as a popular benchmark in the existing literature [20]. In addition, we compare them with PLS basis and fPC basis methods in functional linear regression model. We

conduct our simulations in two settings where the first one is in favor of the fPC basis and the second one is a more general case. Both simulations show superior or comparable performance of our proposed methods.

Simulation I. In this simulation, we adapt the setup in Kato [37] by slightly changing the weights of parameter $\gamma(t)$. In particular, the model is of the form

$$\begin{aligned} Y &= \int_0^1 \gamma(t)Z(t)dt + \varepsilon, \\ \gamma(t) &= \sum_{j=1}^{50} \gamma_j \phi_j(t); \;\; \gamma_1 = 0.5, \gamma_j = \frac{20}{3}(-1)^{j+1} j^{-2}, j \geq 2, \phi_j(t) = 2^{1/2}\cos(j\pi t), \\ Z(t) &= \sum_{j=1}^{50} v_j U_j \phi_j(t); \;\; v_j = (-1)^{j+1} j^{-1.1/2}, U_j \sim U[-3^{1/2}, 3^{1/2}]. \end{aligned}$$

Each $X_i(t)$ was observed at $m = 201$ equally spaced grid points on $[0, 1]$. We choose the sample size n to be 100, 200, and 500. The error ε follows either Gaussian with mean zero and variance one or Cauchy distribution. In this design we have

$$Q_\tau(Y|X) = F_\varepsilon^{-1}(\tau) + \int_0^1 \gamma(t)Z(t)dt,$$

where F_ε is the cumulative distribution function of ε. It should be pointed out that the simulation set up is in favor of fPC basis methods as the functional coefficients lie on the same fPC space of functional covariates. It is expected that fPC basis methods may be superior to other methods.

To facilitate the comparison, we set $\tau = 0.5T$ for QR methods and $\tau_l = l/(1+L)$ with $L = 9$ for CQR methods. One criteria we use is the mean integrated errors

(MISE) of the functional coefficients,

$$\text{MISE} = \frac{1}{S}\sum_{s=1}^{S}\sum_{j=1}^{m}\left(\hat{\gamma}_s(t_j) - \gamma(t_j)\right)^2 = \text{Bias}^2 + \text{Var},$$

where

$$\text{Bias}^2 = \sum_j \left(\frac{1}{S}\sum_s \hat{\gamma}_s(t_j) - \gamma(t_j)\right)^2$$

and

$$\text{Var} = \frac{1}{S}\sum_s\sum_j \left(\hat{\gamma}_s(t_j) - \frac{1}{S}\sum_s \hat{\gamma}_s(t_j)\right)^2.$$

In the simulation, we set the total number of replication $S = 100$ due to the limitation of computational resources. In fact, for a larger S, the result would be very similar. For the first three cutoff levels, Table 2.1 gives us a summary of the different configurations of parameters for the six methods. Although the simulation design is in favour of fPC based methods, for the small number of cutoff levels, the PLS, PQR and PCQR methods perform better regarding the performance measurements of Bias^2 and MISE. Due to the natures of sensitivity against skewness of errors, Figure 2.2 shows that the performances of PLS and fPC are much worse in general compared with the other four methods when the errors follow the Cauchy distribution. On the other hand, when the Gaussian errors are employed, for the lower cutoff levels, the PLS, PQR and PCQR methods are very similar. And when the number of cutoff levels becomes larger, the PCQR performs slightly better than the PLS while PQR performs much better than the PCQR. fPC based methods are similar to each other crossing all cutoff levels.

The averaged mean squared error (MSE) of the responses is another prediction performance criteria we consider. Figure 2.3 indicates that the prediction errors are much lower for PLS, PQR, PCQR methods compared with those for fPC based

methods, due to the fact that fPC based methods are only data driven while the other three methods are both data and response adapted. For the Gaussian errors, although with regard to the functional coefficients estimation PQR is better than both PLS and PCQR methods, taking into consider of the prediction errors, the PLS and PCQR methods perform better than PQR. For the Cauchy errors, PQR performs the best out of the PLS, PQR, PCQR methods which indicates that PQR is more robust against the skewness of error distribution.

Simulation II. In this simulation, we take the $Z_i(t)$s from a real data study, and generate the Y_is according to the linear model of

$$Y = \int_0^1 \gamma(t)Z(t)dt + \varepsilon,$$

where the error ε is taken as Gaussian and Cauchy. The centres of errors are taken as zero while the scales are taken as the empirical standard deviation of the true responses multiplied by $\sqrt{5}$. The Z_is are taken from a benchmark Phoneme dataset (`http://statweb.stanford.edu/~tibs/ElemStatLearn/`). In these data, $Z_i(t)$ represents log-periodogram constructed from recordings of different phone--mes. The periodogram are available at 256 equally-spaced frequencies t, which for simplicity we denote by $0 = t_1 < t_1 < \ldots < t_m = 1$, where $m = 256$ [31]. We used $n = 1717$ data curves $Z_i(t)$ that correspond to the phonemes "aa" as in "dark" and "ao" as in "water". This example can also be found in Delaigle and Hall [20].

Computing the first $J = 20$ empirical fPCbasis functions $\hat{\phi}_1(t), \ldots, \hat{\phi}_J(t)$, we consider four different curves $\gamma(t)$ by taking $\gamma(t) = \sum_{j=1}^{J} a_j\hat{\phi}_i(t)$ for four different sequence of a_js: (i) $a_j = (-1)^j \cdot \mathbf{1}\{0 \le j \le 5\}$; (ii) $a_j = (-1)^j \cdot \mathbf{1}\{6 \le j \le 10\}$; (iii) $a_j = (-1)^j \cdot \mathbf{1}\{11 \le j \le 15\}$; (iv) $a_j = (-1)^j \cdot \mathbf{1}\{16 \le j \le 20\}$. Going through case (i)

to (iv), the models become less favorable for fPC, while we will see the PLS, PQR and PCQR methods manage to capture the interaction between Z and Y using only a few terms.

We take $\tau = 0.5$ and compare the six methods by looking at the MISE, Bias^2 and Var. As shown in Figure 2.4, from case (i) to (iv), PLS, PQR and PCQR methods perform better and better compared with the fPC based methods. In fact, all the PLS, PQR and PCQR methods manage to obtain a very good fitting using only a much small number of components no matter how the errors are distributed. This shows great superiority of our proposed methods when the functional coefficients do not lie on the fPC space.

Figure 2.5 displays the prediction errors MSE when the errors follow Gaussian (left panels) and Cauchy (right panels) distributions. The PLS, PQR and PCQR methods predict better in general compared with fPC based methods. Except for the PLS of Cauchy errors, the MSEs of PQR and PCQR methods decrease immediately with the increase of cutoff levels, while the fPC based methods performed differently under each case. From case (i) to (iv), the MSEs of fPC based methods begin to drop significantly after a larger and larger cutoff level. And for the same cutoff levels, the differences of the prediction errors between the PLS, PQR and PCQR methods and the fPC based methods become more and more signifiant from case (i) to (iv).

One interesting phenomenon here is that although the PCQR method outperforms the PQR method when the errors are Gaussian distributed, the PQR method regains its superiority when the errors are Cauchy distributed. Compared with what we have observed from simulation I, it may indicate that the PCQR method is only a slightly less favourable alternative to the PLS method when the errors are symmetric. On the other hand, when the errors are distributed in an extremely skewed

manner, the PCQR method could not out-perform the PQR method. That is exactly the same situation as the fPC based methods when the CQR method is implemented.

2.6 Real Data Analysis

Real Data Analysis I: ADHD-200 fMRI Data. We apply our proposed method to a dataset on attention deficit hyperactivity disorder (ADHD) from the ADHD-200 Sample Initiative Project. ADHD is the most commonly diagnosed behavioral disorder of childhood, and can continue through adolescence and adulthood. The symptoms include lack of attention, hyperactivity, and impulsive behavior. The dataset we use is the filtered preprocessed resting state data from New York University (NYU) Child Study Center using the Anatomical Automatic Labeling (AAL) [71] atlas. AAL contains 116 Regions of Interests (ROI) fractionated into functional space using nearest-neighbor interpolation. After cleaning the raw data that failed in quality control or has missing data, we include 120 individuals in the analysis.

The response of interest is the ADHD index, Conners' parent rating scale-revised, long version (CPRS-LV), a continuous behavior score reflecting the severity of the ADHD disease. In the AAL atalas data, the mean of the grey scale in each region is calculated for 172 equally spaced time points. We choose six parts of the brain which contain at least 4 ROIs, namely cerebelum, temporal, vermis, parietal, occipital, and frontal. The six functional predictors for each candidate part are computed by taking the average grey scale of the ROIs corresponding to each part, see Figure 2.1 (Left) for some selected subjects at cerebellum. The scalar covariates of primary interest include gender (female/male), age, handedness (continuous between -1 and 1, where -1 denotes totally left-handed and 1 denotes totally right-handed), diagnosis status (categorical with 3 levels: ADHD-combined,

ADHD-inattentative and Control as baseline), medication status (yes/no), Verbal IQ, Performance IQ and Full4 IQ. We build model to predict ADHD index adjusting these 9 scalar covariates (coded with dummy variables) using each of the six functional predictors. We consider the models for each individual functional covariates adjusting for the 9 scalar covariates.

Figure 2.6 displays the changes of MSEs for all six methods, with the increase of the number of cutoff level L for different brain regions. Here the quantile level τ is chosen to be fixed as 0.5. As shown in the figure, PLS, PQR and PCQR methods perform much better than the fPC based methods while PCQR shows a significant superiority. In general, for each method only a few basis functions is capable of predicting the response well and additional basis functions do not decrease MSE much. This is more obvious for PLS, PQR and PCQR methods as they consider information from the response while choose basis functions.

Real Data Analysis II: ADNI DTI Data. We use our model methods to analyze a real DTI data set with $n = 214$ subjects collected from NIH Alzheimer's Disease Neuroimaging Initiative (ADNI) study. Data used in the preparation of this chapter were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early

Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see `www.adni-info.org`. The significance level is an ongoing public-private partnership to test whether genetic, structural and functional neuroimaging, and clinical data can be integrated to assess the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). The structural brain MRI data and corresponding clinical and genetic data from baseline and follow-up were downloaded from the ADNI publicly available database (`https://ida/loni/usc/edu`).

The DTI data were processed by two key steps including a weighted least squares estimation method [4, 86] to construct the diffusion tensors and a FSL TBSS pipeline [67] to register DTIs from multiple subjects to create a mean image and a mean skeleton. Speciffically, maps of fractional anisotropy (FA) were computed for all subjects from the DTI after eddy current correction and automatic brain extraction using FMRIB software library. FA maps were then fed into the TBSS tool,

which is also part of the FSL. In the TBSS analysis, the FA data of all the subjects were aligned into a common space by non-linear registration and the mean FA image were created and thinned to obtain a mean FA skeleton, which represents the centers of all WM tracts common to the group. Subsequently, each subjects aligned FA data were projected onto this skeleton. We focus on the midsagittal corpus callosum skeleton and associated FA curves from all subjects, see Figure 2.1 (Right) for some selected subjects. The corpus callosum (CC) is the largest fiber tract in the human brain and is a topographically organized structure, see Figure 2.7 (Left). It is responsible for much of the communication between the two hemispheres and connects homologous areas in the two cerebral hemispheres. It is important in the transfer of visual, motoric, somatosensory, and auditory information.

We are interested in predicting mini-mental state examination (MMSE) scores, one of the most widely used screening tests, which are used to provide brief, objective measures of cognitive functioning for almost fifty years. The MMSE scores has been seen as a reliable and valid clinical measure quantitatively assessing the severity of cognitive impairment. It was believed that the MMSE scores to be affected by demographic features such as age, education and cultural background, but not gender [70]. After quality control and excluding the missing data, we include 200 subjects from the total 217 subjects. The functional covariate is fractional anisotropy (FA) values along the corpus callosum (CC) fiber tract with 83 equally spaced grid points, which can be treated as a function of arc-length. The scale covariates are the gender variable (coded by a dummy variable indicating for male), the age of the subject (years), the education level (years), an indicator for Alzheimer's disease (AD) status (19.6%) and an indicator for mild cognitive impairment (MCI) status (55.1%), and genotypes for apolipoprotein E ϵ-4 (coded by three indicator variables for four levels).

The MSEs are shown in figure 2.7. In general, PLS, PQR and PCQR methods present consistently better than fPC based methods while PCQR outperforms PQR and PCQR methods. The phenomenon has been observed from the previous read data analysis, which indicates that for brain imaging data PCQR method has a improved prediction accuracy compared with PQR and PCQR methods. With the number of functional basis increases, the MSEs do not decreases much for fPC based methods while constantly decrease for PLS, PQR and PCQR methods. This indicates that the fPC basis is not suitable to do prediction though they may account a large portion of the variations of functional covariates. The PLS, PQR and PCQR methods is capable of explaining a large percentage variation of the response and reducing the MSEs by proving appropriate basis functions. Our proposed methods show great superiority to the fPC based methods and the PLS methods and provide a powerful tool to do prediction in practice.

2.7 Discussion

In this chapter, we first define the concept partial quantile covariance (PQC) to measure the contribution of one covariate to the response. We then propose the partial functional linear quantile regression method to use partial quantile regression (PQR) to extract PQR basis to effectively predict the response. This is motivated by the success of the partial least square (PLS) basis in functional linear regression model. The key idea is to use both information from the functional covariates and the response and therefore both PQR basis and PLS basis can be treated as supervised learning while fPC based methods are semi-supervised learning as they only use information from the functional covariates. The algorithm SIMPQR we developed is analogue to that of SIMPLS. We extend PQC to partial composite quartile

covariance (PCQC) and propose the PCQR basis and its SIMPCQR algorithm under the homoscedasticity condition.

The simulations show that PLS, PQR and PCQR in general perform better than the fPC based methods. However, PQR method is more robust against skewness of error distribution while the PLS and PCQR methods act similarly to each other and perform better than PQR method when the error distribution is symmetric. This advantage from PQR method can be explained by the general nature of quantile method which obtains its robustness by sacrificing certain efficiency. By assuming homoscedasticity, the PCQR method acts similarly to the PLS method when the error distributions are symmetric but retains its robustness when the error distributions are extremely skewed.

Our proposed methods, PQR and PCQR methods, significantly outperform other methods, especially those fPC based methods in both ADHD-200 fMRI data analysis and ADNI DTI data analysis. In ADHD-200 fMRI data analysis, our methods are capable of reducing much more MSEs by using only a few basis while fPC based method are not even by adding more basis. In ADNI DTI data analysis, both PQR and PCQR methods reduce significant amount MSEs with more and more basis. On the other hand, fPC based methods perform poor even with more basis. Overall in the two neuroimaging data analysis, PCQR performs slightly better than PQR though.

The consistency of the PLS methods was proved by Delaigle and Hall [20] where they found an equivalent space with explicit expressed basis functions to the PLS basis space. For PQR and PCQR methods, it is difficult to find such equivalent space and therefore their consistency may not be easy to show. The difficulty of the problem lies on the iterative nature of PQR and PCQR methods where basis is sequentially extracted. One way to overcome that is to find preselected number of

basis simultaneously [83]. Another direction is to impose certain structure on the selected basis, for example, sparsity and smoothness in PLS methods [62]. This can be done for simultaneous basis selection as well [82].

In both simulation studies and real data analysis, only univariate functional covariate case is considered. However, the extension of PQR and PCQR methods to multivariate functional covariates is straightforward. The computation becomes more complex and intensive due to the iterative basis extraction nature. Such complexity is expected to be significantly reduced by applying simultaneous basis selection or imposing certain structure on the selected basis. Further details are out of the scope of this chapter and will be pursuit in the future research.

			fPC			QRfPC			CQRfPC		
Error	L	n	Bias^2	Var	MISE	Bias^2	Var	MISE	Bias^2	Var	MISE
Gaussian	1	100	3.63	0.07*	3.70	3.63	0.09	3.72	3.63	0.07*	3.71
		200	3.64	0.03*	3.67	3.63	0.04	3.68	3.63	0.04	3.67
		500	3.69	0.02*	3.70	3.68	0.02*	3.71	3.69	0.02*	3.70
	2	100	0.78	0.36*	1.14*	0.77	0.39	1.16	0.78	0.36*	1.14*
		200	0.86	0.17*	1.03	0.86	0.18	1.04	0.86	0.17*	1.03
		500	0.86	0.09*	0.95	0.86	0.10	0.96	0.86	0.09*	0.95
	3	100	0.32	0.34*	0.67*	0.33	0.38	0.70	0.32	0.35	0.67*
		200	0.28	0.19*	0.47*	0.28	0.22	0.50	0.28	0.20	0.48
		500	0.29	0.08*	0.38	0.29	0.09	0.38	0.29	0.08*	0.38
Cauchy	1	100	7.32	>100	>100	3.65	0.12	3.76	3.66	0.09*	3.75
		200	4.82	54.42	59.24	3.63	0.10	3.73	3.64	0.08*	3.72
		500	43.78	>100	>100	3.71	0.03	3.74	3.72	0.02*	3.74
	2	100	7.56	>100	>100	0.77	0.40*	1.17*	0.76	0.42	1.18
		200	2.14	>100	>100	0.78	0.21*	0.99	0.77	0.22	1.00
		500	>100	>100	>100	0.81	0.10*	0.91	0.81	0.10*	0.91
	3	100	6.03	>100	>100	0.28	0.48*	0.76*	0.27	0.53	0.80
		200	3.67	>100	>100	0.31	0.24*	0.55*	0.31	0.27	0.58
		500	>100	>100	>100	0.31	0.11*	0.41*	0.31	0.12	0.42

			PLS			PQR			PCQR		
Error	L	n	Bias^2	Var	MISE	Bias^2	Var	MISE	Bias^2	Var	MISE
Gaussian	1	100	0.54	0.82	1.36*	0.50*	0.91	1.41	0.52	0.84	1.36*
		200	0.63	0.20	0.83	0.57*	0.26	0.83	0.60	0.21	0.81*
		500	0.59	0.07	0.66	0.52*	0.10	0.62*	0.56	0.08	0.64
	2	100	0.11*	1.07	1.18	0.15	1.18	1.33	0.12	1.07	1.19
		200	0.12*	0.29	0.41*	0.16	0.36	0.52	0.13	0.29	0.43
		500	0.11*	0.10	0.21*	0.14	0.13	0.27	0.12	0.10	0.23
	3	100	0.08	2.58	2.66	0.07	2.28	2.36	0.06*	2.65	2.71
		200	0.04*	0.87	0.91	0.04*	1.10	1.13	0.04*	0.96	0.99
		500	0.02*	0.26	0.28*	0.02*	0.43	0.45	0.02*	0.30	0.32
Cauchy	1	100	71.08	>100	>100	0.49	1.23	1.72*	0.48*	1.39	1.87
		200	41.24	>100	>100	0.48	0.43	0.91*	0.47*	0.47	0.94
		500	>100	>100	>100	0.46*	0.16	0.62*	0.48	0.17	0.64
	2	100	>100	>100	>100	0.16	1.96	2.12	0.12*	3.01	3.13
		200	>100	>100	>100	0.15	0.67	0.82*	0.12*	0.98	1.09
		500	>100	>100	>100	0.14	0.23	0.37*	0.11*	0.29	0.41
	3	100	>100	>100	>100	0.13*	5.84	5.97	0.20	14.28	14.48
		200	>100	>100	>100	0.06*	2.59	2.65	0.11	5.45	5.56
		500	>100	>100	>100	0.02*	1.01	1.02	0.04	1.67	1.71

Table 2.1: Simulation I: $*$ flags the minimum values of the six methods in each measurement category of Bias^2, Var and MISE.

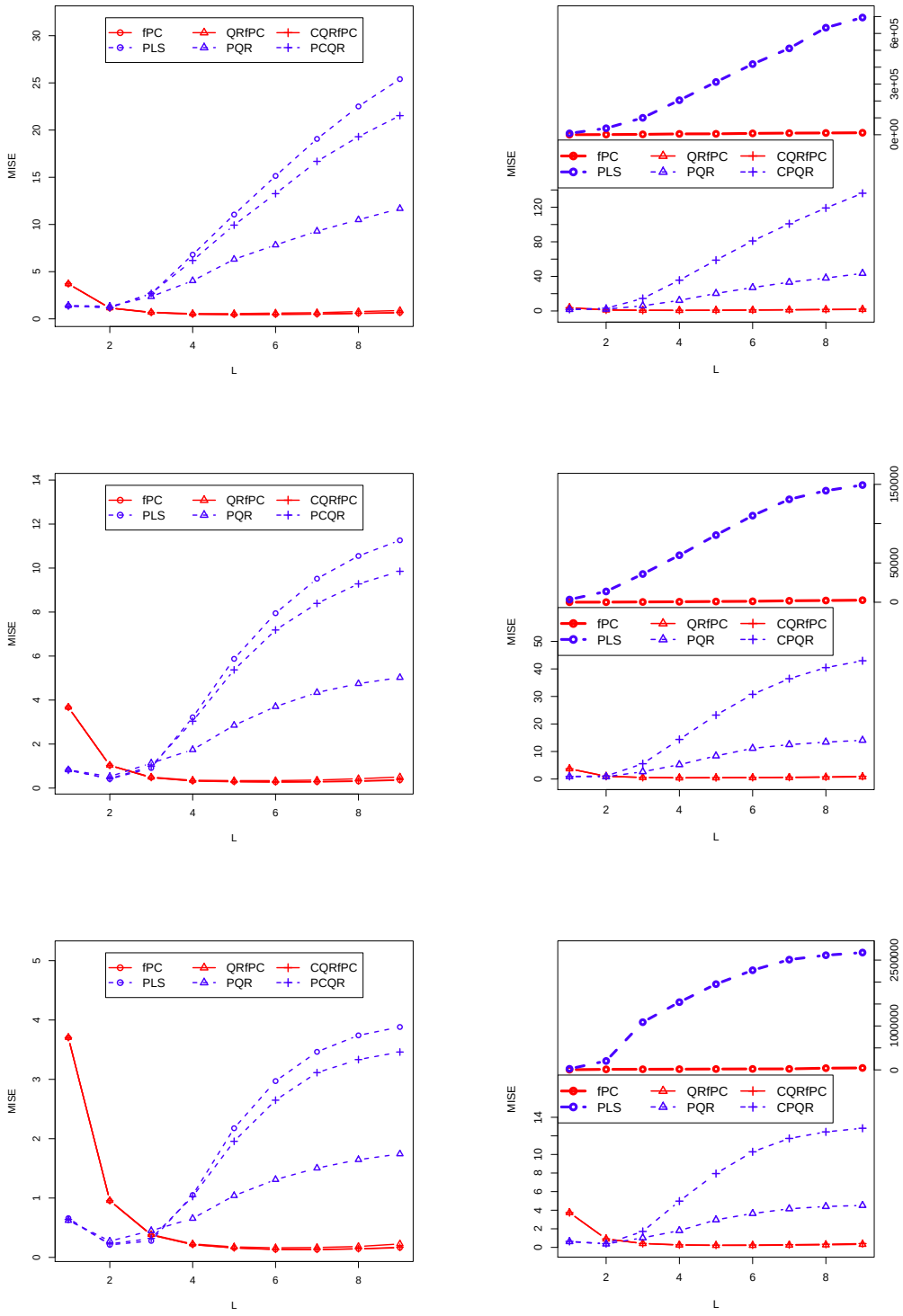


Figure 2.2: Simulation I: the MISEs with Gaussian (left) and Cauchy (right) errors, sample size n = 100, 200, and 500 from up to down.

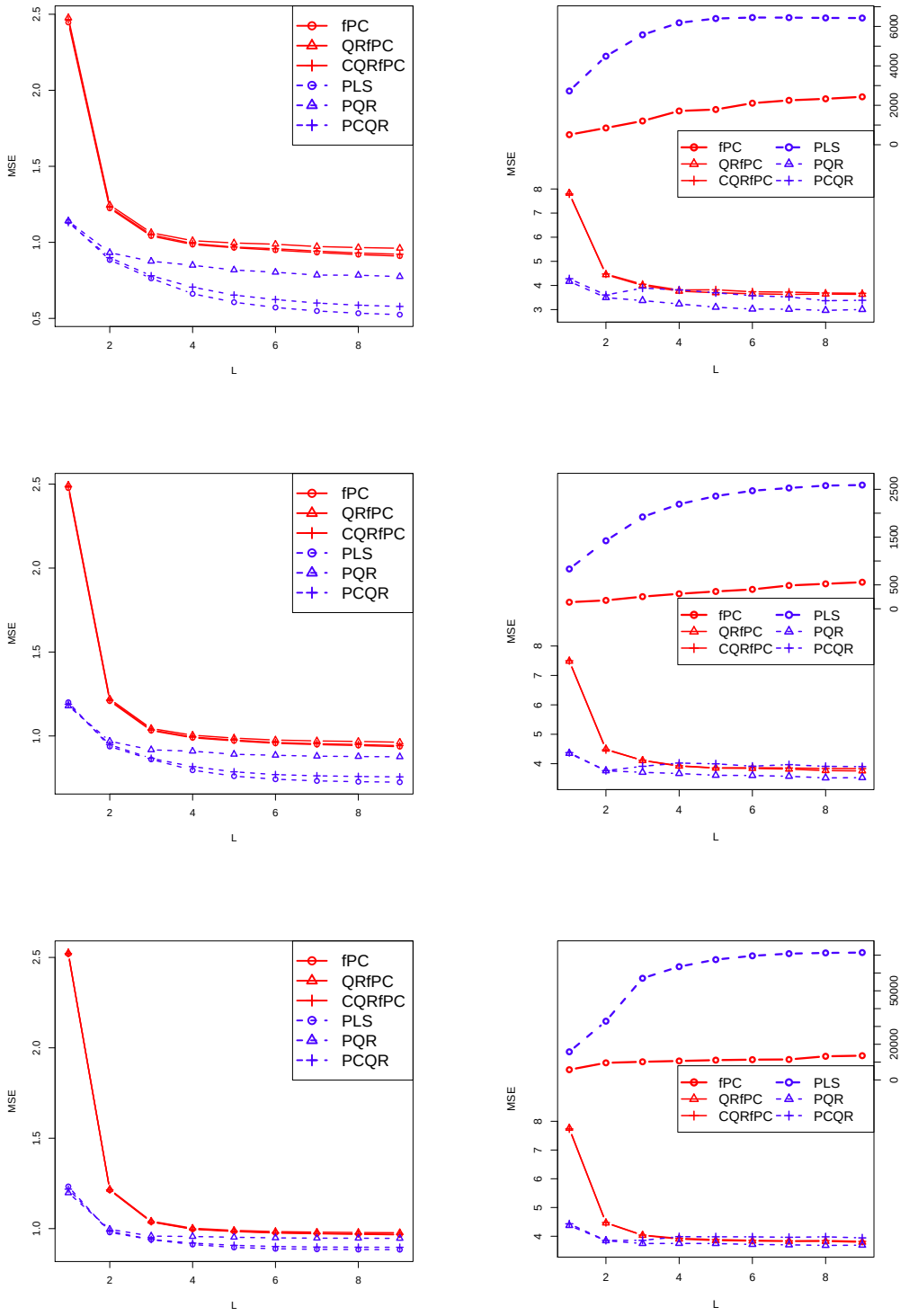


Figure 2.3: Simulation I: the averaged MSEs with Gaussian (left) and Cauchy (right) errors, sample size n = 100, 200, and 500 from up to down.

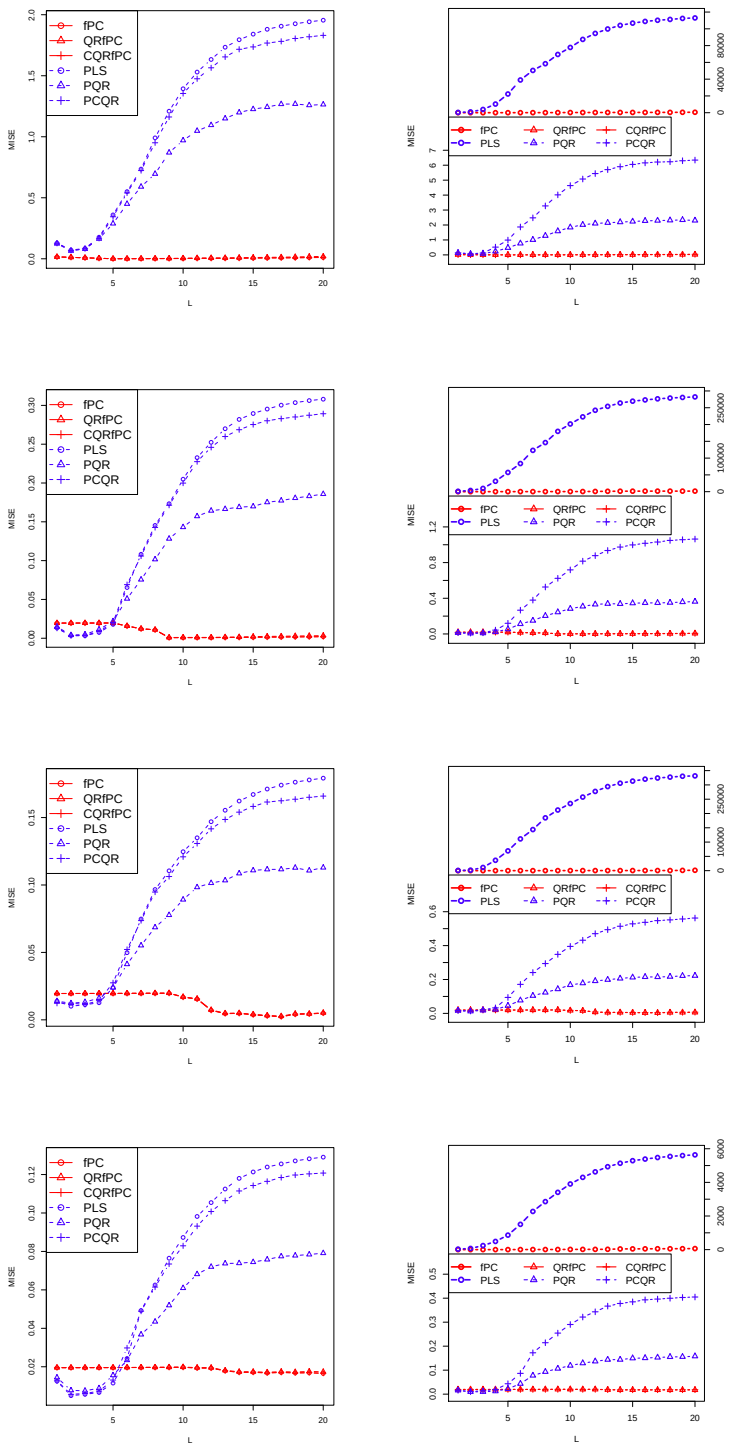


Figure 2.4: Simulation II: the MISEs with Gaussian (left) and Cauchy (right) errors, case I, II, II, and IV from up to down.

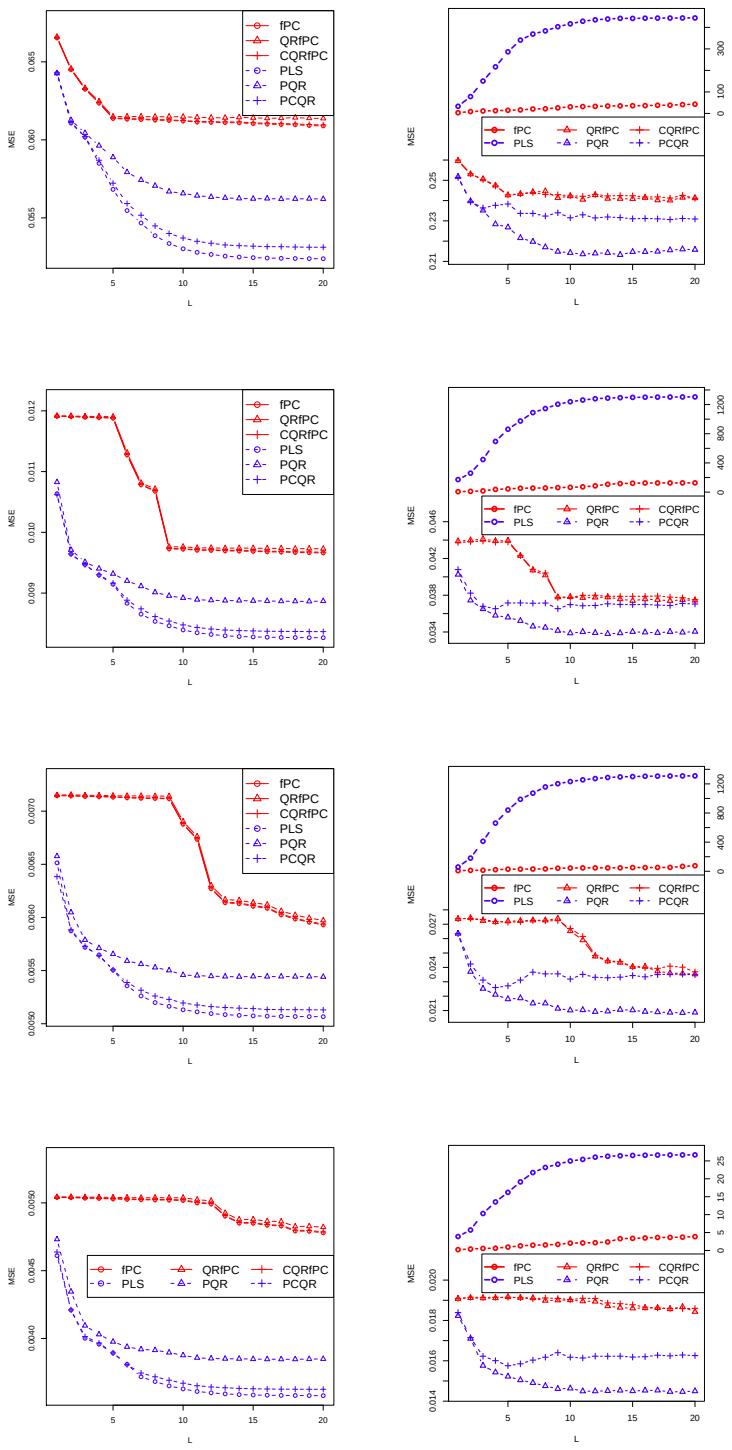


Figure 2.5: Simulation II: the averaged MSEs with Gaussian (left) and Cauchy (right) errors, case I, II, II, and IV from up to down.

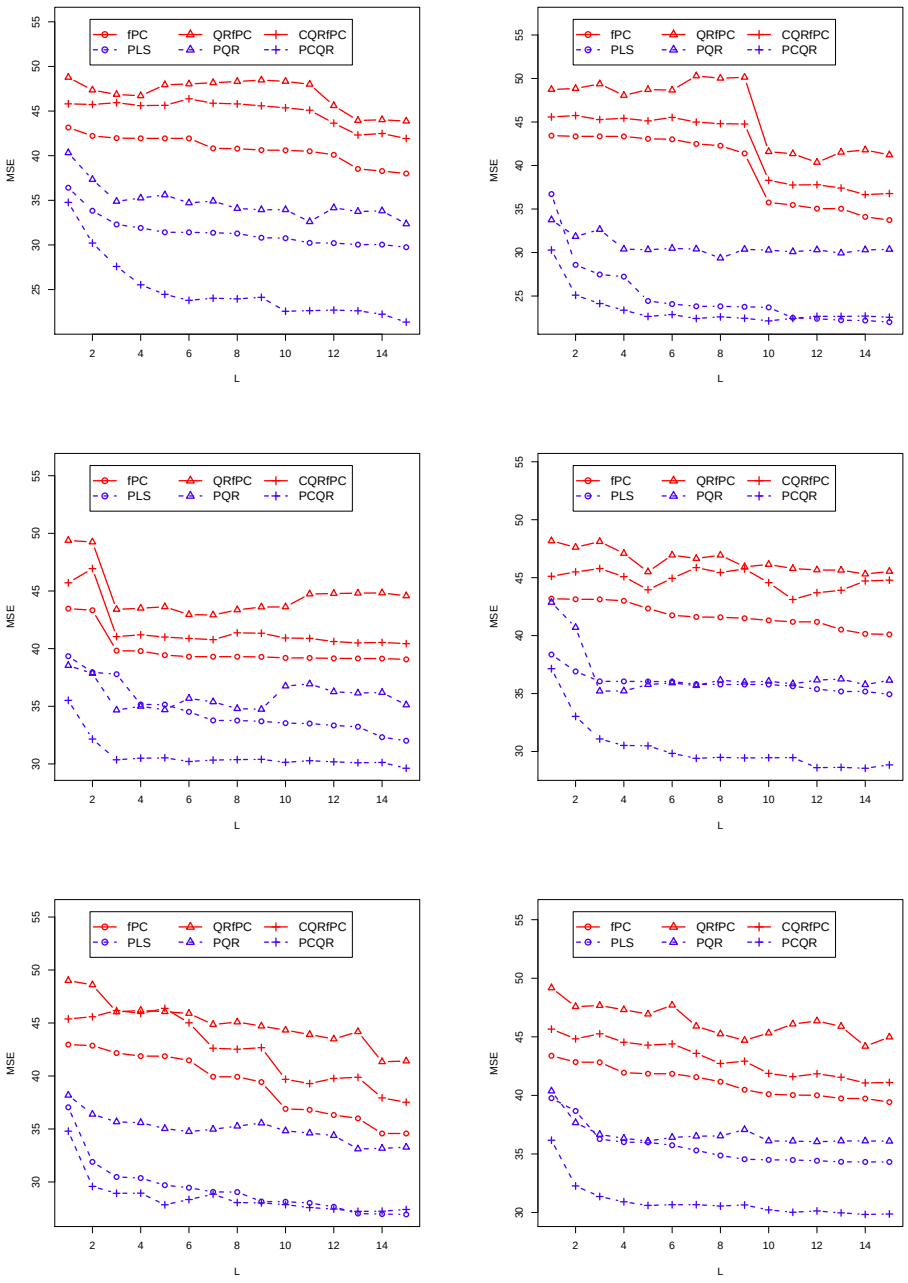


Figure 2.6: Real Data Analysis I: ADHD-200 fMRI data, From up to down there are cerebelum, vermis, and occipital on the left panels and temporal, parietal, and frontal on the right panel.

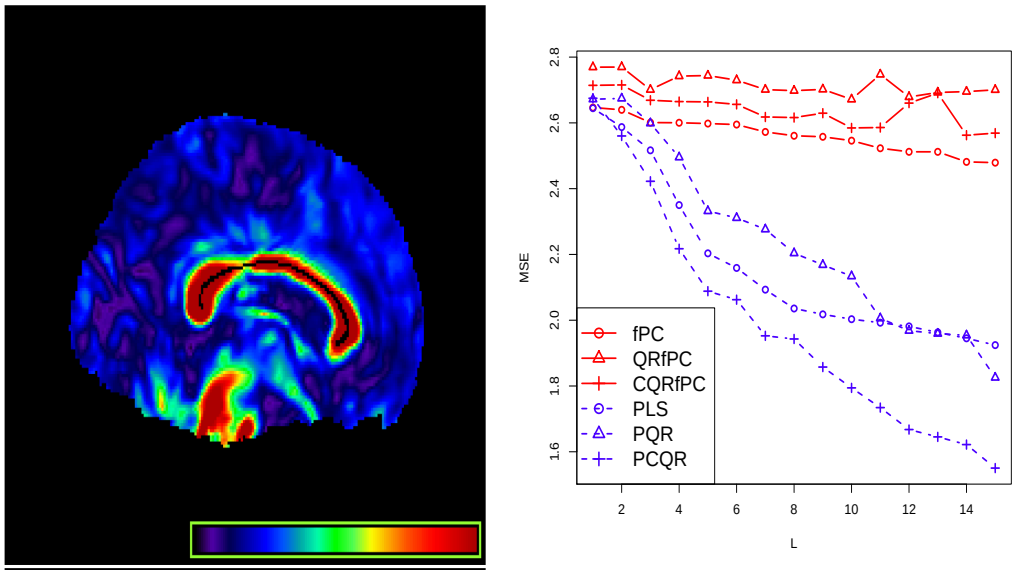


Figure 2.7: Real Data Analysis II: (Left) The midsagittal corpus callosum (CC) skeleton overlaid with fractional anisotropy (FA) from one randomly selected subject and (Right) the MSE of mini-mental state examination (MMSE) at different cut-off levels.

Chapter 3

An Alternative Approach to Functional Linear Partial Quantile Regression

In Chapter 2, we proposed the partial quantile regression (PQR) prediction procedure for functional linear model by using partial quantile covariance techniques and developed the simple partial quantile regression (SIMPQR) algorithm to efficiently extract PQR basis for estimating functional coefficients. However, although the PQR approach is considered as an attractive alternative to projections onto the principal component basis, there are certain limitations to uncovering the corresponding asymptotic properties mainly because of its iterative nature and the non-differentiability of the quantile loss function which prevent many methods with nice properties being applied directly. In this chapter, we propose and implement an alternative formulation of partial quantile regression (APQR) for functional linear model by using block relaxation ideas and finite smoothing techniques. The proposed reformulation leads to insightful results and motivates new theory so that we

can demonstrate consistency and establish convergence rates by applying advanced techniques from empirical process theory. Two simulations and two real data from ADHD-200 sample and ADNI are investigated to show the superiority of our proposed methods.

3.1 Introduction

Partial least squares (PLS) is an iterative procedure for feature extraction of linear regression. The technique was originally developed in high dimensional and collinear multivariate settings and is especially popular in chemometrics [1, 24, 32, 57, 74]. As a supervised dimension reduction technique, PLS projects the original data onto a lower dimensional subspace formed by the linear projects of covariates which best predict the responses. More recently, PLS has been applied in the functional data context by Preda and Saporta [60] using functional linear model, with its consistency and convergence rates established and demonstrated by Delaigle and Hall [20].

Consider a functional linear model with scalar response:

$$y = \alpha + \mathbf{x}^T\boldsymbol{\beta} + \int_I \mathbf{z}^T(t)\boldsymbol{\gamma}(t)dt + \varepsilon, \tag{3.1}$$

where $\mathbf{x}$ and $\mathbf{z}(t)$ are scalar and functional covariates. The contributions of $\mathbf{z}(t)$ towards the variation of y is characterized by the functional coefficient $\boldsymbol{\gamma}(t)$ and change by t. To facilitate the estimation of $\boldsymbol{\gamma}$(t), we usually require that it satisfies certain smoothness conditions and restrict it onto a functional space. For example, we may require that its second derivative exists [36] and $\boldsymbol{\gamma}(t)$ is square integrable [37]. Even in such a situation, the estimation is still an infinite-dimensional prob-

lem.

The common practice is to project $\boldsymbol{\gamma}(t)$ into a space spanned by a finite number of functional basis. There are three major choices of such basis: general basis, functional principal component basis (fPC) and partial least square basis (PLS). There are many options of choosing general basis; for instance, B-spline basis [12] and wavelet basis [79]. In order to provide a good approximation of functional coefficients, a large number of basis are often used. However, this may cause model overfitting, and hence to remedy that, various penalization methods have been proposed [16, 80]. The fPC method has also been extensively studied [29, 44], where the fPCs of $\mathbf{z}(t)$ serve as the basis. Although fPC basis are more data-adapted than the general basis as they use the information of functional covariates and the formed space can explain most of the variation of $\mathbf{z}(t)$, it is not necessary all the fPC basis can contribute to the variation of the responses. Therefore, the PLS basis, which use the information from both covariates and responses, becomes an appealing supplement. In particular, using as basis the linear projects of $\mathbf{z}(t)$ which best predict the responses, PLS method can often capture more relevant information with fewer terms [20].

In recent years, quantile regression, which was introduced by the seminal work of Koenker and Bassett [41], has been well developed and recognized in functional linear regression [37, 76]. In this chapter, we focus on the functional linear quantile regression model:

$$Q_\tau(y|\mathbf{x}, \mathbf{z}(t)) = \alpha_\tau + \mathbf{x}^T\boldsymbol{\beta}_\tau + \int_I \mathbf{z}^T(t)\boldsymbol{\gamma}_\tau(t)dt, \tag{3.2}$$

where $Q_\tau(y|\mathbf{x}, \mathbf{z}(t))$ is the τ-th conditional quantile of response y given scalar covariates $\mathbf{x}$ and functional covariate $\mathbf{z}(t)$ for a fixed quantile level $\tau \in (0, 1)$. As an

alternative to least squares regression, the quantile regression method is more efficient and robust when the responses are non-normal, errors are heavy tailed or outliers are present. It is also capable of dealing with the heteroscedasticity issues and providing a more complete picture of the response [38].

To estimate functional coefficients $\gamma_\tau(t)$, it is convenient to restrict them onto a finite dimensional functional space. Similarly to mean regression, general basis like B-spline can be used to approximate the quantile coefficient functions [11, 68], while fPC basis have also been throughly investigated [37, 52, 69]. Analogue to the PLS in functional linear regression model, partial quantile regression (PQR) basis, utilizing information from both response and covariates, become an attractive alternative. In particular, the PQR basis can be obtained by using simple PQR (SIMPQR) procedure proposed by Yu et al. [76], where the basis are extracted by sequentially maximizing the partial quantile covariance between the response and projections of functional covariate.

Although the functional PQR approach appears to be an enticing choice alternative to the principal component method, there are certain limitations to uncovering the asymptotic properties mainly because of its iterative nature and the non-differentiability of the quantile loss functions. The difficulty due to the iterative formulation used to exist for functional PLS too. To deal with it, an alternative but equivalent PLS formulation (APLS) was proposed by Delaigle and Hall [20]. Based on the fact that there exists an equivalence between fPC space and functional APLS space, it is then possible to verify the consistency and convergence rates. Unfortunately, such equivalence does not exist between PQR and fPC spaces because of the non-additivity of conditional quantiles. In addition, the non-differentiability of the quantile loss function $\rho_\tau(\cdot)$ can prevent many methods and properties being directly applied in quantile regression context [75, 81].

To address these problems, we firstly propose a smoothing approximation for the quantile loss function by applying the finite smoothing techniques [13, 54]. The approximation function should uniformly converge towards the quantile loss function so that the minimizer of the former converges to the minimizer of the latter in a compact set. Replacing ρ_τ by such smoothing approximation, the quantile objective function becomes differentiable. Then for a given fixed K, namely the number of PQR basis, the original PQR formulation can be modified according to the block relaxation ideas [19] which updates and obtains the basis as a "block" instead of one by one sequentially. The value of K can be chosen using BIC or cross validation (CV) as in choosing the number of fPC basis adapted by Kato and other authors [37, 52, 69]. Such modification provides an alternative to the original formulation for PQR (APQR) basis which leads to insightful results and motivates new theory so that we can demonstrate consistency and establish convergence rates by applying advanced techniques from empirical processes theory [72].

The rest of this chapter is organized as follows. In Section 3.2, we define the partial quantile covariance and describe how it can be used to extract PQR basis in functional linear quantile regression model. In Section 3.3, we propose and implement the alternative partial quantile regression (APQR) method and demonstrate its convergence properties. The asymptotic properties including the consistency and convergence rates of the proposed method are established in Section 3.4. In Section 3.5 and 3.6, we use two simulations and two real data from ADHD-200 sample and ADNI to illustrate the superiority of our proposed method. Some discussions and future research directions are made in Section 3.7.

3.2 Alternative Partial Functional Linear Quantile Regression

In model (3.2), we assume without loss generality that $t \in I = [0, 1]$ and restrict the functional coefficient $\boldsymbol{\gamma}_\tau(t) \in L_2[0, 1]$. For simplicity, only one functional covariate is considered here. The extension to multiple functional covariates is straightforward. In general, the estimation of $\boldsymbol{\gamma}_\tau(t)$ is a difficult problem as it lies in an infinite dimensional space. However, if it can be projected into a finite dimensional space, say $H[0, 1]$, it can be easily approximated. In particular, let $\phi_{\tau k}(t)$, $k = 1, \ldots, K$, be a basis of $H[0, 1]$ and $\gamma_\tau(t) = \sum_{k=1}^{K} \gamma_{\tau k}\phi_k(t)$. Model (3.2) can be rewritten as

$$Q_\tau\left(y|\mathbf{x}, \mathbf{z}(t)\right) = \alpha_\tau + \mathbf{x}^T\boldsymbol{\beta}_\tau + \sum_{k=1}^{K} \mathbf{z}_{ik}\boldsymbol{\gamma}_{\tau k}, \tag{3.3}$$

where $\mathbf{z}_{ik} = \int_0^1 \mathbf{z}(t)\phi_k(t)dt$ for $i = 1, \ldots, n$. Model (3.3) is simply a multiple linear regression, which is essentially a linear programming problem and can be solved by many algorithms - for example, the simplex method [6], the interior point method [38], the MM algorithm [35] among many others, already implemented in various statistical softwares like **quantreg** in **R** [40].

In functional linear model (3.1), for a given number of basis, say K, the partial least square (PLS) basis $\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_K(t))$ are chosen to be

$$\arg\min_{\boldsymbol{\phi}(t)} \min_{\alpha,\boldsymbol{\beta},\boldsymbol{\phi}(t)} \mathbb{E}\left(y_i - \alpha - \mathbf{x}_i^T\boldsymbol{\beta} - \sum_{k=1}^{K} \int_0^1 \mathbf{z}_i(t)\phi_k(t)dt\right)^2, \tag{3.4}$$

which is the analogue to the partial least square regression in multivariate analysis. Here $\mathbb{E}$ is the empirical expectation. The essential idea of criteria (3.4) is to

find a group of directions $\boldsymbol{\phi}(t)$ so that the projections of $\mathbf{z}_i(t)$s on them explain as much as possible the variation of the response after adjusting other covariates. It is equivalent to find the basis $\boldsymbol{\phi}(t)$ such that partial covariance

$$COV\left(\mathbf{y} - \alpha - \mathbf{X}\boldsymbol{\beta}, \sum_{k=1}^{K} \int_0^1 \mathbf{Z}(t)\phi_k(t)dt\right) \tag{3.5}$$

is maximized where $\mathbf{y} = (y_1, \ldots, y_n)^T$, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ and $\mathbf{Z}(t) = (z_1(t), \ldots, z_n(t))^T$. Based on this idea, Delaigle and Hall [20] found a equivalent PLS space, in which the consistency and rates of convergence have been established.

Similarly, the parameters in model (3.2) can be estimated by solving

$$\min_{\alpha,\boldsymbol{\beta},\boldsymbol{\phi}(t)} \mathbb{E}\rho_\tau\left(y_i - \alpha - \mathbf{x}_i^T\boldsymbol{\beta} - \sum_{k=1}^{K} \int_0^1 \mathbf{z}_i(t)\phi_k(t)dt\right), \tag{3.6}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the quantile loss function [41] with I as the indicator function. Similarly to the formulation of PLS basis, for given $\tau \in (0, 1)$, a group of directions $\boldsymbol{\phi}_\tau(t)$ can be found so that the projections of $\mathbf{z}_i(t)$s onto them contribute as much as possible to predict the conditional quantile of the response. The concepts of quantile covariance (QC) and partial quantile covariance (PQC) were firstly proposed by Yu et al. [76]. For given $\tau \in (0, 1)$ and a random variable X, the partial quantile covariance (PQC) between two random variables Y and Z is defined as

$$COV_{\mathrm{qr}}(Y, Z|X) = \arg_{\boldsymbol{\gamma}} \inf_{\alpha,\boldsymbol{\beta},\boldsymbol{\gamma}} E\left(\rho_\tau\left(Y - \alpha - \boldsymbol{\beta}X - \boldsymbol{\gamma}Z\right)\right), \tag{3.7}$$

where Z is normalized to have mean zero and variance one. If there is no X, $COV_{\mathrm{qr}}(Y, Z)$ is just quantile covariance (QC) between Y and Z. The quantile covariance measures the contribution of Z to the τ-th quantile of Y. It was first proposed

and studied by Dodge and Whittaker [21] in the context of partial quantile regression. Recently, Li et al. [45] proposed a similar concept of quantile correlation and used it to study quantile autoregressive model, while Ma et al. [53] used the partial quantile correlation to study the ultra-high dimensional variable screening problem.

To find the partial quantile regression basis (PQR), we propose to compute $\boldsymbol{\phi}_\tau(t)$ by maximizing

$$COV_{\mathrm{qr}}\left(\mathbf{y} - \alpha_\tau - \mathbf{X}\boldsymbol{\beta}_\tau, \sum_{k=1}^{K} \int_0^1 \mathbf{Z}(t)\phi_{\tau k}(t)dt\right). \tag{3.8}$$

By projecting functional covariates onto them, the functional linear model (3.2) is approximated by model (3.3), where the parameters can be easily obtained by minimizing

$$\mathbb{E}\rho_\tau\left(y_i - \alpha_\tau - \mathbf{x}_i^T\boldsymbol{\beta}_\tau - \sum_{k=1}^{K} \mathbf{z}_{ik}\boldsymbol{\gamma}_{\tau k}\right). \tag{3.9}$$

The value of K, namely the number of PQR basis, can be chosen using BIC or cross validation (CV) as in choosing the number of fPC basis adopted by Kato and other authors [37, 52, 69].

3.3 Alternative Partial Quantile Regression

Although the original functional PQR approach [76] is regarded as an appealing alternative to the fPC method, there are certain limitations to uncovering the asymptotic properties mainly because of its iterative nature and the non-differentiability of the quantile loss functions. The difficulty due to the iterative formulation used to exist for functional PLS too. To deal with it, an alternative but equivalent PLS formulation (APLS) was proposed by Delaigle and Hall [20]. Based on the fact that there exists an equivalence between fPC and APLS spaces, it is then possible to

verify the consistency and convergence rates. Unfortunately, such equivalence does not exist between PQR and fPC spaces due to the non-additivity of conditional quantiles. In addition, the non-differentiability of the quantile loss function $\rho_\tau(\cdot)$ at the origin can prevent many methods and results from being directly applied in quantile regression context [75, 81].

To address these problems, in this section, we firstly propose a smoothing approximation for the quantile loss function by applying the finite smoothing techniques [13, 54]. The fact that such approximation can uniformly converge towards the quantile loss function guarantees that the minimizer of the former converges to the minimizer of the latter on a compact set. Replacing ρ_τ by its smoothing approximation, the quantile objective function becomes differentiable. Then for a given fixed K, the number of PQR basis, the original PQR formulation can be modified according to the block relaxation ideas [19] which updates and obtains the basis as a "block" instead of one by one sequentially, while the value of K can be chosen using BIC or cross validation (CV) as in choosing the number of fPC basis adapted by Kato and other authors [37, 52, 69]. Such modification provides an alternative to the original formulation for PQR (APQR) basis which leads to our proposed alternative SIMPQR (ASIMPQR) Algorithm.

Finite Smoothing of Quantile Loss Functions As mentioned above, to uncover the asymptotic properties for functional PQR, one difficulty lies in the non-differentia--bility of the quantile loss function $\rho_\tau(\cdot)$ at the origin, which can prevent many methods and results from being directly applied in quantile regression context [75, 81]. To address such problem, we propose for $\rho_\tau(u)$ a smoothing approximation $\rho_{\tau\nu}(u)$ while $\rho_{\tau\nu}(u)$ converges pointwise to $\rho_\tau(u)$, as $\boldsymbol{\nu} \to \boldsymbol{\nu}_0$, where $\boldsymbol{\nu}$ is a vector of smoothing parameters and $\boldsymbol{\nu}_0$ is typically chosen as $\mathbf{0}$. [13, 54]. In addition, if

such convergence is uniform, the minimizer of $\rho_{\tau\nu}(u)$ converges to the minimizer of $\rho_\tau(u)$, for a compact set in $(0, 1)$ [33]. Replacing ρ_τ by $\rho_{\tau\nu}$, the quantile objective function then becomes differentiable which further facilitates defining score, information and identifiability in term of functional linear quantile regression, so that the estimator consistency and convergence rates can be established.

The finite smoothing technique of quantile loss function is a special case from the general problem of non-smooth convex optimization and is both statistically and computationally important in quantile regressions [13, 56, 75, 81]. For a given precision ϵ, by choosing an appropriate smoothing function, one may obtain an optimal efficiency estimate in terms of number of iterations until convergence as $O(\frac{1}{\epsilon})$, which is a significant improvement compared with some popular numerical scheme for non-smooth convex minimization such as subgradient methods whose number of iterations until convergence is $O(\frac{1}{\epsilon^2})$ [56]. There are various options regarding the choices of smoothing functions. Some particular examples include *generalized Huber function* [13] and *iterative least squares smoothing function* [54], both of which converge uniformly towards $\rho_\tau(u)$. We can take *generalized Huber function* [13] as an example:

$$H_{\nu,\tau}(u) = \begin{cases} u(\tau-1) - \frac{1}{2}(\tau-1)^2\nu, & u \le (\tau-1)\nu; \\ \frac{u^2}{2\nu}, & (\tau-1)\nu < u \le \tau\nu; \\ u\tau - \frac{1}{2}\tau^2\nu, & u > \tau\nu. \end{cases}$$

By choosing a non-negative number ν, and let it go to 0, $H_{\nu,\tau}(u)$ converges uniformly towards $\rho_\tau(u)$. Therefore, the minimizer of $H_{\nu,\tau}$ converges to the minimizer of ρ_τ too [33]. The smoothing method is considered to be an important alternative to the methods of simplex and interior point in quantile regression [14].

Empirical Choice of The Number of Basis As discussed previously, it is important to choose K, an adequate number of PQR basis. Given fixed K, the original PQR formulation can be modified according to the block relaxation ideas [19] which can lead to insightful results and motivate new theory by applying the advanced theory of empirical processes [72].

There are various criteria of selecting K. For example, one may select the value of K to minimize the cross validation (CV) prediction errors or BIC as in choosing the number of fPC basis adapted by Kato and other authors (see [37, 52, 69]. In particular, we have

$$\begin{aligned}
\mathrm{CV}_\tau(\mathrm{K}) &= \frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(y_i - \hat{\alpha}_\tau^{(-i)} - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_\tau^{(-i)} - \sum_{k=1}^{K}\hat{\mathbf{z}}_{ki}\hat{\boldsymbol{\gamma}}_{\tau k}^{(-i)}\right), \\
\mathrm{BIC}_\tau(\mathrm{K}) &= \log\left[\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(y_i - \hat{\alpha}_\tau - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_\tau - \sum_{k=1}^{K}\hat{\mathbf{z}}_{ki}\hat{\boldsymbol{\gamma}}_{\tau k}\right)\right] + \frac{(K+1)\log n}{n},
\end{aligned}$$

where $\hat{\boldsymbol{\gamma}}_{\tau k}^{(-i)}$, $k = 1, \ldots, K$, $\hat{\alpha}^{(-i)}$ and $\hat{\boldsymbol{\beta}}_\tau^{(-i)}$ are computed after removing the i-th observation.

Alternative SIMPQR Algorithm For a given fixed K, the original PQR scheme of SIMPQR [76] can be modified according to the ideas of block relaxation [19], which updates and obtains the basis as a "block" instead of one by one sequentially. Such modification provides an alternative to the original formulation for PQR (APQR) basis which can facilitate establishing and demonstrating the asymptotic properties.

Consider the functional linear quantile regression model (3.2). Let (Y, X, Z) be a triplet of scalar random variable Y, scalar random vector X and a random function $Z = ((Z(t))_{t\in[0,1]}$. For given $\tau \in (0, 1)$, let K be the dimension of the functional space,

into which $\boldsymbol{\gamma}_\tau(t)$ is projected, and $\{\phi_{\tau k}(t)\}_{k=1}^K$ be the functional PQR basis. For each $i = 1, \ldots, n$, Y_i is observed as y_i, while $Z_i(t)$ is observed only at d discrete points $\mathcal{T} = \{0 = t_1 < t_2 < \ldots < t_d = 1\}$, that is, we only observe $z_i(t_j)$, $j = 1, \ldots, d$. We can define $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)^T \in \mathbf{R}^{n\times d}$ with $\mathbf{z}_i = (z_i(t_1), \ldots, z_i(t_d))^T \in \mathbf{R}^d$ for $i = 1, \ldots, n$. If $K = 1$, the PQR basis can be computed as a function $\phi_\tau(t)$ that maximizes the partial quantile covariance (3.8), or more specifically, a vector $\mathbf{c}_\tau = (\phi(t))_{t\in\mathcal{T}}^T \in \mathbf{R}^d$, if we restrict the domain of t to be $\mathcal{T}$, that maximizes

$$l(\alpha, \boldsymbol{\beta}, \mathbf{c}) = -\sum_{i=1}^n \rho_\tau \left(y_i - \alpha - \mathbf{x}_i^T\boldsymbol{\beta} - \mathbf{z}_i^T\mathbf{c}\right), \tag{3.10}$$

For a general K, the PQR basis can be obtained as $\mathbf{C}_\tau \in \mathbf{R}^{d\times K}$ by maximizing

$$l(\alpha, \boldsymbol{\beta}, \mathbf{C}) = -\sum_{i=1}^n \rho_\tau \left(y_i - \alpha - \mathbf{x}_i^T\boldsymbol{\beta} - \mathbf{z}_i^T\mathbf{C}\mathbf{1}_K\right), \tag{3.11}$$

where $\mathbf{C}_\tau = (\mathbf{c}_{\tau 1}, \ldots, \mathbf{c}_{\tau K}) \in \mathbf{R}^{d\times K}$ with $\mathbf{c}_{\tau k} = (\phi_{\tau k}(t))_{t\in\mathcal{T}}^T \in \mathbf{R}^d$ for $k = 1, \ldots, K$, i.e. each column of $\mathbf{C}_\tau$ represents the vector of the functional basis restricted to the discrete domain $\mathcal{T}$.

For a nuisance parameter sequence $\{\nu_N\}_{N=1}^\infty$ where $\nu_N \to \nu_0$, we can choose a uniformly smooth approximating functions $\rho_{\tau\nu_N}$ such that $\rho_{\tau\nu_N}(u) \rightrightarrows \rho_\tau(u)$ as $N \to \infty$. For a given fixed K, by replacing ρ_τ by $\rho_{\tau\nu_N}$ in (3.11), we have

$$l_N(\alpha, \boldsymbol{\beta}, \mathbf{C}) = -\sum_{i=1}^n \rho_{\tau\nu_N} \left(y_i - \alpha - \mathbf{x}_i^T\boldsymbol{\beta} - \mathbf{z}_i^T\mathbf{C}\mathbf{1}_K\right). \tag{3.12}$$

where $l_N(\cdot) \rightrightarrows l(\cdot)$, as $N \to \infty$.

A crucial observation here is that both $l(\cdot)$ and $l_N(\cdot)$ are concave blockwisely about $(\alpha, \boldsymbol{\beta})$ and $\mathbf{C}$. For a preset K, we hereby propose an alternative SIMPQR algorithm (ASIMPQR) based on the idea of block relaxation [19]:

1. *Initialization*: Normalize $Z_i(t_j)$ for each j so that it has mean zero and variance one.

2. *Repeat*:

   (a) *Repeat* (for N):

      i. Initialization: Let $\mathbf{C}_\tau^{(0)}$ be a random matrix, and $(\alpha_\tau^{(0)}, \boldsymbol{\beta}_\tau^{(0)}) = \arg_{\alpha,\boldsymbol{\beta}} \max l_N(\alpha, \boldsymbol{\beta}, \mathbf{C}^{(0)})$,

      ii. Repeat (for p):

         A. $\mathbf{C}_\tau^{(p+1)} = \arg_{\mathbf{C}} \max l_N(\alpha_\tau^{(p)}, \boldsymbol{\beta}_\tau^{(p)}, \mathbf{C})$,

         B. $(\alpha_\tau^{(p+1)}, \boldsymbol{\beta}_\tau^{(p+1)}) = \arg_{\alpha,\boldsymbol{\beta}} \max l_N(\alpha, \boldsymbol{\beta}, \mathbf{C}_\tau^{(p+1)})$.

      iii. Stop (for p): when $l_N(\alpha_\tau^{(p+1)}, \boldsymbol{\beta}_\tau^{(p+1)}, \mathbf{C}_\tau^{(p+1)}) - l_N(\alpha_\tau^{(p)}, \boldsymbol{\beta}_\tau^{(p)}, \mathbf{C}_\tau^{(p)}) < \epsilon$.

      iv. Save: $\alpha_\tau = \alpha_\tau^{(p+1)}$, $\boldsymbol{\beta}_\tau = \boldsymbol{\beta}_\tau^{(p+1)}$ and $\mathbf{C}_\tau = \mathbf{C}_\tau^{(p+1)}$.

   (b) *Stop* (for N): when $l_{N+1}(\alpha_\tau, \boldsymbol{\beta}_\tau, \mathbf{C}_\tau) - l_N(\alpha_\tau, \boldsymbol{\beta}_\tau, \mathbf{C}_\tau) < \epsilon$.

   (c) *Output*: α_τ, $\boldsymbol{\beta}_\tau$ and $\mathbf{C}_\tau$.

3. *Project*: $Z_i(\cdot)$ onto the basis $\phi_{\tau k}(\cdot)$ to obtain projected data matrix $\tilde{\mathbf{Z}}^{(K)} = \mathbf{Z}\mathbf{C}_\tau \in \mathbf{R}^{n\times K}$

4. *Model*: retain the projections $\tilde{\mathbf{Z}}^{(K)} = (\tilde{\mathbf{z}}_1, \ldots, \tilde{\mathbf{z}}_K)$ and form the final predictor $\hat{Q}(\tau|\tilde{\mathbf{Z}}_K)_{\mathbf{Y}|\mathbf{Z}}$.

Algorithm 2: Alternative SIMPQR (For a preset K)

Compared with SIMPQR, the advantages of the proposed ASIMPQR procedure are obvious. Firstly, the $\mathbf{C}_\tau$ of ASIMPQR retains by column the PQR basis vectors which are updated altogether unlike the sequential updating scheme of SIMPQR. Secondly, the block relaxation procedure can guarantee the solutions' stability and convergence whenever the objective function, $l_N(\cdot)$, is convex by block and bounded from above which is easy to verify. Last but not least, the convexity and uniform convergence of the objective functions guarantees that the maximizer of $l_N(\cdot)$ converges to the maximizer of $l(\cdot)$ which assures the $\mathbf{C}_\tau$ obtained from optimizing l_N is close to the true $\mathbf{C}_\tau$. Convergence properties of ASIMPQR are summarized in the following Proposition:

Proposition 3.3.1. *If we have $l_N(\boldsymbol{\upsilon}) \rightrightarrows l(\boldsymbol{\upsilon})$ with $\boldsymbol{\upsilon} = (\alpha, \boldsymbol{\beta}, \mathbf{C})$ as $N \to \infty$, where $l_N(\boldsymbol{\upsilon})$ and $l(\boldsymbol{\upsilon})$ are defined as in (3.12) and (3.11). For a given fixed N, if (i) $l_N(\boldsymbol{\upsilon})$ is continuous, and coercive, that is, the set $\{\boldsymbol{\upsilon} : l_N(\boldsymbol{\upsilon}) \geq l_N(\boldsymbol{\upsilon}^{(0)})\}$ is compact and bounded above, (ii) the objective function in each block update of algorithm is strictly concave, and (iii) the set of stationary points of $l_N(\boldsymbol{\upsilon})$ are isolated, for a given fixed K, we have the following results:*

1. *(Global convergence) The sequence $\boldsymbol{\upsilon}^{(p)} = (\alpha^{(p)}, \boldsymbol{\beta}^{(p)}, \mathbf{C}^{(p)})$ generated by the algorithm above converges to a stationary point of $l_N(\boldsymbol{\upsilon})$.*

2. *(Local convergence) Let $\boldsymbol{\upsilon}^{(\infty)} = (\alpha^{(\infty)}, \boldsymbol{\beta}^{(\infty)}, \mathbf{C}^{(\infty)})$ be a strict local maximum of $l_N(\boldsymbol{\upsilon})$. The iterates generated by the algorithm above are locally attracted to $\boldsymbol{\upsilon}^{(\infty)}$ for $\boldsymbol{\upsilon}^{(0)}$ sufficiently close to $\boldsymbol{\upsilon}^{(\infty)}$.*

3. *(Approximation convergence) The convergence points obtained from $l_N(\boldsymbol{\upsilon})$ will converge in probability to the convergence point of $l(\boldsymbol{\upsilon})$ as $N \to \infty$.*

The assumptions above are not hard to verify if it is allowed to impose some

regular conditions on the distribution functions [38]. If assumptions (i) – (iii) stand, the *local convergence* and *global convergence* can be obtained by following the similar discussions as Li et al. [46] and Zhou et al. [83]. In order to obtain the *approximation convergence*, we can use *Lemma 2* of Hjort and Pollard [33], provided that $l_N(\boldsymbol{\upsilon})$ has a unique maximizer.

Firstly, let us check the assumption (i). For a given fixed N, due to the differentiability of $\rho_{\tau\nu_N}(\cdot)$, then $l_N(\alpha, \boldsymbol{\beta}, \mathbf{C})$, or $l_N(\boldsymbol{\upsilon})$, is obviously continuous. And as $\|\boldsymbol{\upsilon}\| \to \infty$, i.e. $\|(\alpha, \boldsymbol{\beta})\| \to \infty$ or $\|\mathbf{C}\| \to \infty$, the function l_N should go $-\infty$, hence is coercive. To check the assumption (ii), we observe that the function of $y_i - \alpha - \mathbf{x}_i^T\boldsymbol{\beta} - \mathbf{z}_i^T\mathbf{C}\mathbf{1}_K$ is an affine function about $\mathbf{C}$. Since $-\rho_\tau$ and its approximation $-\rho_{\tau\nu_N}$ are strictly concave, $l_N(\boldsymbol{\upsilon})$ and $l(\boldsymbol{\upsilon})$ are both strictly concave about $\mathbf{C}$. Since they are also strictly concave about α, we have the strictly concavity of $l_N(\boldsymbol{\upsilon})$ and $l(\boldsymbol{\upsilon})$ about $\boldsymbol{\upsilon}$. The assumption (iii) is an assumption assuring that the locally optimized point is also an isolated one. One important rule to detect the isolated stationary point is that, if the Hessian matrix of the stationary point is nonsingular, the stationary point should be an isolated one [28]. Alternatively, we can impose *Condition A1* of Koenker [38]: the distribution functions $\{F_i\}$ are absolutely continuous, with continuous density $f_i(\xi)$ uniformly bounded away from 0 and ∞ at the point $\xi_i(\tau)$, where F_i is the conditional distribution of y_i given $\mathbf{z}_i$ and $\xi_i(\tau) = F_i^{-1}(\tau)$. *Lemma 2* of Hjort and Pollard [33] states that if we have that $l_N(\boldsymbol{\upsilon})$ is a sequence of convex functions defined on an open convex set S, in $\mathbf{R}^{Kd+1}$, which convergences to $l(v)$, for each $\boldsymbol{\upsilon}$. Then $\sup_{\boldsymbol{\upsilon}\in K} |l_N(\boldsymbol{\upsilon}) - l(\boldsymbol{\upsilon})|$ goes to zero for each compact subset $\mathbf{K}$ of S. As long as the maxima is an unique interior point of set S, we have that the maxima of l_N will go to the maxima of l.

3.4 Asymptotic Properties

We study the statistical properties of the estimator $\hat{\boldsymbol{\upsilon}}$. For simplicity, we may omit α and $\boldsymbol{\beta}$ at certain step, though the conclusions generalize easily to the case with them. To simplify the notations, we suppress the subindex τ. In this chapter, we adopt the asymptotic setup with a fixed number of K and a diverging sample size n, because this is an important first step toward a comprehensive understanding of the theoretical properties of the proposed model.

For a given fixed K, we first establish score, information and identifiability in term of the proposed alternative formulation of PQR (APQR) for functional linear quantile regression. Then the consistency and asymptotic normality can be derived by applying the advanced theory of empirical processes [72] and following the similar discussions as in Li et al. [46] and Zhou et al. [83].

Score and Information We first derive the score and information for partial quantile regression model. As discussed in Yu and Moyeed [77] and Sánchez et al. [64], the minimization of the quantile loss is equivalent to the maximization of a likelihood function formed by independent and identically distributed asymmetric Laplace densities [77]. In fact, $l(\cdot)$ defined as (3.11) is proportional to the log-likelihood. Therefore, we want to derive the score and information for $l(\cdot)$. Since $l(\cdot)$ is not differentiable, we are giving the score and information for its approximation $l_N(\cdot)$ defined as (3.12). As $N \to \infty$, the differences of the scores and information matrices between l_N and l are almost negligible.

The following standard calculus notations are used. For a scalar function f, ∇f is the (column) gradient vector, $df = [\nabla f]^T$ is the differential. As a direct result from *Lemma 2* of Zhou et al. [83], one can easily check the followings.

Lemma 3.4.1. *Let $\eta_i(\mathbf{C}) = \mathbf{z}_i^T \mathbf{C} \mathbf{1}_K$, then the gradient $\nabla\eta_i(\mathbf{C}) \in \mathbf{R}^{Kd}$ is as*

$$\nabla\eta_i(\mathbf{C}) = -\mathbf{1}_K \otimes \mathrm{vec}\,(\mathbf{z}_\mathrm{i})\,.$$

Proposition 3.4.2. *Consider the partial quantile regression model (3.12):*

1. *The score function (or score vector) of l_N is*

$$\nabla l_N(\mathbf{C}) = -\sum_{i=1}^{n} \rho'_{\tau\nu_N}\left(\eta_i(\mathbf{C})\right) \cdot \nabla\eta_i(\mathbf{C}).$$

2. *The Fisher information matrix of l_N is*

$$\mathbf{I}_N(\mathbf{C}) \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} \rho'_{\tau\nu_N}\left(\eta_i(\mathbf{C})\right) \rho'_{\tau\nu_N}\left(\eta_j(\mathbf{C})\right) \nabla\eta_i(\mathbf{C}) d\eta_j(\mathbf{C}).$$

Identifiability Before studying asymptotic properties, we need to deal with the identifiability issue. The parametrization in the partial linear quantile regression model is non-identifiable mainly due to the complication of the indeterminacy of $\mathbf{C}$ due to permutation: $\mathbf{C\Pi}$ for any $K\times K$ permutation matrix $\mathbf{\Pi}$. To fix the permutation indeterminacy, we assume that first row entries of $\mathbf{C}$ are distinct and arranged in descending order $\mathbf{c}_{11} > \cdots > \mathbf{c}_{K1}$. The resulting parameter space $\{\mathbf{C} : \mathbf{c}_{11} > \cdots > \mathbf{c}_{K1}\}$ is open and convex. Next we give a sufficient and necessary condition for local identifiability.

Proposition 3.4.3 (Identifiability)**.** *Given a sequence of iid data points $\{(y_i, \mathbf{z}_i)\}_{i=1}^n$ from the quantile regression model 3.2. Let $\mathbf{C}_0 \in S \subset \mathbf{R}^{d\times K}$ be a parameter point and assume there exists an open neighbourhood of $\mathbf{C}_0$ in which the information matrix has a constant rank. Then $\mathbf{C}_0$ is identifiable up to permutation if and only if*

$$I_N(\mathbf{C}_0) = \sum_{i=1}^{n}\sum_{j=1}^{n} \rho'_{\tau\nu_N}\left(\eta_i(\mathbf{C}_0)\right)\rho'_{\tau\nu_N}\left(\eta_j(\mathbf{C}_0)\right)\nabla\eta_i(\mathbf{C}_0)d\eta_j(\mathbf{C}_0)$$

is nonsingular.

This result can be obtained by using *Theorem 1* of Rothenberg [63]:

Lemma 3.4.4 (Rothenberg [63], Theorem 1)**.** *Let θ_0 be a regular point of the information matrix $I(\theta)$. Then θ_0 is locally identifiable if and only if $I(\theta_0)$ is nonsingular.*

The regularity assumptions for *Lemma 3.4.4* are satisfied by the partial quantile regression model: (1) the parameter space S is open, (2) the density $p(y, \mathbf{z}|\mathbf{C})$ is proper for all $\mathbf{C} \in S$, (3) the support of the density $p(y, \mathbf{z}|\mathbf{C})$ is same for all $\mathbf{C} \in S$, (4) the log density $l_N(\mathbf{C}|y, \mathbf{z}) = lnp(y, \mathbf{z}|\mathbf{C})$ is continuous differentiable, and (5) the information matrix

$$I_N(\mathbf{C}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \rho'_{\tau\nu_N}\left(\eta_i(\mathbf{C})\right)\rho'_{\tau\nu_N}\left(\eta_j(\mathbf{C})\right)\nabla\eta_i(\mathbf{C})d\eta_i(\mathbf{C})$$

is continuous in $\mathbf{C}$ by *Proposition 3.4.2*. Then by *Lemma 3.4.4* of Rothenberg [63], $\mathbf{C}$ is locally identifiable if and only if $I_N(\mathbf{C})$ is nonsingular.

Asymptotics The asymptotics follow from those discussions for MLE or M-estimation. A key observation is that, by adopting the standard tensor notations from Zhou et al. [83], the original model (3.2) can be rewritten as

$$Y = \alpha + \boldsymbol{\beta}^T\mathbf{X} + \langle \mathbf{C}\mathbf{1}_K, \mathbf{Z}\rangle, \tag{3.13}$$

where the collections of polynomials (of degree 1) $\{\langle \mathbf{C}\mathbf{1}_K, \mathbf{Z}\rangle, \mathbf{C} \in S\}$ forms a Vapnik-Cervonenkis (VC) class [73]. Then standard uniform convergence theory for M-estimation [72] applies.

Theorem 3.4.5 (Consistency)**.** *Assume* $\mathbf{C}_0 \in S \subset \mathbf{R}^{K\times m}$ *is (globally) identifiable up to permutation and the array covariates* $\mathbf{Z}_i$ *are iid from a bounded distribution. The M-estimator is consistent, that is,* $\hat{\mathbf{C}}_n$ *converges to* $\mathbf{C}_0$ *(modulo permutation) in probability for quantile regression model (3.13) with a compact parameter space* $S_0 \subset S$.

The consistency can be checked using the theory of empirical processes. By showing that $\{l_N(\mathbf{C}), \mathbf{C} \in S\}$ is a Donsker class [72] under some regularity conditions when the parameter is restricted to a compact set, the Glivenko-Cantelli theorem establish the uniform convergence. Uniqueness is guaranteed by the information equality whenever $\mathbf{C}_0$ is identifiable.

Theorem 3.4.6 (Asymptotic normality)**.** *For an interior point* $\mathbf{C}_0 \in S$ *with nonsingular information matrix* $I_N(\mathbf{C}_n)$ *and* $\hat{\mathbf{C}}_n$ *is consistent,*

$$\sqrt{n}\left[\mathrm{vec}\left(\hat{\mathbf{C}}_\mathrm{n}\right) - \mathrm{vec}\left(\mathbf{C}_0\right)\right]$$

converges in distribution to a normal with mean zero and covariance $I_N^{-1}(\mathbf{C}_0)$.

3.5 Simulation Studies

In this section, we investigate the finite sample performance of our proposed prediction method, namely alternative partial quantile regression (APQR) method. We compare it with the methods of partial quantile regression (PQR), partial least squares (PLS) and fPC basis (QRfPC) in functional linear quantile regression model. We conduct our simulations in two settings where the first one is in favour of the fPC basis and the second one is a more general case. Both simulations show superior or comparable performance of our proposed method.

Simulation I. In this simulation, we adapt the setup in Kato [37] by slightly changing the weights of parameter $\gamma(t)$. In particular, the model is of the form

$$\begin{aligned} Y &= \int_0^1 \gamma(t)Z(t)dt + \epsilon, \\ \gamma(t) &= \sum_{j=1}^{50} \gamma_j \phi_j(t); \gamma = 0.5, \gamma = \frac{20}{3}(-1)^{j+1} j^{-2}, j \geq 2, \phi_j(t) = 2^{1/2}\cos(j\pi t), \\ Z(t) &= \sum_{j=1}^{50} v_j U_j \phi_j(t), v_j = (-1)^{j+1} j^{-1.1/2}, U_j \sim U[-3^{1/2}, 3^{1/2}]. \end{aligned}$$

Each $Z_i(t)$ was observed at $d = 120$ equally spaced grid points on $[0, 1]$. The error ϵ follows either Gaussian with mean zero and variance one or Cauchy distribution. In each case, we set the number of repetitions to be 500. In each repetition we set the total sample size n to be 300 and randomly split the data into training (80% of the sample) and testing (20% of the sample) sets. The mean absolute error (MAE) of the response is the performance criteria we consider. The optimal cutoff levels K are selected to minimize the 10-fold cross validation prediction errors in each sample, where the value of K appears to be relatively insensitive among different samples. Here APQR1, APQR2, APQR3 and APQR4 represent APQR methods with different initial settings as PQR, PLS, QRfPC and random basis respectively.

Although the simulation design is in favour of fPC based methods, for the small number of cutoff levels, the PQR, PLS and APQR methods perform better while for the optimal number of cutff levels, the PQR, PLS and APQR methods perform comparable with QRfPC method. Due to the natures of sensitivity against skewness of errors, Figure 3.1 show that PLS requires more basis ($K = 5$) to achieve a comparable performance when the errors follow the Cauchy distribution. On the other hand, when the Gaussian errors are employed, PQR, PLS and APQR methods require fewer basis ($K = 2$) compared with the QRfPC method ($K = 5$) to achieve

comparable performance. In this simulation, the performance of APQR appears to be insensitive against the different choices of initial settings.

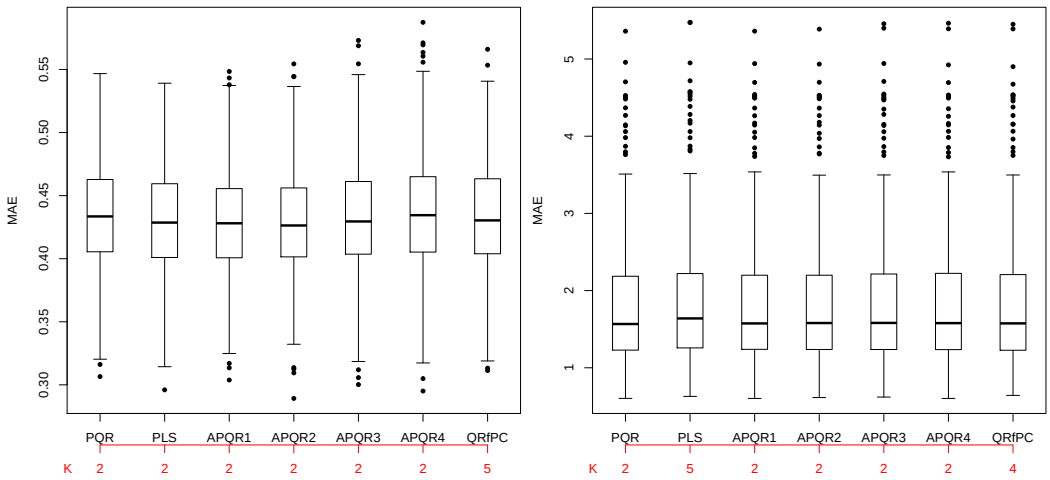


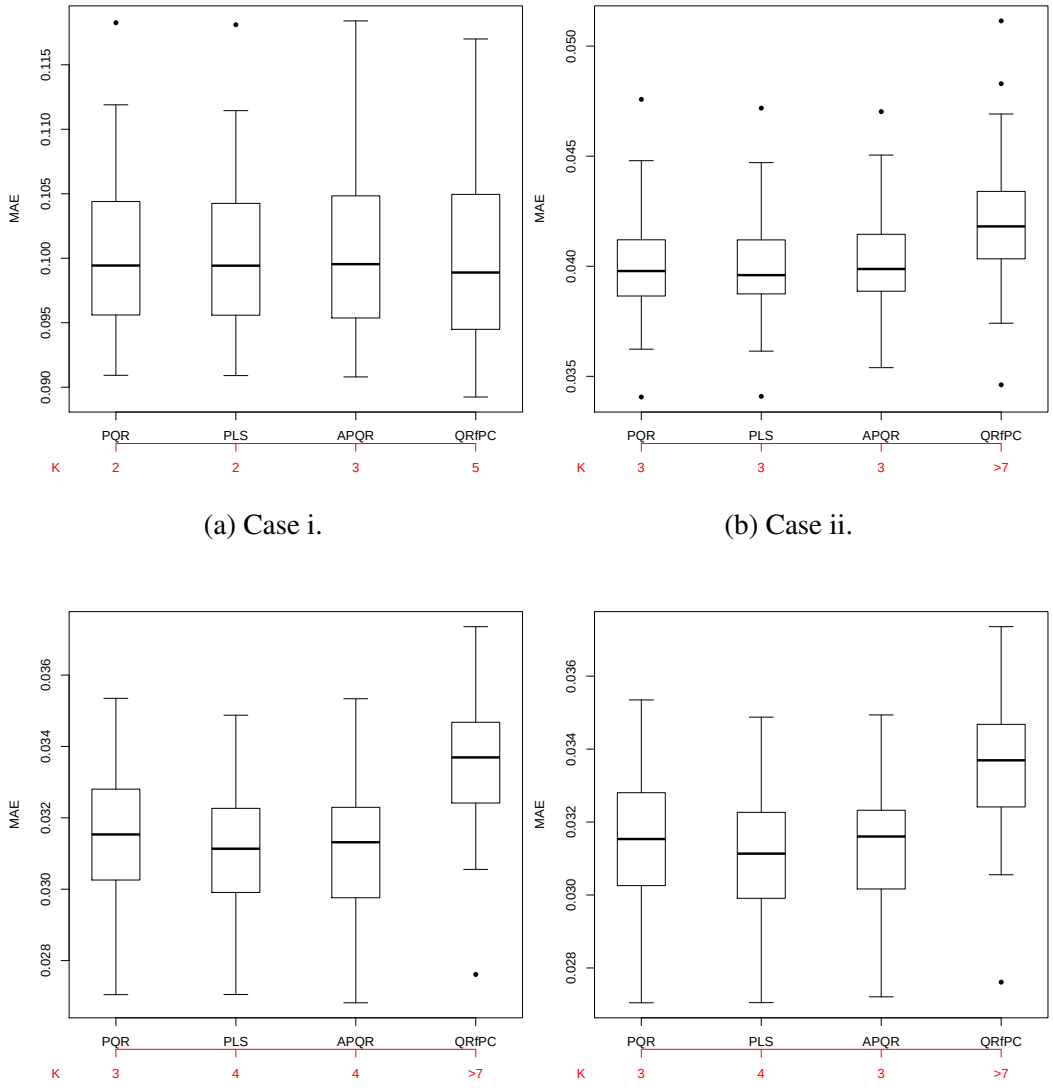
Figure 3.1: Simulation I: the boxplots of mean absolute errors (MAE) with Gaussian (Left) and Cauchy (Right) errors. In each case, there are 500 repetitions with training sample size of 240 and testing sample size of 60. APQR1, APQR2, APQR3 and APQR4 represent APQR methods with different initial settings as PQR, PLS, QRfPC and random basis respectively.

Simulation II. In this simulation, we take the $Z_i(t)$s from a real data study, a benchmark Phoneme dataset, which can be downloaded from `http://statweb.stanford.edu/tibs/ElemStatLearn/`, and generate the Y according to the linear model of

$$Y = \int_0^1 Z(t)\gamma(t)dt + \varepsilon,$$

where the error ε is taken as Gaussian. The centres of errors are taken as zero while the scales are taken as the empirical standard deviation of the true responses multiplied by $\sqrt{5}$. In our setting, each $Z_i(t)$ was observed at $d = 256$ equally spaced grid points on $[0, 1]$ and the total number of observations is 1717. This example together with the detailed background can also be found in Delaigle and Hall [20].

Computing the first $J = 20$ empirical fPC basis functions $\hat{\phi}_1(t), \ldots, \hat{\phi}_J(t)$, we consider four different curves $\gamma(t)$ by taking $\gamma(t) = \sum_{j=1}^{J} a_j \hat{\phi}_j(t)$ for four different sequence of a_js: (i) $a_j = (-1)^j \cdot \mathbf{1}\{0 \leq j \leq 5\}$; (ii) $a_j = (-1)^j \cdot \mathbf{1}\{6 \leq j \leq 10\}$; (iii) $a_j = (-1)^j \cdot \mathbf{1}\{11 \leq j \leq 15\}$; (iv) $a_j = (-1)^j \cdot \mathbf{1}\{16 \leq j \leq 20\}$. Going through case (i) to (iv), the models become less favorable for fPC, while we will see the PLS, PQR and APQR methods manage to capture the interaction between Z and Y using only a few terms.



(a) Case i.

(b) Case ii.

(c) Case iii.

(d) Case iv.

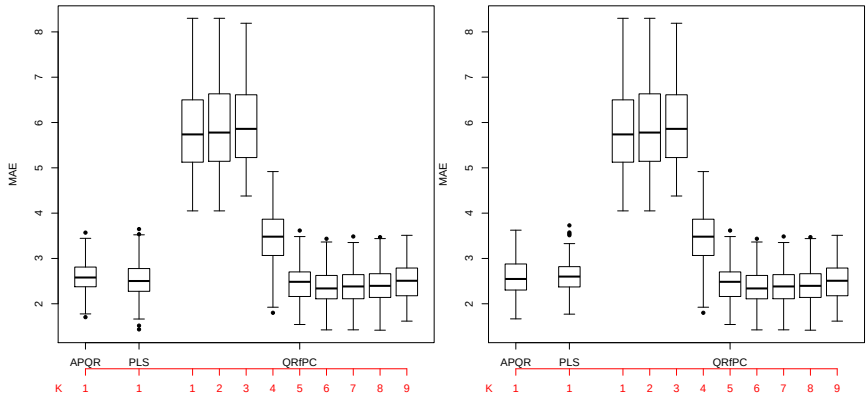
Figure 3.2: Simulation II: the boxplots of mean absolute errors (MAE) with Gaussian errors. The centres of errors are taken as zero while the scales are taken as the empirical standard deviation of the true responses multiplied by $\sqrt{5}$. there are 100 repetitions with training sample size of 1517 and testing sample size of 200.

After responses being generated, we randomly split the data set into training (sample size 1517) and testing (sample size 200) data set and repeat it 100 times. The optimal cutoff levels K are selected using the same criteria as in simulation

I. Figure 3.2 displays the boxplot of the mean absolute prediction errors when the errors follow Gaussian distribution. The PLS, PQR and APQR methods predict better in general compared with QRfPC method except for case (i). While the optimal K are selected consistently around 3 for the PLS, PQR and APQR methods while the QRfPC method require more and more basis from case (i) to (iv). In addition, It is worth noticing that performance of APQR is not always insensitive against the initial values. In fact, the random initial setting of APQR does not perform well in this simulation.

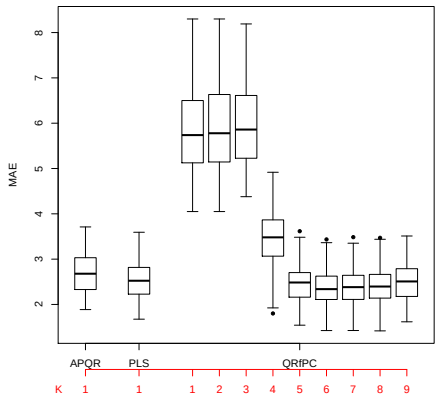
3.6 Real Data Analysis

Real Data I: ADHD-200 fMRI Data. We apply our proposed method to a dataset on attention deficit hyperactivity disorder (ADHD) from the ADHD-200 Sample Initiative Project. The response of interest is ADHD index. The scalar covariates of primary interest include gender, age, handedness, diagnosis status, medication status, Verbal IQ, Performance IQ and Full4 IQ. The functional covariates are the average gray scales of 172 equally spaced time points for cerebelum, temporal, vermis, parietal occipital and frontal out of 116 Regions of Interests (ROI). We randomly split the data set (n=120) into training (80% of the sample size) and testing (20% of the sample size) data set and repeat it 100 times. The optimal cutoff levels K are selected using CV criteria as in simulations. As shown in figure 3.3, our proposed method, APQR, together with the PLS method achieve comparable performance by using only 1 basis while the QRfPC method require more than 5 basis to outperform our method.

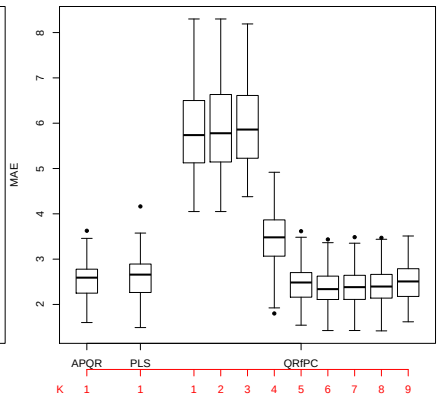


(a) Cerebelum.

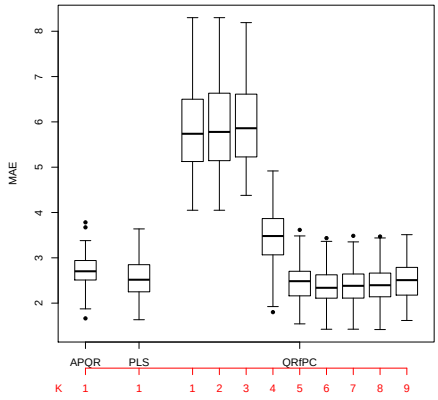
(b) Temporal.



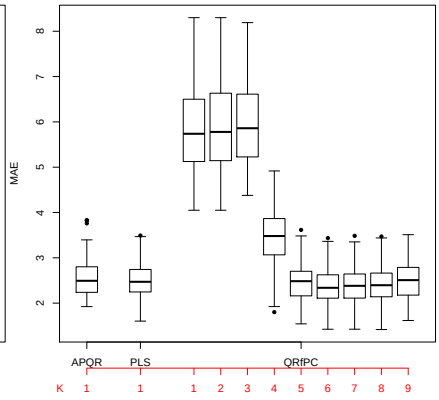
(c) Vermis.



(d) Parietal.



(e) Occipital.



(f) Frontal.

Figure 3.3: Real Data Analysis I: ADHD-200 fMRI data. Comparison of the box-plots of mean absolute errors (MAE) using three different methods of APQR, PLS and QRfPC for each brain part of cerebelum, temporal, vermis, parietal, occipital and frontal.

Real Data II: ADNI DTI Data. We use our model methods to analyze a real DTI data set with 214 subjects collected from NIH Alzheimer's Disease Neuroimaging Initiative (ADNI) study. Data used in the preparation of this chapter were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`adni.loni.ucla.edu`). We are interested in predicting mini-mental state examination (MMSE) scores, one of the most widely used screening tests, which are used to provide brief, objective measures of cognitive functioning for almost fifty years. we include 200 subjects from the total 214 subjects. The functional covariate is fractional anisotropy (FA) values along the corpus callosum (CC) fiber tract with 83 equally spaced grid points, which can be treated as a function of arc-length. The scale covariates are gender, age, education levels, an indicator for Alzheimer's disease (AD) status and an indicator for mild cognitive impairment (MCI) status , and genotypes for apolipoprotein E ϵ-4.

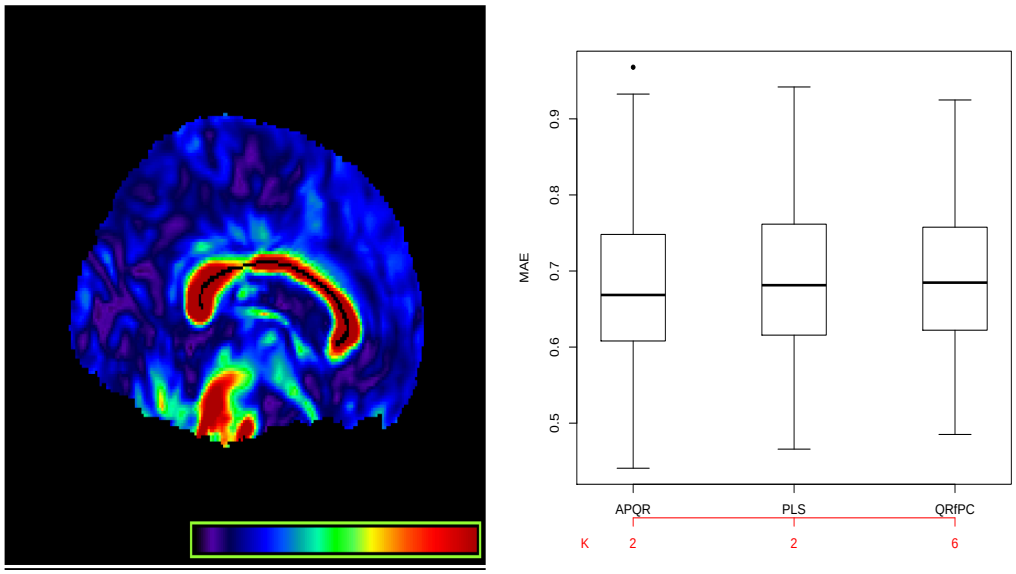


Figure 3.4: Real Data Analysis II: (Left) The midsagittal corpus callosum (CC) skeleton overlaid with fractional anisotropy (FA) from one randomly selected subject and (Right) the boxplot of mean absolute errors of mini-mental state examination (MMSE) with optimal cutoff levels.

We randomly split the data set (n=200) into training (80% of the sample size) and testing (20% of the sample size) data set and repeat it 100 times. The boxplot of MAEs for the optimal cutoff levels K, selected by the same criteria before, is shown in figure 3.4. In general APQR and PLS methods preform consistently better than QRfPC method while APQR outperforms PLS method. Similar to the phenomenon observed from the previous real data analysis, APQR and PLS are able to use fewer number of basis ($K = 2$) and achieve smaller prediction errors compared with the QRfPC method ($K = 6$). This indicates that the fPC basis may not always be suitable to do prediction especially when the number of basis is restricted to be small. In summary, APQR is capable of making better prediction by using fewer basis functions hence provide a powerful tool to do prediction in practice.

3.7 Discussions

In this chapter, we first define the concept of partial quantile covariance (PQC) which measures the contribution of one covariate towards the response, and describe the partial quantile regression (PQR) method to extract PQR basis in functional linear quantile regression model, which is motivated by the success of the partial least square (PLS) basis in functional linear regression model. We then discuss the problems of deriving the asymptotic properties for the original PQR methods and address them by proposing and implementing an alternative formulation to original PQR method (APQR) using finite smoothing techniques and block relaxation ideas for a preset K. In addition, we suggest an empirical guideline of selecting K by using cross validation (CV) and BIC. The proposed APQR can facilitate us to derive score, information and identifiability in term of functional linear quantile regression, so that the estimator consistency and convergence rates of $\sqrt{n}$

can be established and demonstrated by using the advanced theory of empirical processes.

The simulations show that APQR, PQR and PLS in general perform comparable to each other and better than the QRfPC methods. It is also worth noticing that, the performance of APQR is sometimes sensitive against the initial values. Our proposed APQR method can make comparable prediction errors with QRfPC basis method by using an extremely small number of basis in both ADHD-200 fMRI data analysis ($K = 1$) and ADNI DTI data analysis ($K = 2$). In ADNI DTI data analysis, APQR outperforms QRfPC method even though the latter method use more basis ($K = 6$). This indicates that the fPC basis may not always be suitable to do prediction especially when the number of basis is restricted to be small. In total, APQR is capable of making better prediction by using fewer basis functions hence provide a powerful tool to do prediction in practice.

The most important contribution of this chapter is that by proposing an alternative but equivalent formulation for PQR method (APQR), for a given fixed K, we manage to establish and demonstrate the asymptotic properties such as consistency and asymptotic normality, which used to be a very difficult problem in the original PQR setting.

In both simulation studies and real data analysis, only univariate functional covariate case is considered. However, the extension of PQR to multivariate functional covariates is straightforward. In addition, using the similar idea of this chapter, one may establish the alternative formulation of composite partial quantile regression (PCQR) which has been shown to be more robust and efficient than the PQR method. Further details are out of the scope of this chapter and will be pursuit in the future research.

3.8 Appendix

Proof of Theorem 3.4.5

Proof. For simplicity, we omit α and $\boldsymbol{\beta}$, though the conclusions generalize easily to the case with them. we want to show that the consistency of the estimated factor matrix $\hat{\mathbf{C}}_n$. The following well-known theorem is our major tool for establishing consistency.

Lemma 3.8.1 (van der Vaart [72], Theorem 5.7)**.** *Let M_n be random functions and let M be a fixed function of θ such that for every $\epsilon > 0$*

$$\begin{aligned} \sup_{\theta \in \nu} |M_n(\theta) - M(\theta)| &\rightarrow 0 \quad \textit{in probability}, \\ \sup_{\theta : d(\theta, \theta_0) \geq \epsilon} M(\theta) &< M(\theta_0). \end{aligned}$$

Then any sequence of estimators $\hat{\theta}_n$ with $M_n(\hat{\theta}) \geq M_n(\theta_0) - o_P(1)$ converges in probability to θ_0.

To apply *Lemma 3.8.1* in our setting, we take the nonrandom function M to be $\mathbf{C} \mapsto \mathbb{P}_{\mathbf{C}_0}[l_N(Y, Z|\mathbf{C})]$ and the sequence of random functions to be $M_n : \mathbf{C} \mapsto \frac{1}{n}\sum_{i=1}^{n} l_N(y_i, z_i|\mathbf{C}) = \mathbb{P}_n M$, where $\mathbb{P}_n$ denotes the empirical measure under $\mathbf{C}_0$. Then M_n converges to M a.s. by strong law of large number. The second condition requires that $\mathbf{C}_0$ is a well-separated maximum of M. This is guaranteed by the (global) identifiability of $\mathbf{C}_0$ and information inequality. The first uniform convergence condition is most convenient and is verified by the Glivenko-Cantelli theory [72].

The density is $p_{\mathbf{C}}(y|\mathbf{z}) = const \cdot \exp\left[-\rho_{\tau\nu_N}(y - \eta(\mathbf{C}, \mathbf{z}))\right]$ where $\eta(\mathbf{C}, \mathbf{z}) = \langle \mathbf{C}, \mathbf{z} \rangle$.

Take $m_{\mathbf{C}} = ln\left[(p_{\mathbf{C}} + p_{\mathbf{C}_0})/2\right]$. First we show that $\mathbf{C}_0$ is a well-separated maximum of the function $M(\mathbf{C}) := \mathbb{P}_{\mathbf{C}_0} m_{\mathbf{C}}$. The global identifiability of $\mathbf{C}_0$ and information inequality guarantee that $\mathbf{C}_0$ is the unique maximum of M. To show that it is a well-separated maximum, we need to verify that $M(\mathbf{C}_k) \to M(\mathbf{C}_0)$ implies $\mathbf{C}_k \to \mathbf{C}_0$.

Suppose $M(\mathbf{C}_k) \to M(\mathbf{C}_0)$, then $\langle \mathbf{C}_k, \mathbf{Z} \rangle \to \langle \mathbf{C}_0, \mathbf{Z} \rangle$ in probability. If $\mathbf{C}_k$ are bounded, then $\mathbf{E}\left[\langle \mathbf{C}_k - \mathbf{C}_0, \mathbf{Z} \rangle^2\right] \to 0$ and $\mathbf{C}_k \to \mathbf{C}_0$ by nonsigularity of $\mathbf{E}\left[(\text{vec}\mathbf{Z})(\text{vec}\mathbf{Z})^T\right]$. On the other hand, $\mathbf{C}_k$ can not run to infinity. If they do, then $\langle \mathbf{C}_k, \mathbf{Z} \rangle / \|\mathbf{C}_k\| \to 0$ in probability which in turn implies that $\mathbf{C}_k / \|C_k\| \to \mathbf{0}$.

For the uniform convergence, we see that the class of functions $\{\langle \mathbf{C}, \mathbf{Z} \rangle, \mathbf{C} \in S\}$ forms a VC class. This is true because it is collection of number of polynomials of degree 1 and then apply the VC vector space argument (van der Vaart and Wellner [73], 2.6.15). This implies that $\{\eta(\langle \mathbf{C}, \mathbf{Z} \rangle), \mathbf{C} \in S\}$ is a VC class since η is a monotone function (van der Vaart and Wellner [73], 2.6.18).

Now $m_{\mathbf{C}}$ is Lipschitz in η since

$$\begin{aligned}\frac{\partial m_{\mathbf{C}}}{\partial \eta} &= \frac{const \cdot \exp\left[-\rho_{\tau\nu_N}(y-\eta)\right] \cdot \rho'_{\tau\nu_N}(y-\eta)}{const \cdot \exp\left[-\rho_{\tau\nu_N}(y-\eta)\right] + const \cdot \exp\left[-\rho_{\tau\nu_N}(y-\eta_0)\right]} \\ &= \frac{\rho'_{\tau\nu_N}(y-\eta)}{1 + \exp\left[\rho_{\tau\nu_N}(y-\eta) - \rho_{\tau\nu_N}(y-\eta_0)\right]} \leq \sup\left|\rho'_{\tau\nu_N}(\cdot)\right| = const\end{aligned}$$

The last equality holds since $\rho_{\tau\nu_N}(u) \to \rho_\tau(u)$ as $N \to \infty$, which also implies that $\rho'_{\tau\nu_N}(u) \to \rho'_\tau(u)$ as $N \to \infty$ except for $u = 0$. And we know that $\rho'_{\tau\nu_N}(0) = 0$. Similarly we can show that $m_{\mathbf{C}}$ is Lipschitz in η_0. A Lipschitz composition of a Donsker class is still a a Donsker class (van der Vaart [72], 19.20). Therefore $\{\mathbf{C} \mapsto m_{\mathbf{C}}\}$ is a bounded Donsker class with the trivial envelope function 1. A Donsker class is certainly a Glivenko-Cantelli class. Finally the Glivenko-Cantelli theorem establishes the uniform convergence condition required by *Lemma 3.8.1*.

When the parameter is restricted to a compact set, $\eta(\langle \mathbf{C}, \mathbf{Z} \rangle)$ is confined in

a bounded interval and the l_N is Lipschitz on the finite interval. It follows that $\{l_N(\mathbf{C}) = l_N \circ \eta \circ \langle \mathbf{C}, \mathbf{Z} \rangle, \mathbf{C} \in S\}$ is a Donsker class as composition with a monotone or Lipschitz function preserves the Donsker class. Therefore the Glivenko-Cantelli theorem establish the uniform convergence. Compactness of parameter space implies that $\mathbf{C}_0$ is a well separated maximum if it is the unique maximizer of $M(\mathbf{C}) = \mathbb{P}_{\mathbf{C}_0} m_{\mathbf{C}}$ (van der Vaart [72], Exercise 5.27). Uniqueness is guaranteed by the information equality whenever $\mathbf{C}_0$ is identifiable. This verifies the consistency for quantile regression.

□

Lemma 3.8.2. *Tensor quantile linear regression model (3.13) is quadratic mean differentiable (q.m.d.).*

Proof. By a well-known result (van der Vaart [72], Lemma 7.6), it suffices to verify that the density is continuously differentiable in parameter for μ-almost all x and that the Fisher information matrix exists and it continuous. The derivative of density is

$$\nabla l_N(\mathbf{C}) = -\sum_{i=1}^{n} \rho'_{\tau\nu_N}(\eta_i(\mathbf{C})) \cdot \nabla \eta_i(\mathbf{C}),$$

which is well-defined and continuous by Proposition 2. The same proposition shows that the information matrix exists and is continuous. Therefore the tensor quantile linear regression model is q.m.d. □

Proof of Theorem 3.4.6

Proof. The following result relates asymptotic normality to the density that satisfy q.m.d.

Lemma 3.8.3. *At an inner point θ_0 of $\nu \subset \mathbf{R}^k$. Furthermore, suppose that there exists a measurable function $\dot{l}$ with $\mathbf{P}_{\theta_0}\dot{l}^2 < \infty$ such that, for every θ_1 and θ_2 in a neighbourhood of θ_0,*

$$\left|lnp_{\theta_1}(x) - lnp_{\theta_2}(x)\right| \leq \dot{l}(x)\,\|\theta_1 - \theta_2\|\,.$$

If the Fisher information matrix I_{θ_0} is nonsingular and $\hat{\theta}_n$ is consistent, then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = I_{\theta_0}^{-1} \sum_{i=1}^{n} \dot{l}_{\theta_0}(X_i) + o_{P_{\theta_0}}(1).$$

In particular, the sequence $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically normal with mean zero and covariance matrix $I_{\theta_0}^{-1}$.

Lemma 3.8.2 shows that tensor quantile regression linear model is q.m.d. By *proposition 3.4.2* and chain rule, the score function

$$\dot{l}_N(\mathbf{C}) = -\sum_{i=1}^{n} \rho'_{\tau\nu_N}\left(\eta_i(\mathbf{C})\right) \cdot \nabla\eta_i(\mathbf{C})$$

is uniformly bounded in y an $\mathbf{x}$ and continuous in $\mathbf{C}$ for every y and $\mathbf{x}$ with $\mathbf{C}$ ranging over a compact set of S_0. For sufficiently small neighbourhood U of S_0, $\sup_U \left\|\dot{l}_N(\mathbf{C})\right\|$ is square-integrable. Thus the local Lipschitz condition is satisfied and *Lemma 3.8.3* applies.

□

Chapter 4

Partial Quantile Regression for Multidimensional Functional Linear Model

Partial quantile regression (PQR) is a robust procedure for functional linear regression model by using partial quantile covariance techniques to extract supervised basis for estimating functional coefficients. Although it has been previously illustrated that PQR can deal with functional covariates with single variable very well, it is still not clear that how it can be extended and implemented for multidimensional functional covariates. In this chapter, we propose and implement the generalization of PQR procedure to multidimensional functional linear model using tensor decomposition techniques and block relaxation ideas. We also establish and demonstrate the corresponding asymptotic properties by applying advanced techniques from empirical process theory.

4.1 Introduction

Functional data analysis (FDA) is about the analysis of information on curves, images, functions or more general objects where the primary object of observation can be viewed as a function [61]. It has become a major branch of nonparametric statistics while is still fast evolving as more data of larger scale and more complex structure emerge. As a popular tool, functional linear regression model is often considered useful by statisticians to deal with such data [16, 29, 44, 80]. A typical functional linear model with scalar response is

$$y = \alpha + \mathbf{x}^T\boldsymbol{\beta} + \int_0^1 \mathbf{z}^T(t)\boldsymbol{\gamma}(t)dt + \varepsilon, \tag{4.1}$$

where $\mathbf{x}$ and $\mathbf{z}(t)$ are scalar and functional covariates, and the contributions of $\mathbf{z}(t)$ towards the variation of y is characterized by the functional coefficients $\boldsymbol{\gamma}(t)$ and change by t. Note that this is the same model as (3.1) in Chapter 3. It has been many extensions about this model; for instance, Lian [48] considered this model in the setting with multiple functional covariates $\mathbf{z}(t)$, while Kong et al. [43] discussed the situation of ultrahigh-dimensional scalar covariates $\mathbf{x}$.

In recent years, the extension of model (4.1) to quantile regression [41] has been well developed and recognized for functional linear model. In particular, let us consider a functional linear quantile regression model, which is also the same as (3.2) in Chapter 3:

$$Q_\tau(y|\mathbf{x}, \mathbf{z}(t)) = \alpha_\tau + \mathbf{x}^T\boldsymbol{\beta}_\tau + \int_0^1 \mathbf{z}(t)^T\boldsymbol{\gamma}_\tau(t)dt, \tag{4.2}$$

where $Q_\tau(y|\mathbf{x}, \mathbf{z}(t))$ is the τ-th conditional quantile of response y given a functional

covariate $z(t)$ for a fixed quantile level $\tau \in (0, 1)$. As an alternative to least squares regression, the quantile regression method is more efficient and robust when the responses are non-normal, errors are heavy tailed or outliers are present. It is also capable of dealing with the heteroscedasticity issues and providing a more complete picture of the response [38].

To facilitate the estimation of $\boldsymbol{\gamma}(t)$ and $\boldsymbol{\gamma}_\tau(t)$, we usually require that them satisfy certain smoothness conditions and restrict them onto a functional space. For example, we may require that their second derivatives exist [36] and they are square integrable [37]. Even in such a situation, the estimation is still an infinite-dimensional problem. The common practice is to project them into a space spanned by a finite number of functional basis then use these basis to approximate them. For least square regression, three major choices of such basis include: general basis such as B-spline basis and wavelet basis [12, 79], functional principal component basis [8, 10, 43, 55], and partial least square basis [20]. Analogously in quantile regression, general basis like B-spline basis [11, 68], functional principle component basis [37, 52, 69] and partial quantile basis [76] have also been thoroughly investigated.

The functional covariate $\mathbf{z}$ can also be taken as a multivariable function, i.e. its input is made up of multiple numbers. In particular, without loss of generality, let $\mathbf{t} \in [0, 1]^q$ where $q \geq 2$ and $z(\mathbf{t}) \in \mathbf{R}$. For a scalar response, a q-dimensional functional linear regression (qD-FLR) model is of the form:

$$y = \alpha + \mathbf{x}^T\boldsymbol{\beta} + \int_0^1 \cdots \int_0^1 z(t_1, \ldots, t_q)\gamma(t_1, \ldots, t_q)dt_q \cdots dt_1 + \varepsilon, \tag{4.3}$$

where α is the intercept, $\boldsymbol{\beta}$ is a p-dimensional vector of coefficients, $\gamma(t_1, \ldots, t_q)$ is a q-dimensional functional coefficient, and ε is an error term usually with zero mean and finite variance.

Under the context of quantile regression, analogous to model (4.3), and for a given quantile level $\tau \in (0, 1)$, the q-dimensional functional linear quantile regression (qD-FLQR) model is defined as:

$$Q_{y|x,z}(\tau|x, z) = \alpha_\tau + \mathbf{x}^T\boldsymbol{\beta}_\tau + \int_0^1 \cdots \int_0^1 z(t_1, \ldots, t_q)\gamma_\tau(t_1, \ldots, t_q)dt_q \cdots dt_1, \quad (4.4)$$

where $Q_{y|\mathbf{x},z}(\tau|x, z)$ is τ-th conditional quantile of y given scalar covariates $\mathbf{x}$ and multidimensional functional z.

Having $z(\mathbf{t})$ properly discretized, we can use $\mathbf{Z} \in \mathbf{R}^{I_1 \times \cdots \times I_q}$, a q-dimensional array, to represent $z(\mathbf{t})$, where the j-th element of $\mathbf{t}$ is observed at $0 = t_1 < \cdots < t_{I_j} = 1$. Such data of q-dimensional array, also known as order-q tensor, is quite common in medical imaging. A notable example is magnetic resonance imaging (MRI) data where the anatomical MRI images can be observed as a matrix (order-2 tensor) of size 256 by 256. To predict certain clinical outcomes using tensor, a naive attempt to re-arrange $\mathbf{Z}$ into a vector then perform regression. However such practice is evidently unsatisfactory. First, the re-arranged (vectorized) vector of image covariates is of size $256^2 = 65,536$, implicitly requiring large number of regression parameters. Both computational cost and theoretical properties can be severely compromised due to such ultra-high dimensional setting. Furthermore, the vectorized $\mathbf{Z}$ loses information about the original data structure, so the regression model based on it would be lack of efficiency and hard to interpret.

For an order-q random tensor $\mathbf{Z} \in \mathbf{R}^{I_1 \times \cdots \times I_q}$, in order to achieve a similar decomposition as PC decomposition for vectors, i.e. order-1 tensors, Lu et al. [51] proposed a framework to conduct multilinear PC (MPC) decomposition. Its intention was to project an order-q tensor into a lower-dimensional space spanned by the product of a few feature vectors. In general, there are two ways to decompose a

tensor: CANDECOMP/PARAFAC (CP) decomposition [83] and Tucker decomposition [46]. We can use Figure 4.1 adapted from Cichocki et al. [15] to illustrate the ideas. Figure 4.1 (top) shows the way of conducting CP decomposition for an order-

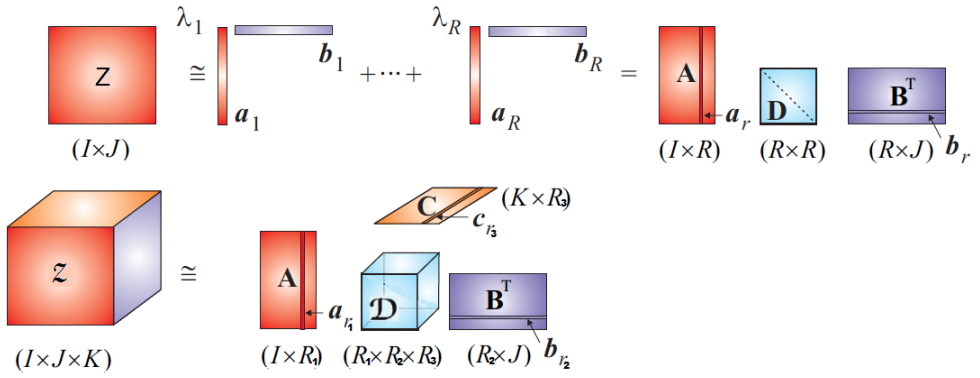


Figure 4.1: Analogy between CP (above) and Tucker (bottom) decompositions.

2 tensor $\mathbf{Z} \in \mathbf{R}^{I \times J}$, which is in fact a rank R representation of $\mathbf{Z}$ by $\sum_{r=1}^{R} d_r \cdot \mathbf{a}_r \circ \mathbf{b}_r$. Here the multidimensional outer product, $\mathbf{g}_1 \circ \cdots \circ \mathbf{g}_q$, of q vectors $\mathbf{g}_j \in \mathbf{R}^{I_j}$, where $j = 1, \ldots, q$, is an $I_1 \times \cdots \times I_q$ array with entries $(\mathbf{b}_1 \circ \cdots \circ \mathbf{b}_q)_{i_1,\cdots,i_q} = \prod_{j=1}^{q} b_{j,i_j}$. As an alternative to CP, Tucker decomposition for an order-3 tensor $\mathbf{Z} \in \mathbf{R}^{I \times J \times K}$ is shown as Figure 4.1 (bottom). It is to approximate a tensor $\mathbf{Z}$ by $\sum_{i=1}^{R_1} \sum_{j=1}^{R_2} \sum_{k=1}^{R_3} d_{ijk} \cdot \mathbf{a}_i \circ \mathbf{b}_j \circ \mathbf{c}_k$ where $\mathbf{a}_i \in \mathbf{R}^I$, $\mathbf{b}_j \in \mathbf{R}^J$ and $\mathbf{c}_k \in \mathbf{R}^K$. We can also abbreviate it by $[\![\mathbf{D}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$, where $[\![\mathbf{D}; \mathbf{G}_1, \ldots, \mathbf{G}_q]\!] = \sum_{r_1=1}^{R_1} \cdots \sum_{r_q=1}^{g_q} d_{r_1 \ldots r_q} \cdot \mathbf{g}_{r_1} \circ \cdots \circ \mathbf{g}_{r_q}$ for $\mathbf{D} = \{d_{r_1 \ldots r_q}\}_{r_1=1,\ldots,r_q=1}^{R_1,\ldots,R_q} \in \mathbf{R}^{I_1 \times \cdots \times I_q}$, and $\mathbf{G}_j = (\mathbf{g}_1, \ldots, \mathbf{g}_{R_j}) \in \mathbf{R}^{I_j \times R_j}$ for $j = 1, \ldots, q$. Here $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are matrices of arranging by column vectors $\mathbf{a}_i$s, $\mathbf{b}_j$s and $\mathbf{c}_k$s respectively, and $\mathbf{D}$ is the loading tensor (core tensor). The column numbers of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are R_1, R_2 and R_3 respectively which are ranks in each dimension and typically much smaller than I, J and K. A CP decomposition is in fact a special case of Tucker decomposition where the core tensor is diagonal. In particular, for $q = 2$, $\mathbf{Z} \approx \sum_{r=1}^{R} d_r \cdot \mathbf{a}_r \circ \mathbf{b}_r = [\![\mathbf{D}; \mathbf{A}, \mathbf{B}]\!]$, where $\mathbf{D} = \{d_{ij}\}_{i=1,j=1}^{R,R} \in \mathbf{R}^{R \times R}$ with $d_{ij} = 0$ if $i \neq j$.

In this chapter, we mainly focus on the more general case of Tucker decomposition, while the discussion for CP decomposition case would be similar.

Since $\mathbf{Z}$ is just a representaion of $z(\mathbf{t})$, the functional MPC decomposition of $z(\mathbf{t})$ is just a natural generalization of the MPC decomposition of $\mathbf{Z}$. In particular, denote by $z(\mathbf{t})$ and $\gamma(\mathbf{t}) \in \mathbf{R}$, the multidimensional function covariates and coefficients, with $\mathbf{t} \in [0,1]^q$. Suppose $\gamma(\mathbf{t})$ admits a Karhunen-Loève(KL) representation by:

$$\sum_{n_1} \cdots \sum_{n_q} \lambda_{n_1,\ldots,n_q} \phi^{(1)}_{n_1}(t_1) \cdot \cdots \cdot \phi^{(q)}_{n_q}(t_q), \tag{4.5}$$

where $\phi^{(d)}_{n_d}(t_d)$'s, for $d = 1,\ldots,q$, are the *eigenfunction*s in d-th dimension, and $\lambda_{n_1,\ldots,n_q}$ are the *eigenvalue*s.

Let $\phi^{(d)}_{n_d}(t_d)$'s, for $d = 1,\ldots,q$, be the *eigenfunction*s in d-th dimension. Then we have

$$\begin{aligned}
&\int_0^1 \cdots \int_0^1 z(t_1,\ldots,t_q)\gamma(t_1,\ldots,t_q)dt_q \cdots dt_1 \\
= &\int_0^1 \cdots \int_0^1 \sum_{n_1} \cdots \sum_{n_q} \lambda_{n_1,\ldots,n_q} \phi^{(1)}_{n_1}(t_1) \cdot \cdots \cdot \phi^{(q)}_{n_q}(t_q) z(t_1,\ldots,t_q)dt_q \cdots dt_1 \\
= &\sum_{n_1} \cdots \sum_{n_q} \lambda_{n_1,\ldots,n_q} \zeta_{n_1,\ldots,n_q},
\end{aligned}$$

where $\zeta_{n_1,\ldots,n_q} = \int_0^1 \cdots \int_0^1 \phi^{(1)}_{n_1}(t_1) \cdot \cdots \cdot \phi^{(q)}_{n_q}(t_q) z(t_1,\ldots,t_q)dt_q \cdots dt_1$. Hence model (4.3) becomes an order-q tensor linear regression (qD-TLR) model:

$$y = \alpha + \mathbf{x}^T\boldsymbol{\beta} + \sum_{n_1} \cdots \sum_{n_q} \lambda_{n_1,\ldots,n_q} \zeta_{n_1,\ldots,n_q} + \varepsilon. \tag{4.6}$$

For a preset $R_1,\ldots,R_q$, namely the ranks for each dimension, denote $\{\zeta_{n_1,\ldots,n_q}\}^{R_1,\ldots,R_q}_{n_1=1,\ldots,n_q=1}$ and $\{\lambda_{n_1,\ldots,n_q}\}^{R_1,\ldots,R_q}_{n_1=1,\ldots,n_q=1}$ by ζ and $\boldsymbol{\Lambda}$ respectively, we can abbreviate model (4.6) by

rewriting it as

$$y \quad = \quad \alpha + \langle \mathbf{x}, \boldsymbol{\beta} \rangle + \langle \zeta, \boldsymbol{\Lambda} \rangle + \varepsilon, \tag{4.7}$$

where the inner product $\langle \cdot, \cdot \rangle$ between two tensors is induced from the inner product of two vectors, $\mathbf{x}$ and $\mathbf{y}$, where $\mathbf{x}$ and $\mathbf{y} \in \mathbf{R}^n$, with $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^{n} x_i y_i$. Then, for two arrays $\mathbf{X}$ and $\mathbf{Y} \in \mathbf{R}^{I_1 \times \cdots \times I_q}$, the inner product is defined as $\langle \mathbf{X}, \mathbf{Y} \rangle = \langle \text{vec}(\mathbf{X}), \text{vec}(\mathbf{Y}) \rangle$, where the operator of vec stacks the entries of a tensor into a column vector in the same manner as defined in Zhou et al. [83], which has also been used across all the tensor literature.

Since in practice, $z(t_1, \ldots, t_q)$ is observed at the grid points $\left\{(t_{n_1}, \ldots, t_{n_q})\right\}_{n_1=1,\ldots,n_q=1}^{N_1,\ldots,N_q}$, we let tensor $\mathbf{Z} = \left\{z_{n_1,\ldots,n_q}\right\} \in \mathbf{R}^{N_1 \times \cdots \times N_q}$ with $z_{n_1,\ldots,n_q} = z(t_1, \ldots, t_q)$. Without loss of generality, we assume, for $d = 1, \ldots, q$, the N_d grid points in each dimension of t_d are equally spaced. To approximate the functional coefficients $\gamma(\mathbf{t})$ when $q = 2$, Caffo et al. [7] proposed to take an *order-2 singular value decomposition* of tensor $\mathbf{Z}$ to get the singular functions (eigenfunctions) in each dimension then project $\gamma(\mathbf{t})$ into the space spanned by the products of them as shown in (4.5). For more general case of $q \geq 3$, it can be dealt with by using higher order singular value decomposition (HOSVD) [18]. In particular, for a preset $R_1, \ldots, R_q$, namely the ranks of each dimension, the HOSVD is in fact a decomposition of $\mathbf{Z}$ such that $\mathbf{Z} \approx \sum_{n_1=1}^{R_1} \cdots \sum_{n_q=1}^{R_q} z_{n_1,\ldots,n_q} \boldsymbol{\gamma}_{n_1} \circ \cdots \circ \boldsymbol{\gamma}_{n_q} = [\![\mathbf{z}; \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_q]\!]$, where $\mathbf{z} = \{z_{n_1,\ldots,n_q}\} \in \mathbf{R}^{R_1 \times \cdots \times R_q}$, and $\boldsymbol{\Gamma}_d = \{\boldsymbol{\gamma}_1, \ldots \boldsymbol{\gamma}_{R_d}\} \in \mathbf{R}^{N_d \times R_d}$ for $d = 1, \ldots, q$, with $\boldsymbol{\gamma}_{n_d} = (\phi_{n_d}^{(d)}(t_1), \ldots, \phi_{n_d}^{(d)}(t_{N_d}))^T \in \mathbf{R}^{N_d}$ for $n_d = 1, \ldots, R_d$.

However, although the singular value decomposition [7] is intuitive and straightforward, the extracted basis are unsupervised in a sense that such dimension reduction does not require information from the response, hence is potentially problem-

atic. To remedy that, Li et al. [46] proposed and implemented a supervised version of Tucker tensor decomposition which can prevent the potential loss of information.

Similarly, we can reduce the q-dimensional functional linear quantile regression (qD-FLQR) model (4.4) into an order-q tensor linear quantile regression (qD-TLQR) model :

$$Q_{y|x,z}(\tau|x,z) = \alpha_\tau + \langle \mathbf{x}, \boldsymbol{\beta}_\tau \rangle + \langle \zeta_\tau, \boldsymbol{\Lambda}_\tau \rangle, \tag{4.8}$$

where $\zeta_{\tau,n_1,\ldots,n_q} = \int_0^1 \cdots \int_0^1 \phi^{(1)}_{\tau n_1}(t_1) \cdot \cdots \cdot \phi^{(q)}_{\tau n_q}(t_q) z(t_1,\ldots,t_q) dt_q \cdots dt_1$ and $\lambda_{\tau,n_1,\ldots,n_q}$ are the elements of the core tensor $\boldsymbol{\Lambda}_\tau$ as in (4.5). Here we denote $\{\zeta_{\tau,n_1,\ldots,n_q}\}^{R_1,\ldots,R_q}_{n_1=1,\ldots,n_q=1}$ and $\{\lambda_{\tau,n_1,\ldots,n_q}\}^{R_1,\ldots,R_q}_{n_1=1,\ldots,n_q=1}$ by ζ_τ and $\boldsymbol{\Lambda}_\tau$ respectively.

In this chapter, we propose a partial quantile regression framework for qD-FLQR model such that the order-q functional basis can be extracted to maximize the partial quantile covariance [76]. We also implement the proposed method by using finite smoothing techniques [13, 54] and block relaxation algorithm [19]. In addition, we establish and demonstrate the asymptotic properties including consistency and establish the convergence rates by applying some advanced techniques from empirical processes theory (for more background of empirical processes, see van der Vaart [72]).

4.2 q-Dimensional Functional Linear Quantile Regression

For a given quantile level $\tau \in (0,1)$, a q-dimensional functional linear quantile regression (qD-FLQR) model is defined as (4.4). Given $\phi^{(d)}_{\tau,n_d}(t_d)$'s, for $d = 1,\ldots,q$, the *eigenfunction*s in d-th dimension, the model can be reduced to a an order-q tensor linear quantile regression qD-TLQR model (4.8).

In one dimensional functional linear quantile regression model (4.2), given a group of functional basis $\phi_{\tau k}(t)$, $k = 1, \ldots, K$ and let $\gamma_\tau(t) = \sum_{k=1}^{K} \gamma_{\tau k}\phi_k(t)$. Model (4.2) can be rewritten as

$$Q_\tau\left(y|\mathbf{x}, \mathbf{z}(t)\right) = \alpha_\tau + \mathbf{x}^T\boldsymbol{\beta}_\tau + \sum_{k=1}^{K} \mathbf{z}_k\boldsymbol{\gamma}_{\tau k}, \tag{4.9}$$

where $\mathbf{z}_k = \int_0^1 \mathbf{z}(t)\phi_k(t)dt$. This is simply a multiple linear regression, which is essentially a linear programming problem and can be solved by many algorithms - for example, the simplex method [6], the interior point method [38], the MM algorithm [35] among many others, already implemented in various statistical softwares like **quantreg** in **R** [40].

Such group of directions $\phi_{\tau k}(t)$ can be found so that the projections of $\mathbf{z}(t)$s onto them contribute as much as possible to predict the conditional quantile of the response. The concepts of quantile covariance (QC) and partial quantile covariance (PQC) were proposed by Yu et al. [76]. For given $\tau \in (0, 1)$ and a random variable X, the partial quantile covariance (PQC) between two random variables Y and Z is defined as

$$COV_{\mathrm{qr}}(Y, Z|X) = \arg_{\boldsymbol{\gamma}} E \inf_{\alpha,\boldsymbol{\beta},\boldsymbol{\gamma}} \left(\rho_\tau\left(Y - \alpha - \boldsymbol{\beta}X - \boldsymbol{\gamma}Z\right)\right), \tag{4.10}$$

where Z is normalized to have mean zero and variance one. If there is no X, $COV_{\mathrm{qr}}(Y, Z)$ is just quantile covariance (QC) between Y and Z. The quantile covariance measures the contribution of Z to the τ-th quantile of Y. It was first proposed and studied by Dodge and Whittaker [21] in the context of partial quantile regression. Recently, Li et al. [45] proposed a similar concept of quantile correlation and used it to study quantile autoregressive model, while Ma et al. [53] used the partial

quantile correlation to study the ultra-high dimensional variable screening problem.

To find the partial quantile regression basis (PQR), the $\phi_{k\tau}(t)$ can be computed by maximizing

$$COV_{\mathrm{qr}}\left(y - \alpha_\tau - \mathbf{x}\boldsymbol{\beta}_\tau, \sum_{k=1}^{K} \int_0^1 \mathbf{z}(t)\phi_{\tau k}(t)dt\right). \tag{4.11}$$

Similarly for a qD-FLQR model (4.4), we can find $\phi^{(d)}_{\tau,n_d}(t_d)$'s, for $d = 1, \ldots, q$, by maximizing

$$COV_{\mathrm{qr}}\left(y - \alpha_\tau - \mathbf{x}\boldsymbol{\beta}_\tau, \sum_{n_1=1}^{R_1} \cdots \sum_{n_q=1}^{R_q} \int_0^1 \cdots \int_0^1 z(t_1, \ldots, t_q) \prod_{d=1}^{q} \phi^{(d)}_{\tau,n_d}(t_d)dt_d\right). \tag{4.12}$$

The values of R_ds, namely the number of PQR basis in each dimension, can be chosen using BIC or cross validation (CV) as in choosing the number of fPC basis adopted by Kato and other authors [37, 52, 69].

4.3 Implementation

Finite Smoothing of Quantile Loss Functions Similarly to the situation of one dimensional function PQR in Chapter 3, to uncover the asymptotic properties for multi-dimensional functional PQR, one difficulty is due to the non-differentiability of the quantile loss function $\rho_\tau(\cdot)$ at the origin. To address this problem, we propose for $\rho_\tau(u)$ a smoothing approximation $\rho_{\tau\boldsymbol{\nu}}(u)$ such that $\rho_{\tau\boldsymbol{\nu}}(u)$ converges pointwise to $\rho_\tau(u)$, as $\boldsymbol{\nu} \to \boldsymbol{\nu}_0$, and $\boldsymbol{\nu}$ is a vector of smoothing parameters with the typical choice of $\boldsymbol{\nu}_0$ as $\mathbf{0}$ [13, 54]. In addition, if such convergence is uniform, the minimizer of $\rho_{\tau\boldsymbol{\nu}}(u)$ converges to the minimizer of $\rho_\tau(u)$, for a compact set in $(0, 1)$ [33]. Replacing ρ_τ by $\rho_{\tau\boldsymbol{\nu}}$, the quantile objective function then becomes differentiable which

further facilitates defining score, information and identifiability in term of functional linear quantile regression, so that the estimator consistency and convergence rates can be established.

The finite smoothing of quantile loss function is a special case from the general problem of non-smooth convex optimization and is a both statistically and computationally important technique in quantile regressions [13, 56, 75, 81]. For a given precision ϵ, by choosing an appropriate smoothing function, one may obtain an optimal efficiency estimate $O(\frac{1}{\epsilon})$ in term of number of iterations until convergence, which significantly improves that of some popular numerical scheme for non-smooth convex minimization such as $O(\frac{1}{\epsilon^2})$ of the subgradient methods [56]. There are various options of the smoothing functions. For example, one may take the *generalized Huber function* [13] or *iterative least squares smoothing function* [54] both of which converge uniformly towards $\rho_\tau(u)$. Here we take *generalized Huber function* [13] as an example:

$$H_{\nu,\tau}(u) = \begin{cases} u(\tau-1) - \frac{1}{2}(\tau-1)^2\nu, & u \le (\tau-1)\nu; \\ \frac{u^2}{2\nu}, & (\tau-1)\nu < u \le \tau\nu; \\ u\tau - \frac{1}{2}\tau^2\nu, & u > \tau\nu. \end{cases}$$

By choosing a non-negative number ν, and let it go to 0, $H_{\nu,\tau}(u)$ converges uniformly towards $\rho_\tau(u)$. In the meantime, the minimizer for $H_{\nu,\tau}$ converge to the minimizer of ρ_τ too [33].

Algorithm To solve the qD-FLQR model (4.4), For a preset $R_1, \ldots, R_q$, namely the ranks of each dimension of $t_1, \ldots, t_q$, we intend to find a $\left(\prod_{j=1}^q R_j\right)$-dimensional space spanned by the products of $\phi_{\tau,n_d}^{(d)}(t_d)$'s by maximizing partial quantile covariance (4.12). In practice, without loss of generality, we assume that both $z(t_1, \ldots, t_q)$

and $\gamma(t_1, \ldots, t_q)$ are observed at grid points $\left\{(t_{n_1}, \ldots, t_{n_q})\right\}_{n_1=1,\ldots,n_q=1}^{N_1,\ldots,N_q}$, where for $d = 1, \ldots, q$, the N_d grid points in each dimension of t_d are equally spaced. Let tensor $\mathbf{Z} = \left\{z_{n_1,\ldots,n_q}\right\} \in \mathbf{R}^{N_1 \times \cdots \times N_q}$ with $z_{n_1,\ldots,n_q} = z(t_1, \ldots, t_q) \Big/ \left(\prod_d^q N_d\right)$, and tensor $\mathbf{\Gamma}_\tau = \left\{\gamma_{\gamma,n_1,\ldots,n_q}\right\} \in \mathbf{R}^{N_1 \times \cdots \times N_q}$ with $\gamma_{\tau,n_1,\ldots,n_q} = \gamma_\tau(t_1, \ldots, t_q)$. The integral in qD-FLQR model (4.4) can be replaced by $\langle \mathbf{Z}, \mathbf{\Gamma}_\tau \rangle$. Suppose that tensor Γ_τ admits Tucker decomposition $\mathbf{\Gamma}_\tau = \sum_{n_1=1}^{R_1} \cdots \sum_{n_q=1}^{R_q} \lambda_{\tau,n_1,\ldots,n_q} \boldsymbol{\gamma}_{\tau,n_1} \circ \cdots \circ \boldsymbol{\gamma}_{\tau,n_q} = [\![\mathbf{\Lambda}_\tau; \mathbf{\Gamma}_{\tau,1}, \ldots, \mathbf{\Gamma}_{\tau,q}]\!]$, where $\mathbf{\Lambda}_\tau = \{\lambda_{\tau,n_1,\ldots,n_q}\} \in \mathbf{R}^{R_1 \times \cdots \times R_q}$, and $\mathbf{\Gamma}_{\tau,d} = \{\boldsymbol{\gamma}_{\tau,1}, \ldots \boldsymbol{\gamma}_{\tau,R_d}\} \in \mathbf{R}^{N_d \times R_d}$ for $d = 1, \ldots, q$, with $\boldsymbol{\gamma}_{\tau,n_d} = (\phi_{\tau,n_d}^{(d)}(t_1), \ldots, \phi_{\tau,n_d}^{(d)}(t_{N_d}))^T \in \mathbf{R}^{N_d}$ for $n_d = 1, \ldots, R_d$.

Hence, model (4.4) is reduced into an order-q tensor linear quantile regression (qD-TLQR) model:

$$Q_{y|x,z}(\tau|x, z) = \alpha_\tau + \langle \mathbf{x}, \boldsymbol{\beta}_\tau \rangle + \langle \mathbf{Z}, \mathbf{\Gamma}_\tau \rangle. \tag{4.13}$$

Using the *duality lemma* of Li et al. [46], we have $\langle \mathbf{Z}, \mathbf{\Gamma}_\tau \rangle = \langle \tilde{\mathbf{Z}}_\tau, \mathbf{\Lambda}_\tau \rangle$, where $\tilde{\mathbf{Z}}_\tau = [\![\mathbf{Z}; \mathbf{\Gamma}_{\tau,1}^T, \ldots, \mathbf{\Gamma}_{\tau,q}^T]\!] \in \mathbf{R}^{R_1 \times \cdots \times R_q}$, and (4.13) is equivalent to

$$Q_{y|x,z}(\tau|x, z) = \alpha_\tau + \langle \mathbf{x}, \boldsymbol{\beta}_\tau \rangle + \langle \tilde{\mathbf{Z}}_\tau, \mathbf{\Lambda}_\tau \rangle. \tag{4.14}$$

Therefore, given $\tau \in (0, 1)$, we can find the appropriate $\Gamma_{\tau,1}, \ldots, \Gamma_{\tau,q}$ representing the PQR basis by maximizing $COV_{\text{qr}}\left(y - \langle \mathbf{x}, \boldsymbol{\beta}_\tau \rangle, \langle \tilde{\mathbf{Z}}_\tau, \mathbf{\Lambda}_\tau \rangle\right)$. In other words, $\mathbf{\Gamma}_{\tau,1}, \ldots, \mathbf{\Gamma}_{\tau,q}$ should be chosen to maximize

$$l(\alpha_\tau, \boldsymbol{\beta}_\tau, \mathbf{\Gamma}_{\tau 1}, \ldots, \mathbf{\Gamma}_{\tau q}, \mathbf{\Lambda}_\tau) = -E\rho_\tau \left[y - \alpha_\tau - \langle \mathbf{x}, \boldsymbol{\beta}_\tau \rangle - \langle \tilde{\mathbf{Z}}_\tau, \mathbf{\Lambda}_\tau \rangle\right]. \tag{4.15}$$

A crucial observation here is that, although $l(\cdot)$ may not be jointly convex by the parameters of $\alpha, \boldsymbol{\beta}, \mathbf{\Gamma}_{\tau 1}, \ldots, \mathbf{\Gamma}_{\tau q}$ and Λ_τ, it is indeed blockwise convex by each block of parameters. It implies that the block relaxation algorithm [19] would be applicable

in such situation.

As mentioned in previous section, to establish and demonstrate the asymptotic properties, we need to replace ρ_τ by a smooth approximation. In particular, for a nuisance parameter sequence $\{\nu_N\}_{N=1}^{\infty}$ where $\nu_N \to \nu_0$, we can choose a uniformly smooth approximating functions $\rho_{\tau\nu_N}$ such that $\rho_{\tau\nu_N}(u) \rightrightarrows \rho_\tau(u)$ as $N \to \infty$. Replacing ρ_τ by $\rho_{\tau\nu_N}$ in (4.15), we have

$$l_N(\alpha_\tau, \boldsymbol{\beta}_\tau, \boldsymbol{\Gamma}_{\tau 1}, \ldots, \boldsymbol{\Gamma}_{\tau q}, \boldsymbol{\Lambda}_\tau) = -E\rho_{\tau\nu_N}\left[y - \alpha_\tau - \langle \mathbf{x}, \boldsymbol{\beta}_\tau \rangle - \langle \tilde{\mathbf{Z}}_\tau, \boldsymbol{\Lambda}_\tau \rangle\right]. \tag{4.16}$$

where $l_N(\cdot) \rightrightarrows l(\cdot)$, as $N \to \infty$.

For preset $R_1, \ldots, R_q$, we hereby propose the block relaxation algorithm for tensor partial quantile regression:

1. *Repeat* (for N):

   (a) *Initialization*: Let $\mathbf{\Gamma}_{\tau d}^{(0)} \in \mathbf{R}^{p_d \times R_d}$ be a random matrix for $d = 1, \ldots, q$, $\mathbf{\Lambda}_{\tau}^{(0)} \in \mathbf{R}^{R_1 \times \cdots \times R_q}$ be a random tensor, and $(\alpha_{\tau}^{(0)}, \boldsymbol{\beta}_{\tau}^{(0)}) = \arg\max_{\alpha, \boldsymbol{\beta}} l_N(\alpha, \boldsymbol{\beta}, \mathbf{\Gamma}_{\tau 1}^{(0)}, \ldots, \mathbf{\Gamma}_{\tau q}^{(0)}, \mathbf{\Lambda}_{\tau}^{(0)})$, and

   (b) *Repeat* (for p):

      i. for $d = 1, \ldots, q$ do
      $\mathbf{\Gamma}_{\tau d}^{(p+1)} =$
      $\arg\max_{\mathbf{\Gamma}_{\tau d}} l_N(\alpha_{\tau}^{(p)}, \boldsymbol{\beta}_{\tau}^{(p)}, \mathbf{\Gamma}_{\tau 1}^{(p)}, \ldots, \mathbf{\Gamma}_{\tau, d-1}^{(p)}, \mathbf{\Gamma}_{\tau d}, \mathbf{\Gamma}_{\tau, d+1}^{(p)}, \ldots, \mathbf{\Gamma}_{\tau q}^{(p)}, \mathbf{\Lambda}_{\tau}^{(p)})$,

      ii. $\mathbf{\Lambda}_{\tau}^{(p+1)} = \arg\max_{\mathbf{\Lambda}_{\tau}} l_N(\alpha, \boldsymbol{\beta}, \mathbf{\Gamma}_{\tau 1}^{(p+1)}, \ldots, \mathbf{\Gamma}_{\tau q}^{(p+1)}, \mathbf{\Lambda}_{\tau})$.

      iii. $(\alpha_{\tau}^{(p+1)}, \boldsymbol{\beta}_{\tau}^{(p+1)}) = \arg\max_{\alpha, \boldsymbol{\beta}} l_N(\alpha, \boldsymbol{\beta}, \mathbf{\Gamma}_{\tau 1}^{(p+1)}, \ldots, \mathbf{\Gamma}_{\tau q}^{(p+1)}, \mathbf{\Lambda}_{\tau}^{(p+1)})$.

   (c) *Stop* (for p): when
   $l_N(\alpha^{(p+1)}, \boldsymbol{\beta}^{(p+1)}, \mathbf{\Gamma}_{\tau 1}^{(p+1)}, \ldots, \mathbf{\Gamma}_{\tau q}^{(p+1)}, \mathbf{\Lambda}_{\tau}^{(p+1)}) -$
   $l_N(\alpha^{(p)}, \boldsymbol{\beta}^{(p)}, \mathbf{\Gamma}_{\tau 1}^{(p)}, \ldots, \mathbf{\Gamma}_{\tau q}^{(p)}, \mathbf{\Lambda}_{\tau}^{(p)}) < \epsilon$.

   (d) *Save*: $\alpha_{\tau} = \alpha_{\tau}^{(p+1)}$, $\boldsymbol{\beta}_{\tau} = \boldsymbol{\beta}_{\tau}^{(p+1)}$, $\mathbf{\Lambda}_{\tau} = \mathbf{\Lambda}_{\tau}^{(p+1)}$, and $\mathbf{\Gamma}_{\tau d} = \mathbf{\Gamma}_{\tau d}^{(p+1)}$ for $d = 1, \ldots, q$.

2. *Stop* (for N): When
   $l_{N+1}(\alpha_{\tau}, \boldsymbol{\beta}_{\tau}, \mathbf{\Gamma}_{\tau 1}, \ldots, \mathbf{\Gamma}_{\tau q}, \mathbf{\Lambda}_{\tau}) - l_N(\alpha_{\tau}, \boldsymbol{\beta}_{\tau}, \mathbf{\Gamma}_{\tau 1}, \ldots, \mathbf{\Gamma}_{\tau q}, \mathbf{\Lambda}_{\tau}) < \epsilon$.

3. *Output*: $\alpha_{\tau}, \boldsymbol{\beta}_{\tau}, \mathbf{\Gamma}_{\tau 1}, \ldots, \mathbf{\Gamma}_{\tau q}$, and $\mathbf{\Lambda}_{\tau}$.

Algorithm 3: Block relaxation algorithm for Tucker PQR (for preset $R_1, \ldots, R_q$)

The advantages of this procedure is obvious. Firstly, $\mathbf{\Gamma}_{\tau d}$s, for $d = 1, \ldots, q$, retain the basis vectors by column which are updated altogether. Secondly, the block relaxation procedure can guarantee the solutions' stability and convergence whenever the objective function, $l_N(\cdot)$, is convex and bounded from above which is easy to verify. Last but not least, due to the convexity and uniform convergence of the objective functions, the maximizer of $l_N(\cdot)$ converges to the maximizer of $l(\cdot)$. Convergence properties of ASIMPQR are summarized in the following Proposition:

Proposition 4.3.1. *If we have* $l_N(\boldsymbol{\upsilon}) \rightrightarrows l(\boldsymbol{\upsilon})$ *with* $\boldsymbol{\upsilon} = (\alpha_\tau, \boldsymbol{\beta}_\tau, \mathbf{\Gamma}_{\tau 1}, \ldots, \mathbf{\Gamma}_{\tau q}, \mathbf{\Lambda}_\tau)$ *as* $N \to \infty$*, where* $l(\boldsymbol{\upsilon})$ *and* $l_N(\boldsymbol{\upsilon})$ *are defined as in (4.15) and (4.16). For a given fixed N, if (i)* $l_N(\boldsymbol{\upsilon})$ *is continuous, and coercive, that is, the set* $\{\boldsymbol{\upsilon} : l_N(\boldsymbol{\upsilon}) \geq l_N(\boldsymbol{\upsilon}^{(0)})\}$ *is compact and bounded above, (ii) the objective function in each block update of algorithm is strictly concave, and (iii) the set of stationary points of* $l_N(\boldsymbol{\upsilon})$ *are isolated, we have the following results:*

1. *(Global convergence) For a give m, the sequence* $\boldsymbol{\upsilon}^{(p)} = (\alpha^{(p)}, \boldsymbol{\beta}^{(p)}, \mathbf{\Gamma}_{\tau 1}^{(p)}, \ldots, \mathbf{\Gamma}_{\tau q}^{(p)}, \mathbf{\Lambda}_\tau^{(p)})$ *generated by the algorithm above converges to a stationary point of* $l_N(\boldsymbol{\upsilon})$*.*

2. *(Local convergence) For a give m, let* $\boldsymbol{\upsilon}^{(\infty)} = (\alpha^{(\infty)}, \boldsymbol{\beta}^{(\infty)}, \mathbf{\Gamma}_{\tau 1}^{(\infty)}, \ldots, \mathbf{\Gamma}_{\tau q}^{(\infty)}, \mathbf{\Lambda}_\tau^{(\infty)})$ *be a strict local maximum of* $l_N(\boldsymbol{\upsilon})$*. The iterates generated by the algorithm above are locally attracted to* $\boldsymbol{\upsilon}^{(\infty)}$ *for* $\boldsymbol{\upsilon}^{(0)}$ *sufficiently close to* $\boldsymbol{\upsilon}^{(\infty)}$*.*

3. *(Approximation convergence) For a given fixed m, the convergence points obtained from* $l_N(\boldsymbol{\upsilon})$ *will converge in probability to the convergence point of* $l(\boldsymbol{\upsilon})$ *as* $N \to \infty$*.*

The assumptions above are not hard to verify if it is allowed to impose some regular conditions on the distribution functions [38]. If assumptions (i) - (iii) stand,

the *local convergence* and *global convergence* can be obtained by following the similar discussions in [46] and [83]. In order to obtain the *approximation convergence*, we can use *Lemma 2* of Hjort and Pollard [33], provided that $l_N(\boldsymbol{\upsilon})$ has a unique $\arg\max_{\boldsymbol{\upsilon}}$. In fact, since the verification is the same as shown in Chapter 3, we skip the details here.

4.4 Asymptotic Properties

We study the statistical properties of the estimator $\hat{\boldsymbol{\upsilon}}$. In this chapter, we adopt the asymptotic setup with fixed ranks, $R_1, \ldots, R_q$, and a diverging sample size n, because this is an important first step toward a comprehensive understanding of the theoretical properties of the proposed model.

For given fixed $R_1, \ldots, R_q$, we first establish score, information and identifiability in term of the proposed q-dimensional PQR (qD-PQR) for q-dimensional functional linear quantile regression. Then the consistency and asymptotic normality can be derived by applying the advanced theory of empirical processes [72] and following the similar discussions as in Chapter 3 [46, 83].

Score and Information We first derive the score and information for partial quantile regression model. As discussed in Yu and Moyeed [77] and Sánchez et al. [64], the minimization of the quantile loss are equivalent to the maximization of a likelihood function formed by independent and identically distributed asymmetric Laplace densities. In fact, $l(\cdot)$ defined as (4.15) is proportional to the log-likelihood. Therefore, we want to derive the score and information for $l(\cdot)$. Since $l(\cdot)$ is not differentiable, we are giving the score and information for its approximation $l_N(\cdot)$ defined as (4.16). As $N \to \infty$, the differences of the scores and information matrices

between l_N and l are almost negligible.

The following standard calculus notations are used. For a scalar function f, ∇f is the (column) gradient vector, $df = [\nabla f]^T$ is the differential. For a multivariate function $g : \mathbf{R}^p \to \mathbf{R}^q$, $Dg \in \mathbf{R}^{p\times q}$ denotes the Jacobian matrix holding partial derivatives $\frac{\partial g_j}{\partial x_i}$. As a direct result from *Lemma 2* of Zhou et al. [83], one can easily check the followings.

Lemma 4.4.1. *Let* $\eta = \alpha_\tau + \langle \mathbf{x}, \boldsymbol{\beta}_\tau \rangle + \langle \mathbf{Z}, \boldsymbol{\Gamma}_\tau \rangle = \alpha_\tau + \langle \mathbf{x}, \boldsymbol{\beta}_\tau \rangle + \langle \tilde{\mathbf{Z}}_\tau, \boldsymbol{\Lambda}_\tau \rangle$*, then the gradient* $\nabla\eta(\boldsymbol{\Gamma}_{\tau 1}, \ldots, \boldsymbol{\Gamma}_{\tau q}) \in \mathbf{R}^{\prod_d Rd + \sum_d N_d R_d}$ *is as*

$$\nabla\eta_i(\boldsymbol{\Lambda}_\tau, \boldsymbol{\Gamma}_{\tau 1}, \ldots, \boldsymbol{\Gamma}_{\tau q}) = -\left[\boldsymbol{\Gamma}_{\tau q} \otimes \cdots \otimes \boldsymbol{\Gamma}_{\tau 1}\ \mathbf{J}_1\ \cdots\ \mathbf{J}_q\right] \mathrm{vec}\,(\mathbf{Z}),$$

where $\mathbf{J}_d \in \mathbf{R}^{\prod_d N_d \times N_d R_d}$ *is the Jacobian*

$$\mathbf{J}_d = D\boldsymbol{\Gamma}(\boldsymbol{\Gamma}_{\tau d}) = \Pi_d \left\{\left[\left(\boldsymbol{\Gamma}_{\tau,q} \otimes \cdots \otimes \boldsymbol{\Gamma}_{\tau,d+1} \otimes \boldsymbol{\Gamma}_{\tau,d-1} \otimes \cdots \otimes \boldsymbol{\Gamma}_{\tau,1}\right) \boldsymbol{\Lambda}_{(d)}^T\right] \otimes I_{N_d}\right\},$$

and Π_d *is the* $\left(\prod_{d=1}^D N_d\right)$ *by* $\left(\prod_{d=1}^D N_d\right)$ *permutation matrix that reorders* $\mathrm{vec}\,\boldsymbol{\Gamma}_{\tau(\mathrm{d})}$ *to obtain* $\mathrm{vec}\,\Gamma_\tau$*, i.e.,* $\mathrm{vec}\ \Gamma_\tau = \Pi_\mathrm{d}\mathrm{vec}\boldsymbol{\Gamma}_{\tau(\mathrm{d})}$.

Here the *mode-d matricization*, $\boldsymbol{\Gamma}_{(d)}$, maps a tensor $\boldsymbol{\Gamma}$ into a $N_d \times \prod_{d'\neq d} N_{d'}$ matrix such that the $(j_1, \ldots, j_q)$ element of the array $\boldsymbol{\Gamma}$ maps to the (j_d, k) element of the matrix $\boldsymbol{\Gamma}_{(d)}$, where $j = 1 + \sum_{d'\neq d}(j_{d'} - 1) \prod_{d''<d', d''\neq d} N_{d''}$.

Proposition 4.4.2. *Consider the partial quantile regression model (4.15):*

1. *The score function (or score vector) of* l_N *is*

$$\nabla l_N(\boldsymbol{\Lambda}_\tau, \boldsymbol{\Gamma}_{\tau 1}, \ldots, \boldsymbol{\Gamma}_{\tau q}) = -E\rho'_{\tau v_N}\left(\eta_i(\boldsymbol{\Lambda}_\tau, \boldsymbol{\Gamma}_{\tau 1}, \ldots, \boldsymbol{\Gamma}_{\tau q})\right) \cdot \nabla\eta(\boldsymbol{\Lambda}_\tau, \boldsymbol{\Gamma}_{\tau 1}, \ldots, \boldsymbol{\Gamma}_{\tau q}).$$

with $\nabla\eta(\boldsymbol{\Lambda}_\tau, \boldsymbol{\Gamma}_{\tau 1}, \ldots, \boldsymbol{\Gamma}_{\tau q})$ *given in Lemma 4.4.1.*

2. *The Fisher information matrix of l_N is*

$$\mathbf{I}_N(\mathbf{\Lambda}_\tau, \mathbf{\Gamma}_{\tau 1}, \ldots, \mathbf{\Gamma}_{\tau q}) \;=\; E\left[\nabla l_N(\mathbf{\Lambda}_\tau, \mathbf{\Gamma}_{\tau 1}, \ldots, \mathbf{\Gamma}_{\tau q}) dl_N(\mathbf{\Lambda}_\tau, \mathbf{\Gamma}_{\tau 1}, \ldots, \mathbf{\Gamma}_{\tau q})\right].$$

Identifiability Before studying asymptotic properties, we need to deal with the identifiability issue. The parametrization in the partial linear quantile regression model is non-identifiable mainly due to the complication of the nonsingular tranformation indeterminacy. That is the tensor decomposition of $\mathbf{\Gamma}_\tau$ of $[\![\mathbf{\Lambda}_\tau; \mathbf{\Gamma}_{\tau,1}, \ldots, \mathbf{\Gamma}_{\tau,q}]\!] \in \mathbf{R}^{R_1 \times \cdots \times R_q}$ is not unique. It is similar to the situation when $q = 1$ such that the vector of $\mathbf{\Gamma}_\tau$ can be decomposed into infinity number of possible ways. To fix the such indeterminacy, we simply assume that first R_d rows of $\mathbf{\Gamma}_{\tau d}$ to be ones

$$\mathit{\Gamma} = \left\{[\![\mathbf{\Lambda}_\tau; \mathbf{\Gamma}_{\tau,1}, \ldots, \mathbf{\Gamma}_{\tau,q}]\!] : \boldsymbol{\gamma}_{j_d}^{(r)} = 1, j_d = 1, \ldots, R_d, d = 1, \ldots, q\right\}.$$

Next we give a sufficient and necessary condition for local identifiability.

Proposition 4.4.3 (Identifiability)**.** *Given a sequence of iid data points $\{(y_i, \mathbf{z}_i)\}_{i=1}^n$ from the quantile regression model. Let $\mathbf{\Gamma}_{\tau 0} \in \mathit{\Gamma}_\tau$ be a parameter point and assume there exists an open neighbourhood of $\mathbf{\Gamma}_{\tau 0}$ in which the information matrix has a constant rank. Then $\mathbf{\Gamma}_{\tau 0}$ is locally identifiable if and only if*

$$I_N(\mathbf{\Gamma}_{\tau 0}) = \mathbb{E}\left[\nabla l_N(\mathbf{\Gamma}_{\tau 0}) dl_N(\mathbf{\Gamma}_{\tau 0})\right]$$

is nonsingular.

This result can be obtained by using *Theorem 1* of Rothenberg [63].

Asymptotics The asymptotics follow from those discussions for MLE or M-estimation. A key observation is that the nonlinear part of tensor model (4.13) is indeed a poly-

nomial of degree q, where the collections of polynomials $\{\langle \mathbf{\Gamma}_\tau, \mathbf{Z}\rangle, \mathbf{\Gamma}_\tau \in \mathit{\Gamma}\}$ forms a Vapnik-Cervonenkis (VC) class. Then standard uniform convergence theory for M-estimation [72] applies.

Theorem 4.4.4 (Consistency)**.** *Assume $\mathbf{\Gamma}_{\tau 0} \in \mathit{\Gamma}$ is (globally) identifiable up to permutation and the array covariates $\mathbf{Z}_i$ are iid from a bounded distribution. The M-estimator is consistent, that is, $\hat{\mathbf{\Gamma}}_n$ converges to $\mathbf{\Gamma}_{\tau 0}$ (modulo permutation) in probability for quantile regression model (4.13) with a compact parameter space $\mathit{\Gamma}_0 \subset \mathit{\Gamma}$.*

The consistency can be checked using the theory of empirical processes. By showing that $\{l_N(\mathbf{\Gamma}_\tau), \mathbf{\Gamma}_\tau \in \mathit{\Gamma}\}$ is a Donsker class when the parameter is restricted to a compact set, the Glivenko-Cantelli theorem establish the uniform convergence. Uniqueness is guaranteed by the information equality whenever $\mathbf{\Gamma}_{\tau 0}$ is identifiable.

Theorem 4.4.5 (Asymptotic normality)**.** *For an interior point $\mathbf{\Gamma}_{\tau 0} \in \mathit{\Gamma}$ with nonsingular information matrix $I_N(\mathbf{\Gamma}_{\tau n})$ and $\hat{\mathbf{\Gamma}}_{\tau n}$ is consistent,*

$$\sqrt{n}\left[\mathrm{vec}\left(\hat{\mathbf{\Gamma}}_{\tau \mathrm{n}}\right) - \mathrm{vec}\left(\mathbf{\Gamma}_{\tau 0}\right)\right]$$

converges in distribution to a normal with mean zero and covariance $I_N^{-1}(\mathbf{\Gamma}_{\tau 0})$.

4.5 Simulation Studies

In this section, we demonstrate that the proposed q-dimensional partial quantile regression (qD-PQR) model, although with substantial reduction in dimension, can manage to identify a range of two dimensional signal shapes with different ranks. As a supervised learning procedure, qD-PQR is more efficient in prediction compared with the unsupervised MPC decomposition [7]. We use a similar setting as Li

et al. [46]. As shown in Figure 4.2 and 4.4, the true 2D signals $\mathbf{\Gamma} \in \mathbf{R}^{64\times64}$ are listed in the first column, along with the estimates by the proposed method in the second to fourth columns with ranks $(1, 1)$, $(2, 2)$ and $(3, 3)$, respectively. As a comparison, Figure 4.3 and 4.5 are listed by using the method proposed by Li et al. [46]. The regular covariate $\mathbf{x}$ and image covariates $\mathbf{Z} \in \mathbf{R}^{64\times64}$ are randomly generated with all elements being independent standard normal. Let $\mu = \alpha + \langle \mathbf{x}, \boldsymbol{\beta} \rangle + \langle \mathbf{Z}, \mathbf{\Gamma} \rangle$, then the response Y is generated from

1) Normal distribution with mean μ and standard deviation $sd(\mu)/\sqrt{10}$;

2) Cauchy distribution with location μ and scale $sd(\mu)/\sqrt{10}$.

The vector of coefficient $\boldsymbol{\beta} = \mathbf{1}_4$ and the coefficient tensor $\mathbf{\Gamma}$ is binary with the signal region equal to one and the rest zero. As shown in Figure 4.2 and 4.3, when the response is generated from normal distribution, our proposed qD-PQR method performs comparable with the Tucker tensor regression model [46]. When the response is generated from Cauchy distribution, as shown in Figure 4.4 and 4.5, the robustness of qD-PQR can result a reasonable recovery of the true signal while the Tucker tensor regression [46] is not designed to handle such situation. As shown in Figure 4.6, even though the true signal is blurred by the heavily skewed Cauchy distribution, by increasing the ranks from $(3, 3)$ to $(4, 4)$, the result can be further improved. The proper ranks can be chosen using either BIC or CV with satisfactory performances as implied by Li et al. [46], whose details are skipped in this chapter.

As mentioned at the beginning, one of the advantages our proposed method is that the basis extracted are supervised by the response, while the unsupervised basis extraction method [7] mainly focus on the decomposition of image covariates $\mathbf{Z}$ which can be totally random. Table lists the mean absolute errors (MAEs) between μ abd $\hat{\mu}$ for different signals where the 2000 observations are randomly divided into training and testing data set with proportions of 80% and 20%. Our

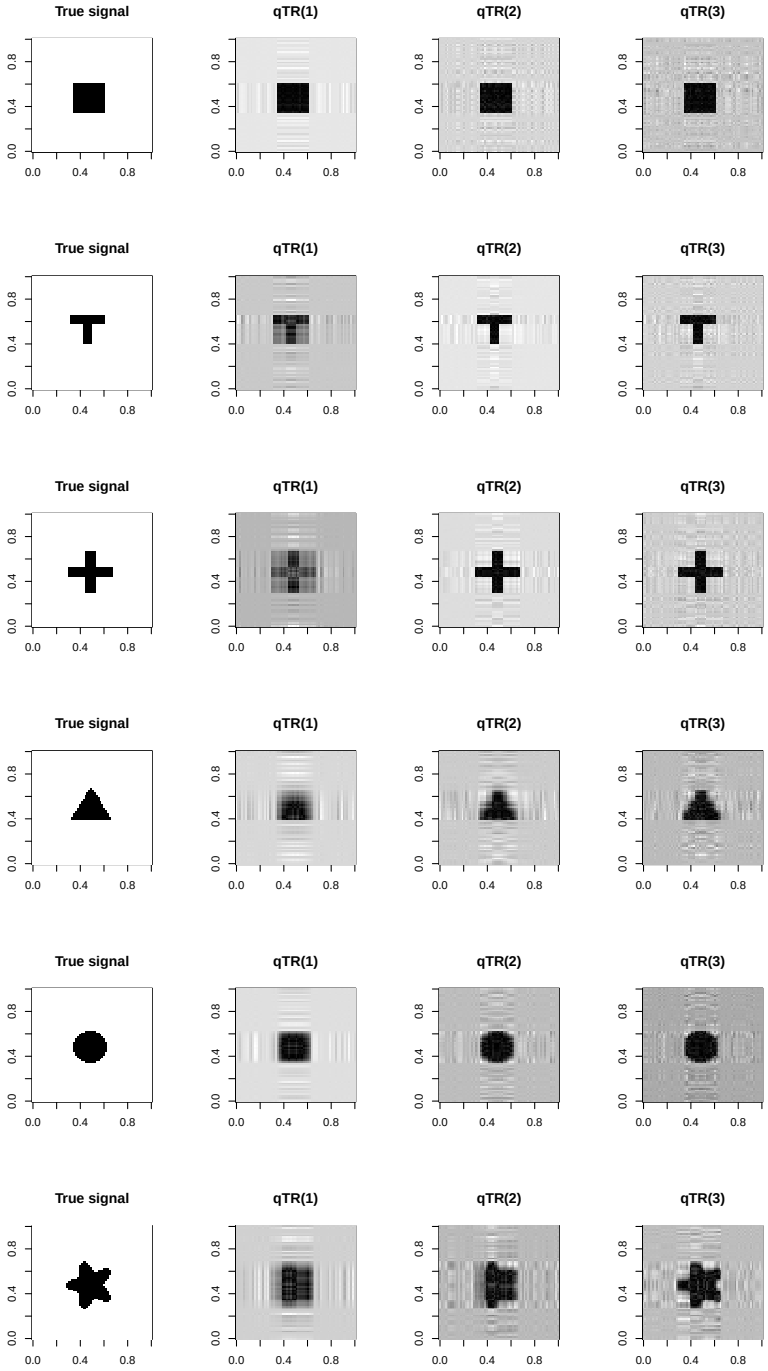


Figure 4.2: Simulation: True and recovered image signals by 2-dimensional partial quantile regression (2D-PQR). The matrix variate has size 64 by 64 with entries generated as independent standard normals. The errors follow normal distributions with mean zero and standard deviation $\mathrm{sd}(\mu) / \sqrt{10}$. The regression coefficient for each entry is either 0 (white) or 1 (black). The sample size is 2000. qTR(r) means the estimate is from 2D-PQR with an r-by-r core tensor.

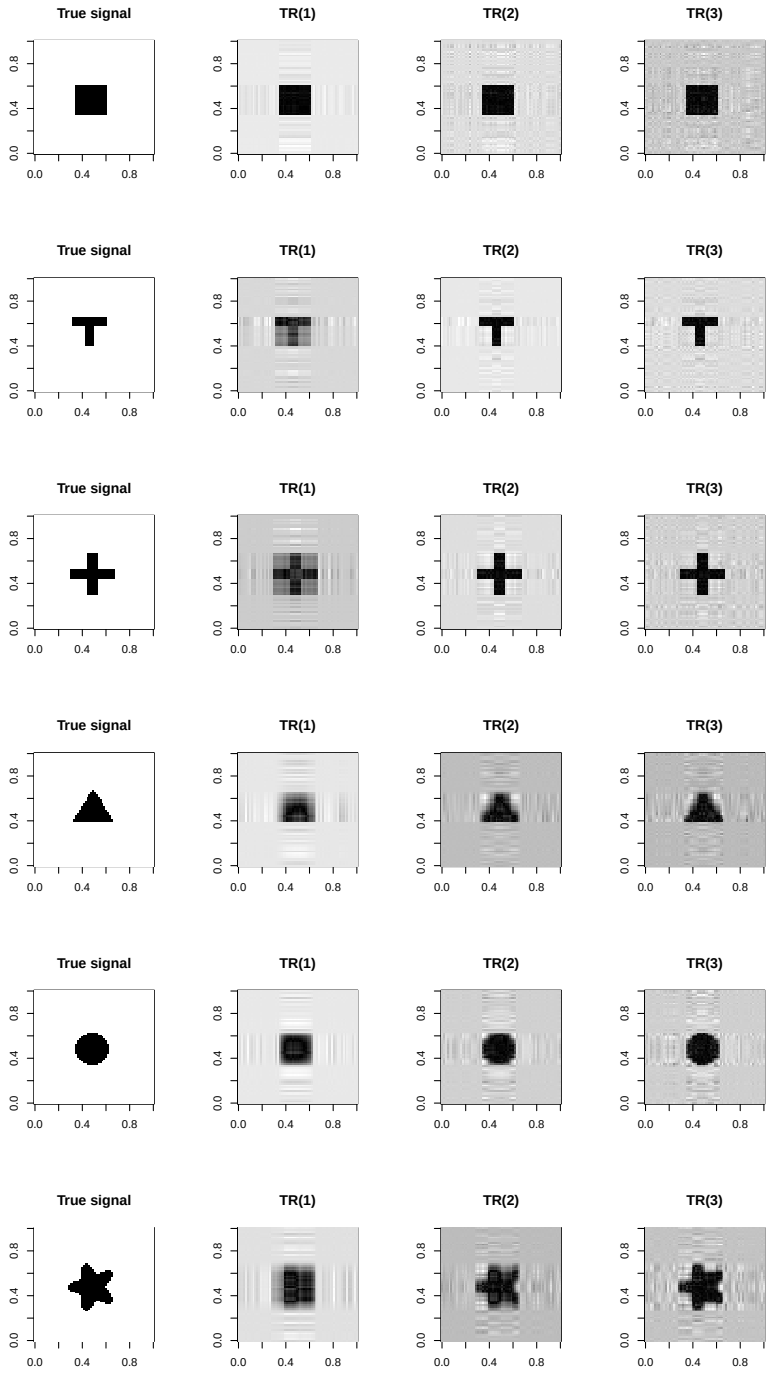


Figure 4.3: Simulation: True and recovered image signals by Tucker tensor regression. The matrix variate has size 64 by 64 with entries generated as independent standard normals. The errors follow normal distributions with mean zero and standard deviation $\mathrm{sd}(\mu) / \sqrt{10}$. The regression coefficient for each entry is either 0 (white) or 1 (black). The sample size is 2000. TR(r) means the estimate is from Tucker regression with an r-by-r core tensor.

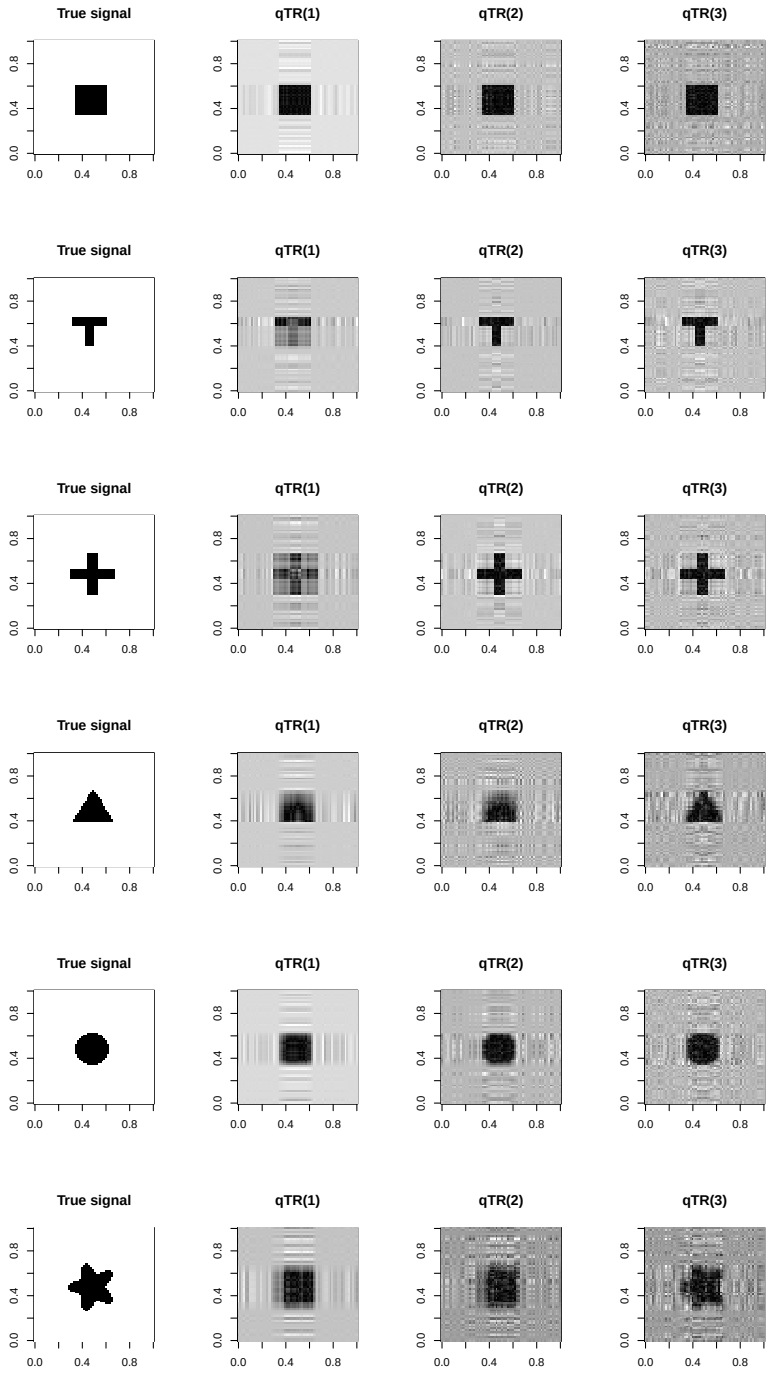


Figure 4.4: Simulation: True and recovered image signals by 2-dimensional partial quantile regression (2D-PQR). The matrix variate has size 64 by 64 with entries generated as independent standard normals. The errors follow Cauchy distributions with location zero and scale $\mathrm{sd}(\mu) / \sqrt{10}$. The regression coefficient for each entry is either 0 (white) or 1 (black). The sample size is 2000. qTR(r) means the estimate is from 2D-PQR with an r-by-r core tensor.

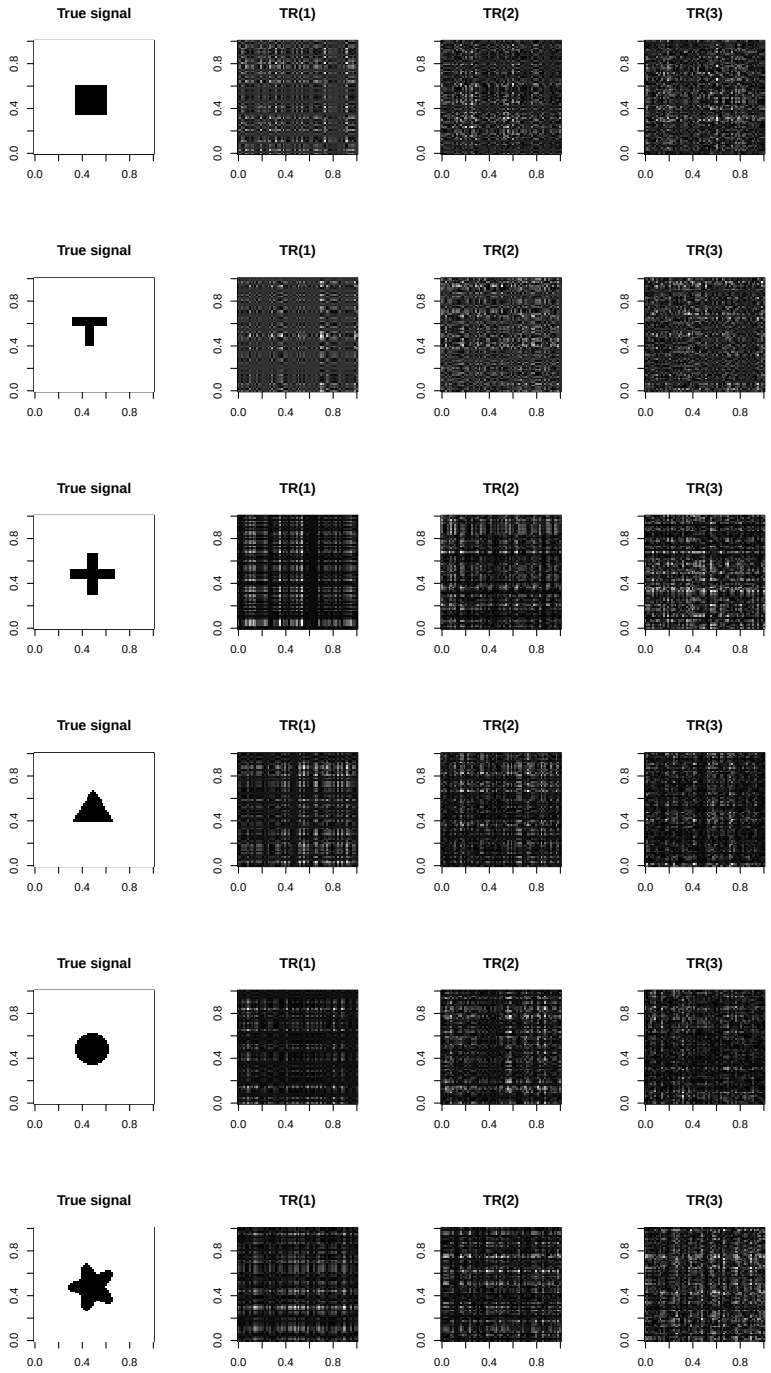


Figure 4.5: Simulation: True and recovered image signals by Tucker tensor regression. The matrix variate has size 64 by 64 with entries generated as independent standard normals. The errors follow Cauchy distributions with location zero and scale $\mathrm{sd}(\mu) / \sqrt{10}$. The regression coefficient for each entry is either 0 (white) or 1 (black). The sample size is 2000. TR(r) means the estimate is from Tucker regression with an r-by-r core tensor.

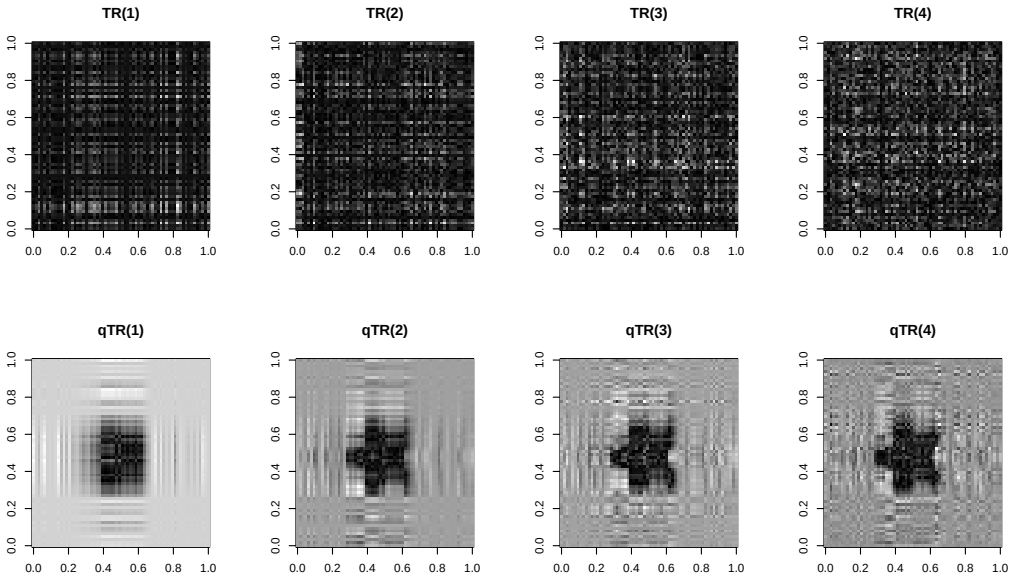


Figure 4.6: Simulation: Comparison of recovered image signals of Tucker tensor regressions with 2-dimensional partial quantile regression (2D-PQR) where the true signal is displayed as the last row of Figure 4.3. The matrix variate has size 64 by 64 with entries generated as independent standard normals. The errors follow Cauchy distributions with location zero and scale $\mathrm{sd}(\mu) / \sqrt{10}$. The regression coefficient for each entry is either 0 (white) or 1 (black). The sample size is 2000.

proposed method outperforms the *order-2 singular value decomposition* (O2SVD) [7] for various signals and errors. One explanation is that unlike the true signal $\boldsymbol{\gamma}$, the image (functional) covariates $\mathbf{Z}$ are randomly generated in this setting, and hence the unsupervised decomposition of the functional covariates brings no useful information for the prediction of μ. In fact, the low rank representation of covariates can even lose some information. On the contrary, our proposed learning method is supervised by y and decompose the signals, instead of covariates, which indeed can be approximated by a low rank representation.

Gaussian	PQR					O2SVD				
	TR(1)	TR(2)	TR(3)	TR(4)	TR(5)	TR(1)	TR(2)	TR(3)	TR(4)	TR(5)
Square	1.59	2.49	2.97	3.53	3.93	12.97	12.93	12.85	13.11	13.18
T	4.77	1.65	2.25	2.69	2.99	9.72	9.69	9.68	9.79	9.88
Cross	6.61	1.99	2.95	3.42	4.06	12.52	12.58	12.76	12.76	12.68
Triangle	5.50	5.20	5.12	5.66	5.68	11.37	11.29	11.42	11.24	11.42
Circle	4.87	4.29	4.19	5.11	5.68	13.62	13.66	13.74	13.79	13.71
Star	7.75	6.66	5.84	6.85	7.14	15.68	15.76	15.78	16.21	16.15
Cauchy	PQR					O2SVD				
	TR(1)	TR(2)	TR(3)	TR(4)	TR(5)	TR(1)	TR(2)	TR(3)	TR(4)	TR(5)
Square	2.12	4.00	4.98	7.26	9.79	13.57	13.67	13.50	13.33	13.89
T	6.82	2.99	4.45	5.24	7.03	9.88	9.90	9.94	9.92	10.17
Cross	7.60	3.18	5.88	6.78	10.65	12.01	12.03	12.07	12.30	12.28
Triangle	6.05	6.89	7.92	9.21	10.54	11.41	11.33	11.32	11.38	11.41
Circle	5.30	6.32	7.08	9.26	10.60	13.10	13.15	13.30	13.25	13.40
Star	7.47	8.17	10.33	11.89	14.46	15.53	15.59	15.52	15.43	15.32

Table 4.1: Simulation: The testing mean absolute errors of 2-dimensional partial quantile regression (2D-PQR) and order-2 singular value decomposition (O2SVD) for different signals and different errors where the 2000 observations are randomly divided into training and testing data set with proportions of 80% and 20%.

4.6 Discussion

In this chapter, we propose and implement the q-dimensional partial quantile regression (qD-PQR) procedure for multidimensional functional linear model using tensor decomposition techniques and block relaxation ideas. We also establish and demonstrate the corresponding asymptotic properties which can be verified by applying advanced techniques from empirical process theory. The proofs of them can be followed in the similar ways as shown in Chapter 3 and Li et al. [46]. As shown in the first part of the numerical studies, when the errors are normally distributed, our proposed method can recover the true signal equally well as the method proposed by Li et al. [46], while the latter can not handle the non-normal errors.

As illustrated in the second part the numerical studies, as a supervised basis extraction method, the proposed qD-PQR procedure is more efficient in prediction than and hence can be considered as an alternative to the unsupervised learning methods such as q-dimensional MPC [7].

Chapter 5

Conclusion and Discussion

In this thesis, we concentrate on the partial quantile regression methods for various functional linear models.

In chapter 2, we propose a new prediction procedure for the functional linear quantile regression model by using partial quantile covariance techniques and develop the simple partial quantile regression (SIMPQR) algorithm inspired by simple partial least regression, SIMPLS [17]. The proposed method is motivated by the success of the partial least square (PLS) method in mean regression problem of functional linear model. We also extend our method to functional composite quantile regression (CQR) [87] which works well under the homoscedasticity assumption.

Although the SIMPQR proposed in chapter 2 works well in practice, there is no theoretical guarantees such as the convergence and asymptotics mainly due to the complications caused by the iterative nature of basis extraction and non-smoothness of quantile loss function. To remedy that, in chapter 3, we propose a smoothing approximation for the quantile loss function by applying the finite smoothing techniques [13, 54]. Since the approximation function is convex and uniformly con-

verges towards the quantile loss function, the minimizer of the former converges to the minimizer of the latter in a compact set. Replacing the quantile loss function by its smoothing approximation, for a given fixed number of PQR basis, the original PQR formulation can be modified to have the algorithm of alternative SIMPQR (ASIMPQR) according to the block relaxation ideas [19] which updates and obtains the basis as a "block" instead of one by one sequentially. Such modification provides an alternative to the original formulation for PQR (APQR) basis which leads to insightful results as what has been shown during the rest of chapter 3, demonstrating consistency and establishing convergence rates by applying advanced techniques from empirical processes theory [72].

Inspired by the successes of the ASIMPQR technique proposed in in chapter 4 in both practice and theory, we further propose and implement the generalization of PQR procedure (qD-PQR) to multidimensional functional linear model in chapter 4, where the functional covariate can be taken as a multivariable function. Analogous to the ASIMPQR method, we also establish and demonstrate the corresponding asymptotic properties [72].

There are several topics that merit further research. In our thesis, we mainly focus on the methodology and theoretical properties of partial quantile regression without penalty. By imposing appropriate penalty terms, it is likely that the corresponding prediction and coefficient estimation could be further improved; for instance, we may use smooth penalty to induce better recovery of true functional coefficients or sparse penalty to overcome the potential issue of overfitting. The theoretical properties of composite partial quantile regression (CPQR) for multidimensional functional linear model is also worth further studying. Moreover, it is also of interest to investigate the performance of our proposed methods for multidimensional functional linear model with multiple functional covariates.

Bibliography

[1] H. Abdi, "Partial least squares regression and projection on latent structure regression (PLS regression)," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 1, pp. 97–106, 2010.

[2] G. Aguirre, E Zarahn, and M D'esposito, "The variability of human, bold hemodynamic responses," *NeuroImage*, vol. 8, no. 4, pp. 360–369, 1998.

[3] G. Aneiros-Pérez and P. Vieu, "Semi-functional partial linear regression," *Statistics & Probability Letters*, vol. 76, no. 11, pp. 1102–1110, 2006.

[4] P. J. Basser, J. Mattiello, and D. LeBihan, "Estimation of the effective self-diffusion tensor from the NMR spin echo," *Journal of Magnetic Resonance, Series B*, vol. 103, no. 3, pp. 247–254, 1994.

[5] —, "MR diffusion tensor spectroscopy and imaging.," *Biophysical journal*, vol. 66, no. 1, pp. 259–267, 1994.

[6] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[7] B. S. Caffo, C. M. Crainiceanu, G. Verduzco, S. Joel, S. H. Mostofsky, S. S. Bassett, and J. J. Pekar, "Two-stage decompositions for the analysis of functional connectivity for fMRI with application to alzheimer's disease risk," *NeuroImage*, vol. 51, no. 3, pp. 1140–1149, 2010.

[8] T. T. Cai and P. Hall, "Prediction in functional linear regression," *The Annals of Statistics*, vol. 34, no. 5, pp. 2159–2179, 2006.

[9] T. T. Cai and M. Yuan, "Minimax and adaptive prediction for functional linear regression," *Journal of the American Statistical Association*, vol. 107, no. 499, pp. 1201–1216, 2012.

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

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

[12] H. Cardot, F. Ferraty, and P. Sarda, "Spline estimators for the functional linear model," *Statistica Sinica*, vol. 13, no. 3, pp. 571–592, 2003.

[13] C. Chen, "A finite smoothing algorithm for quantile regression," *Journal of Computational and Graphical Statistics*, vol. 16, no. 1, pp. 136–164, 2007.

[14] C. Chen and Y. Wei, "Computational issues for quantile regression," *Sankhyā: The Indian Journal of Statistics*, pp. 399–417, 2005.

[15] A. Cichocki, D Mandic, A. Phan, C Caiafa, G Zhou, Q Zhao, and L. De Lathauwer, "Tensor decompositions for signal processing applications from two-way to multiway component analysis," *arXiv preprint arXiv:1403.4462*, 2014.

[16] C. Crambes, A. Kneip, and P. Sarda, "Smoothing splines estimators for functional linear regression," *Annals of Statistics*, pp. 35–72, 2009.

[17] S. De Jong, "Simpls: an alternative approach to partial least squares regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 18, no. 3, pp. 251–263, 1993.

[18] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-(r1, r2,..., rn) approximation of higher-order tensors," *SIAM journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.

[19] J. De Leeuw, "Block-relaxation algorithms in statistics. studies in classification, data analysis, and knowledge organization.," in *Information Systems and Data Analysis*, H. Bock, W. Lenski, and M. Richter, Eds., Springer, Berlin, Heidelberg, 1994, pp. 308–324.

[20] A. Delaigle and P. Hall, "Methodology and theory for partial least squares applied to functional data," *The Annals of Statistics*, vol. 40, no. 1, pp. 322–352, 2012.

[21] Y. Dodge and J. Whittaker, "Partial quantile regression," *Metrika*, vol. 70, no. 1, pp. 35–57, 2009.

[22] K. Domschke and U. Dannlowski, "Imaging genetics of anxiety disorders," *NeuroImage*, vol. 53, no. 3, pp. 822–831, 2010.

[23] L. Fass, "Imaging and cancer: a review," *Molecular Oncology*, vol. 2, no. 2, pp. 115–152, 2008.

[24] L. E. Frank and J. H. Friedman, "A statistical view of some chemometrics regression tools," *Technometrics*, vol. 35, no. 2, pp. 109–135, 1993.

[25] K. J. Friston, "Modalities, modes, and models in functional neuroimaging," *Science*, vol. 326, no. 5951, pp. 399–403, 2009.

[26] J. Goldsmith, J. Bobb, C. M. Crainiceanu, B. Caffo, and D. Reich, "Penalized functional regression," *Journal of Computational and Graphical Statistics*, vol. 20, no. 4, pp. 830–851, 2011.

[27] J. Goldsmith, C. M. Crainiceanu, B. S. Caffo, and D. S. Reich, "Penalized functional regression analysis of white-matter tract profiles in multiple sclerosis," *NeuroImage*, vol. 57, no. 2, pp. 431–439, 2011.

[28] M. Golubitsky and V. Guillemin, *Stable mappings and their singularities*. Springer Science & Business Media, 2012, vol. 14.

[29] P. Hall and J. L. Horowitz, "Methodology and convergence rates for functional linear regression," *Annals of Statistics*, vol. 35, no. 1, pp. 70–91, 2007.

[30] K. Han and H. Shin, "Functional linear regression for functional response via sparse basis selection," *Journal of the Korean Statistical Society*, no. in-press, pp. 00–00, 2014.

[31] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction 2 edition Springer*. Springer New York, 2009.

[32] I. S. Helland, "Partial least squares regression and statistical models," *Scandinavian Journal of Statistics*, pp. 97–114, 1990.

[33] N. L. Hjort and D. Pollard, "Asymptotics for minimisers of convex processes," *arXiv preprint arXiv:1107.3806*, 2011.

[34] S. A. Huettel, A. W. Song, and G. McCarthy, *Functional magnetic resonance imaging*. Sinauer Associates Sunderland, 2004, vol. 1.

[35] D. R. Hunter and K. Lange, "Quantile regression via an mm algorithm," *Journal of Computational and Graphical Statistics*, vol. 9, no. 1, pp. 60–77, 2000.

[36] G. M. James, J. Wang, and J. Zhu, "Functional linear regression that's interpretable," *The Annals of Statistics*, pp. 2083–2108, 2009.

[37] K. Kato, "Estimation in functional linear quantile regression," *Annals of Statistics*, vol. 40, no. 6, pp. 3108–3136, 2012.

[38] R. Koenker, *Quantile regression*. Cambridge university press, 2005.

[39] —, *Quantreg: quantile regression*, R package version 5.05, 2013. [Online]. Available: `http://CRAN.R-project.org/package=quantreg`.

[40] —, "Quantreg: quantile regression," *R package version*, vol. 5, no. 05, 2013.

[41] R. Koenker and G. Bassett, "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.

[42] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[43] D. Kong, K. Xue, F. Yao, and H. H. Zhang, "Partially functional linear regression in high dimensions," *Biometrika*, vol. 103, no. 1, pp. 147–159, 2016.

[44] E. R. Lee and B. U. Park, "Sparse estimation in functional linear regression," *Journal of Multivariate Analysis*, vol. 105, no. 1, pp. 1–17, 2012.

[45] G. Li, Y. Li, and C.-L. Tsai, "Quantile correlations and quantile autoregressive modeling," *Journal of the American Statistical Association*, vol. 110, no. 509, pp. 246–261, 2015.

[46] X. Li, H. Zhou, and L. Li, "Tucker tensor regression and neuroimaging analysis," *arXiv preprint arXiv:1304.5637*, 2013.

[47] H. Lian, "Functional partial linear model," *Journal of Nonparametric Statistics*, vol. 23, no. 1, pp. 115–128, 2011.

[48] —, "Shrinkage estimation and selection for multiple functional regression," *Statistica Sinica*, pp. 51–74, 2013.

[49] M. A. Lindquist, "The statistical analysis of fMRI data," *Statistical Science*, vol. 23, no. 4, pp. 439–464, 2008.

[50] M. A. Lindquist, J. M. Loh, L. Y. Atlas, and T. D. Wager, "Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling," *NeuroImage*, vol. 45, no. 1, S187–S198, 2009.

[51] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Mpca: multilinear principal component analysis of tensor objects," *Neural Networks, IEEE Transactions on*, vol. 19, no. 1, pp. 18–39, 2008.

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

[53] S. Ma, R. Li, and C.-L. Tsai, "Variable screening via quantile partial correlation," *Journal of the American Statistical Association*, no. just-accepted, pp. 00–00, 2016.

[54] V. M. Muggeo, M. Sciandra, and L. Augugliaro, "Quantile regression via iterative least squares computations," *Journal of Statistical Computation and Simulation*, vol. 82, no. 11, pp. 1557–1569, 2012.

[55] H.-G. Müller and F. Yao, "Functional additive models," *Journal of the American Statistical Association*, vol. 103, no. 484, pp. 1534–1544, 2008.

[56] Y. Nesterov, "Smooth minimization of non-smooth functions," *Mathematical Programming*, vol. 103, no. 1, pp. 127–152, 2005.

[57] D. V. Nguyen and D. M. Rocke, "On partial least squares dimension reduction for microarray-based classification: a simulation study," *Computational Statistics & Data Analysis*, vol. 46, no. 3, pp. 407–425, 2004.

[58] E. Niedermeyer and F. L. da Silva, *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[59] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols, *Statistical parametric mapping: the analysis of functional brain images: the analysis of functional brain images*. Academic Press, 2011.

[60] C. Preda and G. Saporta, "Pls regression on a stochastic process," *Computational Statistics & Data Analysis*, vol. 48, no. 1, pp. 149–158, 2005.

[61] J. O. Ramsay, *Functional data analysis*. Wiley Online Library, 2006.

[62] P. T. Reiss and R. T. Ogden, "Functional principal component regression and functional partial least squares," *Journal of the American Statistical Association*, vol. 102, no. 479, pp. 984–996, 2007.

[63] T. J. Rothenberg, "Identification in parametric models," *Econometrica*, pp. 577–591, 1971.

[64] B. Sánchez, H. Lachos, and V. Labra, "Likelihood based inference for quantile regression using the asymmetric laplace distribution," *J. Stat. Comput. Simul*, vol. 81, pp. 1565–1578, 2013.

[65] C. Scharinger, U. Rabl, H. H. Sitte, and L. Pezawas, "Imaging genetics of mood disorders," *NeuroImage*, vol. 53, no. 3, pp. 810–821, 2010.

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

[67] S. M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. E. Nichols, C. E. Mackay, K. E. Watkins, O. Ciccarelli, M. Z. Cader, P. M. Matthews, et al., "Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data," *NeuroImage*, vol. 31, no. 4, pp. 1487–1505, 2006.

[68] Y. Sun, "Semiparametric efficient estimation of partially linear quantile regression models," *Annals of Economics and Finance*, vol. 6, no. 1, p. 105, 2005.

[69] Q. Tang and L. Cheng, "Partial functional linear quantile regression," *Science China Mathematics*, vol. 57, no. 12, pp. 2589–2608, 2014.

[70] T. N. Tombaugh and N. J. McIntyre, "The mini-mental state examination: a comprehensive review.," *Journal of the American Geriatrics Society*, vol. 40, pp. 922–935, 1992.

[71] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, 2002.

[72] A. W. van der Vaart, *Asymptotic statistics*. Cambridge university press, 2000, vol. 3.

[73] A. W. van der Vaart and J. A. Wellner, *Weak Convergence*. Springer, 2000.

[74] H. Wold, "Soft modeling by latent variables: the nonlinear iterative partial least squares approach," *Perspectives in Probability and Statistics, papers in honour of MS Bartlett*, pp. 520–540, 1975.

[75] Y. Wu and Y. Liu, "Variable selection in quantile regression," *Statistica Sinica*, vol. 19, no. 2, p. 801, 2009.

[76] D. Yu, L. Kong, and I. Mizera, "Partial functional linear quantile regression for neuroimaging data analysis," *Neurocomputing*, vol. 195, pp. 74–87, 2016.

[77] K. Yu and R. A. Moyeed, "Bayesian quantile regression," *Statistics & Probability Letters*, vol. 54, no. 4, pp. 437–447, 2001.

[78] M. Yuan and T. T. Cai, "A reproducing kernel hilbert space approach to functional linear regression," *The Annals of Statistics*, vol. 38, no. 6, pp. 3412–3444, 2010.

[79] Y. Zhao, R. T. Ogden, and P. T. Reiss, "Wavelet-based lasso in functional linear regression," *Journal of Computational and Graphical Statistics*, vol. 21, no. 3, pp. 600–617, 2012.

[80] Y. Zhao, H. Chen, and R. T. Ogden, "Wavelet-based weighted lasso and screening approaches in functional linear regression," *Journal of Computational and Graphical Statistics*, vol. 24, no. 3, pp. 655–675, 2015.

[81] S. Zheng, "Gradient descent algorithms for quantile regression with smooth approximation," *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 3, pp. 191–207, 2011.

[82] H. Zhou and L. Li, "Regularized matrix regression," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 76, no. 2, pp. 463–483, 2014.

[83] H. Zhou, L. Li, and H. Zhu, "Tensor regression with applications in neuroimaging data analysis," *Journal of the American Statistical Association*, vol. 108, no. 502, pp. 540–552, 2013.

[84] H. Zhu, L. Kong, R. Li, M. Styner, G. Gerig, W. Lin, and J. H. Gilmore, "Fadtts: functional analysis of diffusion tensor tract statistics," *NeuroImage*, vol. 56, no. 3, pp. 1412–1425, 2011.

[85] H. Zhu, R. Li, and L. Kong, "Multivariate varying coefficient model for functional responses," *Annals of Statistics*, vol. 40, no. 5, pp. 2634–2666, 2012.

[86] H. Zhu, H. Zhang, J. G. Ibrahim, and B. S. Peterson, "Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance imaging data," *Journal of the American Statistical Association*, vol. 102, no. 480, pp. 1085–1102, 2007.

[87] H. Zou and M. Yuan, "Composite quantile regression and the oracle model selection theory," *Annals of Statistics*, vol. 36, no. 3, pp. 1108–1126, 2008.