UNIVERSITY OF CALIFORNIA
RIVERSIDE

The Second-order Bias and MSE of Quantile and Expectile Estimators

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Economics

by

He Wang

September 2018

Dissertation Committee:

Professor Aman Ullah, Co-Chairperson
Professor Tae-Hwy Lee, Co-Chairperson
Professor Goloria Gonzalez-Rivera

The Dissertation of He Wang is approved:

Co-Chairperson

Co-Chairperson

University of California, Riverside

Acknowledgments

First, I would like to thank my two committee chairs, Professor Aman Ullah and Professor Tae-Hwy Lee, for their support, inspiration, guidance, patience, and encouragement during all these years for my studies. They instructed me how to do research, and helped me develop teaching skills in both graduate and undergraduate statistics and econometrics courses. Moreover, They have always been an excellent example of how kind and hard working human beings could be. I am very grateful to my co-advisors, without whose help, I would not have been here.

I also would like to thank my other committee member and professors who helped me through these year. I have many thanks for Professor Gloria Gonzalez-Rivera for her invaluable suggestions on my research and her encourgement. I am also thankful for Professor Ruoyao Shi who always give me a lot of detailed comments for my presentations, and Professor Richard Arnott who gave me many helpful suggestions in preparing for job market.

I would also like to thank my parents who always understand, encourage and love me, my husband who is always there for me when I need him. I am also thankful for all my friends for their support and help all these years.

I couldn't be thankful enough for all my professors, family, and friends. They made my life so easy and made these years in UC Riverside the most wonderful time in my life. I will cherish such colorful and happy memories all the time.

To my parents,

who have always been my motivation to be better.

ABSTRACT OF THE DISSERTATION

The Second-order Bias and MSE of Quantile and Expectile Estimators

by

He Wang

Doctor of Philosophy, Graduate Program in Economics
University of California, Riverside, September 2018
Professor Aman Ullah, Co-Chairperson
Professor Tae-Hwy Lee, Co-Chairperson

This dissertation covers several topics in the second-order bias and mean squared error (MSE) of quantile and expectile estimators.

Chapter one presents the introduction of this dissertation. The finite sample theory using higher order asymptotics provides better approximations of the bias and MSE for a class of estimators. Rilstone, Srivastava and Ullah (1996) provided the second-order bias results of conditional mean regression. The goal of this dissertation is to develops analytical results on the second-order bias and MSE for quantile and expectile estimators.

Chapter two develops new analytical results on the second-order bias up to order $O\left(N^{-1}\right)$ and MSE up to order $O\left(N^{-2}\right)$ of the conditional quantile regression estimators. First, we provide the general results on the second-order bias and MSE of conditional quantile estimators. The second-order bias result enables an improved bias correction and thus to obtain improved quantile estimation. In particular, we show that the second-order bias are much larger towards the tails of the conditional density than near the median, and therefore the benefit of the second order bias correction is greater when we are interested

in the deeper tail quantiles, e.g., for the study of income distribution and financial risk management. The higher order MSE result for the quantile estimation also enables us to better understand the sources of estimation uncertainty. Next, we consider three special cases of the general results, for the unconditional quantile estimation, for the conditional quantile regression with a binary covariate, and for the instrumental variable quantile regression (IVQR). For each of these special cases, we provide the second-order bias and MSE to illustrate their behavior which depends on certain parameters and distributional characteristics. The Monte Carlo simulation indicates that the bias is larger at the extreme low and high tail quantiles, and the second-order bias corrected estimator has better behavior than the uncorrected ones in both conditional and unconditional quantile regression. The second-order bias corrected estimators are numerically much closer to the true estimators of data generating processes. As the higher order bias and MSE decrease as the sample size increases or as the regression error variance decreases, the benefits of the finite sample theory are more apparent when there are larger sampling errors in estimation.

Chapter three develops the second-order asymptotic properties (bias and mean squared error) of the asymmetric least squares (ALS) or expectile estimator, extending the second-order asymptotic results for the symmetric least squares (LS) estimators of Rilstone, Srivastava and Ullah (1996). The LS gives the mean regression function while the ALS gives the "expectile" regression function, a generalization of the usual regression function. The second-order bias result enables an improved bias correction and thus to obtain improved ALS estimation. In particular, we show that the second-order bias is much larger as the asymmetry is stronger, and therefore the benefit of the second-order bias correction is greater when we are interested in extreme expectiles which are used as a risk measure in financial economics. The higher order MSE result for the ALS estimation also enables us to better understand the sources of estimation uncertainty. The Monte Carlo simulation confirms the benefits of the second-order asymptotic theory and indicates that the second-order bias is larger at the extreme low and high expectiles, and the second-order bias correction improves the ALS estimator in bias.

Chapter four introduces the predictive quantile regression and predictive expectile regression. Predictive regression is a fundamental econometric model and widely discussed in finance literature. This chapter focuses on the second-order bias reduction for both regression models, which enable us to obtain a better predictive estimates. An empirical application to stock return prediction using the dividend yield illustrates the benefit of the proposed second-order bias reduction method. We show that the bias is larger at the tails of the stock return distribution.

Chapter five contains the conclusion.

Contents

List of Tables **xi**

1 Introduction **1**

2 The Second-order Bias and Mean Squared Error of Quantile Estimators **6**
2.1 Introduction 6
2.2 Conditional Quantile Estimators 8
2.2.1 Check Loss Function 8
2.2.2 Assumptions 11
2.3 Second-order Bias of Quantile Estimators 19
2.4 The MSE of Quantile Estimators 29
2.5 Illustrations 40
2.5.1 Unconditional Quantile Estimator 40
2.5.2 Conditional Quantile Estimator with Binary Independent Variable . 44
2.5.3 Instrumental Variable Quantile Regression 48
2.6 Monte Carlo Simulation 51
2.6.1 Simulation Design 51
2.6.2 Simulation Results 53
2.7 Appendix 54
2.7.1 Appendix A 54
2.7.2 Appendix B 55
2.7.3 Appendix C 56
2.7.4 Appendix D 67
2.8 References 68

3 The Second-order Asymptotic Properties of Asymmetric Least Squares Estimation **88**
3.1 Introduction 88
3.2 Asymmetric Least Squares Estimation 90
3.2.1 Loss Functions 90
3.2.2 Assumptions 94
3.3 Second-order Bias and MSE of the ALS Estimators 98

3.3.1 Bias 98
3.3.2 MSE 103
3.4 Special Case: Unconditional ALS Model 110
3.4.1 Bias 110
3.4.2 MSE 117
3.5 Monte Carlo Simulation 118
3.6 Appendix 121
3.6.1 Properties of a norm 121
3.6.2 Properties of the Dirac delta function 122
3.7 References 124

4 Second-order Bias Reduction in Predictive Quantile and Exprctile Regressions 128
4.1 Introduction 128
4.2 Predictive Quantile and Expectile Regressions 129
4.3 Second-order Bias and MSE for Quantile and Expectile Estimators 130
4.4 Empirical Application 132
4.5 References 135

5 Conclusions 142

List of Tables

2.1 Bias correction and MSE with x_i generated from exponential distribution, DGP 1, allowing hetroskedasticity . . . 72
2.2 Bias correction and MSE with x_i generated from exponential distribution, DGP 1, $\sigma_u = 0.5$ . . . 74
2.3 Bias correction and MSE with x_i generated from exponential distribution, DGP 1, $\sigma_u = 0.1$ . . . 75
2.4 Bias correction and MSE with x_i generated from mixture normal distribution (skewed unimodal), DGP 1, $\sigma_u = 0.5$ . . . 76
2.5 Bias correction and MSE with x_i generated from mixture normal distribution (strongly skewed), DGP 1, $\sigma_u = 0.5$ . . . 77
2.6 Bias correction and MSE in unconditional quantile model, DGP 1, $\sigma_u = 0.5$ 78
2.7 Bias correction and MSE with binary independent variable, DGP 1, $p = 0.3$, $\sigma_u = 0.5$ . . . 79
2.8 Bias correction for IVQR, $\sigma_u = 0.5$, N=60 . . . 80
2.9 Bias correction and MSE with x_i generated from exponential distribution, DGP 2, $R = 4$ . . . 81
2.10 Bias correction and MSE with x_i generated from exponential distribution, DGP 2, $R = 10$ . . . 83
2.11 Bias correction and MSE with x_i generated from mixture normal distribution (skewed unimodal), DGP 2, $R = 4$ . . . 84
2.12 Bias correction and MSE with x_i generated from mixture normal distribution (strongly skewed), DGP 2, $R = 4$ . . . 86

3.1 Conditional ALS regression . . . 126
3.2 Unconditional ALS regression . . . 127

4.1 Second-order bias reduction in predictive quantile regression, h=1 . . . 136
4.2 Second-order bias reduction in predictive quantile regression, h=3 . . . 137
4.3 Second-order bias reduction in predictive quantile regression, h=12 . . . 138
4.4 Second-order bias reduction in predictive expectile regression, h=1 . . . 139
4.5 Second-order bias reduction in predictive expectile regression, h=3 . . . 140
4.6 Second-order bias reduction in predictive expectile regression, h=12 . . . 141

Chapter 1

Introduction

There has been some significant literature on analytical finite sample properties of econometric estimators and test statistics over the past 60 years. See Nagar (1959), Sargan (1974, 1976), Basmann(1974), Rothenberg (1984) for linear models, and Amemiya (1980), Chesher and Spady (1989), Cordeiro and McCullagh (1991), Newy and Smith (2004), Rilstone, Srivastava and Ullah (1996), Bao and Ullah (2007), and Ullah (2004) for non-linear models. It is well known that the large sample theory properties may not imply the finite sample behavior of econometrics estimators and test statistics. In fact, the use of first-order asymptotic theory results for small or even moderately large samples may give misleading results. The finite sample properties has been developing rapidly in the mean regression models., as well as its applications in improving the inference for finite samples, determining optimal instruments. The finite sample properties permits us to obtain the better approximation of the bias and mean squared error (MSE) of a class of estimators. It also allows us to understand what affects the estimators and allows us to find an approach to improve the finite sample behavior of various estimators and test statistics.[1] Rilstone, Srivastava and Ullah (1996) developed the large-N second-order bias and mean squared error (MSE) of a class of nonlinear estimators in models with i.i.d. samples. Bao and Ullah (2007) analyzed the results for time series dependent observations.

On the other hand, unlike in the mean regression models for which both asymptotic theory and finite sample theory have been fully developed over the past 60 years, there is limited research on finite sample theory for the quantile regression. The literature on the quantile regression models has been either the first-order asymptotic, or the only the order of the second-order representation, see Koenker and Bassett (1978), Bahadur (1966), Kiefer (1967), Jureckova and Sen (1987, 1996), and He and Shao (1996). Quantile regression provides useful tools for a more complete view of the statistical landscape and the relationship among stochastic variables. All of recent research in financial economics uses the large sample theory to study the properties of quantiles (e.g. Value-at-Risk, or VaR) and the variants of VaR, so called the Expected Shortfalls (ES) in financial risk management. Quantile regression methods also have been used to study in many economic applications such as determinants of wages, discrimination effects, and trends in income inequality. There is extensive literature on the first-order asymptotic properties of the quantile regression, which can be improved by considering the higher order asymptotic approximations that are better approximations in finite sample. Phillips (1991) provided the asymptotic theory for LAD estimators using the delta functions, and developed the higher-order asymptotic expansions for LAD estimators. In Chapter 2, we develop the second-order bias results for quantile

[1] We refer the finite sample properties to the higher order asymptotic approximation, in the sense it provides better approximation in small or even moderately large sample. The finite sample properties in this paper is not the exact moment or distributional properties. See Ullah (2004).

estimators, using the higher-order asymptotic expansions for quantile estimators.

The challenge to study the finite sample properties of the quantile estimator in Chapter 2, is due to the non-differentiability of the indicator function inside of the quantile objective function. Dealing with the non-differentiable problem is common in mathematics and physics, this has been rarely explored for the finite sample properties of the quantile regression. See Elliott, Komunjer, and Timmermann (2005). Using the properties of Dirac delta function, we were able to obtain the finite sample properties for quantile regression.

The ALS estimation was first interpreted as a maximum likelihood estimator when the disturbances arise from a normal distribution with unequal weight placed on positive and negative disturbances by Aigner, Amemiya and Poirier (1976). Newey and Powell (1987) proposed the term, ALS, and investigated the estimation and hypothesis tests for coefficients of linear ALS models. The symmetric LS gives the mean regression function while the ALS gives the "expectile" regression function, a generalization of the usual regression function. The ALS model has been used in many economic applications. A lot of recent research in financial economics uses the large sample theory to study the properties of ALS models in financial risk management. Kuan, Yeh and Hsu (2009), proposed an expectile based value-at-risk and extended asymptotic results to allow for stationary and weakly dependent data using a parametric method. Xie, Zhou and Wan (2014) developed a nonparametric varing-coefficient apporach for modeling the expectile-based value-at-risk. However, the literature on the ALS model has been entirely the first-order asymptotic properties. The first-order asymptotic properties of the ALS model can be improved by considering the higher order asymptotic approximations which are better approximations. In Chapter 3, we filled this unexplored area by developing the analytical results of the second-order bias and mean squared error (MSE) for the ALS models.

Predictive regression is a fundamental econometric model in finance. It has been widely discussed in finance literature. Unlike in the mean predictive regression models for which the bias reduction has been actively developed, there is little studies focused on the predictive quantile regression or predictive expectile regression. Quantile regression is capable of modeling the entire conditional distribution. It yields valuable insights of the entire conditional distribution of interest. Expectile regression is also called asymmetric least squares (ALS) regression. It also gives us information of the entire distribution of interest that we would not obtain directly from standard mean regression methods. This is essential for applications such as financial risk management, where we are more interested in modeling the tails of the conditional distribution. In this paper, we develop the predictive quantile and expectile regression models. Since the bias is larger towards the tails of the conditional density than near the median, then the benefit of the bias correction is greater when we are interested in the deeper tail quantiles.

There has been extensive literature on the bias reduction problem. Zhu (2013) proposed a method based on the jackknife technique to reduce the bias for predictive regressions, in order to obtain better predictive estimates. In Chapter 4, we use the finite sample property results to reduce the second-order bias of predictive regression. Rilstone, Srivastava and Ullah (RSU, 1996) developed the second-order bias of a class of nonlinear estimators in models with i.i.d. samples. We analyzed the second-order bias and mean squared error (MSE) in our previous studies. In Chapter 4, we apply the second-order bias and MSE results on the application of stock returns. We are able to calculate the second-order bias of the predictive quantile and expectile estimator and the bias reduction enable us to obtain a better predictive estimates. We illustrate the proposed second-order bias reduction to predict the stock returns by the lagged dividend yield. The data used in this application is from Welch and Goyal (2008). We try both short- and long-horizon regressions for both quantile and expectile models.

Chapter 2

The Second-order Bias and Mean Squared Error of Quantile Estimators

2.1 Introduction

There has been some significant literature on analytical finite sample properties of econometric estimators and test statistics over the past 60 years. The finite sample properties permits us to obtain the better approximation of the bias and mean squared error (MSE) of a class of estimators. It also allows us to understand what affects the estimators and allows us to find an approach to improve the finite sample behavior of various estimators and test statistics. Rilstone, Srivastava and Ullah (1996) developed the large-N second-order bias and mean squared error (MSE) of a class of nonlinear estimators in models with i.i.d. samples. Bao and Ullah (2007) analyzed the results for time series dependent observations. On the other hand, unlike in the mean regression models for which both asymptotic theory and finite sample theory have been fully developed, there is limited research on finite sample theory for the quantile regression. The literature on the quantile regression models has been only the order of the second-order representation. Quantile regression methods also have been used to study in many economic applications such as determinants of wages and financial risk management. There is extensive literature on the first-order asymptotic properties of the quantile regression, which can be improved by considering the higher order asymptotic approximations that are better approximations in finite sample. In this paper, we develop the second-order bias results for quantile estimators, using the higher-order asymptotic expansions for quantile estimators. We discover that while the median is unbiased for a symmetric distribution and the bias of the other quantiles is larger at the tails of any distribution. When the independent variable is generated from symmetric distribution, the bias is zero. When the volatility of the error term is larger, the quantile estimator has larger bias especially at the tails. The Monte Carlo simulations results provide improvement of quantile estimators and quantile prediction.

The paper is organized as follows. In Section 2.2, we present the notation, the moment condition of the quantile regression, the assumptions used in this paper. In Section 2.3, we develop the high-order asymptotic expansion of quantile estimators, and derive the second-order bias of conditional quantile estimators. In Section 2.4, we derive the second-order MSE of conditional quantile estimators. Section 2.5 provides illustrations, including second-order bias and MSE of unconditional and conditional quantile estimators with several different distributions. In Section 3.5, we present Monte Carlo simulations.

2.2 Conditional Quantile Estimators

2.2.1 Check Loss Function

Consider a random variable y from the distribution $F(\cdot)$. Let $f_i(\cdot)$ denote the conditional density, for $i = 1, \cdots, N$, $f_i^{(j)}(\cdot)$ denote the jth order derivative of $f_i(\cdot)$ for $j \geq 1$. The jth order partial derivatives of a matrix $A(\beta)$ is defined as $\nabla_\beta^j A(\beta)$. If $A(\beta)$ is a $k \times 1$ vector, $\nabla_\beta^j A(\beta)$ is a $k \times k^j$ matrix. For a matrix A, $||A||$ denotes the usual norm, $[trace\,(AA')]^{1/2}$. If A is a $k \times 1$ vector, according to the Appendix $A.1$, $||A|| = (A'A)^{1/2}$. The Kronecker product is defined in the usual way. For an $m \times n$ matrix A and a $p \times q$ matrix B, we have $A \otimes B$ as an $mp \times nq$ matrix. The $\overline{X} = E(X)$ denotes the expectation of a random vector X. Given $\alpha \in (0, 1)$, the α-quantile q_α of y with distribution function $F(y)$ is defined as

$$q_\alpha = \inf\{y : F(y) \geq \alpha\}.$$

The quantile can be considered as the inverse of the distribution function. The quantile q_α is the value such that α percent of the mass of the distribution is less than q_α, which can be obtained from

$$q_\alpha = \arg \min_q E[L_\alpha(y - q)],$$

where the check loss function is defined as

$$L_\alpha(y - q) = (\alpha - \mathbf{1}(y - q < 0))\,(y - q)\,.$$

For the random variable (y_i, x_i) with conditional distribution function $F(y|x)$ the conditional quantile function q_α is

$$q_\alpha = \inf\{y : F(y|x) \geq \alpha\}.$$

As a function of x_i, the quantile regression function can be nonlinear. We consider a simple linear model, i.e. $q_\alpha = x_i'\beta_\alpha$, where the quantile estimators β_α varies across α. Then the linear quantile regression model is

$$y_i = x_i'\beta_\alpha + u_i, \tag{2.1}$$

where y_i is a scalar and x_i is a $k \times 1$ vector, u_i is the error defined to be the difference between y_i and its conditional α-quantile $x_i'\beta_\alpha$. To simplify the notation, we use β to denote β_α hereafter.

The $k \times 1$ vector quantile estimators $\widehat{\beta}$ can be obtained by solving

$$\min_\beta E[L_\alpha(\beta)] = E[\left(\alpha - \mathbf{1}(y_i < x_i'\beta)\right)\left(y_i - x_i'\beta\right)]. \tag{2.2}$$

Following the condition A0 in Komunjer (2005) and Elliott, Komunjer, and Timmermann (EKT 2005), we restrict the conditional quantile model that $x_i'\beta$, the conditional α-quantile of y_i, is identified on Θ, i.e. for any $(\beta_1, \beta_2) \in \Theta^2$ we have $x_i'\beta_1 = x_i'\beta_2$ a.s. -P, for all i, if and only if $\beta_1 = \beta_2$. The check loss function $L_\alpha(\beta) = (\alpha - \mathbf{1}(y_i < x_i'\beta))(y_i - x_i'\beta)$ is continuously differentiable on $\Theta \backslash A$, where $A = \{\beta \in \Theta : y_i = x_i'\beta\}$. Let $\nabla_\beta^1 E[L_\alpha(\beta)]$ denote the gradient of $E[L_\alpha(\beta)]$ on $\Theta\backslash A$. By the law of iterated expectations, $E[L_\alpha(\beta)] = E\{E[L_\alpha(\beta)]\}$, so that

$$\nabla_\beta^1 E[L_\alpha(\beta)] = E\{\nabla_\beta^1 L_\alpha(\beta) E[\mathbf{1}(\beta \in A^c)]\} + E\{\nabla_\beta^1 L_\alpha(\beta) E[\mathbf{1}(\beta \in A)]\},$$

where $E[\mathbf{1}(\beta \in A^c)] = 1$, and $E[\mathbf{1}(\beta \in A)] = 0$. Therefore, $E[L_\alpha(\beta)]$ is continuously differentiable on Θ. Then can write the population moment condition as

$$\nabla_{\beta}^{1} E[L_\alpha(\beta)] = E[-\nabla_{\beta}^{1}\mathbf{1}(y_i - x_i'\beta < 0)(y_i - x_i'\beta)] + E[(\alpha - \mathbf{1}(y_i < x_i'\beta))(-x_i)]. \tag{2.3}$$

By the definition of Dirac delta function in Appendix $B.1$, $\mathbf{1}(y_i - x_i'\beta < 0) = \mathbf{1}(x_i'\beta - y_i \geq 0) \equiv \phi(x_i'\beta - y_i)$ is a Heaviside unit step function. Then

$$\nabla_{\beta}^{1}\mathbf{1}(y_i - x_i'\beta < 0) = \nabla_{\beta}^{1}\phi(x_i'\beta - y_i) = \frac{\mathrm{d}\phi(x_i'\beta - y_i)}{\mathrm{d}(x_i'\beta - y_i)}\frac{\mathrm{d}(x_i'\beta - y_i)}{\mathrm{d}\beta} = x_i'\delta(x_i'\beta - y_i).$$

See Gelfand and Shilov (1964). The first term of the equation (2.3)) can be written as $E[x_i'\delta(x_i'\beta - y_i)(y_i - x_i'\beta)]$, which equals zero. According to the property of Dirac delta function in Appendix $B.4$, we have $\delta(x_i'\beta - y_i) = \delta(y_i - x_i'\beta)$. According to the property of Dirac delta function in Appendix $B.3$, we have

$$\begin{aligned}
E[x_i'\delta(x_i'\beta - y_i)(y_i - x_i'\beta)] &= E[x_i'\delta(y_i - x_i'\beta)(y_i - x_i'\beta)] \\
&= E\left[x_i'E\left[\delta(y_i - x_i'\beta)(y_i - x_i'\beta)|x_i\right]\right] \\
&= E\left[x_i'\int_{-\infty}^{+\infty}\delta(y_i - x_i'\beta)(y_i - x_i'\beta)f_i(y_i)\mathrm{d}y_i\right] \\
&= E\left[x_i'(x_i'\beta - x_i'\beta)f_i(x_i'\beta)\right] \\
&= 0.
\end{aligned}$$

where $f_i(x_i'\beta) \equiv f_i(x_i'\beta|x_i)$ is the conditional density of y_i evaluated at $y_i = x_i'\beta$. Thus, the moment condition can be written as

$$\nabla_{\beta}^{1} E[L_\alpha(\beta)] = E[(\alpha - \mathbf{1}(y_i < x_i'\beta))(-x_i)] = E[s_i(\beta)],$$

where the score function $s_i(\beta) = (\alpha - \mathbf{1}(y_i < x_i'\beta))(-x_i)$. The sample moment condition can be written as

$$\Psi_N(\beta) = \frac{1}{N}\sum_{i=1}^{N} s_i(\beta), \tag{2.4}$$

which satisfies equation (2.1).

2.2.2 Assumptions

Rilstone, Srivastava and Ullah (RSU, 1996) developed the second-order bias and MSE of a class of estimators. These results apply for both normal and non-normal errors. A class of estimators $\widehat{\beta}$ can be written as a solution to a set of moment equations of the form

$$\Psi_N(\widehat{\beta}) = \frac{1}{N}\sum_{i=1}^{N} s_i(\widehat{\beta}) = 0, \tag{2.5}$$

where $s_i(\beta) \equiv s(x_i;\beta)$ is a known $k\times 1$ vector-valued function of the observable k-dimensional random vectors x_i and a parameter vector $\beta \in R^k$ with true value β_0 such that $E[s_i(\beta)] = 0$ holds only at $\beta = \beta_0$ for all i. The moment equation $\Psi_N(.)$ can be the first-order condition of some optimization criteria. The estimators can be maximum likelihood (ML), least square (LS) and Generalized Method of Moments (GMM) estimators. In RSU (1996), the Assumption A-C are sufficient for $\widehat{\beta}$ to have an asymptotically normal distribution. To obtain the stochastic expansion of $\widehat{\beta}$, the Assumption A-C are assumed to hold along with the $\sqrt{N}$-consistency of $\widehat{\beta}$. In this section, we give the modified Assumption A-C some remarks for quantile model.

Assumption A. The jth-order derivative of $s_i(\beta)$ exists in a neighborhood of β_0 and is continuous with probability 1, and $E\left[||x_i||^{j+1}\, f_i^{(j-1)}(0|x_i)\right]^2 < \infty$, for $j \geq 1$, where $f_i^{(0)}(0|x_i) = f_i(0|x_i)$ is the conditional density of u_i evaluated at $u_i = 0$.

Assumption B. For some neighborhood of β_0, $\left(E\nabla_\beta^1 \Psi_N(\beta)\right)^{-1} = O(1)$.

Assumption C. For any $\varepsilon \to 0$, $r_j(\beta) = \left\|\nabla_\beta^{j-1} s_i(\beta) - \nabla_\beta^{j-1} s_i(\beta_0) - \nabla_\beta^{j} s_i(\beta_0)\,(\beta-\beta_0)\right\| / \|\beta-\beta_0\| \to$

0 as $\beta \to \beta_0$, $E\left[\sup_{\|\beta-\beta_0\|<\varepsilon} r_j(\beta)\right] < \infty$, with probability 1, and $N^{-1}\sum_{i=1}^{N} \nabla_\beta^j s_i(\beta_0) \xrightarrow{p} E\left[\nabla_\beta^j s_i(\beta_0)\right]$ for $j \geq 1$, where $\nabla_\beta^0 s_i(\beta) = s_i(\beta)$.

In the following we discuss the assumptions under which Theorems, Corollarys and Propositions stated below will be true. We argue that these assumptions encompass a wide variety of conditional quantile models, which means that the analytical results are of wide interest and applicability. In general, Assumption A-C are related to the conditions in Komunjer (2005), but include more primitive conditions. See Huber (1976), Pollard (1985), Pakes and Pollard (1989), Newey and McFadden (1994), Andrews (1994), Chernozhukov and Hong (2003). The most substantial difference is that the conditions in Komunjer (2005) are stated to obtain the asymptotic normality of conditional quantile estimators, which only handles the nonsmoothness of the quantile objective function, while in this paper, to handle higher order stochastic expansion, the Assumption C requires conditions of the higher order stochastic equicontinuity.

First, we discuss the Assumption A. We restrict the conditional quantile model that $x_i'\beta$, the conditional α-quantile of y_i, is identified on Θ, and $E[L(\beta)]$ is continuously differentiable on Θ, then the sample moment conditon $\Psi_N(\beta)$ is continuous differentiable on Θ. In this case, for every $\beta \in \Theta$, $\nabla_\beta^1 \Psi_N(\beta)$ exists and is continuous with probability 1, so that the second-order and third-order derivative of $\Psi_N(\beta)$ exists and continuous with probability 1. By the definition of Dirac delta function in Appendix $B.1$, we have $\nabla_\beta^1 \mathbf{1}(y_i - x_i'\beta < 0) = x_i'\delta(x_i'\beta - y_i)$. Note that β is a $k \times 1$ vector, where x_i is a $k \times 1$ vector, $s_i(\beta)$ is a $k \times 1$ vector, $\delta(x_i'\beta - y_i)$ is a scalar. The derivative of a $k \times 1$ vector $s_i(\beta)$ with respect to a $k \times 1$ vector β is a $k \times k$ matrix $\nabla_\beta^1 s_i(\beta)$. Then the first-order derivative of $s_i(\beta)$ exists and is continuous with probability 1.

$$\begin{aligned}\nabla_\beta^1 s_i(\beta) &= \nabla_\beta^1[(\alpha - \mathbf{1}(y_i < x_i'\beta))(-x_i)] \\ &= x_i \nabla_\beta^1 \phi(x_i'\beta - y_i) \\ &= x_i \frac{\mathrm{d}\phi(x_i'\beta - y_i)}{\mathrm{d}(x_i'\beta - y_i)} \frac{\mathrm{d}(x_i'\beta - y_i)}{\mathrm{d}\beta} \\ &= x_i x_i' \delta(x_i'\beta - y_i).\end{aligned}$$

We can show that locally at any β, the difference between the sample mean of first derivative of score function and its expected value converges in probability to zero, i.e. $\frac{1}{N}\sum_{i=1}^{N} x_i x_i' \delta(x_i'\beta - y_i) - E\left[x_i x_i' \delta(x_i'\beta - y_i)\right] \xrightarrow{p} 0$. Using the the properties in Appendix $A.2, B.3$ and $B.4$, we obtain

$$\begin{aligned}E\left\|\nabla_\beta^1 s_i(\beta_0)\right\|^2 &= E\left[\left\|x_i x_i'\right\| \delta(x_i'\beta_0 - y_i)\right]^2 \\ &= E\left[\left[tr\left(x_i x_i' x_i x_i'\right)\right]^{1/2} \delta(y_i - x_i'\beta_0)\right]^2 \\ &= E\left[\left[tr\left(x_i' x_i x_i' x_i\right)\right]^{1/2} E\left[\delta(y_i - x_i'\beta_0)|x_i\right]\right]^2 \\ &= E\left[\left(x_i' x_i x_i' x_i\right)^{1/2} \int_{-\infty}^{+\infty} \delta(y_i - x_i'\beta_0) f_i(y_i)\mathrm{d}y_i\right]^2 \\ &= E\left[x_i' x_i f_i(x_i'\beta_0)\right]^2 \\ &= E\left[||x_i||^2 f_i(x_i'\beta_0)\right]^2 \\ &< \infty.\end{aligned}$$

The second-order derivative of a $k \times 1$ vector $s_i(\beta)$ with respect to a $k \times 1$ vector β is a $k \times k^2$ matrix $\nabla_\beta^2 s_i(\beta)$. The second order derivative of $s_i(\beta)$ exists and is continuous with probability 1.

$$\nabla_\beta^2 s_i(\beta) = \nabla_\beta^1[x_i x_i' \delta(x_i'\beta - y_i)] = \left(x_i x_i'\right) \otimes \nabla_\beta^1 \delta(x_i'\beta - y_i),$$

where the derivative of a scalar $\delta(x_i'\beta - y_i)$ with respect to a $k \times 1$ vector β is a $1 \times k$ row vector $\nabla_\beta^1 \delta(x_i'\beta - y_i)$. We denote

$$\nabla_\beta^1 \delta(x_i'\beta - y_i) = \frac{\mathrm{d}\delta(x_i'\beta - y_i)}{\mathrm{d}(x_i'\beta - y_i)} \frac{\mathrm{d}(x_i'\beta - y_i)}{\mathrm{d}\beta} = x_i'\delta^{(1)}(x_i'\beta - y_i),$$

where $\delta^{(1)}(x_i'\beta - y_i)$ is a scalar. Then we can rewrite the second-order derivative of $s_i(\beta)$ as

$$\nabla_\beta^2 s_i(\beta) = \left(x_i x_i'\right) \otimes \nabla_\beta^1 \delta(x_i'\beta - y_i) = \left(x_i x_i'\right) \otimes x_i'\delta^{(1)}(x_i'\beta - y_i).$$

We can show that locally at any β, the difference between the sample mean of second derivative of score function and its expected value converges in probability to zero, i.e. $\frac{1}{N}\sum_{i=1}^N (x_i x_i') \otimes x_i'\delta^{(1)}(x_i'\beta - y_i) - E\left[(x_i x_i') \otimes x_i'\delta^{(1)}(x_i'\beta - y_i)\right] \xrightarrow{p} 0$. Using the the properties in Appendix $A.3, B.5$ and $B.6$, we obtain

$$\begin{aligned}
E\left\|\nabla_\beta^2 s_i(\beta_0)\right\|^2 &= E\left\|(x_i x_i') \otimes x_i' \delta^{(1)}(x_i'\beta_0 - y_i)\right\|^2 \\
&= E\left\|\left(x_i x_i'\right) \otimes x_i' E\left[\delta^{(1)}(x_i'\beta_0 - y_i)|x_i\right]\right\|^2 \\
&= E\left\|\left(x_i x_i'\right) \otimes x_i' \left(\int_{-\infty}^{+\infty} \delta^{(1)}(x_i'\beta_0 - y_i) f_i(y_i) \mathrm{d}y_i\right)\right\|^2 \\
&= E\left\|\left(x_i x_i'\right) \otimes x_i' \left(-\int_{-\infty}^{+\infty} \delta^{(1)}(y_i - x_i'\beta_0) f_i(y_i) \mathrm{d}y_i\right)\right\|^2 \\
&= E\left\|\left(x_i x_i'\right) \otimes x_i' \left(\int_{-\infty}^{+\infty} \delta(y_i - x_i'\beta_0) f_i^{(1)}(y_i) \mathrm{d}y_i\right)\right\|^2 \\
&= E\left[f_i^{(1)}(x_i'\beta_0)\left\|\left(x_i x_i'\right) \otimes x_i'\right\|\right]^2 \\
&= E\left[f_i^{(1)}(x_i'\beta_0)\left\{tr\left(\left[\left(x_i x_i'\right) \otimes x_i'\right]\left[\left(x_i x_i'\right) \otimes x_i\right]\right)\right\}^{1/2}\right]^2 \\
&= E\left[f_i^{(1)}(x_i'\beta_0)\left[tr\left(\left(x_i x_i' x_i x_i'\right) \otimes \left(x_i x_i'\right)\right)\right]^{1/2}\right]^2 \\
&= E\left[f_i^{(1)}(x_i'\beta_0)\left[tr\left(x_i' x_i x_i' x_i x_i' x_i\right)\right]^{1/2}\right]^2 \\
&= E\left[f_i^{(1)}(x_i'\beta_0)\left(x_i' x_i\right)\right]^3 \\
&= E\left[f_i^{(1)}(x_i'\beta_0)\,||x_i||^3\right]^2 \\
&< \infty.
\end{aligned}$$

The third-order derivative of a $k \times 1$ vector $s_i(\beta)$ with respect to a $k \times 1$ vector β is a $k \times k^3$ matrix $\nabla_\beta^3 s_i(\beta)$. The third order derivative of $s_i(\beta)$ exists and is continuous with probability 1.

$$\nabla_\beta^3 s_i(\beta) = \nabla_\beta^2[x_i x_i' \delta(x_i'\beta - y_i)] = \left(x_i x_i'\right) \otimes \nabla_\beta^2 \delta(x_i'\beta - y_i),$$

where the derivative of a $1 \times k$ row vector $\nabla_\beta^1 \delta(x_i'\beta - y_i)$ with respect to a $k \times 1$ vector β is a $1 \times k^2$ row vector $\nabla_\beta^2 \delta(x_i'\beta - y_i)$. We denote

$$\nabla_\beta^2 \delta(x_i'\beta - y_i) = \nabla_\beta^1 x_i' \delta^{(1)}(x_i'\beta - y_i) = x_i' \otimes \frac{\mathrm{d}\delta^{(1)}(x_i'\beta - y_i)}{\mathrm{d}(x_i'\beta - y_i)} \frac{\mathrm{d}(x_i'\beta - y_i)}{\mathrm{d}\beta} = x_i' \otimes x_i' \delta^{(2)}(x_i'\beta - y_i),$$

where $\delta^{(2)}(x_i'\beta - y_i)$ is a scalar. Then we can rewrite the third-order derivative of $s_t(\beta)$ as

$$\nabla_\beta^3 s_i(\beta) = \left(x_i x_i'\right) \otimes \nabla_\beta^2 \delta(x_i'\beta - y_i) = \left(x_i x_i'\right) \otimes x_i' \otimes x_i' \delta^{(2)}(x_i'\beta - y_i).$$

We can show that locally at any β, the difference between the sample mean of second derivative of score function and its expected value converges in probability to zero, i.e. $\frac{1}{N}\sum_{i=1}^{N}(x_i x_i') \otimes x_i' \otimes x_i' \delta^{(2)}(x_i'\beta - y_i) - E\left[(x_i x_i') \otimes x_i' \otimes x_i' \delta^{(2)}(x_i'\beta - y_i)\right] \xrightarrow{p} 0$. Using the the properties in Appendix $A.4, B.6$ and $B.7$, we obtain

$$\begin{aligned}
E\left\|\nabla_\beta^3 s_i(\beta_0)\right\|^2 &= E\left\|\left(x_i x_i'\right) \otimes x_i' \otimes x_i' \delta^{(2)}(x_i'\beta_0 - y_i)\right\|^2 \\
&= E\left\|\left(x_i x_i'\right) \otimes x_i' \otimes x_i' E\left[\delta^{(2)}(x_i'\beta_0 - y_i)|x_i\right]\right\|^2 \\
&= E\left\|\left(x_i x_i'\right) \otimes x_i' \otimes x_i' \int_{-\infty}^{+\infty} \delta^{(2)}(y_i - x_i'\beta_0) f_i(y_i) \mathrm{d}y_i\right\|^2 \\
&= E\left\|\left(x_i x_i'\right) \otimes x_i' \otimes x_i' \int_{-\infty}^{+\infty} \delta(y_i - x_i'\beta_0) f_i^{(2)}(y_i) \mathrm{d}y_i\right\|^2 \\
&= E\left\{f_i^{(2)}(x_i'\beta_0)\left\|\left(x_i x_i'\right) \otimes x_i' \otimes x_i'\right\|\right\}^2 \\
&= E\left\{f_i{}^{(2)}(x_i'\beta_0) tr\left(\left[\left(x_i x_i'\right) \otimes x_i' \otimes x_i'\right]\left[\left(x_i x_i'\right) \otimes x_i \otimes x_i\right]\right)^{1/2}\right\}^2 \\
&= E\left[f_i{}^{(2)}(x_i'\beta_0) tr\left[\left(x_i x_i' x_i x_i'\right) \otimes \left(x_i' \otimes x_i'\right)\left(x_i \otimes x_i\right)\right]^{1/2}\right]^2 \\
&= E\left[f_i{}^{(2)}(x_i'\beta_0) tr\left[\left(x_i x_i' x_i x_i'\right) \otimes x_i' x_i \otimes x_i' x_i\right]^{1/2}\right]^2 \\
&= E\left[f_i{}^{(2)}(x_i'\beta_0) tr\left[\left(x_i' x_i x_i' x_i\right) x_i' x_i x_i' x_i\right]^{1/2}\right]^2 \\
&= E\left[f_i{}^{(2)}(x_i'\beta_0)\left(x_i' x_i x_i' x_i\right)\right]^2 \\
&= E\left[f_i{}^{(2)}(x_i'\beta_0)\left(x_i' x_i\right)^2\right]^2 \\
&= E\left[f_i{}^{(2)}(x_i'\beta_0)\left\|x_i\right\|^4\right]^2 \\
&< \infty.
\end{aligned}$$

Since the conditional density of y_i given x_i evaluated at $y_i = x_i'\beta$ is the same as the conditional density of u_i given x_i evaluated at $u_i = 0$. If we use $f_i(0|x_i)$ to denote the conditional density of u_i given x_i evaluated at $u_i = 0$, then the conditions we observe above can be written as

$$E\left\|\nabla_\beta^1 s_i(\beta_0)\right\|^2 = E\left[||x_i||^2 f_i(0|x_i)\right]^2 < \infty,$$

$$E\left\|\nabla_\beta^2 s_i(\beta_0)\right\|^2 = E\left[f_i^{(1)}(0|x_i)\,||x_i||^3\right]^2 < \infty,$$

$$E\left\|\nabla_\beta^3 s_i(\beta_0)\right\|^2 = E\left[f_i{}^{(2)}(0|x_i)\,\|x_i\|^4\right]^2 < \infty.$$

Combinning the conditions in one single equation, we have $E\left[||x_i||^{j+1} f_i^{(j-1)}(0|x_i)\right]^2 < \infty$, and it is easy to show that this condition applies for $j \geq 1$, where $f_i^{(0)}(0|x_i) = f_i(0|x_i)$.

Next, we discuss the Assumption B. For some neighborhood of β_0, $\left(E\nabla_\beta^1 \Psi_N(\beta)\right)^{-1} = O(1)$ is required to obtain the stochastic expansion of $\widehat{\beta} - \beta$ in Section 3. That is

$$\begin{aligned}\left(E\nabla_\beta^1 \Psi_N(\beta)\right)^{-1} &= \left(E\frac{1}{N}\sum_{i=1}^{N} x_i x_i' \delta(x_i'\beta - y_i)\right)^{-1} \\ &= \left(E\left[x_i x_i' \delta(x_i'\beta - y_i)\right]\right)^{-1} \\ &= \left(E\left[x_i x_i' f_i(x_i'\beta)\right]\right)^{-1} \\ &= O(1).\end{aligned}$$

Lastly, we discuss the Assumption C. To derive the second-order bias and MSE of the quantile estimators, we use the higher order Taylor expansion of the gradient $\Psi_N(\beta)$ around β_0, which satisfies $\Psi_N(\widehat{\beta}) = 0$. This approach requires $\Psi_N(\beta)$ and the derivatives of $\Psi_N(\beta)$ to be sufficient smooth, which is not the case with the quantile regression. In general, Assumption C is related to the conditions in Komunjer (2005), which applies the stochastic equicontinuity conditions to handle the expansion of discontinuous and nonsmooth objective function. The discontinuous and nonsmooth problem has been discussed in many literatures, including Huber (1976), Pollard (1985), Newey and McFadden (1994), and Andrews (1994). The basic insight of these papers is that smoothness of the objective function can be replaced by smoothness of the limit if certain reminder terms are small. Therefore, those stochastic equicontinuity conditions do not require differentiablilty of the criterion function, but require that the reminder term of the expansion can be controlled in a particular way over a neighborhood of β_0, and those conditions are sufficient for $\widehat{\beta}$ to have an asymptotically normal distribution. In this paper, to derive the second-order bias and MSE of $\widehat{\beta}$, besides of those stochastic conditions discussed in those literatures mentioned above, we need additional smoothness and dominating conditions for higher moments of quantile objective function. The Assumption C in this paper extends the conditions in Theorem 7.3 in Newey and McFadden (1994), gives a version of the stochastic equicontinuity for Lipschitz moment function, and allows for moments of the objective function to be Lipschitz at β_0 and differentiable with probability 1, ranthe ran continuously differentiable. The Assumption C in this paper restricts the remainder $r_j(\beta)$ to be well behaved uniformly near the true parameter, and this uniformity property requires that higher moments of the objective function be Lipschitz at β_0 with an integrable Lipschitz constant with probability 1. The Assumption C in this paper requires a stronger condition than the conditions in Pollard (1985) and Andrews (1994), but places no restrictions on the dependence of the independent variables. Similarily as the Assumption C in RSU (1996), the additional smoothness and dominating conditions do not require much more from the model. This means that Assumption C in this paper is of wide applicability.

2.3 Second-order Bias of Quantile Estimators

To obtain the second-order bias for quantile estimator, we implement the Taylor's expansion of $\Psi_N(\widehat{\beta}) = 0$ around β_0 up to the second order,

$$0 = \Psi_N + \nabla\Psi_N(\widehat{\beta} - \beta_0) + \frac{1}{2}\nabla^2\Psi_N\left[(\widehat{\beta} - \beta_0) \otimes (\widehat{\beta} - \beta_0)\right] + o_p\left(N^{-1}\right), \tag{2.6}$$

where $\Psi_N = \Psi_N(\beta_0)$. The ordinary stochastic expansion of $\widehat{\beta}$ is obtained from equation (2.6). However, a diffculty arises from the derivatives of the moment condition. Using the properties of the delta function in the appendix or in Phillips (1991, p. 455), we can rewrite (2.6) as

$$\begin{aligned} 0 &= \Psi_N + \overline{\nabla\Psi_N}(\widehat{\beta} - \beta_0) + \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)(\widehat{\beta} - \beta_0) \\ &\quad + \frac{1}{2}\nabla^2\Psi_N\left[(\widehat{\beta} - \beta_0) \otimes (\widehat{\beta} - \beta_0)\right] + o_p\left(N^{-1}\right) \\ &\equiv A_1 + A_2 + A_3 + A_4 + o_p\left(N^{-1}\right), \end{aligned} \tag{2.7}$$

where $\nabla\Psi_N \xrightarrow{p} \overline{\nabla\Psi_N}$, and $\nabla^2\Psi_N \xrightarrow{p} \overline{\nabla^2\Psi_N}$, that is

$$\nabla\Psi_N = \frac{1}{N}\sum_{i=1}^{N} x_i x_i' \delta(x_i'\beta - y_i) \xrightarrow{p} \frac{1}{N}\sum_{i=1}^{N} E\left[x_i x_i' f_i(0|x_i)\right] = \overline{\nabla\Psi_N},$$

$$\nabla^2\Psi_N = \frac{1}{N}\sum_{i=1}^{N}\left(x_i x_i'\right) \otimes x_i'\delta^{(1)}(x_i'\beta - y_i) \xrightarrow{p} \frac{1}{N}\sum_{i=1}^{N} E\left[\left(x_i x_i'\right) \otimes x_i' f_i^{(1)}(0|x_i)\right] = \overline{\nabla^2\Psi_N}.$$

To see the order of each of these terms, we first recall the asymptotic distribution of the quantile regression estimator when the α-quantile is linear in x_i,

$$\sqrt{N}(\widehat{\beta} - \beta_0) \xrightarrow{d} N(0, V_\alpha), \tag{2.8}$$

where

$$V_\alpha = \alpha(1-\alpha)\left[\frac{1}{N}\sum_{i=1}^{N} E[f_i(0|x_i)x_i x_i']\right]^{-1} E(x_i x_i')\left[\frac{1}{N}\sum_{i=1}^{N} E(f_i(0|x_i)x_i x_i')\right]^{-1},$$

and $f_i(0|x_i)$ is the density of u_i conditional on x_i evaluated at $u_i = 0$. See e.g. Koenker (2005). Since the quantile estimator is $\sqrt{N}$-consitent, we can obtain that the orders of both $A_1 = \Psi_N$ and $A_2 = \overline{\nabla\Psi_N}(\widehat{\beta} - \beta_0)$ are $O_p\left(N^{-1/2}\right)$.

We recall the following result. Let

$$\widehat{\beta} - \beta_0 = a_{-1/2} + R_N, \tag{2.9}$$

where $a_{-1/2}$ is a random sequence of $O_p\left(N^{-1/2}\right)$, and R_N is the remainder term of higher order. Bahadur (1966) and Kiefer (1967) established the celebrated results on the order of R_N, that is

$$R_N = O_p\left(n^{-3/4}\left(\log\log n\right)^{3/4}\right). \tag{2.10}$$

See Koenker (2005 pp. 122-123), and also Jureckova and Sen (1987, 1996 pp. 196-202), He and Shao (1996), and van der Vaart (1998 p. 310). Note that (2.10) implies that

$$R_N = O_p\left(N^{-3/4+\varepsilon}\right) \text{ for some small } \varepsilon > 0. \tag{2.11}$$

Below we use this result to obtain Lemma 1(b). In the following Lemma 1 and 2, we discuss A_3 and A_4. Our goal is to obtain the expression of the bias term $E\left(\widehat{\beta} - \beta_0\right)$ *up to* the second-order i.e., of order $O\left(N^{-1}\right)$, which will be disscussed in Lemma 3, 4, and 5.

Lemma 1. Let

$$\begin{aligned} A_3 &= \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)(\widehat{\beta} - \beta_0) \\ &= \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)a_{-1/2} + \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)\left[(\widehat{\beta} - \beta_0) - a_{-1/2}\right] \\ &\equiv A_{31} + A_{32}. \end{aligned} \tag{2.12}$$

Then,

(a) $A_{31} = O_p(N^{-7/6})$,

(b) A_{32} is smaller than $O_p(N^{-1})$, i.e. $A_{32} = o_p\left(N^{-1}\right)$.

Proof:

(a) According to Phillips (1991), Kim and Pollard (1990), and Prakasa Rao (1969), the term $V_N = \nabla\Psi_N - \overline{\nabla\Psi_N} = O_p\left(N^{-1/3}\right)$. We obtain that $\sqrt{N}a_{-1/2} = N(0, V_\alpha)$ is bounded and has zero mean. The term $\sqrt{N}A_{31}$ will contribute to $\sqrt{N}A_3$ through the variance of $\sqrt{N}a_{-1/2}$, and will produce an adjustment of $O_p\left(N^{-2/3}\right)$. Then we observe that A_{31} will be $O_p\left(N^{-7/6}\right)$, and the order of A_{32} is smaller than A_{31}.

(b) By (2.10), R_N is the remainder term of order smaller than $a_{-1/2}$. Since R_N is not of zero mean, because $E\left(R_N\right)$ is the high-order bias of quantile estimators, then $A_{32} = V_N R_N = O_p\left(N^{-1/3-3/4+\varepsilon}\right)$ is smaller than $O_p(N^{-1})$, i.e. $A_{32} = o_p\left(N^{-1}\right)$.

Lemma 2. Let

$$\begin{aligned} A_4 &= \frac{1}{2}\nabla^2\Psi_N\left[(\widehat{\beta} - \beta_0) \otimes (\widehat{\beta} - \beta_0)\right] \\ &= \frac{1}{2}\overline{\nabla^2\Psi_N}\left[(\widehat{\beta} - \beta_0) \otimes (\widehat{\beta} - \beta_0)\right] + \frac{1}{2}\left(\nabla^2\Psi_N - \overline{\nabla^2\Psi_N}\right)\left[(\widehat{\beta} - \beta_0) \otimes (\widehat{\beta} - \beta_0)\right] \\ &\equiv A_{41} + A_{42}. \end{aligned} \tag{2.13}$$

Then,

(a) $A_{41} = O_p(N^{-1})$,

(b) A_{42} is smaller than $O_p(N^{-1})$, i.e. $A_{42} = o_p\left(N^{-1}\right)$.

Proof:

(a) By (2.9), A_{41} can be written as

$$\begin{aligned} A_{41} &= \frac{1}{2}\overline{\nabla^2\Psi_N}\left\{\left[(\widehat{\beta}-\beta_0)-a_{-1/2}+a_{-1/2}\right]\otimes\left[(\widehat{\beta}-\beta_0)-a_{-1/2}+a_{-1/2}\right]\right\} \\ &= \frac{1}{2}\overline{\nabla^2\Psi_N}\left(a_{-1/2}\otimes a_{-1/2}\right) \\ &\quad +\frac{1}{2}\overline{\nabla^2\Psi_N}\left(a_{-1/2}\otimes\left[(\widehat{\beta}-\beta_0)-a_{-1/2}\right]\right) \\ &\quad +\frac{1}{2}\overline{\nabla^2\Psi_N}\left(\left[(\widehat{\beta}-\beta_0)-a_{-1/2}\right]\otimes a_{-1/2}\right) \\ &\quad +\frac{1}{2}\overline{\nabla^2\Psi_N}\left(\left[(\widehat{\beta}-\beta_0)-a_{-1/2}\right]\otimes\left[(\widehat{\beta}-\beta_0)-a_{-1/2}\right]\right), \end{aligned} \tag{2.14}$$

where only the frist term in equation (2.14) is $\frac{1}{2}\overline{\nabla^2\Psi_N}\left(a_{-1/2}\otimes a_{-1/2}\right)=O_p\left(N^{-1}\right)$, and the rest three terms in equation (2.14) are smaller than $O_p\left(N^{-1}\right)$.

(b) Since $\nabla^2\Psi_N-\overline{\nabla^2\Psi_N}$ is smaller than $O_p\left(1\right)$, then A_{42} is smaller than $O_p(N^{-1})$.

Given the Lemma 1 and 2, the equation (2.7) can be written as

$$\begin{aligned} 0 &= A_1+A_2+A_{31}+A_{41}+o_p\left(N^{-1}\right) \\ &= \Psi_N+\overline{\nabla\Psi_N}(\widehat{\beta}-\beta_0)+\left(\nabla\Psi_N-\overline{\nabla\Psi_N}\right)a_{-1/2}+\frac{1}{2}\overline{\nabla^2\Psi_N}\left(a_{-1/2}\otimes a_{-1/2}\right)+o_p\left(N^{-1}\right). \end{aligned} \tag{2.15}$$

The term $\nabla\Psi_N$ in an ordinary Taylor expansion, equation (2.6), is not invertible, because the derivative of moment condition, $\nabla\Psi_N=\frac{1}{N}\sum_{i=1}^{N}x_ix_i'\delta(x_i'\beta-y_i)$, involves delta function and $(\nabla\Psi_N)^{-1}$ is not bounded. Now in the equation (2.15), the Taylor expansion of quantile regression, $\overline{\nabla\Psi_N}$ is invertible, because $\left(\overline{\nabla\Psi_N}\right)^{-1}$ is bounded. In equation (2.15), we keep the term A_{31} even though it is $O_p(N^{-7/6})$ by Lemma 2, because we found that the "expectation" of A_{31} become $O_p(N^{-1})$, which we will discuss in the following Lemma.

Solve for $\widehat{\beta}-\beta_0$ in equation (2.15) to obtain

$$
\begin{aligned}
\widehat{\beta}-\beta_0 &= -\overline{\nabla\Psi_N}^{-1}\Psi_N-\overline{\nabla\Psi_N}^{-1}\left(\nabla\Psi_N-\overline{\nabla\Psi_N}\right)a_{-1/2} \\
&\quad -\frac{1}{2}\overline{\nabla\Psi_N}^{-1}\overline{\nabla^2\Psi_N}\left(a_{-1/2}\otimes a_{-1/2}\right)+o_p\left(N^{-1}\right) \\
&= -Q\Psi_N-QV_Na_{-1/2}-\frac{1}{2}Q\overline{H_2}\left(a_{-1/2}\otimes a_{-1/2}\right)+o_p\left(N^{-1}\right) \qquad (2.16) \\
&\equiv B_1+B_2+B_3+o_p\left(N^{-1}\right),
\end{aligned}
$$

where $H_j=\nabla^j\Psi_N$, for $j=1,2$, $Q=\overline{H_1}^{-1}, V_N=H_1-\overline{H_1}$. Note that multiplying equation (2.16) by $\sqrt{N}$ gives the same as equation (15) of Phillips (1991, p. 457). In order to compute the bias of $\widehat{\beta}$, that is $E\left(\widehat{\beta}-\beta_0\right)$, we now examine the expectations of the three terms B_1, B_2, B_3 in (2.16).

Lemma 3.

(a) $B_1\equiv a_{-1/2}=-Q\Psi_N=O_p\left(N^{-1/2}\right)$, and $E\left(B_1\right)=0$;

(b) $B_2\equiv -QV_Na_{-1/2}=O_p\left(N^{-7/6}\right)$, and $E\left(B_2\right)=O\left(N^{-1}\right)$;

(c) $B_3\equiv -\frac{1}{2}Q\overline{H_2}\left(a_{-1/2}\otimes a_{-1/2}\right)=O_p\left(N^{-1}\right)$, and $E\left(B_3\right)=O\left(N^{-1}\right)$.

Proof: Suppose x_i and u_i are not identically distributed, but independent across $i=1,...,N$. Suppose y_i has conditional density function $f_i\left(y|x\right)$. To simplify the notation, we use $f_i\left(y\right)$ to denote $f_i\left(y|x\right)$.

(a) In equation (2.16), only the first term, B_1, is $O_p\left(N^{-1/2}\right)$, and it should be that $a_{-1/2}=B_1$. Since Ψ_N is the sample moment condition and Q is bounded, then $E\left(B_1\right)=E\left(a_{-1/2}\right)=E\left(-Q\Psi_N\right)=-QE\left(\Psi_N\right)=0$.

(b) By Lemma 1, $A_{31} = \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right) a_{-1/2} = V_N a_{-1/2} = O_p(N^{-7/6})$. Since Q is bounded, then $B_2 \equiv -QV_N a_{-1/2} = O_p\left(N^{-7/6}\right)$. We have

$$H_1 = \nabla_\beta^1 \Psi_N = \nabla_\beta^1 \frac{1}{N}\sum_{i=1}^N s_i = \frac{1}{N}\sum_{i=1}^N \nabla_\beta^1 s_i = \frac{1}{N}\sum_{i=1}^N x_i x_i' \delta(x_i'\beta - y_i),$$

$$\begin{aligned}
\overline{H_1} &= E\nabla_\beta^1 \Psi_N = E\frac{1}{N}\sum_{i=1}^N \left[x_i x_i' \delta(x_i'\beta - y_i)\right] \\
&= \frac{1}{N}\sum_{i=1}^N E\left[x_i x_i' \delta(x_i'\beta - y_i)\right] \\
&= \frac{1}{N}\sum_{i=1}^N E\left[x_i x_i' E\left(\delta(x_i'\beta - y_i)|x_i\right)\right] \\
&= \frac{1}{N}\sum_{i=1}^N E\left[x_i x_i' \int_{-\infty}^{+\infty} \delta(y_i - x_i'\beta_0) f_i(y_i) \mathrm{d}y_i\right] \\
&= \frac{1}{N}\sum_{i=1}^N E\left[x_i x_i' f_i(x_i'\beta)\right],
\end{aligned}$$

$$Q = \left(\overline{H_1}\right)^{-1} = \left(\frac{1}{N}\sum_{i=1}^N E[f_i(x_i'\beta) x_i x_i']\right)^{-1},$$

$$V_N = H_1 - \overline{H_1} = \frac{1}{N}\sum_{i=1}^N x_i x_i' \delta(x_i'\beta - y_i) - \frac{1}{N}\sum_{i=1}^N E[f_i(x_i'\beta) x_i x_i'],$$

Ψ_N, s_i and $a_{-1/2}$ are all $k \times 1$ vectors. H_1, $\overline{H_1}$, Q, and V_N are all $k \times k$ matrixes, H_2, $\overline{H_2}$ and W_N are all $k \times k^2$ matrixes. H_3 and $\overline{H_3}$ are $k \times k^3$ matrixes. Using the the properties in Appendix B.8, we have

$$
\begin{aligned}
E\left(V_N a_{-1/2}\right) &= -E\left[\left(H_1 - \overline{H_1}\right) Q\Psi_N\right] \\
&= -E\left(H_1 Q\Psi_N\right) - E\left(\Psi_N\right) \\
&= -E\left[\frac{1}{N}\sum_{i=1}^{N} x_i x_i' \delta(x_i'\beta - y_i) Q\Psi_N\right] \\
&= -E\left[\frac{1}{N}\sum_{i=1}^{N} x_i x_i' E\left(\delta(x_i'\beta - y_i) Q\Psi_N | x_i\right)\right] \\
&= -\frac{1}{N^2}\sum_{i=1}^{N} E\left[x_i x_i' \int_{-\infty}^{+\infty} \delta(x_i'\beta - y_i) Q(\alpha - \mathbf{1}(y_i < x_i'\beta))(-x_i) f_i(y_i) \mathrm{d}y_i\right] \\
&= -\frac{1}{N^2}\sum_{i=1}^{N} E\left[\begin{array}{c} -x_i x_i' Q x_i \alpha \int_{-\infty}^{+\infty} \delta(x_i'\beta - y_i) f(y_i) \mathrm{d}y_i \\ +x_i x_i' Q x_i \int_{-\infty}^{+\infty} \delta(x_i'\beta - y_i)\phi(x_i'\beta - y_i) f_i(y_i) \mathrm{d}y_i \end{array}\right] \\
&= -\frac{1}{N^2}\sum_{i=1}^{N} E\left[-x_i x_i' Q x_i \alpha f_i(x_i'\beta) + \frac{1}{2} x_i x_i' Q x_i f_i(x_i'\beta)\right] \\
&= -\left(\frac{1}{2} - \alpha\right)\frac{1}{N^2}\sum_{i=1}^{N} E\left[x_i x_i' Q x_i f_i(x_i'\beta)\right].
\end{aligned}
$$

Then, $E\left(B_2\right) = E\left(-QV_N a_{-1/2}\right) = O\left(N^{-1}\right)$.

(c) By Lemma 2, $A_{41} = \frac{1}{2}\overline{\nabla^2\Psi_N}\left(a_{-1/2} \otimes a_{-1/2}\right) = \frac{1}{2}\overline{H_2}\left(a_{-1/2} \otimes a_{-1/2}\right) = O_p\left(N^{-1}\right)$.

Since Q and $\overline{H_2}$ are bounded, then $B_3 = O_p\left(N^{-1}\right)$. We have

$$
H_2 = \nabla_\beta^2 \Psi_N = \frac{1}{N}\sum_{i=1}^{N}\left(x_i x_i'\right) \otimes x_i' \delta^{(1)}(x_i'\beta - y_i),
$$

$$\begin{aligned}
\overline{H_2} &= E\nabla^2_\beta \Psi_N = E\frac{1}{N}\sum_{i=1}^{N}\left[\left(x_i x_i'\right)\otimes x_i'\delta^{(1)}(x_i'\beta - y_i)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left[\left(x_i x_i'\right)\otimes x_i'\delta^{(1)}(x_i'\beta - y_i)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left[\left(x_i x_i'\right)\otimes x_i' E\left(\delta^{(1)}(x_i'\beta - y_i)|x_i\right)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left[\left(x_i x_i'\right)\otimes x_i' \int_{-\infty}^{+\infty}\delta^{(1)}(x_i'\beta - y_i) f_i(y_i)\mathrm{d}y_i\right] \\
&= -\frac{1}{N}\sum_{i=1}^{N} E\left[\left(x_i x_i'\right)\otimes x_i' \int_{-\infty}^{+\infty}\delta^{(1)}(y_i - x_i'\beta) f_i(y_i)\mathrm{d}y_i\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left[\left(x_i x_i'\right)\otimes x_i' \int_{-\infty}^{+\infty}\delta(y_i - x_i'\beta) f_i^{(1)}(y_i)\mathrm{d}y_i\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left[\left(x_i x_i'\right)\otimes x_i' f_i^{(1)}(x_i'\beta)\right],
\end{aligned}$$

$$\begin{aligned}
\overline{a_{-1/2}\otimes a_{-1/2}} &= E\left[\left(Q\Psi_N \otimes Q\Psi_N\right)\right] = E\left[\left(Q\otimes Q\right)\left(\Psi_N\otimes\Psi_N\right)\right] \\
&= \left(Q\otimes Q\right) E\left[\left(\Psi_N\otimes\Psi_N\right)\right] = \frac{1}{N^2}\sum_{i=1}^{N}\left(Q\otimes Q\right) E\left[E\left(s_i\otimes s_i|x_i\right)\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\left(Q\otimes Q\right) E\left[\left(x_i\otimes x_i\right) E\left(\left(\alpha - \mathbf{1}(y_i < x_i'\beta)\right)^2|x_i\right)\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\left(Q\otimes Q\right) E\left(x_i\otimes x_i\right)\left[\left(\alpha-1\right)^2\alpha + \alpha^2\left(1-\alpha\right)\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\alpha(1-\alpha)\left(Q\otimes Q\right) E\left(x_i\otimes x_i\right).
\end{aligned}$$

Then, $E\left(B_3\right) = -\frac{1}{2}Q\overline{H_2}\left(\overline{a_{-1/2}\otimes a_{-1/2}}\right) = O_p\left(N^{-1}\right).$

From equation (2.16), note that the bias of quantile estimators $\widehat{\beta}$ is

$$\begin{aligned}
E\left(\widehat{\beta}-\beta_0\right) &= E\left(B_1\right) + E\left(B_2\right) + E\left(B_3\right) + o\left(N^{-1}\right) \\
&= E\left(-Q\Psi_N\right) + E\left(-QV_N a_{-1/2}\right) + E\left(-\frac{1}{2}Q\overline{H_2}\left(a_{-1/2}\otimes a_{-1/2}\right)\right) + o\left(N^{-1}\right)
\end{aligned} \quad (2.17)$$

Given the above results in Lemma 3, we define the second-order bias of quantile estimators as follows.

Definition 1. Let $E\left(\widehat{\beta}-\beta_0\right)=B(\widehat{\beta})+o\left(N^{-1}\right)$. Then

$$B(\widehat{\beta}) \equiv E\left(-Q\Psi_N\right)+E\left(-QV_N a_{-1/2}\right)+E\left(-\frac{1}{2}Q\overline{H_2}\left(a_{-1/2}\otimes a_{-1/2}\right)\right) \tag{2.18}$$

will be called "the second-order bias of quantile estimators $\widehat{\beta}$ up to $O(N^{-1})$".

Theorem 1. *In the quantile regression model, suppose x_i* and *u_i are not identically distributed, but independent across $i=1,...,N$, the second-order bias, up to $O(N^{-1})$, of the quantile estimators $\widehat{\beta}$ is*

$$\begin{aligned} B(\widehat{\beta}) &= E\left[-QV_N a_{-1/2}-\frac{1}{2}Q\overline{H_2}\left(a_{-1/2}\otimes a_{-1/2}\right)\right] \\ &= \left(\frac{1}{2}-\alpha\right)Q\frac{1}{N^2}\sum_{i=1}^{N}E\left[x_i x_i' Q x_i f_i(0|x_i)\right] \\ &\quad -\frac{\alpha(1-\alpha)}{2}Q\frac{1}{N}\sum_{i=1}^{N}E[(x_i x_i')\otimes x_i' f_i^{(1)}(x_i'\beta)]\frac{1}{N^2}\sum_{i=1}^{N}\left(Q\otimes Q\right)E\left(x_i\otimes x_i\right) \end{aligned} \tag{2.19}$$

where $Q=\left(\frac{1}{N}\sum_{i=1}^{N}E[x_i x_i' f_i(0|x_i)]\right)^{-1}$, $f_i\left(0|x_i\right)$ is the conditional density of u_i given x_i evaluated at $u_i=0$.

Proof: By Lemma 3, the second-order bias of quantile estimators $\widehat{\beta}$ up to $O(N^{-1})$ is

$$\begin{aligned} B(\widehat{\beta}) &= Q\left[-\overline{V_N a_{-1/2}}-\frac{1}{2}\overline{H_2}\left(\overline{a_{-1/2}\otimes a_{-1/2}}\right)\right] \\ &= \left(\frac{1}{2}-\alpha\right)Q\frac{1}{N^2}\sum_{i=1}^{N}E\left[x_i x_i' Q x_i f_i(x_i'\beta)\right] \\ &\quad -\frac{\alpha(1-\alpha)}{2}Q\frac{1}{N}\sum_{i=1}^{N}E[(x_i x_i')\otimes x_i' f_i^{(1)}(x_i'\beta)]\frac{1}{N^2}\sum_{i=1}^{N}\left(Q\otimes Q\right)E\left(x_i\otimes x_i\right), \end{aligned}$$

where $Q = \left(\frac{1}{N}\sum_{i=1}^{N} E[x_i x_i' f_i(x_i'\beta)]\right)^{-1}$. Since the conditional density of y_i given x_i evaluated at $y_i = x_i'\beta$ is the same as the conditional density of u_i given x_i evaluated at $u_i = 0$. If we use $f_i(0|x_i)$ to denote the conditional density of u_i given x_i evaluated at $u_i = 0$, the second-order bias of $\widehat{\beta}$ up to $O(N^{-1})$ can be rewritten as

$$
\begin{aligned}
B(\widehat{\beta}) &= \left(\frac{1}{2}-\alpha\right) Q \frac{1}{N^2}\sum_{i=1}^{N} E\left[x_i x_i' Q x_i f_i(0|x_i)\right] \\
&\quad -\frac{\alpha(1-\alpha)}{2} Q \frac{1}{N^2}\sum_{i=1}^{N} E[(x_i x_i') \otimes x_i' f_i^{(1)}(0|x_i)] \left(Q\otimes Q\right) E\left(x_i \otimes x_i\right),
\end{aligned}
$$

where $Q = \left(\frac{1}{N}\sum_{i=1}^{N} E[x_i x_i' f_i(0|x_i)]\right)^{-1}$.

Corollary 1. When x_i and u_i are i.i.d., the expression of the second-order bias of $\widehat{\beta}$ up to $O(N^{-1})$ can be simplified as

$$
B(\widehat{\beta}) = \frac{1}{N} Q\left[\left(\frac{1}{2}-\alpha\right) E\left(x_i x_i' Q x_i\right) f(0) - \frac{\alpha(1-\alpha)}{2} E[(x_i x_i') \otimes x_i'] f^{(1)}(0) \left(Q\otimes Q\right) E\left(x_i\otimes x_i\right)\right],
$$

where $Q = \left(E\left(x_i x_i'\right) f(0)\right)^{-1}$, and $f(0)$ is the density of u_i evaluated at the $u_i = 0$.

Remark: When x_i and u_i are i.i.d., and $k = 1$, we observe that x_i, Ψ_N, s_i, d, H_1, $\overline{H_1}$, Q, V_N, H_2, $\overline{H_2}$, W_N, H_3, $\overline{H_3}$ are all scalars, and the second-order bias of $\widehat{\beta}$ up to $O(N^{-1})$ can be rewritten as

$$
B(\widehat{\beta}) = \frac{1}{N}\left(\frac{1}{2}-\alpha\right)\frac{E\left(x_i^3\right)}{\left[E\left(x_i^2\right)\right]^2 f(0)} - \frac{1}{N}\frac{\alpha(1-\alpha)}{2}\frac{E\left(x_i^3\right) f^{(1)}(0)}{\left[E\left(x_i^2\right)\right]^2 \left[f(0)\right]^3}.
$$

The quantile estimator $\widehat{\beta}$ is unbiased if x_i follow a symmetric distribution. If u_i follow a symmetric distribution, the median estimator is unbiased. The second-order bias of $\widehat{\beta}$ is larger at the tails of a distribution. The second-order bias of $\widehat{\beta}$ goes to zero as the sample size goes to infinity.

2.4 The MSE of Quantile Estimators

To derive the MSE up to $O\left(N^{-2}\right)$, we take the high order Taylor's expansion as

$$
\begin{aligned}
0 &= \Psi_N + \overline{\nabla\Psi_N}(\widehat{\beta}-\beta_0) + \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)(\widehat{\beta}-\beta_0) + \frac{1}{2}\nabla^2\Psi_N\left[(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\right] \\
&\quad + \frac{1}{6}\nabla^3\Psi_N\left[(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\right] + o_p\left(N^{-3/2}\right) \\
&\equiv A_1 + A_2 + A_3 + A_4 + A_5 + o_p\left(N^{-3/2}\right). \qquad (2.20)
\end{aligned}
$$

Our goal is to obtain the expression of the MSE $E\left(\widehat{\beta}-\beta_0\right)^2$ *up to* the order $O\left(N^{-2}\right)$, therefore, we first need to obtain the stochastic expression of $\widehat{\beta}-\beta_0$ up to the order of $O\left(N^{-3/2}\right)$. By Lemma 3, $\widehat{\beta}-\beta_0 = B_1 + B_2 + B_3 + o_p\left(N^{-1}\right)$, where $B_1 = a_{-1/2} = O_p\left(N^{-1/2}\right)$, $B_2 = O_p\left(N^{-7/6}\right)$, $B_3 = O_p\left(N^{-1}\right)$. Let $B_3 \equiv a_{-1}$, then $\widehat{\beta}-\beta_0 = a_{-1/2} + a_{-1} + O_p\left(N^{-7/6}\right)$. We discuss A_3, A_4, A_5 in equation (2.20) in the following lemmas.

Lemma 4. $A_{32} = \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)[B_2 + B_3] + o_p\left(N^{-3/2}\right)$.

Proof: According to Phillips (1991), $\nabla\Psi_N - \overline{\nabla\Psi_N} = O_p\left(N^{-1/3}\right)$. By Lemma 1, $A_{32} = \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)\left[(\widehat{\beta}-\beta_0) - a_{-1/2}\right]$. By Lemma 3, $\widehat{\beta}-\beta_0 = B_1 + B_2 + B_3 + o_p\left(N^{-1}\right)$. Since $B_1 = a_{-1/2} = O_p\left(N^{-1/2}\right)$, $B_2 = O_p\left(N^{-7/6}\right)$ and $B_3 = a_{-1} = O_p\left(N^{-1}\right)$ are not of zero mean, then we have

$$
\begin{aligned}
A_{32} &= \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)\left[(\widehat{\beta}-\beta_0) - a_{-1/2}\right] \\
&= \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)\left[B_2 + B_3 + o_p\left(N^{-1}\right)\right] \\
&= \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)[B_2 + B_3] + o_p\left(N^{-3/2}\right),
\end{aligned}
$$

where $\left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)B_2 = O_p\left(N^{-4/3}\right)$, and $\left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)B_3 = O_p\left(N^{-3/2}\right)$.

Lemma 5.

(a) $A_{41} = \frac{1}{2}\overline{\nabla^2 \Psi_N}\left(a_{-1/2} \otimes a_{-1/2}\right) + \frac{1}{2}\overline{\nabla^2 \Psi_N}\left[\left(a_{-1/2} \otimes a_{-1}\right) + \left(a_{-1} \otimes a_{-1/2}\right)\right] + o_p\left(N^{-3/2}\right)$,

(b) $A_{42} = \frac{1}{2}\left(\nabla^2 \Psi_N - \overline{\nabla^2 \Psi_N}\right)\left(a_{-1/2} \otimes a_{-1/2}\right) + o_p\left(N^{-3/2}\right)$.

Proof:

(a) By Lemma 3, A_{41} can be written as

$$\begin{aligned}
A_{41} &= \frac{1}{2}\overline{\nabla^2 \Psi_N}\left[(\widehat{\beta} - \beta_0) \otimes (\widehat{\beta} - \beta_0)\right] \\
&= \frac{1}{2}\overline{\nabla^2 \Psi_N}\left\{\left[(\widehat{\beta} - \beta_0) - a_{-1/2} + a_{-1/2}\right] \otimes \left[(\widehat{\beta} - \beta_0) - a_{-1/2} + a_{-1/2}\right]\right\} \\
&= \frac{1}{2}\overline{\nabla^2 \Psi_N}\left(a_{-1/2} \otimes a_{-1/2}\right) \\
&\quad + \frac{1}{2}\overline{\nabla^2 \Psi_N}\left(a_{-1/2} \otimes \left[(\widehat{\beta} - \beta_0) - a_{-1/2}\right]\right) \\
&\quad + \frac{1}{2}\overline{\nabla^2 \Psi_N}\left(\left[(\widehat{\beta} - \beta_0) - a_{-1/2}\right] \otimes a_{-1/2}\right) \\
&\quad + \frac{1}{2}\overline{\nabla^2 \Psi_N}\left(\left[(\widehat{\beta} - \beta_0) - a_{-1/2}\right] \otimes \left[(\widehat{\beta} - \beta_0) - a_{-1/2}\right]\right) \\
&= \frac{1}{2}\overline{\nabla^2 \Psi_N}\left(a_{-1/2} \otimes a_{-1/2}\right) + \frac{1}{2}\overline{\nabla^2 \Psi_N}\left[a_{-1/2} \otimes \left(a_{-1} + O_p\left(N^{-7/6}\right)\right)\right] \\
&\quad + \frac{1}{2}\overline{\nabla^2 \Psi_N}\left[\left(a_{-1} + O_p\left(N^{-7/6}\right)\right) \otimes a_{-1/2}\right] + O_p\left(N^{-2}\right) \\
&= \frac{1}{2}\overline{\nabla^2 \Psi_N}\left(a_{-1/2} \otimes a_{-1/2}\right) \\
&\quad + \frac{1}{2}\overline{\nabla^2 \Psi_N}\left[\left(a_{-1/2} \otimes a_{-1}\right) + \left(a_{-1} \otimes a_{-1/2}\right)\right] + o_p\left(N^{-3/2}\right).
\end{aligned}$$

(b)

$$\begin{aligned}
A_{42} &= \frac{1}{2}\left(\nabla^2 \Psi_N - \overline{\nabla^2 \Psi_N}\right)\left[(\widehat{\beta} - \beta_0) \otimes (\widehat{\beta} - \beta_0)\right] \\
&= \frac{1}{2}\left(\nabla^2 \Psi_N - \overline{\nabla^2 \Psi_N}\right)\left[\left(a_{-1/2} + O_p\left(N^{-1}\right)\right) \otimes \left(a_{-1/2} + O_p\left(N^{-1}\right)\right)\right] \\
&= \frac{1}{2}\left(\nabla^2 \Psi_N - \overline{\nabla^2 \Psi_N}\right)\left(a_{-1/2} \otimes a_{-1/2}\right) + o_p\left(N^{-3/2}\right)
\end{aligned}$$

Since $\nabla^2\Psi_N - \overline{\nabla^2\Psi_N}$ is greater than $O_p\left(N^{-1/2}\right)$, then the first term in A_{42} is greater than $O_p(N^{-3/2})$.

Lemma 6. Let

$$\begin{aligned}
A_5 &= \frac{1}{6}\nabla^3\Psi_N\left[(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\right] \\
&= \frac{1}{6}\overline{\nabla^3\Psi_N}\left[(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\right] \\
&\quad +\frac{1}{6}\left(\nabla^3\Psi_N-\overline{\nabla^3\Psi_N}\right)\left[(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\right] \\
&= \frac{1}{6}\overline{\nabla^3\Psi_N}\left[(a_{-1/2}+O_p\left(N^{-1}\right))\otimes(a_{-1/2}+O_p\left(N^{-1}\right))\otimes(a_{-1/2}+O_p\left(N^{-1}\right))\right] \\
&\quad +\frac{1}{6}\left(\nabla^3\Psi_N-\overline{\nabla^3\Psi_N}\right)\left[(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\right] \\
&= \frac{1}{6}\overline{\nabla^3\Psi_N}\left[a_{-1/2}\otimes a_{-1/2}\otimes a_{-1/2}\right]+o_p\left(N^{-3/2}\right) \\
&= A_{51}+o_p\left(N^{-3/2}\right),
\end{aligned}$$

Proof: Since $\nabla^3\Psi_N - \overline{\nabla^3\Psi_N}$ is smaller than $O_p(1)$, then the results in Lemma 6 follows.

In Lemma 4, 5, and 6, we have discusssed each term in equation (2.20). Now the equation (2.20) can be written as

$$\begin{aligned}
0 &= \Psi_N+\overline{\nabla\Psi_N}(\widehat{\beta}-\beta_0)+\left(\nabla\Psi_N-\overline{\nabla\Psi_N}\right)\left(a_{-1/2}+a_{-1}\right) \\
&\quad +\frac{1}{2}\overline{\nabla^2\Psi_N}\left(a_{-1/2}\otimes a_{-1/2}\right)+\frac{1}{2}\overline{\nabla^2\Psi_N}\left[\left(a_{-1/2}\otimes a_{-1}\right)+\left(a_{-1}\otimes a_{-1/2}\right)\right] \\
&\quad +\frac{1}{2}\left(\nabla^2\Psi_N-\overline{\nabla^2\Psi_N}\right)\left(a_{-1/2}\otimes a_{-1/2}\right) \\
&\quad +\frac{1}{6}\overline{\nabla^3\Psi_N}\left[a_{-1/2}\otimes a_{-1/2}\otimes a_{-1/2}\right]+o_p\left(N^{-3/2}\right). \qquad (2.21)
\end{aligned}$$

The equation (2.21) is invertible as a higher-order Taylor expansion of quantile regression, because $\left(\overline{\nabla\Psi_N}\right)^{-1}$ is bounded. Given the results in Lemma 3(b), we have $B_2 \equiv$

$-QV_N a_{-1/2} = O_p\left(N^{-7/6}\right)$, then $B_2 B_2' = O_p\left(N^{-7/3}\right)$. However, we found that $E\left(B_3 B_3'\right) = O\left(N^{-2}\right)$, which we will discuss in the following Lemma. Solve for $\widehat{\beta} - \beta_0$ in equation (2.15) to obtain

$$\begin{aligned}
\widehat{\beta} - \beta_0 &= -\overline{\nabla\Psi_N}^{-1}\Psi_N - \overline{\nabla\Psi_N}^{-1}\left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)\left(a_{-1/2} + a_{-1}\right) \\
&\quad -\frac{1}{2}\overline{\nabla\Psi_N}^{-1}\overline{\nabla^2\Psi_N}\left(a_{-1/2}\otimes a_{-1/2}\right) \\
&\quad -\frac{1}{2}\overline{\nabla\Psi_N}^{-1}\overline{\nabla^2\Psi_N}\left[\left(a_{-1/2}\otimes a_{-1}\right) + \left(a_{-1}\otimes a_{-1/2}\right)\right] \\
&\quad -\frac{1}{2}\overline{\nabla\Psi_N}^{-1}\left(\nabla^2\Psi_N - \overline{\nabla^2\Psi_N}\right)\left(a_{-1/2}\otimes a_{-1/2}\right) \\
&\quad -\frac{1}{6}\overline{\nabla\Psi_N}^{-1}\overline{\nabla^3\Psi_N}\left[a_{-1/2}\otimes a_{-1/2}\otimes a_{-1/2}\right] + o_p\left(N^{-3/2}\right) \\
&= \left\{-Q\Psi_N\right\} + \left\{-\frac{1}{2}Q\overline{H_2}\left(a_{-1/2}\otimes a_{-1/2}\right)\right\} + \left\{-QV_N a_{-1/2}\right\} \\
&\quad + \left\{-QV_N a_{-1} - \frac{1}{2}QW_N\left(a_{-1/2}\otimes a_{-1/2}\right)\right\} \\
&\quad + \left\{-\frac{1}{2}Q\overline{H_2}\left[\left(a_{-1/2}\otimes a_{-1}\right) + \left(a_{-1}\otimes a_{-1/2}\right)\right] - \frac{1}{6}Q\overline{H_3}\left[a_{-1/2}\otimes a_{-1/2}\otimes a_{-1/2}\right]\right\} \\
&\quad + o_p\left(N^{-3/2}\right), \\
&\equiv B_1 + B_2 + B_3 + B_4 + B_5 + o_p\left(N^{-3/2}\right), \qquad (2.22)
\end{aligned}$$

where $H_j = \nabla^j\Psi_N$, for $j = 1, 2, 3$, $Q = \overline{H_1}^{-1}$, $V_N = H_1 - \overline{H_1}$, $W_N = H_2 - \overline{H_2}$. Note that the equation (2.22) is same as the expression in RSU (1996 p. 390 Eq. A.17).

Lemma 7.

(a) $B_1 = O_p\left(N^{-1/2}\right)$, $B_2 = O_p\left(N^{-1}\right)$, $B_3 = O_p\left(N^{-7/6}\right)$, $B_4 = O_p\left(N^{-4/3}\right)$, and $B_5 = O_p\left(N^{-3/2}\right)$,

(b) $B_1 B_1' = O_p\left(N^{-1}\right)$, and $E\left(B_1 B_1'\right) = O\left(N^{-1}\right)$,

(c) $B_1 B_2' = B_2 B_1' = O_p\left(N^{-3/2}\right)$, and $E\left(B_1 B_2'\right) = E\left(B_2 B_1'\right) = O\left(N^{-2}\right)$,

(d) $B_1B_3' = B_3B_1' = O_p\left(N^{-5/3}\right)$, and $E\left(B_1B_3'\right) = E\left(B_3B_1'\right) = O\left(N^{-2}\right)$,

(e) $B_1B_4' = B_4B_1' = O_p\left(N^{-11/6}\right)$, and $E\left(B_1B_4'\right) = E\left(B_4B_1'\right) = O\left(N^{-2}\right)$,

(f) $B_2B_2' = O_p\left(N^{-2}\right)$, and $E\left(B_2B_2'\right) = O\left(N^{-2}\right)$,

(g) $B_1B_5' = B_5B_1' = O_p\left(N^{-2}\right)$, and $E\left(B_1B_5'\right) = E\left(B_5B_1'\right) = O\left(N^{-2}\right)$,

(h) $B_3B_3' = O_p\left(N^{-7/3}\right)$, and $E\left(B_3B_3'\right) = O\left(N^{-2}\right)$.

Proof: Suppose $k = 1$, x_i and u_i are not identically distributed, but independent across $i = 1, ..., N$.. Let $d = Q\Psi_N = \frac{1}{N}\sum_{i=1}^{N} d_i$, $d_i = Qs_i$, $V_N = \frac{1}{N}\sum_{i=1}^{N}\left(\nabla^1 s_i - \overline{\nabla^1 s_i}\right) = \frac{1}{N}\sum_{i=1}^{N} V_i$, $W_N = \frac{1}{N}\sum_{i=1}^{N}\left(\nabla^2 s_i - \overline{\nabla^2 s_i}\right) = \frac{1}{N}\sum_{i=1}^{N} W_i$, then d_i, V_i, and W_i are not identically distributed, but independent across $i = 1, ..., N$. The expected values of $V_i d_j$, $W_i d_j$, and $V_i W_j$ are all zero for $i \neq j$. Then we have

$$E\left(B_1B_1'\right) = \overline{d_i^2},$$

$$E\left(B_1B_2' + B_2B_1'\right) = Q\overline{H_2}\overline{d_i^3},$$

$$E\left(B_1B_3' + B_3B_1'\right) = -2Q\overline{V_i d_i^2},$$

$$\begin{aligned} E\left(B_1B_4' + B_4B_1'\right) &= 2Q^2\overline{V_i^2}\,\overline{d_i^2} + 4Q^2\overline{V_iV_jd_id_j} - 9Q^2\overline{H_2}\overline{V_id_id_j^2} + 3Q\overline{W_id_id_j^2} \\ &= 2Q^2\overline{V_i^2}\,\overline{d_i^2} + 4Q^2\overline{V_id_i}^2 - 9Q^2\overline{H_2}\;\overline{V_id_i}\overline{d_i^2} + 3Q\overline{W_id_i}\overline{d_i^2} \end{aligned}$$

$$E\left(B_2B_2'\right) = 2Q^2\overline{V_iV_jd_1d_2} + Q^2\overline{V_i^2}\,\overline{d_i^2} = 2Q^2\overline{V_id_i}^2 + Q^2\overline{V_i^2}\,\overline{d_i^2},$$

$$\begin{aligned} E\left(B_1B_5' + B_5B_1'\right) &= 3Q^2\overline{H_2}^2\overline{d_i^2d_j^2} - Q\overline{H_3}\overline{d_i^2d_j^2} \\ &= 3Q^2\overline{H_2}^2\overline{d_i^2}^2 - Q\overline{H_3}\overline{d_i^2}^2 \end{aligned}$$

$$E\left(B_3B_3'\right) = \frac{3}{4}Q^2\overline{H_2}^2\overline{d_i^2d_j^2} - 3Q^2\overline{H_2}\overline{V_id_id_j^2} = \frac{3}{4}Q^2\overline{H_2}^2\overline{d_i^2}^2 - 3Q^2\overline{H_2}\ \overline{V_id_i}\overline{d_i^2},$$

$$\begin{aligned}
\overline{d_i^2} &= Q^2E(\Psi_N^2) \\
&= \frac{1}{N^2}\sum_{i=1}^{N}Q^2E\left[x_i^2(\alpha - \mathbf{1}(y_i < x_i'\beta))^2\right] \\
&= \frac{1}{N}Q^2\left[(\alpha-1)^2\alpha + \alpha^2(1-\alpha)\right]E\left(x_i^2\right) \\
&= \frac{1}{N}\alpha(1-\alpha)Q^2E\left(x_i^2\right),
\end{aligned}$$

$$\begin{aligned}
\overline{d_i^3} &= Q^3E(\Psi_N^3) \\
&= -\frac{1}{N^3}\sum_{i=1}^{N}Q^3E\left[x_i^3(\alpha - \mathbf{1}(y_i < x_i'\beta))^3\right] \\
&= -\frac{1}{N^2}Q^3\left[(\alpha-1)^3\alpha + \alpha^3(1-\alpha)\right]E\left(x_i^3\right) \\
&= -\frac{1}{N^2}\alpha(1-\alpha)(2\alpha-1)Q^3E\left(x_i^3\right),
\end{aligned}$$

$$\begin{aligned}
\overline{V_i^2} &= E\left[\left(H_1 - \overline{H_1}\right)^2\right] \\
&= E\left[H_1^2 - 2H_1\overline{H_1} + \overline{H_1}^2\right] \\
&= E\left(H_1^2\right) - 2\overline{H_1}^2 + \overline{H_1}^2 \\
&= E\left(H_1^2\right) - \overline{H_1}^2 \\
&= \frac{1}{N^2}\sum_{i=1}^{N}E\left[x_i^4\left(\delta(x_i'\beta - y_i)\right)^2\right] - \frac{1}{N^2}\sum_{i=1}^{N}\left(E\left[x_i^2f_i(x_i'\beta)\right]\right)^2 \\
&= \frac{1}{N^2}\sum_{i=1}^{N}E\left[x_i^4\int_{-\infty}^{+\infty}\left(\delta(x_i'\beta - y_i)\right)^2f_i(y_i)\mathrm{d}y_i\right] - \frac{1}{N^2}\sum_{i=1}^{N}\left(E\left[x_i^2f_i(x_i'\beta)\right]\right)^2,
\end{aligned}$$

$$\overline{V_id_i}^2 = \frac{1}{N^3}\sum_{i=1}^{N}\left(\frac{1}{2}-\alpha\right)^2Q^2\left(E\left[x_i^3f(x_i'\beta)\right]\right)^2,$$

$$\overline{V_id_i^2} = E\left[\left(H_1 - \overline{H_1}\right)Q^2\Psi_N^2\right] = Q^2E\left(H_1\Psi_N^2\right) - QE\left(\Psi_N^2\right),$$

$$
\begin{aligned}
E(H_1\Psi_N^2) &= E\left[\left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\delta(x_i'\beta - y_i)\right)\Psi_N^2\right] \\
&= \frac{1}{N^3}\sum_{i=1}^{N} E\left[x_i^2 E\left(\delta(x_i'\beta - y_i)s_i^2|x_i\right)\right] \\
&= \frac{1}{N^3}\sum_{i=1}^{N} E\left[x_i^4\int_{-\infty}^{+\infty}\delta(x_i\beta - y_i)(\alpha - \mathbf{1}(y_i < x_i'\beta))^2 f_i(y_i)\mathrm{d}y\right] \\
&= \frac{1}{N^3}\sum_{i=1}^{N} E\left[\alpha^2 x_i^4\int_{-\infty}^{+\infty}\delta(x_i\beta - y_i)f_i(y_i)\mathrm{d}y_i\right] \\
&\quad + \frac{1}{N^3}\sum_{i=1}^{N} E\left[(1-2\alpha)x_i^4\int_{-\infty}^{+\infty}\delta(x_i\beta - y_i)\phi(x_i\beta - y_i)f_i(y_i)\mathrm{d}y_i\right] \\
&= \frac{1}{N^3}\sum_{i=1}^{N} E\left[\alpha^2 x_i^4 f_i(x_i\beta)\right] + \frac{1}{N^2}\sum_{i=1}^{N} E\left[(1-2\alpha)\frac{1}{2}x_1^4 f_i(x_i\beta)\right] \\
&= \frac{1}{N^3}\sum_{i=1}^{N}\left(\alpha^2 - \alpha + \frac{1}{2}\right)E\left[x_i^4 f_i(x_i'\beta)\right].
\end{aligned}
$$

$$
H_3 = \nabla_\beta^3\Psi_N = \frac{1}{N}\sum_{i=1}^{N} x_i^4\delta^{(2)}(x_i'\beta - y_i),
$$

$$
\begin{aligned}
\overline{H_3} &= E\nabla_\beta^3\Psi_N = E\frac{1}{N}\sum_{i=1}^{N}\left[x_i^4\delta^{(2)}(x_i'\beta - y_i)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left[x_i^4\delta^{(2)}(x_i'\beta - y_i)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left[x_i^4 E\left(\delta^{(2)}(x_i'\beta - y_i)|x_i\right)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left[x_i^4\int_{-\infty}^{+\infty}\delta^{(2)}(x_i'\beta - y_i)f_i(y_i)\mathrm{d}y_i\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left[x_i^4\int_{-\infty}^{+\infty}\delta(x_i'\beta - y_i)f_i^{(2)}(y_i)\mathrm{d}y_i\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E[x_i^4 f_i^{(2)}(x_i'\beta)],
\end{aligned}
$$

$$
W_i = H_2 - \overline{H_2} = \frac{1}{N}\sum_{i=1}^{N} x_i^3\delta^{(1)}(x_i'\beta - y_i) - \frac{1}{N}\sum_{i=1}^{N} E[x_i^3 f_i^{(1)}(x_i'\beta)],
$$

$$
\begin{aligned}
\overline{W_i d_i} &= E\left[\left(H_2 - \overline{H_2}\right) Q\Psi_N\right] \\
&= QE\left(H_2\Psi_N\right) - Q\overline{H_2}E\left(\Psi_N\right) \\
&= \frac{1}{N}\sum_{i=1}^{N} QE\left[x_i^3 \delta^{(1)}(x_i'\beta - y_i)\Psi_N\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N} QE\left[x_i^3 E\left(\delta^{(1)}(x_i'\beta - y_i)\left(\alpha - \mathbf{1}(y_i < x_i'\beta)\right)(-x_i)|x_i\right)\right] \\
&= -\frac{1}{N^2}\sum_{i=1}^{N} QE\left[x_i^4 E\left(\delta^{(1)}(x_i'\beta - y_i)\left(\alpha - \phi\left(x_i\beta - y_i\right)\right)|x_i\right)\right] \\
&= -\frac{1}{N^2}\sum_{i=1}^{N} \alpha QE\left[x_i^4 \int_{-\infty}^{+\infty} \delta^{(1)}(x_i'\beta - y_i) f(y_i)\mathrm{d}y_i\right] \\
&\quad + \frac{1}{N^2}\sum_{i=1}^{N} QE\left[x_i^4 \int_{-\infty}^{+\infty} \delta^{(1)}(x_i'\beta - y_i)\phi\left(x_i\beta - y_i\right) f(y_i)\mathrm{d}y_i\right] \\
&= -\frac{1}{N^2}\sum_{i=1}^{N} \alpha QE\left[x_i^4 f^{(1)}(x_i'\beta)\right] + \frac{1}{N^2}\sum_{i=1}^{N} QE\left[-x_i^4 \int_{-\infty}^{+\infty} \left(\delta(x_i'\beta - y_i)\right)^2 f(y_i)\mathrm{d}y_i\right] \\
&\quad + \frac{1}{N^2}\sum_{i=1}^{N} QE\left[x_i^4 \int_{-\infty}^{+\infty} \delta(x_i'\beta - y_i)\phi\left(x_i\beta - y_i\right) f^{(1)}(y_i)\mathrm{d}y_i\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\left(\frac{1}{2} - \alpha\right) QE\left[x_i^4 f^{(1)}(x_i'\beta)\right] - \frac{1}{N^2}\sum_{i=1}^{N} QE\left[x_i^4 \int_{-\infty}^{+\infty} \left(\delta(x_i'\beta - y_i)\right)^2 f(y_i)\mathrm{d}y_i\right].
\end{aligned}
$$

From equation (2.22), note that the MSE of quantile estimator $\widehat{\beta}$ is

$$
\begin{aligned}
E\left(\widehat{\beta} - \beta_0\right)^2 &= E\left(B_1 B_1'\right) + E\left(B_1 B_2' + B_2 B_1'\right) + E\left(B_1 B_3' + B_3 B_1'\right) \\
&\quad + E\left(B_1 B_4' + B_4 B_1'\right) + E\left(B_2 B_2'\right) + E\left(B_1 B_5' + B_5 B_1'\right) \\
&\quad + E\left(B_3 B_3'\right) + o_p\left(N^{-2}\right). \qquad (2.23)
\end{aligned}
$$

Given the above results in Lemma 7, we define the MSE of quantile estimators as follows.

Definition 2. Let $E\left(\widehat{\beta}-\beta_0\right)^2 = M(\widehat{\beta}) + o_p\left(N^{-2}\right)$. Then

$$
\begin{aligned}
M(\widehat{\beta}) \;=\;& E\left(B_1B_1'\right) + E\left(B_1B_2' + B_2B_1'\right) + E\left(B_1B_3' + B_3B_1'\right) \\
& + E\left(B_1B_4' + B_4B_1'\right) + E\left(B_2B_2'\right) + E\left(B_1B_5' + B_5B_1'\right) + E\left(B_3B_3'\right), \qquad (2.24)
\end{aligned}
$$

will be call "the MSE of quantile estimators $\widehat{\beta}$ up to $O(N^{-2})$".

Theorem 2. *In the quantile regression model, suppose* x_i and u_i *are not identically distributed, but independent across* $i = 1, ..., N$, *when* $k = 1$, *the MSE up to* $O(N^{-2})$, *of the quantile estimator* $\widehat{\beta}$ *is*

$$
\begin{aligned}
M(\widehat{\beta}) \;=\;& \frac{1}{N}\alpha(1-\alpha)Q^2E\left(x_i^2\right) - 2\frac{1}{N^3}\sum_{i=1}^{N}Q^3\left(\alpha^2-\alpha+\frac{1}{2}\right)E\left[x_i^4 f_i(0|x_i)\right] \\
& -\frac{1}{N^3}\sum_{i=1}^{N}\alpha(1-\alpha)Q^2E\left(x_i^2\right) \\
& -\frac{1}{N^3}\sum_{i=1}^{N}\alpha(1-\alpha)(2\alpha-1)Q^4E\left[x_i^3 {f_i}^{(1)}(0|x_i)\right]E\left(x_i^3\right) \\
& +6\frac{1}{N^3}\sum_{i=1}^{N}\left(\frac{1}{2}-\alpha\right)^2Q^4\left(E\left[x_i^3 f_i(0|x_i)\right]\right)^2 \\
& +3\frac{1}{N^3}\sum_{i=1}^{N}\alpha(1-\alpha)Q^4\left(\frac{1}{2}-\alpha\right)E\left[x_i^4 {f_i}^{(1)}(0|x_i)\right]E\left(x_i^2\right) \\
& -12\frac{1}{N^3}\sum_{i=1}^{N}\left(\frac{1}{2}-\alpha\right)\alpha(1-\alpha)Q^5E[x_i^3 {f_i}^{(1)}(0|x_i)]E\left[x_i^3 f_i(0|x_i)\right]E\left(x_i^2\right) \\
& +\frac{15}{4}\frac{1}{N^3}\sum_{i=1}^{N}\alpha^2(1-\alpha)^2Q^6\left(E\left[x_i^3 {f_i}^{(1)}(0|x_i)\right]\right)^2\left(E\left(x_i^2\right)\right)^2 \\
& -\frac{1}{N^3}\sum_{i=1}^{N}\alpha^2(1-\alpha)^2Q^5E[x_i^4 {f_i}^{(2)}(0|x_i)]\left(E\left(x_i^2\right)\right)^2, \qquad (2.25)
\end{aligned}
$$

where $Q = \left(\frac{1}{N}\sum_{i=1}^{N}E[x_i^2 f_i(0|x_i)]\right)^{-1}$.

Proof: For simplicity, we derive the MSE of quantile estimator up to $O(N^{-2})$ for $k = 1$. It follows the same procedure obviously to obtain the MSE for $k > 1$. Suppose x_i and u_i are not identically distributed, but independent across $i = 1, ..., N$. Then s_i, d_i, V_i,and W_i are all independent across i. By the results of Lemma 9, the MSE of the quantile estimator $\widehat{\beta}$ up to $O\left(N^{-2}\right)$ can be written as

$$\begin{aligned} M(\widehat{\beta}) &= \overline{d_i^2} - 2Q\left[\overline{V_i d_i^2} - \frac{1}{2}\overline{H_2}\,\overline{d_i^3}\right] + 6Q^2\overline{V_i d_i}^2 + 3Q^2\overline{V_i^2}\,\overline{d_i^2} \\ &\quad + 3Q\overline{W_i d_i}\,\overline{d_i^2} - 12Q^2\overline{H_2}\;\overline{V_i d_i}\,\overline{d_i^2} + \frac{15}{4}Q^2\overline{H_2}^2\,\overline{d_i^2}^2 - Q\overline{H_3}\,\overline{d_i^2}^2, \end{aligned}$$

Since the conditional density of y_i given x_i evaluated at $y_i = x_i'\beta$ is the same as the conditional density of u_i given x_i evaluated at $u_i = 0$. We use $f_i\left(0|x_i\right)$ to denote the conditional density of u_i given x_i evaluated at $u_i = 0$. The above results complete the proof of the Theorem 2.

Corollary 2.1. The MSE of the quantile estimator $\widehat{\beta}$ up to $O\left(N^{-1}\right)$ equals the asymptotic variance of $\widehat{\beta}$.

Proof: From Theorem 2, we observe that the MSE of $\widehat{\beta}$ up to $O(N^{-1})$ for quantile estimator for i.i.d. case when $k = 1$ can be simplified as

$$MSE(\widehat{\beta}) = \overline{d_i^2} = \frac{1}{N}\alpha(1-\alpha)Q^2 E\left(x_i^2\right).$$

The asymptotic distribution of the quantile regression estimator when the α-quantile is linear in x_i, is given by equation (2.8). We can prove that V_α, the asymptotic variance of $\widehat{\beta}$ equals N times the MSE of $\widehat{\beta}$ up to $O(N^{-1})$. Since $\mathbf{1}(u_i < 0)$ is Bernoulli with mean α and variance $\alpha(1-\alpha)$, then we can have

$$\begin{aligned}
E\left[\Psi_N(\beta)\Psi_N(\beta)'\right] &= E\left[\left(\frac{1}{N}\sum_{i=1}^{N}s_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}s_i'\right)\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}E\left[s_is_i'\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}E\left[(\alpha-\mathbf{1}(u_i<0))^2x_ix_i'\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}E\left[x_ix_i'E[(\alpha-\mathbf{1}(u_i<0))^2|x_i]\right] \\
&= \frac{\alpha(1-\alpha)}{N^2}\sum_{i=1}^{N}E(x_ix_i').
\end{aligned}$$

The MSE of $\widehat{\beta}$ up to $O(N^{-1})$ can be derived by substituting the result above,

$$\begin{aligned}
MSE(\widehat{\beta}) &= E(a_{-1/2}a_{-1/2}') = E(Q\Psi_N\Psi_N'Q) = QE\left[\Psi_N(\beta)\Psi_N(\beta)'\right]Q \\
&= \left[\frac{1}{N}\sum_{i=1}^{N}E(f(0|x_i)x_ix_i')\right]^{-1}\frac{\alpha(1-\alpha)}{N}E(x_ix_i')\left[\frac{1}{N}\sum_{i=1}^{N}E(f(0|x_i)x_ix_i')^{-1}\right] \\
&= \frac{V_\alpha}{N},
\end{aligned}$$

The asymptotic variance

$$V_\alpha = N\times MSE(\widehat{\beta}) = \alpha(1-\alpha)[\frac{1}{N}\sum_{i=1}^{N}E(f(0|x_i)x_ix_i')]^{-1}E(x_ix_i')[\frac{1}{N}\sum_{i=1}^{N}E(f(0|x_i)x_ix_i')]^{-1}.$$

Corollary 2.1. When x_i and u_i are i.i.d., and $k=1$, the expression of the MSE of $\widehat{\beta}$ up to $O(N^{-2})$ can be simplified as

$$\begin{aligned} M(\widehat{\beta}) &= \frac{1}{N}\alpha(1-\alpha)Q^2 E\left(x_i^2\right) - 2\frac{1}{N^2}Q^3\left(\alpha^2-\alpha+\frac{1}{2}\right)E\left(x_i^4\right)f(0) - \frac{1}{N^2}\alpha(1-\alpha)Q^2E\left(x_i^2\right) \\ &+5\frac{1}{N^2}\alpha(1-\alpha)(2\alpha-1)Q^4 f^{(1)}(0)\left(E\left(x_i^3\right)\right)^2 + 6\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)^2 Q^4 f(0)^2\left(E\left(x_i^3\right)\right)^2 \\ &+3\frac{1}{N^2}\alpha(1-\alpha)Q^4\left(\frac{1}{2}-\alpha\right)E\left(x_i^4\right)f^{(1)}(0)E\left(x_i^2\right) \\ &+\frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2Q^6 f^{(1)}(0)\left(E\left(x_i^3\right)\right)^2\left(E\left(x_i^2\right)\right)^2 \\ &-\frac{1}{N^2}\alpha^2(1-\alpha)^2Q^5 f^{(2)}(0)E\left(x_i^4\right)\left(E\left(x_i^2\right)\right)^2, \end{aligned}$$

where $Q = \left(E\left(x_i^2\right)f(0)\right)^{-1}$. When x_i and u_i are i.i.d., the asymptotic variance of $\widehat{\beta}$ is $V_\alpha = N \times MSE(\widehat{\beta}) = \alpha(1-\alpha)E(x_i x_i')/(f(0))^2$.

2.5 Illustrations

In this section, we consider three special cases of the general results on the conditional quantile regression from the previous section: namely, (i) the unconditional quantile estimation, (ii) the conditional quantile regression with a binary independent variable, and (iii) the instrumental variable quantile regression (IVQR). For these cases we illustrate the second-order bias and MSE with several different distributions to highlight the merits of using the higher order terms in bias and MSE.

2.5.1 Unconditional Quantile Estimator

We consider a special case of the model with $x_i = 1$, i.e., the model without any covariate, which gives the unconditional quantile estimator.

Proposition 3. *In the quantile regression model with $x_i = 1$, the second-order bias up to*

$O(N^{-1})$, *of the unconditional quantile estimators* $\widehat{\beta}$ *is*

$$B(\widehat{\beta}) = \frac{1}{N}\left(\frac{1}{2}-\alpha\right)Q - \frac{1}{N}\frac{\alpha(1-\alpha)}{2}Q^3 f^{(1)}(0), \tag{2.26}$$

and the MSE up to $O(N^{-2})$, *of the unconditional quantile estimators* $\widehat{\beta}$ *is*

$$\begin{aligned} M(\widehat{\beta}) &= \frac{1}{N}\alpha(1-\alpha)Q^2 - \frac{1}{N^2}\left(\alpha^2-\alpha-\frac{1}{2}\right)Q^2 - 11\frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2}-\alpha\right)Q^4 f^{(1)}(0) \\ &\quad + \frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2Q^6\left(f^{(1)}(0)\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2Q^5 f^{(2)}(0), \end{aligned} \tag{2.27}$$

where $Q = [f(0)]^{-1}$, $f(0)$ *is the unconditional density of* u_i *evaluated at* $u_i = 0$, $f^{(1)}(0)$ *and* $f^{(2)}(0)$ *are the first and second derivative of the unconditional density of* u_i *evaluated at* $u_i = 0$, *respectively.*

Proof: See appendix C.

Corollary 3.1. In the quantile regression model with $x_i = 1$, and y_i follow normal distribution $N(\mu, \sigma^2)$, the second-order bias up to $O(N^{-1})$, of the unconditional quantile estimators $\widehat{\beta}$ is

$$B(\widehat{\beta}) = \frac{1}{N}\left(\frac{1}{2}-\alpha\right)Q - \frac{1}{N}\frac{\alpha(1-\alpha)}{2}Q^2\left(\frac{-\beta+\mu}{\sigma^2}\right), \tag{2.28}$$

and the MSE up to $O(N^{-2})$, of the unconditional quantile estimators $\widehat{\beta}$ is

$$\begin{aligned} M(\widehat{\beta}) &= \frac{1}{N}\alpha(1-\alpha)Q^2 - 3\frac{1}{N^2}\alpha(1-\alpha)Q^2 + \frac{1}{N^2}\alpha(1-\alpha)(2\alpha-1)Q^3\frac{-\beta+\mu}{\sigma^2} \\ &\quad + \frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2Q^4\left(\frac{-\beta+\mu}{\sigma^2}\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2Q^4\frac{(-\beta+\mu)^2-\sigma^2}{\sigma^4} \end{aligned} \tag{2.29}$$

where $Q = \left[\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(\beta-\mu)^2}{2\sigma^2}\right)\right]^{-1}$.

Proof: If y_i follow normal distribution $N(\mu, \sigma^2)$, then the unconditional density, the first and second derivatives of the unconditional density are

$$f(y_i) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right),$$

$$f^{(1)}(y_i) = \frac{1}{\sqrt{2\pi}\sigma}\left(\frac{-y+\mu}{\sigma^2}\right)\exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right) = \frac{-y_i+\mu}{\sigma^2}f(y_i),$$

$$f^{(2)}(y_i) = \frac{1}{\sqrt{2\pi}\sigma}\left(\frac{(-y_i+\mu)^2-\sigma^2}{\sigma^4}\right)\exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right) = \frac{(-y_i+\mu)^2-\sigma^2}{\sigma^4}f(y_i).$$

Thus,

$$Q = [f(\beta)]^{-1} = \left[\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(\beta-\mu)^2}{2\sigma^2}\right)\right]^{-1}.$$

Based on Proposition 3, the second-order bias of $\widehat{\beta}$ up to $O(N^{-1})$ and the MSE up to $O(N^{-2})$ can be obtained.

Remark: We discover several other interesting properties from the expression of second-order bias: (i) when $|\frac{\beta-\mu}{\sigma}|$ is high, Q is large; (ii) when σ is high, $|B(\widehat{\beta})|$ is large; (iii) when $|\beta-\mu|$ is large, $|B(\widehat{\beta})|$ is large.

Corollary 3.2. If y_i follow a symmetric distribution, then the median estimator is unbiased. When y_i follow normal distribution $N(\mu, \sigma^2)$, the MSE up to $O(N^{-2})$ at the median, of the unconditional quantile estimators $\widehat{\beta}$ is

$$M(\widehat{\beta}) = \frac{\pi\sigma^2}{2N} - \frac{3\pi\sigma^2}{2N^2} + \frac{\pi^2\sigma^2}{4N^2}, \tag{2.30}$$

and non-negative MSE requires $N \geq 1$.

Proof: Since at the median of y_i, we have $\beta = \mu$. It is obvious that the bias is zero at the median. At the median, we also have $f(\beta) = f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$. Then the MSE at the median is

$$\begin{aligned} M(\widehat{\beta}) &= \frac{1}{N}\alpha(1-\alpha)Q^2 - \frac{1}{N^2}(\alpha^2-\alpha-\frac{1}{2})Q^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2Q^4\frac{(-\beta+\mu)^2-\sigma^2}{\sigma^4} \\ &= \frac{1}{N}\frac{1}{4}2\pi\sigma^2 - \frac{1}{N^2}\frac{3}{4}2\pi\sigma^2 + \frac{1}{N^2}\frac{1}{16}4\pi^2\sigma^4\frac{1}{\sigma^2} \\ &= \frac{\pi\sigma^2}{2N} - \frac{3\pi\sigma^2}{2N^2} + \frac{\pi^2\sigma^2}{4N^2}. \end{aligned}$$

Corollary 3.3. In the quantile regression model with $x_i = 1$, and y_i follow exponential distribution with density $f(y_i) = \lambda \exp(-\lambda y_i)$, $\lambda > 0$, the second-order bias up to $O(N^{-1})$, of the unconditional quantile estimators $\widehat{\beta}$ is

$$B(\widehat{\beta}) = \frac{1}{N}\left(\frac{1}{2} - \alpha\right) Q + \frac{1}{N}\frac{\alpha(1-\alpha)}{2}\lambda Q^2, \tag{2.31}$$

which is always non-positive, and the MSE up to $O(N^{-2})$, of the unconditional quantile estimators $\widehat{\beta}$ is

$$M(\widehat{\beta}) = \frac{1}{N}\alpha(1-\alpha)Q^2 - \frac{1}{N^2}(\alpha^2 - \alpha - \frac{1}{2})Q^2 + 11\frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2} - \alpha\right)\lambda Q^3 + \frac{11}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2\lambda^2 Q^4. \tag{2.32}$$

where $Q = [\lambda \exp(-\lambda\beta)]^{-1}$.

Proof: If y_i follow exponential distribution $\exp(\lambda)$, then the unconditional density, the first and second derivatives of the unconditional density are

$$f(y_i) = \lambda \exp(-\lambda y_i),$$

$$f^{(1)}(y_i) = -\lambda^2 \exp(-\lambda y_i) = -\lambda f(y_i),$$

$$f^{(2)}(y_i) = \lambda^3 \exp(-\lambda y_i) = \lambda^2 f(y_i).$$

Thus,

$$Q = [f(\beta)]^{-1} = [\lambda \exp(-\lambda\beta)]^{-1}.$$

Based on Proposition 3, the second-order bias of $\widehat{\beta}$ up to $O(N^{-1})$ and the MSE up to $O(N^{-2})$ can be obtained.

Corollary 3.4. When y_i follow exponential distribution $\exp(\lambda)$ with $\lambda > 0$, at the median, the second-order bias of the unconditional quantile estimator is

$$B(\widehat{\beta}) = \frac{1}{2N\lambda}, \tag{2.33}$$

and the MSE up to $O(N^{-2})$, of the unconditional quantile estimators is

$$M(\widehat{\beta}) = \frac{1}{N\lambda^2} + \frac{3}{N^2\lambda^2} + \frac{1}{4N^2\lambda^2}. \tag{2.34}$$

and non-negative MSE requires $N \geq 1$.

Proof: Since at the median of y_i, we have $\beta = \frac{1}{\lambda}\ln(2)$, $f(\beta) = \lambda \exp(-\ln(2)) = \frac{\lambda}{2}$, then the second-order bias and the MSE at the median can be obtained

2.5.2 Conditional Quantile Estimator with Binary Independent Variable

We consider the conditional quantile regression in Section 2, but now with x_i following the Bernoulli distribution $Bernoulli(p)$.

Proposition 4. *In the quantile regression model with x_i follow Bernoulli distribution $Bernoulli(p)$, the second-order bias up to $O(N^{-1})$, of the conditional quantile estimator $\widehat{\beta}$ is*

$$B(\widehat{\beta}) = \frac{1}{N}\left(\frac{1}{2} - \alpha\right)Q - \frac{1}{N}\frac{\alpha(1-\alpha)}{2}Q^3p^2f^{(1)}(0) \tag{2.35}$$

and the MSE up to $O(N^{-2})$, of the conditional quantile estimators $\widehat{\beta}$ is

$$\begin{aligned} M(\widehat{\beta}) \quad = \quad & \frac{1}{N}\alpha(1-\alpha)Q^2p - \frac{1}{N^2}\left(\alpha\left(1-\alpha\right)\left(4+p\right) + \frac{1}{2}\right)Q^2 - 9\frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2}-\alpha\right)Q^4p^2f^{(1)}(0) \\ & + \frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2Q^6p^4\left(f^{(1)}(0)\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2Q^5p^3f^{(2)}(0), \end{aligned} \tag{2.36}$$

where $Q = [pf(0)]^{-1}$. $f(0) = f(u_i|x_i = 1)$ *evaluated at* $u_i = 0$, $f^{(1)}(0) = f^{(1)}(u_i|x_i = 1)$ *and* $f^{(2)}(0) = f^{(2)}(u_i|x_i = 1)$ *evaluated at* $u_i = 0$.

Proof: See appendix C.

Remark: The second-order bias of $\widehat{\beta}$ is large at tails of a distribution. The second-order bias of $\widehat{\beta}$ goes to zero as $N \to \infty$. When p is small, the second-order bias of $\widehat{\beta}$ is large at tails of a distribution. If u_i follow a symmetric distribution, the median estimator is unbiased.

Corollary 4.1. In the quantile regression model with x_i follow Bernoulli distribution *Bernoulli*(p), and $y_i|x_i$ follow normal distribution $N(\mu, \sigma^2)$, the second-order bias up to $O(N^{-1})$, of the conditional quantile estimators $\widehat{\beta}$ is

$$B(\widehat{\beta}) = \frac{1}{N}\left(\frac{1}{2} - \alpha\right)Q - \frac{1}{N}\frac{\alpha(1-\alpha)}{2}Q^2 p\frac{-\beta+\mu}{\sigma^2} \tag{2.37}$$

and the MSE up to $O(N^{-2})$, of the unconditional quantile estimators $\widehat{\beta}$ is

$$\begin{aligned} M(\widehat{\beta}) &= \frac{1}{N}\alpha(1-\alpha)Q^2 p - \frac{1}{N^2}\left(\alpha(1-\alpha)(4+p) + \frac{1}{2}\right)Q^2 - 9\frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2} - \alpha\right)Q^3 p\frac{-\beta+\mu}{\sigma^2} \\ &\quad + \frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^4 p^2\left[\frac{15}{4}\left(\frac{-\beta+\mu}{\sigma^2}\right)^2 - \frac{(-\beta+\mu)^2 - \sigma^2}{\sigma^4}\right], \end{aligned} \tag{2.38}$$

where $Q = \left[\frac{1}{\sqrt{2\pi}\sigma}p\exp\left(-\frac{(\beta-\mu)^2}{2\sigma^2}\right)\right]^{-1}$.

*Proof***:** If $y_i|x_i$ follow normal distribution $N(\mu, \sigma^2)$, then

$$f(y_i|x_i = 1) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right),$$

$$f^{(1)}(y_i|x_i = 1) = \frac{1}{\sqrt{2\pi}\sigma}\left(\frac{-y_i+\mu}{\sigma^2}\right)\exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right) = \frac{-y_i+\mu}{\sigma^2}f(y_i|x_i = 1),$$

$$f^{(2)}(y_i|x_i=1) = \frac{1}{\sqrt{2\pi}\sigma}\left(\frac{(-y_i+\mu)^2-\sigma^2}{\sigma^4}\right)\exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right) = \frac{(-y_i+\mu)^2-\sigma^2}{\sigma^4}f(y_i|x_i=1).$$

Based on Proposition 4, the second-order bias of $\widehat{\beta}$ up to $O(N^{-1})$ and the MSE up to $O(N^{-2})$ can be obtained

Remark: We discover several other interesting properties from the expression of second-order bias: (i) when $|\frac{\beta-\mu}{\sigma}|$ is high, Q is large; (ii) when σ is high, $|B(\widehat{\beta})|$ is large; (iii) when $|\beta-\mu|$ is large, $|B(\widehat{\beta})|$ is large; and (iv) when p is small, $|B(\widehat{\beta})|$ is large.

Corollary 4.2. If $y_i|x_i$ follow a symmetric distribution, then the median is unbiased. When $y_i|x_i$ follow normal distribution $N(\mu,\sigma^2)$, the MSE up to $O(N^{-2})$ at the median, of the conditional quantile estimators $\widehat{\beta}$ is

$$M(\widehat{\beta}) = \frac{\pi\sigma^2}{2Np} - \frac{\pi\sigma^2}{2N^2p} - \frac{3\pi\sigma^2}{N^2p^2} + \frac{\pi^2\sigma^2}{4N^2p^4}, \tag{2.39}$$

and non-negative MSE requires $N \geq \frac{15}{2p} - \frac{\pi}{2p^3}$.

Proof: Since at the median of y_i, we have $f(\beta) = f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$, then the MSE up to $O(N^{-2})$ is

$$\begin{aligned} M(\widehat{\beta}) &= \frac{1}{N}\alpha(1-\alpha)Q^2p - \frac{1}{N^2}\left(\alpha(1-\alpha)(4+p)+\frac{1}{2}\right)Q^2 + \frac{1}{N^2}\alpha^2(1-\alpha)^2Q^4p^2\left[-\frac{(-\beta+\mu)^2-\sigma^2}{\sigma^4}\right] \\ &= \frac{1}{N}\frac{1}{4}\frac{2\pi\sigma^2}{p} - \frac{1}{N^2}\left(\frac{p}{4}+\frac{3}{2}\right)\frac{2\pi\sigma^2}{p^2} + \frac{1}{N^2}\frac{1}{16}\frac{4\pi^2\sigma^4}{p^4}\frac{1}{\sigma^2} \\ &= \frac{\pi\sigma^2}{2Np} - \frac{\pi\sigma^2}{2N^2p} - \frac{3\pi\sigma^2}{N^2p^2} + \frac{\pi^2\sigma^2}{4N^2p^4}. \end{aligned}$$

Corollary 4.3. In the quantile regression model with x_i follow Bernoulli distribution *Bernoulli*(p), and $y_i|x_i$ follow exponential distribution, $f(y_i|x_i) = \lambda\exp(-\lambda y_i)$ with $\lambda > 0$, the second-order bias up to $O(N^{-1})$, of the conditional quantile estimators $\widehat{\beta}$ is

$$B(\widehat{\beta}) = \frac{1}{N}\left(\frac{1}{2} - \alpha\right) Q + \frac{1}{N}\frac{\alpha(1-\alpha)}{2}\lambda Q^2 p, \tag{2.40}$$

which is always non-positive, and the MSE up to $O(N^{-2})$, of the conditional quantile estimators $\widehat{\beta}$ is

$$\begin{aligned} M(\widehat{\beta}) &= \frac{1}{N}\alpha(1-\alpha)Q^2 p - \frac{1}{N^2}\left(\alpha\left(1-\alpha\right)\left(4+p\right) + \frac{1}{2}\right) Q^2 + 9\frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2} - \alpha\right) Q^3 p\lambda \\ &+ \frac{11}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^4 p^2 \lambda^2, \end{aligned} \tag{2.41}$$

where $Q = \left[p\lambda \exp\left(-\lambda\beta\right)\right]^{-1}$.

Proof: If $y_i|x_i$ follow the exponential distribution, then

$$f(y_i|x_i = 1) = \lambda \exp\left(-\lambda y_i\right),$$

$$f^{(1)}(y_i|x_i = 1) = -\lambda^2 \exp\left(-\lambda y_i\right) = -\lambda f(y_i|x_i = 1),$$

$$f^{(2)}(y_i|x_i = 1) = \lambda^3 \exp\left(-\lambda y_i\right) = \lambda^2 f(y_i|x_i = 1).$$

$$Q = \left(E[x_1^2 f(x_1'\beta)]\right)^{-1} = \left(pE[f(\beta)]\right)^{-1} = \left[pf(\beta)\right]^{-1} = \left[p\lambda \exp\left(-\lambda\beta\right)\right]^{-1}.$$

Based on Proposition 4, the second-order bias of $\widehat{\beta}$ up to $O(N^{-1})$ and the MSE up to $O(N^{-2})$ can obtained

Corollary 4.4. When $y_i|x_i$ follow the exponential distribution, $f\left(y_i|x_i\right) = \lambda \exp(-\lambda y_i)$ with $\lambda > 0$, at the median, the second-order bias of the conditional quantile estimator is

$$B(\widehat{\beta}) = \frac{1}{2Np\lambda}, \tag{2.42}$$

and the MSE up to $O(N^{-2})$ of the unconditional quantile estimators is

$$M(\widehat{\beta}) = \frac{1}{Np\lambda^2} - \frac{1}{N^2 p\lambda^2} - \frac{13}{4N^2 p^2 \lambda^2}, \tag{2.43}$$

and non-negative MSE requires $N \geq 1 + \frac{13}{4p}$.

Proof: Since at the median of y_i, we have $\beta = \frac{1}{\lambda}\ln(2)$, $f(\beta) = \lambda\exp(-\lambda\ln(2)) = \frac{\lambda}{2}$, $Q = [pf(\beta)]^{-1} = \frac{2}{p\lambda}$, then the second-order bias and the MSE at the median of y_i can be obtained.

2.5.3 Instrumental Variable Quantile Regression

Consider the quantile model where the explanatory variable x_i is endogenous, and z_i is the instrumental variable.

$$y_i = x_i'\beta + u_i, \tag{2.44}$$

$$x_i = \Gamma z_i + v_i. \tag{2.45}$$

where y_i is a scalar, x_i is a $k \times 1$ vector, and z_i is an $l \times 1$ vector. We consider the case when $l = k$ below. When $l = k = 1$, the $k \times l$ matrix Γ become a scalar γ.

Proposition 5. *In the instrumental variable quantile regression (IVQR) model, suppose x_i and u_i are i.i.d., the second-order bias, up to $O(N^{-1})$, of the quantile estimators $\widehat{\beta}$ is*

$$B(\widehat{\beta}) = \frac{1}{N}\left[\left(\frac{1}{2} - \alpha\right) QE\left[z_i x_i' Q z_i f(x_i'\beta)\right] - \frac{\alpha(1-\alpha)}{2} E\left[\left(z_i x_i'\right) \otimes x_i' f^{(1)}(0)\right] \left(Q \otimes Q\right) E\left(z_i \otimes z_i\right)\right], \tag{2.46}$$

where $Q = \left(E[z_i x_i' f(0)]\right)^{-1}$. When $k = 1$, the MSE up to $O(N^{-2})$, of the quantile estimator

$\widehat{\beta}$ *is*

$$
\begin{aligned}
M(\widehat{\beta}) \quad = \quad & \frac{1}{N}\alpha(1-\alpha)Q^2E\left(z_i^2\right) - 2\frac{1}{N^2}Q^3\left(\alpha^2-\alpha+\frac{1}{2}\right)E\left[z_i^3x_i\right]f(0) - \frac{1}{N^2}\alpha(1-\alpha)Q^2E\left(z_i^2\right) \\
& + \frac{1}{N^2}\alpha(1-\alpha)(2\alpha-1)Q^4E\left[z_ix_i^2\right]E\left(z_i^3\right)f^{(1)}(0) + 6\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)^2Q^4\left(E\left[z_i^2x_i\right]f(0)\right)^2 \\
& + 3\frac{1}{N^2}\alpha(1-\alpha)Q^4\left(\frac{1}{2}-\alpha\right)E\left[z_i^2x_i^2\right]E\left(z_i^2\right)f^{(1)}(0) \\
& - 12\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)\alpha(1-\alpha)Q^5E[z_ix_i^2]E\left[z_i^2x_i\right]E\left(z_i^2\right)f(0)f^{(1)}(0) \\
& + \frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2Q^6\left(E\left[z_ix_i^2\right]f^{(1)}(0)\right)^2\left(E\left(z_i^2\right)\right)^2 \\
& - \frac{1}{N^2}\alpha^2(1-\alpha)^2Q^5E[z_ix_i^3]\left(E\left(z_i^2\right)\right)^2f^{(2)}(0), \qquad (2.47)
\end{aligned}
$$

where $Q = \left(E[z_ix_i]f(0)\right)^{-1}$.

Proof: See appendix C.

Remark: When $k=1$, $\Gamma=\gamma$, we observe that x_i, Ψ_N, s_i, d_i, H_1, $\overline{H_1}$, Q, V_i, H_2, $\overline{H_2}$, W_i, H_3, $\overline{H_3}$ are all scalars, and the second-order bias of $\widehat{\beta}$ up to $O(N^{-1})$ can be rewritten as

$$
B(\widehat{\beta}) = \frac{1}{N}\left(\frac{1}{2}-\alpha\right)\gamma Q^2E\left[z_i^3f(0)\right] - \frac{1}{N}\frac{\alpha(1-\alpha)}{2}Q^3E[z_ix_i^2f^{(1)}(0)]E\left(z_i^2\right),
$$

where $Q = \left(\gamma E[z_i^2f(0)]\right)^{-1}$.

The second-order bias of $\widehat{\beta}$ is larger at the tails of a distribution. When γ is small, second-order bias of $\widehat{\beta}$ is larger. If u_i follow symmetric distribution, the median estimator is unbiased. The second-order bias of $\widehat{\beta}$ goes to zero as the sample size goes to infinity.

Corollary 5. The MSE of the quantile estimator $\widehat{\beta}$ up to $O(N^{-1})$ equals the asymptotic variance of $\widehat{\beta}$.

Proof: From Theorem 2, we observe that the MSE of $\widehat{\beta}$ up to $O(N^{-1})$ for quantile estimator for i.i.d. case when $k = 1$ can be simplified as

$$M(\widehat{\beta}) = \frac{1}{N}\overline{d_i^2} = \frac{1}{N}Q^2\alpha(1-\alpha)E\left(z_i^2\right).$$

Under the i.i.d. assumption, the asymptotic distribution of the quantile regression estimator when the α-quantile is linear in x_i, is as follows,

$$\sqrt{N}(\widehat{\beta} - \beta) \xrightarrow{d} N(0, V_\alpha),$$

where

$$V_\alpha = \alpha(1-\alpha)[E(f(0)z_i x_i')]^{-1}(E z_i z_i')[E(f(0)z_i x_i')]^{-1},$$

and $f(0|x_i)$ is the density of u_i conditional on x_i evaluated at $u_i = 0$. See Chernozhukov and Hansen (2006). We can prove that V_α, the asymptotic variance of $\widehat{\beta}$, equals the N times the MSE of $\widehat{\beta}$ up to $O(N^{-1})$. Since $\mathbf{1}(u_i < 0)$ is Bernoulli with mean α and variance $\alpha(1-\alpha)$, then we can have

$$\begin{aligned}
E\left[\Psi_N(\beta)\Psi_N(\beta)'\right] &= E\left[\left(\frac{1}{N}\sum_{i=1}^{N} s_i\right)\left(\frac{1}{N}\sum_{i=1}^{N} s_i'\right)\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N} E\left[s_i s_i'\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N} E\left[(\alpha - \mathbf{1}(u_i < 0))^2 z_i z_i'\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N} E\left[z_i z_i' E[(\alpha - \mathbf{1}(u_i < 0))^2 | x_i]\right] \\
&= \frac{\alpha(1-\alpha)}{N^2}\sum_{i=1}^{N} E(z_i z_i').
\end{aligned}$$

Under the i.i.d. assumption, the MSE of $\widehat{\beta}$ up to $O(N^{-1})$ can be derived by substituting the result above,

$$\begin{aligned}
MSE(\widehat{\beta}) &= E(a_{-1/2}a'_{-1/2}) = E(Q\Psi_N\Psi_N'Q) = QE\left[\Psi_N(\beta)\Psi_N(\beta)'\right]Q \\
&= \left[\frac{1}{N}\sum_{i=1}^{N}E(f(0|x_i)z_ix_i')\right]^{-1}\frac{\alpha(1-\alpha)}{N^2}\sum_{i=1}^{N}E(z_iz_i')\left[\frac{1}{N}\sum_{i=1}^{N}E(f(0|x_i)z_ix_i')^{-1}\right] \\
&= [E(f(0)z_ix_i')]^{-1}\frac{\alpha(1-\alpha)}{N}E(z_iz_i')[E(f(0)z_ix_i')]^{-1} \\
&= \frac{V_\alpha}{N},
\end{aligned}$$

where $f(0)$ is the density of u_i evaluated at $u_i = 0$. The asymptotic variance $V_\alpha = N \times MSE(\widehat{\beta}) = \alpha(1-\alpha)[E(f(0)z_iz_i')]^{-1}E(z_iz_i')[E(f(0)z_iz_i')]^{-1}$.

2.6 Monte Carlo Simulation

2.6.1 Simulation Design

Now we give some numerical calculation to present the second-order bias and MSE results. In the quantile regression model $y_i = x_i'\beta + u_i$, the error term u_i satisfies $E\left[\alpha - \mathbf{1}(y_i < x_i'\beta)|x_i\right] = 0$. The α conditional quantile of u_i given x_i is zero.

In the first data generating process (DGP 1), the error term u_i is normally distributed with the CDF $F(\cdot)$, standard deviation σ_u, then the mean equals to $-\Phi^{-1}(\alpha)\sigma_u$, with $\Phi(\cdot)$ denoting the standard normal CDF. We have

$$\begin{aligned}
F(0) &= \int_{-\infty}^{0} f(u)\mathrm{d}u = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}\sigma_u} \exp\left[-\frac{\left\{u-\left(-\Phi^{-1}(\alpha)\sigma_u\right)\right\}^2}{2\sigma_u^2}\right] \mathrm{d}u \\
&= \int_{-\infty}^{\Phi^{-1}(\alpha)\sigma_u} \frac{1}{\sqrt{2\pi}\sigma_u} \exp\left[-\frac{z^2}{2\sigma_u^2}\right] \mathrm{d}z \\
&= \int_{-\infty}^{\Phi^{-1}(\alpha)} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{w^2}{2}\right] \mathrm{d}w \\
&= \Phi(\Phi^{-1}(\alpha)) \\
&= \alpha.
\end{aligned}$$

Therefore, we generate the error term u_i following normal distribution $N(-\Phi^{-1}(\alpha)\sigma_u, \sigma_u^2)$. In the second DGP (DGP 2), the error term u_i is uniformly distributed with the CDF $F(\cdot)$ on $[a, b]$, then $a = \frac{\alpha}{\alpha-1}b$. We have

$$F(0) = \int_{-\infty}^{0} f(u)du = \int_{a}^{0} \frac{1}{b-a}du = -\frac{a}{b-a} = \alpha.$$

Therefore, we generate the error term u_i from uniform distribution on $[a, b]$, where $a = -\alpha R$, $b = R(1-\alpha)$, and the range $R = b - a$. The results tables for DGP 2 are in the appendix D.

We simulate x_i from several different distributions. Then, y_i is simulated from $y_i = x_i'\beta + u_i$. We try with $\alpha = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95$, $\sigma_u = 0.5, 0.1, 0.05$, $\beta = 0$, $N = 50, 60, 100$. We use the Matlab package by Roger Koenker to estimate the models. The results represent the averaged values across 10,000 simulations. Note that when $k = 1$, x_i, Ψ_N, s_i, d, H_1, $\overline{H_1}$, Q, V, H_2, $\overline{H_2}$, W, H_3, $\overline{H_3}$ are all scalars. In all the results tables, for each α, the first row is for bias and the second row is for the MSE of the quantile estimator. For each panel, the first column presents the second-order bias and MSE derived by Theorems, Propositions, and Collaries, the second column presents the

Monte Carlo simulation bias and MSE of quantile estimators $\widehat{\beta}$, the third column presents the Monte Carlo simulation bias and MSE of the bias corrected quantile estimators $\widetilde{\beta}$, where $\widetilde{\beta} \equiv \widehat{\beta} - B\left(\widehat{\beta}\right)$.

Table 1-7 present the results with DGP 1. We use the Matlab package by Roger Koenker to estimate the model. Table 1 shows the results in Theorem 1 and 2, when there is hetroskedasticity. Table 2-5 show the results in Corollary 1, when x_i and u_i are i.i.d.. Table 2 and 3 show the results when $\sigma_u = 0.5, 0.1$, and x_i is generated from exponential distribution, $f(x_i) = \exp(-x_i)$. Table 4 and 5 show the results when $\sigma_u = 0.5$ and when x_i is generated from two different distributions introduced in Marron and Wand (1992). They are mixture normal distributions: Skewed Unimodal Density $\frac{1}{5}N(0,1)+\frac{1}{5}N(\frac{1}{2},(\frac{2}{3})^2)+\frac{3}{5}N(\frac{13}{12},(\frac{5}{9})^2)$, and Strongly Skewed Density ${}_{l=0}^{7}\frac{1}{8}N(3[(\frac{2}{3})^l-1],(\frac{2}{3})^{2l})$. Table 6 shows the results in Proposition 3, with unconditional quantile regression. Table 7 shows the results in Proposition 4, with binary independent variable with $p = 0.3$. Note that the unconditional quantile regression is a special case of conditional quantile regression with binary independent variable with $p = 1$. Table 8 presents the results for IVQR. We use the Matlab package by Chernozhukov and Hansen (2006) to estimate the model. In the simulation of IVQR, u_i is generated the same as DGP 1; v_i is simulated from $v_i = w_i + cu_i$, where w_i is from $N(0, 0.25)$, $c = 0.5, 1$; z_i is from exponential distribution, $f(x_i) = \exp(-x_i)$;.x_i is simulated from $x_i = z_i'\gamma + v_i$, where $\gamma = 0.5, 0.9$; y_i is simulated from $y_i = x_i'\beta + u_i$, where $\beta = 0$.

2.6.2 Simulation Results

From all the results Tables, we find that the derived second-order bias are numerically close to the the Monte Carlo simulation bias, the second-order bias corrected estimators are numerically close to the true estimator zero of the data generation.

The Monte Carlo Simulation results provide improvement of quantile estimators. We find that the second-order bias corrected estimator has better behaviors than the uncorrected ones. From comparing the simulation results, our conclusion are as follows. (i) When x_i is generated from standard normal distribution, the bias are close to zero. That's because the expressions of the second-order bias contain the third-moment of x_i. If the distribution of x_i is symmetric, the bias will goes to zero. Since exponential and the three mixture normal distributions are all asymmetric, bias corrected estimator are more close to the true β. (ii) The bias is larger at the extreme low and high quantiles. The first and second column for each sample size section show that the bias is zero at the median, negative when $\alpha < 0.5$, and positive when $\alpha > 0.5$. There are upward bias at lower quantiles and downward bias at upper quantiles. (iii) When the sample size is increasing, the quantile estimators become closer to the true β, and the bias become smaller. The quantile estimators are asymptotically unbiased. (iv) When σ_u is larger, the quantile estimator has larger bias especially at the tails. (v) In the IVQR, the bias is larger for weak instruments.

2.7 Appendix

2.7.1 Appendix A

This appendix provides a review on a vector norm and the Dirac delta function.a). A.1 Property of norm, if A is a $k \times 1$ vector,

$$||A|| = \left[tr\left(AA'\right)\right]^{1/2} = \left(A'A\right)^{1/2}.$$

$A.2$

$$||AA'|| = \left[tr\left(AA'AA'\right)\right]^{1/2} = \left[tr\left(A'AA'A\right)\right]^{1/2} = \left(A'AA'A\right)^{1/2} = A'A = ||A||^2.$$

$A.3$

$$\begin{aligned}
||(AA') \otimes A'|| &= \left\{tr\left(\left[(AA') \otimes A'\right]\left[(AA') \otimes A\right]\right)\right\}^{1/2} \\
&= \left[tr\left(\left(AA'AA'\right) \otimes \left(A'A\right)\right)\right]^{1/2} \\
&= \left[tr\left(A'AA'AA'A\right)\right]^{1/2} \\
&= \left(A'A\right)^{3/2} \\
&= ||A||^3.
\end{aligned}$$

$A.4$

$$\begin{aligned}
||(AA') \otimes A' \otimes A'|| &= tr\left(\left[(AA') \otimes A' \otimes A'\right]\left[(AA') \otimes A \otimes A\right]\right)^{1/2} \\
&= tr\left[\left(AA'AA'\right) \otimes \left(A' \otimes A'\right)(A \otimes A)\right]^{1/2} \\
&= tr\left[\left(AA'AA'\right) \otimes A'A \otimes A'A\right]^{1/2} \\
&= tr\left[\left(A'AA'A\right) A'AA'A\right]^{1/2} \\
&= \left(A'AA'A\right) \\
&= \left(A'A\right)^2 \\
&= \|A\|^4
\end{aligned}$$

2.7.2 Appendix B

This appendix provides a review on the Dirac delta function.

$B.1$ The Heaviside unit step function is defined as $\phi(z) = 0$ for $z < 0$, $\phi(z) = 1$ for $z \geq 0$. The Dirac delta function is defined as $\delta(z) = \mathrm{d}\phi(z)/\mathrm{d}z$, where $\delta(z) = 0$ for $z < 0$, $\delta(z) = \infty$ for $z = 0$, $\delta(z) = 0$ for $z > 0$.

$B.2$ Property of Dirac delta function $\int_{-\infty}^{+\infty} \delta(z)\mathrm{d}z = 1$.

$B.3$ $\int_{-\infty}^{+\infty} \delta(z-a)f(z)\mathrm{d}z = f(a)$, where $f : R \to R$ is a real function differentiable around $a \in R$.

$B.4$ $\delta(z) = \delta(-z)$.

$B.5$ $\int_{-\infty}^{+\infty} \delta^{(1)}(z-a)f(z)\mathrm{d}z = -\int_{-\infty}^{+\infty} \delta(z-a)f^{(1)}(z)\mathrm{d}z = -f^{(1)}(a)$.

$B.6$ $\delta^{(1)}(-z) = -\delta^{(1)}(z)$, $\delta^{(2)}(-z) = \delta^{(2)}(z)$.

$B.7$ $\int_{-\infty}^{+\infty} \delta^{(n)}(z-a)f(z)\mathrm{d}z = (-1)^n \int_{-\infty}^{+\infty} \delta(z-a)f^{(n)}(z)\mathrm{d}z = (-1)^n f^{(n)}(a)$.

$B.8$ $\phi(z)\delta(z) = \frac{1}{2}\delta(z).a)$.

$B.9$ $\phi(z)\delta^{(1)}(z) = \frac{1}{2}\delta^{(1)}(z) - (\delta(z))^2$.

2.7.3 Appendix C

This appendix provides proofs for some propositions and corollaries.

$C.1$ *Proof of Proposition 3:* If the linear quantile regression model is $y_i = \beta + u_i$, where y_i is a scalar, u_i is the error defined to be the difference between y_i and its α-quantile β, we call $\widehat{\beta}$ as the unconditional quantile estimators. Given the definition of the check loss function, the quantile estimators $\widehat{\beta}$ can be obtained by solving

$$\min_{\beta} E[L_\alpha(\beta)] = E[(\alpha - \mathbf{1}(y_i < \beta))(y_i - \beta)].$$

We can show that $E[L(\beta)]$ is continuously differentiable on Θ. Then can write the population moment condition as

$$\nabla^1_\beta E[L_\alpha(\beta)] = E[-\nabla^1_\beta 1(y_i - \beta < 0)(y_i - \beta)] - E[\alpha - \mathbf{1}(y_i < \beta)].$$

By the definition of Dirac delta function in Appendix B.1,we have $\mathbf{1}(y_i - \beta < 0) = \mathbf{1}(\beta - y_i \geq 0) = \phi(\beta - y_i)$. Then

$$\nabla^1_\beta \mathbf{1}(y_i - \beta < 0) = \delta(\beta - y_i).$$

According to the property of Dirac delta function in Appendix B.4, we have $\delta(\beta - y_i) = \delta(y_i - \beta)$. According to the property of Dirac delta function in Appendix B.3, we have

$$\begin{aligned}
E[\delta(\beta - y_i)(y_i - \beta)] &= E[\delta(y_i - \beta)(y_i - \beta)] \\
&= \int_{-\infty}^{+\infty} \delta(y_i - \beta)(y_i - \beta) f(y_i) \mathrm{d}y_i \\
&= (\beta_\alpha - \beta) f(\beta) \\
&= 0.
\end{aligned}$$

Thus, the moment condition can be written as

$$\nabla^1_\beta E[L_\alpha(\beta)] = -E[\alpha - \mathbf{1}(y_i < \beta)] = E[s_i(\beta)],$$

where $s_i(\beta) = -(\alpha - \mathbf{1}(y_i < \beta))$. The sample moment condition can be written as

$$\Psi_N(\beta) = \frac{1}{N}\sum_{i=1}^{N} s_i(\beta). \tag{2.48}$$

The second-order bias up to $O(N^{-1})$ is

$$B(\widehat{\beta}) = \frac{1}{N} Q \left[\overline{V_i d_i} - \frac{1}{2} \overline{H_2} \left(\overline{d_i \otimes d_i} \right) \right],$$

where

$$H_1 = \nabla^1_\beta s_i = \nabla^1_\beta (\mathbf{1}(y_i < \beta)) = \delta(\beta - y_i),$$

$$H_2 = \nabla^2_\beta s_i = -\delta^{(1)}(\beta - y_i),$$

$$H_3 = \nabla^3_\beta s_i = \delta^{(2)}(\beta - y_i),$$

$$\overline{H_1} = E\nabla^1_\beta s_i = E\left[\delta(\beta - y_i)\right] = \int_{-\infty}^{+\infty} \delta(y_i - \beta) f(y_i) \mathrm{d}y_i = f\left(\beta\right),$$

$$\overline{H_2} = E\nabla^2_\beta s_i = -E\left[\delta^{(1)}(\beta - y_i)\right] = -f^{(1)}\left(\beta\right),$$

$$\overline{H_3} = E\nabla^3_\beta s_i = E\left[\delta^{(2)}(\beta - y_i)\right] = f^{(2)}\left(\beta\right),$$

$$Q = \left(\overline{H_1}\right)^{-1} = [f(\beta)]^{-1},$$

$$V = H_1 - \overline{H_1} = \delta(\beta - y_i) - f(\beta),$$

$$W = H_2 - \overline{H_2} = -\delta^{(1)}(\beta - y_i) + f^{(1)}(\beta),$$

$$d_i = Qs_i = -[f(\beta)]^{-1}(\alpha - \mathbf{1}(y_i < \beta)).$$

$f(\beta)$ is the unconditional density of y_i evaluated at $y_i = \beta$. $f^{(1)}(\beta)$ and $f^{(2)}(\beta)$ are the first and second derivative of the unconditional density of y_i evaluated at $y_i = \beta$, respectively. Since Ψ_N, s_i, d_i, H_1, $\overline{H_1}$, Q, V_i, H_2, $\overline{H_2}$, W_i, H_3, $\overline{H_3}$ are all scalars, then

$$\begin{aligned}
\overline{V_i d_i} &= E\left[\left(H_1 - \overline{H_1}\right) Q s_i\right] \\
&= QE\left(H_1 s_i\right) - E\left(s_i\right) \\
&= Q\left[-\int_{-\infty}^{+\infty} \delta(\beta - y_i)(\alpha - \mathbf{1}(y_i < \beta)) f(y_i) \mathrm{d}y_i\right] \\
&= Q\left[-\int_{-\infty}^{+\infty} \delta(\beta - y_i)\alpha f(y_i) \mathrm{d}y_i + \int_{-\infty}^{+\infty} \delta(\beta - y_i)\mathbf{1}(y_i < \beta) f(y_i) \mathrm{d}y_i\right] \\
&= \left(\frac{1}{2} - \alpha\right) Q\left[f(\beta)\right].
\end{aligned}$$

$$\overline{d_1 \otimes d_1} = Q^2 E\left[s_i^2\right] = Q^2\left[(\alpha - 1)^2 \alpha + \alpha^2 (1 - \alpha)\right] = \alpha(1 - \alpha) Q^2.$$

Therefore, the second-order bias of $\widehat{\beta}$ up to $O(N^{-1})$,of the unconditional quantile estimators $\widehat{\beta}$ can be written as

$$
\begin{aligned}
B(\widehat{\beta}) &= \frac{1}{N} Q \left[\overline{V_i d_i} - \frac{1}{2} \overline{H_2} \left(\overline{d_i \otimes d_i} \right) \right] \\
&= \frac{1}{N} \left(\frac{1}{2} - \alpha \right) Q^2 \left[f(\beta) \right] - \frac{1}{N} \frac{\alpha(1-\alpha)}{2} Q^3 f^{(1)}(\beta) \\
&= \frac{1}{N} \left(\frac{1}{2} - \alpha \right) Q - \frac{1}{N} \frac{\alpha(1-\alpha)}{2} Q^3 f^{(1)}(\beta),
\end{aligned}
$$

where $Q = [f(\beta)]^{-1}$. Since the unconditional density of y_i evaluated at $y_i = \beta$ is the same as the unconditional density of u_i evaluated at $u_i = 0$. If we use $f(0)$ to denote the unconditional density of u_i evaluated at $u_i = 0$, the second-order bias of $\widehat{\beta}$ up to $O(N^{-1})$,of the unconditional quantile estimators $\widehat{\beta}$ can be written as

$$
B(\widehat{\beta}) = \frac{1}{N} \left(\frac{1}{2} - \alpha \right) Q - \frac{1}{N} \frac{\alpha(1-\alpha)}{2} Q^3 f^{(1)}(0),
$$

where $Q = [f(0)]^{-1}$.

If $x_i = 1$, the MSE up to $O(N^{-2})$ of the unconditional quantile estimators $\widehat{\beta}$ can be simplified as

$$
\begin{aligned}
M(\widehat{\beta}) &= \frac{1}{N}\alpha(1-\alpha)Q^2 - 2\frac{1}{N^2}\left(\alpha^2-\alpha+\frac{1}{2}\right)Q^3 f(\beta) + 2\frac{1}{N^2}\alpha(1-\alpha)Q^2 \\
&\quad + \frac{1}{N^2}\alpha(1-\alpha)(2\alpha-1)Q^4 f^{(1)}(\beta) + 6\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)^2 Q^4 \left(f(\beta)\right)^2 \\
&\quad + 3\frac{1}{N^2}\alpha(1-\alpha)Q^4\left(\left(\frac{1}{2}-\alpha\right)f^{(1)}(\beta) - \left(f(\beta)\right)^2\right) - 12\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)\alpha(1-\alpha)Q^5 f^{(1)}(\beta) f(\beta) \\
&\quad + \frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^6\left(f^{(1)}(\beta)\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^5 f^{(2)}(\beta) \\
&= \frac{1}{N}\alpha(1-\alpha)Q^2 - 2\frac{1}{N^2}\left(\alpha^2-\alpha+\frac{1}{2}\right)Q^2 + 2\frac{1}{N^2}\alpha(1-\alpha)Q^2 + \frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2}-\alpha\right)Q^4 f^{(1)}(\beta) \\
&\quad + 3\frac{1}{N^2}\left(\alpha^2-\alpha+\frac{1}{2}\right)Q^2 - 12\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)\alpha(1-\alpha)Q^4 f^{(1)}(\beta) \\
&\quad + \frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^6\left(f^{(1)}(\beta)\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^5 f^{(2)}(\beta) \\
&= \frac{1}{N}\alpha(1-\alpha)Q^2 - \frac{1}{N^2}\left(\alpha^2-\alpha-\frac{1}{2}\right)Q^2 - 11\frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2}-\alpha\right)Q^4 f^{(1)}(\beta) \\
&\quad + \frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^6\left(f^{(1)}(\beta)\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^5 f^{(2)}(\beta),
\end{aligned}
$$

where $Q = [f(\beta)]^{-1}$. Since the unconditional density of y_i evaluated at $y_i = \beta$ is the same as the unconditional density of u_i evaluated at $u_i = 0$. If we use $f(0)$ to denote the unconditional density of u_i evaluated at $u_i = 0$, then we observe the MSE with the expression in Proposition 3.

C.2 Proof of Proposition 4: If x_i follows the Bernoulli distribution $Bernoulli(p)$, then $E(x_i) = E(x_i x_i') = E(x_i x_i' x_i) = E\left((x_i x_i') \otimes x_i'\right) = p$,where $j = 1, 2, 3, \cdots$. Thus,

$$Q = \left(E[x_i x_i' f(x_i'\beta)]\right)^{-1} = (pE[f(\beta)])^{-1} = [pf(\beta)]^{-1},$$

$$E[x_i x_i' x_i f(x_i'\beta)] = pE[f(\beta)] = pf(\beta),$$

$$E[\left(x_i x_i'\right) \otimes x_i' f^{(1)}(x_i'\beta)] = pE[f^{(1)}(\beta)] = pf^{(1)}(\beta).$$

Based on Theorem 1, the second-order bias up to $O(N^{-1})$, of the conditional quantile estimators $\widehat{\beta}$ is

$$\begin{aligned} B(\widehat{\beta}) &= \frac{1}{N}\left(\frac{1}{2}-\alpha\right)Q^2 E\left[x_i x_i' x_i f(x_i'\beta)\right] - \frac{1}{N}\frac{\alpha(1-\alpha)}{2}Q^3 E[(x_i x_i') \otimes x_i' f^{(1)}(x_i'\beta)]E\left(x_i^2\right) \\ &= \frac{1}{N}\left(\frac{1}{2}-\alpha\right)Q - \frac{1}{N}\frac{\alpha(1-\alpha)}{2}Q^3 p^2 f^{(1)}(\beta), \end{aligned}$$

where $Q = [pf(\beta)]^{-1}$. Based on Theorem 2, the MSE up to $O(N^{-2})$, of the conditional quantile estimators $\widehat{\beta}$ is

$$\begin{aligned} M(\widehat{\beta}) &= \frac{1}{N}\alpha(1-\alpha)Q^2 p - 2\frac{1}{N^2}Q^3\left(\alpha^2-\alpha+\frac{1}{2}\right)pf(\beta) + 2\frac{1}{N^2}\alpha(1-\alpha)Q^2 p \\ &\quad + \frac{1}{N^2}\alpha(1-\alpha)(2\alpha-1)Q^4 p^2 f^{(1)}(\beta) + 6\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)^2 Q^4\left(pf(\beta)\right)^2 \\ &\quad + 3\frac{1}{N^2}\alpha(1-\alpha)Q^4\left(\left(\frac{1}{2}-\alpha\right)p^2 f^{(1)}(\beta) - p^3\left(f(\beta)\right)^2\right) \\ &\quad - 12\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)\alpha(1-\alpha)Q^5 p^3 f^{(1)}(\beta)f(\beta) \\ &\quad + \frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^6 p^4\left(f^{(1)}(\beta)\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^5 p^3 f^{(2)}(\beta) \\ &= \frac{1}{N}\alpha(1-\alpha)Q^2 p - 2\frac{1}{N^2}Q\left({}^2\alpha^2-\alpha+\frac{1}{2}\right) + 2\frac{1}{N^2}\alpha(1-\alpha)Q^2 p \\ &\quad + \frac{1}{N^2}\alpha(1-\alpha)(2\alpha-1)Q^4 p^2 f^{(1)}(\beta) + 6\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)^2 Q^2 \\ &\quad + 3\frac{1}{N^2}\alpha(1-\alpha)Q^4\left(\frac{1}{2}-\alpha\right)p^2 f^{(1)}(\beta) - 3\frac{1}{N^2}\alpha(1-\alpha)Q^2 p \\ &\quad - 12\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)\alpha(1-\alpha)Q^4 p^2 f^{(1)}(\beta) \\ &\quad + \frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^6 p^4\left(f^{(1)}(\beta)\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^5 p^3 f^{(2)}(\beta) \\ &= \frac{1}{N}\alpha(1-\alpha)Q^2 p - \frac{1}{N^2}\left(\alpha\left(1-\alpha\right)\left(4+p\right)+\frac{1}{2}\right)Q^2 - 9\frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2}-\alpha\right)Q^4 p^2 f^{(1)}(\beta) \\ &\quad + \frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^6 p^4\left(f^{(1)}(\beta)\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^5 p^3 f^{(2)}(\beta), \end{aligned}$$

where $Q = [pf(\beta)]^{-1}$. Since the conditional density of y_i given x_i evaluated at $y_i = x_i'\beta$ is the same as the conditional density of u_i given x_i evaluated at $u_i = 0$. If we use $f\left(0|x_i\right)$

to denote the conditional density of u_i given x_i evaluated at $u_i = 0$, then we observe the second-order bias and MSE with the expression in Proposition 4.

C.3 Proof of Proposition 5: The moment condition is

$$\Psi_N(\beta) = \frac{1}{N}\sum_{i=1}^{N} s_i(\beta) \tag{2.49}$$

where $s_i(\beta) = (\alpha - \mathbf{1}(y_i < x_i'\beta))(-z_i)$. Since x_i are assumed to be i.i.d., then s_i and d_i are i.i.d. as well. Similarly, V_i and W_i are i.i.d. matrices. We have

$$H_1 = \nabla_\beta^1 s_i = \nabla_\beta^1[(\alpha - \mathbf{1}(y_i < x_i'\beta))(-z_i)] = z_i x_i' \delta(x_i'\beta - y_i),$$

$$H_2 = \nabla_\beta^2 s_i = -\left(z_i x_i'\right) \otimes x_i' \delta^{(1)}(x_i'\beta - y_i),$$

$$H_3 = \nabla_\beta^3 s_i = \left(z_i x_i'\right) \otimes x_i' \otimes x_i' \delta^{(2)}(x_i'\beta - y_i),$$

$$\begin{aligned} \overline{H_1} &= E\nabla_\beta^1 s_i = E\left[z_i x_i' \delta(x_i'\beta - y_i)\right] = E\left[z_i x_i' E\left(\delta(x_i'\beta - y_i)|x_i, z_i\right)\right] \\ &= E\left[z_i x_i' \int_{-\infty}^{+\infty} \delta(y_i - x_i'\beta) f(y_i) dy\right] = E\left[z_i x_i' f(x_i'\beta)\right], \end{aligned}$$

$$\overline{H_2} = E\nabla_\beta^2 s_i = -E\left[\left(z_i x_i'\right) \otimes x_i' \delta^{(1)}(x_i'\beta - y_i)\right] = E\left[\left(z_i x_i'\right) \otimes x_i' f^{(1)}(x_i'\beta)\right],$$

$$\overline{H_3} = E\nabla_\beta^3 s_i = E\left[\left(z_i x_i'\right) \otimes x_i' \otimes x_i' \delta^{(2)}(x_i'\beta - y_i)\right] = E[\left(z_i x_i'\right) \otimes x_i' \otimes x_i' f^{(2)}(x_i'\beta)],$$

$$Q = \left(\overline{H_1}\right)^{-1} = \left(E[z_i x_i' f(x_i'\beta)]\right)^{-1},$$

$$V_1 = H_1 - \overline{H_1} = z_i x_i' \delta(x_i'\beta - y_i) - E[f(x_i'\beta) z_i x_i'],$$

$$W_i = H_2 - \overline{H_2} = -\left(z_i x_i'\right) \otimes x_i' \delta^{(1)}(x_i'\beta - y_i) + E[\left(z_i x_i'\right) \otimes x_i' f^{(1)}(x_i'\beta)],$$

$$d_i = Q s_i = Q(\alpha - \mathbf{1}(y_i < x_i'\beta))(-z_i),$$

where $f(x_1'\beta)$ is the density of $y|x$, at the point $y_1 = x_1'\beta$. We observe that Ψ_N, s_i and d_i are all $k \times 1$ vectors. H_1, $\overline{H_1}$, Q, and V_i are all $k \times k$ matrices, H_2, $\overline{H_2}$ and W_i are all $k \times k^2$ matrices. H_3 and $\overline{H_3}$ are $k \times k^3$ matrices. Then we have

$$\begin{aligned}
\overline{V_i d_i} &= E\left[\left(H_1 - \overline{H_1}\right) Q s_i\right] \\
&= E\left(H_1 Q s_i\right) - E\left(s_i\right) \\
&= E\left[z_i x_i' \delta(x_i'\beta - y_i) Q s_i\right] \\
&= E\left[z_i x_i' E\left(\delta(x_i'\beta - y_i) Q s_i | x_i\right)\right] \\
&= E\left[z_i x_i' \int_{-\infty}^{+\infty} \delta(x_i'\beta - y_i) Q(\alpha - \mathbf{1}(y_i < x_i'\beta))(-z_i) f(y_i) \mathrm{d}y_i\right] \\
&= E\left[-z_i x_i' Q z_i \alpha \int_{-\infty}^{+\infty} \delta(x_i'\beta - y_i) f(y_i) \mathrm{d}y_i + z_i x_i' Q z_i \int_{-\infty}^{+\infty} \delta(x_i'\beta - y_i) \phi(x_i'\beta - y_i) f(y_i) \mathrm{d}y_i\right] \\
&= E\left[-z_i x_i' Q z_i \alpha f(x_i'\beta) + \frac{1}{2} z_i x_i' Q z_i f(x_i'\beta)\right] \\
&= \left(\frac{1}{2} - \alpha\right) E\left[z_i x_i' Q z_i f(x_i'\beta)\right].
\end{aligned}$$

$$\begin{aligned}
\overline{d_i \otimes d_i} &= E\left[\left(Q s_i \otimes Q s_i\right)\right] \\
&= E\left[\left(Q \otimes Q\right)\left(s_i \otimes s_i\right)\right] \\
&= \left(Q \otimes Q\right) E\left[\left(s_i \otimes s_i\right)\right] \\
&= \left(Q \otimes Q\right) E\left[E\left(s_i \otimes s_i | x_i\right)\right] \\
&= \left(Q \otimes Q\right) E\left[\left(z_i \otimes z_i\right) E\left((\alpha - \mathbf{1}(y_i < x_i'\beta))^2 | x_i\right)\right] \\
&= \left(Q \otimes Q\right) E\left(z_i \otimes z_i\right)\left[\left(\alpha - 1\right)^2 \alpha + \alpha^2\left(1 - \alpha\right)\right] \\
&= \alpha(1 - \alpha)\left(Q \otimes Q\right) E\left(z_i \otimes z_i\right).
\end{aligned}$$

Therefore, the second-order bias of $\widehat{\beta}$ up to $O(N^{-1})$ can be rewritten as

$$\begin{aligned} B(\widehat{\beta}) &= \frac{1}{N}Q\left[\overline{V_i d_i} - \frac{1}{2}\overline{H_2}\left(\overline{d_i \otimes d_i}\right)\right] \\ &= \frac{1}{N}Q\left[\left(\frac{1}{2}-\alpha\right)E\left[z_i x_i' Q z_i f(x_i'\beta)\right] - \frac{\alpha(1-\alpha)}{2}E\left[\left(z_i x_i'\right)\otimes x_i' f^{(1)}(x_i'\beta)\right]\left(Q\otimes Q\right)E\left(z_i \otimes z_i\right)\right], \end{aligned}$$

where $Q = \left(E[z_i x_i' f(x_i'\beta)]\right)^{-1}$. When x_i and u_i are i.i.d., $f(0|x_i) = f(0)$. Since the density of y_i evaluated at $y_i = x_i'\beta$ is the same as the density of u_i evaluated at $u_i = 0$. If we use $f(0)$ to denote the conditional density of u_i evaluated at $u_i = 0$, the second-order bias of $\widehat{\beta}$ up to $O(N^{-1})$ can be rewritten as

$$B(\widehat{\beta}) = \frac{1}{N}Q\left[\left(\frac{1}{2}-\alpha\right)E\left[z_i x_i' Q z_i f(0)\right] - \frac{\alpha(1-\alpha)}{2}E\left[\left(z_i x_i'\right)\otimes x_i' f^{(1)}(0)\right]\left(Q\otimes Q\right)E\left(z_i \otimes z_i\right)\right],$$

where $Q = \left(E[z_i x_i' f(0)]\right)^{-1}$.

When $l = k = 1$, the MSE up to $O(N^{-2})$ can be written as

$$\begin{aligned} M(\widehat{\beta}) &= \frac{1}{N}\overline{d_i^2} - 2\frac{1}{N^2}Q\left[\overline{V_i d_i^2} - \frac{1}{2}\overline{H_2}\,\overline{d_i^3}\right] + 6\frac{1}{N^2}Q^2\overline{V_i d_i}^2 + 3\frac{1}{N^2}Q^2\overline{V_i^2}\,\overline{d_i^2} \\ &\quad + 3\frac{1}{N^2}Q\overline{W_i d_i}\,\overline{d_i^2} - 12\frac{1}{N^2}Q^2\overline{H_2}\overline{V_i d_i}\,\overline{d_i^2} + \frac{15}{4}\frac{1}{N^2}Q^2\overline{H_2}^2\,\overline{d_i^2}^2 - \frac{1}{N^2}Q\overline{H_3}\,\overline{d_i^2}^2, \end{aligned}$$

where we have

$$\overline{V_i d_i^2} = E\left[\left(H_1 - \overline{H_1}\right)Q^2 s_i^2\right] = Q^2 E\left(H_1 s_i^2\right) - QE\left(s_i^2\right),$$

$$\begin{aligned} E(H_1 s_i^2) &= E\left[z_i x_i \delta(x_i'\beta - y_i) s_i^2\right] \\ &= E\left[z_i x_i E\left(\delta(x_i'\beta - y_i) s_i^2 | x_i\right)\right] \\ &= E\left[z_i^3 x_i \int_{-\infty}^{+\infty} \delta(x_i\beta - y_i)(\alpha - \mathbf{1}(y_i < x_i'\beta))^2 f(y_i)\mathrm{d}y_i\right] \\ &= E\left[\alpha^2 z_i^3 x_i \int_{-\infty}^{+\infty} \delta(x_i\beta - y_i) f(y_i)\mathrm{d}y_i + (1-2\alpha) z_i^3 x_i \int_{-\infty}^{+\infty} \delta(x_i\beta - y_i)\phi(x_i\beta - y_i) f(y_i)\mathrm{d}y_i\right] \\ &= E\left[\alpha^2 z_i^3 x_i f(x_i\beta)\right] + E\left[(1-2\alpha)\frac{1}{2} z_i^3 x_i f(x_i\beta)\right] \\ &= \left(\alpha^2 - \alpha + \frac{1}{2}\right) E\left[z_i^3 x_i f(x_i'\beta)\right]. \end{aligned}$$

$$
\begin{aligned}
\overline{d_i^2} &= Q^2 E(s_i^2) \\
&= Q^2 E\left[z_i^2(\alpha - \mathbf{1}(y_i < x_i'\beta))^2\right] \\
&= Q^2 \left[(\alpha-1)^2 \alpha + \alpha^2 (1-\alpha)\right] E\left(z_i^2\right) \\
&= \alpha(1-\alpha)Q^2 E\left(z_i^2\right),
\end{aligned}
$$

$$
\begin{aligned}
\overline{d_i^3} &= Q^3 E(s_i^3) \\
&= Q^3 E\left[z_i^3(\alpha - \mathbf{1}(y_i < x_i'\beta))^3\right] \\
&= Q^3 \left[(\alpha-1)^3 \alpha + \alpha^3(1-\alpha)\right] E\left(z_i^3\right) \\
&= \alpha(1-\alpha)(2\alpha-1)Q^3 E\left(z_i^3\right),
\end{aligned}
$$

$$
\overline{V_i d_i}^2 = \left(\frac{1}{2} - \alpha\right)^2 Q^2 \left(E\left[z_i^2 x_i f(x_i'\beta)\right]\right)^2
$$

$$
\begin{aligned}
\overline{V_i^2} &= E\left[\left(H_1 - \overline{H_1}\right)^2\right] \\
&= E\left[H_1^2 - 2H_1\overline{H_1} + \overline{H_1}^2\right] \\
&= E\left(H_1^2\right) - 2\overline{H_1}^2 + \overline{H_1}^2 \\
&= E\left(H_1^2\right) - \overline{H_1}^2 \\
&= E\left[z_i^2 x_i^2 \left(\delta(x_i'\beta - y_i)\right)^2\right] - \left(E\left[z_i x_i f(x_i'\beta)\right]\right)^2 \\
&= E\left[z_i^2 x_i^2 \int_{-\infty}^{+\infty} \left(\delta(x_i'\beta - y_i)\right)^2 f(y_i)\mathrm{d}y_i\right] - \left(E\left[z_i x_i f(x_i'\beta)\right]\right)^2,
\end{aligned}
$$

$$
\begin{aligned}
\overline{W_i d_i} &= E\left[\left(H_2 - \overline{H_2}\right) Q s_i\right] \\
&= QE\left(H_2 s_i\right) - Q\overline{H_2}E\left(s_i\right) \\
&= QE\left[z_i x_i^2 \delta^{(1)}(x_i'\beta - y_i) s_i\right] \\
&= QE\left[z_i x_i^2 E\left(\delta^{(1)}(x_i'\beta - y_i)\left(\alpha - \mathbf{1}(y_i < x_i'\beta)\right)(-z_i)|x_i\right)\right] \\
&= -QE\left[z_i^2 x_i^2 E\left(\delta^{(1)}(x_i'\beta - y_i)\left(\alpha - \phi\left(x_i\beta - y_i\right)\right)|x_i\right)\right] \\
&= -\alpha QE\left[z_i^2 x_i^2 \int_{-\infty}^{+\infty} \delta^{(1)}(x_i'\beta - y_i) f(y_i)\mathrm{d}y_i\right] + QE\left[z_i^2 x_i^2 \int_{-\infty}^{+\infty} \delta^{(1)}(x_i'\beta - y_i)\phi\left(x_i\beta - y_i\right) f(y_i)\mathrm{d}y_i\right] \\
&= -\alpha QE\left[z_i^2 x_i^2 f^{(1)}(x_i'\beta)\right] + QE\left[-z_i^2 x_i^2 \int_{-\infty}^{+\infty} \left(\delta(x_i'\beta - y_i)\right)^2 f(y_i)\mathrm{d}y_i\right] \\
&\quad + QE\left[z_i^2 x_i^2 \int_{-\infty}^{+\infty} \delta(x_i'\beta - y_i)\phi\left(x_i\beta - y_i\right) f^{(1)}(y_i)\mathrm{d}y_i\right] \\
&= \left(\frac{1}{2} - \alpha\right) QE\left[z_i^2 x_i^2 f^{(1)}(x_i'\beta)\right] - QE\left[z_i^2 x_i^2 \int_{-\infty}^{+\infty} \left(\delta(x_i'\beta - y_i)\right)^2 f(y_i)\mathrm{d}y_i\right].
\end{aligned}
$$

Therefore, the MSE up to $O(N^{-2})$ can be written as

$$
\begin{aligned}
M(\widehat{\beta}) &= \frac{1}{N}\alpha(1-\alpha)Q^2 E\left(z_i^2\right) - 2\frac{1}{N^2}Q^3\left(\alpha^2 - \alpha + \frac{1}{2}\right)E\left[z_i^3 x_i\right] f(x_i'\beta) - \frac{1}{N^2}\alpha(1-\alpha)Q^2 E\left(z_i^2\right) \\
&\quad + \frac{1}{N^2}\alpha(1-\alpha)(2\alpha-1)Q^4 E\left[z_i x_i^2\right] E\left(z_i^3\right) f^{(1)}(x_i'\beta) + 6\frac{1}{N^2}\left(\frac{1}{2} - \alpha\right)^2 Q^4\left(E\left[z_i^2 x_i\right] f(x_i'\beta)\right)^2 \\
&\quad + 3\frac{1}{N^2}\alpha(1-\alpha)Q^4\left(\frac{1}{2} - \alpha\right)E\left[z_i^2 x_i^2\right] E\left(z_i^2\right) f^{(1)}(x_i'\beta) \\
&\quad - 12\frac{1}{N^2}\left(\frac{1}{2} - \alpha\right)\alpha(1-\alpha)Q^5 E[z_i x_i^2] E\left[z_i^2 x_i\right] E\left(z_i^2\right) f(x_i'\beta) f^{(1)}(x_i'\beta) \\
&\quad + \frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^6\left(E\left[z_i x_i^2\right] f^{(1)}(0)\right)^2\left(E\left(z_i^2\right)\right)^2 \\
&\quad - \frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^5 E[z_i x_i^3]\left(E\left(z_i^2\right)\right)^2 f^{(2)}(x_i'\beta),
\end{aligned}
$$

where $Q = \left(E\left(z_i x_i\right) f(x_i'\beta)\right)^{-1}$. Since the density of y_i evaluated at $y_i = x_i'\beta$ is the same as the density of u_i evaluated at $u_i = 0$. If we use $f\left(0\right)$ to denote the density of u_i evaluated at $u_i = 0$, then we observe the MSE with the expression in Proposition 5.

2.7.4 Appendix D

This appendix provides the results tables with the second data generating process DGP (DGP 2). In DGP 2, the error term u_i is uniformly distributed with the CDF $F(\cdot)$ on $[a, b]$, then $a = \frac{\alpha}{\alpha - 1} b$. We have

$$F(0) = \int_{-\infty}^{0} f(u)du = \int_{a}^{0} \frac{1}{b-a} du = -\frac{a}{b-a} = \alpha.$$

Therefore, we generate the error term u_i from uniform distribution on $[a, b]$, where $a = -\alpha R$, $b = R(1 - \alpha)$, and the range $R = b - a$.

Table 8 and Table 9 shows the results when $R = 4, 10$, and x_i is generated from exponential distribution, $f(x_i) = \exp(-x_i)$. Table 10 and Table 11 show the results when $R = 4$, and when x_i is generated from two different mixture normal distributions in Marron and Wand (1992).

2.8 References

Amemiya, T., 1980. The n^2-order mean squared errors of the maximum likelihood and the minimum logit chi-square estimators. *Annals of Mathematical Statistics* 8, 488-505.

Andrews, D.W.K., 1994. Empirical process methods in econometrics. In: *Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics* Vol. 4. *Elsevier Science, Amsterdam,* pp. 2247-2294.

Bahadur, R.R., 1966. A note on quantiles in large samples. *Ann. Math. Statist.* 37, 577-581.

Bao, Y., Ullah, A., 2007. The second-order bias and mean squared error of estimators in time-series models. *Journal of Econometrics* 140, 650-669.

Basmann, R.L., 1974. Exact finite sample distribution for some econometric estimators and test statistics: a survey and appraisal. In: Intriligator, M.D., Kendrick, D.A. (Eds.), *Frontiers of Quantitative Economics,* Vol. 2. North-Holland, Amsterdam, pp. 209-288.

Chernozhukov, V., Hansen, C., 2006. Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics* 132, 491-525.

Chernozhukov, V., Hong, H., 2003. An MCMC approach to classical estimation. *Journal of Econometrics* 115, 293-346.

Chesher, A.D., Spady, R., 1989. Asymptotic expansions of the information matrix test statistic. *Econometrica* 59, 787-816.

Cordeiro, G.M., McCullagh, P., 1991. Bias correction in generalized linear models. *Journal of the Royal Statistical Society B* 53, 629-643.

Elliott, G., Komunjer, I., Timmermann, A., 2005. Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies* 72(4): 1107-1125.

Gelfand, I.M., Shilov, G.E., 1964. *Generalized Functions, Vol.1.* New York: Academic Press.

He, X., Shao, Q., 1996. A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *The Annals of Statistics* 24(6): 2608-2630.

Huber, P.J., 1967. The behavior of maximum likelihood estimates under nonstandard conditions. In: *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability, vol.1. University of California Press, Berkeley.*

Jureckova, J., Sen, P.K., 1987. A second-order asymptotic distributional representation of M-estimators with discontinuous score functions. *The Annals of Probability* 15(2): 814-823.

Jureckova, J., Sen, P.K., 1996. *Robust Statistical Procedures: Asymptotics and Interrelations.* New York : Wiley.

Kiefer, J., 1967. On Bahadur representation of sample quantiles. *Ann. Math. Statist.* 38, 1323-1342.

Kim, J., Pollard, D., 1990. Cube root asymptotics. *Annals of Statistics.* 18, 191-219.

Koenker, R., Bassett, G.S., 1978. Regression quantiles. *Econometrica* 46, 33–50.

Koenker, R., 2005. *Quantile regression.* Cambridge university press.

Komunjer, I., 2005. Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics* 128, 137-164.

Marron, J. S., Wand, M. P., 1992. Exact mean integrated squared error. *The Annals of Statistics*, 20, 712-736.

Nagar, A.L., 1959. The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica* 27, 575-595.

Newey, W.K., McFadden, D.L., 1994. Large sample estimation and hypothesie testing. In: *Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics vol. 4. Elsevier Science, Amsterdam,* pp. 2113-2247.

Newey, W.K., Smith, R.J., 2004. Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72, 219-255.

Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. *Econometrica* 57, 1027-1057.

Phillips, P.C.B., 1991. A shortcut to LAD estimator asymptotics. *Econometric Theorey* 7, 450-463.

Pollard, D., 1985. New ways to prove central limit theorems. *Econometric Theorey* 1, 295-314.

Prakasa Rao, B.L.S., 1969. Estimation of a unimodal density. *Sankhya Series A* 31, 23-36.

Raju, C. K., 1982. Products and compositions with the Dirac delta function. *Journal of Physics A: Mathematical and General* 15, 381-396.

Rilstone, P., Srivastava, V. K., Ullah, A., 1996. The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics* 75, 369-395.

Rothenberg, T.J., 1984. Approximating the distribution of econometric estimators and test statistics. In: Intriligator, M.L., Griliches, A. (Eds.), *Handbook of Econometrics*, Vol.2. North-Holland, Amsterdam, pp. 881-935.

Sargan, J.D., 1974. The validity of Nagar's expansion for the moments of econometric estimators. *Econometrica* 42, 169-176.

Sargan, J.D., 1976. Econometric estimators and the Edgeworth approximation. *Econometrica* 44, 421-448.

Ullah, A., 2004. *Finite sample econometrics.* United Kingdom: Oxford University Press.

van der Vaart, A.W., 1998. *Asymptotic Statistics.* New York: Cambridge University Press.

Table 2.1: Bias correction and MSE with x_i generated from exponential distribution, DGP 1, allowing hetroskedasticity

	$\sigma_{ui} = 0.1x_i, N = 50$			$\sigma_{ui} = 0.5x_i, N = 50$		
α	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$
0.05	0.0027	0.0023	-0.0004	0.0135	0.0203	0.0067
	0.0000	0.0018	0.0018	0.0002	0.0447	0.0443
0.1	0.0016	0.0018	0.0002	0.0080	0.0087	0.0008
	0.0006	0.0012	0.0012	0.0155	0.0291	0.0290
0.2	0.0008	0.0013	0.0005	0.0042	0.0081	0.0039
	0.0006	0.0008	0.0008	0.0162	0.0205	0.0204
0.3	0.0005	0.0002	-0.0003	0.0023	0.0025	0.0002
	0.0006	0.0007	0.0007	0.0151	0.0169	0.0169
0.4	0.0002	0.0003	0.0001	0.0011	0.0014	0.0003
	0.0006	0.0007	0.0007	0.0145	0.0159	0.0159
0.5	0.0000	-0.0001	-0.0001	0.0000	0.0005	0.0005
	0.0006	0.0006	0.0006	0.0143	0.0155	0.0155
0.6	-0.0002	-0.0001	0.0001	-0.0011	0.0002	0.0013
	0.0006	0.0006	0.0006	0.0145	0.0160	0.0160
0.7	-0.0005	-0.0005	-0.0001	-0.0023	-0.0018	0.0006
	0.0006	0.0007	0.0007	0.0151	0.0174	0.0174
0.8	-0.0008	-0.0012	-0.0003	-0.0042	-0.0038	0.0003
	0.0006	0.0008	0.0008	0.0162	0.0207	0.0207
0.9	-0.0016	-0.0017	-0.0001	-0.0080	-0.0093	-0.0014
	0.0006	0.0012	0.0012	0.0155	0.0302	0.0301
0.95	-0.0027	-0.0034	-0.0007	-0.0135	-0.0208	-0.0072
	0.0000	0.0017	0.0017	0.0002	0.0434	0.0430

Notes: This table present the simulation results, when u_i is generated from normal distribution, x_i is generated form exponential distribution, when allowing hetroskedasticity. For each level of α, the first row is for bias and the second row is for the MSE of the quantile estimator. For each panel, the first column presents the second-order bias and MSE derived by Theorem 1 and 2, the second column presents the Monte Carlo simulation bias and MSE of quantile estimators $\hat{\beta}$, the third column presents the Monte Carlo simulation bias and

MSE of the bias corrected quantile estimators $\tilde{\beta}$ where $\tilde{\beta} = \hat{\beta} - B(\hat{\beta})$. We set $\beta = 0$ and $N = 50$, and the results are computed from 10,000 Monte Carlo replications.

Table 2.2: Bias correction and MSE with x_i generated from exponential distribution, DGP 1, $\sigma_u = 0.5$

	$\sigma_u = 0.5, N = 60$			$\sigma_u = 0.5, N = 100$		
α	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$
0.05	0.0086	0.0092	0.0005	0.0052	0.0060	0.0009
	0.0038	0.0102	0.0101	0.0036	0.0059	0.0059
0.1	0.0051	0.0051	0.0000	0.0031	0.0032	0.0002
	0.0044	0.0067	0.0067	0.0030	0.0038	0.0038
0.2	0.0027	0.0039	0.0013	0.0016	0.0023	0.0007
	0.0037	0.0046	0.0045	0.0024	0.0026	0.0026
0.3	0.0015	0.0009	-0.0006	0.0009	0.0017	0.0008
	0.0033	0.0039	0.0039	0.0021	0.0023	0.0023
0.4	0.0007	0.0000	-0.0007	0.0004	0.0010	0.0006
	0.0031	0.0036	0.0036	0.0019	0.0021	0.0021
0.5	0.0000	0.0000	0.0000	0.0000	-0.0002	-0.0002
	0.0031	0.0035	0.0035	0.0019	0.0020	0.0020
0.6	-0.0007	0.0002	0.0009	-0.0004	0.0006	0.0010
	0.0031	0.0036	0.0036	0.0019	0.0021	0.0021
0.7	-0.0015	-0.0012	0.0003	-0.0009	-0.0002	0.0007
	0.0033	0.0040	0.0040	0.0021	0.0023	0.0023
0.8	-0.0027	-0.0025	0.0001	-0.0016	-0.0021	-0.0005
	0.0037	0.0045	0.0045	0.0024	0.0027	0.0027
0.9	-0.0051	-0.0051	0.0000	-0.0031	-0.0040	-0.0009
	0.0044	0.0065	0.0065	0.0030	0.0038	0.0038
0.95	-0.0086	-0.0094	-0.0008	-0.0052	-0.0063	-0.0011
	0.0038	0.0101	0.0100	0.0036	0.0059	0.0058

Notes: This table present the simulation results, when u_i is generated from normal distribution with $\sigma_u = 0.5$, x_i is generated form exponential distribution, u_i and x_i are i.i.d.. For each level of α, the first row is for bias and the second row is for the MSE of the quantile estimator. For each panel, the first column presents the second-order bias and MSE derived by by Corollary 1 and 2.2, the second column presents the Monte Carlo simulation bias and MSE of quantile estimators $\hat{\beta}$, the third column presents the Monte Carlo simulation bias and MSE of the bias corrected quantile estimators $\tilde{\beta}$ where $\tilde{\beta} = \hat{\beta} - B(\hat{\beta})$. We set $\beta = 0$ and $N = 60, 100$, and the results are computed from 10,000 Monte Carlo replications.

Table 2.3: Bias correction and MSE with x_i generated from exponential distribution, DGP 1, $\sigma_u = 0.1$

	$\sigma_u = 0.1, N = 60$			$\sigma_u = 0.1, N = 100$		
α	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$
0.05	0.0017	0.0021	0.0004	0.0010	0.0009	-0.0002
	0.0002	0.0004	0.0004	0.0001	0.0002	0.0002
0.1	0.0010	0.0011	0.0001	0.0006	0.0008	0.0002
	0.0002	0.0003	0.0003	0.0001	0.0002	0.0002
0.2	0.0005	0.0007	0.0002	0.0003	0.0004	0.0001
	0.0001	0.0002	0.0002	0.0001	0.0001	0.0001
0.3	0.0003	0.0003	0.0000	0.0002	0.0002	0.0001
	0.0001	0.0002	0.0002	0.0001	0.0001	0.0001
0.4	0.0001	0.0001	0.0000	0.0001	0.0003	0.0002
	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
0.5	0.0000	-0.0002	-0.0002	0.0000	0.0001	0.0001
	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
0.6	-0.0001	0.0000	0.0002	-0.0001	0.0000	0.0001
	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
0.7	-0.0003	-0.0003	0.0000	-0.0002	-0.0002	0.0000
	0.0001	0.0002	0.0002	0.0001	0.0001	0.0001
0.8	-0.0005	-0.0005	0.0001	-0.0003	-0.0003	0.0000
	0.0001	0.0002	0.0002	0.0001	0.0001	0.0001
0.9	-0.0010	-0.0010	0.0000	-0.0006	-0.0007	-0.0001
	0.0002	0.0003	0.0003	0.0001	0.0002	0.0002
0.95	-0.0017	-0.0020	-0.0003	-0.0010	-0.0010	0.0000
	0.0002	0.0004	0.0004	0.0001	0.0002	0.0002

Notes: This table present the simulation results, when u_i is generated from normal distribution with $\sigma_u = 0.1$, x_i is generated form exponential distribution, u_i and x_i are i.i.d.. For each level of α, the first row is for bias and the second row is for the MSE of the quantile estimator. For each panel, the first column presents the second-order bias and MSE derived by by Corollary 1 and 2.2, the second column presents the Monte Carlo simulation bias and MSE of quantile estimators $\hat{\beta}$, the third column presents the Monte Carlo simulation bias and MSE of the bias corrected quantile estimators $\tilde{\beta}$ where $\tilde{\beta} = \hat{\beta} - B(\hat{\beta})$. We set $\beta = 0$ and $N = 60, 100$, and the results are computed from 10,000 Monte Carlo replications.

Table 2.4: Bias correction and MSE with x_i generated from mixture normal distribution (skewed unimodal), DGP 1, $\sigma_u = 0.5$

	$\sigma_u = 0.5, N = 60$			$\sigma_u = 0.5, N = 100$		
α	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$
0.05	0.0058	0.0046	-0.0012	0.0035	0.0027	-0.0007
	0.0119	0.0174	0.0173	0.0079	0.0105	0.0104
0.1	0.0034	0.0018	-0.0016	0.0021	0.0014	-0.0006
	0.0089	0.0111	0.0111	0.0055	0.0065	0.0065
0.2	0.0018	0.0018	0.0000	0.0011	0.0017	0.0006
	0.0065	0.0077	0.0077	0.0040	0.0047	0.0047
0.3	0.0010	0.0013	0.0003	0.0006	0.0000	-0.0006
	0.0057	0.0067	0.0067	0.0034	0.0039	0.0039
0.4	0.0005	0.0015	0.0011	0.0003	0.0010	0.0007
	0.0053	0.0061	0.0061	0.0032	0.0036	0.0036
0.5	0.0000	-0.0014	-0.0014	0.0000	-0.0006	-0.0006
	0.0052	0.0060	0.0060	0.0031	0.0035	0.0035
0.6	-0.0005	-0.0013	-0.0008	-0.0003	-0.0002	0.0001
	0.0053	0.0061	0.0061	0.0032	0.0036	0.0036
0.7	-0.0010	-0.0005	0.0005	-0.0006	-0.0004	0.0002
	0.0057	0.0067	0.0067	0.0034	0.0040	0.0040
0.8	-0.0018	-0.0020	-0.0002	-0.0011	-0.0016	-0.0005
	0.0065	0.0077	0.0077	0.0040	0.0047	0.0047
0.9	-0.0034	-0.0032	0.0003	-0.0021	-0.0018	0.0002
	0.0089	0.0113	0.0113	0.0055	0.0067	0.0067
0.95	-0.0058	-0.0055	0.0003	-0.0035	-0.0029	0.0006
	0.0119	0.0175	0.0174	0.0079	0.0103	0.0102

Notes: This table present the simulation results, when u_i is generated from normal distribution with $\sigma_u = 0.5$, x_i is generated form mixture normal distribution, u_i and x_i are i.i.d.. For each level of α, the first row is for bias and the second row is for the MSE of the quantile estimator. For each panel, the first column presents the second-order bias and MSE derived by Corollary 1 and 2.2, the second column presents the Monte Carlo simulation bias and MSE of quantile estimators $\hat{\beta}$, the third column presents the Monte Carlo simulation bias and MSE of the bias corrected quantile estimators $\tilde{\beta}$ where $\tilde{\beta} = \hat{\beta} - B(\hat{\beta})$. We set $\beta = 0$ and $N = 60, 100$, and the results are computed from 10,000 Monte Carlo replications.

Table 2.5: Bias correction and MSE with x_i generated from mixture normal distribution (strongly skewed), DGP 1, $\sigma_u = 0.5$

	$\sigma_u = 0.5, N = 60$			$\sigma_u = 0.5, N = 100$		
α	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$
0.05	-0.0030	-0.0023	0.0006	-0.0018	-0.0027	-0.0009
	0.0033	0.0070	0.0070	0.0021	0.0040	0.0040
0.1	-0.0017	-0.0012	0.0005	-0.0010	-0.0010	0.0000
	0.0024	0.0043	0.0043	0.0015	0.0026	0.0026
0.2	-0.0009	-0.0008	0.0002	-0.0005	0.0001	0.0007
	0.0017	0.0030	0.0030	0.0010	0.0018	0.0018
0.3	-0.0005	-0.0013	-0.0008	-0.0003	0.0000	0.0003
	0.0015	0.0025	0.0025	0.0009	0.0016	0.0016
0.4	-0.0002	-0.0002	0.0001	-0.0001	-0.0005	-0.0003
	0.0014	0.0024	0.0024	0.0008	0.0014	0.0014
0.5	0.0000	0.0003	0.0003	0.0000	0.0001	0.0001
	0.0013	0.0023	0.0023	0.0008	0.0014	0.0014
0.6	0.0002	0.0010	0.0008	0.0001	0.0000	-0.0001
	0.0014	0.0024	0.0024	0.0008	0.0014	0.0014
0.7	0.0005	0.0001	-0.0004	0.0003	0.0011	0.0008
	0.0015	0.0026	0.0026	0.0009	0.0016	0.0016
0.8	0.0009	0.0003	-0.0006	0.0005	-0.0001	-0.0007
	0.0017	0.0030	0.0030	0.0010	0.0019	0.0019
0.9	0.0017	0.0021	0.0004	0.0010	0.0005	-0.0005
	0.0024	0.0044	0.0043	0.0015	0.0027	0.0027
0.95	0.0030	0.0010	-0.0020	0.0018	0.0015	-0.0003
	0.0033	0.0069	0.0069	0.0021	0.0040	0.0040

Notes: This table present the simulation results, when u_i is generated from normal distribution with $\sigma_u = 0.5$, x_i is generated form exponential distribution, u_i and x_i are i.i.d.. For each level of α, the first row is for bias and the second row is for the MSE of the quantile estimator. For each panel, the first column presents the second-order bias and MSE derived by Corollary 1 and 2.2, the second column presents the Monte Carlo simulation bias and MSE of quantile estimators $\hat{\beta}$, the third column presents the Monte Carlo simulation bias and MSE of the bias corrected quantile estimators $\tilde{\beta}$ where $\tilde{\beta} = \hat{\beta} - B(\hat{\beta})$. We set $\beta = 0$ and $N = 60, 100$, and the results are computed from 10,000 Monte Carlo replications.

Table 2.6: Bias correction and MSE in unconditional quantile model, DGP 1, $\sigma_u = 0.5$

	$\sigma_u = 0.5, N = 60$			$\sigma_u = 0.5, N = 100$		
α	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$
0.05	0.0058	-0.0166	-0.0223	0.0035	-0.0149	-0.0183
	0.0183	0.0177	0.0179	0.0110	0.0114	0.0115
0.1	0.0034	0.0117	0.0083	0.0020	0.0039	0.0019
	0.0126	0.0105	0.0104	0.0075	0.0068	0.0068
0.2	0.0018	0.0141	0.0124	0.0011	0.0080	0.0070
	0.0090	0.0082	0.0081	0.0053	0.0050	0.0050
0.3	0.0010	0.0104	0.0094	0.0006	0.0054	0.0048
	0.0078	0.0070	0.0070	0.0045	0.0043	0.0043
0.4	0.0005	0.0073	0.0069	0.0003	0.0053	0.0050
	0.0072	0.0066	0.0066	0.0042	0.0040	0.0040
0.5	0.0000	0.0002	0.0002	0.0000	0.0006	0.0006
	0.0070	0.0060	0.0060	0.0041	0.0036	0.0036
0.6	-0.0005	-0.0087	-0.0083	-0.0003	-0.0045	-0.0042
	0.0072	0.0067	0.0066	0.0042	0.0040	0.0039
0.7	-0.0010	-0.0083	-0.0073	-0.0006	-0.0064	-0.0058
	0.0078	0.0070	0.0070	0.0045	0.0043	0.0043
0.8	-0.0018	-0.0148	-0.0130	-0.0011	-0.0087	-0.0077
	0.0090	0.0083	0.0083	0.0053	0.0050	0.0050
0.9	-0.0034	-0.0100	-0.0066	-0.0020	-0.0045	-0.0024
	0.0126	0.0108	0.0108	0.0075	0.0066	0.0066
0.95	-0.0058	0.0147	0.0205	-0.0035	0.0133	0.0167
	0.0183	0.0177	0.0179	0.0110	0.0112	0.0113

Notes: This table present the simulation results for unconditional quantile regression with $x_i = 1$, when u_i is generated from normal distribution with $\sigma_u = 0.5$, u_i is i.i.d.. For each level of α, the first row is for bias and the second row is for the MSE of the quantile estimator. For each panel, the first column presents the second-order bias and MSE derived by Proposition 3, the second column presents the Monte Carlo simulation bias and MSE of quantile estimators $\hat{\beta}$, the third column presents the Monte Carlo simulation bias and MSE of the bias corrected quantile estimators $\tilde{\beta}$ where $\tilde{\beta} = \hat{\beta} - B(\hat{\beta})$. We set $\beta = 0$ and $N = 60, 100$, and the results are computed from 10,000 Monte Carlo replications.

Table 2.7: Bias correction and MSE with binary independent variable, DGP 1, $p = 0.3$, $\sigma_u = 0.5$

	$\sigma_u = 0.5, N = 60$			$\sigma_u = 0.5, N = 100$		
α	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$
0.05	0.0192	-0.0138	-0.0330	0.0115	0.0169	0.0054
	0.0402	0.0656	0.0665	0.0294	0.0364	0.0362
0.1	0.0113	0.0116	0.0003	0.0068	0.0099	0.0031
	0.0337	0.0414	0.0413	0.0219	0.0246	0.0246
0.2	0.0059	0.0115	0.0056	0.0035	0.0075	0.0040
	0.0260	0.0286	0.0285	0.0162	0.0168	0.0168
0.3	0.0033	0.0058	0.0025	0.0020	0.0036	0.0016
	0.0228	0.0253	0.0253	0.0140	0.0150	0.0150
0.4	0.0015	0.0064	0.0048	0.0009	0.0033	0.0024
	0.0214	0.0233	0.0232	0.0130	0.0136	0.0136
0.5	0.0000	0.0012	0.0012	0.0000	0.0003	0.0003
	0.0209	0.0204	0.0204	0.0128	0.0124	0.0124
0.6	-0.0015	-0.0091	-0.0076	-0.0009	-0.0028	-0.0019
	0.0214	0.0229	0.0228	0.0130	0.0136	0.0136
0.7	-0.0033	-0.0052	-0.0019	-0.0020	-0.0051	-0.0031
	0.0228	0.0248	0.0247	0.0140	0.0146	0.0146
0.8	-0.0059	-0.0109	-0.0050	-0.0035	-0.0096	-0.0060
	0.0260	0.0291	0.0290	0.0162	0.0174	0.0174
0.9	-0.0113	-0.0146	-0.0033	-0.0068	-0.0090	-0.0022
	0.0337	0.0405	0.0403	0.0219	0.0250	0.0249
0.95	-0.0192	0.0140	0.0332	-0.0115	-0.0160	-0.0045
	0.0402	0.0668	0.0677	0.0294	0.0366	0.0364

Notes: This table present the simulation results when u_i is generated from normal distribution with $\sigma_u = 0.5$, x_i is binary and x_i=1 with probability 0.3, u_i is i.i.d.. For each level of α, the first row is for bias and the second row is for the MSE of the quantile estimator. For each panel, the first column presents the second-order bias and MSE derived by Proposition 4, the second column presents the Monte Carlo simulation bias and MSE of quantile estimators $\hat{\beta}$, the third column presents the Monte Carlo simulation bias and MSE of the bias corrected quantile estimators $\tilde{\beta}$ where $\tilde{\beta} = \hat{\beta} - B(\hat{\beta})$. We set $\beta = 0$ and $N = 60, 100$, and the results are computed from 10,000 Monte Carlo replications.

Table 2.8: Bias correction for IVQR, $\sigma_u = 0.5$, N=60

	$\gamma = 0.5$			$\gamma = 0.9$		
α	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$
0.05	0.0155	0.1260	0.1105	0.0094	0.0215	0.0121
0.1	0.0093	0.0700	0.0607	0.0055	0.0070	0.0015
0.2	0.0050	0.0355	0.0305	0.0029	0.0025	-0.0004
0.3	0.0029	0.0235	0.0206	0.0017	0.0015	-0.0002
0.4	0.0014	0.0220	0.0206	0.0008	0.0020	0.0012
0.5	0.0000	0.0255	0.0255	0.0000	0.0005	0.0005
0.6	-0.0014	-0.0176	-0.0162	-0.0008	-0.0041	-0.0033
0.7	-0.0029	-0.0084	-0.0055	-0.0016	-0.0107	-0.0091
0.8	-0.0043	-0.0425	-0.0383	-0.0028	-0.0102	-0.0074
0.9	-0.0014	-0.0499	-0.0485	-0.0046	-0.0177	-0.0131
0.95	-0.0255	-0.0859	-0.0604	-0.0052	-0.0380	-0.0328

Notes: This table present the simulation results for IVQR, when u_i is generated from normal distribution with $\sigma_u = 0.5$; v_i is generated by $v_i = w_i + cu_i$, where w_i is from N(0,0.25), c=0.5; z_i is from exponential distribution with mean 1; x_i is generated from $x_i = z_i\gamma + u_i$, where $\gamma = 0.5, 0.9$; y_i is generated from $y_i = x_i\beta + u_i$, where $\beta = 0$. For each level of α, the numbers are bias of IVQR estimator. For each panel, the first column presents the second-order bias derived by Proposition 5, the second column presents the Monte Carlo simulation bias IVQR estimators $\hat{\beta}$, the third column presents the Monte Carlo simulation bias of the bias corrected IVQR estimators $\tilde{\beta}$ where $\tilde{\beta} = \hat{\beta} - B(\hat{\beta})$. We set N=60, 100, and the results are computed from 1,000 Monte Carlo replications.

Appendix D

Table 2.9: Bias correction and MSE with x_i generated from exponential distribution, DGP 2, $R = 4$

	$R=4, N=60$			$R=4, N=100$		
α	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$
0.05	0.0450	0.0369	-0.0081	0.0270	0.0243	-0.0027
	0.0063	0.0014	0.0098	0.0038	0.0014	0.0051
0.1	0.0400	0.0359	-0.0041	0.0240	0.0200	-0.0040
	0.0105	0.0013	0.0148	0.0066	0.0013	0.0079
0.2	0.0300	0.0267	-0.0033	0.0180	0.0190	0.0010
	0.0173	0.0010	0.0221	0.0114	0.0010	0.0131
0.3	0.0200	0.0145	-0.0055	0.0120	0.0096	-0.0024
	0.0222	0.0008	0.0271	0.0147	0.0008	0.0160
0.4	0.0100	0.0093	-0.0007	0.0060	0.0076	0.0016
	0.0251	0.0006	0.0312	0.0167	0.0006	0.0187
0.5	0.0000	0.0039	0.0039	0.0000	-0.0005	-0.0005
	0.0261	0.0004	0.0318	0.0174	0.0004	0.0195
0.6	-0.0100	-0.0115	-0.0015	-0.0060	-0.0085	-0.0025
	0.0251	0.0003	0.0313	0.0167	0.0003	0.0184
0.7	-0.0200	-0.0154	0.0046	-0.0120	-0.0102	0.0018
	0.0222	0.0001	0.0275	0.0147	0.0001	0.0165
0.8	-0.0300	-0.0264	0.0036	-0.0180	-0.0161	0.0019
	0.0173	0.0001	0.0222	0.0114	0.0001	0.0130
0.9	-0.0400	-0.0352	0.0048	-0.0240	-0.0214	0.0026
	0.0105	0.0000	0.0145	0.0066	0.0000	0.0081
0.95	-0.0450	-0.0393	0.0057	-0.0270	-0.0235	0.0035
	0.0063	0.0000	0.0100	0.0038	0.0000	0.0054

Notes: This table present the simulation results, when u_i is generated from uniform distribution with the range $R = 4$, x_i is generated form exponential distribution, u_i and x_i are i.i.d.. For each level of α, the first row is for bias and the second row is for the MSE of the quantile estimator. For each panel, the first column presents the second-order bias and MSE derived by Corollary 1 and 2.2, the second column presents the Monte Carlo simulation bias and MSE of quantile estimators $\hat{\beta}$, the third column presents the Monte Carlo simulation bias and MSE of the bias corrected quantile estimators $\tilde{\beta}$ where $\tilde{\beta} = \hat{\beta} - B(\hat{\beta})$. We set $\beta = 0$

and $N = 60, 100$, and the results are computed from 10,000 Monte Carlo replications.

Table 2.10: Bias correction and MSE with x_i generated from exponential distribution, DGP 2, $R = 10$

	$R = 10, N = 60$			$R = 10, N = 100$		
α	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$
0.05	0.1125	0.0945	-0.0180	0.0675	0.0576	-0.0099
	0.0394	0.0090	0.0627	0.0237	0.0090	0.0311
0.1	0.1000	0.0842	-0.0158	0.0600	0.0585	-0.0015
	0.0654	0.0081	0.0915	0.0416	0.0081	0.0518
0.2	0.0750	0.0656	-0.0094	0.0450	0.0427	-0.0023
	0.1082	0.0064	0.1418	0.0710	0.0064	0.0815
0.3	0.0500	0.0453	-0.0047	0.0300	0.0323	0.0023
	0.1388	0.0049	0.1726	0.0920	0.0049	0.1051
0.4	0.0250	0.0201	-0.0049	0.0150	0.0148	-0.0002
	0.1571	0.0036	0.1944	0.1046	0.0036	0.1163
0.5	0.0000	-0.0026	-0.0026	0.0000	-0.0034	-0.0034
	0.1632	0.0025	0.1995	0.1088	0.0025	0.1208
0.6	-0.0250	-0.0191	0.0059	-0.0150	-0.0079	0.0071
	0.1571	0.0016	0.1915	0.1046	0.0016	0.1147
0.7	-0.0500	-0.0427	0.0073	-0.0300	-0.0266	0.0034
	0.1388	0.0009	0.1773	0.0920	0.0009	0.1039
0.8	-0.0750	-0.0709	0.0041	-0.0450	-0.0427	0.0023
	0.1082	0.0004	0.1382	0.0710	0.0004	0.0798
0.9	-0.1000	-0.0863	0.0137	-0.0600	-0.0534	0.0066
	0.0654	0.0001	0.0912	0.0416	0.0001	0.0506
0.95	-0.1125	-0.0951	0.0174	-0.0675	-0.0588	0.0087
	0.0394	0.0000	0.0618	0.0237	0.0000	0.0321

Notes: This table present the simulation results, when u_i is generated from uniform distribution with the range $R = 10$, x_i is generated form exponential distribution, u_i and x_i are i.i.d.. For each level of α, the first row is for bias and the second row is for the MSE of the quantile estimator. For each panel, the first column presents the second-order bias and MSE derived by Corollary 1 and 2.2, the second column presents the Monte Carlo simulation bias and MSE of quantile estimators $\hat{\beta}$, the third column presents the Monte Carlo simulation bias and MSE of the bias corrected quantile estimators $\tilde{\beta}$ where $\tilde{\beta} = \hat{\beta} - B(\hat{\beta})$. We set $\beta = 0$ and $N = 60, 100$, and the results are computed from 10,000 Monte Carlo replications.

Table 2.11: Bias correction and MSE with x_i generated from mixture normal distribution (skewed unimodal), DGP 2, $R = 4$

	$R = 4, N = 60$			$R = 4, N = 100$		
α	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$
0.05	0.0304	0.0246	-0.0059	0.0182	0.0157	-0.0025
	0.0088	0.0014	0.0142	0.0056	0.0014	0.0079
0.1	0.0271	0.0252	-0.0019	0.0162	0.0156	-0.0006
	0.0174	0.0013	0.0228	0.0109	0.0013	0.0139
0.2	0.0203	0.0203	0.0000	0.0121	0.0092	-0.0029
	0.0314	0.0010	0.0379	0.0196	0.0010	0.0219
0.3	0.0135	0.0103	-0.0033	0.0081	0.0055	-0.0026
	0.0415	0.0008	0.0472	0.0258	0.0008	0.0293
0.4	0.0068	0.0085	0.0017	0.0040	0.0033	-0.0008
	0.0475	0.0006	0.0543	0.0296	0.0006	0.0333
0.5	0.0000	0.0005	0.0005	0.0000	-0.0010	-0.0010
	0.0495	0.0004	0.0569	0.0308	0.0004	0.0339
0.6	-0.0068	-0.0069	-0.0001	-0.0040	-0.0029	0.0012
	0.0475	0.0003	0.0545	0.0296	0.0003	0.0343
0.7	-0.0135	-0.0131	0.0004	-0.0081	-0.0087	-0.0006
	0.0415	0.0001	0.0499	0.0258	0.0001	0.0291
0.8	-0.0203	-0.0174	0.0029	-0.0121	-0.0124	-0.0003
	0.0314	0.0001	0.0366	0.0196	0.0001	0.0227
0.9	-0.0271	-0.0252	0.0018	-0.0162	-0.0165	-0.0003
	0.0174	0.0000	0.0224	0.0109	0.0000	0.0132
0.95	-0.0304	-0.0266	0.0039	-0.0182	-0.0179	0.0003
	0.0088	0.0000	0.0146	0.0056	0.0000	0.0079

Notes: This table present the simulation results, when u_i is generated from uniform distribution with the range $R = 4$, x_i is generated form mixture normal distribution, u_i and x_i are i.i.d.. For each level of α, the first row is for bias and the second row is for the MSE of the quantile estimator. For each panel, the first column presents the second-order bias and MSE derived by Corollary 1 and 2.2, the second column presents the Monte Carlo simulation bias and MSE of quantile estimators $\hat{\beta}$, the third column presents the Monte Carlo simulation bias and MSE of the bias corrected quantile estimators $\tilde{\beta}$ where $\tilde{\beta} = \hat{\beta} - B(\hat{\beta})$. We set $\beta = 0$ and $N = 60, 100$, and the results are computed from 10,000

Monte Carlo replications.

Table 2.12: Bias correction and MSE with x_i generated from mixture normal distribution (strongly skewed), DGP 2, $R = 4$

	$R = 4, N = 60$			$R = 4, N = 100$		
α	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}_{formula}$	$\hat{\beta}$	$\tilde{\beta}$
0.05	-0.0154	-0.0142	0.0012	-0.0092	-0.0086	0.0006
	0.0029	0.0014	0.0062	0.0017	0.0014	0.0030
0.1	-0.0137	-0.0109	0.0028	-0.0082	-0.0077	0.0005
	0.0050	0.0013	0.0089	0.0030	0.0013	0.0055
0.2	-0.0103	-0.0124	-0.0021	-0.0062	-0.0078	-0.0016
	0.0086	0.0010	0.0150	0.0052	0.0010	0.0094
0.3	-0.0068	-0.0064	0.0004	-0.0041	-0.0032	0.0009
	0.0111	0.0008	0.0184	0.0068	0.0008	0.0116
0.4	-0.0034	-0.0016	0.0019	-0.0021	-0.0040	-0.0019
	0.0126	0.0006	0.0223	0.0078	0.0006	0.0135
0.5	0.0000	0.0000	0.0000	0.0000	-0.0002	-0.0002
	0.0131	0.0004	0.0226	0.0081	0.0004	0.0139
0.6	0.0034	0.0028	-0.0006	0.0021	0.0032	0.0011
	0.0126	0.0003	0.0220	0.0078	0.0003	0.0131
0.7	0.0068	0.0097	0.0028	0.0041	0.0053	0.0012
	0.0111	0.0001	0.0192	0.0068	0.0001	0.0115
0.8	0.0103	0.0079	-0.0024	0.0062	0.0061	0.0000
	0.0086	0.0001	0.0148	0.0052	0.0001	0.0091
0.9	0.0137	0.0126	-0.0011	0.0082	0.0081	-0.0001
	0.0050	0.0000	0.0088	0.0030	0.0000	0.0051
0.95	0.0154	0.0139	-0.0015	0.0092	0.0076	-0.0016
	0.0029	0.0000	0.0062	0.0017	0.0000	0.0030

Notes: This table present the simulation results, when u_i is generated from uniform distribution with the range $R = 4$, x_i is generated form mixture normal distribution, u_i and x_i are i.i.d.. For each level of α, the first row is for bias and the second row is for the MSE of the quantile estimator. For each panel, the first column presents the second-order bias and MSE derived by Corollary 1 and 2.2, the second column presents the Monte Carlo simulation bias and MSE of quantile estimators $\hat{\beta}$, the third column presents the Monte Carlo simulation bias and MSE of the bias corrected quantile estimators $\tilde{\beta}$ where $\tilde{\beta} = \hat{\beta} - B(\hat{\beta})$. We set $\beta = 0$ and $N = 60, 100$, and the results are computed from 10,000

Monte Carlo replications.

Chapter 3

The Second-order Asymptotic Properties of Asymmetric Least Squares Estimation

3.1 Introduction

The higher-order asymptotic properties permit us to obtain better approximation of the bias of estimators, and allow us to find an approach to improve the behavior of estimators and test statistics. In this paper, we extend the second-order asymptotic results for the symmetric least squares (LS) estimators to asymmetric least squares (ALS) estimators. Newey and Powell (1987) proposed the term, ALS, and investigated the estimation and hypothesis tests for coefficients of linear ALS models. The symmetric LS gives the mean regression function while the ALS gives the "expectile" regression function, a generalization of the usual regression function. The ALS model has been used in many financial applications. However, the literature on the ALS model has been entirely the first-order asymptotic properties. The first-order asymptotic properties of the ALS model can be improved by considering the higher order asymptotic approximations which are better approximations. In this paper, we filled this unexplored area by developing the analytical results of the second-order bias and mean squared error (MSE) for the ALS models. We show that the second-order bias is much larger as the asymmetry is stronger, and therefore the benefit of the second-order bias correction is greater when we are interested in extreme expectiles. The higher order MSE result for the ALS estimation enables us to better understand the sources of estimation uncertainty. The Monte Carlo simulations results present that the second-order bias corrected estimator has better behavior than the uncorrected one.

The paper is organized as follows. In Section 3.2, we review Newey and Powell (1987) to introduce the ALS estimator, and present the moment condition of the ALS regression and the assumptions used in this paper. In Section 3.3, we derive the second-order bias and MSE of the conditional ALS regression estimators. In Section 3.4, a special case of the ALS regression model without a covariate is considered, which gives the unconditional ALS estimator. In Section 3.5, we present Monte Carlo simulations.

In this paper, $f_i(\cdot) \equiv f_i(\cdot|x_i)$ denotes the density of y_i conditional on x_i, and $f_i^{(j)}(\cdot)$ denote the jth order derivative of $f_i(\cdot)$ for $j \geq 1$. The jth-order partial derivative of a matrix $A(\beta)$ is defined as $\nabla_\beta^j A(\beta)$. For a matrix A, $\|A\|$ denotes the usual norm, $[\text{trace}\,(AA')]^{1/2}$. If A is a $k \times 1$ vector, then $||A|| = (A'A)^{1/2}$. The Kronecker product is defined in the usual way. For an $m \times n$ matrix A and a $p \times q$ matrix B, we have $A \otimes B$ as an $mp \times nq$ matrix. The $\overline{X} = E(X)$ denotes the expectation of a random vector X.

3.2 Asymmetric Least Squares Estimation

3.2.1 Loss Functions

Consider a random variable y from distribution $F(\cdot)$. Then the linear regression model is

$$y_i = x_i'\beta + u_i, \tag{3.1}$$

where y_i is a scalar, x_i is a $k \times 1$ vector, and u_i is a scalar, $i = 1, \ldots, N$.

Given $\theta \in (0, 1)$, the quantile regression estimators $\hat{\beta}(\theta)$ proposed by Koenker and Bassett (1987), are obtained by minimizing

$$Q_N(\beta; \theta) = \sum_{i=1}^{N} r_\theta \left(y_i - x_i'\beta \right), \tag{3.2}$$

where $r_\theta(\cdot)$ is the check loss function,

$$r_\theta(\lambda) \equiv |\theta - \mathbf{1}(\lambda < 0)| \cdot |\lambda|. \tag{3.3}$$

Newey and Powell (1987) considered a similar class of estimators. Given $\tau \in (0, 1)$, the asymmetric least squares (ALS) estimators $\hat{\beta}(\tau)$ can be obtained by minimizing

$$R_N(\beta; \tau) = \sum_{i=1}^{N} \rho_\tau \left(y_i - x_i'\beta \right), \tag{3.4}$$

where it replaces the check loss function by the following asymmetric least squares loss function,

$$\rho_\tau(\lambda) \equiv |\tau - \mathbf{1}(\lambda < 0)| \cdot \lambda^2. \tag{3.5}$$

ALS gives weight of τ and $(1-\tau)$ to the squared errors depending upon the sign of errors u_i. A value of $\tau = 0.5$ reproduces ordinary least squares (OLS) estimation. Newey and Powell (1987) showed that ALS estimators can be computed by iterated weighted least squares,

$$\widehat{\beta}(\tau) = \left[\sum_{i=1}^{N} \left|\tau - \mathbf{1}(y_i < x_i'\widehat{\beta}(\tau))\right| x_i x_i'\right]^{-1} \sum_{i=1}^{N} \left|\tau - \mathbf{1}(y_i < x_i'\widehat{\beta}(\tau))\right| x_i y_i. \tag{3.6}$$

We follow Newey and Powell (1987) and refer to $\mu(\tau) = x_i'\beta$ as the τ-conditional expectile of y_i. There is an extensive literature on the relationship and difference of quantile and expectile. In general, an expectile $\mu(\tau)$ is related to a quantile $q(\theta)$. Yao and Tong (1996) showed that for any $\theta \in (0,1)$, there is a relationship that $\mu(\tau(\theta)) = q(\theta)$. Kuan et al. (2009) showed that an expectile with a given τ corresponds to quantiles with different θ under distinct distributios, for example, for a given $\theta < 0.5$, $\tau(\theta)$ is larger for the distribution with thicker tails. The quantile depends only on the probability of tails but not their magnitude. Therefore, quantile is insensitive to the magnitude of extreme tails. Unlike quantile, the expectile is sensitive to magnitude of extreme tails.

Unlike the check loss function $r_\theta(\lambda)$, which is not continuously differentiable, the advantage of ALS regression is that the asymmetric least squares loss function $\rho_\tau(\lambda)$ is differentiable in λ, so that $\rho_\tau(y_i - x_i'\beta)$ is differentiable in β. See Pagan and Ullah (1999, pp. 240-241). Newey and Powell (1987) investigated the moment conditions and asymptotic distribution of the ALS estimators. In this paper, we use an alternative approach with the use of delta (generalized) function to derive moment conditions. Our approach gives the idential results for the moment conditions and their derivatives to those in Newey and Powell (1987). Given the asymmetric least squares loss function, the $k \times 1$ vector expectile estimators $\widehat{\beta}(\tau)$ can be obtained by solving

$$\min_{\beta} E[\rho_\tau(y_i - x_i'\beta)] = E\left[\left|\tau - \mathbf{1}(y_i < x_i'\beta)\right| \cdot \left(y_i - x_i'\beta\right)^2\right]. \tag{3.7}$$

Equation (3.7) reduces to the standard least squares objective function when $\tau = 0.5$.

Newey and Powell (1987) indicated that $\rho_\tau(y_i - x_i'\beta)$ is continuously differentiable in β. Then the population moment condition is

$$\begin{aligned} & \nabla_\beta^1 E\left[\rho_\tau(y_i - x_i'\beta)\right] \\ = \; & E\left[\nabla_\beta^1 \rho_\tau(y_i - x_i'\beta)\right] \\ = \; & E\left[\nabla_\beta^1 \left|\tau - \mathbf{1}(y_i < x_i'\beta)\right| \cdot (y_i - x_i'\beta)^2\right] + 2E[|\tau - \mathbf{1}(y_i < x_i'\beta)|(y_i - x_i'\beta)(-x_i)] \end{aligned} \tag{3.8}$$

By the definition of delta function in Appendix $B.1$, we have $\mathbf{1}(y_i - x_i'\beta < 0) = \mathbf{1}(x_i'\beta - y_i \geq 0) = \phi(x_i'\beta - y_i)$. See Gelfand and Shilov (1964). Then

$$\nabla_\beta^1 1(y_i - x_i'\beta < 0) = \nabla_\beta^1 \phi(x_i'\beta - y_i) = \frac{\mathrm{d}\phi(x_i'\beta - y_i)}{\mathrm{d}(x_i'\beta - y_i)} \frac{\mathrm{d}(x_i'\beta - y_i)}{\mathrm{d}\beta} = x_i'\delta(x_i'\beta - y_i).$$

The first term of the Equation (3.8) can be written as $E[x_i'\delta(x_i'\beta - y_i)(y_i - x_i'\beta)^2]$, which equals zero, because according to the property of Dirac delta function in Appendix $B.3$ and $B.4$, we have

$$\begin{aligned} E[x_i'\delta(x_i'\beta - y_i)(y_i - x_i'\beta)^2] &= E[x_i'\delta(y_i - x_i'\beta)(y_i - x_i'\beta)^2] \\ &= E\left[x_i' E\left[\delta(y_i - x_i'\beta)(y_i - x_i'\beta)^2 | x_i\right]\right] \\ &= E\left[x_i' \int_{-\infty}^{+\infty} \delta(y_i - x_i'\beta)(y_i - x_i'\beta)^2 f_i(y_i)\mathrm{d}y_i\right] \\ &= E\left[x_i'(x_i'\beta - x_i'\beta)^2 f_i(x_i'\beta)\right] \\ &= 0, \end{aligned}$$

where $f_i(x_i'\beta) \equiv f_i(x_i'\beta|x_i)$ is the conditional density of y_i evaluated at $y_i = x_i'\beta$, which equals to the conditional density of the error evaluated at zero, i.e. $f_i(0|x_i)$. Under the assumptions that we will state shortly, the moment condition can be written as

$$\begin{aligned} \nabla_\beta^1 E\left[\rho_\tau(y_i - x_i'\beta)\right] &= 2E[|\tau - \mathbf{1}(y_i < x_i'\beta)|(y_i - x_i'\beta)(-x_i)] \\ &\equiv E\left[s_i(\beta)\right], \end{aligned} \tag{3.9}$$

where $s_i(\beta) \equiv -2|\tau - \mathbf{1}(y_i < x_i'\beta)|(y_i - x_i'\beta)x_i$ is the score function. This is the same as $g_i(\beta)$ in Newey and Powell (1987, p. 844, line 2).

To get rid of the absolute value in (3.9), we first rewrite the score function as

$$\begin{aligned} s_i(\beta) &= 2|\tau - \mathbf{1}(y_i < x_i'\beta)|(y_i - x_i'\beta)(-x_i) \\ &= 2\left(\mathbf{1}(y_i < x_i'\beta) - \tau\right) x_i \left|y_i - x_i'\beta\right|. \end{aligned}$$

Since $\mathbf{1}(y_i \geq x_i'\beta) = 1 - \mathbf{1}(y_i < x_i'\beta)$, we then have

$$\begin{aligned} \left|y_i - x_i'\beta\right| &= \mathbf{1}(y_i \geq x_i'\beta)\left(y_i - x_i'\beta\right) + \mathbf{1}(y_i < x_i'\beta)\left(y_i - x_i'\beta\right) \\ &= \left[1 - \mathbf{1}(y_i < x_i'\beta)\right]\left(y_i - x_i'\beta\right) + \mathbf{1}(y_i < x_i'\beta)\left(y_i - x_i'\beta\right) \\ &= \left(y_i - x_i'\beta\right)\left[1 - 2 \cdot \mathbf{1}(y_i < x_i'\beta)\right]. \end{aligned}$$

Thus, the score function can be rewritten as

$$\begin{aligned} s_i(\beta) &= 2\left(\mathbf{1}(y_i < x_i'\beta) - \tau\right) x_i \left|y_i - x_i'\beta\right| \\ &= 2\left(\mathbf{1}(y_i < x_i'\beta) - \tau\right) x_i \left(y_i - x_i'\beta\right)\left[1 - 2 \cdot \mathbf{1}(y_i < x_i'\beta)\right] \\ &= 2x_i\left(y_i - x_i'\beta\right)\left[(2\tau - 1)\,\mathbf{1}(y_i < x_i'\beta) - \tau\right]. \end{aligned}$$

The sample moment condition for (3.9) is denoted as

$$\Psi_N(\beta) = \frac{1}{N}\sum_{i=1}^{N} s_i(\beta). \tag{3.10}$$

3.2.2 Assumptions

Now we discuss the assumptions under which theorems and corollaries stated below will be true. We argue that these assumptions encompass a wide variety of ALS models, which means that the analytical results are of wide interest and applicability. The first-order asymptotic properties of the ALS model has been investigated by Newey and Powell (1987). To develop the higher-order asymptotic properties of the ALS model, we follow Assumptions A-C in RSU (1996), which are similar to some of the assumptions in Newey and Powell (1987). Assumptions A-C of RSU (1996) is stated as follows.

Assumption A. The jth-order derivative of score function $s_i(\beta)$ exists in a neighborhood of β_0, $i = 1, 2, \ldots$, and $E\left\|\nabla_\beta^j s_i(\beta_0)\right\|^2 < \infty$.

Assumption B. For some neighborhood of $\beta_0, (\nabla\Psi_N(\beta))^{-1} = O_p(1)$.

Assumption C. $||\nabla^j q_i(\beta) - \nabla^j q_i(\beta_0)||||\beta - \beta_0||M_i$ for some neighborhood of β_0, where $E|M_i|C\infty, i = 1, 2, \ldots.$

Assumption A implies that for the ALS mode, the jth-order derivative of $s_i(\beta)$ exists in a neighborhood of β_0, and $E\,||x_i||^4 < \infty$, $E\left[||x_i||^{j+2}\, f_i^{(j-1)}(0|x_i)\right]^2 < \infty$, for $j \geq 1, 2$, where $f_i^{(0)}(0|x_i) = f_i(0|x_i)$ is the conditional density of u_i given x_i evaluated at zero. Assumption A for ALS model requires that the conditional density of y_i given x_i is continuous, and slightly higher than fouth moments of x_i are bounded, which are the same as Assumptions 2 and 3 in Newey and Powell (1987). In the following, we present how we derive the specific exprecession in Assumption A for the ALS model. Note that β is a $k \times 1$ vector, where x_i is a $k \times 1$ vector, $s_i(\beta)$ is a $k \times 1$ vector, $\delta(x_i'\beta - y_i)$ is a scalar.

The derivative of a $k \times 1$ vector $s_i(\beta)$ with respect to a $k \times 1$ vector β is a $k \times k$ matrix $\nabla^1_\beta s_i(\beta)$. Then the first-order derivative of $s_i(\beta)$ exists,

$$\begin{aligned}
\nabla^1_\beta s_i(\beta) &= \nabla^1_\beta \left[2x_i \left(y_i - x_i'\beta\right)\left[(2\tau-1)\,\mathbf{1}(y_i < x_i'\beta) - \tau\right]\right] \\
&= -2x_ix_i'\left[(2\tau-1)\,\mathbf{1}(y_i < x_i'\beta) - \tau\right] + 2\,(2\tau-1)\,x_ix_i'\left(y_i - x_i'\beta\right)\delta(x_i'\beta - y_i) \\
&= -2\,(2\tau-1)\,x_ix_i'\mathbf{1}(y_i < x_i'\beta) + 2\tau x_ix_i' + 2\,(2\tau-1)\,x_ix_i'\left(y_i - x_i'\beta\right)\delta(x_i'\beta - y_i).
\end{aligned}$$

Using the the properties in Appendix $A.2, B.3$ and $B.4$, we obtain

$$\begin{aligned}
E\left\|\nabla^1_\beta s_i(\beta_0)\right\| &= E\left[\left\|x_ix_i'\right\|\left[-2\,(2\tau-1)\,\mathbf{1}(y_i < x_i'\beta) + 2\tau + 2\,(2\tau-1)\left(y_i - x_i'\beta\right)\delta(x_i'\beta - y_i)\right]\right] \\
&= E\left[\left\|x_ix_i'\right\|\left[-2\,(2\tau-1)\,\mathbf{1}(y_i < x_i'\beta) + 2\tau + 0\right]\right] \\
&= E\left[\left\|x_ix_i'\right\| E\left[\left(-2\,(2\tau-1)\,\mathbf{1}(y_i < x_i'\beta) + 2\tau\right)|x_i\right]\right] \\
&= E\left[\left\|x_ix_i'\right\| E\left[(-2\,(2\tau-1) + 2\tau)\,\tau + 2\tau\,(1-\tau)\right]\right] \\
&= 4\tau(1-\tau)E\left\|x_i\right\|^2 \\
&< \infty.
\end{aligned}$$

which is the same results as the derivative $\nabla^2_\beta R\,(\beta;\tau)$ in Newey and Powell (1987, p. 844 equation A.11). The second-order derivative of a $k \times 1$ vector $s_i(\beta)$ with respect to a $k \times 1$ vector β is a $k \times k^2$ matrix $\nabla^2_\beta s_i(\beta)$.

The second order derivative of $s_i(\beta)$ exists,

$$\begin{aligned}
\nabla^2_\beta s_i(\beta) &= \nabla^1_\beta\left[-2\,(2\tau-1)\,x_ix_i'\mathbf{1}(y_i < x_i'\beta) + 2\tau x_ix_i' + 2\,(2\tau-1)\,x_ix_i'\left(y_i - x_i'\beta\right)\delta(x_i'\beta - y_i)\right] \\
&= -2\,(2\tau-1)\left(x_ix_i'\right)\otimes x_i'\delta(x_i'\beta - y_i) + 2\,(2\tau-1)\left(x_ix_i'\right)\otimes\left(-x_i'\right)\delta(x_i'\beta - y_i) \\
&\quad +2\,(2\tau-1)\left(x_ix_i'\right)\otimes x_i'\left(y_i - x_i'\beta\right)\delta^{(1)}(x_i'\beta - y_i) \\
&= -4\,(2\tau-1)\left(x_ix_i'\right)\otimes x_i'\delta(x_i'\beta - y_i) + 2\,(2\tau-1)\left(x_ix_i'\right)\otimes x_i'\left(y_i - x_i'\beta\right)\delta^{(1)}(x_i'\beta - y_i),
\end{aligned}$$

where the derivative of a scalar $\delta(x_i'\beta - y_i)$ with respect to a $k \times 1$ vector β is a $1 \times k$ row vector $\nabla_\beta^1 \delta(x_i'\beta - y_i)$. We denote

$$\nabla_\beta^1 \delta(x_i'\beta - y_i) = \frac{\mathrm{d}\delta(x_i'\beta - y_i)}{\mathrm{d}(x_i'\beta - y_i)} \frac{\mathrm{d}(x_i'\beta - y_i)}{\mathrm{d}\beta} = x_i'\delta^{(1)}(x_i'\beta - y_i),$$

where $\delta^{(1)}(x_i'\beta - y_i)$ is a scalar. Using the the properties in Appendix $A.3, B.5$ and $B.6$, we obtain

$$\begin{aligned}
E\left\|\nabla_\beta^2 s_i(\beta_0)\right\|^2 &= (2\tau - 1)\, E\left\|\left(x_i x_i'\right) \otimes x_i' \left[2\left(y_i - x_i'\beta\right)\delta^{(1)}(x_i'\beta - y_i) - 4\delta(x_i'\beta - y_i)\right]\right\| \\
&= (2\tau - 1)\, E\left\|\left(x_i x_i'\right) \otimes x_i' \left[2E\left[\left(y_i - x_i'\beta\right)\delta^{(1)}(x_i'\beta - y_i)|x_i\right] - 4E\left[\delta(x_i'\beta - y_i)|x_i\right]\right]\right\| \\
&= (2\tau - 1)\, E||\left(x_i x_i'\right) \otimes x_i'[-2\int \delta^{(1)}(y_i - x_i'\beta)\left(y_i - x_i'\beta\right) f_i(y_i)\mathrm{d}y_i \\
&\quad -4\int \delta(y_i - x_i'\beta) f_i(y_i)\mathrm{d}y_i]|| \\
&= (2\tau - 1)\, E||\left(x_i x_i'\right) \otimes x_i'[2\int \delta(y_i - x_i'\beta)\left(f_i(y_i) + (y_i - x_i'\beta) f_i^{(1)}(y_i)\right)\mathrm{d}y_i \\
&\quad -4 f_i(x_i'\beta)]|| \\
&= (2\tau - 1)\, E\left\|\left(x_i x_i'\right) \otimes x_i' \left[2\int \delta(y_i - x_i'\beta) f_i(y_i)\mathrm{d}y_i + 0 - 4 f_i(x_i'\beta)\right]\right\| \\
&= (2\tau - 1)\, E\left\|\left(x_i x_i'\right) \otimes x_i' \left[2 f_i(x_i'\beta) + 0 - 4 f_i(x_i'\beta)\right]\right\| \\
&= -2\,(2\tau - 1)\, E\left[f_i(x_i'\beta)\left\|x_i\right\|^3\right] \\
&< \infty.
\end{aligned}$$

The third-order derivative of a $k \times 1$ vector $s_i(\beta)$ with respect to a $k \times 1$ vector β

is a $k \times k^3$ matrix $\nabla^3_\beta s_i(\beta)$. The third order derivative of $s_i(\beta)$ exists,

$$\begin{aligned}
\nabla^3_\beta s_t(\gamma) &= \nabla^1_\beta \left[-4\left(2\tau-1\right)\left(x_i x_i'\right)\otimes x_i'\delta(x_i'\beta-y_i)+2\left(2\tau-1\right)\left(x_i x_i'\right)\otimes x_i'\left(y_i-x_i'\beta\right)\delta^{(1)}(x_i'\beta-y_i)\right] \\
&= -4\left(2\tau-1\right)\left(x_i x_i'\right)\otimes x_i'\otimes x_i'\delta^{(1)}(x_i'\beta-y_i)+2\left(2\tau-1\right)\left(x_i x_i'\right)\otimes\left(-x_i'\right)\otimes x_i'\delta^{(1)}(x_i'\beta-y_i) \\
&\quad +2\left(2\tau-1\right)\left(x_i x_i'\right)\otimes x_i'\otimes x_i'\left(y_i-x_i'\beta\right)\delta^{(2)}(x_i'\beta-y_i) \\
&= \left(x_i x_i'\right)\otimes x_i'\otimes x_i'\left[-6\left(2\tau-1\right)\delta^{(1)}(x_i'\beta-y_i)+2\left(2\tau-1\right)\left(y_i-x_i'\beta\right)\delta^{(2)}(x_i'\beta-y_i)\right],
\end{aligned}$$

where the derivative of a $1\times k$ row vector $\nabla^1_\beta\delta(x_i'\beta-y_i)$ with respect to a $k\times 1$ vector β is a $1\times k^2$ row vector $\nabla^2_\beta\delta(x_i'\beta-y_i)$. We denote

$$\nabla^2_\beta\delta(x_i'\beta-y_i)=\nabla^1_\beta x_i'\delta^{(1)}(x_i'\beta-y_i)=x_i'\otimes\frac{\mathrm{d}\delta^{(1)}(x_i'\beta-y_i)}{\mathrm{d}(x_i'\beta-y_i)}\frac{\mathrm{d}(x_i'\beta-y_i)}{\mathrm{d}\beta}=x_i'\otimes x_i'\delta^{(2)}(x_i'\beta-y_i),$$

where $\delta^{(2)}(x_i'\beta-y_i)$ is a scalar. Using the the properties in Appendix $A.4, B.6$ and $B.7$, we obtain

$$\begin{aligned}
E\left\|\nabla^3_\beta s_i(\beta_0)\right\| &= \left(2\tau-1\right)E\left\|\left(x_i x_i'\right)\otimes x_i'\otimes x_i'\left[-6\delta^{(1)}(x_i'\beta-y_i)+2\delta^{(2)}(x_i'\beta-y_i)(y_i-x_i'\beta)\right]\right\| \\
&= \left(2\tau-1\right)E||\left(x_i x_i'\right)\otimes x_i'\otimes x_i'[-6E\left[\delta^{(1)}(x_i'\beta-y_i)|x_i\right] \\
&\quad +2E\left[\delta^{(2)}(x_i'\beta-y_i)(y_i-x_i'\beta)|x_i\right]]|| \\
&= \left(2\tau-1\right)E||\left(x_i x_i'\right)\otimes x_i'\otimes x_i'[6\int\delta^{(1)}(y_i-x_i'\beta)f_i(y_i)\mathrm{d}y_i \\
&\quad +2\int\delta^{(2)}(x_i'\beta-y_i)(y_i-x_i'\beta)f_i(y_i)\mathrm{d}y_i]|| \\
&= \left(2\tau-1\right)E||\left(x_i x_i'\right)\otimes x_i'\otimes x_i'[-6\int\delta(y_i-x_i'\beta){f_i}^{(1)}(y_i)\mathrm{d}y_i \\
&\quad +2\int\delta(y_i-x_i'\beta)\left[2{f_i}^{(1)}(y_i)+(y_i-x_i'\beta)f_i^{(2)}(y_i)\right]\mathrm{d}y_i]|| \\
&= \left(2\tau-1\right)E\left\|\left(x_i x_i'\right)\otimes x_i'\otimes x_i'\left[-6f_i^{(1)}(x_i'\beta)+4f_i^{(1)}(x_i'\beta)+0\right]\right\| \\
&= -2\left(2\tau-1\right)E\left[{f_i}^{(1)}(x_i'\beta)\left\|x_i\right\|^4\right] \\
&< \infty
\end{aligned}$$

Next, we discuss Assumption B. For ALS models, Assumption B requires $\operatorname*{p\lim}_{N\to\infty}\left(\nabla_{\beta}^{1}\Psi_{N}(\beta)\right)^{-1}=\left(\lim_{N\to\infty}E\nabla_{\beta}^{1}\Psi_{N}(\beta)\right)^{-1}=\left(\lim_{N\to\infty}4\tau(1-\tau)E\left(x_{i}x_{i}'\right)\right)^{-1}=O(1)$, that implies $E\left(x_{i}x_{i}'\right)$ is non-singular, which is the same as Assumption 4 of Newey and Powell (1987).

3.3 Second-order Bias and MSE of the ALS Estimators

The assumptions in RSU (1996) are necessary to obtain the stochastic expansion of $\widehat{\beta}$, based on which we derive the second-order bias of the ALS esitmator. For the bias results in Theorems 1 and 3 we allow that x_i and u_i are not identically distributed but independent across $i=1,...,N$. For independent and identically distributed (i.i.d.) x_i and u_i, the second-order bias and MSE can be further simplified since most of the cross-terms in the matrix multiplications drop out, which will be stated in corresponding Corollaries 1 and 3.

3.3.1 Bias

Theorem 1. *Under Assumptions A-C, the second-orer bias of the ALS estimators* $\widehat{\beta}\left(\tau\right)$ *up to* $O(N^{-1})$ *is*

$$\begin{aligned}
B\left(\widehat{\beta}\left(\tau\right)\right) &= \frac{1}{N^2}\sum_{i=1}^{N}4Q\left\{\left(2\tau-1\right)E\left[x_ix_i'Qx_iu_i\mathbf{1}(u_i<0)\right]-\tau^2E\left[x_ix_i'Qx_iu_i\right]\right\}\\
&\quad+\frac{1}{N}\sum_{i=1}^{N}\left(2\tau-1\right)QE\left[\left(x_ix_i'\right)\otimes x_i'f_i(0|x_i)\right]\\
&\quad\times\frac{1}{N^2}\sum_{i=1}^{N}4\left(Q\otimes Q\right)\left\{\begin{array}{c}-\left(2\tau-1\right)E\left[\left(x_i\otimes x_i\right)u_i^2\mathbf{1}(u_i<0)\right]\\ +\tau^2E\left[\left(x_i\otimes x_i\right)u_i^2\right]\end{array}\right\}, \qquad (3.11)
\end{aligned}$$

where $Q=\left(4\tau(1-\tau)\frac{1}{N}\sum_{i=1}^{N}E\left[x_ix_i'\right]\right)^{-1}$.

Proof: Suppose x_i and u_i are not identically distributed, but independent across $i = 1, ..., N$. Suppose y_i has conditional density function $f_i(y|x)$. To simplify the notation, we use $f_i(y)$ to denote $f_i(y|x)$. As in Bao and Ullah (2007), the second-order bias of the ALS estimators $\widehat{\beta}(\tau)$ up to $O(N^{-1})$ is

$$B(\widehat{\beta}) = Q\left[\overline{Vd} - \frac{1}{2}\overline{H_2}\left(\overline{d \otimes d}\right)\right].$$

We have

$$\Psi_N(\beta) = \frac{1}{N}\sum_{i=1}^{N} s_i(\beta),$$

$$s_i(\beta) = 2x_i\left(y_i - x_i'\beta\right)\left[(2\tau - 1)\,\mathbf{1}(y_i < x_i'\beta) - \tau\right],$$

$$\begin{aligned} H_1 &= \nabla_\beta^1 \Psi_N = \nabla_\beta^1 \frac{1}{N}\sum_{i=1}^{N} s_i = \frac{1}{N}\sum_{i=1}^{N} \nabla_\beta^1 s_i \\ &= \frac{1}{N}\sum_{i=1}^{N}\left[-2(2\tau-1)x_i x_i'\mathbf{1}(y_i < x_i'\beta) + 2\tau x_i x_i' + 2(2\tau-1)x_i x_i'\left(y_i - x_i'\beta\right)\delta(x_i'\beta - y_i)\right], \end{aligned}$$

$$\begin{aligned} H_2 &= \nabla_\beta^2 \Psi_N = \nabla_\beta^2 \frac{1}{N}\sum_{i=1}^{N} s_i = \frac{1}{N}\sum_{i=1}^{N} \nabla_\beta^2 s_i \\ &= \frac{1}{N}\sum_{i=1}^{N}\left[(2\tau-1)\left(x_i x_i'\right)\otimes x_i'\left[-4\delta(x_i'\beta - y_i) + 2\left(y_i - x_i'\beta\right)\delta^{(1)}(x_i'\beta - y_i)\right]\right], \end{aligned}$$

$$\begin{aligned} H_3 &= \nabla_\beta^3 \Psi_N = \nabla_\beta^3 \frac{1}{N}\sum_{i=1}^{N} s_i = \frac{1}{N}\sum_{i=1}^{N} \nabla_\beta^3 s_i \\ &= \frac{1}{N}\sum_{i=1}^{N}\left[(2\tau-1)\left(x_i x_i'\right)\otimes x_i' \otimes x_i'\left[-6\delta^{(1)}(x_i'\beta - y_i) + 2\left(y_i - x_i'\beta\right)\delta^{(2)}(x_i'\beta - y_i)\right]\right], \end{aligned}$$

$$\overline{H_1} = E\nabla_\beta^1 \Psi_N = 4\tau(1-\tau)\frac{1}{N}\sum_{i=1}^{N} E\left(x_i x_i'\right),$$

$$\overline{H_2} = E\nabla_\beta^2 \Psi_N = -2(2\tau-1)\frac{1}{N}\sum_{i=1}^{N} E\left[\left(x_i x_i'\right)\otimes x_i' f_i(x_i'\beta)\right],$$

$$\overline{H_3} = E\nabla^3_\beta \Psi_N = -2(2\tau - 1)\frac{1}{N}\sum_{i=1}^{N} E\left[(x_i x_i') \otimes x_i' \otimes x_i') f_i{}^{(1)}(x_i'\beta)\right],$$

$$Q = \left(\overline{H_1}\right)^{-1} = \left(4\tau(1-\tau)\frac{1}{N}\sum_{i=1}^{N} E\left[x_i x_i'\right]\right)^{-1},$$

$$V = H_1 - \overline{H_1},$$

$$W = H_2 - \overline{H_2},$$

and

$$d = Q\Psi_N,$$

where Ψ_N, s_i and d are all $k \times 1$ vectors. H_1, $\overline{H_1}$, Q, and V are all $k \times k$ matrixes, H_2, $\overline{H_2}$ and W are all $k \times k^2$ matrixes. H_3 and $\overline{H_3}$ are $k \times k^3$ matrixes. Using the the properties in Appendix B.8,

$$\begin{aligned}
\overline{Vd} &= E\left[\left(H_1 - \overline{H_1}\right) Q\Psi_N\right] \\
&= E\left(H_1 Q \Psi_N\right) - E\left(\Psi_N\right) \\
&= E\left[\frac{1}{N}\sum_{i=1}^{N}\left[-2\left(2\tau-1\right) x_i x_i' \mathbf{1}(y_i < x_i'\beta) + 2\tau x_i x_i' + 2\left(2\tau - 1\right) x_i x_i' \left(y_i - x_i'\beta\right) \delta(x_i'\beta - y_i)\right] Q\Psi_N\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N} E[-4(2\tau-1)^2 x_i x_i' Q x_i (y_i - x_i'\beta)\mathbf{1}(y_i < x_i'\beta) \\
&\quad +4\tau\left(2\tau-1\right) x_i x_i' Q x_i (y_i - x_i'\beta)\mathbf{1}\left(y_i < x_i'\beta\right) \\
&\quad +4\tau\left(2\tau-1\right) x_i x_i' Q x_i (y_i - x_i'\beta)\mathbf{1}\left(y_i < x_i'\beta\right) \\
&\quad -4\tau^2 x_i x_i' Q x_i (y_i - x_i'\beta) \\
&\quad +4\tau\left(2\tau-1\right)^2 x_i x_i' Q x_i (y_i - x_i'\beta)^2 \delta(x_i'\beta - y_i)\mathbf{1}\left(y_i < x_i'\beta\right) \\
&\quad -4\tau\left(2\tau-1\right) x_i x_i' Q x_i (y_i - x_i'\beta)^2 \delta\left(y_i - x_i'\beta\right)] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\left\{4\left(2\tau-1\right) E\left[x_i x_i' Q x_i (y_i - x_i'\beta)\mathbf{1}(y_i < x_i'\beta)\right] - 4\tau^2 E\left[x_i x_i' Q x_i (y_i - x_i'\beta)\right]\right\},
\end{aligned}$$

and

$$\begin{aligned}
\overline{d \otimes d} &= E\left[(Q\Psi_N \otimes Q\Psi_N)\right] = E\left[(Q \otimes Q)(\Psi_N \otimes \Psi_N)\right] \\
&= (Q \otimes Q)\, E\left[(\Psi_N \otimes \Psi_N)\right] = \frac{1}{N^2}\sum_{i=1}^{N}(Q \otimes Q)\, E\,(s_i \otimes s_i) \\
&= \frac{1}{N^2}\sum_{i=1}^{N}(Q \otimes Q)\, E\left[4\,(x_i \otimes x_i)\,(y_i - x_i'\beta)^2\left[(2\tau-1)\,\mathbf{1}(y_i < x_i'\beta) - \tau\right]^2\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N} 4\,(Q \otimes Q)\left\{\begin{array}{c} \tau^2 E\left[(x_i \otimes x_i)\,(y_i - x_i'\gamma)^2\right] \\ -\,(2\tau-1)\,E\left[(x_i \otimes x_i)\,(y_i - x_i'\beta)^2\mathbf{1}(y_i < x_i'\beta)\right] \end{array}\right\}.
\end{aligned}$$

Therefore, the second-order bias of $\widehat{\beta}$ up to $O(N^{-1})$ can be rewriten as

$$\begin{aligned}
B\left(\widehat{\beta}\,(\tau)\right) &= Q\left[\overline{Vd} - \frac{1}{2}\overline{H_2}\left(\overline{d \otimes d}\right)\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N} 4Q\left\{(2\tau-1)\,E\left[x_i x_i' Q x_i (y_i - x_i'\beta)\mathbf{1}(y_i < x_i'\beta)\right] - 4\tau^2 Q E\left[x_i x_i' Q x_i (y_i - x_i'\beta)\right]\right\} \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}(2\tau-1)\,QE\left[\left(x_i x_i'\right) \otimes x_i' f_i(x_i'\beta)\right] \\
&\quad \times \frac{1}{N^2}\sum_{i=1}^{N} 4\,(Q \otimes Q)\left\{\begin{array}{c} -\,(2\tau-1)\,E\left[(x_i \otimes x_i)\,(y_i - x_i'\beta)^2\mathbf{1}(y_i < x_i'\beta)\right] \\ +\tau^2 E\left[(x_i \otimes x_i)\,(y_i - x_i'\beta)^2\right] \end{array}\right\},
\end{aligned}$$

where $Q = \left(4\tau(1-\tau)\frac{1}{N}\sum_{i=1}^{N} E\left[x_i x_i'\right]\right)^{-1}$. Since the conditional density of y_i given x_i evaluated at $y_i = x_i'\beta$ is the same as the conditional density of u_i given x_i evaluated at $u_i = 0$. We use $f_i\,(0|x_i)$ to denote the conditional density of u_i given x_i evaluated at $u_i = 0$, which completes the proof of Theorem 1.

Corollary 1. *Under Assumptions A-C, when x_i and u_i are i.i.d., the second-order bias of*

$\widehat{\beta}(\tau)$ *up to* $O(N^{-1})$ *is*

$$
\begin{aligned}
B\left(\widehat{\beta}(\tau)\right) &= \frac{1}{N}4Q\left\{(2\tau-1)E\left[x_i x_i' Q x_i u_i \mathbf{1}(u_i<0)\right] - \tau^2 E\left[x_i x_i' Q x_i u_i\right]\right\} \\
&\quad + \frac{1}{N}4(2\tau-1)QE\left[\left(x_i x_i'\right)\otimes x_i' f(0)\right](Q\otimes Q)\left\{\begin{array}{c} -(2\tau-1)E\left[(x_i\otimes x_i)u_i^2\mathbf{1}(u_i<0)\right] \\ +\tau^2 E\left[(x_i\otimes x_i)u_i^2\right]\end{array}\right\},
\end{aligned}
$$

where $Q=(4\tau(1-\tau)E\left[x_i x_i'\right])^{-1}$.

Since $f_i(0|x_i)$ denotes the conditional density of u_i evaluated at the $u_i=0$. When x_i and u_i are i.i.d, these $f_i(\cdot)$s are identical, and we use $f(\cdot)$ to denote the conditional density of u_i. When x_i and u_i are i.i.d., the conditional density of u_i is the same as unconditional density, $f(0|x_i)=f(0)$.

Remark 1.1. When x_i and u_i are i.i.d., and $k=1$, we observe that x_i, Ψ_N, s_i, d, H_1, $\overline{H_1}$, Q, V, H_2, $\overline{H_2}$, W, H_3, $\overline{H_3}$ are all scalars, and the second-order bias of $\widehat{\beta}(\tau)$ up to $O(N^{-1})$ can be rewritten as

$$
\begin{aligned}
B\left(\widehat{\beta}(\tau)\right) &= \frac{1}{N}\left\{\frac{(2\tau-1)E\left[x_i^3 u_i\mathbf{1}(u_i<0)\right]}{4\tau^2(1-\tau)^2\left[E\left(x_i^2\right)\right]^2} - \frac{E\left[x_i^3 u_i\right]}{4(1-\tau)^2\left[E\left(x_i^2\right)\right]^2}\right\} \\
&\quad + \frac{1}{N}\left\{\frac{(2\tau-1)E\left(x_i^3\right)f(0)E\left[x_i^2 u_i^2\right]}{16\tau(1-\tau)^3\left[E\left(x_i^2\right)\right]^3} - \frac{(2\tau-1)^2E\left(x_i^3\right)f(0)E\left[x_i^2u_i^2\mathbf{1}(u_i<0)\right]}{16\tau^3(1-\tau)^3\left[E\left(x_i^2\right)\right]^3}\right\}.
\end{aligned}
$$

Remark 1.2. The second-order bias of $\widehat{\beta}(\tau)$ is larger at the extreme expectiles a distribution, because at the extreme expectiles Q is larger, and the second term in (3.11) dominant the other terms. The second-order bias of $\widehat{\beta}(\tau)$ goes to zero as the sample size goes to infinity.

Remark 1.3. The objective function of ALS model reduces to the standard least-squares objective function when $\tau=0.5$. In this case, the second-order bias of $\widehat{\beta}(\tau)$ up to $O(N^{-1})$

equals the second-order bias of OLS estimator. The OLS estimator is unbiased because $E\left(u_i|x_i\right)=0$.

Now, we derive the MSE of the ALS estimator of order up to $O\left(N^{-2}\right)$ in Theorem 2. For simplicity, we make an additional assumption that x_i and u_i are not only identially distributed but also independent and $k=1$. The MSE result when x_i and u_i are independent but not identically distributed as we did for the bias result in Theorem 1 can be easily obtained using the same method but not presented here for simplicity.

3.3.2 MSE

Theorem 2. *Under Assumptions A-C, in the ASL regression model, suppose* x_i and u_i *are i.i.d. and k=1, the MSE of the ALS estimator* $\widehat{\beta}(\tau)$ *up to* $O(N^{-2})$ *is*

$$\begin{aligned} M\left(\widehat{\beta}(\tau)\right) &= \frac{1}{N}4Q^2C_1-\frac{1}{N^2}16Q^3C_3+\frac{1}{N^2}8Q^4C_1-\frac{1}{N^2}Q^4 16(2\tau-1)E\left[x_i^3 f(0)\right]C_4 \\ &\quad +\frac{1}{N^2}96Q^4C_2^2+\frac{1}{N^2}48Q^4\left\{\tau(1-\tau)E\left(x_i^4\right)-4\tau^2(1-\tau)^2\left[E\left(x_i^2\right)\right]^2\right\}C_1 \\ &\quad +\frac{1}{N^2}384Q^5(2\tau-1)E\left[x_i^3 f(0)\right]C_1C_2+\frac{1}{N^2}240Q^6(2\tau-1)^2\left[E\left[x_i^3 f(0)\right]\right]^2 C_1^2 \\ &\quad +\frac{1}{N^2}32Q^5(2\tau-1)E\left[x_i^4 f^{(1)}(0)\right]C_1^2, \end{aligned} \tag{3.12}$$

where

$$\begin{aligned} Q &= \left(4\tau(1-\tau)E\left[x_i^2\right]\right)^{-1}, \\ C_1 &= E\left[-(2\tau-1)x_i^2u_i^2\mathbf{1}(u_i<0)\right]+E\left(\tau^2x_i^2u_i^2\right), \\ C_2 &= E\left[(2\tau-1)x_i^3u_i\mathbf{1}(u_i<0)\right]-E\left(\tau^2x_i^3u_i\right), \\ C_3 &= E\left[-(2\tau-1)\left(\tau^2-\tau+1\right)x_i^4u_i^2\mathbf{1}(u_i<0)\right]+E\left(\tau^3x_i^4u_i^2\right), \\ C_4 &= E\left[(2\tau-1)\left(\tau^2-\tau+1\right)x_i^3u_i^3\mathbf{1}(u_i<0)\right]-E\left(\tau^3x_i^3u_i^3\right), \end{aligned}$$

and $f(0)$ is the density of u_i evaluated at $u_i = 0$, $f^{(1)}(0)$ is the first derivative of the density of u_i evaluated at $u_i = 0$.

Proof: Following RSU (1996), the MSE of the ALS estimator $\widehat{\beta}(\tau)$ up to $O\left(N^{-2}\right)$ is

$$M(\widehat{\beta}) = \frac{1}{N}A_1 + \frac{1}{N^2}\left(A_2 + A_2'\right) + \frac{1}{N^2}\left(A_3 + A_4 + A_4'\right)$$

where $\frac{1}{N}A_1 = E(a_{-1/2}a'_{-1/2})$, $\frac{1}{N^2}\left(A_2 + A_2'\right) = E(a_{-1}a'_{-1/2}+a_{-1/2}a'_{-1})$, $\frac{1}{N^2}\left(A_3 + A_4 + A_4'\right) = E(a_{-1}a'_{-1} + a_{-3/2}a'_{-1/2} + a_{-1/2}a'_{-3/2})$.

Suppose x_i and u_i are i.i.d.. For ALS estimators when $k = 1$, $i = 1, \ldots, N$, and $j = 1, \ldots, N$, we have

$$A_1 = \overline{d_i^2},$$

$$A_2 = -Q\overline{V_i d_i^2} + \frac{1}{2}Q\overline{H_2}\,\overline{d_i^3},$$

$$A_3 = 2Q^2\overline{V_i V_j d_i d_j} + Q^2\overline{V_i^2}\,\overline{d_i^2} + \frac{3}{4}Q^2\overline{H_2}^2\overline{d_i^2 d_j^2} - 3Q^2\overline{H_2}\,\overline{V_i d_i d_j^2},$$

$$A_4 = Q^2\overline{V_i^2}\,\overline{d_i^2} + 2Q^2\overline{V_i V_j d_i d_j} - \frac{9}{2}Q^2\overline{H_2}\,\overline{V_i d_i d_j^2} + \frac{3}{2}Q\overline{W_i d_i d_j^2} + \frac{3}{2}Q^2\overline{H_2}^2\overline{d_i^2 d_j^2} - \frac{1}{2}Q\overline{H_3}\,\overline{d_i^2 d_j^2},$$

where $H_1 = \nabla^1_\beta s_i$, $\overline{H_1} = \overline{\nabla^1_\beta s_i}$, $H_2 = \nabla^2_\beta s_i$, $\overline{H_2} = \overline{\nabla^2_\beta s_i}$, $H_3 = \nabla^3_\beta s_i$, $\overline{H_3} = \overline{\nabla^3_\beta s_i}$, $Q = \left(\overline{H_1}\right)^{-1}$, and

$$d = Q\Psi_N = \frac{1}{N}\sum_{i=1}^{N} d_i = \frac{1}{N}\sum_{i=1}^{N} Q s_i,$$

$$V = \nabla^1_\beta\Psi_N - \overline{\nabla^1_\beta\Psi_N} = \frac{1}{N}\sum_{i=1}^{N} V_i = \frac{1}{N}\sum_{i=1}^{N}\left(\nabla^1_\beta s_i - \overline{\nabla^1_\beta s_i}\right),$$

$$W = \nabla^2_\beta\Psi_N - \overline{\nabla^2_\beta\Psi_N} = \frac{1}{N}\sum_{i=1}^{N} W_i = \frac{1}{N}\sum_{i=1}^{N}\left(\nabla^2_\beta s_i - \overline{\nabla^2_\beta s_i}\right).$$

If x_i and u_i are i.i.d., then s_i, d_i, V_i,and W_i are all i.i.d.. Since $\overline{V_i V_j d_i d_j} = \overline{V_i d_i}^2$, $\overline{V_i d_i d_j^2} = \overline{V_i d_i}\,\overline{d_i^2}$, and $\overline{d_i^2 d_j^2} = \overline{d_i^2}^2$, then A_3 and A_4 can be simplified as

$$A_3 = 2Q^2\overline{V_i d_i}^2 + Q^2\overline{V_i^2}\,\overline{d_i^2} + \frac{3}{4}Q^2\overline{H_2}^2\overline{d_i^2}^2 - 3Q^2\overline{H_2}\;\overline{V_i d_i}\,\overline{d_i^2},$$

$$A_4 = Q^2\overline{V_i^2}\,\overline{d_i^2} + 2Q^2\overline{V_i d_i}^2 - \frac{9}{2}Q^2\overline{H_2}\ \overline{V_i d_i}\overline{d_i^2} + \frac{3}{2}Q\overline{W_i d_i}\overline{d_i^2} + \frac{3}{2}Q^2\overline{H_2}^2\overline{d_i^2}^2 - \frac{1}{2}Q\overline{H_3}\overline{d_i^2}^2.$$

Then the MSE up to $O(N^{-2})$ can be written as

$$\begin{aligned} M(\widehat{\beta}) &= \frac{1}{N}\overline{d_i^2} - \frac{1}{N^2}2Q\left[\overline{V_i d_i^2} - \frac{1}{2}\overline{H_2}\overline{d_i^3}\right] + \frac{1}{N^2}6Q^2\overline{V_i d_i}^2 + \frac{1}{N^2}3Q^2\overline{V_i^2}\,\overline{d_i^2} \\ &\quad + \frac{1}{N^2}3Q\overline{W_i d_i}\overline{d_i^2} - \frac{1}{N^2}12Q^2\overline{H_2}\ \overline{V_i d_i}\overline{d_i^2} + \frac{1}{N^2}\frac{15}{4}Q^2\overline{H_2}^2\overline{d_i^2}^2 - \frac{1}{N^2}Q\overline{H_3}\overline{d_i^2}^2, \end{aligned}$$

where we have

$$\overline{V_i d_i^2} = E\left[\left(H_1 - \overline{H_1}\right)Q^2 s_i^2\right] = E\left(H_1 Q^2 s_i^2\right) - QE\left(s_i^2\right),$$

$$\begin{aligned} E\left(H_1 Q^2 s_i^2\right) &= E\left[\left(\left[-2\left(2\tau-1\right)x_i^2\mathbf{1}(y_i < x_i'\beta) + 2\tau x_i^2 + 2\left(2\tau-1\right)x_i^2\left(y_i - x_i'\beta\right)\delta(x_i'\beta - y_i)\right]\right)Q^2 s_i^2\right] \\ &= Q^2 E\left\{\begin{array}{c} 8\left(2\tau-1\right)^2 x_i^4 u_i^2\mathbf{1}(u_i<0) - 8\tau^2\left(2\tau-1\right)x_i^4 u_i^2\mathbf{1}(u_i<0) \\ -8\tau\left(2\tau-1\right)x_i^4 u_i^2\mathbf{1}(u_i<0) + 8\tau^2 x_i^4 u_i^2 \\ -8\left(2\tau-1\right)^2 x_i^4 u_i^3\delta(x_i'\beta - y_i)\mathbf{1}(u_i<0) \\ +8\tau^2\left(2\tau-1\right)x_i^4 u_i^3\delta(x_i'\beta - y_i) \end{array}\right\} \\ &= Q^2 E\left[8\left(2\tau-1\right)\left(-\tau^2+\tau-1\right)x_i^4 u_i^2\mathbf{1}(u_i<0)\right] + Q^2 E\left(8\tau^3 x_i^4 u_i^2\right). \end{aligned}$$

We also observe

$$\begin{aligned} \overline{d_i^2} &= Q^2 E(s_i^2) \\ &= Q^2 E\left[4x_i^2(y_i - x_i'\beta)^2\left[\left(2\tau-1\right)\mathbf{1}(y_i < x_i'\beta) - \tau\right]^2\right] \\ &= Q^2 E\left[-4\left(2\tau-1\right)x_i^2 u_i^2\mathbf{1}(u_i<0)\right] + Q^2 E\left(4\tau^2 x_i^2 u_i^2\right), \end{aligned}$$

$$\begin{aligned} \overline{d_i^3} &= Q^3 E(s_i^2) \\ &= Q^3 E\left[8x_i^3(y_i - x_i'\beta)^2\left[\left(2\tau-1\right)\mathbf{1}(y_i < x_i'\beta) - \tau\right]^3\right] \\ &= Q^3 E\left[8\left(2\tau-1\right)\left(\tau^2-\tau+1\right)x_i^3 u_i^3\mathbf{1}(u_i<0)\right] - Q^3 E\left(8\tau^3 x_i^3 u_i^3\right), \end{aligned}$$

$$\overline{V_i d_i}^2 = 16Q^2 \left\{ (2\tau - 1) E\left[x_i^3 u_i \mathbf{1}(u_i < 0)\right] - \tau^2 E\left[x_i^3 u_i\right] \right\}^2,$$

$$\begin{aligned}
\overline{V_i^2} &= E\left[\left(H_1 - \overline{H_1}\right)^2\right] \\
&= E\left[H_1^2 - 2H_1\overline{H_1} + \overline{H_1}^2\right] \\
&= E\left[H_1^2\right] - 2\overline{H_1}^2 + \overline{H_1}^2 \\
&= E\left[H_1^2\right] - \overline{H_1}^2 \\
&= E\left[\left[-2(2\tau-1)x_i^2\mathbf{1}(y_i < x_i'\beta) + 2\tau x_i^2 + 2(2\tau-1)x_i^2\left(y_i - x_i'\beta\right)\delta(x_i'\beta - y_i)\right]^2\right] \\
&\quad - \left[4\tau(1-\tau)E\left(x_i^2\right)\right]^2 \\
&= E\left\{\begin{array}{c} 4(2\tau-1)^2 x_i^4 \mathbf{1}(y_i < x_i'\beta) + 4\tau^2 x_i^4 - 8\tau(2\tau-1)x_i^4\mathbf{1}(y_i < x_i'\beta) \\ +4(2\tau-1)^2 x_i^4 (y_i - x_i'\beta)^2 \left(\delta(x_i'\beta - y_i)\right)^2 \\ -8(2\tau-1)^2 x_i^4 (y_i - x_i'\beta)\mathbf{1}(y_i < x_i'\beta)\delta(x_i'\beta - y_i) \\ +8\tau(2\tau-1)x_i^4(y_i - x_i'\beta)\delta(x_i'\beta - y_i) \end{array}\right\} \\
&\quad -16\tau^2(1-\tau)^2\left[E\left(x_i^2\right)\right]^2 \\
&= 4\tau(1-\tau)E\left(x_i^4\right) + E\left[4(2\tau-1)^2 x_i^4 \int \left(y_i - x_i'\beta\right)^2 \left(\delta(x_i'\beta - y_i)\right)^2 f(y_i)\mathrm{d}y_i\right] \\
&\quad -16\tau^2(1-\tau)^2\left[E\left(x_i^2\right)\right]^2,
\end{aligned}$$

$$
\begin{aligned}
\overline{W_i d_i} &= E\left[\left(H_2 - \overline{H_2}\right) Q s_i\right] \\
&= E\left(H_2 Q s_i\right) - Q\overline{H_2} E\left(s_i\right) \\
&= E\left\{\left[\left(2\tau - 1\right) x_i^3 \left[-4\delta(x_i'\beta - y_i) + 2\left(y_i - x_i'\beta\right)\delta^{(1)}(x_i'\beta - y_i)\right]\right] Q s_i\right\} \\
&= 2\left(2\tau - 1\right) QE \left\{ \begin{array}{c} -4\left(2\tau - 1\right) x_i^4 \left(y_i - x_i'\beta\right)\delta(x_i'\beta - y_i)\mathbf{1}(y_i < x_i'\beta) \\ +4\tau x_i^4 \left(y_i - x_i'\beta\right)\delta(x_i'\beta - y_i) \\ +2\left(2\tau - 1\right) x_i^4 \left(y_i - x_i'\beta\right)^2 \delta^{(1)}(x_i'\beta - y_i)\mathbf{1}(y_i < x_i'\beta) \\ -2\tau x_i^4 \left(y_i - x_i'\beta\right)^2 \delta^{(1)}(x_i'\beta - y_i) \end{array} \right\} \\
&= 2\left(2\tau - 1\right) QE \left\{ \begin{array}{c} \left(2\tau - 1\right) x_i^4 \left(y_i - x_i'\beta\right)^2 \delta^{(1)}(x_i'\beta - y_i) - 2\left(2\tau - 1\right) x_i^4 \left(y_i - x_i'\beta\right)^2 \left(\delta(x_i'\beta - y_i)\right)^2 \\ -2\tau x_i^4 \left(y_i - x_i'\beta\right)^2 \delta^{(1)}(x_i'\beta - y_i) \end{array} \right\} \\
&= 2\left(2\tau - 1\right) QE \left\{-x_i^4 \left(y_i - x_i'\beta\right)^2 \delta^{(1)}(x_i'\beta - y_i) - 2\left(2\tau - 1\right) x_i^4 \left(y_i - x_i'\beta\right)^2 \left(\delta(x_i'\beta - y_i)\right)^2\right\} \\
&= 2\left(2\tau - 1\right) QE \left\{ \begin{array}{c} x_i^4 \int \left(y_i - x_i'\beta\right)^2 \delta^{(1)}(y_i - x_i'\beta) f(y_i)\mathrm{d}y_i \\ -2\left(2\tau - 1\right) x_i^4 \int \left(y_i - x_i'\beta\right)^2 \left(\delta(x_i'\beta - y_i)\right)^2 f(y_i)\mathrm{d}y_i \end{array} \right\} \\
&= 2\left(2\tau - 1\right) QE \left\{ \begin{array}{c} x_i^4 \int \left(y_i - x_i'\beta\right)^2 \delta(y_i - x_i'\beta)\left[-2x_i\left(y_i - x_i'\beta\right) f(y_i) + \left(y_i - x_i'\beta\right)^2 f^{(1)}(y_i)\right]\mathrm{d}y_i \\ -2\left(2\tau - 1\right) x_i^4 \int \left(y_i - x_i'\beta\right)^2 \left(\delta(x_i'\beta - y_i)\right)^2 f(y_i)\mathrm{d}y_i \end{array} \right\} \\
&= -4\left(2\tau - 1\right)^2 QE\left[x_i^4 \int \left(y_i - x_i'\beta\right)^2 \left(\delta(x_i'\beta - y_i)\right)^2 f(y_i)\mathrm{d}y_i\right].
\end{aligned}
$$

Since the conditional density of y_i given x_i evaluated at $y_i = x_i'\beta$ is the same as the conditional density of u_i given x_i evaluated at $u_i = 0$. Then the MSE up to $O(N^{-2})$ can be written as

$$
\begin{aligned}
M\left(\widehat{\beta}(\tau)\right) &= \frac{1}{N}\overline{d_i^2} - \frac{1}{N^2}2Q\left[\overline{V_i d_i^2} - \frac{1}{2}\overline{H_2}\,\overline{d_i^3}\right] + \frac{1}{N^2}6Q^2\overline{V_i d_i}^2 + \frac{1}{N^2}3Q^2\overline{V_i^2}\,\overline{d_i^2} \\
&\quad + \frac{1}{N^2}3Q\overline{W_i d_i}\,\overline{d_i^2} - \frac{1}{N^2}12Q^2\overline{H_2}\;\overline{V_i d_i}\,\overline{d_i^2} + \frac{1}{N^2}\frac{15}{4}Q^2\overline{H_2}^2\,\overline{d_i^2}^2 - \frac{1}{N^2}Q\overline{H_3}\,\overline{d_i^2}^2 \\
&= \frac{1}{N}Q^2E\left[-4\left(2\tau-1\right)x_i^2u_i^2\mathbf{1}(u_i<0)\right] + \frac{1}{N}Q^2E\left(4\tau^2x_i^2u_i^2\right) \\
&\quad -\frac{1}{N^2}2Q\left\{\begin{array}{c} Q^2E\left[8\left(2\tau-1\right)\left(-\tau^2+\tau-1\right)x_i^4u_i^2\mathbf{1}(u_i<0)\right] + Q^2E\left(8\tau^3x_i^4u_i^2\right) \\ -QE\left[-4\left(2\tau-1\right)x_i^2u_i^2\mathbf{1}(u_i<0)\right] - QE\left(4\tau^2x_i^2u_i^2\right)\end{array}\right\} \\
&\quad -\frac{1}{N^2}2Q(2\tau-1)E\left[x_i^3f(0)\right]\left\{Q^3E\left[8\left(2\tau-1\right)\left(\tau^2-\tau+1\right)x_i^3u_i^3\mathbf{1}(u_i<0)\right] - Q^3E\left(8\tau^3x_i^3u_i^3\right.\right. \\
&\quad +\frac{1}{N^2}6Q^2 16Q^2\left\{\left(2\tau-1\right)E\left[x_i^3u_i\mathbf{1}(u_i<0)\right] - \tau^2E\left[x_i^3u_i\right]\right\}^2 \\
&\quad +\frac{1}{N^2}3Q^2\left\{\begin{array}{c} 4\tau\left(1-\tau\right)E\left(x_i^4\right) + E\left[4\left(2\tau-1\right)^2x_i^4\int\left(y_i-x_i'\beta\right)^2\left(\delta(x_i'\beta-y_i)\right)^2f(y_i)\mathrm{d}y_i\right] \\ -16\tau^2(1-\tau)^2\left[E\left(x_i^2\right)\right]^2\end{array}\right\} \\
&\quad \times\left[Q^2E\left[-4\left(2\tau-1\right)x_i^2u_i^2\mathbf{1}(u_i<0)\right] + Q^2E\left(4\tau^2x_i^2u_i^2\right)\right] \\
&\quad -\frac{1}{N^2}12Q\left(2\tau-1\right)^2QE\left[x_i^4\int\left(y_i-x_i'\beta\right)^2\left(\delta(x_i'\beta-y_i)\right)^2f(y_i)\mathrm{d}y_i\right] \\
&\quad \times\left[Q^2E\left[-4\left(2\tau-1\right)x_i^2u_i^2\mathbf{1}(u_i<0)\right] + Q^2E\left(4\tau^2x_i^2u_i^2\right)\right] \\
&\quad +\frac{1}{N^2}12Q^2 2(2\tau-1)E\left[x_i^3f(0)\right]\left\{4\left(2\tau-1\right)QE\left[x_i^3u_i\mathbf{1}(u_i<0)\right] - 4\tau^2QE\left[x_i^3u_i\right]\right\} \\
&\quad \times\left[Q^2E\left[-4\left(2\tau-1\right)x_i^2u_i^2\mathbf{1}(u_i<0)\right] + Q^2E\left(4\tau^2x_i^2u_i^2\right)\right] \\
&\quad +\frac{1}{N^2}\frac{15}{4}Q^2 4(2\tau-1)^2\left[E\left[x_i^3f(0)\right]\right]^2\left[Q^2E\left[-4\left(2\tau-1\right)x_i^2u_i^2\mathbf{1}(u_i<0)\right] + Q^2E\left(4\tau^2x_i^2u_i^2\right)\right]^2 \\
&\quad +\frac{1}{N^2}Q2(2\tau-1)E\left[x_i^4f^{(1)}(0)\right]\left[Q^2E\left[-4\left(2\tau-1\right)x_i^2u_i^2\mathbf{1}(u_i<0)\right] + Q^2E\left(4\tau^2x_i^2u_i^2\right)\right]^2.
\end{aligned}
$$

This is as stated in Theorem 2.

Corollary 2. The asymptotic variance of $\widehat{\beta}(\tau)$ of Newey and Powell (1987) is the N times the first-order term of $M\left(\widehat{\beta}(\tau)\right)$ in (**??**) in Theorem 2.

Proof: By the MSE expression in (**??**) of Theorem 2, the first-order term of $M\left(\widehat{\beta}(\tau)\right)$ is $\overline{d_i^2}$. Then we have

$$N \times \overline{d_i^2} = 4Q^2 E\left[-\left(2\tau-1\right) x_i^2 u_i^2 \mathbf{1}(u_i<0)\right] + Q^2 E\left(\tau^2 x_i^2 u_i^2\right) + O\left(N^{-1}\right).$$

Newey and Powell (1987) derived the first-order asymptotic distribution of the ALS estimator as follows

$$\sqrt{N}(\widehat{\beta}-\beta_0) \xrightarrow{d} N(0, W^{-1}VW^{-1}),$$

where $w_i(\tau) = \left|\tau - \mathbf{1}(u_i<0)\right|$, $W = E\left[w_i(\tau) x_i^2\right]$, $V = E\left[w_i^2 x_i^2 u_i^2\right]$.

We want to show that the asymptotic variance of $\widehat{\beta}(\tau)$ equals N times the MSE of $\widehat{\beta}(\tau)$ up to $O(N^{-1})$. We have

$$\begin{aligned} W &= E\left[w_i(\tau) x_i^2\right] \\ &= E\left[x_i^2 E\left(w_i(\tau) | x_i\right)\right] \\ &= E\left[x_i^2 E\left(\left|\tau - \mathbf{1}(u_i<0)\right| | x_i\right)\right] \\ &= E\left[x_i^2\left[(1-\tau)\tau + \tau(1-\tau)\right]\right] \\ &= 2\tau(1-\tau) E\left(x_i^2\right) \\ &= \frac{1}{2}Q^{-1}, \end{aligned}$$

and

$$\begin{aligned} V &= E\left[w_i^2 x_i^2 u_i^2\right] \\ &= E\left[\left|\tau - \mathbf{1}(u_i<0)\right| x_i^2 u_i^2\right] \\ &= E\left[-\left(2\tau-1\right) x_i^2 u_i^2 \mathbf{1}(u_i<0)\right] + E\left(\tau^2 x_i^2 u_i^2\right). \end{aligned}$$

Therefore, the asymptotic variance $W^{-1}VW^{-1} = N \times \overline{d_i^2}$.

3.4 Special Case: Unconditional ALS Model

In this section, we consider a special case of the ALS regression model with $x_i = 1$, i.e., the ALS model without any covariate, which gives the unconditional ALS estimator. Consider a random variable y from distribution $F(\cdot)$. Then the unconditional ALS model is

$$y_i = \beta + u_i, \tag{3.13}$$

where y_i is a scalar and u_i is a scalar, $i = 1, \ldots, N$. Given $\tau \in (0, 1)$, ALS estimators $\hat{\beta}(\tau)$ can be obtained by minimizing

$$R_N(\beta; \tau) = \sum_{i=1}^{N} \rho_\tau (y_i - \beta),$$

where the asymmetric least squares loss function is

$$\rho_\tau(\lambda) \equiv |\tau - \mathbf{1}(\lambda < 0)| \cdot \lambda^2.$$

For this simpler case, we now present the bias result in Theorem 3 and the MSE result in Theorem 4.

3.4.1 Bias

Theorem 3. *Under Assumptions A-C, suppose that u_i is independent but not identically distributed, the second-order bias of the unconditional ALS estimator $\widehat{\beta}(\tau)$ up to $O(N^{-1})$ is*

$$\begin{aligned} B\left(\widehat{\beta}(\tau)\right) &= \frac{1}{N^2}\sum_{i=1}^{N} 4Q^2 \left\{(2\tau - 1) E[u_i \mathbf{1}(u_i < 0)] - \tau^2 E(u_i)\right\} \\ &+ \frac{1}{N}\sum_{i=1}^{N} 4(2\tau - 1) Q^3 f_i(0) \frac{1}{N^2}\sum_{i=1}^{N} \left\{-(2\tau - 1) E\left[u_i^2 \mathbf{1}(u_i < 0)\right] + \tau^2 E\left(u_i^2\right)\right\}, \end{aligned} \tag{3.14}$$

where $Q = [4\tau(1-\tau)]^{-1}$, $f(0)$ *is the density of* u_i *evaluated at* $u_i = 0$.

Proof: Consider the linear ALS regression model $y_i = \beta + u_i$, where y_i is a scalar, u_i is the error defined to be the difference between y_i and its τ-expectile β, we call $\widehat{\beta}(\tau)$ as the unconditional ALS estimator. Given the asymmetric least squares loss function, the ALS estimators $\widehat{\beta}(\tau)$ can be obtained by solving

$$\min_{\beta} E[\rho_\tau(y_i - \beta)] = E\left[|\tau - \mathbf{1}(y_i < \beta)| \cdot (y_i - \beta)^2\right].$$

Then the population moment condition is

$$\begin{aligned} \nabla^1_\beta E\left[\rho_\tau(y_i - \beta)\right] &= E\left[\nabla^1_\beta \rho_\tau(y_i - \beta)\right] \\ &= E\left[\nabla^1_\beta |\tau - \mathbf{1}(y_i < \beta)| \cdot (y_i - \beta)^2\right] - 2E[|\tau - \mathbf{1}(y_i < \beta)|(y_i - \beta)]. \end{aligned}$$

By the definition of Dirac delta function in Appendix $B.1$, we have $1(y_i - \beta < 0) = 1(\beta - y_i \geq 0) = \phi(\beta - y_i)$. Then

$$\nabla^1_\beta 1(y_i - \beta < 0) = \delta(\beta - y_i).$$

According to the property of Dirac delta function in Appendix $B.4$, we have $\delta(\beta - y_i) = \delta(y_i - \beta)$. According to the property of Dirac delta function in Appendix $B.3$, we have

$$\begin{aligned} E[\delta(\beta - y_i)(y_i - \beta)] &= E[\delta(y_i - \beta)(y_i - \beta)] \\ &= \int_{-\infty}^{+\infty} \delta(y_i - \beta)(y_i - \beta) f(y_i) \mathrm{d}y_i \\ &= (\beta - \beta) f(\beta) \\ &= 0. \end{aligned}$$

Thus, the moment condition can be written as

$$\nabla^1_\beta E[\rho_\tau(y_i - \beta)] = -2E[|\tau - \mathbf{1}(y_i < \beta)|(y_i - \beta)] \equiv E[s_i(\beta)],$$

where $s_i(\beta)$ is the score function. To get rid of the absolute value, first, we can rewrite the score function as

$$\begin{aligned} s_i(\beta) &= -2|\tau - \mathbf{1}(y_i < \beta)|(y_i - \beta) \\ &= 2\left(\mathbf{1}(y_i < \beta) - \tau\right)|y_i - \beta|. \end{aligned}$$

Since $\mathbf{1}(y_i \geq x_i'\beta) = 1 - \mathbf{1}(y_i < x_i'\beta)$, we have

$$\begin{aligned} |y_i - \beta| &= \mathbf{1}(y_i \geq \beta)(y_i - \beta) + \mathbf{1}(y_i < \beta)(y_i - \beta) \\ &= [1 - \mathbf{1}(y_i < \beta)](y_i - \beta) + \mathbf{1}(y_i < \beta)(y_i - \beta) \\ &= (y_i - \beta)[1 - 2 \cdot \mathbf{1}(y_i < \beta)]. \end{aligned}$$

Then, the score function can be written as

$$\begin{aligned} s_i(\beta) &= 2\left(\mathbf{1}(y_i < \beta) - \tau\right)|y_i - \beta| \\ &= 2\left(\mathbf{1}(y_i < \beta) - \tau\right)(y_i - \beta)[1 - 2 \cdot \mathbf{1}(y_i < \beta)] \\ &= 2(y_i - \beta)[(2\tau - 1)\mathbf{1}(y_i < \beta) - \tau]. \end{aligned}$$

Therefore, the sample moment condition can be written as

$$\Psi_N(\beta) = \frac{1}{N}\sum_{i=1}^{N} s_i(\beta), \tag{3.15}$$

where $s_i(\beta) = 2(y_i - \beta)[(2\tau - 1)\mathbf{1}(y_i < \beta) - \tau]$.

The second-order bias up to $O(N^{-1})$ is

$$B(\widehat{\beta}) = Q\left[\overline{V}d - \frac{1}{2}\overline{H_2}\left(\overline{d \otimes d}\right)\right],$$

where

$$
\begin{aligned}
H_1 &= \nabla^1_\beta \Psi_N \\
&= \frac{1}{N}\sum_{i=1}^{N} \nabla^1_\beta \left[2\left(y_i-\beta\right)\left[\left(2\tau-1\right)\mathbf{1}(y_i<\beta)-\tau\right]\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} \left[-2\left(2\tau-1\right)\mathbf{1}(y_i<\beta)+2\tau+2\left(2\tau-1\right)\left(y_i-\beta\right)\delta(\beta-y_i)\right],
\end{aligned}
$$

$$
\begin{aligned}
H_2 &= \nabla^2_\beta \Psi_N \\
&= \frac{1}{N}\sum_{i=1}^{N} \nabla^1_\beta \left[-2\left(2\tau-1\right)\mathbf{1}(y_i<\beta)+2\tau+2\left(2\tau-1\right)\left(y_i-\beta\right)\delta(\beta-y_i)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} \left[-4\left(2\tau-1\right)\delta(\beta-y_i)+2\left(2\tau-1\right)\left(y_i-\beta\right)\delta^{(1)}(\beta-y_i)\right],
\end{aligned}
$$

$$
\begin{aligned}
H_3 &= \nabla^3_\beta \Psi_N \\
&= \frac{1}{N}\sum_{i=1}^{N} \nabla^3_\beta \left[-4\left(2\tau-1\right)\delta(\beta-y_i)+2\left(2\tau-1\right)\left(y_i-\beta\right)\delta^{(1)}(\beta-y_i)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} \left[-6\left(2\tau-1\right)\delta^{(1)}(\beta-y_i)+2\left(2\tau-1\right)\left(y_i-\beta\right)\delta^{(2)}(\beta-y_i)\right],
\end{aligned}
$$

$$
\begin{aligned}
\overline{H_1} &= E\nabla^1_\beta \Psi_N \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left[-2\left(2\tau-1\right)\mathbf{1}(y_i<\beta)+2\tau+2\left(2\tau-1\right)\left(y_i-\beta\right)\delta(\beta-y_i\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left[-2\left(2\tau-1\right)\mathbf{1}(y_i<\beta)+2\tau+0\right] \\
&= \left(-2\left(2\tau-1\right)+2\tau\right)\tau+2\tau\left(1-\tau\right) \\
&= 4\tau\left(1-\tau\right),
\end{aligned}
$$

$$\begin{aligned}
\overline{H_2} &= E\nabla_\beta^2 \Psi_N \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left[-4\,(2\tau-1)\,\delta(\beta-y_i)+2\,(2\tau-1)\,(y_i-\beta)\,\delta^{(1)}(\beta-y_i)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[-4\,(2\tau-1)\int\delta(y_i-\beta)f_i(y_i)\mathrm{d}y_i-2\,(2\tau-1)\int\delta^{(1)}(y_i-\beta)\,(y_i-\beta)\,f_i(y_i)\mathrm{d}y_i\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[-4\,(2\tau-1)\,f_i(\beta)+2\,(2\tau-1)\int\delta(y_i-\beta)\left(f_i(y_i)+(y_i-\beta)f_i^{(1)}(y_i)\right)\mathrm{d}y_i\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[-4\,(2\tau-1)\,f_i(\beta)+2\,(2\tau-1)\,f_i(\beta)+0\right] \\
&= -2\,(2\tau-1)\,\frac{1}{N}\sum_{i=1}^{N} f_i(\beta),
\end{aligned}$$

$$\begin{aligned}
\overline{H_3} &= E\nabla_\beta^3 \Psi_N \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left[-6\,(2\tau-1)\,\delta^{(1)}(\beta-y_i)+2\,(2\tau-1)\,(y_i-\beta)\,\delta^{(2)}(\beta-y_i)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[\begin{array}{c} 6\,(2\tau-1)\int\delta^{(1)}(y_i-\beta)f_i(y_i)\mathrm{d}y_i \\ +2\,(2\tau-1)\int\delta^{(2)}(y_i-\beta)\,(y_i-\beta)\,f_i(y_i)\mathrm{d}y_i \end{array}\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[\begin{array}{c} -6\,(2\tau-1)\int\delta(y_i-\beta)f_i^{(1)}(y_i)\mathrm{d}y_i \\ +2\,(2\tau-1)\int\delta(y_i-\beta)\left(2f_i^{(1)}(y_i)+(y_i-\beta)f_i^{(2)}(y_i)\right)\mathrm{d}y_i \end{array}\right] \\
&= \frac{1}{N}\left[\sum_{i=1}^{N}-6\,(2\tau-1)\,f_i^{(1)}(\beta)+2\,(2\tau-1)\left[2f_i^{(1)}(\beta)+0\right]\right] \\
&= -2\,(2\tau-1)\,\frac{1}{N}\sum_{i=1}^{N} f_i^{(1)}(\beta),
\end{aligned}$$

$$Q=\left(\overline{H_1}\right)^{-1}=\left[4\tau\,(1-\tau)\right]^{-1},$$

$$V=H_1-\overline{H_1},$$

$$W=H_2-\overline{H_2},$$

and

$$d = Q\Psi_N.$$

$f_i(\beta)$ is the density of y_i evaluated at $y_i = \beta$. $f_i^{(1)}(\beta)$ and $f_i^{(2)}(\beta)$ are the first and second derivative of the density of y_i evaluated at $y_i = \beta$, respectively. Since Ψ_N, s_i, d, H_1, $\overline{H_1}$, Q, V, H_2, $\overline{H_2}$, W, H_3, $\overline{H_3}$ are all scalars, then

$$\begin{aligned}
\overline{Vd} &= E\left[\left(H_1 - \overline{H_1}\right) Q\Psi_N\right] \\
&= E\left(H_1 Q\Psi_N\right) - E\left(\Psi_N\right) \\
&= E\left[\frac{1}{N}\sum_{i=1}^{N}\left[-2\left(2\tau-1\right)\mathbf{1}(y_i<\beta) + 2\tau + 2\left(2\tau-1\right)\left(y_i-\beta\right)\delta(\beta-y_i)\right] Q\Psi_N\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N} E[-4(2\tau-1)^2 Q(y_i - x_i'\beta)\mathbf{1}(y_i<\beta) + 4\tau\left(2\tau-1\right)Q(y_i-\beta)\mathbf{1}\left(y_i<\beta\right) \\
&\quad +4\tau\left(2\tau-1\right)Q(y_i-\beta)\mathbf{1}\left(y_i<\beta\right) - 4\tau^2 Q(y_i-\beta) \\
&\quad +4\tau\left(2\tau-1\right)^2 Q(y_i-\beta)^2\delta(\beta-y_i)\mathbf{1}\left(y_i<\beta\right) - 4\tau\left(2\tau-1\right)Q(y_i-\beta)^2\delta\left(y_i-\beta\right)] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\left\{4\left(2\tau-1\right)QE\left[(y_i-\beta)\mathbf{1}(y_i<\beta)\right] - 4\tau^2 QE(y_i-\beta)\right\},
\end{aligned}$$

$$\begin{aligned}
\overline{d\otimes d} &= \frac{1}{N^2}\sum_{i=1}^{N} Q^2 E\left[s_i^2\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N} Q^2 E\left[4(y_i-\beta)^2\left[\left(2\tau-1\right)\mathbf{1}(y_i<\beta) - \tau\right]^2\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N} 4Q^2\left\{\tau^2 E\left[(y_i-\beta)^2\right] - \left(2\tau-1\right)E\left[(y_i-\beta)^2\mathbf{1}(y_i<\beta)\right]\right\}.
\end{aligned}$$

Therefore, the second-order bias of $\widehat{\beta}$ up to $O(N^{-1})$,of the unconditional ALS estimators $\widehat{\beta}$

can be written as

$$
\begin{aligned}
B\left(\widehat{\beta}\left(\tau\right)\right) &= Q\left[\overline{Vd}-\frac{1}{2}\overline{H_2}\left(\overline{d\otimes d}\right)\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}4Q^2\left\{\left(2\tau-1\right)E\left[u_i\mathbf{1}(u_i<0)\right]-\tau^2E(u_i)\right\} \\
&\quad +\frac{1}{N}\sum_{i=1}^{N}\left(2\tau-1\right)Qf_i(\beta)\frac{1}{N^2}\sum_{i=1}^{N}4Q^2\left\{-\left(2\tau-1\right)E\left[u_i^2\mathbf{1}(u_i<0)\right]+\tau^2E\left(u_i^2\right)\right\},
\end{aligned}
$$

where $Q=\left[4\tau\left(1-\tau\right)\right]^{-1}$. Since the unconditional density of y_i evaluated at $y_i=\beta$ is the same as the unconditional density of u_i evaluated at $u_i=0$. We use $f_i\left(0\right)$ to denote the unconditional density of u_i evaluated at $u_i=0$, which completes the proof of Theorem 3.

Corollary 3. *Under Assumptions A-C, when u_i are i.i.d., the second-order bias of the unconditional ALS estimator $\widehat{\beta}\left(\tau\right)$ up to $O(N^{-1})$ is*

$$
\begin{aligned}
B\left(\widehat{\beta}\left(\tau\right)\right) &= \frac{1}{N}\left\{\frac{\left(2\tau-1\right)E\left[u_i\mathbf{1}(u_i<0)\right]}{4\tau^2(1-\tau)^2}-\frac{E\left(u_i\right)}{4(1-\tau)^2}\right\} \\
&\quad +\frac{1}{N}\left\{\frac{\left(2\tau-1\right)f(0)E\left(u_i^2\right)}{16\tau(1-\tau)^3}-\frac{\left(2\tau-1\right)^2f(0)E\left[u_i^2\mathbf{1}(u_i<0)\right]}{16\tau^3(1-\tau)^3}\right\}.
\end{aligned}
$$

Since $f_i(0)$ denotes the unconditional density of u_i evaluated at the $u_i=0$. When u_i are i.i.d, these $f_i(\cdot)$s are identical, and we use $f(\cdot)$ to denote the unconditional density of u_i.

Remark 3.1. The second-order bias of $\widehat{\beta}\left(\tau\right)$ is larger at extreme expectiles of a distribution, because at the extreme expectiles Q is larger, and the second term in (3.14) dominant the other terms. The second-order bias of $\widehat{\beta}\left(\tau\right)$ goes to zero as the sample size goes to infinity.

Remark 3.2. The objective function of ALS model reduces to the standard least-squares objective function when $\tau=0.5$. In this case, the second-order bias of $\widehat{\beta}\left(\tau\right)$ up to $O(N^{-1})$

equals the second-order bias of OLS estimator. The OLS estimator is unbiased because $E(u_i) = 0$.

3.4.2 MSE

Theorem 4. *Under Assumptions A-C, suppose u_i is i.i.d., the MSE of the unconditional ALS estimators $\widehat{\beta}(\tau)$ up to $O(N^{-2})$ is*

$$\begin{aligned} M\left(\widehat{\beta}(\tau)\right) &= \frac{1}{N}4Q^2C_1 - \frac{1}{N^2}16Q^3C_3 + \frac{1}{N^2}8Q^2C_1 - \frac{1}{N^2}Q^4 16(2\tau-1)E\left[f_i(0)\right]C_4 \\ &+\frac{1}{N^2}96Q^4C_2^2 + \frac{1}{N^2}48Q^4\left\{\tau(1-\tau)\left(4\tau^2-4\tau+1\right)\right\}C_1 \\ &+\frac{1}{N^2}384Q^5(2\tau-1)E\left[f(0)\right]C_1C_2 \\ &+\frac{1}{N^2}240Q^6(2\tau-1)^2\left[E\left[f(0)\right]\right]^2C_1^2 + \frac{1}{N^2}32Q^5(2\tau-1)E\left[f^{(1)}(0)\right]C_1^2. \end{aligned}$$

where

$$\begin{aligned} Q &= \left[4\tau(1-\tau)\right]^{-1}, \\ C_1 &= E\left[-(2\tau-1)u_i^2\mathbf{1}(u_i<0)\right] + E\left(\tau^2u_i^2\right), \\ C_2 &= E\left[(2\tau-1)u_i\mathbf{1}(u_i<0)\right] - E\left(\tau^2u_i\right), \\ C_3 &= E\left[-(2\tau-1)\left(\tau^2-\tau+1\right)u_i^2\mathbf{1}(u_i<0)\right] + E\left(\tau^3u_i^2\right), \\ C_4 &= E\left[(2\tau-1)\left(\tau^2-\tau+1\right)u_i^3\mathbf{1}(u_i<0)\right] - E\left(\tau^3u_i^3\right), \end{aligned}$$

and $f(0)$ is the density of u_i evaluated at $u_i = 0$, $f^{(1)}(0)$ is the first derivative of the density of u_i evaluated at $u_i = 0$.

Proof: By Theorem 2, when $x_i = 1$, the MSE of the unconditional ALS estimator $\widehat{\beta}(\tau)$ up to $O\left(N^{-2}\right)$ is

$$
\begin{aligned}
M\left(\widehat{\beta}\left(\tau\right)\right) \quad = \quad & \frac{1}{N}4Q^2\left\{E\left[-\left(2\tau-1\right)u_i^2\mathbf{1}(u_i<0)\right]+E\left(\tau^2u_i^2\right)\right\} \\
& -\frac{1}{N^2}16Q^3\left\{E\left[-\left(2\tau-1\right)\left(\tau^2-\tau+1\right)u_i^2\mathbf{1}(u_i<0)\right]+E\left(\tau^3u_i^2\right)\right\} \\
& +\frac{1}{N^2}8Q^2\left\{E\left[-\left(2\tau-1\right)u_i^2\mathbf{1}(u_i<0)\right]+E\left(\tau^2u_i^2\right)\right\} \\
& -\frac{1}{N^2}Q^416(2\tau-1)E\left[f_i(0)\right]\left\{E\left[\left(2\tau-1\right)\left(\tau^2-\tau+1\right)u_i^3\mathbf{1}(u_i<0)\right]-E\left(\tau^3u_i^3\right)\right\} \\
& +\frac{1}{N^2}96Q^4\left\{E\left[\left(2\tau-1\right)u_i\mathbf{1}(u_i<0)\right]-E\left[\tau^2u_i\right]\right\}^2 \\
& +\frac{1}{N^2}48Q^4\left\{\tau\left(1-\tau\right)\left(4\tau^2-4\tau+1\right)\right\}\left\{E\left[-\left(2\tau-1\right)u_i^2\mathbf{1}(u_i<0)\right]+E\left(\tau^2u_i^2\right)\right\} \\
& -\frac{1}{N^2}48Q^4\left(2\tau-1\right)^2\left\{E\left[-\left(2\tau-1\right)u_i^2\mathbf{1}(u_i<0)\right]+E\left(\tau^2u_i^2\right)\right\} \\
& +\frac{1}{N^2}384Q^5(2\tau-1)E\left[f(0)\right]\left\{E\left[\left(2\tau-1\right)u_i\mathbf{1}(u_i<0)\right]-E\left[\tau^2u_i\right]\right\} \\
& \times\left\{E\left[-\left(2\tau-1\right)u_i^2\mathbf{1}(u_i<0)\right]+E\left(\tau^2u_i^2\right)\right\} \\
& +\frac{1}{N^2}240Q^6(2\tau-1)^2\left[E\left[f(0)\right]\right]^2\left\{E\left[-\left(2\tau-1\right)u_i^2\mathbf{1}(u_i<0)\right]+E\left(\tau^2u_i^2\right)\right\}^2 \\
& +\frac{1}{N^2}32Q^5(2\tau-1)E\left[f^{(1)}(0)\right]\left\{E\left[-\left(2\tau-1\right)u_i^2\mathbf{1}(u_i<0)\right]+E\left(\tau^2u_i^2\right)\right\}^2.
\end{aligned}
$$

This is as stated in Theorem 4.

3.5 Monte Carlo Simulation

Now we give some numerical calculations to present the second-order bias results by Sections 3 and 4. The goal of the data generating process (DGP) is to let the error term u_i, in the ALS regression model $y_i = x_i'\beta + u_i$, satisfies that the τ-conditional expectile of u_i given x_i is zero. Newey and Powell (1987, p. 823) and Kuan, Yeh, and Hsu (2009) showed that the first order condition of minimizing $R_N(\beta;\tau)$ is

$$\tau\int_{\mu(\tau)}^{\infty}(y-\mu(\tau))\,dF(y)+(\tau-1)\int_{-\infty}^{\mu(\tau)}(\mu(\tau)-y)\,dF(y)=0, \tag{3.16}$$

so that the expectile $\mu(\tau)=x_i'\beta(\tau)$ satisfies

$$\frac{\tau}{1-\tau}=\frac{\int_{-\infty}^{\mu(\tau)}(\mu(\tau)-y)\,\mathrm{d}F(y)}{\int_{\mu(\tau)}^{\infty}(y-\mu(\tau))\,\mathrm{d}F(y)}. \tag{3.17}$$

If we set the true β to be zero, then y_i have the same distribution as u_i. To generate u_i from uniform distribution on $[a,b]$ and $\mu(\tau)=0$, we have $\int_{-\infty}^{\mu(\tau)}(\mu(\tau)-y)\mathrm{d}F(y)=\int_{-\infty}^{0}(-y)\mathrm{d}F(y)=\int_{a}^{0}(-y)\frac{1}{b-a}\mathrm{d}y=\frac{a^2}{2(b-a)}$, and $\int_{\mu(\tau)}^{\infty}(y-\mu(\tau))\mathrm{d}F(y)=\int_{0}^{\infty}y\mathrm{d}F(y)=\int_{0}^{b}y\frac{1}{b-a}\mathrm{d}y=\frac{b^2}{2(b-a)}$. Then we can get the relationship between a and b, i.e. $a=-\sqrt{\frac{\tau}{1-\tau}}b$. In the DGP, we generate the error term u_i from uniform distribution on $[a,b]$, where $a=-\frac{\sqrt{\frac{\tau}{1-\tau}}R}{1+\sqrt{\frac{\tau}{1-\tau}}}$, $b=\frac{R}{1+\sqrt{\frac{\tau}{1-\tau}}}$, and the range $R=b-a$. For example, $R=4,\tau=0.1$, implies that u_i is generated from $U[-1,3]$, the mean of u_i is 1, variance of u_i is $\frac{4}{3}$, and $\mu(0.1)=0$. The DGP of u_i guarantees that the 0.1 conditional expectile of u_i given x_i is zero. In addition, we can verify the relationship of quantile and expectile, that is if u_i follows $U[-1,3]$, then $\mu(0.1)=q(0.25)=0$. We simulate x_i from exponential distribution, $f(x_i)=\exp(-x_i)$. Then, y_i is simulated from $y_i=x_i'\beta+u_i$. In this setup, $k=1$, $\beta=0$, $R=4$, $N\in\{100,300\}$.

Following Newey and Powell (1987) and Kuan et al. (2009), we use the iterated weighted least squares algorithm to compute the ALS estimator, in equation (3.6). We use the OLS estimates as the initial value of $\widehat{\beta}$ for the terated weighted least squares estimates and iterate until the estimates converge. The convergence was quick and did not depend on the choice of initial value of $\widehat{\beta}$. We repeat the Monte Carlo simulations 10,000 times and take the average.

Table 1 presents the simulation results when x_i is generated form exponential distribution. Table 2 presents the simulation results when $x_i = 1$. For each τ, the first row is for bias and the second row is for the mean squared error of the ALS estimator. For each panel, the first column presents the Monte Carlo (MC) simulation bias and MSE of ALS estimators $\widehat{\beta}$, the second column presents the second-order bias and MSE derived by Theorems (Thm), the third column presents the Monte Carlo (MC) simulation bias and MSE of the bias-corrected ALS estimators $\widetilde{\beta}$, where $\widetilde{\beta} \equiv \widehat{\beta} - B\left(\widehat{\beta}\right)$. The Monte Carlo results are summarized as follows: (i) $\widetilde{\beta}$ is numerically closer to the true value $\beta = 0$ than $\widehat{\beta}$, as the bias in $\widehat{\beta}$ has been substantially corrected; (ii) the magnitude of bias and MSE is larger in extreme expectiles; (iii) the estimator is unbiased when $\tau = 0.5$, because the ALS model reduces to the OLS model; and (iv) there are upward bias at lower expectiles and downward bias at upper expectiles.

3.6 Appendix

3.6.1 Properties of a norm

$A.1$ If A is a $k \times 1$ vector,

$$||A|| = \left[tr\left(AA'\right)\right]^{1/2} = \left(A'A\right)^{1/2}.$$

$A.2$

$$\left|\left|AA'\right|\right| = \left[tr\left(AA'AA'\right)\right]^{1/2} = \left[tr\left(A'AA'A\right)\right]^{1/2} = \left(A'AA'A\right)^{1/2} = A'A = ||A||^2.$$

$A.3$

$$\begin{aligned}
\left|\left|\left(AA'\right) \otimes A'\right|\right| &= \left\{tr\left(\left[\left(AA'\right) \otimes A'\right]\left[\left(AA'\right) \otimes A\right]\right)\right\}^{1/2} \\
&= \left[tr\left(\left(AA'AA'\right) \otimes \left(A'A\right)\right)\right]^{1/2} \\
&= \left[tr\left(A'AA'AA'A\right)\right]^{1/2} \\
&= \left(A'A\right)^{3/2} \\
&= ||A||^3.
\end{aligned}$$

$A.4$

$$
\begin{aligned}
\left|\left|(AA') \otimes A' \otimes A'\right|\right| &= tr\left(\left[(AA') \otimes A' \otimes A'\right]\left[(AA') \otimes A \otimes A\right]\right)^{1/2} \\
&= tr\left[(AA'AA') \otimes (A' \otimes A')(A \otimes A)\right]^{1/2} \\
&= tr\left[(AA'AA') \otimes A'A \otimes A'A\right]^{1/2} \\
&= tr\left[(A'AA'A)\, A'AA'A\right]^{1/2} \\
&= (A'AA'A) \\
&= (A'A)^2 \\
&= \|A\|^4
\end{aligned}
$$

3.6.2 Properties of the Dirac delta function

$B.1$ The Heaviside unit step function is defined as $\phi(z) = 0$ for $z < 0$, $\phi(z) = 1$ for $z \geq 0$. The Dirac delta function is defined as $\delta(z) = \mathrm{d}\phi(z)/\mathrm{d}z$, where $\delta(z) = 0$ for $z < 0$, $\delta(z) = \infty$ for $z = 0$, $\delta(z) = 0$ for $z > 0$.

$B.2$ $\int_{-\infty}^{+\infty} \delta(z)\mathrm{d}z = 1$.

$B.3$ $\int_{-\infty}^{+\infty} \delta(z-a)f(z)\mathrm{d}z = f(a)$, where $f : R \to R$ is a real function differentiable around $a \in R$.

$B.4$ $\delta(z) = \delta(-z)$.

$B.5$ $\int_{-\infty}^{+\infty} \delta^{(1)}(z-a)f(z)\mathrm{d}z = -\int_{-\infty}^{+\infty} \delta(z-a)f^{(1)}(z)\mathrm{d}z = -f^{(1)}(a)$.

$B.6$ $\delta^{(1)}(-z) = -\delta^{(1)}(z)$, $\delta^{(2)}(-z) = \delta^{(2)}(z)$.

$B.7$ $\int_{-\infty}^{+\infty} \delta^{(n)}(z-a)f(z)\mathrm{d}z = (-1)^n \int_{-\infty}^{+\infty} \delta(z-a)f^{(n)}(z)\mathrm{d}z = (-1)^n f^{(n)}(a)$.

$B.8$ $\phi(z)\delta(z) = \frac{1}{2}\delta(z)$.

$B.9$ $\phi(z)\delta^{(1)}(z) = \frac{1}{2}\delta^{(1)}(z) - (\delta(z))^2$.

3.7 References

Aigner, D.J., Amemiya, T., and Poirier, D., 1976. On the estimation of production frontiers: maximum likelihood estimation of the parameters of a discontinuous density function. *International Economic Review* 17, 377-396.

Bao, Y., Ullah, A., 2007. The second-order bias and mean squared error of estimators in time-series models. *Journal of Econometrics* 140, 650-669.

Gelfand, I.M., Shilov, G.E., 1964. *Generalized Functions* Vol. 1. New York: Academic Press.

Koenker, R., Bassett, G.S., 1978. Regression quantiles. *Econometrica* 46, 33–50.

Kuan, C., Yeh, J., and Hsu, Y., 2009. Assessing value at risk with CARE, the conditional autoregressive expectile models. *Journal of Econometrics* 150, 261-270.

Newey, W.K., Powell, J.L., 1987. Asymmetric Least Squares Estimation and Testing. *Econometrica* Vol. 55, No. 4, 819-847.

Pagan, A., Ullah, A., 1999. *Nonparametric Econometrics.* Cambridge university press.

Raju, C. K., 1982. Products and compositions with the Dirac delta function. *Journal of Physics A: Mathematical and General* 15, 381-396.

Rilstone, P., Srivastava, V. K., Ullah, A., 1996. The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics* 75, 369-395.

Xie, S., Zhou, Y., and Wan, A., 2014, A varying-coefficient expectile model for estimating value at risk. *Journal of Business and Economic Statistics* 32:4, 576-592.

Yao, Q., Tong, H., 1996. Asymmetric least squares regression estimation: a nonparametric approach. *Nonparametric Statistics* 6, 273-292.

Table 3.1: Conditional ALS regression

		$N = 100$			$N = 300$		
τ		$\hat{\beta}_{MC}$	$\hat{\beta}_{Thm}$	$\tilde{\beta}_{MC}$	$\hat{\beta}_{MC}$	$\hat{\beta}_{Thm}$	$\tilde{\beta}_{MC}$
0.1	bias	0.0100	0.0185	-0.0085	0.0031	0.0062	-0.0031
	MSE	0.0057	0.0192	0.0057	0.0018	0.0052	0.0018
0.2	bias	0.0076	0.0091	-0.0015	0.0022	0.0030	-0.0008
	MSE	0.0064	0.0106	0.0063	0.0020	0.0032	0.0020
0.3	bias	0.0048	0.0047	0.0001	0.0002	0.0016	-0.0014
	MSE	0.0068	0.0082	0.0068	0.0022	0.0026	0.0022
0.4	bias	0.0028	0.0021	0.0007	0.0013	0.0007	0.0006
	MSE	0.0073	0.0073	0.0073	0.0023	0.0023	0.0023
0.5	bias	-0.0005	0.0000	-0.0005	0.0006	0.0000	0.0006
	MSE	0.0070	0.0070	0.0070	0.0023	0.0023	0.0023
0.6	bias	-0.0007	-0.0021	0.0013	-0.0007	-0.0007	0.0000
	MSE	0.0072	0.0073	0.0072	0.0022	0.0023	0.0022
0.7	bias	-0.0036	-0.0047	0.0012	-0.0006	-0.0016	0.0010
	MSE	0.0068	0.0082	0.0068	0.0022	0.0026	0.0022
0.8	bias	-0.0057	-0.0091	0.0034	-0.0029	-0.0030	0.0001
	MSE	0.0064	0.0106	0.0064	0.0020	0.0032	0.0020
0.9	bias	-0.0099	-0.0185	0.0086	-0.0037	-0.0062	0.0025
	MSE	0.0058	0.0192	0.0058	0.0018	0.0052	0.0018

Notes: Table 1 presents the simulation results when x_i is generated form exponential distribution. Table 2 presents the simulation results when $x_i = 1$. For each τ, the first row is for bias and the second row is for the mean squared error of the ALS estimator. The results are presented in the following manner in each corresponding cell.

		Monte Carlo (MC)	from theorems (Thm)	Monte Carlo (MC)
τ	bias	$\frac{1}{J}\sum_{j=1}^{J}\left(\widehat{\beta}_j(\tau)-\beta(\tau)\right)$	$B\left(\widehat{\beta}(\tau)\right)$	$\frac{1}{J}\sum_{j=1}^{J}\left(\widetilde{\beta}_j(\tau)-\beta(\tau)\right)$
	MSE	$\frac{1}{J}\sum_{j=1}^{J}\left(\widehat{\beta}_j(\tau)-\beta(\tau)\right)^2$	$M\left(\widehat{\beta}(\tau)\right)$	$\frac{1}{J}\sum_{j=1}^{J}\left(\widetilde{\beta}_j(\tau)-\beta(\tau)\right)^2$

where the true value $\beta(\tau) = 0$ for all τ. The bias-corrected estimate is $\widetilde{\beta}(\tau) = \widehat{\beta}(\tau) - B\left(\widehat{\beta}(\tau)\right)$. The subscript j denotes the jth Monte Carlo replication ($j = 1, \ldots, J$). We replicate $J = 10,000$ times in the Monte Carlo.

Table 3.2: Unconditional ALS regression

		$N = 100$			$N = 300$		
τ		$\hat{\beta}_{MC}$	$\hat{\beta}_{Thm}$	$\tilde{\beta}_{MC}$	$\hat{\beta}_{MC}$	$\hat{\beta}_{Thm}$	$\tilde{\beta}_{MC}$
0.1	bias	0.0064	0.0123	-0.0059	0.0027	0.0041	-0.0014
	MSE	0.0102	0.0099	0.0102	0.0034	0.0073	0.0034
0.2	bias	0.0055	0.0061	-0.0005	0.0010	0.0020	-0.0010
	MSE	0.0120	0.0165	0.0120	0.0039	0.0059	0.0039
0.3	bias	0.0023	0.0032	-0.0009	0.0000	0.0011	-0.0010
	MSE	0.0130	0.0148	0.0130	0.0044	0.0050	0.0044
0.4	bias	0.0018	0.0014	0.0004	0.0009	0.0005	0.0005
	MSE	0.0134	0.0137	0.0134	0.0044	0.0046	0.0044
0.5	bias	-0.0016	0.0000	-0.0016	-0.0005	0.0000	-0.0005
	MSE	0.0134	0.0133	0.0134	0.0044	0.0044	0.0044
0.6	bias	-0.0026	-0.0014	-0.0012	-0.0007	-0.0005	-0.0003
	MSE	0.0133	0.0137	0.0133	0.0044	0.0046	0.0044
0.7	bias	-0.0026	-0.0032	0.0006	-0.0014	-0.0011	-0.0004
	MSE	0.0126	0.0148	0.0126	0.0043	0.0050	0.0043
0.8	bias	-0.0050	-0.0061	0.0011	-0.0013	-0.0020	0.0008
	MSE	0.0119	0.0165	0.0119	0.0039	0.0059	0.0039
0.9	bias	-0.0090	-0.0123	0.0033	-0.0026	-0.0041	0.0016
	MSE	0.0101	0.0099	0.0101	0.0034	0.0073	0.0034

Chapter 4

Second-order Bias Reduction in Predictive Quantile and Exprctile Regressions

4.1 Introduction

Predictive regression is a fundamental econometric model in finance. It has been widely discussed in finance literature. Unlike in the mean predictive regression models for which the bias reduction has been actively developed, there is little studies focused on the predictive quantile regression or predictive expectile regression. In this paper, we develop the predictive quantile and expectile regression models. We apply the second-order bias and MSE results on the application of stock returns. We are able to calculate the second-order bias of the predictive quantile and expectile estimator and the bias reduction enable us to obtain a better predictive estimates. We illustrate the proposed second-order bias reduction to predict the stock returns by the lagged dividend yield. The data used in this application is from Welch and Goyal (2008). We try both short- and long-horizon regressions for both quantile and expectile models. We discover that the bias is larger at the tails of the stock return distribution.

The paper is organized as follows. In Section 4.2, we introduce the predictive quantile and expectile regression model. In Section 4.3, we present the second-order bias and MSE up to $O(N^{-2})$ of both quantile and expectile estimators. In Section 4.4, we present the application of stock returns.

4.2 Predictive Quantile and Expectile Regressions

Consider a simple predictive regression model for stock or portfolio returns using a lagged predictor variable.

$$y_{t+h} = x_t'\beta + u_{t+1}, \tag{4.1}$$

where y_t is returns, and x_t is a $k \times 1$ vector of predictor variables, such as dividend yield or the T-bill rate, which is a first-order autoregressive process, $t = 1, \ldots, T$.

Given $\alpha \in (0,1)$, the predictive quantile regression estimator $\hat{\beta}(\alpha)$ is obtained by minimizing

$$Q_T(\beta; \alpha) = \sum_{t=1}^{T} r_\alpha \left(y_{t+h} - x_t'\beta \right), \tag{4.2}$$

where $r_\alpha(\cdot)$ is the check loss function,

$$r_\alpha(\lambda) \equiv |\alpha - \mathbf{1}(\lambda < 0)| \cdot |\lambda|. \tag{4.3}$$

The $k \times 1$ vector quantile estimators $\widehat{\beta}(\alpha)$ can be obtained by minimizing the expected check loss. The sample moment condition of the predictive quantile regression is

$$\Psi_T(\beta(\alpha)) = \frac{1}{T}\sum_{i=1}^{T} s_t(\beta(\alpha)). \tag{4.4}$$

where

$$s_t(\beta(\alpha)) = (\alpha - \mathbf{1}(y_{t+h} < x_t'\beta))(-x_t) \tag{4.5}$$

Given $\tau \in (0,1)$, the predictive asymmetric least squares (ALS) or predictive expectile regression estimator $\hat{\beta}(\tau)$ is obtained by minimizing

$$R_T(\beta;\tau) = \sum_{t=1}^{T} \rho_\tau\left(y_{t+h} - x_t'\beta\right), \tag{4.6}$$

where $\rho_\tau(\cdot)$ is the asymmetric least squares loss function,

$$\rho_\tau(\lambda) \equiv |\tau - \mathbf{1}(\lambda < 0)| \cdot \lambda^2. \tag{4.7}$$

The $k \times 1$ vector expectile estimators $\widehat{\beta}(\tau)$ can be obtained by minimizing the asymmetric least squares loss. The sample moment condition of the predictive expectile regression is

$$\Psi_T(\beta(\tau)) = \frac{1}{T}\sum_{t=1}^{T} s_t(\beta(\tau)). \tag{4.8}$$

where

$$s_t(\beta(\alpha)) = 2x_t\left(y_{t+h} - x_t'\beta\right)\left[(2\tau - 1)\mathbf{1}(y_{t+h} < x_t'\beta) - \tau\right]. \tag{4.9}$$

4.3 Second-order Bias and MSE for Quantile and Expectile Estimators

In the quantile regression model, suppose x_t and u_{t+h} are not identically distributed, but independent across $t = 1, ..., T$, when $k = 1$, the second-order bias of the quantile estimators $\widehat{\beta}(\alpha)$ up to $O(N^{-1})$ is

$$B\left(\widehat{\beta}(\alpha)\right) = \left(\frac{1}{2}-\alpha\right)Q^2\frac{1}{T^2}\sum_{t=1}^{T}E\left[x_t^3 f_t(0|x_t)\right] - \frac{\alpha(1-\alpha)}{2}Q\frac{1}{T}\sum_{t=1}^{T}E[x_t^3 f_t^{(1)}(x_t'\beta)] \times \frac{1}{T^2}\sum_{t=1}^{T}Q^2E\left(x_t^2\right), \quad (4.10)$$

and the MSE up to $O(N^{-2})$, of the quantile estimator $\widehat{\beta}$ is

$$\begin{aligned} M\left(\widehat{\beta}\right) = & \frac{1}{T^2}\sum_{t=1}^{T}\alpha(1-\alpha)Q^2E\left(x_t^2\right) - 2\frac{1}{T^3}\sum_{t=1}^{T}Q^3\left(\alpha^2-\alpha+\frac{1}{2}\right)E\left[x_t^4 f_t(0|x_t)\right] - \frac{1}{T^3}\sum_{t=1}^{T}\alpha(1-\alpha)Q^2E\left(\right. \\ & -\frac{1}{T^3}\sum_{t=1}^{T}\alpha(1-\alpha)(2\alpha-1)Q^4E\left[x_t^3 f_t^{(1)}(0|x_t)\right]E\left(x_t^3\right) + 6\frac{1}{T^3}\sum_{t=1}^{T}\left(\frac{1}{2}-\alpha\right)^2Q^4\left(E\left[x_t^3 f_t(0|x_t)\right]\right. \\ & +3\frac{1}{T^3}\sum_{t=1}^{T}\alpha(1-\alpha)Q^4\left(\frac{1}{2}-\alpha\right)E\left[x_t^4 f_t^{(1)}(0|x_t)\right]E\left(x_t^2\right) \\ & -12\frac{1}{T^3}\sum_{t=1}^{T}\left(\frac{1}{2}-\alpha\right)\alpha(1-\alpha)Q^5E[x_t^3 f_t^{(1)}(0|x_t)]E\left[x_t^3 f_t(0|x_t)\right]E\left(x_t^2\right) \\ & +\frac{15}{4}\frac{1}{T^3}\sum_{t=1}^{T}\alpha^2(1-\alpha)^2Q^6\left(E\left[x_t^3 f_t^{(1)}(0|x_t)\right]\right)^2\left(E\left(x_t^2\right)\right)^2 \\ & -\frac{1}{T^3}\sum_{t=1}^{T}\alpha^2(1-\alpha)^2Q^5E[x_t^4 f_t^{(2)}(0|x_t)]\left(E\left(x_t^2\right)\right)^2, \quad (4 \end{aligned}$$

where $Q = \left(\frac{1}{T}\sum_{t=1}^{T}E[x_t^2 f_t(0|x_t)]\right)^{-1}$, $f_t(0|x_t)$ is the conditional density of u_{t+h} given x_t evaluated at $u_i = 0$, $f_t^{(1)}(0|x_t)$ and $f_t^{(2)}(0|x_t)$ are the first and second derivative of the conditional density of u_{t+h} given x_t evaluated at $u_{i+h} = 0$.

In the expectile regression model, suppose x_t and u_{t+h} are not identically distributed, but independent across $t = 1, ..., T$, when $k = 1$, the second-orer bias of the expectile estimators $\widehat{\beta}(\tau)$ up to $O(N^{-1})$ is

$$B\left(\widehat{\beta}(\tau)\right) = \frac{1}{T^2}\sum_{t=1}^{T}4Q^2C_2 + \frac{1}{T}\sum_{t=1}^{T}(2\tau-1)QE\left[x_t^3 f_t(0|x_t)\right] \times \frac{1}{T^2}\sum_{t=1}^{T}4Q^2C_1, \quad (4.12)$$

and the MSE of the ALS estimator $\widehat{\beta}(\tau)$ up to $O(N^{-2})$ is

$$\begin{aligned} M\left(\widehat{\beta}\right) &= \frac{1}{T^2}\sum_{t=1}^{T} 4Q^2 C_1 - \frac{1}{T^3}\sum_{t=1}^{T} 16Q^3 C_3 + \frac{1}{T^3}\sum_{t=1}^{T} 8Q^4 C_1 - \frac{1}{T^3}\sum_{t=1}^{T} Q^4 16(2\tau-1)E\left[x_t^3 f_t(0|x_t)\right] C_4 \\ &+ \frac{1}{T^3}\sum_{t=1}^{T} 96Q^4 C_2^2 + \frac{1}{T^3}\sum_{t=1}^{T} 48Q^4 \left\{\tau(1-\tau)E\left(x_t^4\right) - 4\tau^2(1-\tau)^2\left[E\left(x_t^2\right)\right]^2\right\} C_1 \\ &+ \frac{1}{T^3}\sum_{t=1}^{T} 384Q^5(2\tau-1)E\left[x_t^3 f_t(0|x_t)\right] C_1 C_2 + \frac{1}{T^3}\sum_{t=1}^{T} 240Q^6(2\tau-1)^2\left[E\left[x_t^3 f_t(0|x_t)\right]\right]^2 C_1^2 \\ &+ \frac{1}{T^3}\sum_{t=1}^{T} 32Q^5(2\tau-1)E\left[x_t^4 f_t^{(1)}(0|x_t)\right] C_1^2, \end{aligned} \tag{4.13}$$

where

$$\begin{aligned} Q &= \left(4\tau(1-\tau)\frac{1}{T}\sum_{t=1}^{T} E\left[x_t^2\right]\right)^{-1}, \\ C_1 &= \frac{1}{T}\sum_{t=1}^{T}\left[E\left[-(2\tau-1)x_t^2 u_{t+h}^2 \mathbf{1}(u_{t+h}<0)\right] + E\left(\tau^2 x_t^2 u_{t+h}^2\right)\right], \\ C_2 &= \frac{1}{T}\sum_{t=1}^{T}\left[E\left[(2\tau-1)x_t^3 u_{t+h}\mathbf{1}(u_{t+h}<0)\right] - E\left(\tau^2 x_t^3 u_{t+h}\right)\right], \\ C_3 &= \frac{1}{T}\sum_{t=1}^{T}\left[E\left[-(2\tau-1)\left(\tau^2-\tau+1\right)x_t^4 u_{t+h}^2\mathbf{1}(u_i<0)\right] + E\left(\tau^3 x_t^4 u_{t+h}^2\right)\right], \\ C_4 &= \frac{1}{T}\sum_{t=1}^{T}\left[E\left[(2\tau-1)\left(\tau^2-\tau+1\right)x_t^3 u_{t+h}^3\mathbf{1}(u_{t+h}<0)\right] - E\left(\tau^3 x_t^3 u_{t+h}^3\right)\right], \end{aligned}$$

and $f_t(0|x_t)$ is the conditional density of u_{t+h} given x_t evaluated at $u_i = 0$, $f_t^{(1)}(0|x_t)$ is the first derivative of the conditional density of u_{t+h} given x_t evaluated at $u_{i+h} = 0$.

4.4 Empirical Application

As an empirical application, our second-order bias reduction approach is illustrated through a predictive model in finance. We predict the stock returns by the lagged dividend yielde. There is extensive literature on the stock return prediction. See Lewellen (2004) and Zhu (2013). In the empirical application, we aim to illustrate the proposed method and show the effect of bias reduction. The data are monthly return series of the S&P 500 Index from Amit Goyal's website. Welch and Goyal (2008) provide detailed descriptions of the data. The dividend yield is the ratio of the previous 12-month sum of dividends paid on the S&P 500 Index. According to Ang and Bekaert (2007), Paye and Timmermann (2006) and Goyal and Welch (2003), the interested rate data are hard to interpret before the 1951 Treasury Accord, and the dividend yield predictability is not robust with the 1990s. Therefore, we use the period from 01/1952 to 12/1989 (total 450 months), which is the post-Accord period. The application uses a rolling window sample of $T = 100$ observations. We construct a predictive quantile regression and a predictive expectile regression. We predict future h-period returns onto dividend yield. In the equation (4.1), we try with different horizon, that is $h = 1, 3$, and 12. Table 1, Table 2 and Table 3 present the predictive quantile results based on 344 replietitions for each value of horizon h. Table 4, Table 5 and Table 6 present the predictive expectile results based on 344 replietitions for each value of horizon h. For each level of α or τ, the first column presents the quantile estimator of $\widehat{\beta}$. The second column presents the second-order bias $B\left(\widehat{\beta}\right)$ derived in equation (4.10) for Table 1 to 3, and equation (4.12) for Table 4 to 6. The third column presents the second-order bias corrected quantile estimators $\widetilde{\beta} = \widehat{\beta} - B\left(\widehat{\beta}\right)$. The fourth column presents the the mean squared error $M\left(\widehat{\beta}\right)$ up to $O(N^{-1})$ obtained by the first term in equation (4.11) for Table 1 to 3, and equation (4.13) for Table 4 to 6. The last column presents the the mean squared error $M\left(\widehat{\beta}\right)$ up to $O(N^{-2})$ derived in equation (4.11) for Table 1 to 3, and equation (4.13) for Table 4 to 6.

The summary statistics of the stock return and dividend yield are presented in the Table 6 in Zhu (2013). Zhu (2013) pointed out that the stock return distribution has fat tails as evidenced by the large excess kurtosis. In our results tables below, we find that (i) the magnitude of the second-order bias and MSE is larger towards the tails of the stock return distribution; (ii) there are upward bias at lower quantiles and downward bias at upper quantiles; (iii) the MSE up to $O(N^{-1})$ is smaller than the MSE up to $O(N^{-2})$.

4.5 References

Amihud, Y., Hurvich, C.M., 2007. Stock return predictability: Is it there? *Review of Financial Sttudies* 20: 651-707.

Goyal, A., Welch. I, 2003. Predicting the equity premium with dividend ratios. *Management Science* 49: 639-654.

Goyal, A., Welch. I, 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21(4) 1455-1508.

Paye, B., Timmermann, A., 2006. Instability of return prediction models. *Jourmal of Empirical Finance* 13: 274-315.

Zhu, M., 2013. Jackknife for bias reduction in predictive regressions. *Journal of Financial Econometrics* 11(1), 193-220.

Rilstone, P., Srivastava, V. K., Ullah, A., 1996. The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics* 75, 369-395.

Table 4.1: Second-order bias reduction in predictive quantile regression, h=1

α	$\hat{\beta}$	$B(\hat{\beta})$	$\tilde{\beta}$	$AsyMSE$	$MSE(\hat{\beta})$
0.05	21.4330	0.0418	21.3912	0.0806	0.6401
0.1	22.6019	0.0254	22.5765	0.0652	0.3325
0.2	24.2561	0.0133	24.2428	0.0550	0.2426
0.3	25.4255	0.0075	25.4180	0.0529	0.2550
0.4	26.4911	0.0033	26.4878	0.0499	0.2494
0.5	27.4588	-0.0001	27.4589	0.0497	0.2756
0.6	28.4041	-0.0037	28.4078	0.0498	0.2975
0.7	29.5814	-0.0078	29.5891	0.0557	0.4831
0.8	30.7923	-0.0139	30.8062	0.0585	0.9343
0.9	32.4173	-0.0239	32.4412	0.0667	2.7309
0.95	33.4169	-0.0363	33.4532	0.0648	3.1295

Notes: For each level of α, the first column presents the quantile estimators $\widehat{\beta}$. The second column presents the second-order bias $B\left(\widehat{\beta}\right)$ derived in equation (4.10). The third column presents the second-order bias corrected quantile estimators $\widetilde{\beta} \equiv \widehat{\beta} - B\left(\widehat{\beta}\right)$. The fourth column presents the MSE up to . The last column presents the MSE up to derived in equation equation (4.11).

Table 4.2: Second-order bias reduction in predictive quantile regression, h=3

α	$\hat{\beta}$	$B(\hat{\beta})$	$\tilde{\beta}$	$AsyMSE$	$MSE(\hat{\beta})$
0.05	21.6167	0.0421	21.5746	0.0796	0.5784
0.1	22.7921	0.0257	22.7665	0.0656	0.3275
0.2	24.4771	0.0136	24.4636	0.0558	0.2485
0.3	25.6484	0.0076	25.6408	0.0539	0.2628
0.4	26.7110	0.0034	26.7076	0.0512	0.2596
0.5	27.6863	-0.0002	27.6865	0.0510	0.2873
0.6	28.6523	-0.0038	28.6562	0.0510	0.3198
0.7	29.8833	-0.0079	29.8912	0.0567	0.5243
0.8	31.1202	-0.0139	31.1341	0.0597	0.9344
0.9	32.6784	-0.0240	32.7024	0.0663	2.6087
0.95	33.7131	-0.0367	33.7499	0.0665	3.1980

Notes: See notes for Table 1.

Table 4.3: Second-order bias reduction in predictive quantile regression, h=12

α	$\hat{\beta}$	$B(\hat{\beta})$	$\tilde{\beta}$	$AsyMSE$	$MSE(\hat{\beta})$
0.05	22.5823	0.0443	22.5380	0.0847	0.5048
0.1	23.9541	0.0254	23.9287	0.0684	0.3481
0.2	25.6044	0.0137	25.5907	0.0559	0.2746
0.3	26.7643	0.0078	26.7565	0.0534	0.2974
0.4	27.7779	0.0035	27.7744	0.0522	0.3282
0.5	28.7469	-0.0003	28.7472	0.0522	0.3738
0.6	29.7894	-0.0039	29.7932	0.0526	0.4567
0.7	30.9788	-0.0079	30.9868	0.0560	0.6295
0.8	32.3023	-0.0141	32.3164	0.0666	1.1107
0.9	33.9143	-0.0247	33.9390	0.0676	2.5825
0.95	34.9662	-0.0375	35.0037	0.0677	3.0623

Notes: See notes for Table 1.

Table 4.4: Second-order bias reduction in predictive expectile regression, h=1

α	$\hat{\beta}$	$B(\hat{\beta})$	$\tilde{\beta}$	$AsyMSE$	$MSE(\hat{\beta})$
0.05	23.3676	-0.2223	23.5899	0.6815	1.7870
0.1	24.3002	-0.0454	24.3456	0.3831	0.4348
0.2	25.3977	-0.0041	25.4018	0.2238	0.2203
0.3	26.1668	0.0007	26.1662	0.1733	0.1696
0.4	26.8118	0.0014	26.8104	0.1530	0.1498
0.5	27.4116	0.0015	27.4101	0.1478	0.1448
0.6	28.0142	0.0017	28.0125	0.1542	0.1514
0.7	28.6678	0.0031	28.6648	0.1755	0.1729
0.8	29.4468	0.0090	29.4377	0.2266	0.2270
0.9	30.5533	0.0490	30.5043	0.3827	0.4595
0.95	31.4819	0.2143	31.2676	0.6545	1.9325

Notes: For each level of τ, the first column presents the expectile estimators $\widehat{\beta}$. The second column presents the second-order bias $B\left(\widehat{\beta}\right)$ derived in equation (4.12). The third column presents the second-order bias corrected quantile estimators $\widetilde{\beta} \equiv \widehat{\beta} - B\left(\widehat{\beta}\right)$. The fourth column presents the MSE up to . The last column presents the MSE up to derived in equation equation (4.13).

Table 4.5: Second-order bias reduction in predictive expectile regression, h=3

α	$\hat{\beta}$	$B(\hat{\beta})$	$\tilde{\beta}$	$AsyMSE$	$MSE(\hat{\beta})$
0.05	23.5536	-0.2231	23.7767	0.6918	1.8158
0.1	24.5021	-0.0465	24.5486	0.3907	0.4449
0.2	25.6157	-0.0044	25.6201	0.2286	0.2252
0.3	26.3939	0.0006	26.3933	0.1770	0.1733
0.4	27.0470	0.0014	27.0456	0.1564	0.1532
0.5	27.6550	0.0015	27.6535	0.1511	0.1481
0.6	28.2660	0.0017	28.2643	0.1577	0.1548
0.7	28.9291	0.0031	28.9260	0.1795	0.1768
0.8	29.7183	0.0093	29.7090	0.2321	0.2324
0.9	30.8399	0.0515	30.7885	0.3948	0.4743
0.95	31.7838	0.2265	31.5574	0.6841	2.0507

Notes: See notes for Table 4.

Table 4.6: Second-order bias reduction in predictive expectile regression, h=12

α	$\hat{\beta}$	$B(\hat{\beta})$	$\tilde{\beta}$	$AsyMSE$	$MSE(\hat{\beta})$
0.05	24.5971	-0.2331	24.8303	0.7443	2.0015
0.1	25.6018	-0.0551	25.6568	0.4170	0.4936
0.2	26.7453	-0.0076	26.7529	0.2389	0.2376
0.3	27.5324	-0.0008	27.5332	0.1832	0.1798
0.4	28.1934	0.0006	28.1928	0.1613	0.1580
0.5	28.8095	0.0009	28.8086	0.1555	0.1524
0.6	29.4297	0.0012	29.4285	0.1621	0.1590
0.7	30.1027	0.0026	30.1002	0.1846	0.1813
0.8	30.9069	0.0085	30.8984	0.2391	0.2372
0.9	32.0596	0.0476	32.0119	0.4068	0.4592
0.95	33.0362	0.2088	32.8274	0.6965	1.7265

Notes: See notes for Table 4.

Chapter 5

Conclusions

Chapter two developed analytical results on the finite sample properties of quantile estimators. We have provided the general results on the second-order bias and MSE of quantile estimators. We discover that while the median is unbiased for a symmetric distribution, and the bias of the other quantiles is larger at the tails of any distribution. The Monte Carlo simulations results indicate the improvement of quantile estimators and quantile prediction. The theoretical results are illustrated in quantile estimation of the impact of schooling on earnings, and the effect of smoking and prenatal care on birthweight. We find that the second-order bias corrected estimator has better behaviors than the uncorrected ones, and the bias is larger at the extreme low and high earning and birthweight quantiles. The prediction error becomes smaller with the second order bias correction, which implies that the bias correction improves the accuracy of estimation and prediction of quantile regressions.

Chapter three provides the results on the second-order bias and MSE of ALS regression models. The second-order bias result enables an improved bias correction and thus to obtain improved ALS estimations. We show that the second-order bias is much larger as the asymmetry is stronger, and therefore the benefit of the second-order bias correction is greater when we are interested in extreme expectiles. The higher order MSE result for the ALS estimation also enables us to better understand the sources of estimation uncertainty. The Monte Carlo simulation indicates that the second-order bias corrected ALS estimator has better behaviou than the uncorrected ones.

Chapter four illustrates the second-order bias reduction in the predictive quantile and expectile regressions, which enables an improved predictive quantile and expectile estimates. We show that the second-order bias are much larger towards the tails of the conditional density than near the median, and therefore the benefit of the second-order bias reduction is greater when we are interested in the deeper tail quantiles. The empirical application of stock returns highlights the benefit of the proposed approach.