ABSTRACT

WANG, KEHUI. Combined Estimation for Quantile Regression. (Under the direction of Huixia Judy Wang and Howard Bondell.)

Quantile regression offers a convenient tool to assess the relationship between a response and covariates in a comprehensive way and it is appealing especially in applications where interests are on the tails of the response distribution. However, due to data sparsity, the finite sample estimation at tail quantiles often suffers from high variability. To improve the tail estimation efficiency, we consider modeling multiple quantiles jointly for the cases where the quantile slope coefficients appear constant within a small quantile region at tails.

This dissertation comprehensively investigates various topics in the joint quantile regression. Chapter 1 gives an introduction to quantile regression. Chapter 2 studies the advantage and how to select the optimal weight when combining information across multiple quantile levels in the context of extreme quantiles. The data is assumed to be independent. Chapter 3 studies the joint quantile regression for estimating covariate dependent Value-at-Risk, an important measurement in finance. The time-dependency of the financial data makes the theory development fundamentally different from that in Chapter 2. Chapter 4 considers functional data which is essentially infinite-dimensional. We propose a Wald-type test statistic for testing the common slope hypothesis in functional linear regression model, which is an essential step for composite estimation.
Combined Estimation for Quantile Regression

by
Kehui Wang

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Statistics

Raleigh, North Carolina

2015

APPROVED BY:

Ana-Maria Staicu

Yichao Wu

Huixia Judy Wang
Co-chair of Advisory Committee

Howard Bondell
Co-chair of Advisory Committee
DEDICATION

To my family.
BIOGRAPHY

Kehui Wang was born on April 10th, 1988 in Ji’an of Jiangxi, China to two wonderful parents Yanming Wang and Haoyin Chen, along with her twin sister Keying. After living in Beijing for two years (1990 – 1992), her family moved to Guangzhou, which is the place she spent the next 18 years. She attended the National Scientific Base Class in Mathematics in Sun Yat-sen University in 2006, where she met her husband Meng Li. Kehui graduated with a Bachelor degree in Statistics in 2010. She then joined the Department of Statistics at North Carolina State University to pursue a doctoral degree in August 2010. Under the direction of Dr. Huixia (Judy) Wang, she will earn her Ph.D. degree in 2015.
ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Huixia (Judy) Wang. I am grateful for her immense help, great patience, and inspiring work attitude. This dissertation would not have been possible without her guidance. I have the highest regard for her deep understanding of statistics and the qualities of being professional and self-discipline. It has been a great pleasure working with her, and I have learned so much beyond research but more broadly for my future career.

I would also like to thank all my committee members, Howard Bondell Ana-Maria Staicu and Yichao Wu, for being supportive during the entire process. Their careful review, valuable comments and suggestions helped strengthen this dissertation. I also thank Huaiyu Dai for serving on my committee as the graduate school representative.

I would like to extend my sincere thanks to Ana-Maria Staicu and Arnab Maity, with whom we collaborate in the functional quantile linear regression project. I am grateful for all the mentorship and advices provided by them. My appreciation also goes to the DGPs of the department, Sujit Ghosh, Howard Bondell and Keem Weems, who are always ready to help. I should also thank all the staff, for their excellent service to the department.

My appreciation goes to Ellen Lentz, my supervisor at Genentech, and Veena Somayaji, my supervisor at Pfizer. They led me into the pharmaceutical industry and provided invaluable guidance to my future career.

I thank Yifang Li, Tian Chen, Liwei Wang and Bo Zhang and Changning Niu, for being my true friends of life.

Special thank to my father, who inspired me in the first place for higher education, and made me never give up. The same goes to my mom and sister, without whose unconditional love and accompany I couldn’t come this far. Last but not least, I have the heartiest gratitude to my husband and best friend Meng Li. We have been through all the good times and bad times together. Life has been changed, but I’m glad that you are always there for me.

This research was supported in part by the National Science Foundation Grant DMS-1149355.
# TABLE OF CONTENTS

List of Tables ................................................................. vii

List of Figures ................................................................. ix

Chapter 1 Introduction ....................................................... 1
   1.1 Introduction to Quantile Regression ................................. 1
   1.2 Introduction to Combined Quantile Regression ...................... 2
   1.3 Introduction to Functional Data Analysis ......................... 4

Chapter 2 Combined Estimation for Tail Quantile Regression .......... 6
   2.1 Introduction .............................................................. 6
   2.2 Proposed Methods ....................................................... 8
      2.2.1 Joint Quantile Regression Model ............................... 8
      2.2.2 Optimally Weighted Quantile Average Estimator .......... 10
      2.2.3 Weighted Composite Quantile Estimator .................. 12
      2.2.4 Estimation of EVI ................................................ 16
      2.2.5 Comparison of Asymptotic Efficiency ....................... 17
   2.3 Simulation Study ....................................................... 19
   2.4 Application to Chicago Precipitation Data ....................... 22
   2.5 Proof ................................................................. 24

Chapter 3 Combined Estimation for Conditional Value-at-Risk .......... 35
   3.1 Introduction .............................................................. 35
   3.2 CAViaR Model ............................................................ 37
   3.3 Combined Quantile Regression ....................................... 38
      3.3.1 Composite Quantile Loss Function ......................... 38
      3.3.2 Asymptotic Property ............................................. 39
   3.4 Computation ............................................................. 40
   3.5 VQR test ............................................................... 42
   3.6 Simulation ............................................................. 45
      3.6.1 Choices of Initial Values for the Conventional CAViaR Estimator 45
      3.6.2 Comparison of Local (Conventional) Estimation and Composite Estimation ................................. 47
      3.6.3 Assessment of Goodness of Fit via VQR Test ............... 48
   3.7 Data Application ....................................................... 49
   3.8 Proof of Theorem 3.3.1 ................................................. 53

Chapter 4 Inference in Functional Linear Quantile Regression .......... 60
   4.1 Introduction .............................................................. 60
   4.2 Methodology ............................................................ 62
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.1 Hypothesis tests in functional quantile regression model</td>
<td>62</td>
</tr>
<tr>
<td>4.2.2 Estimation procedure</td>
<td>65</td>
</tr>
<tr>
<td>4.2.3 Wald-type test statistic</td>
<td>66</td>
</tr>
<tr>
<td>4.3 Asymptotic results</td>
<td>66</td>
</tr>
<tr>
<td>4.3.1 Assumptions</td>
<td>66</td>
</tr>
<tr>
<td>4.3.2 Asymptotic distributions</td>
<td>67</td>
</tr>
<tr>
<td>4.3.3 Estimation of $p_0$</td>
<td>69</td>
</tr>
<tr>
<td>4.4 Simulation</td>
<td>70</td>
</tr>
<tr>
<td>4.5 Application</td>
<td>74</td>
</tr>
<tr>
<td>4.6 Proofs</td>
<td>77</td>
</tr>
<tr>
<td>4.6.1 Proof of Theorem 4.3.1</td>
<td>77</td>
</tr>
<tr>
<td>4.6.2 Proofs of Lemmas</td>
<td>80</td>
</tr>
</tbody>
</table>

References                                                              | 87   |
LIST OF TABLES

Table 2.1 The asymptotic relative efficiency of QAE, CRQ, WCRQ+ and OWCRQ with respect to OWQAE (OWCRQ) for five types of distributions. ........................................ 18
Table 2.2 Optimal weights for the WCRQ and WQAE methods at $\tau_k = 0.95$, $0.96$, $0.97$, $0.98$ and $0.99$. ......................................................... 18
Table 2.3 The $10^3 \times $MSE of different estimators of $\beta$ in Example 1. The values in the parentheses are the standard errors of $10^3 \times $MSE. ............. 21
Table 2.4 The $10^3 \times $MSE of different estimators of $\beta$ in Example 2. The values in the parentheses are the standard errors of $10^3 \times $MSE. ............. 21
Table 2.5 The $10^3 \times $MISE of different estimators in Example 3. The values in the parentheses are the standard errors of $10^3 \times $MISE. ............. 22
Table 2.6 The upper part of the table summarizes the estimated common slope $\hat{\beta}$ for the proposed methods and the bootstrap standard error (s.e.). The lower part of the table presents the quantile-specific weights given by different methods based on the estimated extreme value index $\hat{\xi} = 0.29$. .................................................. 24
Table 2.7 The prediction error of different methods at quantile levels $0.990$, $0.992$ and $0.995$. The values in the parentheses are the standard errors. 25

Table 3.1 The MSE, Bias1 and Bias2 of 4 methods using different initial values. 47
Table 3.2 The $1000 \times $MSE of the estimated $0.05$-th quantiles for composite and local estimations. The $1000 \times $standard errors are reported in the parentheses. ........................................ 49
Table 3.3 The $1000 \times $ MSE of the estimated $0.1$-th quantiles for composite and local estimations. The $1000 \times $ standard errors are reported in the parentheses. ........................................ 49
Table 3.4 The $100 \times $MSE of the composite and local estimations of the common parameter $\beta$. The $100 \times $ standard errors are reported in the parentheses. ........................................ 50
Table 3.5 The size of VQR test at $5\%$ and $10\%$ quantiles. The nominal levels are $0.1$ and $0.05$. ......................................................... 50
Table 3.6 The p-values of the VQR test for different methods on three data sets. 51
Table 3.7 Local and composite CAViaR estimators for the General Motor data. 52
Table 3.8 Local and composite CAViaR estimators for the IBM data. ........ 52
Table 3.9 Local and composite CAViaR estimators for the S&P 500 data. .... 52

Table 4.1 Type I error rates of the functional Wald-type test at given significance levels $\alpha$ when considering $U_1 = \{0.1, 0.2, 0.3, 0.4\}$. ............. 72
<table>
<thead>
<tr>
<th>Table 4.2</th>
<th>Type I error rates of the functional Wald-type test at given significance levels $\alpha$ when considering $U_2 = {0.1, 0.2, 0.6, 0.7}$.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 4.3</td>
<td>Type I errors when using the method of the naive multivariate quantile regression (NMQR), the mean as the scalar summary (SSQR) and Multivariate PCA (MPCA). When one method returns error (due to singularity of the design matrix) in more than 20% replications, we report it as “NA”.</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>Bootstrap standard errors of the estimates of 4 coefficients including the intercept $\beta_0$ from the QAE, CRQ and the local quantile regression estimation at the 0.9-th quantile (RQ). The method of FPCA selects 3 PC’s for all 1000 bootstrap samples.</td>
</tr>
<tr>
<td>Table 4.5</td>
<td>Mean prediction errors from different methods over 1000 cross-validations.</td>
</tr>
</tbody>
</table>
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Local quantile regression estimates of the slope at upper quantiles for the precipitation data in Aurora station.</td>
<td>23</td>
</tr>
<tr>
<td>4.1</td>
<td>Powers of FPCA and SSQR in various scenarios.</td>
<td>73</td>
</tr>
<tr>
<td>4.2</td>
<td>Hourly bike rentals for casual users. The $x$-axis is the hour and the $y$-axis is the hourly total count of bike rentals for casual users.</td>
<td>75</td>
</tr>
<tr>
<td>4.3</td>
<td>$eta(t)$ at various quantile levels.</td>
<td>75</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Introduction to Quantile Regression

In statistics, regression is an important technique to model the relationship between response (dependent variable) and predictors (covariates or independent variables). The traditional estimation is the ordinary least squares (OLS), which estimates the parameters by minimizing the sum of squared errors. Consider a linear regression model

\[ y_i = \mathbf{x}_i^T \beta + \epsilon_i, \quad i = 1, \ldots, n, \]

where \( y_i \) is the response, \( \mathbf{x}_i \) is the \( p \)-dimensional predictor, \( \beta \) is the \( p \)-dimensional parameter, and \( \epsilon_i \) is the error with mean 0. The OLS estimation of \( \beta \) is defined as:

\[ \hat{\beta} = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \beta)^2. \]

The OLS estimator has a closed form \( \hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \), where \( \mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T \) is the design matrix and \( \mathbf{Y} = (y_1, \ldots, y_n)^T \) is the vector of responses.

Even though OLS is the most popular regression estimation, it has some limitations. First, the OLS is not robust to outliers. When we have extreme observations or the response distribution has a heavy tail, the OLS estimation can be easily distorted. Second, the OLS only models the mean of the response as a function of the predictor, but in lot of cases, people are concern about the tail behavior of the response distribution. Examples can be found in many different research areas such as in environmental studies (Uppala...
et al., 2005) and in financial risk management (Chernozhukov and Du, 2008).

Quantile regression, first introduced by Koenker and Bassett (1978), was a valuable alternative to the OLS. Different from the OLS, quantile regression models the conditional quantile of the response as a function of the predictor. Let $Y$ be the scalar response variable, $X$ be the $p$-dimensional vector of predictors, and $\{(y_i, x_i)\}_1^n$ be a random sample of $(Y, X)$. For a given quantile level $\tau \in (0, 1)$, consider the linear quantile regression model

$$Q_Y(\tau|x_i) = \alpha_0(\tau) + x_i^T \beta_0(\tau),$$

where $Q_Y(\tau|x) = \inf\{y : F_Y(y|x) \geq \tau\}$ is the $\tau$-th conditional quantile of $Y$ given $X = x$, and $F_Y(\cdot|x)$ is the conditional distribution function. The conventional quantile regression estimator of the parameters $\alpha_0(\tau)$ and $\beta(\tau)$ is defined as:

$$\{\hat{\alpha}(\tau), \hat{\beta}(\tau)\} = \arg\min_{(\alpha, \beta) \in \mathbb{R}^{p+1}} \sum_{i=1}^n \rho_\tau(y_i - \alpha - x_i^T \beta),$$

(1.1)

where $\rho_\tau(t) = t(\tau - I(t \leq 0))$ is the quantile loss function and $I(\cdot)$ is the indicator function. The objective function is minimized via linear programming.

Quantile regression offers a convenient tool to access the relationship between a response and covariates in a comprehensive way by varying the quantile level $\tau$. Consequently, quantile regression is appealing especially in applications where interests are on the tails of the response distribution.

### 1.2 Introduction to Combined Quantile Regression

Combined estimation was first proposed by Hogg (1980), and some asymptotic properties of the combined estimators were studied by Koenker (1984). Consider the classic independent and identically distributed (i.i.d.) error model

$$y_i = \alpha_0 + x_i^T \beta_0 + \epsilon_i, \ i = 1, \ldots, n,$$

which implies

$$Q_Y(\tau|x_i) = \alpha_0(\tau) + x_i^T \beta_0,$$

(1.2)

where $\alpha_0(\tau) = \alpha_0 F^{-1}_\epsilon(\tau)$ and $F^{-1}_\epsilon(\cdot)$ is the cumulative distribution function (CDF) of $\epsilon_i$. 

2
Note that the above model (1.2) implies that the quantile intercept $\alpha_0(\tau)$ is varying across different $\tau$, while the slope is a constant. This equal slope feature motivates the estimation of the common slope by combining information across multiple quantiles.

Consider $K$ quantile levels $0 < \tau_1 < \ldots, \tau_K < 1$, the vector of true parameters is $\theta_0 = \{\alpha_0(\tau_1), \ldots, \alpha_0(\tau_K), \beta_0^T\}^T$. There are two ways to aggregate information across quantile levels. Koenker (1984) proposed the weighted quantile average estimator (WQAE). First, we obtain quantile regression estimators by minimizing the quantile loss function in (1.1) at $\tau_1, \ldots, \tau_K$ separately. Each $\hat{\beta}(\tau_k), \ k = 1, \ldots, K$ is a consistent estimator of $\beta_0$, and the WQAE estimator of the slope is defined as

$$\hat{\beta}_{WQAE} = \sum_{k=1}^K \omega_k \hat{\beta}(\tau_k),$$

where $\omega_k$ is the weight assigned to quantile level $\tau_k$ and $\sum_k \omega_k = 1$.

Hogg (1980) proposed the weighted composite regression of quantile (WCRQ) estimator, which is obtained by minimizing the weighted average of the quantile loss function

$$(\hat{\alpha}(\tau_1), \ldots, \hat{\alpha}(\tau_K), \hat{\beta}) = \arg \min_{(\alpha_1, \ldots, \alpha_K, \beta) \in \mathbb{R}^{K+p}} \sum_{k=1}^K \sum_{i=1}^n \omega_k \rho_{\tau_k}(y_i - \alpha_k - x_i^T \beta), \quad (1.3)$$

where $\omega_k$ is the weight corresponding to $\tau_k$ and $\sum_k \omega_k = 1$.

Koenker (1984) derived the asymptotic distributions of the WQAE and WCRQ estimators under i.i.d. error assumption. The optimal weights of these estimators can be chosen by minimizing the asymptotic variances. Koenker (1984) showed that these two estimators achieve the same asymptotic efficiency when the optimal weights are used. This encouraging result suggests the potential efficiency gain by combining information across quantiles. However, the current literature on combined quantile regression only focused on the i.i.d. error models or on central quantiles. We study combined quantile regression at extreme quantiles for independent data in Chapter 2. In Chapter 3, we investigate combined quantile regression for estimating covariate-dependent Value-at-Risk (VaR) for time series data with autocorrelation. In Chapter 4, we propose a method for testing the common slope assumption in functional linear quantile regression models, which is an essential step before carrying out combined quantile estimation.
1.3 Introduction to Functional Data Analysis

Recent development of technology and computational capacity allows us to record data with repeated measurement in high frequency, making it natural to conceptually regard data as a function. In functional data analysis (FDA), the data is viewed as a realization of an underlying random process \(X(t)\) in \(L^2(\mathcal{T})\), where \(L^2(\mathcal{T})\) is the \(L^2\) space on \(\mathcal{T}\) and \(\mathcal{T} \subset \mathbb{R}\) is a bounded closed interval.

Suppose we have \(n\) i.i.d. observations \(\{X_1(t), X_2(t), \ldots, X_n(t) : t \in \mathcal{T}\}\), where \(E\{X_i(t)\} = \mu(t)\) and \(\text{Cov}\{X_i(s), X_i(t)\} = G(s, t), i = 1, \ldots, n\). It is well known that the covariance kernel \(K(s, t)\) has orthogonal expansion \(K(s, t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s) \phi_j(t)\), where the eigenvalues \(\lambda_j\) are nondecreasing and nonnegative, and \(\phi_j\)'s are the associated eigenfunctions. Here \(\{\phi_j : j = 1, \ldots, \infty\}\) forms an orthonormal basis of \(L^2(\mathcal{T})\). Then by Karhunen-Loève expansion, the random process \(X_i(t)\) can be represented as

\[
X_i(t) = \mu(t) + \sum_{j=1}^{\infty} z_{i,j} \phi_j(t),
\]

(1.4)

where \(z_{i,j} = \int_0^1 \{X_i(t) - \mu(t)\} \phi_j(t) dt\) is the so-called principle scores (PC’s) for \(X_i\) satisfying that \(E(z_{i,j}) = 0, \text{Var}(z_{i,j}) = \lambda_j\) and \(E(z_{i,j} z_{i,j'}) = 0 (j \neq j')\).

For each individual functional data, what we actually observe are \(m_i\) discrete measurements with measurement error at time point \(t_1, \ldots, t_{m_i} \in \mathcal{T}\), denoted as

\[
W_{i,j} = X_i(t_{i,j}) + \epsilon_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m_i,
\]

where the measurement error \(\epsilon_{i,j}\) has mean 0 and variance \(\sigma^2_w\).

Dimension reduction is critical in FDA, and among various methods, functional principal component analysis (FPCA) is the most widely used technique. Given the observed data \(\{W_{ij}, i = 1, \ldots, n, j = 1, \ldots, m_i\}\), one can first estimate the mean and covariance kernel function via commonly used smoothing techniques such as local polynomial. Then with the estimated eigenfunctions \(\hat{\phi}_j(t)\), we can recover the scores \(z_{i,j}\) by \(\hat{z}_{i,j}\) with approximate numerical integrals. Truncation is also needed to make the problem feasible. In equation (1.4), we need to select the number of PC’s according to some criteria such as the percentage of variance, AIC-type criteria or cross validation. With the truncation and the recovered scores, we are able to describe the functional data approximately in a
multivariate setting.

Functional quantile regression (fQR) model is an extension of the standard linear quantile regression to functional covariates. We consider the i.i.d. data of \((Y_i, X_i)\), where \(Y_i\) is a scalar response variable and \(X_i = X_i(t), t \in \mathcal{T}\) is a random covariate function. Then we assume that \((Y_i, X_i)\) is generated from the following model

\[
Y_i = \alpha(\tau) + \int_0^1 \beta(t, \tau) X_i^c(t) dt + \epsilon_i
\]  \hspace{1cm} (1.5)

where \(\beta(\cdot, \tau) \in L^2(\mathcal{T}), X_i^c(t) = X_i(t) - E\{X_i(t)\}\) and the \(\tau\)th quantile of the i.i.d. random error \(\epsilon_i\) is 0. Let \(Q_{Y|X}(\tau)\) be the conditional \(\tau\)-th quantile function of \(Y\) given \(X\) for a given \(\tau\), then equation (1.5) implies that

\[
Q_{Y|X_i}(\tau) = \alpha(\tau) + \int_0^1 \beta(t, \tau) X_i^c(t) dt,
\]  \hspace{1cm} (1.6)

which is the fQR model.

The coefficient function \(\beta(t, \tau)\) can be expanded by the eigenfunction basis as \(\beta(t, \tau) = \sum_{k=1}^{\infty} \beta_j(\tau) \phi_j(t)\), where \(\beta_j(\tau) = \int_0^1 \beta(t, \tau) \phi_j(t) dt\). Suppose we have \(p_n\) PC’s to be selected, then the fQR model is reduced to

\[
Q_{Y_i|X_i}^{(p_n)}(\tau) = \alpha(\tau) + \sum_{j=1}^{p_n} \beta_j(\tau) z_{i,j}.
\]

Even though it seems to be similar, the fQR model is fundamentally different than the case of the regular multivariate QR model, because the predictors (the scores) are estimated rather than observed. One may view the differences between the estimated predictors and true scores as measurement errors in predictors, however, these differences are dependent across subjects and correlated with the true scores, violating the common independence assumption in the classic covariate measurement error literature. This makes it challenging to establish the asymptotic theory allowing for inference. In chapter 4, we investigate the asymptotic distribution of proposed estimators under the fQR model, and show that the estimators are still unbiased but the variance is inflated compared to the case of multivariate QR models. A Wald-type test to test the coefficient constancy across quantile levels is also considered, where we show the inflated variance will not have an effect on the null distribution of the test statistic.
Chapter 2

Combined Estimation for Tail Quantile Regression

2.1 Introduction

Quantile regression has generated tremendous research interests after being introduced by Koenker and Bassett (1978). Different from the conventional least squares regression, quantile regression models the $\tau$th conditional quantile of the response as a function of the covariates, where $0 < \tau < 1$ is the prespecified quantile level of interest.

An important problem in many fields is to model rare and extreme phenomena, which correspond to the lower or upper tails of the variable. For instance, in environmental studies, extremely low or high precipitations are of more importance than average precipitations, since hydrologists may want to model the chance of draught to help with the design of reservoirs, and model the chance of heavy rainfall to help with the design of flood drains (Pandey and Nguyen, 1999; Friederichs and Hense, 2007; Friederichs, 2010; Wang et al., 2012). In the analysis of infant birth weights (Abrevaya, 2002), extremely low birthweights are associated with various health problems and higher infant mortality rates. In financial risk management, investors are more interested in forecasting extremely large financial losses of an institution’s portfolio given today’s available information (Chernozhukov and Du, 2008). Without loss of generality, in this paper, we focus on high quantile regression.

The main challenge for tail quantile regression is data sparsity, especially for heavy-tailed distributions where the disturbance of errors remains strong in the far tail. By
fitting the regression model at one quantile level at a time, the \textit{conventional quantile regression}, also referred to as the \textit{local quantile regression} thereafter, is often unstable at tails. In applications where the covariate effects have some common features across quantile levels in the tail region, it is desirable to aggregate information across multiple quantiles to improve the estimation efficiency over the local quantile estimator (i.e. conventional quantile-specific estimator). For instance, Figure 4.1 in Section 4 plots the conventional quantile regression estimates of slopes at multiple quantiles, where the response is the observed daily precipitation in Chicago area and the predictor is the simulated daily precipitation from the ERA-40 reanalysis model. The estimated quantile slopes appear to be constant in the upper quantiles \( \tau \in (0.990, 0.995) \). For such data sets, we could utilize the commonality of quantile slopes to improve the estimation efficiency at tails by pooling information across tail quantiles.

Consider a linear regression model with quantile-invariant covariate effects, there exist two plausible ways to combine information across quantiles: combining the local quantile estimators or the criterion functions involved in the estimation procedure at different quantiles. The first strategy leads to the weighted quantile average estimator (WQAE) introduced by Koenker and Bassett (1978), which is the weighted average of quantile-specific slope estimators. The second strategy leads to the weighted composite regression of quantiles (WCRQ) estimator first proposed by Hogg (1980), which minimizes the combined quantile objective function across quantiles. For central quantiles with \( 0 < \tau < 1 \), Koenker (1984) studied the asymptotic properties of these two estimators and showed their asymptotic equivalency when optimal weights are used. In recent years, combined quantile regression has been studied in various setups with more work focusing on the first strategy. The composite quantile regression idea was employed for estimation and variable selection in linear regression (Zou and Yuan, 2008; Tang et al., 2012a), nonlinear regression (Jiang et al., 2012c), nonparametric regression (Kai et al., 2010, 2011; Guo et al., 2012; Jiang et al., 2012b, 2013b), and linear regression with censored data (Jiang et al., 2012a; Tang et al., 2012b). Based on the combined quantile loss function, Jiang et al. (2013a) proposed two penalization methods that perform estimation, detection of the interquantile slope commonality and variable selection simultaneously. Zhao and Xiao (2014) discussed both WCRQ and WQAE methods for linear and nonparametric regression models.

Previous works on combined quantile regression have been restricted to central quant-
tiles with \( \tau \in [\epsilon, 1 - \epsilon] \), where \( 0 < \epsilon < 1 \) is some positive constant, and this rules out the study of extreme tails of the response distribution. To our knowledge, there exists no discussion about how to optimally combine information across tail quantiles. At the tails with quantile level \( \tau \to 1 \) as the sample size \( n \to \infty \), the convergence rate of quantile regression estimator depends on the heaviness of the tails of the response distribution and is slower than root-\( n \). Therefore, the methods and theory developed in Koenker (1984) and Zhao and Xiao (2014) are not applicable for tail quantiles. In this paper, using the tools of extreme value theory, we establish the asymptotic properties of the weighted composite and weighted quantile average estimators for tail quantile regression, and propose a procedure for estimating the optimal weights for both estimators.

The rest of this chapter is organized as follows. In Section 2, we propose two weighted estimators for tail quantile regression with constant slopes, present their asymptotic properties and discuss the construction of optimal weights. The numerical performance of the proposed estimators is assessed through a simulation study in Section 3 and the analysis of a daily precipitation data in Chicago area in Section 4. All technical details are provided in the Appendix.

2.2 Proposed Methods

2.2.1 Joint Quantile Regression Model

Let \( Y \) be the scalar response variable, \( X \) be the \( p \)-dimensional vector of covariates, and \( \{(y_i, x_i)\}_{i=1}^n \) be a random sample of \((Y, X)\). Suppose we are interested in regression at the upper tails with quantile level \( \tau \in \mathcal{T} = [1 - \epsilon_n, 1] \), where \( \epsilon_n \to 0 \) as \( n \to \infty \). Let \( F_Y(\cdot|\mathbf{x}) \) denote the conditional distribution function of \( Y \) given \( X = \mathbf{x} \). The linear quantile regression model assumes that

\[
Q_Y(\tau|\mathbf{x}) = \alpha_0(\tau) + \mathbf{x}^T \beta_0(\tau), \quad \tau \in \mathcal{T},
\]

(2.1)

where \( Q_Y(\tau|\mathbf{x}) = \inf\{y : F_Y(y|x) \geq \tau\} \) is the \( \tau \)th conditional quantile of \( Y \) given \( X = \mathbf{x} \), and \( \alpha_0(\tau) \in \mathbb{R} \) and \( \beta_0(\tau) \in \mathbb{R}^p \) are the unknown quantile coefficients associated with the \( \tau \)th quantile level.

At a given quantile level \( \tau \), the conventional quantile regression estimator of
\((\alpha(\tau), \beta(\tau))\) is defined as

\[
\left(\hat{\alpha}(\tau), \hat{\beta}(\tau)\right) = \text{argmin}_{(\alpha, \beta) \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \rho_{\tau}(y_i - \alpha - x_i^T \beta),
\]

(2.2)

where \(\rho_{\tau}(t) = t\{\tau - I(t \leq 0)\}\) is the quantile loss function and \(I(\cdot)\) is the indicator function (Koenker, 2005). The conventional quantile regression method estimates the quantile coefficient at each quantile level of interest separately, and the resulting local slope estimator \(\hat{\beta}(\tau)\) can vary freely in \(\tau\). However, in data-sparse area such as extreme tails, the variability of local estimates is often overly high. In some applications it might be reasonable to assume the slope coefficient \(\beta(\tau)\) to share some common features, for instance, to be constant or locally linear, within a certain region of quantiles. By utilizing this commonality, we can aggregate information across quantiles to improve the estimation efficiency. In this paper, we focus on linear quantile regression with constant slopes at the upper tails, but the proposed method can also be easily extended to lower tails, and also be adapted to accommodate other common features such as local linearity, and cases where only a subset of the components of \(\beta(\tau)\) are locally constant.

We assume the following linear quantile regression model at the upper tails

\[
Q_Y(\tau | x) = \alpha_0(\tau) + x^T \beta_0, \quad \tau \in \mathcal{T}.
\]

(2.3)

Different from model (2.1), here the quantile slope is assumed to be constant at the upper tails across \(\tau \in \mathcal{T}\), but \(\alpha(\tau)\) is still an increasing function of \(\tau\).

A number of researchers have considered estimation for models similar to (2.3), but with constant slopes across \(\tau \in (0, 1)\); for instance, see Kai et al. (2010), Kai et al. (2011), Guo et al. (2012), Jiang et al. (2012a), Tang et al. (2012b), Jiang et al. (2012b), Jiang et al. (2013b). These work have focused on central quantiles with \(\tau\) strictly greater than zero and less than one. In contrast, we focus on upper tail quantiles that are allowed to approach one as sample size increases. Based on the extreme value theory, we develop two optimally combined estimators for tail quantile regression, which are obtained by aggregating information across \(K\) quantiles \(\tau_1 < \tau_2 < \cdots < \tau_K \in \mathcal{T}\).
2.2.2 Optimally Weighted Quantile Average Estimator

For model (2.3) at $K$ upper quantiles, the unknown parameters consist of $K$ distinct intercepts $\alpha_0(\tau_k)$ and one common slope vector $\beta_0$. Denote the vector of true parameters $\theta_0 = (\alpha_{0,1}, \ldots, \alpha_{0,K}, \beta_0^T)^T$, where $\alpha_{0,k} = \alpha_0(\tau_k), k = 1, \ldots, K$.

Let $\varpi_k$ be the weight assigned to the quantile level $\tau_k$, $k = 1, \ldots, K$. For the sake of identifiability, we assume that $1^T_K \varpi = 1$, where $\varpi = (\varpi_1, \ldots, \varpi_K)^T$, and $1_K$ denotes a $K$-dimensional vector of ones. We define the weighted quantile average estimator of $\theta$ as

$$\hat{\beta}_{WQAE} = \sum_{k=1}^{K} \varpi_k \hat{\beta}(\tau_k), \tag{2.4}$$

where $\hat{\beta}(\tau_k)$ is the local quantile slope estimator obtained by minimizing the objective function in (2.2) at the $\tau_k$th quantile. The WQAE is the weighted average over conventional estimators at multiple high quantiles, and thus can be viewed as a special case of the L-estimator with discrete weights. Portnoy and Koenker (1989) and Koenker (1984) studied the L-estimator for the slope in the location-shift linear model, which implies that $\beta(\tau)$ is constant across the entire quantile region $\tau \in (0, 1)$.

The asymptotic properties of WQAE at central quantiles have been studied by Koenker (1984) and Zhao and Xiao (2014). In the following Theorem 2.2.1, we establish the asymptotic distribution of $\hat{\beta}_{WQAE}$ at tails.

For any sequences $a(z)$ and $b(z)$, throughout the paper, we use the notation $a(z) \sim b(z)$ to mean that $a(z)/b(z) \to 1$ as a specified limit is taken over $z$. We make the following assumptions.

A1. The distribution function $F_Y(y|x_i)$ is absolutely continuous with continuous density $f_Y(y|x_i)$, which is uniformly bounded away from zero and infinity and has a bounded first derivative around $Q_Y(\tau_k|x_i)$ for any $k = 1, \ldots, K$.

A2. The distribution of $X$ has a compact support $X$. The expectation $E(X) = 0$, and $E(XX^T) = D$ exists and is positive definite.

A3. Let $U = Y - X^T \beta_0$. There exists some distribution function $F_0(\cdot)$ in the maximum domain of attraction with extreme value index $\xi$, such that $1 - F_U(z|x) \sim 1 - F_0(z)$ as $z \to s_u$ uniformly in $x \in X$, where $s_u$ is the upper end-point of $U$.

A4. $\partial F_0^{-1}(1 - \tau)/\partial \tau$ is regularly varying at 0 with exponent $-\xi - 1$. 

10
A5. For $k = 1, \ldots, K$, $(1 - \tau_k)/(1 - \tau) = l_k + o(1)$ for some $\tau \to 1$ and $n(1 - \tau) \to \infty$, where $l_k > 0$ are constants.

Condition A1 specifies some conditions on the conditional distribution of $Y$, which are standard in quantile regression. Condition A2 assumes that the covariate vector $X$ has mean zero. In general, for cases where $E(X)$ is not zero, one can simply center the covariate to have mean zero before carrying out the data analysis. Conditions A3-A4 specify some assumptions on the tail behaviors of the conditional distribution of $Y$, which are needed to establish the properties of the quantile coefficient estimator at tails. Specifically, A3 requires that after the linear transformation, $U = Y - X^T \beta_0$ is tail equivalent to a distribution $F_0$ in the maximum domain of attraction. The domain of attraction assumption is not very restrictive since it covers most of the common distributions such as Gaussian, Beta, and $t$ distribution, and so on; see de Haan and Ferreira (2006) for more details about the domain of attraction. Condition A4 is a von Mises type condition, which specifies how the density function decays at the right tail. The von Mises condition is basic for a distribution to belong to the maximum domain of attraction.

Condition A5 restricts our attention to the intermediate order of extreme quantiles with $\tau_k \to 1$ and $n(1 - \tau_k) \to \infty$ as $n \to \infty$.

Remark 1. Extreme value index $\xi$ is a measurement of the heaviness of the tail distribution. According to the signs of the extreme value indices, distributions in the domain of attraction can be categorized into 3 classes: heavy-tailed, light-tailed and short-tailed. A distribution with $\xi > 0$ has infinite right endpoint and a heavy tail. While a distribution with $\xi = 0$ also has infinite right endpoint but the tail is light. As for $\xi < 0$, the right endpoint of the distribution is $-1/\xi$, so the distribution has a short tail. For more details about extreme value index, we refer to de Haan and Ferreira (2006, chapter 1).

Throughout this paper, we define

$$a_n = \sqrt{(1 - \tau)n F^{-1}_0(\tau) - F^{-1}_0(\tilde{\tau}_m)},$$

where $\tilde{\tau}_m = 1 - m(1 - \tau)$ for some $m > 1$.

**Theorem 2.2.1** Suppose that model (2.3) and conditions A1-A5 hold. Let $\omega = \ldots$
$(\varpi_1, \ldots, \varpi_K)^T$, as $n \to \infty$, we have

$$a_n(\hat{\beta}_{WQAE} - \beta) \xrightarrow{d} N \left(0, \sigma^2_{WQAE}(\varpi) \left(\frac{m^{-\xi} - 1}{-\xi}\right)^{-2} D^{-1}\right),$$

$$\sigma^2_{WQAE}(\varpi) = \varpi^T \Phi^{-1}(\xi) \Gamma \Phi^{-1}(\xi) \varpi,$$

where $\Gamma$ is a $K \times K$ matrix with the $(k,k')$th element as $\min(l_k, l_{k'})$, $\Phi(\xi) = \text{diag}\{(l_1^{\xi+1}, \ldots, l_K^{\xi+1})\}$, and $D = E(XX^T)$.

Theorem 2.2.1 suggests that the asymptotic covariance of $\hat{\beta}_{WQAE}$ depends on the weights $\varpi$ only through a scalar $\sigma^2_{WQAE}(\varpi)$, which is a function of $\varpi$. Therefore, the optimal weights that maximize the efficiency of $\hat{\beta}_{WQAE}$ is

$$\varpi_{\text{opt}} = \arg\min_{\varpi} \varpi^T \Phi^{-1}(\xi) \Gamma \Phi^{-1}(\xi) \varpi, \text{ subject to } 1_K^T \varpi = 1. \quad (2.5)$$

The minimization in (2.5) is a standard constrained optimization problem. Direct application of the Lagrange multipliers method gives

$$\varpi_{\text{opt}} = \Phi(\xi) \Gamma^{-1} \{\phi^T(\xi) \Gamma \phi(\xi)\}^{-1}. \quad (2.6)$$

Therefore, the minimal value of $\sigma^2_{WQAE}$ is $\{\phi^T(\xi) \Gamma \phi(\xi)\}^{-1}$. We refer to the optimal WQAE of $\beta_0$ based on $\varpi_{\text{opt}}$ as $\hat{\beta}_{\text{OWQAE}}$.

### 2.2.3 Weighted Composite Quantile Estimator

An alternative estimator of the common slope $\beta$ is the weighted composite regression of quantiles (WCRQ) defined as

$$\hat{\theta}_{WCRQ} = \text{argmin}_{\alpha_1, \ldots, \alpha_K, \beta} \sum_{k=1}^K \varpi_k \sum_{i=1}^n \rho_{\tau_k}(y_i - \alpha_k - x_i^T \beta), \quad (2.7)$$

where $\varpi_k$ is the weight assigned to the quantile level $\tau_k$. The estimator $\hat{\theta}_{WCRQ}$ is referred to as the weighted composite regression of quantiles (WCRQ) estimator, since it minimizes the weighted joint quantile objective function across quantile levels.

The weighted composite estimator $\hat{\theta}_{WCRQ}$ depends on the prespecified weights $\varpi_k$. Focusing on central quantiles, Zou and Yuan (2008), Kai et al. (2011), Jiang et al. (2013a)
considered a combined quantile objective function that is similar to the one in (2.7) but assigns equal weights to different quantiles. However, since neighboring quantiles are correlated especially at the tails, in general assigning equal weights is not an efficient way to combine information across quantiles. In some cases, composite estimator based on equal weights could even be less efficient than the local quantile estimator obtained at a single quantile level. Therefore, it is important to select appropriate weights \( \omega_k \) in order to achieve efficiency gain.

Note that when the components in \( \omega_{opt} \) are non-negative, the combined objective function in (2.7) is convex. The optimization in (2.7) can be recast into a linear programming problem and solved by using existing software such as the function “make.lp” in the R package \texttt{lpSolveAPI} or function “rq.fit.fnb” in the R package \texttt{quantreg}. But when some weights are negative, the objective function in (2.7) might not be convex and will lead to difficulty in minimization. Therefore, to avoid potential computational difficulty, we first consider WCRQ with nonnegative weights. The following theorem establishes the asymptotic properties of the weighted composite estimator at tails with given nonnegative weights.

\textbf{Theorem 2.2.2} Suppose that model (2.3) and conditions A1-A4 hold, and \( \omega_k \geq 0 \) for \( k = 1, \ldots, K \), as \( n \to \infty \),

\[
\frac{a_n(\hat{\beta}_{WCRQ} - \beta_0)}{\sqrt{n} \sigma^2_{WCRQ}(\omega)} \xrightarrow{d} N \left( 0, \sigma^2_{WCRQ}(\omega) \left( \frac{m^{-\xi} - 1}{-\xi} \right)^{-2} D^{-1} \right),
\]

\[
\sigma^2_{WCRQ}(\omega) = \{ \omega^T \phi(\xi) \}^{-2} \omega^T \Gamma \omega,
\]

where \( \phi(\xi) = (l_1^{\xi+1}, \ldots, l_K^{\xi+1})^T \).

\textbf{Sub-optimal Estimator.} Similar to WQAE, the optimal weights for WCRQ can be obtained by minimizing the asymptotic variance of \( \hat{\beta}_{WCRQ} \), which depends on the weights only through a scalar \( \sigma^2_{WCRQ}(\omega) \). We define

\[
\omega_{sub} = \arg\min_{\omega_1 \geq 0, \ldots, \omega_K \geq 0} \{ \omega^T \phi(\xi) \}^{-2} \omega^T \Gamma \omega, \quad \text{subject to } \mathbf{1}_K^T \omega = 1. \quad (2.8)
\]

We refer to \( \omega_{sub} \) as the sub-optimal vector of weights since it minimizes the asymptotic variance of \( \hat{\beta}_{WCRQ} \) under the nonnegative restriction. The optimization in (2.8) is a standard quadratic programming problem, and thus can be solved by existing quadratic pro-
gramming software such as the function “solve.QP” in the R package quadprog. Throughout, we denote the minimizer of (2.7) based on $\omega_{\text{sub}}$ as $\hat{\theta}_{\text{WCRQ+}}$.

Compared with OWQAE, the additional restriction in WCRQ+ estimator may cause some loss of efficiency for the estimation of $\beta$ as indicated in Proposition 2.2.1; also see some numerical evidences in Section 2.5. If we release the nonnegative constraint, we can obtain the “optimal” weight that minimizes $\sigma^2_{\text{WCRQ}}$ as

$$\omega_{\text{opt}} = \Gamma^{-1}\phi(\xi) / \{1^T_k\Gamma^{-1}\phi(\xi)\}. \quad (2.9)$$

Recall that the minimum value of $\sigma^2_{\text{WQAE}}$ is $\{\phi^T(\xi)\Gamma\phi(\xi)\}^{-1}$, which is the same as that of the OWQAE estimator. Note that $\omega_{\text{opt}}^T\phi(\xi) = \phi^T(\xi)\Gamma^{-1}\phi(\xi) \geq 0$. Following the similar argument as in the proof of Theorem 2.2.2, we can show that the asymptotic normality in Theorem 2.2.2 still holds for the WCRQ estimator based on the weights $\omega_{\text{opt}}$, and this estimator has the same asymptotic variance with OWQAE. Thereafter, we refer to the WCRQ estimator based on $\omega_{\text{opt}}$ as the OWCRQ (optimally weighted composite regression of quantile).

**Proposition 2.2.1** For any weight $\omega \in \mathbb{R}^K$, $\sigma^2_{\text{WCRQ}}(\omega) \geq \{\phi^T(\xi)\Gamma\phi(\xi)\}^{-1}$, and the equality holds if and only if $\omega = \omega_{\text{opt}}$.

**Remark 2.** In some situations, upper quantiles may be negatively correlated, resulting in negative optimal weights. In such cases, using WCRQ based on sub-optimal weights is not as as efficient as the OWCRQ based on the optimal weights or the OWQAE estimator.

For tail quantile regression, we find out that the occurrence of negative optimal weights depends on the heaviness of the tail of the response distribution. Using linear algebra, we can show that when the response distribution is heavy-tailed with $\xi > 0$, the least extreme quantile will receive positive weight and all the others receive negative weights. For light-tailed distributions with $\xi = 0$, as $\tau \to 1$, all the weights will be put into the least extreme quantile level $\tau_1$ while other higher quantiles receive zero weight. For short-tailed distributions with $\xi < 0$, all the weights at high quantiles are positive. Some justification for the above findings are provided in the Appendix. For illustration, Table 2.2 presents the optimal weights for normal, $t_2$, $t_1$, Beta(2, 5) and Beta(2, 2.5) distributions, which are in the domain of attractions with $\xi = 0, 0.5, 1, -0.2$ and -0.4, respectively.

**One-step Estimator.** When $\omega_{\text{opt}}$ contains negative weights, the weighted objective function in (2.7) may be non-convex, making it difficult to be minimized to obtain the
OWCRQ estimator. To utilize the optimal weights $\omega_{\text{opt}}$ and meanwhile avoid the non-convex optimization, we consider an alternative one-step estimator. Let $z_{i,k} = (e_k^T, x_i^T)^T$, where $e_k$ is a $K$-dimensional vector with the $k$th entry equals 1 and the others equal 0. Denote

$$A(\theta) = \sum_{k=1}^K \sum_{i=1}^n \omega_k^{(o)} z_{i,k} \{I(y_i - z_{i,k}^T \theta < 0) - \tau_k\}, B(\theta) = \sum_{k=1}^K \sum_{i=1}^n \omega_k^{(o)} z_{i,k} z_{i,k}^T f_Y(z_{i,k}^T \theta | x_i),$$

where $\omega_k^{(o)}$ is the $k$th element of $\omega_{\text{opt}}$. The one-step estimator of $\theta_0$ is defined as

$$\hat{\theta}_{\text{OS}} = \tilde{\theta} - \left\{ B(\tilde{\theta}) \right\}^{-1} A(\tilde{\theta}), \quad (2.10)$$

where $\tilde{\theta}$ is any $a_n$-consistent estimator of $\theta_0$.

The one-step estimating approach was first discussed by Bickel (1975) for estimating the location parameter in a linear model. Recently, Bradic et al. (2011) studied the one-step method for variable selection based on penalized composite quasi-likelihood. Following the similar arguments as in Bradic et al. (2011), we can show that for tail quantiles, the one-step slope estimator $\hat{\beta}_{\text{OS}}$ achieves the same asymptotic efficiency as the OWQAE and OWCRQ estimator as long as the initial estimator $\tilde{\beta}$ is consistent of the same rate $a_n$.

**Theorem 2.2.3** Suppose conditions $A1$-$A5$ hold, if $||\tilde{\theta} - \theta_0|| = O_p(1/a_n)$, then we have

$$a_n(\hat{\beta}_{\text{OS}} - \beta_0) \overset{d}{\to} N\left(0, \frac{1}{\phi^T(\xi) \Gamma \phi(\xi)} \left(\frac{m^{-\xi} - 1}{-\xi}\right)^{-2} D^{-1}\right).$$

Note that to achieve the above asymptotic efficiency, the initial estimator $\tilde{\theta}$ can be any $a_n$-consistent estimator such as the conventional quantile regression estimator or $\hat{\theta}_{\text{WCRQ}}$ with any given nonnegative weights. In the implementation, we use $\hat{\theta}_{\text{WCRQ+}}$ as the initial estimator.

In (2.10), $B(\theta)$ involves the unknown quantity $f_Y(z_{i,k}^T \theta | x_i) = f_Y\{Q_Y(\tau_k | x_i) | x_i\}$, which has to be estimated in practice. In general, we can estimate $f_Y\{Q_Y(\tau | x) | x\}$ by using the difference quotient method proposed by Hendricks and Koenker (1992), or
estimate the quantity
\[ \sum_{i=1}^{n} z_{i,k} z_{i,k}^T f_Y(z_{i,k}^T \theta | x_i) \]
by using the kernel method proposed by Powell (1991). Our numerical investigation suggests that these two methods work well for large sample sizes but they sometimes lead to unstable results for small samples. Note that model (2.3) and assumption \( A3 \) imply that as \( \tau \to \infty \), \( f_Y\{Q_Y(\tau|x)|x\} = f_U\{\alpha(\tau)|x\} \sim f_0\{\alpha(\tau)\} \), which is common across \( x \). Therefore, we suggest to use the nonparametric kernel density estimation based on the estimated residuals \( \hat{\epsilon}_i = y_i - x_i^T \hat{\beta}_{WCRQ} \) as in Zhao and Xiao (2014).

Different from the WCRQ estimator that requires minimizing the weighted combined quantile objective function, the calculation of WQAE only requires minimizing the convex quantile objective function in (2.2) at each quantile level \( \tau_k \) separately. Therefore, negative weights in \( \omega_{opt} \) do not cause any computational difficulty. In contrast, the WCRQ method requires solving the combined objective function of \( K + p \) parameters, and a one-step iteration is needed for calculating the optimal estimator when some of the optimal weights are negative. Despite the possible computational complication, the WCRQ method has the following advantages. First, the weighted composite quantile regression method can be used to accommodate general interquantile commonality in a more direct way, for instance, locally linear quantile slopes with \( \beta(\tau_k) = \beta(\tau_1) + (\tau_k - \tau_1)\gamma \), where \( \gamma \) is an unknown parameter. Second, penalization can be incorporated in the weighted composite quantile loss function for variable selection and inter-quantile shrinkage; see for instance Bradic et al. (2011), Jiang et al. (2013a) and Jiang et al. (2014).

### 2.2.4 Estimation of EVI

Both the optimal and sub-optimal weights involve the unknown extreme value index \( \xi \). In this paper we estimate \( \xi \) by the maximum likelihood estimator. Suppose \( z_1, \ldots, z_n \) is a random sample of \( Z \) that has distribution function \( F \). Pickands III (1975) showed at if \( F \) is in the domain of attraction with the extreme value index \( \xi \), then as \( u \to z_F \), the endpoint of \( Z \), \( P(Z \leq u + z|Z > u) \) converges to the distribution function of the generalized Pareto distribution (GPD) with parameters \( \Psi = (\xi, \sigma, u)^T \). The density function of the GPD
can be written as
\[
g(z|\Psi) = \begin{cases} 
\frac{1}{\sigma} \left( 1 + \frac{\xi (z-u)}{\sigma} \right)^{-(1+\xi)/\xi}, & \text{if } \xi \neq 0, \\
\frac{1}{\sigma} \exp\{- (z-u)/\sigma\}, & \text{if } \xi = 0,
\end{cases}
\]
where \( z > u \) for \( \xi \geq 0 \) and \( 0 \leq z - u \leq -\sigma/\xi \) for \( \xi < 0 \).

In our setup, we define \( u_i = \hat{Q}_Y(\tau_0|x_i), i = 1, \ldots, n \), where \( \tau_0 \to 1 \) as \( n \to \infty \), and \( \hat{Q}_Y(\tau_0|x) \) is any consistent estimator of \( Q_Y(\tau_0|x) \), for instance, the conventional quantile regression estimator. We can then estimate \( \xi \) by maximizing the GPD likelihood based on the exceedances \{\( y_i - u_i : y_i \geq u_i, i = 1, \ldots, n \)\}. If \( F_0(\cdot) \) satisfies the second-order condition as defined on page 44 of de Haan and Ferreira (2006) and \( \xi > -1/2 \), it then follows by Smith (1987) and Theorem 3.4.2 in de Haan and Ferreira (2006) that the maximum likelihood estimator of \( \xi \) is consistent and asymptotically normal. Throughout our numerical studies, we choose \( \tau_0 = 0.95 \) and obtain the maximum likelihood estimator of \( \xi \) by using the function “gpd.fit” in the R package ismev.

### 2.2.5 Comparison of Asymptotic Efficiency

Theorems 2.2.1 and 2.2.2 suggest that the OWCRQ and OWQAE estimators are asymptotically equivalent. The one-step estimator \( \hat{\beta}_{O/S} \) achieves the same asymptotic efficiency as OWCRQ. The WCRQ+ estimator is the same as OWCRQ when the optimal weights are all nonnegative, but the former is asymptotically less efficient when some of the optimal weights are negative.

We assess the efficiency gain of using optimal weights by comparing the asymptotic efficiency of OWCRQ/OWQAE, WCRQ+ with CRQ and QAE, the composite and quantile average estimators based on equal weights. We consider five different cases: \( \xi = 0 \) corresponding to light-tailed such as exponential and normal distributions, \( \xi = 0.5 \) and 1 corresponding to heavy-tailed distributions such as \( t_2 \) and \( t_1 \), \( \xi = -0.2 \) and -0.4 corresponding to short-tailed distributions such as Beta(2, 5) and Beta(2, 2.5). We consider five quantiles with \( \tau_k = 0.95, 0.96, 0.97, 0.98 \) and 0.99. Table 2.1 summarizes the asymptotic relative efficiency of QAE, CRQ and WCRQ+ with respect to OWQAE (or equivalently OWCRQ), and Table 2.2 presents the optimal weights for both the WCRQ and WQAE methods.

Table 2.2 suggests that for the optimal WCRQ and WQAE method, less weights
Table 2.1: The asymptotic relative efficiency of QAE, CRQ, WCRQ+ and OWCRQ with respect to OWQAE (OWCRQ) for five types of distributions.

<table>
<thead>
<tr>
<th>Tail type</th>
<th>QAE</th>
<th>CRQ</th>
<th>WCRQ+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light (ξ = 0)</td>
<td>0.64</td>
<td>0.81</td>
<td>1.00</td>
</tr>
<tr>
<td>Heavy (ξ = 0.5)</td>
<td>0.23</td>
<td>0.52</td>
<td>0.90</td>
</tr>
<tr>
<td>Heavy (ξ = 1)</td>
<td>0.06</td>
<td>0.33</td>
<td>0.77</td>
</tr>
<tr>
<td>Short (ξ = -0.2)</td>
<td>0.84</td>
<td>0.92</td>
<td>1.00</td>
</tr>
<tr>
<td>Short (ξ = -0.4)</td>
<td>0.95</td>
<td>0.95</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 2.2: Optimal weights for the WCRQ and WQAE methods at \( \tau_k = 0.95, 0.96, 0.97, 0.98 \) and 0.99.

<table>
<thead>
<tr>
<th>Tail type</th>
<th>( \omega_{opt} ) for WCRQ</th>
<th>( \omega_{opt} ) for WQAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light (ξ = 0)</td>
<td>1.00 0.00 0.00 0.00 0.00</td>
<td>1.00 0.00 0.00 0.00 0.00</td>
</tr>
<tr>
<td>Heavy (ξ = 0.5)</td>
<td>2.96 -0.34 -0.39 -0.48 -0.74</td>
<td>1.26 -0.10 -0.08 -0.05 -0.03</td>
</tr>
<tr>
<td>Heavy (ξ = 1)</td>
<td>7.41 -1.60 -1.60 -1.60 -1.60</td>
<td>1.35 -0.19 -0.11 -0.05 -0.12</td>
</tr>
<tr>
<td>Short (ξ = -0.2)</td>
<td>0.61 0.03 0.04 0.07 0.25</td>
<td>0.80 0.03 0.04 0.04 0.09</td>
</tr>
<tr>
<td>Short (ξ = -0.4)</td>
<td>0.35 0.03 0.05 0.10 0.47</td>
<td>0.53 0.05 0.06 0.09 0.28</td>
</tr>
</tbody>
</table>

(and sometimes even negative weights) are put on coefficients corresponding to the more extreme quantiles. This pattern agrees with our expectation stated in Remark 2. When the quantile level gets more extreme, the local estimator becomes more unstable, therefore assigning less weights can reduce the variance of the weighted estimator.

Table 2.1 shows that the optimal estimators have higher efficiency across different types of distributions when compared to the estimators QAE and CRQ with equal weights. The efficiency gains of OWCRQ and OWQAE also depend on the heaviness of the tail distribution, and particularly they are more pronounced for heavy-tailed distributions. One explanation is that for heavy-tailed error distributions, the data sparsity leads to larger variances to the local quantile regression estimator at more extreme tails. In contrast, the optimally weighted estimators improves the efficiency by assigning more weights to the less extreme quantiles. For cases with short-tailed and light-tailed distri-
butions, estimators based on equal weights lose some efficiency but the efficiency loss is not as substantial as for heavy-tailed distributions. Moreover, note that for distributions with light and short tails, all the optimal weights are non-negative, so OWCRQ and WCRO+ are the same. However, for heavy-tailed distributions, it is more beneficial to consider optimal weights than the sub-optimal weights.

2.3 Simulation Study

To demonstrate the finite sample performance of the proposed methods, we conduct a simulation study. We consider three examples: (i) univariate predictor with constant quantile slope; (ii) univariate predictor with constant quantile slope only at upper quantiles; (iii) multivariate predictor with constant slopes across quantiles.

Example 1. The data is generated from

\[ y_i = x_i \beta + \epsilon_i, \quad i = 1, \ldots, n, \quad \text{(2.11)} \]

where \( x_i \sim N(0, 1), \beta = 1, \) and \( \epsilon_i \) are independent and identically distributed (i.i.d.) errors. In this model, the \( \tau \)th conditional quantile of \( Y \) is \( Q_Y(\tau|x) = F_{\epsilon}^{-1}(\tau) + x \), where \( F_{\epsilon} \) is the cumulative distribution function of \( \epsilon_i \).

Example 2. This is an example where the slope is constant only at the upper quantiles with \( \tau > 0.9 \). The quantile function is defined as \( Q_Y(\tau|x) = \alpha(\tau) + \beta(\tau)x \), where \( \alpha(\tau) = F_{\epsilon}^{-1}(\tau) \) and

\[ \beta(\tau) = \begin{cases} \beta - F_{\epsilon}^{-1}(0.90) + F_{\epsilon}^{-1}(\tau) & \text{if } 0 < \tau < 0.90 \\ \beta & \text{if } 0.90 \leq \tau < 1 \end{cases} \quad \text{(2.12)} \]

with \( \beta = 1 \). To generate the data, we first generate \( x_i \sim U(0, 1) \) and quantile levels \( u_i \sim U(0, 1) \), and then let \( y_i = \alpha(u_i) + \beta(u_i)x_i, \quad i = 1, \ldots, n \). Therefore, \( \beta(\tau) \) varies for \( \tau < 0.9 \), but it remains constant for \( \tau \geq 0.9 \).

Example 3. In this example, we generate data from the following model with two predictors

\[ y_i = x_{i,1} \beta_1 + x_{i,2} \beta_2 + \epsilon_i, \quad i = 1, \ldots, n, \quad \text{(2.13)} \]

where \( x_{i,1} \sim N(0, 1), \quad x_{i,2} \sim N(0, 1), \quad \beta_1 = 1 \quad \text{and} \quad \beta_2 = 2, \) and \( \epsilon_i \) are i.i.d. errors with distribution \( F_{\epsilon} \). Therefore, the \( \tau \)th conditional quantile of \( Y \) given \( x_1 \) and \( x_2 \) is
$Q_Y(\tau|x_1, x_2) = F^{-1}_\epsilon(\tau) + x_1 + 2x_2$, and the slopes are constant across $\tau \in (0, 1)$.

For all three examples, we consider the following error distributions for $\epsilon$: the standard normal $N(0, 1)$, $t_2$ and Beta$(2, 5)$. In each example, we consider two sample sizes: $n=500$ and 1000. Since the main interest of this study is on extreme quantiles, we choose the equally spaced extreme upper quantiles as $\tau_k = 1 - (6 - k) \times n^{-3/4}$, where $k = 1, \ldots, 5$. The simulation is repeated 500 times for each scenario.

The following five estimators of $\beta$ are included for comparison: the quantile average estimator with equal weights (QAE), the composite estimator with equal weights (CRQ), the optimally weighted quantile average estimator (OWQAE), the one-step composite estimator (referred to as OWCRQ as the two have the same asymptotic efficiency), and the weighted composite estimator based on the sub-optimal nonnegative weights (WCRQ+). For OWQAE, OWCRQ and WCRQ+ methods, the extreme value index $\xi$ is estimated by the maximum likelihood estimator introduced in Section 2.3.

Tables 2.3 and 2.4 summarize the mean squared error (MSE) of different estimators of $\beta$ in Examples 1 and 2, respectively. For Example 3, we report the mean integrated squared error (MISE) defined as

$$
\text{MISE} = \frac{1}{500} \sum_{j=1}^{500} \{(\hat{\beta}_{j1} - \beta_1)^2 + (\hat{\beta}_{j2} - \beta_2)^2\},
$$

where $\hat{\beta}_{j1}$ and $\hat{\beta}_{j2}$ are the estimators of $\beta_1$ and $\beta_2$ from the $j$th simulation. Tables 2.3-2.5 suggest that, for all three types of distributions, OWCRQ and OWQAE have similar performance, which agrees with the asymptotic theory. For the heavy-tailed $t_2$ distribution, OWQAE and OWCRQ are more efficient than WCRQ+, which loses some efficiency due to the nonnegative weights constraint employed. In addition, for the $t_2$ distribution, the three methods involving weight estimation perform significantly better than the two methods QAE and CRQ based on equal weights. For normal errors, optimally weighted methods still have some efficiency gain over the equally weighted estimators but the gain is less obvious as that for the $t_2$ distribution. For the short-tailed Beta$(2, 5)$ distribution, the optimally weighted methods have no obvious efficiency gain, and this agrees with our asymptotic efficiency comparison in Section 2.5.
Table 2.3: The $10^3 \times $MSE of different estimators of $\beta$ in Example 1. The values in the parentheses are the standard errors of $10^3 \times $MSE.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>n</th>
<th>QAE</th>
<th>OWQAE</th>
<th>CRQ</th>
<th>WCRQ+</th>
<th>OWCRQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>500</td>
<td>11.54</td>
<td>10.29</td>
<td>10.52</td>
<td>9.96</td>
<td>9.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.76)</td>
<td>(0.72)</td>
<td>(0.74)</td>
<td>(0.66)</td>
<td>(0.66)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>8.82</td>
<td>7.16</td>
<td>7.78</td>
<td>7.19</td>
<td>7.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.61)</td>
<td>(0.51)</td>
<td>(0.54)</td>
<td>(0.52)</td>
<td>(0.52)</td>
</tr>
<tr>
<td>$t_2$</td>
<td>500</td>
<td>349.89</td>
<td>101.01</td>
<td>173.39</td>
<td>104.41</td>
<td>100.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(20.08)</td>
<td>(6.43)</td>
<td>(11.01)</td>
<td>(6.71)</td>
<td>(6.40)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>628.92</td>
<td>148.91</td>
<td>302.17</td>
<td>167.87</td>
<td>159.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(39.72)</td>
<td>(8.78)</td>
<td>(19.09)</td>
<td>(10.22)</td>
<td>(9.69)</td>
</tr>
<tr>
<td>Beta(2, 5)</td>
<td>500</td>
<td>0.39</td>
<td>0.39</td>
<td>0.37</td>
<td>0.39</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.03)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.26</td>
<td>0.25</td>
<td>0.24</td>
<td>0.24</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
</tbody>
</table>

Table 2.4: The $10^3 \times $MSE of different estimators of $\beta$ in Example 2. The values in the parentheses are the standard errors of $10^3 \times $MSE.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>n</th>
<th>QAE</th>
<th>OWQAE</th>
<th>CRQ</th>
<th>WCRQ+</th>
<th>OWCRQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>500</td>
<td>129.68</td>
<td>124.94</td>
<td>114.67</td>
<td>119.63</td>
<td>119.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8.48)</td>
<td>(8.22)</td>
<td>(7.54)</td>
<td>(7.96)</td>
<td>(7.90)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>99.75</td>
<td>88.31</td>
<td>89.78</td>
<td>85.21</td>
<td>85.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.94)</td>
<td>(5.60)</td>
<td>(5.28)</td>
<td>(5.29)</td>
<td>(5.25)</td>
</tr>
<tr>
<td>$t_2$</td>
<td>500</td>
<td>5493.83</td>
<td>1459.47</td>
<td>2452.45</td>
<td>1498.51</td>
<td>1467.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(406.81)</td>
<td>(95.32)</td>
<td>(161.92)</td>
<td>(99.53)</td>
<td>(97.95)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>8044.79</td>
<td>2116.36</td>
<td>3807.19</td>
<td>2241.42</td>
<td>2182.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(518.23)</td>
<td>(144.46)</td>
<td>(223.98)</td>
<td>(145.35)</td>
<td>(143.02)</td>
</tr>
<tr>
<td>Beta(2, 5)</td>
<td>500</td>
<td>4.25</td>
<td>4.54</td>
<td>3.95</td>
<td>4.54</td>
<td>4.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.28)</td>
<td>(0.29)</td>
<td>(0.26)</td>
<td>(0.30)</td>
<td>(0.30)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>2.97</td>
<td>2.88</td>
<td>2.86</td>
<td>2.79</td>
<td>2.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.18)</td>
<td>(0.18)</td>
<td>(0.17)</td>
<td>(0.17)</td>
<td>(0.16)</td>
</tr>
</tbody>
</table>
Table 2.5: The $10^3 \times \text{MISE}$ of different estimators in Example 3. The values in the parentheses are the standard errors of $10^3 \times \text{MISE}$.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$n$</th>
<th>QAE</th>
<th>OWQAE</th>
<th>CRQ</th>
<th>WCRQ+</th>
<th>OWCRQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>500</td>
<td>22.68</td>
<td>21.15</td>
<td>20.66</td>
<td>20.66</td>
<td>20.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.10)</td>
<td>(1.03)</td>
<td>(0.98)</td>
<td>(1.00)</td>
<td>(0.99)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>17.40</td>
<td>13.63</td>
<td>15.16</td>
<td>13.46</td>
<td>13.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.75)</td>
<td>(0.63)</td>
<td>(0.67)</td>
<td>(0.63)</td>
<td>(0.63)</td>
</tr>
<tr>
<td>$t_2$</td>
<td>500</td>
<td>780.16</td>
<td>231.48</td>
<td>428.49</td>
<td>249.95</td>
<td>240.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(38.45)</td>
<td>(11.24)</td>
<td>(22.78)</td>
<td>(12.00)</td>
<td>(11.58)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1007.67</td>
<td>290.43</td>
<td>537.08</td>
<td>312.2</td>
<td>298.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(42.22)</td>
<td>(12.88)</td>
<td>(24.15)</td>
<td>(13.55)</td>
<td>(12.97)</td>
</tr>
<tr>
<td>Beta(2, 5)</td>
<td>500</td>
<td>0.76</td>
<td>0.77</td>
<td>0.70</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.52</td>
<td>0.45</td>
<td>0.48</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
</tbody>
</table>

2.4 Application to Chicago Precipitation Data

An important topic in climate studies is quantifying extremal phenomena such as heavy precipitation or high temperature, for which quantile regression serves as a promising tool. For illustration, we apply the proposed methods to the statistical downscaling of daily precipitations in the Aurora station of Chicago, IL. The response $Y$ is the observed daily precipitation (inch) at the station from 1957 to 2002, and the covariate $X$ is the simulated daily precipitation from the ERA-40 reanalysis model introduced in Uppala et al. (2005).

Since we are interested in estimating the extremely heavy precipitation conditioning on $X$, we only include the wet days data. In this data set 30% of the days are wet with $y_i > 0$. Since in climate studies it is commonly assumed that the percentage of wet days in the future is the same as in the past and prediction is based on the simulated daily precipitation, we define the wet days data as the pairs of $(y_i, x_i)$ with $x_i$ exceeding its 70th sample percentile. This yields a data set of 4816 observations.

As a preliminary analysis, we apply conventional quantile regression at high quantiles $\tau > 0.99$. The slope estimates appear to be stable from $\tau = 0.990$ to $\tau = 0.995$; see Figure 4.1. We apply the Wald-type test (Koenker, 2005, page 76) for testing the null
Figure 2.1: Local quantile regression estimates of the slope at upper quantiles for the precipitation data in Aurora station.

hypothesis $H_0 : \beta(0.990) = \beta(0.991) = \ldots = \beta(0.995)$ and obtain a $p$-value=0.794. However, the test for $H_0 : \beta(0.990) = \ldots = \beta(0.996) = \beta(0.997) = \beta(0.998)$ yields a $p$-value of $4^{-10}$, suggesting a possible violation of the common slope assumption for $\tau \geq 0.996$. Therefore, we choose five quantile levels 0.990, 0.991, \ldots, 0.995 to apply the proposed combined estimation methods.

Table 2.6 summarizes the estimated slopes from different methods. The values in the parentheses are the bootstrap standard errors based on 1000 bootstrap samples obtained by sampling the paired observations $(y_i, x_i)$ with replacement. It is noticed that $\hat{\beta}_{QAE}$ and $\hat{\beta}_{CRQ}$ are larger than the other estimates, which is partially due to the fact that these two methods enforce equal weights at the six quantiles including the higher quantile $\tau = 0.994$, at which the quantile slope estimation $\hat{\beta}(0.994)$ is the largest among the six slope estimations. In addition, the estimators $\hat{\beta}_{QAE}$ and $\hat{\beta}_{CRQ}$ have larger variances, which lead to insignificance in the slope. In contrast, the OWQAE, OWCRQ and WCRQ+ estimators have smaller variances, and all three showed that at the high quantiles, the simulated daily precipitation has a significant positive effect on the observed precipitation and thus can serve as a good predictor for high precipitation.

We conduct a cross-validation study to evaluate the prediction accuracy of each method. We randomly choose about half of the data set as the training data (total 2409 observations) and the other half as the validation data (total 2407 observations). For each method, we use training data to obtain the estimated coefficients $\hat{\alpha}(\tau)$ and $\hat{\beta}(\tau)$ and predict the $\tau$th conditional quantile of $Y$ as $\hat{Q}_Y(\tau|x_i) = \hat{\alpha}(\tau) + x_i\hat{\beta}(\tau)$. The prediction
Table 2.6: The upper part of the table summarizes the estimated common slope $\hat{\beta}$ for the proposed methods and the bootstrap standard error (s.e.). The lower part of the table presents the quantile-specific weights given by different methods based on the estimated extreme value index $\xi = 0.29$.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>QAE</th>
<th>CRQ</th>
<th>OWQAE</th>
<th>WCRQ+</th>
<th>OWCRQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}$</td>
<td>115.07</td>
<td>132.63</td>
<td>91.21</td>
<td>96.25</td>
<td>96.78</td>
</tr>
<tr>
<td>Bootstrap s.e.</td>
<td>(103.08)</td>
<td>(117.21)</td>
<td>(30.21)</td>
<td>(40.93)</td>
<td>(37.65)</td>
</tr>
<tr>
<td>$\tau_1 = 0.990$</td>
<td>1/6</td>
<td>1/6</td>
<td>1.22</td>
<td>1</td>
<td>1.55</td>
</tr>
<tr>
<td>$\tau_2 = 0.991$</td>
<td>1/6</td>
<td>1/6</td>
<td>-0.03</td>
<td>0</td>
<td>-0.05</td>
</tr>
<tr>
<td>$\tau_3 = 0.992$</td>
<td>1/6</td>
<td>1/6</td>
<td>-0.03</td>
<td>0</td>
<td>-0.05</td>
</tr>
<tr>
<td>$\tau_4 = 0.993$</td>
<td>1/6</td>
<td>1/6</td>
<td>-0.03</td>
<td>0</td>
<td>-0.06</td>
</tr>
<tr>
<td>$\tau_5 = 0.994$</td>
<td>1/6</td>
<td>1/6</td>
<td>-0.03</td>
<td>0</td>
<td>-0.06</td>
</tr>
<tr>
<td>$\tau_6 = 0.995$</td>
<td>1/6</td>
<td>1/6</td>
<td>-0.10</td>
<td>0</td>
<td>-0.32</td>
</tr>
</tbody>
</table>

The prediction error (PE) is defined as:

$$PE = \sum_{i=1}^{2407} \rho_{\tau} \{ y_i - \hat{Q}_Y(\tau|x_i) \}, \quad \tau \in \{0.990, \ldots, 0.995\},$$

where $\{(y_i, x_i), i = 1, \ldots, 2407\}$ are in the validation set. The cross validation is repeated 500 times and the mean of PE is reported at quantile levels 0.990, 0.992 and 0.995 in Table 2.7. Results show that the proposed optimally weighted methods perform generally better than the equally-weighted methods QAE and CRQ and the local quantile regression method (RQ). Moreover, the gain in the prediction accuracy of the optimally weighted methods increases as the quantile level gets more extreme.

2.5 Proof

We first introduce some notations. Denote

$$a_n = \frac{\sqrt{(1 - \tau)n}}{F_0^{-1}(\tau) - F_0^{-1}(\tilde{\tau}_m)} \quad \text{and} \quad a_k = \frac{\sqrt{l_k(1 - \tau)n}}{F_0^{-1}(\tau_k) - F_0^{-1}(\tilde{\tau}_{mk})},$$

where $m > 1, \tilde{\tau}_m = 1 - m(1 - \tau)$ and $\tilde{\tau}_{mk} = 1 - m(1 - \tau_k)$. For notational simplicity, we denote $F_i = F_Y(\cdot|x_i)$ and $f_i = f_Y(\cdot|x_i)$. 
Table 2.7: The prediction error of different methods at quantile levels 0.990, 0.992 and 0.995. The values in the parentheses are the standard errors.

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>RQ</th>
<th>QAE</th>
<th>CRQ</th>
<th>OWQAE</th>
<th>WCRQ+</th>
<th>OWCRQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.990</td>
<td>65.27</td>
<td>66.75</td>
<td>68.74</td>
<td>65.27</td>
<td>65.27</td>
<td>65.25</td>
</tr>
<tr>
<td></td>
<td>(0.34)</td>
<td>(0.30)</td>
<td>(0.29)</td>
<td>(0.33)</td>
<td>(0.34)</td>
<td>(0.33)</td>
</tr>
<tr>
<td>0.992</td>
<td>57.10</td>
<td>57.92</td>
<td>60.29</td>
<td>56.77</td>
<td>56.79</td>
<td>56.76</td>
</tr>
<tr>
<td></td>
<td>(0.32)</td>
<td>(0.30)</td>
<td>(0.29)</td>
<td>(0.33)</td>
<td>(0.33)</td>
<td>(0.33)</td>
</tr>
<tr>
<td>0.995</td>
<td>45.01</td>
<td>43.76</td>
<td>46.47</td>
<td>42.31</td>
<td>42.38</td>
<td>42.37</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td>(0.26)</td>
<td>(0.3)</td>
<td>(0.32)</td>
<td>(0.32)</td>
<td>(0.32)</td>
</tr>
</tbody>
</table>

Proof of Theorem 2.2.1. At the \( \tau_k \)th quantile, the local quantile estimator of the coefficients in model (2.1) is defined as

\[
(\hat{\alpha}_k, \hat{\beta}_k) = \arg\min_{(\alpha, \beta)} \sum_{i=1}^{n} \rho_{\tau_k}(y_i - \alpha - x_i^T \beta).
\]

Denote \( \hat{t}_{n,k} = a_k(\hat{\alpha}_k - \alpha_{0,k}) \) and \( \hat{z}_{n,k} = a_k(\hat{\beta}_k - \beta_0), k = 1, \ldots, K \). By Theorem 5.1 of Chernozhukov (2005), we have

\[
\left( \hat{t}_{n,1}, \hat{z}_{n,1}, \ldots, \hat{t}_{n,K}, \hat{z}_{n,K}^T \right)^T \overset{d}{\rightarrow} N \left( 0, \left( \frac{m-\xi-1}{\xi} \right)^{-2} \tilde{\Gamma} \otimes \left( \begin{array}{cc} 1 & 0 \\ 0 & D \end{array} \right)^{-1} \right),
\]

where \( \tilde{\Gamma} \) is a \( K \times K \) matrix with the \( (k, k') \)th element as \( \min(l_k, l_{k'})/\sqrt{l_k l_{k'}} \). Therefore, we have

\[
a_n(\hat{\beta}_{WQAE} - \beta_0) = \sum_{k=1}^{K} \frac{\varpi_k a_k}{a_k} \hat{z}_{n,k} \overset{d}{\rightarrow} N \left( 0, \varpi^T \Phi^{-1}(\xi) \Gamma \Phi^{-1}(\xi) \varpi \left( \frac{m-\xi-1}{\xi} \right)^{-2} D^{-1} \right).
\]

Lemma 2.5.1 For a sequence of quantiles \( \tau_1, \ldots, \tau_K \) with \( \tau_k \rightarrow 1 \) and \( (1 - \tau_k)n \rightarrow \infty \),

\[
\frac{\min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'}}{1 - \tau} \sim \min(l_k, l_{k'}),
\]

where \( \tau \rightarrow 1 \), \( (1 - \tau)n \rightarrow \infty \), and \( (1 - \tau_k)/(1 - \tau) \rightarrow l_k \) for \( k = 1, \ldots, K \).
Proof. Let $\tau^* = 1 - \tau$, $\tau^*_k = 1 - \tau_k$. Therefore, $\tau^*_k/\tau^* \to l_k$, $k = 1, \ldots, K$, and $\tau^* \to 0$. It’s easy to show that $\min(\tau_k, \tau_k') - \tau_k \tau_k' = \min(\tau^*_k, \tau^*_k') - \tau^*_k \tau^*_k$. For any $k, k' = 1, \ldots, K$,

$$
\frac{\min(\tau_k, \tau_k') - \tau_k \tau_k'}{1 - \tau} = \frac{\min(\tau^*_k, \tau^*_k') - \tau^*_k \tau^*_k}{\tau^*} = \frac{\tau^*[\min(l_k, l_k') + o(1) - \tau^*\{l_k + o(1)\}\{l_k' + o(1)\}]}{\tau^*} = \min\{l_k, l_k' + o(1)\} - \tau^*\{l_k + o(1)\}\{l_k' + o(1)\}
$$

$\sim \min(l_k, l_k')$.

Lemma 2.5.2 Under conditions A3-A5, $a_k/a_n \to l_k^{\xi+1/2}$ for any $k = 1, \ldots, K$.

Proof. Since $\partial F_0^{-1}(\tau)/\partial \tau = 1/f_0\{F_0^{-1}(\tau)\}$, A4 means that for any $x > 0$

$$
\frac{f_0\{F_0^{-1}(1 - \tau^*)\}}{f_0\{F_0^{-1}(1 - x\tau^*)\}} \sim x^{-\xi-1}, \text{ as } \tau^* \to 0. \tag{2.15}
$$

For any $\delta > 0$, note that $dF_0^{-1}(1 - s\delta)/ds = -\delta \left[ f_0\{F_0^{-1}(1 - s\delta)\} \right]^{-1}$. Therefore,

$$
\int_1^m \frac{1}{f_0\{F_0^{-1}(1 - s\delta)\}} ds = \frac{F_0^{-1}(1 - \delta) - F_0^{-1}(1 - m\delta)}{\delta}. \tag{2.16}
$$

Combining (2.15) and (2.16) gives

$$
\frac{F_0^{-1}(\tau_k) - F_0^{-1}(\tau_{mk})}{l_k(1 - \tau)/f_0\{F_0^{-1}(\tau_k)\}} \sim f_0\{F_0^{-1}(\tau_k)\} \left[ F_0^{-1}\{1 - (1 - \tau_k)\} - F_0^{-1}\{1 - m(1 - \tau_k)\} \right] \frac{1}{1 - \tau_k}
$$

$$
= f_0\{F_0^{-1}(\tau_k)\} \int_1^m \frac{1}{f_0\{F_0^{-1}\{1 - s(1 - \tau_k)\}\}} ds
$$

$$
= \int_1^m f_0\{F_0^{-1}\{1 - (1 - \tau_k)\}\} ds
$$

$$
= \int_1^m f_0\{F_0^{-1}\{1 - s(1 - \tau_k)\}\} ds
$$

$$
\sim \int_1^m s^{-\xi-1} ds = \frac{m^{-\xi} - 1}{-\xi} \text{ (lnm if } \xi = 0). \tag{2.17}
$$
Therefore, applying (2.15) and (2.17), we have

\[
\frac{a_k}{a_n} = \frac{\sqrt{n l_k (1 - \tau)}}{\sqrt{n (1 - \tau)}} \left\{ F_0^{-1}(\tau) - F_0^{-1}(\tilde{\tau}_m) \right\} \\
= \frac{\sqrt{l_k}}{(1 - \tau)} \frac{F_0^{-1}(\tau) - F_0^{-1}(\tilde{\tau}_m)}{\{ F_0^{-1}(\tau) \}} \left[ \frac{(1 - \tau)}{f_0(F_0^{-1}(\tau))} \right] \left[ l_k (1 - \tau) / f_0(F_0^{-1}(\tau_k)) \right] \\
\sim \sqrt{l_k} \left( \frac{m^{-\xi} - 1}{-\xi} \right) \left( \frac{1}{l_k^{\xi+1}} \right) \left( \frac{-\xi}{m^{-\xi} - 1} \right) = l_k^{\xi+\frac{1}{2}}.
\]

**Proof of Theorem 2.2.2.** For notational simplicity, we write \( \hat{\theta}_{WCRQ} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_K, \hat{\beta}^T)^T \) in this proof. Let \( \hat{u}_{n,k} = a_k(\hat{\alpha}_k - \alpha_{0,k}) \), \( k = 1, \ldots, K \), and \( \hat{u}_n = a_n(\hat{\beta} - \beta_0) \), where \((\alpha_0, \ldots, \alpha_{0,K}, \beta_0)\) are the true parameters. From (2.7), it is clear that \((\hat{u}_{n,1}, \ldots, \hat{u}_{n,K}, \hat{u}_n)\) is the minimizer of

\[
L_n = \frac{a_n}{\sqrt{n(1 - \tau)}} \sum_{k=1}^{K} \omega_k \sum_{i=1}^{n} \left\{ \rho_{\tau_k} \left( y_i - x_i^T \beta_0 - \alpha_{0,k} - \frac{x_k}{a_k} - \frac{x_i^T u}{a_n} \right) - \rho_{\tau_k} (y_i - x_i^T \beta_0 - \alpha_{0,k}) \right\}
\]

with respect to \((u_1, \ldots, u_K, u)\). Using Knight’s identity (Knight, 1998),

\[
\rho_{\tau}(u - v) - \rho_{\tau}(u) = -v\{\tau - I(u < 0)\} + \int_{0}^{v} \{I(u \leq s) - I(u < 0)\} ds,
\]

we can rewrite \( L_n \) as \( L_n = L_{n,1} + L_{n,2} \), where

\[
L_{n,1} = \frac{a_n}{\sqrt{n(1 - \tau)}} \sum_{k=1}^{K} \omega_k \sum_{i=1}^{n} \left( \frac{u_k}{a_k} + \frac{x_i^T u}{a_n} \right) \{I(\epsilon_{i,k} < 0) - \tau_k\},
\]

\[
L_{n,2} = \frac{a_n}{\sqrt{n(1 - \tau)}} \sum_{k=1}^{K} \omega_k \sum_{i=1}^{n} \int_{0}^{u_k + \frac{x_i^T u}{a_n}} \{I(\epsilon_{i,k} \leq s) - I(\epsilon_{i,k} < 0)\} ds,
\]

and \( \epsilon_{i,k} = y_i - x_i^T \beta_0 - \alpha_{0,k} \). Denoting \( \psi_{i,k} = I(\epsilon_{i,k} < 0) - \tau_k \), we have

\[
L_{n,1} = \frac{a_n}{\sqrt{n(1 - \tau)}} \sum_{k=1}^{K} \omega_k \sum_{i=1}^{n} \psi_{i,k} u_k + \frac{a_n}{\sqrt{n(1 - \tau)}} \sum_{k=1}^{K} \omega_k \sum_{i=1}^{n} \psi_{i,k} x_i^T u
\]

\[
= \sum_{k=1}^{K} W_{n,k} u_k + W_n^T u \triangleq \tilde{W}_n^T \tilde{U}, \tag{2.18}
\]

27
where

\[
W_{n,k} = \frac{\omega_k}{\sqrt{n(1-\tau)}} \frac{a_n}{a_k} \sum_{i=1}^{n} \psi_{i,k}, \quad W_n = \frac{1}{\sqrt{n(1-\tau)}} \sum_{k=1}^{K} \sum_{i=1}^{n} \omega_k \psi_{i,k} x_i,
\]

\[
\tilde{W}_n = (W_{n,1}, \ldots, W_{n,K}, W_n^T)^T, \quad \text{and} \quad \tilde{U} = (u_{n,1}, \ldots, u_{n,K}, u_n^T)^T.
\]

We next derive the limiting distribution of \(\tilde{W}_n\). Denote

\[
T_i = \left( \frac{\omega_1}{\sqrt{(1-\tau)}} \frac{a_n}{a_1} \psi_{i,1}, \ldots, \frac{\omega_K}{\sqrt{(1-\tau)}} \frac{a_n}{a_K} \psi_{i,K}, S_i^T \right)^T,
\]

where \(S_i = (1-\tau)^{-1/2} \sum_{k=1}^{K} \omega_k \psi_{i,k} x_i\). Note that \(T_i\) are i.i.d. with mean 0 and covariance matrix \(V_n\). For \(k, k' = 1, \ldots, K\), the \((k, k')\)th element of \(V_n\) is,

\[
V_n(k, k') = \text{cov} \left( \frac{\omega_k}{\sqrt{(1-\tau)}} \frac{a_n}{a_k} \psi_{i,k}, \frac{\omega_{k'}}{\sqrt{(1-\tau)}} \frac{a_n}{a_{k'}} \psi_{i,k'} \right)
= \frac{\omega_k \omega_{k'}}{(1-\tau)} \frac{a_n^2}{a_k a_{k'}} \{ \min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'} \}
\rightarrow \omega_k \omega_{k'} l_k^{\frac{\epsilon-\frac{1}{2}}{l_{k'}^{\frac{\epsilon-\frac{1}{2}}{}}}} \min(l_k, l_{k'}),
\]

(2.19)

where Lemmas 2.5.1 and 2.5.2 are used to prove the last step. Under condition \(\text{A2}\),

\[
Var(S_i) = E \{ Var(S_i | x_i) \} + Var \{ E(S_i | x_i) \}
= E \left\{ x_i x_i^T \sum_{k=1}^{K} \sum_{k'=1}^{K} \omega_k \omega_{k'} \frac{\min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'}}{1-\tau} \right\} + 0
\]

\[
= D \sum_{k=1}^{K} \sum_{k'=1}^{K} \omega_k \omega_{k'} \frac{\min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'}}{1-\tau} \rightarrow D \omega^T \Gamma \omega.
\]

In addition, for any \(k = 1, \ldots, K\),

\[
\text{cov} \left( \frac{\omega_k}{\sqrt{(1-\tau)}} \frac{a_n}{a_k} \psi_{i,k}, S_i \right) = \frac{\omega_k}{(1-\tau)} \frac{a_n}{a_k} E \left\{ \psi_{i,k} \left( \sum_{j=1}^{K} \omega_j \psi_{i,j} x_i \right) \right\}
= \frac{\omega_k}{(1-\tau)} \frac{a_n}{a_k} E \left\{ E \left( \psi_{i,k} \sum_{j=1}^{K} \omega_j \psi_{i,j} x_i \mid x_i \right) \right\} = 0,
\]

(2.20)
where the last step is due to the assumption that $E(X) = 0$.

Combining (2.19)-(2.20) gives the limit of $V_n$

$$V_n \to V = \begin{pmatrix} V_1 & 0 \\ 0 & D\omega^T\Gamma\omega \end{pmatrix},$$

where $V_1$ is a $K \times K$ matrix with the $(k,k')$th element $\omega_k\omega_{k'}l_k^{-\frac{\xi}{2}}l_{k'}^{-\frac{\xi}{2}}\min(l_kl_{k'})$, and $k,k' = 1,\ldots,K$. Applying the multivariate Central Limit Theorem and Slutsky theorem to $T_i$, we can show that

$$\tilde{W}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} T_i \overset{d}{\to} N(0, V).$$

Now we consider the second part of the objective function $L_n, L_{n,2}$. By the definitions of $a_n$ and $a_k$, we have

$$L_{n,2} = \sum_{k=1}^{K} \frac{F_{0}^{-1}(\tau_k) - F_{0}^{-1}(\tilde{\tau}_m)}{F_{0}^{-1}(\tau) - F_{0}^{-1}(\tilde{\tau}_m)} G_{k}^n,$$

where

$$G_{k}^n = \frac{a_k}{\sqrt{l_k(1-\tau)n}} \sum_{i=1}^{n} \int_{0}^{u_k + \frac{x^T_i u}{a_n}} \{I(\epsilon_{i,k} \leq s) - I(\epsilon_{i,k} \leq 0)\} ds$$

$$= \sum_{i=1}^{n} \int_{0}^{u_k + \frac{a_k x^T_i u}{a_n}} I(\epsilon_{i,k} \leq \frac{s}{a_k}) - I(\epsilon_{i,k} \leq 0) \frac{1}{\sqrt{l_k(1-\tau)n}} ds.$$
Furthermore, we have

$$E(G_n^k) = nE \left[ \int_0^{u_k + \frac{a_k}{a_n}x_i^T u} \frac{F_i \left\{ F_i^{-1}(\tau_k) + s/a_k \right\} - F_i \left\{ F_i^{-1}(\tau_k) \right\}}{\sqrt{k(1 - \tau)n}} \, ds \right] \quad \text{(iterated expectations)}$$

$$= nE \left[ \int_0^{u_k + \frac{a_k}{a_n}x_i^T u} \frac{F_i \left\{ F_i^{-1}(\tau_k) \right\} s}{a_k \sqrt{k(1 - \tau)n}} \, ds \right] \quad \text{(i)}$$

$$= nE \left[ \int_0^{u_k + \frac{a_k}{a_n}x_i^T u} \frac{F_i \left\{ F_i^{-1}(\tau_k) \right\} s}{a_k \sqrt{k(1 - \tau)n}} \, ds \right] \quad \text{(ii)}$$

$$= nE \left[ \frac{1}{2} \left( u_k + \frac{a_k}{a_n}x_i^T u \right)^2 \frac{F_i \left\{ F_i^{-1}(\tau_k) \right\}}{a_k \sqrt{k(1 - \tau)n}} \right]$$

$$= E \left[ \frac{1}{2} \left( u_k + \frac{a_k}{a_n}x_i^T u \right)^2 \frac{F_i \left\{ F_i^{-1}(\tau_k) \right\}}{a_k \sqrt{k(1 - \tau)n}} \right] \quad \text{(iii)}$$

$$\sim E \left[ \frac{1}{2} \left( u_k + \frac{a_k}{a_n}x_i^T u \right)^2 K(x_i)_{-\xi} \right] \quad \text{(2.23)}$$

By Taylor expansion and the fact that \((1 - \tau)n \to \infty\),

$$s/a_k = s\{F_i^{-1}(\tau_k) - F_i^{-1}(\tau_{mk})\}/\sqrt{k(1 - \tau)n} = o\{F_i^{-1}(\tau_k) - F_i^{-1}(\tau_{mk})\},$$

then equation (i) in (2.23) is proven. The equation (ii) holds because

$$f_i \left\{ F_i^{-1}(\tau_k) + o \left( F_i^{-1}(\tau_k) - F_i^{-1}(\tau_{mk}) \right) \right\} \sim f_i \left\{ F_i^{-1}(\tau_k) \right\}$$

as \(\tau_k \to 1\), which is derived following the same arguments as in the proof of Lemma 9.6 in Chernozhukov (2005). The equation (iii) is proven as follows. By condition \textbf{A3}, as \(\tau \to 1\),

$$f_i \left\{ F_i^{-1}(\tau) \right\} = f_U(\alpha(\tau) | x_i) \sim f_0 \left\{ F_0^{-1}(\tau) \right\}.$$

Therefore,

$$\frac{F_0^{-1}(\tau_k) - F_0^{-1}(\tau_{mk})}{l_k(1 - \tau)/f_i \left\{ F_i^{-1}(\tau_k) \right\}} \sim \frac{F_0^{-1}(\tau_k) - F_0^{-1}(\tau_{mk})}{l_k(1 - \tau)/f_0 \left\{ F_0^{-1}(\tau_k) \right\}}. \quad \text{(2.24)}$$

Combining (2.24) and (2.17), we have

$$\frac{F_0^{-1}(\tau_k) - F_0^{-1}(\tau_{mk})}{l_k(1 - \tau)/f_i \left\{ F_i^{-1}(\tau_k) \right\}} \sim \frac{m^{-\xi} - 1}{-\xi}, \quad \text{(2.25)}$$
which together with Lemma 2.5.2 proves equation (iii). Furthermore, we can show that $\text{Var}(G_k^n) \to 0$ by following the same arguments as in the proof of Lemma 9.6 in Chernozhukov (2005). In Lemma 2.5.2 we showed that $\{F_0^{-1}(\tau) - F_0^{-1}(\tilde{\tau}_m)\}/\{F_0^{-1}(\tau_k) - F_0^{-1}(\tilde{\tau}_{mk})\} \sim l_k^\xi$. Therefore, we have

$$L_{n,2} \overset{p}{\to} E \left\{ \sum_{k=1}^{K} \omega_k l_k^{-\xi} \frac{1}{2} (u_k + l_k^{\xi+\frac{1}{2}} x_i^T u)^2 \left( \frac{m^{-\xi} - 1}{-\xi} \right) \right\} = \left( \frac{m^{-\xi} - 1}{-\xi} \right) \sum_{k=1}^{K} \omega_k \left( \frac{1}{2} u_k^2 l_k^{-\xi} + \frac{1}{2} l_k^{\xi+1} u^T D u \right).$$

Combining (2.18), (2.22) and (2.26), we get

$$L_n \overset{d}{\to} L_{\infty} \equiv \sum_{k=1}^{K} W_k u_k + W^T u + \left( \frac{m^{-\xi} - 1}{-\xi} \right) \sum_{k=1}^{K} \omega_k \left( \frac{1}{2} u_k^2 l_k^{-\xi} + \frac{1}{2} l_k^{\xi+1} u^T D u \right),$$

where $\tilde{W} = (W_1, \ldots, W_K, W^T)^T$ is a random vector following the distribution $N(0, V)$ with $V$ defined in (2.21). Since the objective function $L_{\infty}$ is quadratic in $\tilde{U}$, the minimizer of $L_{\infty}$ is

$$u_{k,\infty} = \left\{ \left( \frac{m^{-\xi} - 1}{\xi} \right) \omega_k l_k^{-\xi} \right\}^{-1} W_k, \text{ for } k = 1, \ldots, K,$n

$$u_{\infty} = \left( \frac{m^{-\xi} - 1}{-\xi} \right)^{-1} \{\phi^T(\xi)\omega\}^{-1} D^{-1} W,$$

where $\phi(\xi) = (l_1^{\xi+1}, \ldots, l_K^{\xi+1})^T$. By the definition of $W$, we have

$$u_{\infty} \sim N \left( 0, \frac{\omega^T \Gamma \omega}{\{\phi^T(\xi)\omega\}^2} \left( \frac{m^{-\xi} - 1}{-\xi} \right)^{-2} \right).$$

Note that $\omega_k \geq 0$, $k = 1, \ldots, K$, then the application of the convexity lemma in Pollard (1991) gives

$$a_n(\hat{\beta}_{WCRQ} - \beta_0) = \tilde{u}_n \overset{d}{\to} u_{\infty}.$$

The proof of the statements in Remark 2 relies on the following Lemma 2.5.3.

**Lemma 2.5.3** Let $f(x) = (a^{\xi+1} - x^{\xi+1})/(a - x)$ for $a > 0$, $x > 0$ and $x \neq a$, then (i) when $\xi > 0$, $f(x)$ is an increasing function; (ii) when $-1/2 < \xi < 0$, $f(x)$ is a decreasing function.
Therefore, the optimal weights \( \hat{\Gamma} \) has the same sign as that of \( (a/x)^{\xi+1} - (\xi + 1)a/x + \xi \). Consider the function \( s(t) = t^{\xi+1} - (\xi + 1)t + \xi, \quad t > 0 \). For \( \xi > 0 \), \( s''(t) = \xi(\xi + 1)t^{\xi-1} > 0 \), so \( s(t) \) is a convex function that achieves its minimum at \( t = 1 \). Since \( s(1) = 0 \), \( s(t) \) and \( f'(x) \) are both nonnegative. Thus \( f(x) \) is an increasing function for \( \xi > 0 \). To prove (ii), we can use the same technique to show that \( s(t) \) is a concave function achieving its maximum at \( t = 1 \), and thus \( f'(x) \leq 0 \) for all \( x > 0 \).

**Proof of Remark 2.** Recall that the matrix \( \Gamma \) is a \( K \times K \) matrix with the \((k,k')\)th element defined as \( \min(l_k,l_{k'}) \). Then it can be shown that \( \Gamma^{-1} \) is a band matrix with the following form

\[
\Gamma^{-1} = \begin{pmatrix}
\frac{1}{l_1-l_2} & -\frac{1}{l_1-l_2} & 0 & 0 \\
-\frac{1}{l_1-l_2} & \frac{1}{l_2-l_3} + \frac{1}{l_3-l_4} & -\frac{1}{l_2-l_3} & 0 \\
0 & -\frac{1}{l_2-l_3} & \frac{1}{l_3-l_4} & 0 \\
0 & 0 & -\frac{1}{l_3-l_4} & \frac{1}{l_2-l_3} + \frac{1}{l_3-l_4} \\
0 & 0 & 0 & -\frac{1}{l_3-l_4} \\
0 & 0 & 0 & 0 \\
\end{pmatrix}
\]

Therefore, the optimal weights \( \Gamma^{-1}\phi(\xi) / 1^T_K \Gamma^{-1}\phi(\xi) = (\omega_k)_{k=1}^K \), where

\[
\omega_1 = c \left( \frac{l_{\xi+1}^{\xi+1}}{l_1 - l_2} - \frac{l_{\xi+1}^{\xi+1}}{l_1 - l_2} \right), \quad \omega_K = c \left\{ \frac{l_{K-1}^{\xi}}{l_{K-1}^{\xi} - l_{K-1}^{\xi}} \right\},
\]

and \( \omega_k = c \left( \frac{l_{\xi+1}^{\xi+1} - l_{k+1}^{\xi+1}}{l_k - l_{k+1}} - \frac{l_{k+1}^{\xi+1} - l_k^{\xi+1}}{l_{k-1} - l_k} \right), \) for \( k = 2, \ldots, K - 1 \),

where \( c = 1^T_K \Gamma^{-1}\phi(\xi) \) is a positive constant. We consider the three different cases separately.

(i) Case 1 \((\xi > 0)\). Note that \( l_1 > l_2 > \ldots > l_K \). Obviously \( \omega_1 > 0, \omega_K < 0 \). For any \( k = 2, \ldots, K - 1 \), let \( f(x) = \frac{x^{\xi+1} - x^{\xi+1}}{l_k - x} \), then \( \omega_k = c\{ f(l_{k+1}) - f(l_k) \} < 0 \) by (i) in Lemma
2.5.3.

(ii) Case 2 ($\xi = 0$). It is easy to show that $\omega_1 = 1$ and $\omega_2 = \ldots = \omega_K = 0$.

(iii) Case 3 ($-1/2 < \xi < 0$). By Lemma 2.5.3 (ii) and the similar technique as used in the proof for case 1, we can show that $\omega_k > 0$ for $k = 1, \ldots, K$.

The proof of Proposition 2.2.1 relies on the following lemma.

**Lemma 2.5.4 (Lemma 2 of Zhao and Xiao, 2013)** Let $S$ be a $K \times K$ symmetric positive-definite matrix and $v$ be any non-zero $K \times 1$ column vector. Define $M = v^T S^{-1} v S - vv^T$. Then (i) for any column vector $z$, $z^T M z \geq 0$; and (ii) $z^T M z = 0$ holds if and only if $z = c S^{-1} v$ for some real constant $c$.

**Proof of Proposition 2.2.1.** Since the matrix $\Gamma$ is positive-definite and symmetric, and $\phi(\xi)$ is non-zero, by Lemma 2.5.4 (i), we have

$$
\omega^T \phi^T(\xi) \Gamma^{-1} \phi(\xi) \Gamma \omega - \omega^T \phi(\xi) \phi^T(\xi) \omega \geq 0, \quad \text{for any } \omega \in \mathbb{R}^K. \quad (2.28)
$$

Note that $\phi^T(\xi) \Gamma \phi(\xi)$ is a scalar, (2.28) can be expressed as

$$
\frac{\omega^T \Gamma \omega}{\omega^T \phi(\xi) \phi^T(\xi) \omega} = \sigma^2_{WCRQ}(\omega) \geq \{\phi^T(\xi) \Gamma \phi(\xi)\}^{-1}, \quad \text{for any } \omega \in \mathbb{R}^K.
$$

Lemma 2.5.4 (ii) then implies that the equality in (2.28) holds if and only if $\omega = c \Gamma^{-1} \phi(\xi)$ for some constant $c$.

**Proof of Theorem 2.2.3.** By the definitions, $B(\theta)$ is the first derivative of $E\{A(\theta)\}$. With the Taylor series expansion, we get

$$
E\{A(\tilde{\theta})\} = E\{A(\theta_0)\} + B(\theta)(\tilde{\theta} - \theta_0), \quad (2.29)
$$

where $\tilde{\theta}$ lies between $\theta_0$ and $\bar{\theta}$. Define

$$
r_n(\delta) = A(\theta + \delta) - A(\theta) = \sum_{k=1}^{K} \sum_{i=1}^{n} \omega^{(i)} z_{i,k} [I\{y_i - z_{i,k}^T(\theta + \delta) < 0\} - I(y_i - z_{i,k}^T \theta < 0)].
$$
Applying Lemma 4.1 of He and Shao (1996), we have the uniform approximation

$$
\sup_{\delta:||\delta|| \leq C} ||r_n(\delta) - E\{r_n(\delta)\}|| = O_p(\sqrt{n} \log n ||\delta||^{1/2}), \text{ for some constant } C.
$$

Since $\tilde{\theta}$ is an $a_n$-consistent estimator of $\theta_0$,

$$
||\{A(\tilde{\theta}) - A(\theta_0)\} - [E\{A(\tilde{\theta})\} - E\{A(\theta_0)\}]|| = O_p(\sqrt{n} \log n ||\tilde{\theta} - \theta_0||^{1/2}). \quad (2.30)
$$

Combining (2.29) and (2.30) gives

$$
\hat{\theta}_{OS} - \theta_0 = -B(\tilde{\theta})A(\theta_0) - R_n,
$$

where $R_n = B(\tilde{\theta})^{-1}\{B(\tilde{\theta}) - B(\theta)\}(\tilde{\theta} - \theta_0) + B(\tilde{\theta})^{-1}O_p(\sqrt{n} \log n ||\tilde{\theta} - \theta_0||^{1/2})$. Following similar arguments as in the proof of Theorem 3 in Bradic et al. (2011), we can show that $R_n$ is $o_p(1/a_n)$. Consequently, to prove Theorem 2.2.3, we only need to consider $a_n\{B(\tilde{\theta})A(\theta_0)\}$. Since $\tilde{\theta}$ is a consistent estimator, by Slusky theorem, it is sufficient to show the asymptotic normality of $a_n\{B(\theta_0)A(\theta_0)\}$. By the regularly varying property in (2.25), Lemma 2.5.2 and the multivariate CLT, we can show that

$$
a_n\{B(\theta_0)A(\theta_0)\} \overset{d}{\to} N\left(0, T^{-1}JT^{-1}\right),
$$

where

$$
T = \left(\frac{m^{-\xi} - 1}{-\xi}\right)^{\frac{1}{2}} \begin{pmatrix}
\omega_1^{(o)}\xi^{+1} & 0^T \\
\vdots & \ddots & \vdots \\
\omega_K^{(o)}\xi^{+1} & 0^T \\
0 & \ldots & 0 & \omega_{opt}^T\phi(\xi)D
\end{pmatrix}, \quad J = \begin{pmatrix}
J_1 & 0 \\
0 & D\omega_{opt}^T\Gamma_{opt}\phi(\xi)
\end{pmatrix}
$$

and $J_1$ is a $K \times K$ matrix with the $(k, k')$th element $\omega_k\omega_{k'}\min(l_k, l_{k'})$. By some linear algebra, we can show that the lower right $p \times p$ block of $T^{-1}JT^{-1}$ is

$$
\frac{\omega_{opt}^T\Gamma_{opt}}{\{\phi^T(\xi)\omega_{opt}\}^2} \left(\frac{m^{-\xi} - 1}{-\xi}\right)^{-2} D^{-1}.
$$

The proof is completed by plugging in $\omega_{opt} = \Gamma^{-1}\phi(\xi)/\{1_k^T\Gamma^{-1}\phi(\xi)\}$. 

34
Chapter 3

Combined Estimation for Conditional Value-at-Risk

3.1 Introduction

Value-at-Risk (VaR) is a widely used risk measure in financial risk management. The concept of VaR was introduced in 1990’s, and soon became the standard measure in many financial institution. Given a probability \( \tau \in (0, 1) \), the VaR of a portfolio is a threshold number such that the probability that the loss of the portfolio exceeds this value is \( \tau \). From a statistical point of view, the VaR is just the negative of the conditional \( \tau \)-th quantile of the returns. Given the information at various time points, the goal is to estimate the VaR at each time point and also predict the VaR in the future. Different models of VaR has been intensively discussed in the literature. Generally, the models can be categorized into three classes: parametric models, nonparametric models and semi-parametric models. Parametric models such as RiskMetrics (Longerstaey and Spencer, 1996) and GARCH model (Bollerslev, 1986) estimate the VaR through the conditional volatility of returns. This type of models usually specify the residuals distribution as normal, and require the i.i.d. residual assumption to estimate the parameters. However, the well-recognized features of return distribution are heavy-tailed and autocorrelated (Fama, 1965), which makes the parametric models undesirable. For nonparametric models, the most famous ones are historical simulation, which uses sample quantile within a rolling window to obtain the VaR forecast, and the hybrid method (Richardson et al., 1997), which improves over historical simulation by assigning exponentially inclining weights
to less recent returns. These nonparametric methods are free of distribution assumption and easy to implement. But as mentioned by Manganelli and Engle (2001), these methods suffer from jumps in predicted VaR caused by some extreme returns. In addition, as far as we know, there is no rule of thumb to choose the window length. Semiparametric methods consist of those based on Extreme Value Theory (EVT), which estimate the tail parameter of the return distribution, and quantile regression, which model the conditional quantiles directly. EVT is an excellent tool to handle the heavy-tailedness nature of the return distribution; works done in this direction include Chernozhukov and Umantsev (2001), McNeil and Frey (2000), Manganelli and Engle (2001), and Brooks et al. (2005). However, EVT-based methods often require independence assumption, which contradicts the autoregressive nature of the return data. Estimating VaR via quantile regression is a natural thought and has several advantages. Compared with other types of models, quantile regression is robust against extreme returns and free of distributional assumption. Chernozhukov and Du (2008) investigated quantile regression in extreme order quantiles and applied the method for VaR estimation. Xiao and Koenker (2009a) relaxed the normality assumption in GARCH models by incorporating quantile regression with GARCH models. Koenker and Xiao (2006a) considered the quantile autoregression. Among the approaches based on quantile regression, the conditional autoregressive value at risk by regression quantiles (CAVIAr) proposed by Engle and Manganelli (2004) is the most popular one. CAVIAr model captures the autocorrelation of the VaR, and also takes the previous information such as returns and other covariates into account. Its appealing feature has been also recognized in the literature, e.g., in Manganelli and Engle (2001), Bao et al. (2006), and Kuester et al. (2006).

However, as pointed out by Xiao and Koenker (2009a), Engle and Manganelli (2004) focused on the construction of models rather than the estimating process. Since CAVIAr is nonlinear and autoregressive, the estimation is rather complicated. In Engle and Manganelli (2004), they proposed choosing a starting value by grid search, and then use quasi-Newton and simplex algorithm iteratively. We observed that in practical applications some parameters in CAVIAr model are invariant to the quantile levels (i.e. global) while some may be varying across quantiles. Composite quantile regression based on the common slope feature is promising to improve the efficiency of the parameter estimation, especially when the return distribution has heavy tails (Wang and Wang, 2015). In this chapter, we propose a new composite CAVIAr estimator. In addition, we suggest using a
working model for the initial estimation, which can largely reduce the computation time without losing accuracy of the estimates.

3.2 CAViaR Model

Suppose $y_t$ is the observed return at time $t = 1, \ldots, T$. By definition, the VaR is essentially the conditional $\tau$-th quantile of the return at time $t$:

$$\Pr(y_t < -VaR_t|\mathcal{F}_{t-1}) = \Pr(y_t < Q_{t,\tau}|\mathcal{F}_{t-1}) = \tau.$$igion

Among all the methods on VaR estimation the CAViaR model proposed by Engle and Manganelli (2004) drew the most attention. In fact, CAViaR is an analogue of Generalized autoregressive conditional heteroskedasticity (GARCH) model (Bollerslev, 1986). GARCH model allows the current volatility, i.e., variance, to depend on previous volatility. Similarly, CAViaR models the current quantile using the historical information including quantiles at previous time points, leading to an autoregressive type of modelling approach. This improves the flexibility of the model to allow the consideration of the autocorrelation structure of VaR.

Let $\theta_{\tau} = (\alpha_{\tau}, \beta_{\tau}', \gamma_{\tau}')$ be the vector of parameters in a CAViaR model. The CAViaR model has the following form:

$$Q_{t,\tau}(\theta_{\tau}) = \alpha_{\tau} + \sum_{i=1}^{p} \beta_{i,\tau} Q_{t-i,\tau}(\theta_{\tau}) + \sum_{j=1}^{q} \gamma_{j,\tau} l(x_{t-j}), \quad (3.1)$$

where $p + q + 1$ is the dimension of $\theta_{\tau}$ and $l$ is a function of previous information $x_{t-j}$. Usually, $x_{t-j}$ is chosen as previous returns $y_{t-j}$. Depending on the values of $p, q$ and the function $l$, we have the following models that are often used:

Symmetric absolute value:

$$Q_{t,\tau}(\theta_{\tau}) = \alpha_{\tau} + \beta_{\tau} Q_{t-1,\tau}(\theta_{\tau}) + \gamma_{\tau} |y_{t-1}|.$$}

Asymmetric slope:

$$Q_{t,\tau}(\theta_{\tau}) = \alpha_{\tau} + \beta_{\tau} Q_{t-1,\tau}(\theta_{\tau}) + \gamma_{1,\tau} (y_{t-1})^+ + \gamma_{2,\tau} (y_{t-1})^-.$$
Indirect GARCH(1,1):

\[ Q_{t,\tau}^2(\theta_\tau) = \alpha_\tau + \beta_\tau Q_{t-1,\tau}^2(\theta_\tau) + \gamma_\tau y_{t-1}^2. \]

Under a specific CAViaR model, we can estimate the unknown parameter \( \theta_\tau \) at a given quantile level \( \tau \) by

\[ \hat{\theta}_\tau = \arg \min_{\theta_\tau} \sum_{t=1}^{T} \rho_\tau \{ y_t - Q_{t,\tau}(\theta_\tau) \}. \]

3.3 Combined Quantile Regression

3.3.1 Composite Quantile Loss Function

The conventional CAViaR estimator is obtained by fitting the nonlinear quantile regression at a given quantile level \( \tau \). However, in some applications, some components of \( \theta \) may be constant across quantile levels, and in such cases the conventional local estimator of \( \theta \) may not be efficient. For example, suppose the returns follow a linear GARCH \((p,q)\) model (Taylor, 2007), i.e.,

\[ y_t = \sigma_t \epsilon_t, \]
\[ \sigma_t = \alpha + \sum_{i=1}^{p} \beta_i \epsilon_{t-i} + \sum_{j=1}^{q} \gamma_j |y_{t-j}|. \]  
(3.2)

Then the \( \tau \)-th conditional quantile of \( y_t \) has the CAViaR representation:

\[ Q_{t,\tau}(\theta_\tau) = \alpha_\tau + \sum_{i=1}^{p} \beta_{i,\tau} Q_{t-1,\tau}(\theta_\tau) + \sum_{j=1}^{q} \gamma_{j,\tau} |y_{t-j}|, \]  
(3.3)

where \( \alpha_\tau = \alpha F^{-1}_\epsilon(\tau), \beta_{i,\tau} = \beta_i, \gamma_{j,\tau} = \gamma_j F^{-1}_\epsilon(\tau) \) and \( F_\epsilon(\cdot) \) is the cumulative distribution function of \( \epsilon_i \). We can see that \( \beta_i \)'s in model (3.3) do not depend on \( \tau \), while \( \alpha_\tau \) and \( \gamma_{j,\tau} \) still depend on \( \tau \). In cases where some parameters are indeed global (constant across \( \tau \)), it is natural to borrow the combined estimation idea introduced in Chapter 2, to improve the estimation efficiency of the global parameters, which could consequently improve the prediction.

We assume that within a quantile region \( T \), the vector of parameters \( \beta_\tau = \beta \) in
model (3.3) is constant. Consider a set of quantile levels \( \tau_1, \tau_2, \ldots, \tau_K \in \mathcal{T} \), and let the quantile-specific parameters be \( \alpha_k, \quad k=1, \ldots, K \). Let \( \theta = (\alpha'_1, \ldots, \alpha'_K, \beta')' \) denote the vector of unknown parameters, and let \( \theta_0 = (\alpha'_{1,0}, \ldots, \alpha'_{K,0}, \beta'_0)' \) denote the truth of \( \theta \).

We propose to estimate the parameter vector \( \theta \) by minimizing the composite quantile loss function:

\[
\hat{\theta} = \arg \min_{\theta \in \mathbb{R}^{K(q+1)+p}} \sum_{t=1}^{T} \sum_{k=1}^{K} \rho_{\tau_k} \{ y_t - Q_{t,\tau_k}(\theta) \}.
\]

Hereafter we refer to \( \hat{\theta} \) as the composite CAViaR estimator.

### 3.3.2 Asymptotic Property

The parameters in the composite CAViaR model are estimated via quantile regression. Different from conventional quantile regression, the models implied by CAViaR are autoregressive and might be nonlinear in parameters. Based on White (1996)’s work on nonlinear quantile regression, Engle and Manganelli (2004) established the asymptotic distribution of CAViaR estimator. In this paper, we extend the asymptotic result to multi-quantile case.

For any \( k = 1, \ldots, K \), the vector \( \alpha_{k,0} \) is not just an intercept, but it includes all the quantile-specific parameters (like \( \gamma_k \) in the GARCH model example) corresponding to \( \tau_k \).

We define:

\[
\begin{align*}
\phi_{t,k} &= I\{ y_t < Q_{t,\tau_k}(\theta) \} - \tau_k, \\
g_{t,k} &= \phi_{t,k}(\theta) \nabla Q_{t,\tau_k}(\theta), \\
\lambda_{T,k}(\theta) &= \frac{1}{T} \sum_{t=1}^{T} E\{ g_{t,k}(\theta) \}, \\
\mu_{t,k}(\theta, d) &= \sup_{||\theta - \tilde{\theta}|| < d} ||g_{t,k}(\theta) - g_{t,k}(\tilde{\theta})||, \quad k = 1, \ldots, K.
\end{align*}
\]

**Theorem 3.3.1** If model (3.1) and conditions **C1-C7** and **AN1-A4** (defined in Section 3.8) hold, we have

\[
A_T^{-1/2} D_T \sqrt{T} (\hat{\theta} - \theta_0) \overset{d}{\rightarrow} N(0, I),
\]
where

\[
\begin{align*}
A_T &= \frac{1}{T} \sum_k \sum_{k'} \{\min(\tau_k \tau_{k'}) - \tau_k \tau_{k'}\} E \left\{ \sum_{t=1}^{T} \nabla' Q_{t,\tau_k}(\theta_0) \nabla Q_{t,\tau_{k'}}(\theta_0) \right\}, \\
D_T &= \frac{1}{T} \sum_k E \left[ \sum_{t=1}^{T} f_t \{Q_{t,\tau_k}(\theta_0) | \Omega_t \} \nabla' Q_{t,\tau_k}(\theta_0) \nabla Q_{t,\tau_k}(\theta_0) \right].
\end{align*}
\]

The covariance matrices involved in the covariance matrix of \(\hat{\theta}\) can be estimated by

\[
\begin{align*}
\hat{A}_T &= \frac{1}{T} \sum_k \sum_{k'} \{\min(\tau_k \tau_{k'}) - \tau_k \tau_{k'}\} \sum_{t=1}^{T} \nabla' Q_{t,\tau_k}(\hat{\theta}) \nabla Q_{t,\tau_{k'}}(\hat{\theta}), \\
\hat{D}_T &= \frac{1}{T} \sum_k \sum_{t=1}^{T} \frac{1}{2\hat{c}_{T,k}} I(|y_t - Q_t(\hat{\theta})| < \hat{c}_{T,k}) \{Q(\tau_k) | \Omega_t \} \nabla' Q_{t,\tau_k}(\hat{\theta}) \nabla Q_{t,\tau_k}(\hat{\theta}),
\end{align*}
\]

where \(\hat{c}_{T,k}\) is the bandwidth defined by the Variance-Covariance matrix estimation assumptions in the Variance-Covariance assumptions of Section 3.8. In their implementation, Engle and Manganelli (2004) defined the residuals as \(\epsilon_t = y_t - Q_{t,\tau}(\hat{\theta}_r)\), and they chose the bandwidth to be the 40-th smallest value of \(\epsilon_t\)'s when \(\tau = 0.01\), and the 60-th smallest value of \(\epsilon_t\) when \(\tau = 0.05\).

3.4 Computation

Note that the CAViaR model in (3.1) has the recursive construction, which makes the optimization techniques in conventional quantile regression not applicable. Engle and Manganelli (2004) proposed to use the following algorithm. First they generated a large number (usually 5000 or 10000) of vectors with length \(K \times (q + 1) + p\) from independent Uniform (0,1). For each initial value, they used the Nelder-Mead simplex and the quasi-Newton method iteratively until the convergence criterion is met (tolerance levels for the function and the parameters are 10\(^{-10}\)). At last, the vector that yields the smallest quantile loss is chosen to be the estimate.

Our empirical studies show that the CAViaR estimator obtained from the above procedure is very sensitive to the initial value. The above grid search method is computationally intensive even for single quantile level, and it becomes worse when considering \(K\) quantile levels, where the parameter dimension increases from \(p + q + 1\) to \(K \times (q + 1) + p\).
We propose to obtain the initial estimates by combining a quantile regression and grid search (referred to as Mix). Specifically, we first estimate $\alpha(\tau)$ and $\gamma(\tau)$ by $\hat{\alpha}(\tau)$ and $\hat{\gamma}(\tau)$, the estimated intercept and slope from fitting the quantile regression of $y_t$ on $|y_{t-1}|$, and then perform grid search around $\hat{\alpha}(\tau)$ and $\hat{\gamma}(\tau)$. For $\beta$, we still grid search on $(0,1)$. When using simplex and quasi-Newton algorithm, we add an non-negative restriction on $\beta$.

Alternatively, we can use some existing (but simpler) working models to obtain some initial estimations of $Q_t,\tau$’s first, and then estimate the initial value by performing quantile regression of $y_t$ on the estimated conditional quantiles and other previous information. We consider the following working models.

- **QGARCH**: Xiao and Koenker (2009b) proposed a two-step estimator for the conditional quantiles under the assumption that $y_t$ follows a linear GARCH process in (3.2). Under some regularity assumption, the volatility $\sigma_t$ has the $\text{ARCH}(\infty)$ representation:
  \[
  \sigma_t = 1 + \sum_{j=1}^{\infty} a_j |y_{t-j}|, \tag{3.5}
  \]
  where $a_j$’s satisfy the summability conditions (Bollerslev (1986)). By plugging (3.5) into (3.2), we have
  \[
  Q_t,\tau = Q_{\epsilon_t,\tau} + \sum_{j=1}^{\infty} \alpha_j |y_{t-j}|, \tag{3.6}
  \]
  where $\alpha_j = a_j Q_{\epsilon_t,\tau}$, and $Q_{\epsilon_t,\tau}$ is the $\tau$-th quantile of $\epsilon_t$. Without losing too much accuracy, Xiao and Koenker (2009a) considered the $m$-truncated quantile autoregression
  \[
  Q_t,\tau = Q_{\epsilon_t,\tau} + \sum_{j=1}^{m} \alpha_j |y_{t-j}|. \tag{3.7}
  \]
  The first step is to estimate $a_j$, $j = 1, ..., m$, by $\hat{\alpha}_j/\hat{Q}_{\epsilon_t,\tau}$, where $\hat{\alpha}_j$ and $\hat{Q}_{\epsilon_t,\tau}$ are obtained from quantile regression of $y_t$ on $|y_{t-j}|$. And the volatility is estimated by
  \[
  \hat{\sigma}_t = 1 + \sum_{j=1}^{\infty} \hat{a}_j |y_{t-j}|. \tag{3.8}
  \]
  The second step is to perform quantile regression of $y_t$ on the covariates $(1, \hat{\sigma}_{t-1}, ..., \hat{\sigma}_{t-p}, |y_{t-1}|, ..., |y_{t-q}|)$, and predict the conditional quantile $Q_t,\tau$ by plug-
ging in the estimated coefficients.

- **QAR(p)**: We use quantile autoregression proposed by Koenker and Xiao (2006b) with order $p$. Specifically, we perform quantile regression of $y_t$ on $y_{t-1}, \ldots y_{t-p}$ to have a rough estimation of the conditional $\tau$-th quantile.

### 3.5 VQR test

In practice, it is important to test the model sufficiency and conduct model selection, especially when a number of VaR models are considered. Many backtests have been discussed in the literature; see Kupiec (1995), Christoffersen (1998), Berkowitz et al. (2011), Gaglianone et al. (2011), Engle and Manganelli (2004). Among them, Gaglianone et al. (2011) proposed to test the Value-at-Risk using quantile regression (VQR test). In this section, we provide the theoretical justification of the VQR test. For a given model, let $Q_{t,\tau}^*$ be the estimated $\tau$-th conditional quantile of returns given previous returns. Consider the following model,

$$y_t = \alpha_0(U_t) + \alpha_1(U_t)Q_{t,\tau}^*, \quad (3.6)$$

where $Q_{t,\tau}^*$ is $\Omega_{t-1}$-measurable, $U_t \sim$ i.i.d. Uniform(0,1), $y_t$ is an increasing function of $U_t$. Consequently, the conditional $\tau$-th quantile of $y_t$ can be represented as

$$Q_{t,\tau} = \alpha_0(\tau) + \alpha_1(\tau)Q_{t,\tau}^*, \quad \tau \in (0,1). \quad (3.7)$$

Intuitively, if a model is correctly specified and the conditional quantiles are accurately estimated, the coefficients $\alpha_0$ and $\alpha_1$ in model (3.7) should be close to 0 and 1. Let $\lambda(\tau) = \{\alpha_0(\tau), \alpha_1(\tau) - 1\}'$, and $\hat{\lambda}(\tau)$ is the quantile regression estimator of $\lambda(\tau)$, then the null hypothesis of the VQR test is $H_0 : \lambda(\tau) = 0$. Now the backtest problem is transformed into a quantile regression problem, where $Q_{t,\tau}^*$ is the predictor and $y_t$ is the response.

The VQR test has the advantage of easy implementation and it also utilizes more information compared to the backtests based on the binary variable $\text{Hit}_{t,\tau} = I\{y_t < \hat{Q}_{t,\tau}\}$ (Gaglianone et al., 2011). In addition, Gaglianone et al. (2011) showed that the VQR test has decent size and power for large sample sizes.

The Wald-type test proposed in Proposition 2 of Gaglianone et al. (2011) is a direct result of the asymptotic distribution of $\hat{\lambda}(\tau)$ which was obtained from fitting the model.
However, the original proof in the paper did not consider the dependency among data. Gaglianone et al. (2011) made the following assumptions:

B1. Let $Q_{t,\tau}^*$ be $\Omega_{t-1}$-measurable and $\{y_t, Q_{t,\tau}^*\}$ be as strictly stationary process.

B2. Let $y_t$ have conditional distribution function $F_t$ with continuous Lebesgue density $f_t$ uniformly bounded away from 0 and $\infty$ at the points $Q_{t,\tau}$ for all $\tau \in (0, 1)$.

B3. There exist positive definite matrices $J$ and $H_\tau$, such that for all $\tau \in (0, 1)$:

$$J = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} x_t x_t', \quad (3.8)$$

$$H_\tau = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} x_t x_t' f_t(Q_{t,\tau}), \quad (3.9)$$

where $x_t = (1, Q_{t,\tau}^*)'$.

B4. $\max_{t=1,...,T} ||x_t||/\sqrt{T} \overset{p}{\to} 0$.

**Theorem 3.5.1** If model (3.1), and Assumptions B1-B4 hold, we have

$$\sqrt{T} \{ \hat{\lambda}(\tau) - \lambda(\tau) \} \overset{d}{\to} N\{0, \tau(1-\tau)H^{-1}_\tau J H^{-1}_\tau\}, \text{ as } T \to \infty.$$  

Furthermore, the Wald-type VQR test statistic is defined as

$$\zeta_{\text{VQR}} = T \left[ \hat{\lambda}(\tau)' \{ \tau(1-\tau)H^{-1}_\tau J H^{-1}_\tau\}^{-1} \hat{\lambda}(\tau) \right],$$

and under null hypothesis $H_0 : \lambda(\tau) = 0$, the Wald-type VQR test statistic $\zeta_{\text{VQR}}$ is asymptotically chi-square distributed with 2 degrees of freedom.

Gaglianone et al. (2011) proved Theorem 3.5.1 by stating that “A sketch of the proof of this CLT, via a Bahadur representation, is also presented in Hendricks and Koenker (1992)”. However, Hendricks and Koenker (1992) made the key assumption that the responses are independent across time, which allows the application of the Lindeberg CLT to the Bahadur representation. In the case of model (4.1) and general VaR models in the literature, the responses are recursive and thus not independent. The predictors are not independent either. Therefore, the original proofs in Gaglianone et al. (2011) are
not rigorous. We next sketch a modified proof based on the Knight’s identity and the martingale CLT.

To get around the problem, we use the martingale CLT proposed by Pollard (1984) and used in Pollard (1991). Pollard (1991) considered $\Omega_{t-1}$-measurable predictor $x_t$, and i.i.d. regression errors. To prove the asymptotic result in (3.5.1), we can follow the proof of Theorem 4.1 in Koenker (2005, page 120). For $Z_{1,T}(\delta) = -T^{-1/2} \sum_{t=1}^{T} x_t' \delta \phi_t$, instead of applying Lindeberg CLT, now we apply martingale CLT to get $Z_{1,T}(\delta) \overset{d}{\to} -\delta^T W$, where $W \sim N(0, \tau(1-\tau)J)$.

We write $Z_{2,T}(\delta)$ as

$$Z_{2,T}(\delta) = \sum_{t=1}^{T} E\{Z_{2,Tt}(\delta)|\Omega_{t-1}\} + \sum_{t=1}^{T} [Z_{2,Tt}(\delta) - E\{Z_{2,Tt}(\delta)|\Omega_{t-1}\}].$$

We can show that

$$\sum_{t=1}^{T} E\{Z_{2,Tt}(\delta)|\Omega_{t-1}\} \overset{p}{\to} \frac{1}{2} \delta^T H \delta. \quad (3.10)$$

Let $\Delta_t = Z_{2,Tt}(\delta) - E\{Z_{2,Tt}(\delta)|\Omega_{t-1}\}$. Then $\Delta_t$’s are martingale differences, i.e., $E(\Delta_t|\Omega_{t-1}) = 0$. It is easy to see that

$$\sum_t Var(\Delta|\Omega_{t-1}) = \sum_t Var[Z_{2,Tt}(\delta) - E\{Z_{2,Tt}(\delta)|\Omega_{t-1}\}|\Omega_{t-1}]$$

$$= \sum_t Var\{Z_{2,Tt}(\delta)|\Omega_{t-1}\}$$

$$\leq \sum_t E[Z_{2,Tt}(\delta)^2|\Omega_{t-1}]$$

$$\leq \frac{1}{\sqrt{T}} \max(x_t'\delta) \sum_{t=1}^{T} E\{Z_{2,Tt}(\delta)|\Omega_{t-1}\}$$

$$\overset{p}{\to} 0.$$

Following the same arguments as in Pollard (1991, proof of theorem 2), the condition $\sum_t Var(\Delta|\Omega_{t-1}) \overset{p}{\to} 0$ implies $\sum_t \Delta_t \overset{p}{\to} 0$. Therefore, we have

$$\sum_{t=1}^{T} [Z_{2,Tt}(\delta) - E\{Z_{2,Tt}(\delta)|\Omega_{t-1}\}] \overset{p}{\to} 0. \quad (3.11)$$
Combining (3.10) and (3.11) gives
\[ Z_{2T}(\delta) \overset{p}{\to} \frac{1}{2}\delta^T H \delta. \]

Then the asymptotic result in Theorem 3.5.1 follows immediately.

### 3.6 Simulation

#### 3.6.1 Choices of Initial Values for the Conventional CAViaR Estimator

We generate data from the following symmetric absolute value model:
\[ Q_t(\tau) = \alpha(\tau) + \beta Q_{t-1}(\tau) + \gamma(\tau)|y_{t-1}|, \tag{3.12} \]
where \( \tau = 0.05, \alpha(\tau) = \alpha \Phi^{-1}(\tau), \gamma(\tau) = \gamma \Phi^{-1}(\tau), \) and \( \Phi(\cdot) \) is the cumulative distribution of \( N(0,1) \). We consider three sets of parameter values:

- **P1**: \( \alpha = 0.1, \beta = 0.5, \gamma = 0.3. \)
- **P2**: \( \alpha = 0.1, \beta = 0.8, \gamma = 0.1. \)
- **P3**: \( \alpha = 0.1, \beta = 0.9, \gamma = 0.05. \)

Under the first set of parameters, the generated GARCH process is stationary. But when \( \beta + \gamma \) gets closer to 1 as in P2 and P3, the generated process is close to nonstationary, making the estimation more challenging. We consider sample sizes \( T = 100 \) and 500. So there are 6 settings in total:

- **Set1**: P1, \( T=100. \)
- **Set2**: P1, \( T=500. \)
- **Set3**: P2, \( T=100. \)
- **Set4**: P2, \( T=500. \)
- **Set5**: P3, \( T=100. \)
More specifically, the data generating process is as follows. First, we generate a random variable \( y_1 \) from the standard normal distribution. We then generate an i.i.d. random sample \( u_1, \ldots, u_n \) from Uniform \((0,1)\). Then according to model (3.12), we generate \( y_t \) from
\[
y_t = \alpha (u_{t-1}) + \beta Q_{t-1}(u_{t-1}) + \gamma (u_{t-1})|y_{t-1}|, \quad t = 2, \ldots, t,
\]
where \( Q_t(u) = \Phi^{-1}(u) \).

We use the simplex algorithm and quasi-Newton algorithm iteratively to estimate the parameters. We compare four methods for selecting the initial values: Grid search, QAR(5), QGARCH with \( m = 3T^{1/4} \), and Mix (see Section 3.3 for details). For each method and setting, the number of repetition is 50. Let \( Q_{t,\tau}^{(i)} \), \( i = 1, \ldots, 50 \) be the true \( \tau \)-th quantile at time \( t \) from the \( i \)-th iteration, and \( \hat{Q}_{t,\tau}^{(i)} \) be the estimation of \( Q_{t,\tau}^{(i)} \). We use MSE and bias of the estimated conditional quantile defined below as the comparison measures:

\[
MSE = \frac{1}{T \times n_{sim}} \sum_{i=1}^{n_{sim}} \sum_{t=1}^{T} (\hat{Q}_{t,\tau}^{(i)} - Q_{t,\tau}^{(i)})^2,
\]

\[
Bias_1 = \text{median} \left\{ \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} (\hat{Q}_{t,\tau}^{(i)} - Q_{t,\tau}^{(i)}) \right\},
\]

\[
Bias_2 = \frac{2}{T} \sum_{t=T/2+1}^{T} \left\{ \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} (\hat{Q}_{t,\tau}^{(i)} - Q_{t,\tau}^{(i)}) \right\}.
\]

We define Bias1 and Bias2 as robust summaries to overcome the instability in the first several timepoints.

We’d like to point out that the GRID and MIX methods are both time-consuming because of the grid search. Based on our numerical studies, on average, the computation time of QGARCH and QAR method is less than 1/10 of the GRID and MIX methods.

The MIX method performs uniformly better than the GRID method and is a good option for obtaining conventional CAViaR estimator at a single quantile level. But when we consider composite estimation, the computation time is a major concern, since every \( \tau_k \) will bring in \( q + 1 \) additional dimensions to the parameter space.
Table 3.1: The MSE, Bias1 and Bias2 of 4 methods using different initial values.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Method</th>
<th>Set1</th>
<th>Set2</th>
<th>Set3</th>
<th>Set4</th>
<th>Set5</th>
<th>Set6</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>grid</td>
<td>0.0495</td>
<td>0.0075</td>
<td>0.1394</td>
<td>0.0247</td>
<td>0.4326</td>
<td>0.1135</td>
</tr>
<tr>
<td></td>
<td>qgarch</td>
<td>0.0489</td>
<td>0.0077</td>
<td>0.1206</td>
<td>0.0258</td>
<td>$\infty$</td>
<td>0.1010</td>
</tr>
<tr>
<td></td>
<td>qar</td>
<td>0.0468</td>
<td>0.0085</td>
<td>0.1116</td>
<td>0.0267</td>
<td>0.3688</td>
<td>0.1038</td>
</tr>
<tr>
<td></td>
<td>mix</td>
<td>0.0437</td>
<td>0.0072</td>
<td>0.1087</td>
<td>0.0188</td>
<td>0.3574</td>
<td>0.0893</td>
</tr>
<tr>
<td>Bias1</td>
<td>grid</td>
<td>-0.0118</td>
<td>-0.0019</td>
<td>-0.0530</td>
<td>-0.0052</td>
<td>-0.1257</td>
<td>-0.0472</td>
</tr>
<tr>
<td></td>
<td>qgarch</td>
<td>-0.0017</td>
<td>-0.0029</td>
<td>-0.0158</td>
<td>-0.0013</td>
<td>$\infty$</td>
<td>-0.0223</td>
</tr>
<tr>
<td></td>
<td>qar</td>
<td>-0.0064</td>
<td>-0.0024</td>
<td>-0.0301</td>
<td>-0.0019</td>
<td>-0.1852</td>
<td>-0.0320</td>
</tr>
<tr>
<td></td>
<td>mix</td>
<td>-0.0102</td>
<td>-0.0022</td>
<td>-0.0174</td>
<td>-0.0026</td>
<td>-0.0766</td>
<td>-0.0213</td>
</tr>
<tr>
<td>Bias2</td>
<td>grid</td>
<td>-0.0109</td>
<td>-0.0035</td>
<td>-0.0496</td>
<td>-0.0081</td>
<td>-0.0588</td>
<td>-0.0260</td>
</tr>
<tr>
<td></td>
<td>qgarch</td>
<td>-0.0013</td>
<td>-0.0037</td>
<td>-0.0286</td>
<td>-0.0039</td>
<td>-0.0048</td>
<td>-0.0218</td>
</tr>
<tr>
<td></td>
<td>qar</td>
<td>-0.0110</td>
<td>-0.0038</td>
<td>-0.0277</td>
<td>-0.0046</td>
<td>-0.1328</td>
<td>-0.0184</td>
</tr>
<tr>
<td></td>
<td>mix</td>
<td>-0.0102</td>
<td>-0.0022</td>
<td>-0.0174</td>
<td>-0.0033</td>
<td>-0.0766</td>
<td>-0.0199</td>
</tr>
</tbody>
</table>

In Tables 3.1, the numbers in bold are the cases when QAR outperforms GRID. In four out of the six cases, QAR method gives better estimation, and in other cases, the differences are not substantial. This indicates that QAR is a better choice for selecting initial values, when both computation time and numerical performance are considered.

The QGARCH method is a computationally efficient method, but it gives very large MSE and Bias1 in Set5. This is due to the fact that in 2 simulated data, the $\hat{\beta}$ estimated by QGARCH has absolute value greater than 1, and this makes the recursively predicted quantiles explode. If we exclude these two simulated data, the QGARCH method has good performance in terms of MSE, Bias1 and Bias2. Noting that this poor performance of QGARCH may be due to small sample size and relatively extreme quantile levels, we still keep this QGARCH method in our later simulation study for composite estimation.

### 3.6.2 Comparison of Local (Conventional) Estimation and Composite Estimation

In this section we investigate the potential efficiency gain we may obtain from using composite estimation in CAViaR models with some constant global parameters. We consider two methods QGARCH and QAR($p$) for selecting the initial values, where the order of autoregression $p = 5$. For composite estimation, we use quantile levels $\tau = (0.05, 0.075, 0.1)$, and refer to the composite estimators with initial values selected by QGARCH and QAR.
methods as CQGARCH and CQAR, respectively. We also conduct local estimation at \( \tau = 0.05 \) and \( \tau = 0.1 \). We still follow the data generating process described in Subsection 3.6.1, but the sample sizes are increased to 500 and 1000. We consider total 6 settings:

Set1: P1: \( \alpha = 0.1, \beta = 0.5, \gamma = 0.3 \), T=500.
Set2: P1: \( \alpha = 0.1, \beta = 0.5, \gamma = 0.3 \), T=1000.
Set3: P2: \( \alpha = 0.1, \beta = 0.8, \gamma = 0.1 \), T=500.
Set4: P2: \( \alpha = 0.1, \beta = 0.8, \gamma = 0.1 \), T=1000.
Set5: P3: \( \alpha = 0.1, \beta = 0.9, \gamma = 0.05 \), T=500.
Set6: P3: \( \alpha = 0.1, \beta = 0.9, \gamma = 0.05 \), T=1000.

Note that in the data generating model (3.12), the parameters \( \alpha \) and \( \gamma \) are local and \( \beta \) is global.

Tables 3.2 and 3.3 summarize the MSE of the estimated conditional 0.05-th and 0.01-th quantiles for different methods. In most cases, the composite estimations are more efficient than their local counterpart. In addition, we observe unusually large MSE of QAR in Sets 3 and 5, and of QGARCH in Set 5. This is again due to the wild estimation of autoregressive coefficient \( \beta \) involved in the step of initial value calculation. In contrast, the composite estimators appear to be more stable with different initial value calculation methods. In Table 3.4, we report the MSE of \( \hat{\beta} \). In terms of parameter estimation, composite estimations are more advantageous, since in most cases composite estimates yield smaller MSE than their local counterparts. Between the two composite estimates, CQGARCH performs slightly better than CQAR. In Table 3.4, we report the MSE of different estimates of the common slope \( \beta \).

### 3.6.3 Assessment of Goodness of Fit via VQR Test

In this subsection we use the VQR test to assess the accuracy of estimated quantile for the CAViaR model and the composite CAViaR (abbreviated as CCAViaR) model. Data is generated from GARCH(1,1) model with parameter \( \alpha_0 = 0.1, \beta_0 = 0.8, \gamma_0 = 0.1 \). Sample size is 500 and the number of repetition is 400. For a local CAViaR model, we estimate the conditional quantile at \( \tau = 0.05 \) and \( \tau = 0.1 \). For composite CAViaR model, we combine information from three quantile levels 0.05, 0.075 and 0.1.
Table 3.2: The $1000 \times$ MSE of the estimated 0.05-th quantiles for composite and local estimations. The $1000 \times$ standard errors are reported in the parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Set1</th>
<th>Set2</th>
<th>Set3</th>
<th>Set4</th>
<th>Set5</th>
<th>Set6</th>
</tr>
</thead>
<tbody>
<tr>
<td>CQGARCH</td>
<td>7.5507</td>
<td>3.6908</td>
<td>21.5822</td>
<td>9.9801</td>
<td>91.0908</td>
<td>56.8035</td>
</tr>
<tr>
<td></td>
<td>(0.6049)</td>
<td>(0.3985)</td>
<td>(2.5930)</td>
<td>(1.6504)</td>
<td>(9.2859)</td>
<td>(5.5831)</td>
</tr>
<tr>
<td>QGARCH</td>
<td>7.6824</td>
<td>3.8467</td>
<td>25.7801</td>
<td>11.0349</td>
<td>100.9768</td>
<td>58.6562</td>
</tr>
<tr>
<td></td>
<td>(0.6578)</td>
<td>(0.4329)</td>
<td>(2.8331)</td>
<td>(2.1372)</td>
<td>(9.8759)</td>
<td>(7.0314)</td>
</tr>
<tr>
<td>CQAR</td>
<td>7.4757</td>
<td>3.6831</td>
<td>25.4672</td>
<td>12.6670</td>
<td>106.7804</td>
<td>61.2632</td>
</tr>
<tr>
<td></td>
<td>(0.6099)</td>
<td>(0.3958)</td>
<td>(2.5491)</td>
<td>(1.6388)</td>
<td>(9.5076)</td>
<td>(5.3615)</td>
</tr>
<tr>
<td>QAR</td>
<td>8.4642</td>
<td>3.8528</td>
<td>26.7245</td>
<td>13.4048</td>
<td>103.7974</td>
<td>66.5043</td>
</tr>
<tr>
<td></td>
<td>(0.8223)</td>
<td>(0.4326)</td>
<td>(2.6550)</td>
<td>(1.7989)</td>
<td>(10.3422)</td>
<td>(6.5903)</td>
</tr>
</tbody>
</table>

Table 3.3: The $1000 \times$ MSE of the estimated 0.1-th quantiles for composite and local estimations. The $1000 \times$ standard errors are reported in the parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Set1</th>
<th>Set2</th>
<th>Set3</th>
<th>Set4</th>
<th>Set5</th>
<th>Set6</th>
</tr>
</thead>
<tbody>
<tr>
<td>CQGARCH</td>
<td>5.0353</td>
<td>2.0509</td>
<td>15.0719</td>
<td>5.6978</td>
<td>64.8572</td>
<td>34.1092</td>
</tr>
<tr>
<td></td>
<td>(0.3874)</td>
<td>(0.1428)</td>
<td>(1.4125)</td>
<td>(0.8350)</td>
<td>(5.4403)</td>
<td>(3.5807)</td>
</tr>
<tr>
<td>QGARCH</td>
<td>4.8224</td>
<td>2.137</td>
<td>16.923</td>
<td>6.9666</td>
<td>75.9479</td>
<td>33.5711</td>
</tr>
<tr>
<td></td>
<td>(0.3723)</td>
<td>(0.1599)</td>
<td>(1.6444)</td>
<td>(1.0610)</td>
<td>(6.1501)</td>
<td>(3.6068)</td>
</tr>
<tr>
<td>CQAR</td>
<td>4.9287</td>
<td>2.0579</td>
<td>17.4951</td>
<td>7.5006</td>
<td>74.6232</td>
<td>35.6876</td>
</tr>
<tr>
<td></td>
<td>(0.3949)</td>
<td>(0.1420)</td>
<td>(1.5492)</td>
<td>(1.0030)</td>
<td>(5.9385)</td>
<td>(3.1848)</td>
</tr>
<tr>
<td>QAR</td>
<td>5.7896</td>
<td>2.1751</td>
<td>17.4387</td>
<td>8.5547</td>
<td>70.8524</td>
<td>39.7151</td>
</tr>
<tr>
<td></td>
<td>(0.6399)</td>
<td>(0.1597)</td>
<td>(1.6054)</td>
<td>(1.1226)</td>
<td>(5.7889)</td>
<td>(3.6900)</td>
</tr>
</tbody>
</table>

Table 3.5 gives the empirical sizes of the VQR tests. Note that both CAViaR and composite CAViaR (CCAViaR) specify the correct model, therefore we expect that the rejection rates to be close to the corresponding nominal levels 0.1 and 0.05. The test is a bit conservative especially at the tail quantile $\tau = 0.05$ and the test based on the CAViaR estimator is in general more conservative than that based on the CCAViaR estimator.

### 3.7 Data Application

In this section, we apply the proposed methods to the daily returns (100 times the difference of log of the close price) for General Motors (GM), IBM and the S&P 500; note that the same data was used in Engle and Manganelli (2004). The range of the data is from April 7, 1986, to April 7, 1999. The total number of observations is 3392, and we
Table 3.4: The 100×MSE of the composite and local estimations of the common parameter $\beta$. The 100× standard errors are reported in the parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Set1</th>
<th>Set2</th>
<th>Set3</th>
<th>Set4</th>
<th>Set5</th>
<th>Set6</th>
</tr>
</thead>
<tbody>
<tr>
<td>CQGARCH</td>
<td>5.7878 (1.7376)</td>
<td>1.3422 (0.2565)</td>
<td>34.5972 (9.0187)</td>
<td>8.9639 (4.5824)</td>
<td>58.3162 (13.3363)</td>
<td>52.5675 (13.0009)</td>
</tr>
<tr>
<td>QGARCH 0.05</td>
<td>9.0258 (3.0122)</td>
<td>1.3844 (0.2594)</td>
<td>64.2022 (14.1066)</td>
<td>19.8102 (10.4990)</td>
<td>71.4435 (15.7666)</td>
<td>58.5599 (16.3280)</td>
</tr>
<tr>
<td>QGARCH 0.1</td>
<td>5.6086 (1.9482)</td>
<td>2.1197 (0.6646)</td>
<td>53.5842 (11.8748)</td>
<td>24.2963 (10.3880)</td>
<td>79.5611 (14.3228)</td>
<td>48.385 (12.0294)</td>
</tr>
<tr>
<td>CQAR</td>
<td>8.6804 (4.3287)</td>
<td>1.3202 (0.2513)</td>
<td>58.4682 (12.3350)</td>
<td>26.2178 (7.8837)</td>
<td>90.2008 (13.4702)</td>
<td>62.6788 (12.5698)</td>
</tr>
<tr>
<td>QAR 0.05</td>
<td>14.3784 (5.4943)</td>
<td>1.3788 (0.2592)</td>
<td>75.8893 (15.0754)</td>
<td>37.6583 (11.5378)</td>
<td>93.0184 (16.6505)</td>
<td>86.4276 (17.8021)</td>
</tr>
<tr>
<td>QAR 0.1</td>
<td>17.9122 (6.7082)</td>
<td>2.1462 (0.6572)</td>
<td>70.2374 (13.8758)</td>
<td>38.4841 (12.2773)</td>
<td>85.5924 (14.6420)</td>
<td>72.7857 (14.7542)</td>
</tr>
</tbody>
</table>

Table 3.5: The size of VQR test at 5% and 10% quantiles. The nominal levels are 0.1 and 0.05.

<table>
<thead>
<tr>
<th>Nominal level</th>
<th>5% quantile</th>
<th>10% quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAViaR</td>
<td>CCAViaR</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0225 (0.0074)</td>
<td>0.0475 (0.0106)</td>
</tr>
<tr>
<td>0.05</td>
<td>0.0175 (0.0066)</td>
<td>0.04 (0.0098)</td>
</tr>
</tbody>
</table>

use the first 2892 returns to estimate the CAViaR model, and use the last 500 returns to calculate the out of sample Hit, which is defined as $I\{y_t < \hat{Q}_{t,\tau}\}$, $t = 2893, \ldots, 3392$.

We use the Symmetric Absolute Value model in (3.12) and select the initial values via QAR(5). We are interested in the 0.01-th and the 0.05-th quantiles. For the composite CAViaR model, we consider combining information at two quantile levels 0.01 and 0.05 and at five quantile levels 0.01, 0.02, 0.03, 0.04 and 0.05. The resulting estimators are referred to as CQAR1 and CQAR2, respectively.

Table 3.6 reports the p-values of the VQR test for different methods on the three data sets. The large p-values suggest that the Symmetric Absolute Value model provides an adequate fit to the data. Tables 3.7, 3.8 and 3.9 present the estimated parameters and corresponding standard errors, as well as the percentages of in-sample Hits and out-of-
sample Hits for four different methods. For all the three data sets, the composite estimator of $\beta$ has smaller standard error, indicating higher estimation efficiency. Take the S&P 500 data as an example, we highlight the standard errors of $\hat{\beta}$ in Table 3.9. Methods CQAR1 and CQAR2 give smaller standard errors than the local QAR estimators and Engle’s estimator. We observe that even combining information from two quantiles (CQAR1) can be beneficial. All the in-sample Hits are very close to the nominal quantile levels. This is reasonable because in-sample Hits are directly related to the quantile function to be minimized. The aggregation across quantile levels improves the efficiency of the estimation, thus benefits the predictability. Compared to the Engle’s original method, the composite methods CQAR1 and CQAR2 either have better out-of-sample Hits (at the 0.01-th quantile) or produce similar hits (at the 0.05-th quantile).

Table 3.6: The p-values of the VQR test for different methods on three data sets.

<table>
<thead>
<tr>
<th></th>
<th>1% quantile</th>
<th>5% quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QAR</td>
<td>CQAR1</td>
</tr>
<tr>
<td>General Moto</td>
<td>1.0000</td>
<td>0.9990</td>
</tr>
<tr>
<td>IBM</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Table 3.7: Local and composite CAViaR estimators for the General Motor data.

<table>
<thead>
<tr>
<th></th>
<th>1% quantile</th>
<th></th>
<th>5% quantile</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QAR</td>
<td>CQAR1</td>
<td>CQAR2</td>
<td>Engle</td>
</tr>
<tr>
<td>$\hat{\alpha}$</td>
<td>-0.467</td>
<td>-0.248</td>
<td>-0.248</td>
<td>-0.451</td>
</tr>
<tr>
<td></td>
<td>(0.269)</td>
<td>(0.154)</td>
<td>(0.171)</td>
<td>(0.203)</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>0.786</td>
<td>0.857</td>
<td>0.856</td>
<td>0.826</td>
</tr>
<tr>
<td></td>
<td>(0.093)</td>
<td>(0.038)</td>
<td>(0.043)</td>
<td>(0.083)</td>
</tr>
<tr>
<td>$\hat{\gamma}$</td>
<td>-0.350</td>
<td>-0.289</td>
<td>-0.293</td>
<td>-0.331</td>
</tr>
<tr>
<td></td>
<td>(0.165)</td>
<td>(0.089)</td>
<td>(0.086)</td>
<td>(0.169)</td>
</tr>
<tr>
<td>Hits.in(%)</td>
<td>0.968</td>
<td>1.003</td>
<td>0.968</td>
<td>1.003</td>
</tr>
<tr>
<td>Hits.out(%)</td>
<td>1.200</td>
<td>1.200</td>
<td>1.200</td>
<td>1.400</td>
</tr>
</tbody>
</table>

Table 3.8: Local and composite CAViaR estimators for the IBM data.

<table>
<thead>
<tr>
<th></th>
<th>1% quantile</th>
<th></th>
<th>5% quantile</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QAR</td>
<td>CQAR1</td>
<td>CQAR2</td>
<td>Engle</td>
</tr>
<tr>
<td>$\hat{\alpha}$</td>
<td>-0.727</td>
<td>-0.176</td>
<td>-0.121</td>
<td>-0.126</td>
</tr>
<tr>
<td></td>
<td>(0.653)</td>
<td>(0.183)</td>
<td>(0.125)</td>
<td>(0.093)</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>0.722</td>
<td>0.910</td>
<td>0.931</td>
<td>0.948</td>
</tr>
<tr>
<td></td>
<td>(0.163)</td>
<td>(0.026)</td>
<td>(0.010)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>$\hat{\gamma}$</td>
<td>-0.352</td>
<td>-0.171</td>
<td>-0.139</td>
<td>-0.113</td>
</tr>
<tr>
<td></td>
<td>(0.081)</td>
<td>(0.148)</td>
<td>(0.104)</td>
<td>(0.119)</td>
</tr>
<tr>
<td>Hits.in(%)</td>
<td>1.003</td>
<td>1.003</td>
<td>1.003</td>
<td>0.968</td>
</tr>
<tr>
<td>Hits.out(%)</td>
<td>1.800</td>
<td>1.400</td>
<td>1.600</td>
<td>1.600</td>
</tr>
</tbody>
</table>

Table 3.9: Local and composite CAViaR estimators for the S&P 500 data.

<table>
<thead>
<tr>
<th></th>
<th>1% quantile</th>
<th></th>
<th>5% quantile</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QAR</td>
<td>CQAR1</td>
<td>CQAR2</td>
<td>Engle</td>
</tr>
<tr>
<td>$\hat{\alpha}$</td>
<td>-0.009</td>
<td>-0.023</td>
<td>-0.037</td>
<td>-0.203</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.012)</td>
<td>(0.042)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>0.955</td>
<td>0.946</td>
<td>0.932</td>
<td>0.873</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.006)</td>
<td>(0.008)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>$\hat{\gamma}$</td>
<td>-0.150</td>
<td>-0.160</td>
<td>-0.188</td>
<td>-0.382</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.008)</td>
<td>(0.069)</td>
<td>(0.277)</td>
</tr>
<tr>
<td>Hits.in(%)</td>
<td>0.968</td>
<td>1.003</td>
<td>1.003</td>
<td>1.003</td>
</tr>
<tr>
<td>Hits.out(%)</td>
<td>1.200</td>
<td>1.200</td>
<td>1.400</td>
<td>1.800</td>
</tr>
</tbody>
</table>
3.8 Proof of Theorem 3.3.1

Before proving the Theorem 3.3.1, we first present all the assumptions and Lemma 3.8.1, which gives the consistency of $\hat{\theta}$.

**Consistency Assumptions**

C0. $(\Omega, F, P)$ is a complete probability space, and $\{y_t, x_t\}, t = 1, 2, \ldots,$ are random vectors on this space.

C1. For $k = 1, \ldots, K$ and $t = 1, 2, \ldots$, the function $Q_{t,\tau_k}(\theta)$ is such that for each $\theta \in \Theta^{K(q+1)+p}$, $Q_{t,\tau_k}(\theta)$ is measurable with respect to the information set $\Omega_t$ and $Q_{t,\tau_k}(\cdot)$ is continuous in $\Theta$ for a given choice of explanatory variables $\{y_{t-1}, x_{t-1}, \ldots, y_1, x_1\}$.

C2. Conditional on all of the past information $\Omega_t$, the returns $y_t$ form a stationary process, with continuous conditional density $f_t(y|\Omega_t)$.

C3. There exists $h > 0$ such that for all $t$ and $k$, $f_t\{Q_{t,\tau_k}(\theta)|\Omega_t\} \leq h$.

C4. $|Q_{t,\tau_k}(\theta)| < K(\Omega_t)$ for each $\theta \in \Theta$ and for all $t$ and $k$, where $K(\Omega_t)$ is some stochastic function of variables that belong to the information set, such that $E(|K(\Omega_T)|) \leq K_0 < \infty$, for some constant $K_0$.

C5. $E(|y_t - Q_{t,\tau_k}(\theta)|) < \infty$ for all $t$ and $k$.

C6. For $k = 1, \ldots, K$, $[\tau_k - I\{y_t < Q_{t,\tau_k}(\theta)\}]\{y_t - Q_{t,\tau_k}(\theta)\}$ obeys the uniform law of large numbers.

C7. For every $\epsilon > 0$, there exists a $\delta_k > 0$ such that if $||\theta - \theta_0|| \geq \epsilon$, then

$$\liminf_{T \to \infty} T^{-1} \sum_t P\{|Q_{t,\tau_k}(\theta) - Q_{t,\tau_k}(\theta_0)| > \delta_k\} > 0, \; k = 1, \ldots, K.$$  

**Asymptotic Assumptions**

AN1. $Q_{t,\tau_k}(\theta)$ is differentiable in $\Theta$, a compact subset of $\mathbb{R}^{K(q+1)+p}$, and for all $\theta$ and $\tilde{\theta}$ in a neighborhood of $\theta_0$, such that $||\theta - \tilde{\theta}|| \leq d$ for $d$ sufficiently small and for all $t$ and $k$: 

53
(a) $||\nabla Q_{t,\tau_k}(\theta)|| \leq F(\Omega_t)$, where $F(\Omega_t)$ is some stochastic function of variables that belong to the information set and $E\{F(\Omega_t)^3\} \leq F_0 \leq \infty$, for some constant $F_0$.

(b) $||\nabla Q_{t,\tau_k}(\theta) - \nabla Q_{t,\tau_k}(\tilde{\theta})|| \leq M(\Omega_t, \theta, \tilde{\theta}) = O(||\theta - \tilde{\theta}||)$, where $M(\Omega_t, \theta, \tilde{\theta})$ is some function such that $E\{M(\Omega_t, \theta, \tilde{\theta})\}^2 \leq M_0||\theta - \tilde{\theta}|| \leq \infty$ and $E\{M(\Omega_t, \theta, \tilde{\theta})\} \leq M_1||\theta - \tilde{\theta}|| \leq \infty$ for some constants $M_0$ and $M_1$.

AN2. $f_t(y|\Omega_T)$ is bounded uniformly for any $t$, and satisfies the Lipschitz condition $|f_t(y_1|\Omega_t) - f_t(y_2|\Omega_t)| \leq L|y_1 - y_2|$ for some constant $L \leq \infty$.

AN3. The matrices $A_T$ defined in Theorem 3.3.1 and $D_{T,k} \equiv E\left\{\frac{1}{T}\sum_{t=1}^T f_t(Q_{t,\tau_k}(\theta_{0k})|\Omega_t) \times \nabla' Q_{t,\tau_k}(\theta_{0k}) \nabla Q_{t,\tau_k}(\theta_{0k})\right\}$, $k = 1,\ldots,K$, have the smallest eigenvalues bounded below by a positive constant for $T$ sufficiently large.

AN4. The sequence $T^{-1/2}\sum_{t=1}^T \sum_{k=1}^K \sum_{t=1}^T [I\{y_t < Q_{t,\tau_k}(\theta)\} - \tau_k] \nabla' Q_{t,\tau_k}(\theta)$ obeys the central limit theorem.

Variance-Covariance Assumptions

VC1. For $k = 1,\ldots,K$, $\hat{c}_{T,k}/c_{T,k} \xrightarrow{p} 1$, where the nonstochastic positive sequence $c_{T,k}$ satisfies $c_{T,k} = o(1)$ and $c_{T,k}^{-1} = o(T^{1/2})$.

VC2. $E(|F(\Omega_t)|^4) \leq F_1 < \infty$ for all $t$ and for some constant $F_1$, where $F(\Omega_t)$ has been defined in assumption AN1(a).

VC3. $T^{-1}\sum_k \sum_{k'} \{\min(\tau_k \tau_{k'}) - \tau_k \tau_{k'}\} \left\{\sum_{t=1}^T \nabla' Q_{t,\tau_k}(\theta_0) \nabla Q_{t,\tau_{k'}}(\theta_0)\right\} - A_T \xrightarrow{p} 0$, and $T^{-1}\sum_k \sum_{t=1}^T f_t(Q_{t,\tau_k}(\theta)|\Omega_t) \nabla' Q_{t,\tau_k}(\theta_0) \nabla Q_{t,\tau_k}(\theta_0) - D_T \xrightarrow{p} 0$.

Lemma 3.8.1 Under Assumptions C0–C7 and Model (3.1), we have $\hat{\theta} \xrightarrow{p} \theta_0$.

We need to verify conditions of Corollary 5.12 of White (1996), specifically assumptions 3.1 and 3.2 in White (1996), since the others are easy to obtain. The consistency of CAViaR estimator at a single quantile level model has been proved (Theorem 1 in Engle and Manganelli (2004)); noting that our objective function in equation (3.4) is the summation of $K$ objective functions at each single quantile level, the proof of Lemma

54
3.8.1 goes verbatim following the arguments in Engle and Manganelli (2004) by using the linearity of expectation.

**Proof of Theorem 3.3.1**

Consider \( \tau_1, \tau_2, \ldots, \tau_K \), then the parameters are
\[
\theta = (\alpha_1', \ldots, \alpha_K', \beta'),
\]
where \( \alpha_k \) is a \( q \)-dimensional vector corresponding to \( \tau_k \) and \( \beta \) is a \( p \)-dimensional vector of common parameters. We first define the following notations:

\[
\phi_{t,k} = I\{y_t < Q_{t,k}(\hat{\theta})\} - \tau_k,
\]
\[
g_{t,k} = \phi_{t,k}(\theta)\nabla Q_{t,k}(\theta),
\]
\[
\lambda_{T,k}(\theta) = \frac{1}{T} \sum_{t=1}^{T} E\{g_{t,k}(\theta)\},
\]
\[
\mu_{t,k}(\theta, d) = \sup_{||\theta - \tilde{\theta}|| < d} ||g_{t,k}(\theta) - g_{t,k}(\tilde{\theta})||, \quad k = 1, \ldots, K.
\]

In order to apply the Huber’s Theorem (Theorem 3 in Huber (1967)), we first verify the following Huber’s conditions:

1. \( T^{-1/2} \sum_{k=1}^{K} \sum_{t=1}^{T} g_{t,k}(\hat{\theta}) = o_p(1) \).
2. \( \sum_{k=1}^{K} \lambda_{T,k}(\theta_0) = T^{-1} \sum_{k=1}^{K} \sum_{t=1}^{T} E\{g_{t,k}(\theta_0)\} = 0 \), where \( \theta_0 \) is the true vector of parameters.
3. (3i) \( || \sum_{k=1}^{K} \lambda_{T,k}(\theta)|| \geq a \cdot ||\theta - \theta_0|| \), for \( ||\theta - \theta_0|| \leq d_0 \);
   (3ii) \( E\{\mu_t(\theta, d)\} = E\left\{ \sup_{||\theta - \tilde{\theta}|| \leq d} || \sum_{k} g_{t,k}(\theta) - \sum_{k} g_{t,k}(\tilde{\theta}) || \right\} \leq b \cdot d \), for \( ||\theta - \theta_0|| + d \leq d_0 \);
   (3iii) \( E\{\mu_t(\theta, d)^2\} \leq c \cdot d \) for \( ||\theta - \theta_0|| + d \leq d_0 \).

For Condition 1, let \( \{e_j\}_{j=1}^{K(q+1)+p} \) be the standard basis of \( R^{K(q+1)+p} \), and

\[
L_j(a) = -\frac{1}{\sqrt{T}} \sum_{k} \sum_{t} \rho_{t,k} \{y_t - Q_{t,\tau_k}(\hat{\theta} + ae_j)\},
\]

where \( a \) is a scalar. Then the one-sided directional derivative of \( L_j(a) \) is given by

\[
G_j(a) = -\frac{1}{\sqrt{T}} \sum_{k} \sum_{t} \phi_{t,k}(\hat{\theta} + ae_j)\nabla Q_{t,\tau_k}(\hat{\theta} + ae_j).
\]
Since \( L_j(a) \) is continuous in \( a \) and achieves maximum at 0, it must be that in a neighborhood of 0, and for some \( \xi > 0 \),

\[
|G_j(0)| \leq G_j(-\xi) + \{-G_j(\xi)\}.
\]

Let \( \xi \to 0 \), then according to AN1(a) we obtain

\[
\left| \frac{1}{\sqrt{T}} \sum_{k} \sum_{t} g_{t,k}(\hat{\theta}) \right| = |G_j(0)| \leq o_p(1).
\]

Condition 2 is obtained by observing that

\[
E\{g_{t,k}(\theta_0)\} = E\{\phi_{t,k}(\theta_0) \nabla' f_{t,k}(\theta_0)\} = E\left[ E\{\phi_{t,k}(\theta_0) \nabla' f_{t,k}(\theta_0) | \Omega_t\} \right] = E\{\nabla' f_{t,k} E(\phi_{t,k}|\Omega_t)\} = 0.
\]

We next verify Condition 3(i). Denote \( \theta_1 = (\alpha'_1, \beta'_1)', ..., \theta_K = (\alpha'_K, \beta'_K)', \) and let \( \theta_k^* \) be their upscaling version, i.e., \( \theta_1^* = (\alpha'_1, 0', ..., \beta'), ..., \theta_K^* = (0', ..., \alpha'_K, \beta')' \). Similarly, define

\[
D_{T,k}^* = E \left\{ \frac{1}{T} \sum_{t=1}^{T} f_t(Q_{t,\tau_{t}}(\theta_{0k}) | \Omega_t) \nabla' Q_{t,\tau_{t}}(\theta_{0k}^*) \nabla Q_{t,\tau_{t}}(\theta_{0k}) \right\},
\]

\[
D_{T,k} = E \left\{ \frac{1}{T} \sum_{t=1}^{T} f_t(Q_{t,\tau_{t}}(\theta_{0k}) | \Omega_t) \nabla' Q_{t,\tau_{t}}(\theta_{0k}) \nabla Q_{t,\tau_{t}}(\theta_{0k}) \right\}, \quad k = 1, ..., K.
\]

Engle and Manganelli (2004) showed that for \( k = 1, ..., K \),

\[
\lambda_{T,k}(\theta_k) = D_{T,k}(\theta_k - \theta_{0k}) + O(||\theta_k - \theta_{0k}||^2).
\] \hspace{1cm} (3.13)

And by AN3, \( D_{T,k} \) is positive definite and the smallest eigen value is bounded away from 0. Therefore

\[
||\lambda_{T,k}(\theta_k)|| \geq a_k ||\theta_k - \theta_{0k}||.
\]
where $a_k$ is the smallest eigen value of $D_{T,k}$. Then we have

$$
\left\| \sum_{k=1}^{K} \lambda_{T,k}(\theta) \right\| = \left\| \sum_{k=1}^{K} \lambda_{T,k}(\theta^*_k) \right\|
$$

$$
\succeq \left\| \begin{pmatrix} D_{*,1}^* & \cdots & D_{*,K}^* \\ \vdots & \ddots & \vdots \\ D_{T,1} & \cdots & D_{T,K} \end{pmatrix} \begin{pmatrix} \theta_1^* - \theta_{01}^* \\ \vdots \\ \theta_K^* - \theta_{0K}^* \end{pmatrix} \right\|
$$

$$
= \left\| \begin{pmatrix} D_{T,1} & \cdots & D_{T,K} \\ \vdots & \ddots & \vdots \\ D_{T,1} & \cdots & D_{T,K} \end{pmatrix} \begin{pmatrix} \theta_1 - \theta_{01} \\ \vdots \\ \theta_K - \theta_{0K} \end{pmatrix} \right\|
$$

$$
\geq \min\{a_k\} \|\theta - \theta_0\|.
$$

For Condition (3ii), we first notice that

$$
\sup_{\|\theta - \theta_0\| \leq d} \left\| \sum_k g_{t,k}(\theta) - \sum_k g_{t,k}(\bar{\theta}) \right\| \leq \sup_{\|\theta - \theta_0\| \leq d} \sum_k \|g_{t,k}(\theta) - g_{t,k}(\bar{\theta})\|
$$

$$
\leq \sum_k \sup_{\|\theta - \theta_0\| \leq d} \|g_{t,k}(\theta) - g_{t,k}(\bar{\theta})\|
$$

$$
= \sum_k \sup_{\|\theta - \theta_0\| \leq d} \|g_{t,k}(\theta_k) - g_{t,k}(\bar{\theta}_k)\|.
$$

Since in Engle and Manganelli (2004),

$$
E \left\{ \sup_{\|\theta - \theta_0\| \leq d} \|g_{t,k}(\theta_k) - g_{t,k}(\bar{\theta}_k)\| \leq b_k \cdot d \right\},
$$

and in addition, $\|\theta - \theta_0\| + d \leq d_0$ implies $\|\theta_k - \theta_{0k}\| + d \leq d_0$, therefore there exists $b > 0$ such that

$$
E \left\{ \sup_{\|\theta - \theta_0\| \leq d} \left\| \sum_k g_{t,k}(\theta) - \sum_k g_{t,k}(\bar{\theta}) \right\| \leq b \cdot d, \right\}
$$

for $\|\theta - \theta_0\| + d \leq d_0$. 

57
For the last Condition 3(iii), we have
\[
    u_t(\theta, d)^2 = \left\{ \sup_{||\theta - \bar{\theta}|| \leq d} \left\| \sum_k g_t(k) - \sum_k g_t(k, \bar{\theta}) \right\| \right\}^2 \leq \left\{ \sum_k \sup_{||\theta - \bar{\theta}|| \leq d} \left\| g_t(k) - g_t(k, \bar{\theta}) \right\| \right\}^2.
\]
\[
    \leq K \sum_k \left\{ \sup_{||\theta - \bar{\theta}|| \leq d} \left\| g_t(k) - g_t(k, \bar{\theta}) \right\| \right\}^2 \quad \text{(Cauchy-Schwartz inequality)}.
\]

In Engle and Manganelli (2004), it was shown that for \( k = 1, ..K \),
\[
    E \left\{ \sup_{||\theta - \bar{\theta}|| \leq d} \left\| g_t(k) - g_t(k, \bar{\theta}) \right\| \right\} \leq c_k \cdot d.
\]

Therefore, we can find a constant \( c > 0 \) such that
\[
    E \{ \mu_t(\theta, d)^2 \} \leq c \cdot d \text{ for } ||\theta - \theta_0|| + d \leq d_0.
\]

Now that the Huber’s conditions are verified, by applying Huber’s Theorem, we have
\[
    \frac{1}{\sqrt{T}} \sum_{k=1}^K \sum_{t=1}^T g_t(k, \theta_0) + \sqrt{T} \sum_{k=1}^K \lambda_{T,k}(\bar{\theta}) = o_p(1). \quad (3.14)
\]

Recall that
\[
    A_T = \frac{1}{T} \sum_k \sum_{k'} \left\{ \min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'} \right\} E \left\{ \sum_{t=1}^T \nabla' Q_{t, \tau_k}(\theta) \nabla Q_{t, \tau_{k'}}(\theta) \right\},
\]
\[
    D_T = \frac{1}{T} \sum_k E \left[ \sum_{t=1}^T f_t(Q_{t, \tau_k}(\theta) | \Omega_t) \nabla' Q_{t, \tau_k}(\theta) \nabla Q_{t, \tau_k}(\theta) \right],
\]
where \( f_t(Q_{t, \tau_k}(\theta)) \) is the conditional density of \( y_t \) evaluated at \( Q_{t, \tau_k}(\theta_0) \). The result (3.13) implies that
\[
    \sum_{k=1}^K \lambda_{T,k}(\theta) = D_T(\theta - \theta_0) + O(||\theta - \theta_0||^2). \quad (3.15)
\]
By the consistency of $\hat{\theta}$ and by applying Slutsky’s theorem to (3.15), we get

$$D_T \sqrt{T} (\hat{\theta} - \theta_0) = \frac{1}{\sqrt{T}} \sum_{k=1}^{K} \sum_{t=1}^{T} \lambda_{T,k}(\hat{\theta}) + o_p(1), \quad (3.16)$$

which together with (3.14) gives

$$D_T \sqrt{T} (\hat{\theta} - \theta_0) = -\frac{1}{\sqrt{T}} \sum_{k=1}^{K} \sum_{t=1}^{T} g_{t,k}(\theta_0) + o_p(1).$$

By condition AN4, $T^{-1/2} \sum_{k=1}^{K} \sum_{t=1}^{T} g_{t,k}(\theta_0)$ obeys CLT, which implies the following result

$$A_T^{-1/2} D_T \sqrt{T} (\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I).$$
Chapter 4

Inference in Functional Linear Quantile Regression

The content in this chapter is based on the independent work of Kehui Wang through her collaboration with Meng Li, Dr. Ana-Maria Staicu and Dr. Arnab Maity.

4.1 Introduction

The improvement of modern measurement instruments allow people to obtain data that has a functional feature. Functional data is a realization, usually with measurement error, of an underlying random process. This sort of data takes functions (curves, surfaces or images) as the basic unit of observation, containing much more information than the traditional single point data, and thus is called functional data. The functions are often about time, but may also be about spatial location, wavelength, probability, etc. One functional data is observed at a number of points. Depending on the sampling design, the repeated measurements from individuals could be dense or sparse, and could be on regular or irregular time grids. For example, in Section 4.5, we analyzed the daily bike rental data. The covariate is the hourly bike rental count, which can be viewed as functional data. One interest of the study is to predict the total count of bike rental in the next day. This data motivates functional quantile regression, where the predictor is functional, and our interest is in describing the tail behavior of the response distribution.

Dimension reduction is a crucial step for analyzing infinite dimensional functional data. A common way to reduce the data dimension is using basis expansion. By projecting
the original functional data into the linear space spanned by the basis system, we are allowed to conduct the estimation and inference based on the basis coefficients. Functional principal component analysis (FPCA), which takes the eigenfunctions as basis system, is the most important and widely used analysis method in FDA (Ramsay and Silverman, 2005). The principal components pick out the strongest mode of variation among the curves, and provide better interpretation. Classical FPCA is only applicable to dense and regular functional data without measurement error. The shrinkage estimates improved FPCA by taking measurement errors into consideration (Yao et al., 2003). Reduced rank mixed effect model (James et al., 2000) and principal component analysis through conditional expectation (PACE) (Yao et al., 2005) can address sparse and irregular data. FPCA is also extendable to non-Gaussian data such as binary response (Hall et al., 2008) or pointwise skewed data (Staicu et al., 2012).

While quantile regression is appealing by allowing us to describe the entire distribution at various quantile levels, the investigation of quantile regression in the functional data context is limited. For functional linear quantile regression models (detailed in next section), Cardot et al. (2005) used smoothing splines to expand the functional covariates and establishes a convergence rate result; Kato (2012) studied PCA-based estimation and establishes a sharp convergence rate. Ferraty et al. (2005) and Chen and Müller (2012) estimated conditional quantile functions by inverting the corresponding distribution functions, which are referred as “indirect” methods. However, as far as we know, there exists no hypothesis testing methods in functional quantile regression.

In this chapter, we consider quantile regression for functional covariates and scalar responses, and develop the asymptotic results allowing inference for quantile regression. We establish the asymptotic distribution of the coefficient estimator for both one single quantile level and multiple quantile levels. The quantile regression on functional covariates is fundamentally different from the case of multi-covariates, because the predictors (the scores) are estimated rather than known. Different from the view of traditional measurement errors in predictors, which are typically independent with the true predictors or at least independent between errors across subjects, the functional data setting produces errors that are dependent with the true predictors and across subjects. We show that under this setting, the asymptotic distribution of the estimators are still unbiased but the variance is inflated. A Wald-type test to test the coefficient constancy across quantile levels is also considered. We show the inflated variance will not have an effect...
on the asymptotic null distribution of the test statistic. We conduct simulation studies to confirm the performance of the proposed Wald-type test. We also study the composite quantile regression when applying to a bike rental data.

The rest of this chapter is organized as follows. Section 4.2 demonstrates the estimation procedure and the proposed Wald-type test. Section 4.3 includes the main theorem regarding the asymptotic normality. The estimation of the number of principal components is discussed in Section 4.4. The simulation studies are presented in Section 4.4, and Section 4.5 applies the proposed test to a bike rental data. All proofs are deferred to Section 4.6.

4.2 Methodology

4.2.1 Hypothesis tests in functional quantile regression model

Suppose we observe data \((Y_i, W_{i,j}), i = 1, \ldots, n, j = 1, \ldots, m_i\), where \(Y_i\) is a scalar response variable, and \(W_{i,j}\) is the discrete realization of a latent and smooth process \(X(t)\) observed at \(t_{i,j} \in \mathcal{T}\) (a bounded closed interval). The observed functional covariate is perturbed by white noise \(w(t)\), i.e. \(W_{i,j} = X_i(t_{i,j}) + w(t_{i,j})\), where \(w(t_{i,j})\) has mean 0 and variance \(\sigma_w^2\). We assume that the true functional signal \(X(t) \in L^2(\mathcal{T})\), where \(L^2(\mathcal{T})\) is the \(L^2\) space on \(\mathcal{T}\). Without loss of generality, we use \(\mathcal{T} = [0, 1]\).

Let \(Q_Y(\tau|X)\) be the conditional \(\tau\)-th quantile function of a generic response \(Y\) given the true generic signal \(X(\cdot)\), where \(\tau \in \mathcal{U} \subset (0, 1)\) and \(\mathcal{U}\) is the quantile region of interest. Typically, the quantile region \(\mathcal{U}\) could be a single point (single quantile level), a finite set (multiple quantile levels) or a bounded closed interval. For each \(\tau \in \mathcal{U}\), we assume the following functional linear quantile regression (fQR) model:

\[
Q_{Y|X}(\tau) = \beta_0(\tau) + \int_0^1 \beta(t, \tau) X^c(t) dt,
\]

where for each \(\tau\), \(\beta_0(\tau) \in \mathbb{R}\) and \(\beta(\cdot, \tau) \in L^2[0, 1]\). Here \(X^c(t) = X(t) - E\{X(t)\}\).

Model (4.1) is an extension of the standard linear quantile regression model (Koenker, 2005) to functional covariates. It was formulated first by Cardot et al. (2005), and followed by many others (Chen and Müller, 2012; Kato, 2012).

Previous literature in functional quantile regression model has been focused on the prediction of \(Q_{Y|X}(\tau)\) with a given \(\tau\). Our goal is to test whether the covariate affects the
response differently at some specified quantile levels. This is an important question in its own right, but the test can also be helpful for developing more accurate estimation of the quantile functions. For example, in the case of vector predictor, there has been a lot of interests in developing more accurate estimation techniques. If the coefficient function in model (4.1) is a constant with respect to \( \tau \) within a quantile region \( \mathcal{U} \), we can aggregate information across \( \mathcal{U} \) to improve estimation efficiency via composite quantile regressions; see Koenker (1984); Zou and Yuan (2008); Zhao and Xiao (2014); Jiang et al. (2014); Wang and Wang (2015).

Let \( \mathbf{\tau} = \{\tau_1, \cdots, \tau_K\} \) be the quantile levels of interest, and \( \tau_1 < \cdots < \tau_K \in \mathcal{U} \). For example, we can choose \( \mathcal{U} \) to be a neighborhood quantile region when a specific quantile is of interest. Our goal is to test the null hypothesis:

\[
H_0 : \beta(\cdot, \tau_1) = \cdots = \beta(\cdot, \tau_K), \tag{4.2}
\]

against the alternative hypothesis

\[
H_a : \beta(\cdot, \tau_k) \neq \beta(\cdot, \tau_k'), \text{ for some } k \neq k' \in 1, \ldots, K.
\]

One approach to test the null hypothesis is to treat the discretely observed functional covariates as high dimensional covariates and apply traditional testing procedures for vector covariates. This approach, however, will result in large type I error rate, due to the fact that the functional covariate \( X_i \) is observed at discrete time points. Since the discrete observations are highly correlated, and also measured with errors, dimension reduction and the usage of the smoothness of the covariates become critical.

In our proposed method, we will take advantage of the functional principal component analysis (FPCA) technique to reduce the dimensionality of \( X(t) \) thus transforming the model in (4.1) and hypothesis test in (4.4) to some analogs in multi-covariate quantile regression.

Denote the mean function of \( X_i(t) \) as \( \mathbb{E}X_i(t) = \mu(t) \) and the covariance kernel function \( \text{Cov}\{X_i(s), X_i(t)\} = G(s, t) \). Mercer’s theorem gives the following spectral decomposition of the covariance \( G(s, t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s)\phi_j(t) \), where \( \{\phi_j, \lambda_j\} \) is the pair of eigenfunction and eigenvalue. The eigenvalues \( \lambda_j \)'s are nondecreasing and nonnegative and the eigenfunctions \( \phi_j \)'s are mutually orthogonal functions in \( L^2[0, 1] \), i.e. \( \int \phi_j(t)\phi_{j'}(t)dt = 0 \). Using the Karhunen-Loéve expansion, we can represent the ran-
dom process $X_i$ as $X_i(t) = \mu(t) + \sum_{j=1}^{\infty} z_{i,j} \phi_j(t)$, where $z_{i,j} = \int_0^1 \{X_i(t) - \mu(t)\} \phi_j(t) dt$. Notice that $\{z_{i,j} : j = 1, \cdots, \infty\}$ are functional principal component (FPC) scores of $X_i$, satisfying that $E(z_{i,j}) = 0$, $\text{Var}(z_{i,j}) = \lambda_j$ and uncorrelated over $j$.

We use the eigenfunction basis of $X_i$'s to represent the coefficient function $\beta(t, \tau)$ as $\beta(t, \tau) = \sum_{j=1}^{\infty} \beta_j(\tau) \phi_j(t)$. Therefore the original model (4.1) can be transformed to a regular quantile regression model but with infinite covariates:

$$Q_{Y_i|X_i}(\tau) = \beta_0(\tau) + \sum_{j=1}^{\infty} \beta_j(\tau) z_{i,j}.$$  (4.3)

Furthermore, the equality $\beta(\cdot, \tau_k) = \beta(\cdot, \tau_{k'})$ is equivalent to $\beta_j(\tau_k) = \beta_j(\tau_{k'})$, $j = 1, 2, \ldots$. Therefore, the null hypothesis (4.2) can be written as

$$H_0 : \begin{pmatrix} \beta_1(\tau_1) \\ \beta_2(\tau_1) \\ \vdots \\ \beta_p(\tau_1) \\ \beta_1(\tau_2) \\ \beta_2(\tau_2) \\ \vdots \\ \beta_p(\tau_2) \\ \vdots \\ \beta_1(\tau_K) \\ \beta_2(\tau_K) \\ \vdots \\ \beta_p(\tau_K) \end{pmatrix} = \begin{pmatrix} \beta_1(\tau_1) \\ \beta_2(\tau_1) \\ \vdots \\ \beta_p(\tau_1) \\ \beta_1(\tau_2) \\ \beta_2(\tau_2) \\ \vdots \\ \beta_p(\tau_2) \\ \vdots \\ \beta_1(\tau_K) \\ \beta_2(\tau_K) \\ \vdots \\ \beta_p(\tau_K) \end{pmatrix}$$  (4.4)

In practice, the functional covariate is observed at discrete time points, and some kind of truncation to the expansion $X(t)$ and $G(s,t)$ is widely used. For example, we can select the number of principal components by the percentage of variance, AIC-type criteria or cross-validation. Suppose we have $p$ principal components to select. Then the model (4.3) is reduced to

$$Q^{(p)}_{Y_i|X_i}(\tau) = \beta_0(\tau) + \sum_{j=1}^{p} \beta_j(\tau) z_{i,j}.$$  (4.3)'

The hypothesis (4.4) is reduced to

$$H_0 : \begin{pmatrix} \beta_1(\tau_1) \\ \beta_2(\tau_1) \\ \vdots \\ \beta_p(\tau_1) \\ \beta_1(\tau_2) \\ \beta_2(\tau_2) \\ \vdots \\ \beta_p(\tau_2) \\ \vdots \\ \beta_1(\tau_K) \\ \beta_2(\tau_K) \\ \vdots \\ \beta_p(\tau_K) \end{pmatrix} = \begin{pmatrix} \beta_1(\tau_1) \\ \beta_2(\tau_1) \\ \vdots \\ \beta_p(\tau_1) \\ \beta_1(\tau_2) \\ \beta_2(\tau_2) \\ \vdots \\ \beta_p(\tau_2) \\ \vdots \\ \beta_1(\tau_K) \\ \beta_2(\tau_K) \\ \vdots \\ \beta_p(\tau_K) \end{pmatrix}.$$  (4.5)

This is similar to quantile regression for vector covariates with two important differences: 1) the vector covariate, namely, the FPC scores are not directly observable; 2) the dimension of the vector covariate $p$ is usually unknown and has to be estimated.
Our proposed testing procedure is Wald-type based on the asymptotic distribution of the quantile estimators $\hat{\beta}_0(\tau_k), \ldots, \hat{\beta}_p(\tau_k)$, $k = 1, \ldots, K$. The proposed test statistic is presented in Section 4.2.3.

**4.2.2 Estimation procedure**

In the following we describe the quantile estimators. To illustrate ideas, we first assume the FPC scores are observed. At the end of the section we discuss the adjustments to account for the fact that the functional covariate is estimated instead. We detail the estimation of $\beta_j(\tau)$’s for each quantile level $\tau$ in this section. We first describe the estimation when the scores are known and then discuss the extension to the case that scores are estimated from the noisy and discrete functional covariates.

For convenience, let $z_{i,0} = 1$ and $z_i = (z_{i,0}, z_{i,1}, \ldots, z_{i,p})^T$, $i = 1, \ldots, n$, be the full vector of $z_{i,j}$’s for $j = 0, 1, \ldots, p$; also let $\theta(\tau) = \{\hat{\beta}_0(\tau), \hat{\beta}_1(\tau), \ldots, \hat{\beta}_p(\tau)\}^T$ be the vector of unknown parameters at the specified quantile level $\tau$.

When $z_i$’s are known, $\theta(\tau)$ can be estimated by the standard quantile regression technique, i.e.

$$\hat{\theta}_n(\tau) = \arg\min_{b \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \rho_\tau(y_i - z_i^T b),$$

(4.6)

where $\rho_\tau(x) = x\{\tau - I(x < 0)\}$ is the quantile loss function and $I(\cdot)$ is the indicator function.

This estimation procedure is not feasible since the scores $z_i$’s are not known. Therefore, we first estimate the scores using the standard FPCA techniques, say $\hat{z}_i^T = (1, \hat{z}_{i,1}, \ldots, \hat{z}_{i,p})$, where $\hat{z}_{i,j} = \int X(t)\hat{\phi}_j(t)dt$ and $\hat{\phi}_j(t)$ is the estimated eigenfunction based on fully observed $X_i$’s. For notational convenience, we let $\hat{z}_{i,0} = 1$. Then the estimators are

$$\hat{\theta}_n(\tau) = \arg\min_{b \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \rho_\tau(y_i - \hat{z}_i^T b).$$

(4.7)

However in practice the signal $X_i$ is rarely known, instead a noisy and discrete version of this is observed. For example, in our setting $\{W_{i,j} : j = 1, \ldots, m_i\}_{i=1}^{n}$ is observed. There exist extensive studies of how to obtain the estimates $\hat{X}_i$ and $\hat{\phi}_j$ in FPCA, including those that involve univariate smoothing for the covariate function and bivariate smoothing for the corresponding kernel; for example, see Yao et al. (2005); Hall and Hosseini-Nasab (2006); Hall et al. (2006); Zhang and Chen (2007); Li et al. (2013) and discussion therein.
Specifically, we follow the “smooth first, then perform estimation” procedure justified by Zhang and Chen (2007). We use a local linear regression with bandwidth $h_X$ to smooth each curve (observed at discrete time points with measurement errors), then use the sample mean $\hat{\mu}(\cdot)$ and sample covariance $\hat{G}(\cdot, \cdot)$ to estimate the mean curve and covariance kernel. Let $(\hat{\lambda}_k, \hat{\phi}_k)$ be the estimated eigenvalue/eigenfunction pairs by discretization of $\hat{G}(\cdot, \cdot)$. Finally the scores $\hat{z}_i$’s are estimated using numerical integral predictor for dense designs as in Section 4.2 of Li et al. (2010).

Note that the dependency of the estimated scores makes theoretical developments much more challenging. We shall detail and discuss more on the theoretical aspects in Section 4.3.

4.2.3 Wald-type test statistic

Consider $\tau_1, \ldots, \tau_K$ as $K$ quantiles of interest. Let $\zeta = \{\theta(\tau_1)^T, \ldots, \theta(\tau_K)^T\}^T$ denote the full parameters, and $\hat{\zeta}$ is the quantile regression estimator of $\zeta$. The null hypothesis 4.4 can be written equivalently as $H_0 : R\zeta = 0$. The constraint matrix $R = R_1 \otimes R_2$, where $\otimes$ is the Kronecker product. The matrix $R_1$ is a $(K-1) \times K$ matrix whose $(k, k)$th entry is -1 and $(k, k+1)$th entry is 1 for $k = 1, \ldots, K-1$. The matrix $R_2$ is a $p \times (p+1)$ matrix defined as $[0; I_p]$ where $I_p$ is the $p$-dimensional identity matrix. Using the analogy from testing in quantile regression with vector covariates, the Wald-type test statistic is

$$T_n = (R\hat{\zeta})^T [RV^*R]^{-1} (R\hat{\zeta}),$$

where $V^*$ is the $K(p+1) \times K(p+1)$ covariance matrix of $\hat{\zeta}$. The $(k, k')$th block of $V^*$ is the covariance matrix between $\hat{\theta}_n(\tau_k)$ and $\hat{\theta}_n(\tau_{k'})$, which will be defined in equation (4.10). The following section discusses the asymptotic null distribution of this test statistic.

4.3 Asymptotic results

4.3.1 Assumptions

Let $F_i(y) = P(Y_i < y|X_i(\cdot))$, and $f_i(\cdot)$ be the corresponding density function. We make the following assumptions:

A1. $\{Y_i, X_i(\cdot), w_i\}_{i=1}^n$ is i.i.d. with $\{Y, X(\cdot), w(\cdot)\}$ and $X(\cdot)$ is independent with the white
noise \( w(\cdot) \) where \( \text{E}_w(t) = 0 \) and \( \text{E}_w(t)^2 = \sigma_w^2 < \infty \) for any \( t \).

A2. The conditional distribution \( F_i(\cdot) \) is twice continuously differentiable and the corresponding density function \( f_i(\cdot) \) is uniformly bounded away from 0 and \( \infty \) at points \( Q_{Y_i|X_i}(\tau) \).

A3. The functional covariates \( X(\cdot) \) satisfy that \( \text{E}\{X(t)^4\} < \infty \) uniformly for \( t \).

A4. There exists a finite number \( p_0 \) such that \( \lambda_j = 0 \) if \( j \geq p_0 \).

The i.i.d. assumption in A1 and assumption on the conditional distribution in A2 are standard in quantile regression when the covariates are vectors; for example see Chapter 4 of Koenker (2005). A1 assumes that the functional covariates \( X_i(\cdot) \) is observed with independent measurement errors \( w_i(\cdot) \). The involvement of measurement errors makes the model more realistic, compared to error free assumptions made by Kato (2012). A3 holds for Gaussian processes and is common in FDA literature; for example, see Hall et al. (2006); Li et al. (2010). A4 means that the functional covariates \( X(\cdot) \) has a finite number of PC’s.

The following assumptions are commonly used to describe a dense sampling design (Zhang and Chen, 2007; Li et al., 2010). For convenient mathematical derivations, we assume that there are the same number of observations per subject, i.e. \( m_i = m \) for all \( i \).

B1. The time points \( t_{i,j} \overset{i.i.d.}{\sim} g(\cdot) \) for \( i = 1, \ldots, n; j = 1, \ldots, \), where the density \( g(\cdot) \) has bounded support \([0,1]\) and is continuously differentiable.

B2. \( m \geq C n^{c_m} \) where \( c_m > 5/4 \) and \( C \) is some constant.

For our theoretical development, we require the following condition for the kernel bandwidth \( h_X \).

C1. \( h_X = O(n^{-c_m/5}) \).

### 4.3.2 Asymptotic distributions

The following theorem gives the asymptotic distribution of the quantile estimator. Kato (2012) gave the minimax rate of the coefficient function estimation when there is no measurement error on the discrete functional covariates. The author assumed that the
number of eigenvalues is infinite instead of finite as in our Assumption A4. Assuming finite number of non-zero eigenvalues is critical since otherwise the quantile regression problem is ill-posed and the estimation of the coefficient function cannot achieve the root-

\( n \) rate. We denote \( D_0 \) as the diagonal matrix whose diagonal entries are \((1, \lambda_1, \cdots, \lambda_{p_0})\) and \( D_1(\tau) = E[f_i\{T_{Y_i|x_i}(\tau)\}z_iz_i^T] \) which is positive definite. We now assume that \( p_0 \) is known and shall defer the discussion of estimating \( p_0 \) to Section 4.3.3.

**Theorem 4.3.1** Let \( \Sigma_0^+ = (\sigma_{j,k}) \) be a \( p_0 \times p_0 \) matrix where \( \sigma_{j,j} = \sum_{j' \neq j} (\lambda_j - \lambda_{j'})^{-2} \lambda_{j'j'} \) and \( \sigma_{j,j'} = -(\lambda_j - \lambda_{j'})^{-2} \lambda_{j'j}, \ j \neq j' \). Let \( \Sigma_0 \) be a \((p_0 + 1) \times (p_0 + 1)\) matrix whose bottom right block is \( \Sigma_0^+ \) and other entries are zeros. Then under conditions A1–A4, B1–B2 and C1, we have the following asymptotic normality for the estimator defined in equation (4.7):

\[
\sqrt{n}\{ \hat{\theta}(\tau) - \theta(\tau) \} \xrightarrow{d} N \left\{ 0, \tau(1 - \tau)D_1^{-1}(\tau)D_0D_1^{-1}(\tau) + \Theta(\tau)\Sigma_0(\tau) \right\}, \quad (4.9)
\]

where \( \Theta(\tau) \) is a diagonal matrix with \( \theta(\tau) \) being the diagonal entries. The asymptotic covariance matrix for \( \theta(\tau_k), \theta(\tau_{k'}) \) for \( 1 \leq k, k' \leq K \) is

\[
A_{\text{cov}} \left[ \sqrt{n}\{ \hat{\theta}(\tau_k) - \theta(\tau_k) \}, \sqrt{n}\{ \hat{\theta}(\tau_{k'}) - \theta(\tau_{k'}) \} \right] = [\min(\tau_k, \tau_{k'}) - \tau_k\tau_{k'}]D_1^{-1}(\tau_k)D_0D_1^{-1}(\tau_{k'}) + \Sigma(\tau_k, \tau_{k'}), \quad (4.10)
\]

where \( \Sigma(\tau_k, \tau_{k'}) = \Theta(\tau_k)\Sigma_0(\tau_{k'}) \).

When the true scores \( z_i \)'s are observed, then the asymptotic variance of \( \hat{\theta}(\tau) \) is \( \tau(1 - \tau)D_1^{-1}(\tau)D_0D_1^{-1}(\tau) \), and the asymptotic covariance matrix for \( \theta(\tau_k), \theta(\tau_{k'}) \) is the same as in (4.10) but without the term \( \Sigma(\tau_k, \tau_{k'}) \) (Pollard, 1991; Koenker, 2005). Therefore, Theorem 4.3.1 explicitly shows the effect of the uncertainty in the covariates to the estimate. It indicates that the asymptotic distribution of \( \hat{\theta}(\tau) \) is unbiased but the variance is inflated for functional data. Furthermore, the estimated scores from FDA techniques are not independent across subjects, making the proof of Theorem 4.3.1 much more challenging. The proof of Theorem 4.3.1 is presented in Section 4.6. The arguments rely on the observation that \( \hat{z}_i \)'s are identically distributed across \( i \) and can be approximated by linear combinations of random vectors that are independent across subjects.

Section 4.2.3 discussed the Wald-type test in equation (4.8) for testing the hypothesis in equation (4.5) or equivalently \( R\zeta = 0 \). One may expect that the inflated term in
covariance matrix in equation (4.10) is still present in the covariance matrix $V^*$. For a general linear hypothesis, one should use this inflated covariance matrix $V^*$. However, we surprisingly find out that for testing the common slope, the inflation term will be canceled out. Let the “un-inflated” matrix $V$ be the counterpart of $V^*$ by letting $\Sigma(\tau_k, \tau_{k'}) = 0$, it can be shown that $RV^*R = RV R$. Therefore, we consider the following test statistic

$$T_n = (R\hat{\zeta})^T [RV R]^{-1} (R\hat{\zeta}),$$

(4.11)

where $V$ is the $K(p_0 + 1) \times K(p_0 + 1)$ matrix with the $(k, k')$th block defined by

$$[\min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'}]D^{-1}_1(\tau_k)D_1^{-1}(\tau_{k'}).$$

Then as a consequence of Theorem 4.3.1, we have the following result.

**Theorem 4.3.2** Assume that $R\zeta = 0$ and conditions A1–A4, B1–B2 and C1 hold, then the asymptotic distribution of $T_n$ is $\chi^2_{p_0}$.

Theorem 4.3.2 indicates that even though the uncertainty in the estimated scores inflate the estimation, it makes no difference for a Wald-type test for testing the hypothesis of common slopes. In practice, we estimate $D_0$ by $\hat{D}_0 = \sum_{i=1}^n \hat{z}_i \hat{z}_i^T / n$ and estimate $D_1(\tau)$ by $\hat{D}_1(\tau) = \sum_{i=1}^n \hat{f}_i \{Q(\tau | \hat{z}_i)\} \hat{z}_i \hat{z}_i^T / n$, which are both consistent estimators by the law of large numbers and the closedness between $\hat{z}_i$ and $z_i$ presented in Lemma 4.6.1.

### 4.3.3 Estimation of $p_0$

The number of principal components $p_0$ is unknown and should be selected in practice. The selection of $p_0$ has been studied intensively in the literature such as the criterion of percentage of variance explained (PVE). Recently Li et al. (2013) proposed a Bayesian information criterion (BIC) based on marginal modeling that can consistently select the number of principal components for both sparse and dense functional data. The method was shown to be consistent, i.e. the resulting estimator $\hat{p}_n$ converges to $p_0$ in probability. Using this result, we can show that the test statistic $T_n$ with $p_0$ replaced by the consistent estimator $\hat{p}_0$ still follows $\chi^2_{\hat{p}_n}$ asymptotically under $H_0: R\zeta = 0$. 


4.4 Simulation

We generate the data from the following heteroscedastic model:

\[ Y = \int X^c(t)dt + \{1 + \gamma \int X^c(t)t^2dt\}\epsilon, \quad \epsilon \sim N(0, \sigma^2_\epsilon = 1). \]

which leads to the quantile regression model:

\[ Q_\tau(X) = \Phi^{-1}(\tau) + \int_0^1 X^c(t)[t + \gamma t^2\Phi^{-1}(\tau)]dt. \quad (4.12) \]

We generate \( X(t) \) by the Karhunen-Loève expansion:

\[ X(t) = \mu(t) + \sum_{j=1}^{\infty} \xi_j \phi_j(t). \]

Instead of observing \( X(\cdot) \), we assume the covariates have measurement errors, i.e., we observe that \( W(t) = X(t) + w(t) \) at 100 equal-spaced time points in [0, 1], where \( w(t) \) is the i.i.d. white noise. Here the eigenfunctions \( \phi_k(\cdot) \) are orthonormal Legendre polynomials on [0, 1]:

\[
\begin{align*}
\phi_1(t) &= \sqrt{3}(2t^2 - 1), \\
\phi_2(t) &= \sqrt{5}(6t^2 - 6t + 1), \\
\phi_3(t) &= \sqrt{7}(20t^3 - 30t^2 + 12t - 1).
\end{align*}
\]

The eigenvalues are set as \( \lambda_k = (1/2)^{k-1} \) for \( k = 1, 2, 3 \), and \( \lambda_k = 0 \) for \( k \geq 4 \). We use the zero mean function \( \mu(t) = 0 \), and the independent white noise process \( w(t) \sim N(0, \sigma^2_w) \). It is clear that model (4.12) is a linear functional quantile model as in model 4.1 with \( \beta_0(\tau) = \Phi^{-1}(\tau), \) and \( \beta(t, \tau) = t + \gamma t^2\Phi^{-1}(\tau). \) Here the scalar \( \gamma \) controls the heteroscedasticity and determines how the coefficient function \( \beta(\cdot, \tau)(\tau \in \mathcal{U}) \) varies across \( \tau \). When \( \gamma = 0 \), the coefficient function \( \beta(\cdot, \tau) \) is independent of \( \tau \), which leads to the case of \( H_0 \); with various values of \( \gamma \) (non-zero), we can investigate the power of the proposed Wald-type test.

In the simulation design, we have three factors: the sample size \( n \) with five levels (\( n = 100, 500, 1000, 2000, 5000 \)), the white noise standard deviations \( \sigma_w \) with two levels (\( \sigma_w = 0.05, 1 \)), and the specified quantiles \( \mathcal{U} \) with two levels (\( \mathcal{U}_1 = \{0.1, 0.2, 0.3, 0.4\} \) for one-sided quantile levels and \( \mathcal{U}_2 = \{0.1, 0.2, 0.6, 0.7\} \) for two-sided quantile levels). For each scenario, we apply the Wald-type test of specified sizes at selected values of \( \gamma \). In addition to the proposed functional Wald-type test referred to as FPCA, we consider three other test procedures to compare: 1) the Naive Multivariate Quantile Regression model (NMQR) which treats the observed covariates \( \{W_{i,1}, \ldots, W_{i,100}\} \) as high-dimensional vectors; 2). a Scalar Summary of the observed \( W(t) \), i.e. the mean and then conduct Quantile Regression (SSQR); 3). Multivariate Principal Component Analysis (MPCA). For the MPCA approach, we select the number of components such that it explains at least 95%
of the variance. For the methods of NMQR and MPCA, we use the exiting Wald-type test in multi-covariate quantile regression (Koenker, 2005, Chapter 3.2.3).

Tables 4.1 and 4.2 summarize the Type I error rates (when $\gamma = 0$) of FPCA corresponding to different significance levels $\alpha$. For the other three test procedures, the Type I error rates are presented in Table 4.3. All results are based on 5000 simulations. Table 4.3 indicates that the method of NMQR has poor performance. This is because NMQR does not conduct dimension reduction. Using highly correlated covariates leads to the singularity of the design matrix. Therefore the NMQR method generates many NA’s in the table, and does not preserve the Type I error rates around the nominal levels. The method of MPCA performs poorly when $\sigma_w$ is large, producing many NA’s in Table 4.1 and Table 4.2. That is because the MPCA method chooses a large number of PC’s in this case, and thus suffers from the singularity of the design matrix as in NMQR. For example, suppose the true eigenvalues of the covariance kernel are $(1, 1/2, 1/4)$. When $\sigma_w = 1$, MPCA will estimate $m$ eigenvalues as $(2, 3/2, 5/4, 1, 1, \ldots, 1)$. If we use 0.95 as the PVE level, MPCA turns out to select far more than 3 PC’s. When $\sigma_w$ is close to 0, MPCA will choose an appropriate number of PC’s. Similar to the FPCA method, MPCA gives reasonable Type I error rates in this case.

The hypothesis $H_0$ in equation (4.5) implies that any linear combination of the slope coefficients has the same effect across quantile levels, particularly for the mean summary as used in the method of SSQR. This explains why the method of SSQR preserves the Type I error rates reasonably close to the nominal levels.

We further investigate the powers of the tests based on FPCA and SSQR. Figure 4.1 plots the estimated powers of two methods based on 2000 simulation repetitions. We can see that the method of FPCA is much more powerful than SSQR. In Figure 4.1, the left column corresponds to $\gamma = 0.5$. When the coefficient function has a small deviation from $H_0$, the FPCA method achieves high power as sample size increases to 2000. On the contrary, the SSQR method fails to detect the small deviation from $H_0$ even with large sample size 5000. This is because the method SSQR only uses a scalar summary of the functional covariate while the method of FPCA recovers the scores based on the entire functional observation. The right column of Figure 4.1 gives the power result when the coefficient function has largest deviation from $H_0$. In this case, SSQR would be able to achieve high power at a large sample size 5000, but FPCA method has large powers at sample size less than 1000.
Table 4.1: Type I error rates of the functional Wald-type test at given significance levels \( \alpha \) when considering \( U_1 = \{0.1, 0.2, 0.3, 0.4\} \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \sigma_w )</th>
<th>( \alpha )</th>
<th>( \alpha )</th>
<th>( \alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.01</td>
<td>0.05</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.05</td>
<td>0.021</td>
<td>0.062</td>
<td>0.101</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>0.014</td>
<td>0.056</td>
<td>0.106</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>0.013</td>
<td>0.050</td>
<td>0.100</td>
</tr>
<tr>
<td>2000</td>
<td>0.05</td>
<td>0.012</td>
<td>0.053</td>
<td>0.108</td>
</tr>
<tr>
<td>5000</td>
<td>0.05</td>
<td>0.012</td>
<td>0.053</td>
<td>0.103</td>
</tr>
</tbody>
</table>

Table 4.2: Type I error rates of the functional Wald-type test at given significance levels \( \alpha \) when considering \( U_2 = \{0.1, 0.2, 0.6, 0.7\} \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \sigma_w )</th>
<th>( \alpha )</th>
<th>( \alpha )</th>
<th>( \alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.01</td>
<td>0.05</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.05</td>
<td>0.028</td>
<td>0.077</td>
<td>0.13</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>0.018</td>
<td>0.067</td>
<td>0.126</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>0.011</td>
<td>0.056</td>
<td>0.108</td>
</tr>
<tr>
<td>2000</td>
<td>0.05</td>
<td>0.012</td>
<td>0.059</td>
<td>0.108</td>
</tr>
<tr>
<td>5000</td>
<td>0.05</td>
<td>0.012</td>
<td>0.048</td>
<td>0.096</td>
</tr>
</tbody>
</table>

Table 4.3: Type I errors when using the method of the naive multivariate quantile regression (NMQR), the mean as the scalar summary (SSQR) and Multivariate PCA (MPCA). When one method returns error (due to singularity of the design matrix) in more than 20% replications, we report it as “NA”.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \sigma_w )</th>
<th>( u )</th>
<th>NMQR</th>
<th>SSQR</th>
<th>MPCA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.01</td>
<td>0.05</td>
<td>0.10</td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td>100</td>
<td>0.05</td>
<td>( U_1 )</td>
<td>NA</td>
<td>NA</td>
<td>0.009</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>( U_1 )</td>
<td>NA</td>
<td>NA</td>
<td>0.011</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>( U_1 )</td>
<td>NA</td>
<td>NA</td>
<td>0.009</td>
</tr>
<tr>
<td>2000</td>
<td>0.05</td>
<td>( U_1 )</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>5000</td>
<td>0.05</td>
<td>( U_1 )</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>100</td>
<td>0.05</td>
<td>( U_2 )</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>( U_2 )</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>( U_2 )</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>2000</td>
<td>0.05</td>
<td>( U_2 )</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>5000</td>
<td>0.05</td>
<td>( U_2 )</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>( U_1 )</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>( U_1 )</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>1000</td>
<td>1</td>
<td>( U_1 )</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>2000</td>
<td>1</td>
<td>( U_1 )</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>5000</td>
<td>1</td>
<td>( U_1 )</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>( U_2 )</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>( U_2 )</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>1000</td>
<td>1</td>
<td>( U_2 )</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>2000</td>
<td>1</td>
<td>( U_2 )</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>5000</td>
<td>1</td>
<td>( U_2 )</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Figure 4.1: Powers of FPCA and SSQR in various scenarios.
4.5 Application

In this section, we apply the proposed estimating and testing procedure to a bike rental data set. The bike data (Fanaee-T and Gama, 2014) is based on the two-year historical log (the years of 2011 and 2012) from Capital Bikeshare System (CBS), Washington D.C., USA, which is available at http://capitalbikeshare.com/system-data. The data contains aggregated hourly rented bikes for regular users with membership and casual users. As the new generation of bike rentals, bike sharing systems possess membership, rental and return automatically. With currently over 500 bike-share programs around the world (Larsen, 2013) and the fast growing trend, data analysis on these systems regarding the effects to public traffic and the environment has become popular. In particular, it is important to predict the total count of bike rental in a day for the company to arrange the facility and the related agencies to control the public traffic. We use the hourly bike rental counts as the functional covariates, and predict higher 0.9-th quantiles of the next-day’s total count. All discussions are restricted to casual users since the randomness comes more from casual users which are important to analyze, while the registered users typically have strong patterns.

Figure 4.2 plotted the data set containing 731 days of records. We apply the proposed method to estimate the coefficient functions at various quantile levels \( U = (0.8, 0.825, 0.85, 0.875, 0.9) \) as shown in Figure 4.3. The corresponding Wald-type test leads to a \( p \)-value = 0.466, which suggests that the quantile coefficients are not significantly different across the quantile levels.

When some quantile coefficients are constant, we may improve the estimator’s efficiency by borrowing information from neighboring quantiles to estimate the common coefficients. We consider combined quantile regression at \( U \) by using the methods of quantile average estimator (QAE) and composite regression of quantiles (CRQ) with equal weights; see Chapter 2 for more technical details. By using 99% variance explained, we select 3 PC’s therefore the coefficient vectors have 4 elements including the intercept and 3 slopes, denoted as \((\beta_0, \beta_1, \beta_2, \beta_3)\). The number of PC’s is consistently selected as 3 when we use 1000 bootstrap samples. Table 4.4 reports the bootstrap standard errors of the estimates of the 4 coefficients when using QAE, CRQ and the single quantile regression estimation at the 0.9-th quantile (RQ). The QAE and CRQ estimators have smaller standard errors, indicating the efficiency improvement.

Furthermore, we conduct a cross-validation by randomly selecting 50% of the data
Figure 4.2: Hourly bike rentals for casual users. The $x$-axis is the hour and the $y$-axis is the hourly total count of bike rentals for casual users.

Figure 4.3: $\hat{\beta}(t)$ at various quantile levels.
as the training data set and using the other half as the testing data set. We use 1000 replications and for each replication, we calculate the prediction error as follows:

$$\text{PE} = \sum_i \rho_r(y_i - \hat{z}_i^T \hat{\beta}),$$

(4.13)

where the estimated coefficients $\hat{\beta}$ are based on the training data and the summation is over the test data. The mean prediction errors based on 1000 replications are reported in Table 4.5. We can see that the application of QAE and CRQ improves the prediction significantly for the 0.875-th and 0.9-th quantiles; differences among the three methods are not significant at the lower quantiles.

Table 4.4: Bootstrap standard errors of the estimates of 4 coefficients including the intercept $\beta_0$ from the QAE, CRQ and the local quantile regression estimation at the 0.9-th quantile (RQ). The method of FPCA selects 3 PC’s for all 1000 bootstrap samples.

<table>
<thead>
<tr>
<th></th>
<th>QAE</th>
<th>CRQ</th>
<th>RQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>-3.5478</td>
<td>-3.493</td>
<td>-3.867</td>
</tr>
<tr>
<td></td>
<td>(0.2963)</td>
<td>(0.2576)</td>
<td>(0.5446)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>5.7056</td>
<td>5.9869</td>
<td>5.056</td>
</tr>
<tr>
<td></td>
<td>(1.0529)</td>
<td>(1.0158)</td>
<td>(1.4642)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>5.0734</td>
<td>5.0083</td>
<td>5.9553</td>
</tr>
<tr>
<td></td>
<td>(2.5792)</td>
<td>(2.5561)</td>
<td>(3.4227)</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-14.1971</td>
<td>-14.9916</td>
<td>-18.2351</td>
</tr>
<tr>
<td></td>
<td>(3.9551)</td>
<td>(3.9144)</td>
<td>(4.4743)</td>
</tr>
</tbody>
</table>
Table 4.5: Mean prediction errors from different methods over 1000 cross-validations.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>QAE</th>
<th>CRQ</th>
<th>RQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>154.1625</td>
<td>153.0729</td>
<td>152.3959</td>
</tr>
<tr>
<td></td>
<td>(0.2772)</td>
<td>(0.2604)</td>
<td>(0.2720)</td>
</tr>
<tr>
<td>0.825</td>
<td>146.1629</td>
<td>145.5977</td>
<td>145.5041</td>
</tr>
<tr>
<td></td>
<td>(0.2679)</td>
<td>(0.2603)</td>
<td>(0.2682)</td>
</tr>
<tr>
<td>0.85</td>
<td>137.0283</td>
<td>136.7577</td>
<td>137.0707</td>
</tr>
<tr>
<td></td>
<td>(0.2555)</td>
<td>(0.2591)</td>
<td>(0.2577)</td>
</tr>
<tr>
<td>0.875</td>
<td>126.1382</td>
<td>125.9492</td>
<td>126.8194</td>
</tr>
<tr>
<td></td>
<td>(0.2386)</td>
<td>(0.2519)</td>
<td>(0.2400)</td>
</tr>
<tr>
<td>0.9</td>
<td>112.7738</td>
<td>112.842</td>
<td>113.8227</td>
</tr>
<tr>
<td></td>
<td>(0.2185)</td>
<td>(0.2379)</td>
<td>(0.2233)</td>
</tr>
</tbody>
</table>

4.6 Proofs

In this section, we sketch the proof for the main Theorem 4.3.1 in Section 4.6.1 and then prove all lemmas in Section 4.6.2. The notation $||\cdot||$ is the $L^2$-norm for a function or the Euclidean norm for a vector.

4.6.1 Proof of Theorem 4.3.1

The proof of Theorem 4.3.1 is composed of three steps. In step 1, we approximate the estimated scores $\hat{z}_i$’s by linear combinations of $z_i$’s. In step 2, we obtain the asymptotic distribution of $\hat{\theta}(\tau)$ at a single quantile level. In step 3, we extend the results in step 2 to multiple quantile levels.

**Step 1. Approximation of the estimated scores.** Existing literature has provided numerous error bound regarding the estimation of the eigenvalues and eigenfunctions; see for example Hall and Hosseini-Nasab (2006, 2009) and discussion therein. However, we here need to investigate the behavior of the estimated scores which are the covariates with measurement error in quantile regression.

**Lemma 4.6.1** Under assumptions B1, B2 and C1, we have

$$E||\hat{z}_i - z_i||^2 = o(n^{-1/2}).$$ \hspace{1cm} (4.14)
In addition,
\[ \hat{z}_i - z_i = n^{-1/2}B z_i + O_p(n^{-1}), \]  
(4.15)
where \(B\) is a \((p_0 + 1) \times (p_0 + 1)\) zero matrix but the bottom right block is replaced by \(B^+\), and \(B^+ = (b_{j,k})\) is a \(p_0 \times p_0\) random matrix such that \(b_{j,k} = 0\) if \(j = k\) and \(b_{j,j'} = n^{-1/2}(\lambda_j - \lambda_{j'})^{-1} (\sum_{i=1}^n z_{i,j} z_{i,j'})\) if \(j \neq j'\).

Equation (4.15) is a convenient bound for later usage (although not sharp). Equation 4.15 shows that the main part of \(\hat{z}_i - z_i\) is \(n^{-1/2}B z_i\), which is a linear combination of \(z_i\) and the random matrix \(B\) does not depend on \(i\). The proof of this lemma is in Section 4.6.2.

Step 2. Quantile regression on estimated scores. For any \(\delta \in \mathbb{R}^{p_0+1}\), let
\[ Z_n(\delta) = \sum_{i=1}^n \{ \rho_\tau(\hat{u}_i - \hat{z}_i^T \delta / \sqrt{n}) - \rho_\tau(\hat{u}_i) \}, \]  
(4.16)
where \(\hat{u}_i = y_i - \hat{z}_i^T \theta(\tau)\). Then \(Z_n(\delta)\) is convex and minimized at \(\hat{\delta}_n = \sqrt{n} \{ \hat{\theta}_n(\tau) - \theta(\tau) \}\). Therefore, the asymptotic distribution of \(\hat{\delta}_n\) is determined by the limiting behavior of \(Z_n(\delta)\) (Pollard, 1991). Let \(\psi_\tau(t) = \tau - I(t < 0)\), then according to the Knight’s identity (Knight, 1998), we can decompose \(Z_n(\delta)\) into two parts: \(Z_n(\delta) = Z_{1n}(\delta) + Z_{2n}(\delta)\), where
\[ Z_{1n}(\delta) = -\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{z}_i^T \delta \psi_\tau(\hat{u}_i) \]  
(4.17)
\[ Z_{2n}(\delta) = \sum_{i=1}^n \int_0^\infty \{ I(\hat{u}_i \leq s) - I(\hat{u}_i \leq 0) \} ds = \sum_{i=1}^n Z_{2ni}(\delta). \]  
(4.18)
In order to show (4.9), it is sufficient to prove that
\[ Z_n(\delta) \xrightarrow{d} -\delta^T W(\tau) + \frac{1}{2} \delta^T D_1 \delta, \]  
(4.19)
where \(W(\tau) \sim N \{ 0, \tau(1 - \tau) D_0 + D_1(\tau) \Sigma_0(\tau) D_1(\tau) \}\). This is because equation (4.19) is a quadratic form of \(\delta\), therefore we can use the convexity lemma (Pollard, 1991) to obtain the asymptotic distribution of \(\hat{\theta}_n(\tau)\).

We shall derive the limiting distributions of \(Z_{1n}(\delta)\) and \(Z_{2n}(\delta)\) separately. Similarly
to equation (4.17), we define $Z^*_{1n}(\delta)$ using the true scores $z_i$:

$$Z^*_{1n}(\delta) = -\frac{1}{\sqrt{n}} \sum_i z_i^T \delta \psi_r(u_i), \quad (4.20)$$

where $u_i = y_i - Q_{Y_i|X_i}(\tau) = y_i - z_i^T \theta(\tau) = y_i - \sum_{j=0}^{p_0} z_{i,j} \theta_j(\tau)$.

By a direct application of CLT, we have the asymptotic distribution of $Z^*_{1n}(\delta)$. However, when the predictors are estimated with errors, the difference $Z_{1n}(\delta) - Z^*_{1n}(\delta)$ is not negligible. In fact, $Z_{1n}(\delta)$ has the following representation and asymptotic distribution:

**Lemma 4.6.2** Under assumptions B1, B2 and C1,

$$Z_{1n}(\delta) = \delta^T \left[-\frac{1}{\sqrt{n}} \sum_i \{z_i \psi_r(u_i) - D_1(\tau) d_i\} \right] + o_p(1), \quad (4.21)$$

where $d_i = (0, d_{i,1}, \ldots, d_{i,p_0})^T$ and $d_{i,j} = \sum_{j' \neq j} (\lambda_j - \lambda_{j'})^{-1} z_{i,j} z_{i,j'} \theta_{j'}$, $j \geq 1$.

Since $z_i \psi_r(u_i) - D_1 d_i$ are i.i.d., Lemma 4.6.2 actually allows us to directly apply Linderberg’s CLT to obtain the asymptotic distribution of $Z_{1n}(\delta)$. Note that $E\{z_i \psi_r(u_i)\} = 0$ and $\text{Var}\{z_i \psi_r(u_i)\} = \tau(1 - \tau) D_0$. In addition, $Ed_i = 0$ because $z_{i,j}$ and $z_{i,j'}$ are uncorrelated and have mean 0. For the variance of $d_i$, simple calculation leads to

$$\text{Var}(d_{i,j}) = \sum_{k \neq j} (\lambda_j - \lambda_{j'})^{-2} \lambda_j \lambda_{j'} \theta^2_{j'}, \quad \text{Cov}(d_{i,j}, d_{i,j'}) = - (\lambda_j - \lambda_{j'})^{-2} \lambda_j \lambda_{j'} \theta_j \theta_{j'}, \quad (4.22)$$

and further

$$\text{Cov}\{z_i \psi_r(u_i), d_i\} = E \psi_r(u_i) z_i^T d_i = 0.$$ 

Let the matrix $\Sigma$ be the covariance matrix of $d_i$ given by equation (4.22), then

$$-\frac{1}{\sqrt{n}} \sum_i \{z_i \psi_r(u_i) - D_1 d_i\} \overset{d}{\to} N(0, \tau(1 - \tau) D_0 + D_1(\tau) \Sigma(\tau) D_1(\tau)).$$

Equivalently, let $W(\tau) \sim N(0, \tau(1 - \tau) D_0 + D_1(\tau) \Sigma(\tau) D_1(\tau))$, then $Z_{1n}(\delta) \overset{d}{\to} -\delta^T W(\tau)$. Equation (4.19) is then obtained by the following result for $Z_{2n}(\delta)$.

**Lemma 4.6.3** Under assumptions B1, B2 and C1,

$$Z_{2n}(\delta) = \frac{1}{2} \delta^T D_1(\tau) \delta + o_p(1).$$
Step 3. Asymptotic distributions across quantile levels. When considering various quantile levels, the same arguments can be made via a convex optimization and the limiting distribution of the objective function. The asymptotic covariance in equation (4.10) is obtained by the covariance between $z_i \psi_{\tau_k}(u_i) + D_1(\tau_k)d_i(\tau_k)$ and $z_i \psi_{\tau_{k'}}(u_i) + D_1(\tau_{k'})d_i(\tau_{k'})$.

4.6.2 Proofs of Lemmas

Proof of Lemma 4.6.1. Equation (4.14) is a result following standard bounds for the estimations of the eigenfunctions and covariance kernel in FDA literature. According to Theorem 1 in Hall and Hosseini-Nasab (2006), we have

$$||\hat{\phi}_j - \phi_j|| \leq 8^{1/2} s_j^{-1} ||\hat{G} - G||,$$

where $s_j = \min_{k \leq j}(\lambda_k - \lambda_{k+1})$ and $||\hat{G} - G|| = \left[\int_0^1 \int_0^1 (\hat{G}(u, v) - G(u, v))^2 du dv\right]^{1/2}$. Therefore,

$$|\tilde{z}_{i,j} - z_{i,j}| = \left|\int_0^1 X_i(t)(\hat{\phi}_j(t) - \phi_j(t))dt\right| \leq ||X_i|| \cdot ||\hat{\phi}_j - \phi_j|| \leq \text{constant} \cdot ||X_i|| s_j^{-1} ||\hat{G} - G||,$$

which leads to $||\tilde{z}_i - z_i|| \leq \text{constant} \cdot ||X_i|| s_j^{-1} ||\hat{G} - G||$. Therefore, for any $k > 0$,

$$E||\tilde{z}_i - z_i||^k \leq \text{constant} \cdot s_j^{-k} (E||\hat{G} - G||^{2k})^{1/2} \leq \text{constant} \cdot s_j^{-k} n^{-k/2},$$

by noting that $E||\hat{G} - G||^C \leq \text{constant} \cdot n^{-C/2}$ for any $C > 0$ (Hall and Hosseini-Nasab, 2009, Lemma 3.3). Thus for finite $p_0$, we have $E||\tilde{z}_i - z_i||^k = o(n^{-k/4})$; in particular, $\sqrt{n} E||\tilde{z}_j^T - z_j^T||^2 = o(1)$.

Next we prove equation (4.15). Let $\tilde{G}$ be the estimate of the kernel $G$ based on the fully observed covariate $X_i(\cdot)$, and recall that $\hat{G}$ is the estimate based on the discretized $W_{i,j}$ with measurement error. Denote $\tilde{Z} = \sqrt{n}(\tilde{G} - G)$ and $\hat{Z} = \sqrt{n}(\hat{G} - G)$. We use the notation $\int \hat{Z} \phi_j \phi_k$ to denote $\int_0^1 \int_0^1 \hat{Z}(u, v)\phi_j(u)\phi_k(v)du dv$.

Since $\{\phi_k : k = 1, \ldots, \infty\}$ forms a basis of the $L^2$ space on $[0, 1]$, we have $\hat{\phi}_j = \sum_{k=1}^\infty a_{j,k}\phi_k$, where $j = 1, \ldots, p_0$ and the generalized Fourier coefficients $a_{j,k} =$

80
\[ \int_0^1 \hat{\phi}_j(t)\hat{\phi}_k(t)dt. \] Furthermore, we have the following expansion for \( a_{j,k} \)’s:

\[ a_{jj} = 1 + O_p(n^{-1}); \quad a_{j,k} = n^{-1/2}(\lambda_j - \lambda_k)^{-1}\int \hat{Z}\phi_j\phi_k + O_p(n^{-1}) \text{ if } k \neq j, \]

according to equations (2.6) and (2.7) in Hall and Hosseini-Nasab (2006). Therefore, for \( j = 1, \ldots, p_0 \), we have

\[ \int_0^1 X_i(t)\{ \hat{\phi}_j(t) - \phi_j(t) \}dt = \sum_{k=1}^{p_0} \{a_{j,k} - I(k = j)\}z_{i,k} \]

\[ = \sum_{k \neq j, k \leq p_0} n^{-1/2}(\lambda_j - \lambda_k)^{-1}\int \hat{Z}\phi_j\phi_kz_{i,k} + O_p(n^{-1}). \]

A direct calculation gives that

\[ \int \hat{Z}\phi_j\phi_k = n^{-1/2}\sum_{i} z_{i,j}z_{i,k} - n^{1/2}\bar{z}_j\bar{z}_k \]

for \( j, k = 1, \ldots, p_0 \) and \( j \neq k \), where \( \bar{z}_j = n^{-1}\sum_{i=1}^{n} z_{i,j} \). Since \( n^{1/2}\bar{z}_j\bar{z}_k = n^{-1/2} \cdot (n^{1/2}\bar{z}_j) \cdot (n^{1/2}\bar{z}_k) = n^{-1/2} \cdot O_p(1) \cdot O_p(1) = O_p(n^{-1/2}) \), we have \( \int \hat{Z}\phi_j\phi_k = n^{-1/2}\sum_{i} z_{i,j}z_{i,k} + O_p(n^{-1/2}) \). The same approximation holds when using \( \hat{Z} \) since \( \hat{Z} - \bar{Z} \) is uniformly \( o_p(n^{-1/2}) \) as shown by Zhang and Chen (2007). Consequently,

\[ \int_0^1 X_i(t)\{ \hat{\phi}_j(t) - \phi_j(t) \}dt = n^{-1}(\lambda_j - \lambda_k)^{-1}\left( \sum_{i=1}^{n} z_{i,j}z_{i,k} \right)z_{i,k} + O_p(n^{-1}). \quad (4.23) \]

This approximation will not be affected if we use \( \hat{X}_i(\cdot) \) instead of the true curve \( X_i(\cdot) \) because the difference \( \hat{X}_i(\cdot) - X_i(\cdot) \) is negligible uniformly for all \( i \) (e.g., see Theorem 2 in Zhang and Chen (2007) or Lemma 1 in Zhu et al. (2014)). Let a \( p_0 \)-dimension random matrix \( B^+ = (b_{j,k}) \) where \( b_{j,k} = 0 \) if \( j = k \) and \( b_{j,k} = n^{-1/2}(\lambda_j - \lambda_k)^{-1} \left( \sum_{i=1}^{n} z_{i,j}z_{i,k} \right) \) if \( j \neq k \). Let \( B \) be a \((p_0 + 1) \times (p_0 + 1)\) zero matrix but the bottom right block is replaced by \( B^+ \), then the right hand side in equation (4.23) becomes \( n^{-1/2}Bz_i + O_p(n^{-1}) \). Consequently, we have

\[ \hat{z}_i - z_i = n^{-1/2}Bz_i + O_p(n^{-1}). \quad (4.24) \]

**Proof of Lemma 4.6.2.** We first decompose the difference between \( Z_{1n}(\delta) \) and \( Z_{1n}^*(\delta) \)
into three parts $S_1, S_2$ and $S_3$ as follows.

\[
Z_{1n}(\delta) - Z_{1n}^*(\delta) = - \frac{1}{\sqrt{n}} \sum_i \hat{z}_i^T \delta \psi_r(\hat{u}_i) + \frac{1}{\sqrt{n}} \sum_i z_i^T \delta \psi_r(u_i)
\]

\[
= - \frac{1}{\sqrt{n}} \sum_i (\hat{z}_i - z_i^T) \delta \{ \psi_r(\hat{u}_i) - \psi_r(u_i) \} - \frac{1}{\sqrt{n}} \sum_i (\hat{z}_i - z_i^T) \delta \psi_r(u_i) - \frac{1}{\sqrt{n}} \sum_i z_i^T \delta \{ \psi_r(\hat{u}_i) - \psi_r(u_i) \}
\]

\[=: S_1 + S_2 + S_3.\]

We then show that the first two terms are small, i.e. $S_2 = o_p(1)$ (Step i) and $S_1 = o_p(1)$ (Step ii). Step i and Step ii imply that the limiting distribution of $Z_{1n}(\delta)$ is the same as $S_3$, however, $S_3$ is still challenging to analyze since the function of $\psi_r(\cdot)$ is not differentiable. We shall approximate the term $S_3$ using properties of the function $\psi_r(\cdot)$ in Step iii. For the first two steps, it is sufficient to prove that $E(S_2^2) = o(1)$ and $E|S_1| = o(1)$ by applying Chebyshev’s inequality.

**Step i.** First notice that $E\{\psi_r(u_i)|z_i, \hat{z}_i\} = 0$ and $E\{\psi_r(u_i)^2|z_i, \hat{z}_i\} = \tau - \tau^2$. Therefore, we have $E(S_2) = 0$, and further

\[
E(S_2^2) = E \left\{ \frac{1}{\sqrt{n}} \sum_i (\hat{z}_i^T \delta - z_i^T \delta) \psi_r(u_i) \right\}^2
\]

\[= \frac{1}{n} \sum_j \sum_k E \left\{ (\hat{z}_j^T - z_j^T) \delta \psi_r(u_j) \cdot (\hat{z}_k^T - z_k^T) \delta \psi_r(u_k) \right\}.
\]

For $j = k$,

\[
E \left\{ (\hat{z}_j^T - z_j^T) \delta \psi_r(u_j) \cdot (\hat{z}_j^T - z_j^T) \delta \psi_r(u_j) \right\} = E \left\{ (\hat{z}_j^T \delta - z_j^T \delta) \psi_r(u_j) \right\}^2
\]

\[= E \left\{ (\hat{z}_j^T \delta - z_j^T \delta)^2 E\{\psi_r^2(u_j)|z_j, \hat{z}_j\} \right\} = \tau(1 - \tau) E \left\{ (\hat{z}_j^T \delta - z_j^T \delta)^2 \right\}.
\]

Since $\hat{z}_i$ is identically distributed for all $i$, we have $E \left\{ (\hat{z}_j^T \delta - z_j^T \delta)^2 \right\} =
\[ E \left\{ (\hat{z}_i^T \delta - z_i^T \delta)^2 \right\} \]. For \( j \neq k \), we have

\[ E \left\{ (\hat{z}_j^T - z_j^T) \delta \psi_r(u_j) \cdot (\hat{z}_k^T - z_k^T) \delta \psi_r(u_k) \right\} = 0 \]

since

\[ E \{ \psi_r(u_j) \psi_r(u_k) | z_j, \hat{z}_j, z_k, \hat{z}_k \} = E \{ \psi_r(u_j) | z_j, \hat{z}_j, z_k, \hat{z}_k \} \cdot E \{ \psi_r(u_k) | z_j, \hat{z}_j, z_k, \hat{z}_k \} = 0. \]

Therefore \( E(S_2^2) = \tau (1 - \tau) E \left\{ (\hat{z}_1^T \delta - z_1^T \delta)^2 \right\} = O(E||\hat{z}_i - z_i||^2) = o(1). \)

**Step ii.** For \( S_1 \), we first introduce the notation

\[ \Delta_i = E(\psi_r(\hat{u}_i) | z_i, \hat{z}_i) = \tau - F_i(\hat{z}_i^T \theta) = F_i(z_i^T \theta) - F_i(\hat{z}_i^T \theta). \]  

(4.25)

In addition, the random variable \( \Delta_i \) satisfy that

\[ \Delta_i = E(\psi_r(\hat{u}_i) - \psi_r(u_i) | z_i, \hat{z}_i), \]  

(4.26)

\[ |\Delta_i| = E(|\psi_r(\hat{u}_i) - \psi_r(u_i)| | z_i, \hat{z}_i). \]  

(4.27)

Equation (4.26) is obtained by noting that \( \psi_r(u_i) \) has mean 0 conditional on \( z_i \); equation (4.27) is obtained by noting that \( |\psi_r(\hat{u}_i) - \psi_r(u_i)| = I\{\min(\hat{z}_i \theta, z_i \theta) < y_i < \max(\hat{z}_i \theta, z_i \theta)\}\).

By Taylor’s theorem, for any \( a, b \in \mathbb{R} \), we have

\[ F(a + b) - F(a) = f(a)b + b^2 \int_0^1 f'(a + tb)(1 - t)dt =: f(a)b + \frac{b^2}{2} R(a, b), \]

where \( |R(a, b)| \leq C_0 \). Therefore,

\[ \Delta_i = - (\hat{z}_i^T \theta - z_i^T \theta) f_i(z_i^T \theta) + (\hat{z}_i^T \theta - z_i^T \theta)^2 R(\hat{z}_i^T, z_i^T), \]

where \( |R(\hat{z}_i^T, z_i^T)| \leq 2C_0 \). We also have the bound

\[ E\Delta_i^2 \leq \text{constant} \cdot E||\hat{z}_i - z_i||^2 = o(n^{-1/2}). \]
Therefore, $|S_1| \leq \frac{1}{\sqrt{n}} \sum_i |(\hat{z}_i^T \delta - z_i^T \delta)\cdot (\psi_\tau(\hat{u}_i) - \psi_\tau(u_i))|$ and consequently

$$E|S_1| \leq \frac{1}{\sqrt{n}} E\left\{ \sum_i |(\hat{z}_i^T \delta - z_i^T \delta)||\Delta_i| \right\}$$

$$= \sqrt{n}E|\hat{z}_i - z_i|^2 E\Delta_i^2 \leq o(1).$$

**Step iii.** Define

$$R_n(t) = \sum_{i=1}^n z_i \{\psi_\tau(u_i - z_i^T t) - \psi_\tau(u_i)\},$$

for any vector such that $||t|| \leq C$ for some constant $C$. Then the uniform approximation (He and Shao 1996) indicates that

$$\sup ||R_n(t) - E\{R_n(t)\}|| = O_p(n^{1/2}(\log n)||t||^{1/2}).$$

On the other hand,

$$E\{R_n(t)\} = \sum_i E[z_i \{F_i(z_i^T \theta) - F_i(z_i^T \theta - z_i^T t)\}] = nE[z_1 \{F_1(z_1^T \theta) - F_1(z_1^T \theta - z_1^T t)\}]$$

$$= -nEz_1z_1^T f_1(z_1^T \theta) t + O(n||z_1||^3||t||^2)$$

$$= -nD_1 t + O(n||t||^2).$$

Therefore,

$$R_n(t) = -nD_1 t + O(n||t||^2) + O_p\left(n^{1/2}(\log n)||t||^{1/2}\right). \quad (4.28)$$

Note that $\hat{u}_i = u_i + z_i^T \theta - \hat{z}_i^T \theta$ and $\hat{z}_i - z_i = Bz_i$ up to a negligible term $O_p(n^{-1})$, let $t = n^{-1/2}B\theta$, then

$$- \frac{1}{\sqrt{n}} \sum_i z_i^T \delta \{\psi_\tau(\hat{u}_i) - \psi_\tau(u_i)\} = -n^{-1/2}R_n(n^{-1/2}B\theta) + o_p(1), \quad (4.29)$$

where the term $o_p(1)$ is caused by the remainder term $\hat{z}_i - z_i - Bz_i = O_p(n^{-1})$ and obtained by using the same technique in Step ii via conditional expectation and Taylor theorem. Combining equation 4.28 and the fact that $||n^{-1/2}B\theta|| = O_p(n^{-1/2})$, we obtain
that $R_n(n^{-1/2} \mathbf{B\theta}) = -n^{1/2} \mathbf{D}_1 \mathbf{B\theta} + O(1) + O_p(n^{1/4} \log n)$. Therefore,

$$\frac{1}{\sqrt{n}} \sum_i z_i^T \delta \{ \psi_r(\hat{u}_i) - \psi_r(u_i) \} = \delta^T \mathbf{D}_1 \mathbf{B\theta} + o_p(1). \quad (4.30)$$

Consequently,

$$Z_{1n}(\delta) = \frac{1}{\sqrt{n}} \sum z_i^T \delta \psi_r(u_i) + \delta^T \mathbf{D}_1 \mathbf{B\theta} + o_p(1). \quad (4.31)$$

Let $d_i = (0, d_{i,1}, \ldots, d_{i,p_0})$ and $d_{i,j} = \sum_{k \neq j} (\lambda_j - \lambda_k)^{-1} z_{i,j} z_{i,k} \theta_k$, then the statement of Lemma 4.6.2 follows.

**Proof of Lemma 4.6.3.** Recall that $Z_{2n} = \sum_i Z_{2ni}$, where

$$Z_{2ni}(\delta) = \int_0^{\tilde{z}_i^T \delta / \sqrt{n}} \{ I(\hat{u}_i \leq s) - I(\hat{u}_i \leq 0) \} ds.$$

First, we have

$$E[Z_{2ni}(\delta) | z_i, \hat{z}_i] = \int_0^{\tilde{z}_i^T \delta / \sqrt{n}} F_i(\tilde{z}_i^T \theta + s) - F_i(\tilde{z}_i^T \theta) ds = \frac{1}{\sqrt{n}} \int_0^{\tilde{z}_i^T \delta} F_i(\tilde{z}_i^T \theta + t \sqrt{n}) - F_i(\tilde{z}_i^T \theta) dt. \quad (4.30)$$

Therefore, by Taylor’s theorem, we have

$$E[Z_{2ni}(\delta) | z_i, \hat{z}_i] = \frac{1}{\sqrt{n}} \int_0^{\tilde{z}_i^T \delta} f_i(\tilde{z}_i^T \theta) \frac{t}{\sqrt{n}} + \frac{t^2}{2n} R(\tilde{z}_i^T \delta, \frac{t}{\sqrt{n}}) dt = \frac{1}{2n} \delta^T \tilde{z}_i f_i(\tilde{z}_i^T \theta) \tilde{z}_i^T \delta + R_{n,i},$$

where $R_{n,i}$ is the remainder satisfying that $|R_{n,i}| \leq cn^{-3/2} |\tilde{z}_i^T \delta|^3$. Consequently,

$$E[Z_{2ni}(\delta) | z_i, \hat{z}_i] = \frac{1}{2} \cdot \frac{1}{n} \delta^T \tilde{z}_i f_i(\tilde{z}_i^T \theta) \tilde{z}_i^T \delta + R_{n,i}.$$

Therefore, the unconditional expectation of $Z_{2ni}(\delta)$ is

$$E[Z_{2ni}(\delta)] = E[E[Z_{2ni}(\delta) | z_i, \hat{z}_i]] = \frac{1}{2} \cdot \frac{1}{n} \delta^T \left( \frac{1}{n} \tilde{z}_i f_i(\tilde{z}_i^T \theta) \tilde{z}_i^T \right) \delta + ER_n = \frac{1}{2} \cdot \frac{1}{n} \delta^T \mathbf{D}_1 \delta + E(R_{n,i}),$$

leading to

$$E[Z_{2n}] = \frac{1}{2} \delta^T \mathbf{D}_1 \delta + \sum_{i=1}^n E(R_{n,i}).$$

85
The second term $\sum_{i=1}^{n} E(R_{n,i})$ is negligible because

$$|\sum_{i=1}^{n} E(R_{n,i})| \leq \sum_{i=1}^{n} E|R_{n,i}| \leq cn^{-3/2} \sum_{i=1}^{n} E|\hat{z}_{i}^{T} \delta|^{3} = cn^{-1/2} E|\hat{z}_{i}^{T} \delta|^{3} \leq O(n^{-1/2}) \cdot (E|\hat{z}_{1}|^{3}) \|\delta\|_{2}^{3} = O(n^{-1/2}) \cdot O(1) = o(1),$$

where the last step is due to the fact that $|\hat{z}_{1}|_{2} = ||\hat{X}_{1}||_{2} \leq ||\hat{X}_{1} - X_{1}||_{2} + ||X_{1}||_{2}$. We next will show that $\max_{i=1,...,n} ||z_{i}||/\sqrt{n} \overset{p}{\to} 0$. Note that $||z_{i}||^{2} = 1 + z_{i,1}^{2} + \ldots + z_{i,p_{0}}^{2}, i = 1, \ldots, n$, and $||z_{i}||^{2}$'s are i.i.d. with a finite second moment $E||z_{i}||^{2} = 1 + \lambda_{1} + \ldots + \lambda_{p_{0}} < \infty$. For any $\epsilon > 0$, we have

$$P\left(\max_{i=1,...,n} ||z_{i}||/\sqrt{n} > \epsilon\right) \leq \sum_{i=1}^{n} P(||z_{i}|| > \sqrt{n}\epsilon) \leq \frac{1}{n\epsilon^{2}} \sum_{i=1}^{n} E\{|z_{i}|^{2}I(||z_{i}|| > \sqrt{n}\epsilon)\} = \frac{1}{\epsilon^{2}} E\{|z_{1}|^{2}I(||z_{1}|| > \sqrt{n}\epsilon)\} \to 0,$$

according to the dominated convergence theorem. It implies that $\max_{i=1,...,n} ||\hat{z}_{i}||/\sqrt{n} = o_{p}(1)$ since $||\hat{z}_{i} - z_{i}|| = o_{p}(1)$ uniformly for all $i$'s. Consequently, $\text{Var}(Z_{2n}|z_{i}$'s, $\hat{z}_{i}$'s) $\leq \max_{i=1,...,n} ||\hat{z}_{i}^{T} \delta||/\sqrt{n} \cdot \sum_{i=1}^{n} E(Z_{2n}|z_{i}$'s, $\hat{z}_{i}$'s) = $o_{p}(1)$, i.e., the conditional variance converges to 0 in probability. Therefore, following the martingale argument in the proof of Theorem 2 in Pollard (1991), we have $Z_{2n} - E(Z_{2n}) = o_{p}(1)$, which completes the proof.
REFERENCES


Hogg, R. (1980). Personal communication with Koenker, R.


