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

Tae-Hwy Lee[†] Aman Ullah[‡] He Wang[§]

August 28, 2023

Abstract

The finite sample theory using higher order asymptotics provides better approximations of the bias and mean squared error (MSE) for a class of estimators. Rilston, Srivastava and Ullah (1996) provided the second-order bias results of conditional mean regression. This paper 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 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 estimation. The second-order bias corrected estimators are numerically much closer to the true parameters of the 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. The empirical application of the theory to the predictive quantile regression model in finance highlights the benefit of the proposed second-order bias reduction.

Key Words: Check-loss, Dirac delta function, Quantile regression, Second-order bias, MSE.
JEL Classification: C13, C33, C52

---

*We thank the seminar participants at the California Econometrics Conference (CEC) at Stanford University and the International Conference on Computational and Financial Econometrics (CFE) in London for many useful comments.

[†]Department of Economics, University of California, Riverside, CA 92521. E-mail: tae.lee@ucr.edu

[‡]Department of Economics, University of California, Riverside, CA 92521. E-mail: aman.ullah@ucr.edu

[§]Department of Insurance, University of International Business and Economics, Beijing. E-mail: hewang72@gmail.com

1 Introduction

Over the last six decades, Professor Kosaraju Leela Krishna, popularly known as "KLK" among his students and colleagues at the Delhi School of Economics (DSE), contributed immensely through his teaching and research interests in the fields of Applied Econometrics, Industrial Economics, Economics of Productivity, and Empirics of Trade, and wrote research papers on a variety of topics. In fact, a product of University of Chicago, KLK has been a founding econometrician guiding and mentoring both Ph.D. and M.Phil. students at the DSE, and has been a charismatic guide showing how to use econometrics tools for doing high quality practical work and answering deeper economic and policy questions. All these made him very popular among all scholars. He is amongst the most distinguished economists in India, which is also reflected in the honors and awards he received from many organizations, including founding Managing Editor of the *Journal of Quantitative Economics*, President of The Indian Econometric Society (1996-1997), PJ Thomas Foundation Economist of the Year Award (2015-2016), and Distinguished Service Award from University of Delhi, among others. Our paper contributes to the finite sample behavior of the quantile estimators, which are robust and in recent years frequently used in applied economics and econometrics work, instead of estimating only the mean estimator.

It is well known that the large sample properties of an estimator and a test statistic may not imply their finite sample behavior. In fact, the use of the first-order asymptotic theory for small or even moderately large samples may give misleading results. There has been significant literature on analytical "finite sample properties" of econometric estimators and test statistics over the past six decades.[1] See, among others, Nagar (1959), Sargan (1974, 1976), Basmann (1974), Phillips (1977), Rothenberg (1984) for linear models, and Amemiya (1980), Chesher and Spady (1989), Cordeiro and McCullagh (1991), Newey and Smith (2004), Rilstone, Srivastava and Ullah (RSU, 1996), Bao and Ullah (2007), and Ullah (2004) for non-linear models.

The finite sample theory has been developed extensively for the mean regression models, which provides a better approximation of the bias and mean squared error (MSE) and thus improves finite sample inference. It also enables us to examine quality of instruments and to understand

[1] We refer the finite sample properties to the higher order asymptotic approximations, 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 exact distributional properties. See Ullah (2004).

what affects the behavior of estimators and how to improve it from correcting for the higher order bias. In particular, RSU (1996) developed the second-order bias and mean squared error (MSE) of a class of nonlinear estimators in conditional mean regression models with i.i.d. samples. Bao and Ullah (2007) extended the RSU results for time series dependent observations.

However, unlike in the mean regression models for which both the first-order asymptotic theory and the finite sample (higher-order asymptotic) theory have been fully developed, the quantile regression literature has been almost entirely based on the first-order asymptotic theory. The quantile literature has been either on the first-order asymptotic expansion (Koenker and Bassett 1978) or on determining the order of the higher order remainder term in the first-order asymptotic expansion of the quantile estimators (Bahadur 1966, Kiefer 1967, Jureckova and Sen 1987, 1996, He and Shao 1996, De Angelis, Hall, and Young 1993, and Chapter 4 of David and Nagaraja 2003).

In this paper, unlike in the existing quantile literature mentioned above, we extend the RSU results for the quantile regressions focusing on the second-order terms. We derive the analytical expressions of the second-order bias up to the order $O\left(N^{-1}\right)$ and the MSE up to the order $O\left(N^{-2}\right)$ for quantile regression estimators, using the higher-order asymptotic expansions. The challenge to study the finite sample properties of quantile estimators is due to the non-differentiability of the objective function for the quantile estimation. While 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. Phillips (1991) used the Dirac delta functions for the median regression estimators. Whang (2006) and Otsu (2008) used moment smoothing for empirical likelihood quantile regression. The related idea of smoothing non-differentiable objective functions has been used for quantile regression by Kaplan and Sun (2017) and Fernandes et al. (2017). We also use the properties of the Dirac delta function and obtain the finite sample properties for the quantile regression estimators in the second-order bias and MSE. We show that the second-order bias result enables an improved bias correction and thus to obtain improved quantile estimation and prediction. We also 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. Among many interesting findings, we find that the second-order bias is 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 financial risk management. The higher order MSE result for the quantile estimation also enables us to better understand the sources of estimation uncertainty.

The paper is organized as follows. In Section 2, we present the notations, the moment condition of the quantile regression, and the assumptions used in this paper. In Section 3, we develop the high-order asymptotic expansion of quantile estimators, and derive the second-order bias of conditional quantile estimators. In Section 4, we derive the second-order MSE of conditional quantile estimators. Section 5 provides three examples for illustrations, which include the unconditional quantile estimation, the conditional quantile regression with a binary covariate, and the IVQR estimation. Section 6 presents Monte Carlo simulations. In Section 7, an empirical application is presented for the predictive quantile regression model for the financial returns. Section 8 contains the conclusion.

2 Conditional Quantile Estimators

2.1 Check Loss Function

Consider a random variable y from the distribution $F(\cdot)$. Let $f_i(\cdot)$ denote the conditional density, for $i = 1, \ldots, N$. $f_i^{(j)}(\cdot)$ denotes 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, $[\text{trace}(AA')]^{1/2}$. If A is a $k \times 1$ vector, according to Appendix A, $\|A\| = (A'A)^{1/2}$. 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

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

where the check loss function is defined as

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

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

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

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

$$y_i = x_i'\beta_\alpha + u_i, \tag{1}$$

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 coefficients β can be obtained by solving

$$\min_\beta E\left[L_\alpha(\beta)\right] = E\left[\alpha - \mathbf{1}(y < x'\beta)\right] \cdot \left(y - x'\beta\right). \tag{2}$$

Following Elliott, Komunjer, and Timmermann (2005), we assume that the conditional α-quantile of y, $x'\beta$, is identified on the parameter space Θ, that is, for any $\beta_1, \beta_2 \in \Theta$ we have $x'\beta_1 = x'\beta_2$ a.s. $-P$, if and only if $\beta_1 = \beta_2$. The check loss function $L_\alpha(\beta) = [\alpha - \mathbf{1}(y < x'\beta)]\,(y - x'\beta)$ is continuously differentiable on $\Theta\backslash A$, where $A = \{\beta \in \Theta : y = x'\beta\}$. Let $\nabla^1_\beta E[L_\alpha(\beta)]$ denote the gradient of $E[L_\alpha(\beta)]$ on $\Theta\backslash A$. By the law of iterated expectations,

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

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 we can write the population moment condition as

$$\nabla^1_\beta E[L_\alpha(\beta)] = E[-\nabla^1_\beta \mathbf{1}(y - x'\beta < 0)(y - x'\beta)] + E[(\alpha - \mathbf{1}(y < x'\beta))(-x)]. \tag{3}$$

Let $\mathbf{1}(y - x'\beta < 0) \equiv \phi(x'\beta - y)$ a Heaviside unit step function. Then by the definition of the Dirac delta function in Appendix B,

$$\nabla^1_\beta \mathbf{1}(y - x'\beta < 0) = \nabla^1_\beta \phi(x'\beta - y) = \frac{\mathrm{d}\phi(x'\beta - y)}{\mathrm{d}(x'\beta - y)} \frac{\mathrm{d}(x'\beta - y)}{\mathrm{d}\beta} = x'\delta(x'\beta - y).$$

See Gelfand and Shilov (1964). The first term of the equation (3) can be written as $E[x'\delta(x'\beta - y)(y - x'\beta)]$, which equals zero.

According to the property of Dirac delta function in Appendix B, we have $\delta(x'\beta - y) = \delta(y - x'\beta)$ and

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

where $f(x'\beta) \equiv f(x'\beta|x)$ is the conditional density of y evaluated at $y = x'\beta$.

Thus, the moment condition can be written as

$$\nabla^1_\beta E[L_\alpha(\beta)] = E[(\alpha - \mathbf{1}(y < x'\beta))(-x)] = E[s(\beta)] = 0, \tag{4}$$

where the score function $s(\beta) \equiv [\alpha - \mathbf{1}(y < x'\beta)]\,(-x)$. The score function $s(\beta) \equiv [\alpha - \mathbf{1}(y < x'\beta)]\,(-x)$ is a special case of the score of the form $s(\beta) \equiv [\alpha - \mathbf{1}(y < x'\beta)]\,(-z)$ with some instrument variable $z = x$ for an IVQR. With $z = x$, the moment condition gives the conditional quantile regression. With $z = 1$, the moment condition gives the unconditional quantile regression. The main results of the paper are of the second-order bias and MSE for various quantile estimators satisfying the above moment condition (4).

2.2 Assumptions

Denoting $s_i(\beta) \equiv [\alpha - \mathbf{1}(y_i < x_i'\beta)]\,(-x_i)$, the sample moment condition can be written as

$$\Psi_N(\beta) = \frac{1}{N}\sum_{i=1}^{N} s_i(\beta). \tag{5}$$

A class of estimators $\widehat{\beta}$ can be written as a solution to a set of moment equations of the form

$$\Psi_N(\widehat{\beta}) = \frac{1}{N}\sum_{i=1}^{N} s_i(\widehat{\beta}) = 0, \tag{6}$$

where $s_i(\beta)$ is a known $k \times 1$ vector-valued function of the observable k-dimensional random vectors x_i and a parameter vector $\beta \in \mathbb{R}^k$ with true value β_0 such that $E[s_i(\beta)] = 0$ holds only at $\beta = \beta_0$ for all i.

In this paper, we assume the moment condition $\Psi_N(\widehat{\beta}) = 0$ holds exactly as in Phillips (1991, Equation 4). Also, following Phillips (1991, Assumption $\mathfrak{A}_1$), we require the conditional density $f(y)$ to be analytic. These assumptions are usually unimportant for the first-order asymptotic theory but could matter for higher-order results. Nevertheless there will be gains to make these strong assumptions. Not only do they help in developing generalized Taylor series but also they facilitate the derivations of the higher-order results, for example, to demonstrate that the second-order bias is much larger towards the tails of the conditional density than near the median and therefore provide useful insights on the finite sample bias when we are interested in the deeper tail quantiles.

RSU (1996) developed the second-order bias and MSE of a class of estimators. These results apply for both normal and non-normal errors. The moment equation $\Psi_N(\cdot)$ can be the first-order condition of some optimization criteria. The estimators can be the maximum likelihood (ML), least square (LS), or Generalized Method of Moments (GMM) estimators. In RSU (1996), their Assumptions A-C are sufficient for $\widehat{\beta}$ to have an asymptotically normal distribution. To obtain the stochastic expansion of $\widehat{\beta}$, the RSU's Assumptions A-C are assumed to hold along with the $\sqrt{N}$-consistency of $\widehat{\beta}$. For the RSU's (1996) results to hold for the quantile model, we make some modifications to their Assumptions A-C as follows.

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 $\nabla_\beta^0 s_i(\beta) = s_i(\beta)$.

Assumptions A-C are related to the conditions in Komunjer (2005) but include other primitive conditions. The conditions in Komunjer (2005) are stated to obtain the asymptotic normality of conditional quantile estimators to handle the non-smoothness of the quantile objective function. See also Huber (1976), Pollard (1985), Pakes and Pollard (1989), Newey and McFadden (1994), Andrews (1994), Chen, Linton, and van Keilegom (2003), and Chernozhukov and Hong (2003). In this paper, Assumption C requires conditions of the higher order stochastic equicontinuity for the higher order stochastic expansion. Let us discuss these assumptions in some details.

First, we discuss 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 condition $\Psi_N(\beta)$ is continuously 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 derivatives of $\Psi_N(\beta)$ exist and are continuous with probability 1. By the definition of the Dirac delta function in Appendix B, 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 first-order derivative of a $k \times 1$ vector $s_i(\beta)$ with respect to a $k \times 1$ vector β is a $k \times k$ 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 the first derivative of the 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 properties in Appendixes A and B, we obtain

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

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 the second derivative of the score function and its expected value converges in probability to zero, i.e., $\frac{1}{N}\sum_{i=1}^N \left(x_i x_i'\right) \otimes x_i' \delta^{(1)}(x_i'\beta - y_i) - E\left[\left(x_i x_i'\right) \otimes x_i' \delta^{(1)}(x_i'\beta - y_i)\right] \xrightarrow{p} 0$. Using the properties in Appendixes A and B, we obtain

$$
\begin{aligned}
E\left\|\nabla_{\beta}^{2}s_{i}(\beta_{0})\right\|^{2} &= E\left\|(x_{i}x_{i}')\otimes x_{i}'\delta^{(1)}(x_{i}'\beta_{0}-y_{i})\right\|^{2}\\
&= E\left\|\left(x_{i}x_{i}'\right)\otimes x_{i}'E\left[\delta^{(1)}(x_{i}'\beta_{0}-y_{i})|x_{i}\right]\right\|^{2}\\
&= E\left\|\left(x_{i}x_{i}'\right)\otimes x_{i}'\left(\int_{-\infty}^{+\infty}\delta^{(1)}(x_{i}'\beta_{0}-y_{i})f_{i}(y_{i})\mathrm{d}y_{i}\right)\right\|^{2}\\
&= E\left\|\left(x_{i}x_{i}'\right)\otimes x_{i}'\left(-\int_{-\infty}^{+\infty}\delta^{(1)}(y_{i}-x_{i}'\beta_{0})f_{i}(y_{i})\mathrm{d}y_{i}\right)\right\|^{2}\\
&= E\left\|\left(x_{i}x_{i}'\right)\otimes x_{i}'\left(\int_{-\infty}^{+\infty}\delta(y_{i}-x_{i}'\beta_{0})f_{i}^{(1)}(y_{i})\mathrm{d}y_{i}\right)\right\|^{2}\\
&= E\left[f_{i}^{(1)}(x_{i}'\beta_{0})\left\|\left(x_{i}x_{i}'\right)\otimes x_{i}'\right\|\right]^{2}\\
&= E\left[f_{i}^{(1)}(x_{i}'\beta_{0})\left\{\mathrm{tr}\left(\left[\left(x_{i}x_{i}'\right)\otimes x_{i}'\right]\left[\left(x_{i}x_{i}'\right)\otimes x_{i}\right]\right)\right\}^{1/2}\right]^{2}\\
&= E\left[f_{i}^{(1)}(x_{i}'\beta_{0})\left[\mathrm{tr}\left(\left(x_{i}x_{i}'x_{i}x_{i}'\right)\otimes\left(x_{i}x_{i}'\right)\right)\right]^{1/2}\right]^{2}\\
&= E\left[f_{i}^{(1)}(x_{i}'\beta_{0})\left[\mathrm{tr}\left(x_{i}'x_{i}x_{i}'x_{i}x_{i}'x_{i}\right)\right]^{1/2}\right]^{2}\\
&= E\left[f_{i}^{(1)}(x_{i}'\beta_{0})\left(x_{i}'x_{i}\right)^{3/2}\right]^{2}\\
&= E\left[f_{i}^{(1)}(x_{i}'\beta_{0})\left\|x_{i}\right\|^{3}\right]^{2}\\
&< \infty.
\end{aligned}
$$

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 $s_i(\beta)$ 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. $\frac{1}{N}\sum_{i=1}^{N}\left(x_ix_i'\right)\otimes x_i'\otimes$

$x_i'\delta^{(2)}(x_i'\beta - y_i) - E\left[(x_ix_i') \otimes x_i' \otimes x_i'\delta^{(2)}(x_i'\beta - y_i)\right] \xrightarrow{p} 0$. Using the the properties in Appendixes A and B, we obtain

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

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

$$\begin{aligned}
E\left\|\nabla_\beta^1 s_i(\beta_0)\right\|^2 &= E\left[\left\|x_i\right\|^2 f_i(0|x_i)\right]^2 < \infty, \\
E\left\|\nabla_\beta^2 s_i(\beta_0)\right\|^2 &= E\left[\left\|x_i\right\|^3 f_i^{(1)}(0|x_i)\right]^2 < \infty, \\
E\left\|\nabla_\beta^3 s_i(\beta_0)\right\|^2 &= E\left[\left\|x_i\right\|^4 f_i{}^{(2)}(0|x_i)\right]^2 < \infty.
\end{aligned}$$

Combining 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 $j \geq 1$, with $f_i^{(0)}(0|x_i) = f_i(0|x_i)$.

Next, let us discuss 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

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

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 sufficiently smooth, which is not the case with the quantile regression. In general, Assumption C requires the stochastic equicontinuity conditions to handle the expansion of discontinuous and non-smooth objective function. This problem has been discussed in many papers in the literature, 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 its limit if the remainder term is small enough. Therefore, those stochastic equicontinuity conditions do not require differentiability of the objective function, but require that the remainder term of the expansion can be controlled in a particular way over a neighborhood of β_0. Besides of those stochastic conditions discussed in the literature mentioned above, in this paper we need additional smoothness and dominating conditions for higher moments of the quantile objective function. 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 the Lipschitz moment function, and allows for moments of the objective function to be Lipschitz at β_0 and differentiable with probability 1, rather than continuously differentiable. Assumption C in this paper restricts the remainder to be well behaved uniformly near the true parameter β_0, 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.

3 Second-order Bias of Quantile Estimators

Following RSU(1996), we define the second-order bias for a class of estimators in general as follows. For a class of estimators β, the second-order bias is the expection of the asymptotic distribution of

$\left(\widehat{\beta}-\beta\right)$ *up to* the second-order, i.e., of order $O\left(N^{-1}\right)$.

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,

$$0=\Psi_N+\nabla\Psi_N(\widehat{\beta}-\beta_0)+\frac{1}{2}\nabla^2\Psi_N\left[(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\right]+o_p\left(N^{-1}\right), \tag{7}$$

where $\Psi_N=\Psi_N\left(\beta_0\right)$. The ordinary stochastic expansion of $\widehat{\beta}$ is obtained from equation (7). However, a difficulty arises from the derivatives of the moment condition. Using the properties of the delta function in Appendix B or in Phillips (1991, p. 455), we can rewrite (7) as

$$\begin{aligned} 0 &= \Psi_N+\overline{\nabla\Psi_N}(\widehat{\beta}-\beta_0)+\left(\nabla\Psi_N-\overline{\nabla\Psi_N}\right)(\widehat{\beta}-\beta_0)+\frac{1}{2}\nabla^2\Psi_N\left[(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\right]+o_p\left(N^{-1}\right) \\ &\equiv A_1+A_2+A_3+A_4+o_p\left(N^{-1}\right), \end{aligned} \tag{8}$$

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

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

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

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,

$$\sqrt{N}(\widehat{\beta}-\beta_0)\overset{d}{\to}N(0,V_\alpha), \tag{9}$$

where

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

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 $\sqrt{N}$-consistent, 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

$$\widehat{\beta}-\beta_0=a_{-1/2}+R_N, \tag{10}$$

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

$$R_N = O_p\left(n^{-3/4}(\log\log n)^{3/4}\right). \tag{11}$$

See Koenker (2005 pp. 122-123), and also Jureckova and Sen (1987, 1996 pp. 196-202), He and Shao (1996), van der Vaart (1998 p. 310), and Portnoy (2012). Note that (11) implies that

$$R_N = O_p\left(N^{-3/4+\varepsilon}\right) \quad \text{for some small } \varepsilon > 0. \tag{12}$$

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 of this section 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 discussed in Lemma 3.

Lemma 1. Let

$$\begin{aligned} A_3 &= \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)(\widehat{\beta} - \beta_0) \\ &= \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)a_{-1/2} + \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)\left[(\widehat{\beta} - \beta_0) - a_{-1/2}\right] \\ &\equiv A_{31} + A_{32}. \end{aligned} \tag{13}$$

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, p. 457), 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 $O_p\left(N^{-1/3}N^{-1/3}\right)$, that is $\sqrt{N}A_{31} = O_p\left(N^{-2/3}\right)$. Then A_{31} is $O_p\left(N^{-7/6}\right)$.

(b) By (11), 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

$$\begin{aligned} A_4 &= \frac{1}{2}\nabla^2\Psi_N\left[(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\right] \\ &= \frac{1}{2}\overline{\nabla^2\Psi_N}\left[(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\right]+\frac{1}{2}\left(\nabla^2\Psi_N-\overline{\nabla^2\Psi_N}\right)\left[(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\right] \\ &\equiv A_{41}+A_{42}. \end{aligned} \tag{14}$$

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 (10), A_{41} can be written as

$$\begin{aligned} A_{41} &= \frac{1}{2}\overline{\nabla^2\Psi_N}\left\{\left[(\widehat{\beta}-\beta_0)-a_{-1/2}+a_{-1/2}\right]\otimes\left[(\widehat{\beta}-\beta_0)-a_{-1/2}+a_{-1/2}\right]\right\} \\ &= \frac{1}{2}\overline{\nabla^2\Psi_N}\left(a_{-1/2}\otimes a_{-1/2}\right) \\ &\quad+\frac{1}{2}\overline{\nabla^2\Psi_N}\left(a_{-1/2}\otimes\left[(\widehat{\beta}-\beta_0)-a_{-1/2}\right]\right) \\ &\quad+\frac{1}{2}\overline{\nabla^2\Psi_N}\left(\left[(\widehat{\beta}-\beta_0)-a_{-1/2}\right]\otimes a_{-1/2}\right) \\ &\quad+\frac{1}{2}\overline{\nabla^2\Psi_N}\left(\left[(\widehat{\beta}-\beta_0)-a_{-1/2}\right]\otimes\left[(\widehat{\beta}-\beta_0)-a_{-1/2}\right]\right), \end{aligned} \tag{15}$$

Recall that $\overline{\nabla^2\Psi_N}=E\left[(x_ix_i')\otimes x_i'f_i^{(1)}(0|x_i)\right]=O(1)$. Only the first term in equation (15) 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 (15) 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 (8) can be written as

$$\begin{aligned} 0 &= A_1+A_2+A_{31}+A_{41}+o_p\left(N^{-1}\right) \\ &= \Psi_N+\overline{\nabla\Psi_N}(\widehat{\beta}-\beta_0)+\left(\nabla\Psi_N-\overline{\nabla\Psi_N}\right)a_{-1/2}+\frac{1}{2}\overline{\nabla^2\Psi_N}\left(a_{-1/2}\otimes a_{-1/2}\right)+o_p\left(N^{-1}\right) \end{aligned} \tag{16}$$

The term $\nabla\Psi_N$ in an ordinary Taylor expansion, equation (7), 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 the delta function and $(\nabla\Psi_N)^{-1}$

is not bounded. Now in the equation (16), the Taylor expansion of quantile regression, $\overline{\nabla\Psi_N}$ is invertible, because $\left(\overline{\nabla\Psi_N}\right)^{-1}$ is bounded by Assumption B. In equation (16), we keep the term A_{31} even though it is $O_p(N^{-7/6})$ by Lemma 1, 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 (16) to obtain

$$\begin{aligned}
\widehat{\beta} - \beta_0 &= -\overline{\nabla\Psi_N}^{-1}\Psi_N - \overline{\nabla\Psi_N}^{-1}\left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)a_{-1/2} - \frac{1}{2}\overline{\nabla\Psi_N}^{-1}\overline{\nabla^2\Psi_N}\left(a_{-1/2}\otimes a_{-1/2}\right) + o_p\left(N^{-1}\right) \\
&= -Q\Psi_N - QV_N a_{-1/2} - \frac{1}{2}Q\overline{H_2}\left(a_{-1/2}\otimes a_{-1/2}\right) + o_p\left(N^{-1}\right) \qquad (17) \\
&\equiv B_1 + B_2 + B_3 + o_p\left(N^{-1}\right),
\end{aligned}$$

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 (17) by $\sqrt{N}$ gives a generalization of equation (15) of Phillips (1991, p. 457) for general α. 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 (17).

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) $B_2 \equiv -QV_N a_{-1/2} = O_p\left(N^{-7/6}\right)$, 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 both are not identically distributed, but independent across $i = 1, ..., N$. Suppose y_i has the 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 (17), 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

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

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

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

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

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

(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

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

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

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

From equation (17), note that the bias of quantile estimators $\widehat{\beta}$ is

$$\begin{aligned}
E\left(\widehat{\beta} - \beta_0\right) &= E\left(B_1\right) + E\left(B_2\right) + E\left(B_3\right) + o\left(N^{-1}\right) \\
&= E\left(-Q\Psi_N\right) + E\left(-QV_N a_{-1/2}\right) + E\left(-\frac{1}{2}Q\overline{H_2}\left(a_{-1/2} \otimes a_{-1/2}\right)\right) + o\left(N^{-1}\right) \quad (18) \\
&\equiv B\left(\widehat{\beta}\right) + o\left(N^{-1}\right).
\end{aligned}$$

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

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

$$\begin{aligned}
B\left(\widehat{\beta}\right) &= E\left[-QV_N a_{-1/2} - \frac{1}{2}Q\overline{H_2}\left(a_{-1/2}\otimes a_{-1/2}\right)\right] \\
&= \left(\frac{1}{2}-\alpha\right)Q\frac{1}{N^2}\sum_{i=1}^{N}E\left[x_i x_i' Q x_i f_i(0|x_i)\right] \\
&\quad -\frac{\alpha(1-\alpha)}{2}Q\frac{1}{N}\sum_{i=1}^{N}E[\left(x_i x_i'\right)\otimes x_i' f_i^{(1)}(x_i'\beta)]\frac{1}{N^2}\sum_{i=1}^{N}\left(Q\otimes Q\right)E\left(x_i\otimes x_i\right), \qquad (19)
\end{aligned}$$

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

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

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

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

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

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

Remark: When x_i˜$i.i.d$ and u_i˜$i.i.d$, 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

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

The quantile estimator $\widehat{\beta}$ is unbiased if x_i follows a symmetric distribution with $E\left(x_i^3\right)=0$. If u_i follows 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.

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

$$\begin{aligned} 0 &= \Psi_N+\overline{\nabla\Psi_N}(\widehat{\beta}-\beta_0)+\left(\nabla\Psi_N-\overline{\nabla\Psi_N}\right)(\widehat{\beta}-\beta_0)+\frac{1}{2}\nabla^2\Psi_N\left[(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\right] \\ &\quad +\frac{1}{6}\nabla^3\Psi_N\left[(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\otimes(\widehat{\beta}-\beta_0)\right]+o_p\left(N^{-3/2}\right) \\ &\equiv A_1+A_2+A_3+A_4+A_5+o_p\left(N^{-3/2}\right). \end{aligned} \tag{20}$$

Our goal of this section 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 (20) in the following lemmas.

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

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

$$\begin{aligned} A_{32} &= \left(\nabla\Psi_N-\overline{\nabla\Psi_N}\right)\left[(\widehat{\beta}-\beta_0)-a_{-1/2}\right] \\ &= \left(\nabla\Psi_N-\overline{\nabla\Psi_N}\right)\left[B_2+B_3+o_p\left(N^{-1}\right)\right] \\ &= \left(\nabla\Psi_N-\overline{\nabla\Psi_N}\right)\left[B_2+B_3\right]+o_p\left(N^{-4/3}\right), \end{aligned}$$

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

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 2, 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 $O_p\left(N^{-1/2}\right)$, 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 discussed each term in equation (20). Now the equation (20) can be written as

$$
\begin{aligned}
0 &= \Psi_N+\overline{\nabla\Psi_N}(\widehat{\beta}-\beta_0)+\left(\nabla\Psi_N-\overline{\nabla\Psi_N}\right)\left(a_{-1/2}+a_{-1}\right) \\
&\quad +\frac{1}{2}\overline{\nabla^2\Psi_N}\left(a_{-1/2}\otimes a_{-1/2}\right)+\frac{1}{2}\overline{\nabla^2\Psi_N}\left[\left(a_{-1/2}\otimes a_{-1}\right)+\left(a_{-1}\otimes a_{-1/2}\right)\right] \\
&\quad +\frac{1}{2}\left(\nabla^2\Psi_N-\overline{\nabla^2\Psi_N}\right)\left(a_{-1/2}\otimes a_{-1/2}\right) \\
&\quad +\frac{1}{6}\overline{\nabla^3\Psi_N}\left[a_{-1/2}\otimes a_{-1/2}\otimes a_{-1/2}\right]+o_p\left(N^{-3/2}\right).
\end{aligned} \tag{21}
$$

The equation (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 $B_2\equiv -QV_Na_{-1/2}=O_p\left(N^{-7/6}\right)$, then $B_2B_2'=O_p\left(N^{-7/3}\right)$. However, we found that $E\left(B_2B_2'\right)=O\left(N^{-2}\right)$, which we will discuss in the following Lemma.

Solve for $\widehat{\beta}-\beta_0$ in equation (16) to obtain

$$\begin{aligned}
\widehat{\beta}-\beta_0 &= -\overline{\nabla\Psi_N}^{-1}\Psi_N-\overline{\nabla\Psi_N}^{-1}\left(\nabla\Psi_N-\overline{\nabla\Psi_N}\right)\left(a_{-1/2}+a_{-1}\right) \\
&\quad -\frac{1}{2}\overline{\nabla\Psi_N}^{-1}\overline{\nabla^2\Psi_N}\left(a_{-1/2}\otimes a_{-1/2}\right)-\frac{1}{2}\overline{\nabla\Psi_N}^{-1}\overline{\nabla^2\Psi_N}\left[\left(a_{-1/2}\otimes a_{-1}\right)+\left(a_{-1}\otimes a_{-1/2}\right)\right] \\
&\quad -\frac{1}{2}\overline{\nabla\Psi_N}^{-1}\left(\nabla^2\Psi_N-\overline{\nabla^2\Psi_N}\right)\left(a_{-1/2}\otimes a_{-1/2}\right) \\
&\quad -\frac{1}{6}\overline{\nabla\Psi_N}^{-1}\overline{\nabla^3\Psi_N}\left[a_{-1/2}\otimes a_{-1/2}\otimes a_{-1/2}\right]+o_p\left(N^{-3/2}\right) \\
&= \left\{-Q\Psi_N\right\}+\left\{-QV_Na_{-1/2}\right\}+\left\{-\frac{1}{2}Q\overline{H_2}\left(a_{-1/2}\otimes a_{-1/2}\right)\right\}+\left\{-QV_Na_{-1}-\frac{1}{2}QW_N\left(a_{-1/2}\otimes a_{-1/2}\right)\right\} \\
&\quad +\left\{-\frac{1}{2}Q\overline{H_2}\left[\left(a_{-1/2}\otimes a_{-1}\right)+\left(a_{-1}\otimes a_{-1/2}\right)\right]-\frac{1}{6}Q\overline{H_3}\left[a_{-1/2}\otimes a_{-1/2}\otimes a_{-1/2}\right]\right\} \\
&\quad +o_p\left(N^{-3/2}\right), \\
&\equiv B_1+B_2+B_3+B_4+B_5+o_p\left(N^{-3/2}\right), \qquad (22)
\end{aligned}$$

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 (22) is the same as the expression in RSU (1996 p. 390 Eq. A.17).

Lemma 7.

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

(b) $B_1B_1'=O_p\left(N^{-1}\right)$, and $E\left(B_1B_1'\right)=O\left(N^{-1}\right)$,

(c) $B_1B_2'=B_2B_1'=O_p\left(N^{-5/3}\right)$, and $E\left(B_1B_2'\right)=E\left(B_2B_1'\right)=O\left(N^{-2}\right)$,

(d) $B_1B_3'=B_3B_1'=O_p\left(N^{-3/2}\right)$, 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) $B_2B_2'=O_p\left(N^{-7/3}\right)$, 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) $B_3B_3'=O_p\left(N^{-2}\right)$, 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

$$E\left(B_1 B_1'\right) = \overline{d_i^2},$$

$$E\left(B_1 B_2' + B_2 B_1'\right) = -2Q\overline{V_i d_i^2},$$

$$E\left(B_1 B_3' + B_3 B_1'\right) = Q\overline{H_2}\overline{d_i^3},$$

$$\begin{aligned} E\left(B_1 B_4' + B_4 B_1'\right) &= 2Q^2\overline{V_i^2}\,\overline{d_i^2} + 4Q^2\overline{V_i V_j d_i d_j} - 9Q^2\overline{H_2}\overline{V_i d_i d_j^2} + 3Q\overline{W_i d_i d_j^2} \\ &= 2Q^2\overline{V_i^2}\,\overline{d_i^2} + 4Q^2\overline{V_i d_i}^2 - 9Q^2\overline{H_2}\;\overline{V_i d_i}\overline{d_i^2} + 3Q\overline{W_i d_i}\overline{d_i^2} \end{aligned}$$

$$E\left(B_2 B_2'\right) = 2Q^2\overline{V_i V_j d_1 d_2} + Q^2\overline{V_i^2}\,\overline{d_i^2} = 2Q^2\overline{V_i d_i}^2 + Q^2\overline{V_i^2}\,\overline{d_i^2},$$

$$\begin{aligned} E\left(B_1 B_5' + B_5 B_1'\right) &= 3Q^2\overline{H_2}^2\overline{d_i^2 d_j^2} - Q\overline{H_3}\overline{d_i^2 d_j^2} \\ &= 3Q^2\overline{H_2}^2\overline{d_i^2}^2 - Q\overline{H_3}\overline{d_i^2}^2 \end{aligned}$$

$$E\left(B_3 B_3'\right) = \frac{3}{4}Q^2\overline{H_2}^2\overline{d_i^2 d_j^2} - 3Q^2\overline{H_2}\overline{V_i d_i d_j^2} = \frac{3}{4}Q^2\overline{H_2}^2\overline{d_i^2}^2 - 3Q^2\overline{H_2}\;\overline{V_i d_i}\overline{d_i^2},$$

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

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

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

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

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

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

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

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

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

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

■

From equation (22), note that the MSE of quantile estimator $\widehat{\beta}$ is

$$E\left(\widehat{\beta}-\beta_0\right)\left(\widehat{\beta}-\beta_0\right)' = E\left(B_1B_1'\right)+E\left(B_1B_2'+B_2B_1'\right)+E\left(B_1B_3'+B_3B_1'\right)+E\left(B_1B_4'+B_4B_1'\right) +E\left(B_2B_2'\right)+E\left(B_1B_5'+B_5B_1'\right)+E\left(B_3B_3'\right)+o_p\left(N^{-2}\right) \equiv M\left(\widehat{\beta}\right)+o_p\left(N^{-2}\right). \quad (23)$$

Given the above results in Lemma 7, we define the MSE of quantile estimators up to $O(N^{-2})$ as follows.

Definition 2. Let $E\left(\widehat{\beta}-\beta_0\right)\left(\widehat{\beta}-\beta_0\right)' = M\left(\widehat{\beta}\right)+o_p\left(N^{-2}\right)$. Then $M\left(\widehat{\beta}\right)$ will be called "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 *both 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*

$$\begin{aligned} M\left(\widehat{\beta}\right) = {} & \frac{1}{N}\alpha(1-\alpha)Q^2E\left(x_i^2\right)-2\frac{1}{N^3}\sum_{i=1}^{N}Q^3\left(\alpha^2-\alpha+\frac{1}{2}\right)E\left[x_i^4f_i(0|x_i)\right]+2\frac{1}{N^3}\sum_{i=1}^{N}\alpha(1-\alpha)Q^2E\left(x_i^2\right) \\ & -\frac{1}{N^3}\sum_{i=1}^{N}\alpha(1-\alpha)(2\alpha-1)Q^4E\left[x_i^3f_i^{(1)}(0|x_i)\right]E\left(x_i^3\right)+6\frac{1}{N^3}\sum_{i=1}^{N}\left(\frac{1}{2}-\alpha\right)^2Q^4\left(E\left[x_i^3f_i(0|x_i)\right]\right)^2 \\ & +3\frac{1}{N^3}\sum_{i=1}^{N}\alpha(1-\alpha)Q^4\left(\frac{1}{2}-\alpha\right)E\left[x_i^4f_i^{(1)}(0|x_i)\right]E\left(x_i^2\right) \\ & -3\frac{1}{N^3}\sum_{i=1}^{N}\alpha(1-\alpha)Q^4\left(E\left[x_i^2f_i(0|x_i)\right]\right)^2E\left(x_i^2\right) \\ & -12\frac{1}{N^3}\sum_{i=1}^{N}\left(\frac{1}{2}-\alpha\right)\alpha(1-\alpha)Q^5E\left[x_i^3f_i^{(1)}(0|x_i)\right]E\left[x_i^3f_i(0|x_i)\right]E\left(x_i^2\right) \\ & +\frac{15}{4}\frac{1}{N^3}\sum_{i=1}^{N}\alpha^2(1-\alpha)^2Q^6\left(E\left[x_i^3f_i^{(1)}(0|x_i)\right]\right)^2\left(E\left(x_i^2\right)\right)^2 \\ & -\frac{1}{N^3}\sum_{i=1}^{N}\alpha^2(1-\alpha)^2Q^5E\left[x_i^4f_i^{(2)}(0|x_i)\right]\left(E\left(x_i^2\right)\right)^2, \end{aligned} \quad (24)$$

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

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 7, the MSE of the quantile estimator $\widehat{\beta}$ up to $O\left(N^{-2}\right)$ can be written as

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

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

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

The asymptotic distribution of the quantile regression estimator when the α-quantile is linear in x_i, is given by equation (9). 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

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

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

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

The asymptotic variance is

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

■

Corollary 2.2. *When x_i˜i.i.d. and u_i˜i.i.d., and $k=1$, the expression of the MSE of $\widehat{\beta}$ up to $O(N^{-2})$ can be simplified as*

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

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

5 Illustrations

In this section, we consider three special cases of the general results on the conditional quantile regression from the previous section: namely, (1) the unconditional quantile estimation, (2) the conditional quantile regression with a binary independent variable, and (3) 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.

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*

$$B\left(\widehat{\beta}\right) = \frac{1}{N}\left(\frac{1}{2} - \alpha\right)Q - \frac{1}{N}\frac{\alpha(1-\alpha)}{2}Q^3 f^{(1)}(0), \tag{25}$$

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

$$\begin{aligned} M\left(\widehat{\beta}\right) &= \frac{1}{N}\alpha(1-\alpha)Q^2 + \frac{1}{N^2}\left(7\alpha^2 - 7\alpha - \frac{3}{2}\right)Q^2 - 7\frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2}-\alpha\right)Q^4 f^{(1)}(\beta) \\ &\quad + \frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^6\left(f^{(1)}(\beta)\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^5 f^{(2)}(\beta), \end{aligned} \tag{26}$$

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 derivatives 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 following the normal distribution $N(\mu, \sigma^2)$, the second-order bias up to $O(N^{-1})$, of the unconditional quantile estimators $\widehat{\beta}$ is*

$$B\left(\widehat{\beta}\right) = \frac{1}{N}\left(\frac{1}{2} - \alpha\right)Q - \frac{1}{N}\frac{\alpha(1-\alpha)}{2}Q^2\left(\frac{-\beta+\mu}{\sigma^2}\right), \tag{27}$$

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

$$\begin{aligned} M\left(\widehat{\beta}\right) &= \frac{1}{N}\alpha(1-\alpha)Q^2 + \frac{1}{N^2}\left(7\alpha^2 - 7\alpha - \frac{3}{2}\right)Q^2 - 7\frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2}-\alpha\right)Q^3\frac{-\beta+\mu}{\sigma^2} \\ &\quad + \frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^4\left(\frac{-\beta+\mu}{\sigma^2}\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^4\frac{(-\beta+\mu)^2-\sigma^2}{\sigma^4}, \end{aligned} \tag{28}$$

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 follows 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 $\left|\frac{\beta-\mu}{\sigma}\right|$ is high, Q is large; (ii) when σ is high, $\left|B(\widehat{\beta})\right|$ is large; (iii) when $|\beta-\mu|$ is large, $\left|B\left(\widehat{\beta}\right)\right|$ is large.

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

$$M\left(\widehat{\beta}\right) = \frac{\pi\sigma^2}{2N} - \frac{13\pi\sigma^2}{2N^2} + \frac{\pi^2\sigma^2}{4N^2}. \tag{29}$$

and non-negative MSE requires $N \geq 13-\pi/2$. □

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

$$\begin{aligned} M(\widehat{\beta}) &= \frac{1}{N}\alpha(1-\alpha)Q^2 + \frac{1}{N^2}(7\alpha^2-7\alpha-\frac{3}{2})Q^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2Q^4\frac{(-\beta+\mu)^2-\sigma^2}{\sigma^4} \\ &= \frac{1}{N}\frac{1}{4}2\pi\sigma^2 - \frac{1}{N^2}\frac{13}{4}2\pi\sigma^2 + \frac{1}{N^2}\frac{1}{16}4\pi^2\sigma^4\frac{1}{\sigma^2} \\ &= \frac{\pi\sigma^2}{2N} - \frac{13\pi\sigma^2}{2N^2} + \frac{\pi^2\sigma^2}{4N^2}. \end{aligned}$$

■

Corollary 3.3. *In the quantile regression model with $x_i = 1$ and y_i following the 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*

$$B\left(\widehat{\beta}\right) = \frac{1}{N}\left(\frac{1}{2}-\alpha\right)Q + \frac{1}{N}\frac{\alpha(1-\alpha)}{2}\lambda Q^2, \tag{30}$$

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

$$M\left(\widehat{\beta}\right) = \frac{1}{N}\alpha(1-\alpha)Q^2 + \frac{1}{N^2}(7\alpha^2 - 7\alpha - \frac{3}{2})Q^2 + 7\frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2} - \alpha\right)\lambda Q^3 + \frac{11}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2\lambda^2 Q^4. \tag{31}$$

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

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

$$\begin{aligned} f(y_i) &= \lambda \exp(-\lambda y_i), \\ f^{(1)}(y_i) &= -\lambda^2 \exp(-\lambda y_i) = -\lambda f(y_i), \\ f^{(2)}(y_i) &= \lambda^3 \exp(-\lambda y_i) = \lambda^2 f(y_i). \end{aligned}$$

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 follows the exponential distribution $\exp(\lambda)$ with $\lambda > 0$, at the median, the second-order bias of the unconditional quantile estimator is*

$$B\left(\widehat{\beta}\right) = \frac{1}{2N\lambda}, \tag{32}$$

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

$$M\left(\widehat{\beta}\right) = \frac{1}{N\lambda^2} - \frac{13}{N^2\lambda^2} + \frac{11}{4N^2\lambda^2}. \tag{33}$$

□

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. ■

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 following the Bernoulli distribution $Bernoulli(p)$, the second-order bias up to $O(N^{-1})$, of the conditional quantile estimator $\widehat{\beta}$ is*

$$B\left(\widehat{\beta}\right)=\frac{1}{N}\left(\frac{1}{2}-\alpha\right)Q-\frac{1}{N}\frac{\alpha(1-\alpha)}{2}Q^3p^2f^{(1)}(0) \tag{34}$$

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

$$\begin{aligned} M\left(\widehat{\beta}\right) &= \frac{1}{N}\alpha(1-\alpha)Q^2p-\frac{1}{N^2}\left(\alpha\left(1-\alpha\right)\left(4+p\right)+\frac{1}{2}\right)Q^2-7\frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2}-\alpha\right)Q^4p^2f^{(1)}(0) \\ &\quad +\frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2Q^6p^4\left(f^{(1)}(0)\right)^2-\frac{1}{N^2}\alpha^2(1-\alpha)^2Q^5p^3f^{(2)}(0), \end{aligned} \tag{35}$$

where $Q=[pf(0)]^{-1}$. $f(0)=f\left(u_i|x_i=1\right)$ evaluated at $u_i=0$, $f^{(1)}(0)=f^{(1)}\left(u_i|x_i=1\right)$ and $f^{(2)}(0)=f^{(2)}\left(u_i|x_i=1\right)$ 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 follows a symmetric distribution, the median estimator is unbiased.

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

$$B\left(\widehat{\beta}\right)=\frac{1}{N}\left(\frac{1}{2}-\alpha\right)Q-\frac{1}{N}\frac{\alpha(1-\alpha)}{2}Q^2p\frac{-\beta+\mu}{\sigma^2} \tag{36}$$

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

$$\begin{aligned} M\left(\widehat{\beta}\right) &= \frac{1}{N}\alpha(1-\alpha)Q^2p-\frac{1}{N^2}\left(\alpha\left(1-\alpha\right)\left(4+p\right)+\frac{1}{2}\right)Q^2-7\frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2}-\alpha\right)Q^3p\frac{-\beta+\mu}{\sigma^2} \\ &\quad +\frac{1}{N^2}\alpha^2(1-\alpha)^2Q^4p^2\left[\frac{15}{4}\left(\frac{-\beta+\mu}{\sigma^2}\right)^2-\frac{\left(-\beta+\mu\right)^2-\sigma^2}{\sigma^4}\right], \end{aligned} \tag{37}$$

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

Proof: If $y_i|x_i$ follows the 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 the second-order bias: (i) when $\left|\frac{\beta-\mu}{\sigma}\right|$ is high, Q is large; (ii) when σ is high, $\left|B\left(\widehat{\beta}\right)\right|$ is large; (iii) when $|\beta - \mu|$ is large, $\left|B\left(\widehat{\beta}\right)\right|$ is large; and (iv) when p is small, $\left|B\left(\widehat{\beta}\right)\right|$ is large.

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

$$M\left(\widehat{\beta}\right) = \frac{\pi\sigma^2}{2Np} - \frac{\pi\sigma^2}{2N^2p} - \frac{3\pi\sigma^2}{N^2p^2} + \frac{\pi^2\sigma^2}{4N^2p^4}, \tag{38}$$

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

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

■

Corollary 4.3. *In the quantile regression model with x_i following the Bernoulli distribution Bernoulli(p) and $y_i|x_i$ following the 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*

$$B\left(\widehat{\beta}\right) = \frac{1}{N}\left(\frac{1}{2}-\alpha\right)Q + \frac{1}{N}\frac{\alpha(1-\alpha)}{2}\lambda Q^2 p, \tag{39}$$

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

$$\begin{aligned} M\left(\widehat{\beta}\right) &= \frac{1}{N}\alpha(1-\alpha)Q^2p - \frac{1}{N^2}\left(\alpha(1-\alpha)(4+p)+\frac{1}{2}\right)Q^2 + 7\frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2}-\alpha\right)Q^3p\lambda \\ &\quad + \frac{11}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2Q^4p^2\lambda^2, \end{aligned} \tag{40}$$

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

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

$$\begin{aligned} f(y_i|x_i=1) &= \lambda\exp(-\lambda y_i), \\ f^{(1)}(y_i|x_i=1) &= -\lambda^2\exp(-\lambda y_i) = -\lambda f(y_i|x_i=1), \\ f^{(2)}(y_i|x_i=1) &= \lambda^3\exp(-\lambda y_i) = \lambda^2 f(y_i|x_i=1), \end{aligned}$$

and

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

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 $y_i|x_i$ follows the exponential distribution, $f(y_i|x_i) = \lambda\exp(-\lambda y_i)$ with $\lambda > 0$, at the median, the second-order bias of the conditional quantile estimator is*

$$B\left(\widehat{\beta}\right) = \frac{1}{2Np\lambda}, \tag{41}$$

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

$$M\left(\widehat{\beta}\right) = \frac{1}{Np\lambda^2} - \frac{1}{N^2p\lambda^2} - \frac{13}{4N^2p^2\lambda^2}, \tag{42}$$

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. ■

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

$$\begin{aligned} y_i &= x_i'\beta + u_i, & (43) \\ x_i &= \Gamma z_i + v_i, & (44) \end{aligned}$$

where y_i is a scalar, x_i is a $k \times 1$ vector, and z_i is an $l \times 1$ vector. $E\left(x_i u_i\right) \neq 0$, $E\left(u_i | z_i\right) = 0$. We consider the case when $l = k$ below. When $l = k = 1$, the $k \times l$ matrix Γ becomes a scalar γ. The $k \times 1$ vector quantile estimators $\widehat{\beta}$ can be written as a solution to a set of moment equations of the form

$$\Psi_N(\widehat{\beta}) = \frac{1}{N} \sum_{i=1}^{N} s_i(\widehat{\beta}) = 0, \tag{45}$$

where $s_i(\beta) \equiv \left[\alpha - \mathbf{1}(y_i < x_i'\beta)\right](-z_i)$.

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

$$B\left(\widehat{\beta}\right) = \frac{1}{N} Q \left[\left(\frac{1}{2} - \alpha\right) E\left[z_i x_i' Q z_i f(0)\right] - \frac{\alpha(1-\alpha)}{2} E\left[\left(z_i x_i'\right) \otimes x_i' f^{(1)}(0)\right] (Q \otimes Q) E\left(z_i \otimes z_i\right) \right], \tag{46}$$

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

$$\begin{aligned} M\left(\widehat{\beta}\right) = {} & \frac{1}{N}\alpha(1-\alpha) Q^2 E\left(z_i^2\right) - 2\frac{1}{N^2} Q^3 \left(\alpha^2 - \alpha + \frac{1}{2}\right) E\left[z_i^3 x_i\right] f(0) + 2\frac{1}{N^2}\alpha(1-\alpha) Q^2 E\left(z_i^2\right) \\ & - \frac{1}{N^2}\alpha(1-\alpha)(2\alpha-1) Q^4 E\left[z_i x_i^2\right] E\left(z_i^3\right) f^{(1)}(0) + 6\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)^2 Q^4 \left(E\left[z_i^2 x_i\right] f(0)\right)^2 \\ & + 3\frac{1}{N^2}\alpha(1-\alpha) Q^4 \left(\frac{1}{2}-\alpha\right) E\left[z_i^2 x_i^2\right] E\left(z_i^2\right) f^{(1)}(0) \\ & - 3\frac{1}{N^2}\alpha(1-\alpha) Q^4 \left(E\left[x_i^2\right] f(0)\right)^2 E\left(z_i^2\right) & (47) \\ & - 12\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)\alpha(1-\alpha) Q^5 E[z_i x_i^2] E\left[z_i^2 x_i\right] E\left(z_i^2\right) f(0) f^{(1)}(0) \\ & + \frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^6 \left(E\left[z_i x_i^2\right] f^{(1)}(0)\right)^2 \left(E\left(z_i^2\right)\right)^2 \\ & - \frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^5 E[z_i x_i^3] \left(E\left(z_i^2\right)\right)^2 f^{(2)}(0), & (48) \end{aligned}$$

where $Q = (E[z_i x_i] f(0))^{-1}$. □

Proof: See Appendix C. ■

Remark: When $k = 1$, 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

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

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

The second-order bias of $\widehat{\beta}$ is larger at the tails of a distribution. When the instrumental variable is weak, the second-order bias of $\widehat{\beta}$ is larger. If u_i follows a 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 the quantile estimator for the i.i.d. case when $k = 1$ can be simplified as

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

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,

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

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

6 Monte Carlo Simulation

6.1 Simulation Design

Now we give some numerical calculation to the second-order bias and MSE. 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. We consider two data generating processes (DGP).

In the first DGP (DGP1), the error term u_i is normally distributed with the CDF $F(\cdot)$, whose standard deviation is σ_u and mean is set to be $-\Phi^{-1}(\alpha)\sigma_u$, with $\Phi(\cdot)$ denoting the standard normal

CDF. Then we note that

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

Therefore, we generate the error term u_i from the normal distribution $N(-\Phi^{-1}(\alpha)\sigma_u, \sigma_u^2)$. To allow hetroskedasticity, u_i can be set depending on x_i.

In the second DGP (DGP2), the error term u_i is uniformly distributed with the CDF $F(\cdot)$ on $[a, b]$ with $a = \frac{\alpha}{\alpha-1}b$. Then we note that

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

Therefore, we generate the error term u_i from the uniform distribution on $[a, b]$, where $a = -\alpha R$, $b = R(1-\alpha)$, and the range $R = b - a$. Also to allow hetroskedasticity, u_i can be set depending on x_i.

We simulate x_i from several different distributions. Then, y_i is simulated from $y_i = x_i\beta + u_i$. We consider $\alpha \in \{0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95\}$, $\beta = 0$, $N \in \{60, 100\}$. We use the Matlab package by Roger Koenker to estimate the models. The results are presented with 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 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 corollaries, the second column presents the bias and MSE of $\widehat{\beta}$, and the third column presents the bias and MSE of the bias-corrected estimator $\widetilde{\beta} \equiv \widehat{\beta} - B\left(\widehat{\beta}\right)$.

Tables 1-8 present the results for DGP1. We use the Matlab package by Roger Koenker to estimate the model. Table 1 shows the results in Theorems 1 and 2 when there is hetroskedasticity, $\sigma_{ui} = 0.1x_i, 0.5x_i,$. Tables 2-5 show the results in Corollary 1, when $x_i \tilde{} i.i.d.$ and $u_i \tilde{} i.i.d.$. Tables

2 and 3 compares the results when $\sigma_u \in \{0.1, 0.5\}$. In Tables 1, 2, 3, x_i is generated from the exponential distribution with $f(x_i) = \exp(-x_i)$. Tables 4 and 5 show the results when x_i is generated from two different normal mixture distributions in Marron and Wand (1992): in Table 4, x_i is generated from the Skewed Unimodal Density $\frac{1}{5}N(0,1) + \frac{1}{5}N\left(\frac{1}{2}, \left(\frac{2}{3}\right)^2\right) + \frac{3}{5}N\left(\frac{13}{12}, \left(\frac{5}{9}\right)^2\right)$, while in Table 5 x_i is generated from the Strongly Skewed Density $\sum_{l=0}^{7} \frac{1}{8}N\left(3\left[\left(\frac{2}{3}\right)^l - 1\right], \left(\frac{2}{3}\right)^{2l}\right)$. See Marron and Wand (1992, page 717, Fig 1, #2 and #3) for shapes of these two normal mixture densities.

Table 6 shows the simulation results for 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 estimation is a special case of the conditional quantile estimation with binary independent variable with $p = 1$.

Table 8 presents the results for IVQR for which we use the Matlab package by Chernozhukov and Hansen (2006) to estimate the IVQR model. In the simulation of IVQR, u_i is generated from DGP1 (normal); v_i is simulated from $v_i = w_i + cu_i$ such that v is contaminated by the structural error u and thus v becomes endogenous, where w_i is from $N(0, 0.25)$, $c = 0.5$; z_i is from the exponential distribution, $f(x_i) = \exp(-x_i)$; x_i is simulated from $x_i = z_i\gamma + v_i$, where $\gamma \in \{0.5, 0.9\}$; y_i is simulated from $y_i = x_i\beta + u_i$, where $\beta = 0$.

6.2 Simulation Results

From the results for DGP1 reported in Table 1-8, we find that the analytically derived second-order bias is numerically close to the Monte Carlo simulated bias, the estimator $\widetilde{\beta} \equiv \widehat{\beta} - B\left(\widehat{\beta}\right)$ with the second-order-bias-correction is numerically close to the true parameter value $\beta = 0$ of the data generation. The Monte Carlo simulation results show that the second-order bias corrected estimator has better behavior than the uncorrected estimator $\widehat{\beta}$. The results for DGP2 are qualitatively similar to those for DGP1. The results for DGP2 are made available in the Supplemental Appendix (Tables 10-13) on the authors' website. From the simulation results, our conclusions are summarized as follows:

1. If x_i is generated from the standard normal distribution, the bias is close to zero. That is because the expressions of the second-order bias contain the third-moment of x_i. If the

distribution of x_i is symmetric, the second-order bias will go to zero. Therefore, we simulate x_i from several asymmetric distributions. Since the exponential distribution and the two mixture normal distributions are all asymmetric, the bias-corrected estimator $\widetilde{\beta}$ is closer to the true β_0 than the uncorrected estimator $\widehat{\beta}$.

2. The first column for each sample size N shows that the bias is zero at the median, and the bias is larger at deeper tail quantiles.

3. When the sample size is increasing, the bias becomes smaller. The quantile estimators are asymptotically unbiased.

4. When σ_u is larger, the quantile estimator has larger bias.

5. In the IVQR, the bias is larger with weaker instruments.

7 Empirical Application

In this section, we demonstrate the benefit of the second-order-bias-correction in the predictive quantile regression model for financial returns. We predict conditional quantiles of the stock returns conditioning on the lagged dividend yields. There is extensive literature on the stock return prediction. See Lewellen (2004) and Zhu (2013) among many others. To examine the effect of the second-order-bias-correction in the predictive quantile regression, we consider a linear predictive quantile model for the h-period ahead portfolio return y_{t+h}

$$y_{t+h} = x_t'\beta + u_{t+h}, \qquad t = 1, \ldots, T \tag{49}$$

where y_t is the return, and x_t is a $k \times 1$ vector of predictor variables such as dividend yield or the T-bill rate. We consider $k = 1$ here. Given $\alpha \in (0, 1)$, the predictive quantile regression estimator $\widehat{\beta}\,(\alpha)$ is obtained by solving

$$\min_{\beta} E[L_\alpha(\beta)] = E[\left(\alpha - \mathbf{1}(y_{t+h} < x_t'\beta)\right)\left(y_{t+h} - x_t'\beta\right)]. \tag{50}$$

The data are monthly from Amit Goyal's website. Welch and Goyal (2008) provide detailed descriptions of the data. y_{t+h} is the future h-period returns on the S&P 500 Index, defined as $y_{t+h} = (P_{t+h} - P_t)/P_t$, where P_t is the the S&P 500 Index. The dividend yield x_t is the ratio of the previous 12-month sum of dividends paid on the S&P 500 Index. Following Ang and Bekaert (2007), Paye and Timmermann (2006) and Goyal and Welch (2003), we use the data after the 1951 Treasury Accord period, from January 1952 to December 1989 (total 456 months). The application uses a rolling window sample of $T = 100$ observations. We predict future h-period returns using the dividend yield. We consider three different horizons $h \in \{1, 3, 12\}$. Table 9 presents the predictive quantile results for $h = 1$ in the out-of-sample average over the 344 rolling windows ($344 = 456 - 100 - 12$, with 12 observations used for taking lags for $h = 12$). The results for $h = 3, 12$ are made available in the Supplemental Appendix (Tables 14, 15) for space reason as they are very similar to Table 9.

For each level of α, 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 Theorem 1, where the conditional density and derivative of density are estimated by nonparametric approach. 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 $O(N^{-1})$ (which is the asymptotic variance) obtained by the first term in Theorem 2. The last column presents the the mean squared error $M\left(\widehat{\beta}\right)$ up to $O(N^{-2})$ derived in Theorem 2, where the conditional density and derivatives of density are estimated by nonparametric approach. There are literatures discussed ways to estimating the distribution function. A whole methodology known as kernel distribution function estimation (KDFE) has been explored since Nadaraya (1964). An improvement of this kind of method can be found in Sheather and Jones (1991).

From these results our findings are summarized as follows:

1. The magnitude of the second-order bias and MSE are larger towards the tails of the stock return distribution.

2. There are upward bias at lower quantiles and downward bias at upper quantiles.

3. The MSE up to $O(N^{-1})$ is smaller than the MSE up to $O(N^{-2})$. Hence, the statistical significance of the predictive ability of the predictor (dividend yield) can be over-stated if the (first-order) asymptotic variance is used.

8 Conclusions

This paper develops analytical results on the second-order bias and MSE of the quantile regression estimators. The results show that while the median is unbiased for a symmetric distribution, and the other quantiles are biased, with larger bias at deeper tails of any distribution. The higher order MSE gives further insights on how the efficiency of quantile estimators behave. The Monte Carlo simulation indicates the improvement of quantile estimation and quantile prediction by the second-order-bias-correction. The theoretical results are applied to the predictive quantile regression model for financial returns. We find that the quantile estimators with the second-order bias correction behave better than the uncorrected ones, and the bias is larger at extremely low and high quantiles of stock returns. It is shown that the second-order-bias-correction improves the accuracy of estimation and prediction of the conditional quantiles, especially in tails.

9 Appendix

9.1 Appendix A. Properties of a vector norm

Let A be a $k \times 1$ vector.

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

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

3. $||(AA') \otimes A'|| = \{\mathrm{tr}\,([(AA') \otimes A'] [(AA') \otimes A])\}^{1/2} = [\mathrm{tr}\,((AA'AA') \otimes (A'A))]^{1/2} = [\mathrm{tr}\,(A'AA'AA'A)]^{1/2}$
$= (A'A)^{3/2} = ||A||^3$.

4. $||(AA') \otimes A' \otimes A'|| = \mathrm{tr}([(AA') \otimes A' \otimes A'] [(AA') \otimes A \otimes A])^{1/2} = \mathrm{tr}[(AA'AA') \otimes (A' \otimes A') (A \otimes A)]^{1/2}$
$= \mathrm{tr}[(AA'AA') \otimes A'A \otimes A'A]^{1/2} = \mathrm{tr}[(A'AA'A)\, A'AA'A]^{1/2} = (A'AA'A) = (A'A)^2 = \|A\|^4$

9.2 Appendix B. Properties of the Dirac delta function

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 $z > 0$. The Dirac delta function $\delta(z)$ has the following properties.

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

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

3. $\int_{-\infty}^{+\infty} \delta^{(n)}(z-a)f(z)\mathrm{d}z = (-1)^n \int_{-\infty}^{+\infty} \delta(z-a)f^{(n)}(z)\mathrm{d}z = (-1)^n f^{(n)}(a)$, for $n = 1, 2, \ldots$.

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

5. $\phi(z)\delta(z) = \frac{1}{2}\delta(z)$. See Raju (1982).

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

9.3 Appendix C: Proofs

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

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

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

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

According to properties of the Dirac delta function in Appendix B, we have $\delta(\beta - y_i) = \delta(y_i - \beta)$ and

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

Thus, the moment condition can be written as

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

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

$$\Psi_N(\beta) = \frac{1}{N}\sum_{i=1}^{N} s_i(\beta). \tag{51}$$

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

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

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

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

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 $u_i = 0$, then the second-order bias of $\widehat{\beta}$ up to $O(N^{-1})$ of the unconditional quantile estimators $\widehat{\beta}$ can be written as

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

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

$$\begin{aligned}
M\left(\widehat{\beta}\right) &= \frac{1}{N}\alpha(1-\alpha)Q^2 - 2\frac{1}{N^2}\left(\alpha^2-\alpha+\frac{1}{2}\right)Q^3 f(\beta) + 2\frac{1}{N^2}\alpha(1-\alpha)Q^2 \\
&\quad -\frac{1}{N^2}\alpha(1-\alpha)(2\alpha-1)Q^4 f^{(1)}(\beta) + 6\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)^2 Q^4\left(f(\beta)\right)^2 \\
&\quad +3\frac{1}{N^2}\alpha(1-\alpha)Q^4\left(\left(\frac{1}{2}-\alpha\right)f^{(1)}(\beta) - (f(\beta))^2\right) - 12\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)\alpha(1-\alpha)Q^5 f^{(1)}(\beta)f(\beta) \\
&\quad +\frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^6\left(f^{(1)}(\beta)\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^5 f^{(2)}(\beta) \\
&= \frac{1}{N}\alpha(1-\alpha)Q^2 - 2\frac{1}{N^2}\left(\alpha^2-\alpha+\frac{1}{2}\right)Q^2 + 2\frac{1}{N^2}\alpha(1-\alpha)Q^2 - \frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2}-\alpha\right)Q^4 f^{(1)}(\beta) \\
&\quad +3\frac{1}{N^2}\left(3\alpha^2-3\alpha+\frac{1}{2}\right)Q^2 - 12\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)\alpha(1-\alpha)Q^4 f^{(1)}(\beta) \\
&\quad +\frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^6\left(f^{(1)}(\beta)\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^5 f^{(2)}(\beta) \\
&= \frac{1}{N}\alpha(1-\alpha)Q^2 + \frac{1}{N^2}\left(7\alpha^2-7\alpha-\frac{3}{2}\right)Q^2 - 7\frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2}-\alpha\right)Q^4 f^{(1)}(\beta) \\
&\quad +\frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^6\left(f^{(1)}(\beta)\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2 Q^5 f^{(2)}(\beta),
\end{aligned}$$

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. ■

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

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

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

$$\begin{aligned}
M\left(\widehat{\beta}\right) &= \frac{1}{N}\alpha(1-\alpha)Q^2p - 2\frac{1}{N^2}Q^3\left(\alpha^2-\alpha+\frac{1}{2}\right)pf(\beta) + 2\frac{1}{N^2}\alpha(1-\alpha)Q^2p \\
&\quad -\frac{1}{N^2}\alpha(1-\alpha)(2\alpha-1)Q^4p^2f^{(1)}(\beta) + 6\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)^2 Q^4\left(pf(\beta)\right)^2 \\
&\quad +3\frac{1}{N^2}\alpha(1-\alpha)Q^4\left(\left(\frac{1}{2}-\alpha\right)p^2f^{(1)}(\beta) - p^3\left(f(\beta)\right)^2\right) \\
&\quad -12\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)\alpha(1-\alpha)Q^5p^3f^{(1)}(\beta)f(\beta) \\
&\quad +\frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2Q^6p^4\left(f^{(1)}(\beta)\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2Q^5p^3f^{(2)}(\beta) \\
&= \frac{1}{N}\alpha(1-\alpha)Q^2p - 2\frac{1}{N^2}Q^2\left(\alpha^2-\alpha+\frac{1}{2}\right) + 2\frac{1}{N^2}\alpha(1-\alpha)Q^2p \\
&\quad -\frac{1}{N^2}\alpha(1-\alpha)(2\alpha-1)Q^4p^2f^{(1)}(\beta) + 6\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)^2 Q^2 \\
&\quad +3\frac{1}{N^2}\alpha(1-\alpha)Q^4\left(\frac{1}{2}-\alpha\right)p^2f^{(1)}(\beta) - 3\frac{1}{N^2}\alpha(1-\alpha)Q^2p \\
&\quad -12\frac{1}{N^2}\left(\frac{1}{2}-\alpha\right)\alpha(1-\alpha)Q^4p^2f^{(1)}(\beta) \\
&\quad +\frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2Q^6p^4\left(f^{(1)}(\beta)\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2Q^5p^3f^{(2)}(\beta) \\
&= \frac{1}{N}\alpha(1-\alpha)Q^2p - \frac{1}{N^2}\left(\alpha\left(1-\alpha\right)\left(4+p\right)+\frac{1}{2}\right)Q^2 - 7\frac{1}{N^2}\alpha(1-\alpha)\left(\frac{1}{2}-\alpha\right)Q^4p^2f^{(1)}(\beta) \\
&\quad +\frac{15}{4}\frac{1}{N^2}\alpha^2(1-\alpha)^2Q^6p^4\left(f^{(1)}(\beta)\right)^2 - \frac{1}{N^2}\alpha^2(1-\alpha)^2Q^5p^3f^{(2)}(\beta),
\end{aligned}$$

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 $f(0|x_i)$ 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. ■

Proof of Proposition 5: The moment condition is

$$\Psi_N(\beta) = \frac{1}{N}\sum_{i=1}^{N} s_i(\beta) \tag{52}$$

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

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

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

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

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

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

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

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

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

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

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

where $f(x_1'\beta)$ is the density of $y|x, z$, 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

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

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

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

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

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

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

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

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

where we have

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

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

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

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

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

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

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

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

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

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. ■

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.

Ang, A., Bekaert, G., 2007. Stock return predictability: Is it there? *Review of Financial Studies* 20: 651-707.

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.

Chen, X., Linton, O., van Keilegom, I., 2003. Estimation of semi-parametric models when the criterion function is not smooth. *Econometrica* 71: 1591-1608.

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.

David, H. A., Nagaraja, H. N., 2003. *Order Statistics*, 3rd Edition. Wiley, New York.

De Angelis, D., Hall, P., Young, G. A., 1993. Analytical and bootstrap approximations to estimator distributions in L1 regression. *Journal of the American Statistical Association* 88 (424), 1310-1316.

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

Fernandes, M., Guerre, E., & Horta, E. 2021. Smoothing quantile regressions. *Journal of Business & Economic Statistics*, 39(1), 338-357.

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

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.

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.

Horowitz, J. L., 1998. Bootstrap methods for median regression models. *Econometrica.* 66, 1327-1351.

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.

Kaplan, D. M., Sun, Y., 2017. Smoothed estimating equations for instrumental variables quantile regression. *Econometric Theory* 33 (1), 105-157.

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

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.

Lewellen, J. 2004. Predicting returns with financial ratios. *Journal of Financial Economics* 74: 209-235.

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

Nadaraya, E. A., 1964. Some new estimates for distribution functions. *Theory of Probability & Its Applications* 9 (3), 497-500.

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 hypothesis 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.

Otsu, T., 2008. Conditional empirical likelihood estimation and inference for quantile regression models. *Journal of Econometrics* 142 (1), 508-538.

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

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

Phillips, P.C.B., 1977. A general theorem in the theory of asymptotic expansion as approximations to finite sample distributions of econometric estimators. *Econometrica* 45: 1517-1534.

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

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

Portnoy, S., 2012. Nearly root-n approximation for regression quantile processes. *Annals of Statistics* 40 (3), 1714-1736.

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.

Sheather, S. J., Jones, M. C., 1991. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 53 (3), 683{690.

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

Whang, Y.-J., 2006. Smoothed empirical likelihood methods for quantile regression models. *Econometric Theory* 22 (2), 173-205.

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

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

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

	$\sigma_{ui} = 0.1x_i, N = 60$			$\sigma_{ui} = 0.5x_i, N = 60$		
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 $N = 60$, and the results are computed from 10,000 Monte Carlo replications.

Table 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 both 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 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 both 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 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 both 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 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 mixture normal distribution, u_i and x_i both 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 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 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 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.

Table 9: Application of second-order bias reduction to the 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 Theorem 2. 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 $O\left(N^{-1}\right)$. The last column presents the MSE up to $O\left(N^{-2}\right)$ derived in Theorem 2.

10 Supplemental Appendix

This appendix is *not* intended for publication and will be made available on the authors' webpage. It presents additional results that are not reported in the paper for space concern. Included in the Supplemental Appendix are:

1. the simulation results (Tables 10, 11, 12, 13) for DGP2 as described in the Monte Carlo section (Section 6), and

2. the empirical results (Tables 14, 15) to predictive quantile regression for forecast horizons $h = 3, 12$ as discussed in the application section (Section 7).

10.1 Additional Monte Carlo results for DGP2

In DGP2, 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 simulation results for DGP2 are presented in Tables 10, 11, 12, 13.

1. Tables 10-11 show the results with the range $R = 4, 10$, respectively. For these two tables, x_i is generated from exponential distribution, $f(x_i) = \exp(-x_i)$. These two tables are to be compared with Tables 2-3 in the paper. As we see from Tables 2-3 that the quantile regression estimator has larger bias when σ_u is larger ($\sigma_u = 0.5$ in Table 2 is larger than $\sigma_u = 0.1$ in Table 3), the same is observed from Tables 10-11 ($\sigma_u = \sqrt{\frac{R^2}{12}} = \sqrt{\frac{10^2}{12}} \approx 2.89$ in Table 11 is larger than $\sigma_u = \sqrt{\frac{4^2}{12}} \approx 1.15$ in Table 10).

2. Table 12 shows the results with $R = 4$ when x_i is generated from the Skewed Unimodal Density $\frac{1}{5}N(0,1) + \frac{1}{5}N\left(\frac{1}{2}, \left(\frac{2}{3}\right)^2\right) + \frac{3}{5}N\left(\frac{13}{12}, \left(\frac{5}{9}\right)^2\right)$, one of mixture normal distributions in Marron and Wand (1992). This table is to be compared with Table 4 in the paper.

3. Table 13 shows the results with $R = 4$ when x_i is generated from the Strongly Skewed Density $\sum_{l=0}^{7} \frac{1}{8}N\left(3\left[\left(\frac{2}{3}\right)^l - 1\right], \left(\frac{2}{3}\right)^{2l}\right)$, another mixture normal distribution in Marron and Wand (1992). This table is to be compared with Table 5 in the paper.

Table 10: 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	0.0384	-0.0066	0.0270	0.0226	-0.0044
	0.0063	0.0114	0.0100	0.0038	0.0054	0.0049
0.1	0.0400	0.0342	-0.0058	0.0240	0.0221	-0.0019
	0.0105	0.0156	0.0144	0.0066	0.0086	0.0081
0.2	0.0300	0.0256	-0.0044	0.0180	0.0176	-0.0004
	0.0173	0.0228	0.0222	0.0114	0.0132	0.0129
0.3	0.0200	0.0197	-0.0003	0.0120	0.0120	0.0000
	0.0222	0.0279	0.0275	0.0147	0.0164	0.0163
0.4	0.0100	0.0089	-0.0011	0.0060	0.0064	0.0004
	0.0251	0.0304	0.0303	0.0167	0.0188	0.0188
0.5	0.0000	-0.0008	-0.0008	0.0000	-0.0002	-0.0002
	0.0261	0.0310	0.0310	0.0174	0.0189	0.0189
0.6	-0.0100	-0.0090	-0.0010	-0.0060	-0.0048	0.0012
	0.0251	0.0303	0.0302	0.0167	0.0185	0.0185
0.7	-0.0200	-0.0174	0.0026	-0.0120	-0.0122	0.-0002
	0.0222	0.0273	0.0270	0.0147	0.0166	0.0164
0.8	-0.0300	-0.0266	0.0034	-0.0180	-0.0171	0.0009
	0.0173	0.0223	0.0216	0.0114	0.0138	0.0135
0.9	-0.0400	-0.0341	0.0059	-0.0240	-0.0224	0.0016
	0.0105	0.0156	0.0145	0.0066	0.0088	0.0083
0.95	-0.0450	-0.0364	0.0086	-0.0270	-0.0245	0.0025
	0.0063	0.0107	0.0094	0.0038	0.0058	0.0053

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 both 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 11: 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	0.0967	-0.0158	0.0675	0.0583	-0.0092
	0.0394	0.0712	0.0621	0.0237	0.0342	0.0309
0.1	0.1000	0.0873	-0.0127	0.0600	0.0574	-0.0026
	0.0654	0.0989	0.0915	0.0416	0.0544	0.0512
0.2	0.0750	0.0680	-0.0070	0.0450	0.0426	-0.0024
	0.1082	0.1426	0.1380	0.0710	0.0847	0.0829
0.3	0.0500	0.0477	-0.0023	0.0300	0.0290	-0.0010
	0.1388	0.1766	0.1743	0.0920	0.1018	0.1009
0.4	0.0250	0.0192	-0.0058	0.0150	0.0164	0.0014
	0.1571	0.1882	0.1879	0.1046	0.1115	0.1112
0.5	0.0000	-0.0003	-0.0003	0.0000	-0.0051	-0.0051
	0.1632	0.1960	0.1960	0.1088	0.1208	0.1208
0.6	-0.0250	-0.0314	-0.0064	-0.0150	-0.0083	0.0067
	0.1571	0.1886	0.1877	0.1046	0.1155	0.1155
0.7	-0.0500	-0.0464	0.0036	-0.0300	-0.0324	-0.0024
	0.1388	0.1721	0.1700	0.0920	0.1040	0.1030
0.8	-0.0750	-0.0671	0.0079	-0.0450	-0.0369	0.0081
	0.1082	0.1445	0.1401	0.0710	0.0818	0.0805
0.9	-0.1000	-0.0903	0.0097	-0.0600	-0.0550	0.0050
	0.0654	0.0984	0.0904	0.0416	0.0544	0.0514
0.95	-0.1125	-0.0947	0.0178	-0.0675	-0.0605	0.0070
	0.0394	0.0695	0.0608	0.0237	0.0356	0.0319

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 both 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 12: 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	0.0303	0.0276	-0.0027	0.0181	0.0159	-0.0022
	0.0088	0.0159	0.0152	0.0056	0.0081	0.0078
0.1	0.0269	0.0246	-0.0023	0.0161	0.0151	-0.0010
	0.0173	0.0228	0.0222	0.0109	0.0132	0.0130
0.2	0.0202	0.0196	-0.0006	0.0121	0.0095	-0.0025
	0.0314	0.0388	0.0384	0.0197	0.0219	0.0218
0.3	0.0135	0.0100	-0.0034	0.0080	0.0072	-0.0009
	0.0414	0.0479	0.0478	0.0259	0.0292	0.0292
0.4	0.0067	0.0060	-0.0007	0.0040	0.0030	-0.0010
	0.0474	0.0547	0.0547	0.0296	0.0344	0.0344
0.5	0.0000	-0.0014	-0.0014	0.0000	0.0011	0.0011
	0.0494	0.0569	0.0569	0.0309	0.0349	0.0349
0.6	-0.0067	-0.0033	0.0034	-0.0040	-0.0016	0.0024
	0.0474	0.0537	0.0537	0.0296	0.0326	0.0326
0.7	-0.0135	-0.0121	0.0014	-0.0080	-0.0085	-0.0005
	0.0414	0.0478	0.0476	0.0259	0.0286	0.0286
0.8	-0.0202	-0.0225	-0.0023	-0.0121	-0.0095	0.0025
	0.0314	0.0379	0.0374	0.0197	0.0220	0.0219
0.9	-0.0269	-0.0249	0.0020	-0.0161	-0.0145	0.0016
	0.0173	0.0234	0.0228	0.0109	0.0135	0.0133
0.95	-0.0303	-0.0258	0.0045	-0.0181	-0.0173	0.0008
	0.0088	0.0155	0.0148	0.0056	0.0080	0.0078

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 both 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 13: 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	-0.0153	-0.0128	0.0025	-0.0092	-0.0081	0.0011
	0.0029	0.0062	0.0061	0.0017	0.0031	0.0030
0.1	-0.0136	-0.0110	0.0026	-0.0082	-0.0085	-0.0003
	0.0050	0.0091	0.0090	0.0030	0.0050	0.0049
0.2	-0.0102	-0.0105	-0.0002	-0.0062	-0.0053	0.0009
	0.0085	0.0156	0.0155	0.0052	0.0087	0.0087
0.3	-0.0068	-0.0051	0.0017	-0.0041	-0.0027	0.0014
	0.0110	0.0192	0.0192	0.0068	0.0116	0.0116
0.4	-0.0034	-0.0035	-0.0001	-0.0021	-0.0008	0.0012
	0.0125	0.0222	0.0222	0.0078	0.0133	0.0133
0.5	0.0000	0.0001	0.0001	0.0000	-0.0009	-0.0009
	0.0130	0.0226	0.0226	0.0081	0.0138	0.0138
0.6	0.0034	0.0026	-0.0008	0.0021	0.0015	-0.0005
	0.0125	0.0222	0.0222	0.0078	0.0131	0.0131
0.7	0.0068	0.0067	-0.0001	0.0041	0.0024	-0.0017
	0.0110	0.0194	0.0194	0.0068	0.0116	0.0116
0.8	0.0102	0.0087	-0.0015	0.0062	0.0064	0.0002
	0.0085	0.0151	0.0150	0.0052	0.0088	0.0087
0.9	0.0136	0.0130	-0.0006	0.0082	0.0070	-0.0012
	0.0050	0.0090	0.0089	0.0030	0.0052	0.0051
0.95	0.0153	0.0139	-0.0014	0.0092	0.0082	-0.0010
	0.0029	0.0059	0.0057	0.0017	0.0031	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 both 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.

10.2 Additional results in empirical application of predictive quantile regressions for $h = 3, 12$

In Section 7 for the empirical application on the predictive quantile regression for stock returns, we present the results only for forecast horizon $h = 1$ (one month ahead) to save space in the paper. See Table 9. Here in this appendix, we present additional results for forecast horizon $h = 3$ (three months ahead) in Table 14 and for $h = 12$ (12 months ahead) in Table 15.

Table 14: 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: 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 Theorem 2. 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 $O\left(N^{-1}\right)$. The last column presents the MSE up to $O\left(N^{-2}\right)$ derived in Theorem 2.

Table 15: 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: 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 Theorem 2. 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 $O\left(N^{-1}\right)$. The last column presents the MSE up to $O\left(N^{-2}\right)$ derived in Theorem 2.