The Second-order Bias of Quantile Estimators*

Tae-Hwy Lee†, Aman Ullah‡, and He Wang§

First version: March 2018
Revised 1: June 2018
Revised 2: August 2018
This version: September 2018

Abstract

The finite sample theory using higher-order asymptotics provides better approximations of the bias for a class of estimators. Phillips (1991) demonstrated the higher-order asymptotic expansions for LAD estimators. Rilstone, Srivastava and Ullah (1996) provided the second-order bias results of conditional mean regression estimators. This paper develops new analytical results on the second-order bias of the conditional quantile regression estimators, which enables an improved bias correction and thus to obtain improved quantile estimation. In particular, we show that the second-order bias is larger towards the tails of the conditional density than near the median, and therefore the benefit of the second-order bias correction is greater when we are interested in the deeper tail quantiles, e.g., for the study of income distribution and financial risk management. The Monte Carlo simulation confirms the theory that the bias is larger at the tail quantiles, and the second-order bias correction improves the behavior of the quantile estimators.

Key Words: Delta function, Quantile regression, Second-order bias.
JEL Classification: C1, C2, C13

*We thank an anonymous referee for many useful comments and suggestions.
†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 Economics, University of California, Riverside, CA 92521. E-mail: he.wang001@email.ucr.edu

1 Introduction

The finite sample properties have been almost entirely developing for the mean regression models.[1] Rilstone, Srivastava and Ullah (RSU, 1996) developed the second-order bias of a class of nonlinear estimators in models with i.i.d. samples. Bao and Ullah (2007) analyzed the RSU results for time series dependent observations. On the other hand, there is little finite sample results in the quantile regression although there is extensive literature on the first-order asymptotic results, see Koenker and Bassett (1978), and Koenker (2005). The literature on higher-order distributional properties focused on the order of the remainder term of the expansion of $\sqrt{N}\left(\widehat{\beta}-\beta\right)$, that is often referred to as the second-order asymptotic (distributional) representations (SOADR), see Bahadur (1966), Kiefer (1967), and Jureckova and Sen (1996). Unlike these literature, our goal is to derive the explicit expression of the second-order bias up to $O(N^{-1})$, rather than only studying the asymptotic distribution of the remainder term. Portnoy (2012) provided an alternative approximation expansion for the quantile process with the remainder bound of nearly $\sqrt{N}$, beyond that provided by the Bahadur representation, and made the bias up to $O(N^{-1})$.

The challenge to study the high-order asymptotic properties of quantile estimators is due to the non-differentiability of the objective function for the quantile estimation. Horowitz (1998) smoothed the objective function to deal with the non-differentiability. Alternatively, Komunjer (2005) and and Elliott, Komunjer, and Timmermann (2005) focused on a family of conditional quantile models with the 'smooth' objective functions that are continuously differentiable.[2] Instead of smoothing an objective function (Horowitz 1998) or of using a smooth objective function (Komunjer 2005), Phillips (1991) overcame the non-differentiability of least absolute deviation (LAD) regression by using the *generalized function* (or Dirac delta function).[3] In this paper, we follow Phillips (1991) and Chernozhukov, Fernandez-Val, and Galichon (2007), with noting that Chernozhukov, Fernandez-

[1] We refer to the higher-order asymptotic properties as the finite sample properties. The finite sample properties in this paper is not the exact moment or distributional properties. See Ullah (2004).

[2] Komunjer (2005, page 147) states, "The non-differentiability problem has prompted several authors to develop asymptotic normality results under a weaker set of assumptions, generally requiring that $\nabla_{\theta} L_T(\theta)$ exist with probability one. Examples include: Daniels (1961), Huber (1967), Pollard (1985), Newey and McFadden (1994). In this paper, we focus on conditional quantile models that are continuously differentiable on Θ (A0), so that the log-likelihood function $L_T(\theta)$ is continuously differentiable a.s.-P_0 on Θ."

[3] Phillips (1991, p. 451) states, "If the criterion function has nonregularities like discontinuities in its derivatives, these may be accommodated directly by the use of generalized functions, provided the discontinuities are smoothed out asymptotically."

Val, and Galichon (2007, Appendix C) state that it is an informal justification in using the Dirac delta function.[4]

We develop the second-order bias of quantile estimators using the Dirac delta function. We discover that while the median is unbiased for symmetric error distributions and the bias of the other quantiles is larger at the tails of any distribution. The Monte Carlo simulations results present that the second-order bias corrected estimator has better behavior than the uncorrected ones.

The paper is organized as follows. In Section 2, we present 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 quantile estimators. In Section 4, we present Monte Carlo simulations. Section 5 concludes.

Notation: The notation used in the paper is summarized here. $f_{y|x}(\cdot)$ denotes the density of y conditional on x, $f^{(j)}_{y|x}(\cdot)$ denotes the jth-order derivative of $f_{y|x}(\cdot)$, $f_u(\cdot)$ denotes the density of u, and $f^{(j)}_u(\cdot)$ denotes the jth-order derivative of $f_u(\cdot)$. The jth-order partial derivative of a matrix $A(\beta)$ is defined as $\nabla^j_\beta A(\beta)$. If $A(\beta)$ and β are both $k \times 1$ vectors, then $\nabla^j_\beta A(\beta)$ 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, then $\|A\| = (A'A)^{1/2}$. The Kronecker product is defined in the usual way. For an $m \times n$ matrix A and a $p \times q$ matrix B, we have $A \otimes B$ as an $mp \times nq$ matrix. $\overline{X}$ denotes the expectation $E(X)$ of a random vector X.

Generalized function: Let $\phi(z) = \mathbf{1}\,(z \geq 0)$ is a step function. The delta function is defined as $\delta(z) = \mathrm{d}\phi(z)/\mathrm{d}z$. See Gelfand and Shilov (1964, p. 4). Denote the derivatives of delta function by $\delta^{(j)}(\cdot)$ for $j = 1, 2, \ldots$. The properties of the delta function are critical in this paper, which are summarized here: (i) $\delta(-z) = \delta(z)$, (ii) $\delta^{(1)}(-z) = -\delta^{(1)}(z)$, (iii) $\delta^{(2)}(-z) = \delta^{(2)}(z)$, (iv) $\int_{-\infty}^{+\infty} \delta(z-a)f(z)\mathrm{d}z = f(a)$, and (v) $\int_{-\infty}^{+\infty} \delta^{(n)}(z-a)f(z)\mathrm{d}z = (-1)^n \int_{-\infty}^{+\infty} \delta(z-a)f^{(n)}(z)\mathrm{d}z = (-1)^n f^{(n)}(a)$, where $f : \mathbb{R} \to \mathbb{R}$ is a real function differentiable around $a \in \mathbb{R}$. See Gelfand and Shilov (1964, pp. 4, 5, 26). More properties of the delta function used in this paper are: (vi) $\phi(z)\delta(z) = \frac{1}{2}\delta(z)$.

[4] We thank a referee for this point and the reference.

2 Conditional Quantile Estimators

Consider a random variable y from distribution $F(\cdot)$. Given $\alpha \in (0,1)$, consider a general linear regression quantile model, $q_\alpha = x'\beta_\alpha$, where q_α is the conditional α-quantile of y, the quantile estimators β_α vary across α. Then the location-scale version of the linear regression quantile model is $y_i = x_i'\beta_\alpha + u_i$, where y_i is a scalar, x_i is a $k \times 1$ vector, and u_i is the error defined as the difference between y_i and its conditional α-quantile. For simplicity, we set x_i and u_i to be i.i.d. in this paper.[5] To simplify the notation, we use β to denote β_α hereafter. The $k \times 1$ vector quantile estimators $\widehat{\beta}$ can be obtained by solving $\min_\beta E[L_\alpha(\beta)] = E[(\alpha - \mathbf{1}(y_i < x_i'\beta))\,(y_i - x_i'\beta)]$. The moment condition can be written as $\nabla^1_\beta E[L_\alpha(\beta)] = E[(\alpha - \mathbf{1}(y_i < x_i'\beta))(-x_i)] = E[s_i(x_i, \beta)]$, where the score function $s_i(x_i,\beta) \equiv s_i(\beta) = (\alpha - \mathbf{1}(y_i < x_i'\beta))(-x_i)$ 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. The sample moment condition can be written as

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

An estimator $\widehat{\beta}$ is a solution to a set of moment equations of the form

$$\Psi_N(\widehat{\beta}) = \frac{1}{N}\sum_{i=1}^{N} s_i(\widehat{\beta}) = 0. \tag{2}$$

Equation (2) is the first-order conditions for the quantile estimator $\widehat{\beta}$, which is the analogous to equation (4) in Phillips (1991, p. 452) for the LAD estimator.[6]

RSU (1996) developed the second-order bias of a class of nonlinear estimators in models with i.i.d. samples. Assumptions in RSU (1996) are sufficient to obtain the stochastic expansion of $\widehat{\beta}$. Now we give the modified high-level Assumptions A-C for quantile models as follows and some remarks are made.

Assumption A. The jth-order derivative of $s_i(\beta)$ exists in a neighborhood of β_0, and $E\left[||x_i||^{(j+1)}\, f_u^{(j-1)}(0)\right]^2 < \infty$, for $j = 1, 2$, with $f_u^{(0)}(0) = f_u(0)$.

[5] In mean regression, Bao and Ullah (2007) show that the RSU results continue to hold for non-i.i.d cases. The same may be the case in quantile regression, which we leave it to our future work.

[6] We note that the empirical moment equation for quantile regression may not be exactly zero but $\Psi_N(\widehat{\beta}) = o_p\left(N^{-1/2}\right)$ as discussed in Angrist, Chernozhukov, and Fernandez-Val (2006, Appendix). We thank a referee for this reference. This may affect the order of the remainder term of $\widehat{\beta} - \beta_0$ for quantile regression as in Bahadur (1966). See equation (8). Nevertheless, we will show in the next section that this would not affect the second-order bias as long as equation (9) holds.

RSU (1996) assumes that the jth-order derivative of score function $s_i(\beta)$ exists in a neighborhood of β_0, and $E\left\|\nabla_\beta^j s_i(\beta_0)\right\|^2 < \infty$, for $j \geq 1$. We modify this assumption for quantile models. To derive the second-order bias of $\widehat{\beta}$, we require $j = 1, 2$. Noting that $\mathbf{1}(y_i - x_i'\beta < 0) = \mathbf{1}(x_i'\beta - y_i \geq 0) \equiv \phi(x_i'\beta - y_i)$ and $\delta(z) = \mathrm{d}\phi(z)/\mathrm{d}z$, the first 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, $\nabla_\beta^1 s_i(\beta) = \nabla_\beta^1[(\alpha - \mathbf{1}(y_i < x_i'\beta))(-x_i)] = x_i x_i'\delta(x_i'\beta - y_i)$, and

$$E\left\|\nabla_\beta^1 s_i(\beta_0)\right\|^2 = E\left[\|x_i\|^2 f_{y|x}(x_i'\beta_0)\right]^2 < \infty. \tag{3}$$

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

$$E\left\|\nabla_\beta^2 s_i(\beta_0)\right\|^2 = E\left[\|x_i\|^3 f_{y|x}^{(1)}(x_i'\beta_0)\right]^2 < \infty. \tag{4}$$

Since the conditional density of y_i given x_i evaluated at $y_i = x_i'\beta_0$ is the same as the conditional density of u_i given x_i evaluated at $u_i = 0$, and since u_i and x_i are independent, we have $f_{y|x}(x_i'\beta_0) = f_u(0)$. Then the above boundedness conditions on the derivatives can be rewritten as shown in Assumption A.

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

Note that $\left(E\nabla_\beta^1\Psi_N(\beta)\right)^{-1} = \left(E(x_i x_i') f_u(0)\right)^{-1} = O(1)$ from (3). Under Assumption B, we will be able to rewrite (12) as (13) to obtain the second-order bias in the next section.

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

Assumption C(i) gives the modified Lipschitz condition for a quantile model. To derive the second-order bias of the quantile estimators, we use the high-order Taylor expansion of $\Psi_N(\beta)$ 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. Assumption C requires not only the stochastic equicontinuity to handle the nonsmooth objective function but also the higher-order stochastic equicontinuity to handle nonsmooth derivatives of the objective function. This problem has been discussed by Newey and McFadden (1994, Theorem 7.3), Horowitz (1998), Komunjer (2005), and Elliott, Komunjer, and Timmermann (2005). The basic insight is that smoothness of a function can be replaced by the smoothness of its limit if the remainder term is small enough. Therefore, the stochastic equicontinuity conditions do not require differentiability of the criterion function but require that the remainder term of the expansion can be controlled in a particular way over a neighborhood of β_0. Assumption C(ii) gives the weak law of large numbers condition. This condition is stated and discussed in Phillips (1991, pp. 453-455) and it requires that the right-hand-side, $E\left[\nabla_\beta^j s_i(\beta_0)\right]$ for $j \geq 1$, be bounded, which we verified in the discussion of Assumption A above.

3 Second-order Bias of Quantile Estimators

To obtain the second-order bias for quantile estimators which is to be summarized in Theorem 1 below, let us begin with taking the Taylor's expansion of $\Psi_N(\widehat{\beta}) = 0$ around β_0,

$$0 = \Psi_N + \nabla\Psi_N \left(\widehat{\beta} - \beta_0\right) + \frac{1}{2}\nabla^2\Psi_N \left[(\widehat{\beta} - \beta_0) \otimes (\widehat{\beta} - \beta_0)\right] + o_p\left(N^{-1}\right), \tag{5}$$

where $\Psi_N = \Psi_N(\beta_0)$. The ordinary stochastic expansion of $\widehat{\beta}$ can be obtained from equation (5). However, a difficulty arises from the derivatives of the moment condition (1). Using the properties of the delta function summarized earlier at the end of Section 1 or in Phillips (1991, p. 455), it can be shown that $\nabla\Psi_N \xrightarrow{p} \overline{\nabla\Psi_N}$, i.e., $\frac{1}{N}\sum_{i=1}^N x_i x_i' \delta(x_i'\beta - y_i) \xrightarrow{p} E\left(x_i x_i'\right) f_u(0)$. See Gelfand and Shilov (1964, p. 26). Then, similar to Phillips (1991), we rewrite (5) as

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

To see the order of each of these terms, we recall the asymptotic distribution of the quantile estimators when x_i and u_i are i.i.d.

$$\sqrt{N}(\widehat{\beta} - \beta_0) \xrightarrow{d} N\left(0, \frac{\alpha(1-\alpha)}{[f_u(0)]^2}\left[E\left(x_i x_i'\right)\right]^{-1}\right). \tag{7}$$

See, e.g., Koenker (2005), and also Phillips (1991) for the LAD estimator with $\alpha = 0.5$. As this textbook result states that the quantile estimator $\widehat{\beta}$ is $\sqrt{N}$-consistent estimator, using the same argument in Phillips (1991, p. 455), we can obtain that the orders of both $A_1 = \Psi_N$ and

$A_2 = \overline{\nabla\Psi_N}(\widehat{\beta} - \beta_0)$ are $O_p\left(N^{-1/2}\right)$. In the following Lemma 1 and Lemma 2, we discuss the orders of A_3 and A_4.

Before doing that, it is important to recall the following result in this literature. Let $\widehat{\beta} - \beta_0 = a_{-1/2} + R_N$, where $a_{-1/2}$ is a random sequence of $O_p\left(N^{-1/2}\right)$ with zero mean $E\left(a_{-1/2}\right) = 0$ and R_N is the remainder term of higher order. Bahadur (1966) and Kiefer (1967) established the celebrated results on the order of R_N, that is

$$R_N = O_p\left(N^{-3/4}\left(\log\log N\right)^{3/4}\right). \tag{8}$$

See Koenker (2005, pp. 122-123), and also Jureckova and Sen (1996, pp. 196-202), and van der Vaart (1998 p. 310). Note that (8) implies that

$$R_N = O_p\left(N^{-3/4+\varepsilon}\right) \text{ for some small } \varepsilon > 0. \tag{9}$$

Below we use this result to obtain Lemma 1(b). Our goal is to obtain the expression of the bias term $E\left(R_N\right) = E\left(\widehat{\beta} - \beta_0\right)$ *up to* the second-order i.e., of order $O\left(N^{-c}\right)$ with $c \leq 1$. We first state five lemmas whose proofs are made available in supplemental appendix.

Lemma 1. Let

$$\begin{aligned} A_3 &= \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)(\widehat{\beta} - \beta_0) \\ &= \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right) a_{-1/2} + \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right)\left[(\widehat{\beta} - \beta_0) - a_{-1/2}\right] \\ &\equiv A_{31} + A_{32}. \end{aligned} \tag{10}$$

Then, (a) $A_{31} = O_p(N^{-7/6})$, and (b) A_{32} is smaller than $O_p(N^{-1})$. □

Lemma 2. Let

$$\begin{aligned} A_4 &= \frac{1}{2}\nabla^2\Psi_N \left[(\widehat{\beta} - \beta_0) \otimes (\widehat{\beta} - \beta_0)\right] \\ &= \frac{1}{2}\overline{\nabla^2\Psi_N}\left[(\widehat{\beta} - \beta_0) \otimes (\widehat{\beta} - \beta_0)\right] + \frac{1}{2}\left(\nabla^2\Psi_N - \overline{\nabla^2\Psi_N}\right)\left[(\widehat{\beta} - \beta_0) \otimes (\widehat{\beta} - \beta_0)\right] \\ &\equiv A_{41} + A_{42}, \end{aligned} \tag{11}$$

Then, (a) $A_{41} = O_p(N^{-1})$, and (b) A_{42} is smaller than $O_p(N^{-1})$. □

Given Lemmas 1-2, we can now rewrite equation (6) as

$$\begin{aligned} 0 &= A_1 + A_2 + A_{31} + A_{41} + o_p\left(N^{-1}\right) \quad (12) \\ &= \Psi_N + \overline{\nabla\Psi_N}(\widehat{\beta} - \beta_0) + \left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right) a_{-1/2} + \frac{1}{2}\overline{\nabla^2\Psi_N}\left(a_{-1/2} \otimes a_{-1/2}\right) + o_p\left(N^{-1}\right). \end{aligned}$$

In equation (12), it is important to note that we keep the term A_{31} even though it is $O_p(N^{-7/6})$, because we find that the "expectation" of A_{31} becomes $O(N^{-1})$ so that $E(A_{31})$ is a part of the second-order bias, as we will show shortly.

Solve for $\widehat{\beta} - \beta_0$ in equation (12) to obtain

$$\begin{aligned} \widehat{\beta} - \beta_0 &= -\overline{\nabla\Psi_N}^{-1}\Psi_N - \overline{\nabla\Psi_N}^{-1}\left(\nabla\Psi_N - \overline{\nabla\Psi_N}\right) a_{-1/2} - \frac{1}{2}\overline{\nabla\Psi_N}^{-1}\overline{\nabla^2\Psi_N}\left(a_{-1/2} \otimes a_{-1/2}\right) + o_p\left(N^{-1}\right) \\ &= -Q\Psi_N - QV_N a_{-1/2} - \frac{1}{2}Q\overline{H_2}\left(a_{-1/2} \otimes a_{-1/2}\right) + o_p\left(N^{-1}\right) \quad (13) \\ &\equiv B_1 + B_2 + B_3 + o_p\left(N^{-1}\right), \end{aligned}$$

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 (13) by $\sqrt{N}$ gives the same as equation (15) of Phillips (1991, p. 457). In order to compute the bias of $\widehat{\beta}$, that is $E\left(\widehat{\beta} - \beta_0\right)$, we now examine the expectations of the three terms B_1, B_2, B_3 in (13). Lemma 3 shows that $E(B_1)$ is the first-order bias which is zero, while Lemmas 4 and 5 show the second-order bias $E(B_2 + B_3)$.

Lemma 3. Let $B_1 = -Q\Psi_N$. Then, (a) $B_1 = O_p(N^{-1/2})$ and (b) $E(B_1) = 0$. □

Lemma 4. Let

$$B_2 = -QV_N a_{-1/2} = Q\left(H_1 - \overline{H_1}\right)Q\Psi_N = QH_1Q\Psi_N - Q\overline{H_1}Q\Psi_N \equiv B_{21} + B_{22}. \quad (14)$$

Then, (a) $B_{21} = O_p(N^{-7/6})$, (b) $E(B_{21}) = O(N^{-1})$, (c) $E(B_{22}) = 0$, and (d) $E(B_2) = O(N^{-1})$.□

Lemma 5. Let $B_3 = -\frac{1}{2}Q\overline{H_2}\left(a_{-1/2} \otimes a_{-1/2}\right)$. Then, (a) $B_3 = O_p(N^{-1})$ and (b) $E(B_3) = O(N^{-1})$. □

Given Lemmas 3-5, and from equation (13), the bias of quantile estimators $\widehat{\beta}$ is

$$\begin{aligned} E\left(\widehat{\beta} - \beta_0\right) &= E(-Q\Psi_N) + E(QH_1Q\Psi_N) + E\left(-\frac{1}{2}Q\overline{H_2}\left(a_{-1/2} \otimes a_{-1/2}\right)\right) + o\left(N^{-1}\right) \quad (15) \\ &= E(B_1) + E(B_{21}) + E(B_3) + o\left(N^{-1}\right) \\ &\equiv B\left(\widehat{\beta}\right) + o\left(N^{-1}\right), \end{aligned}$$

where $B\left(\widehat{\beta}\right)$ is the "second-order bias" of quantile estimators $\widehat{\beta}$ *up to* $O(N^{-1})$. We now summarize the above as a theorem:

Theorem 1. *Suppose Assumptions A, B, C hold. In the quantile regression model, suppose* x_i and u_i *are i.i.d., the second-order bias of the quantile estimators* $\widehat{\beta}$ *up to* $O(N^{-1})$ *is*

$$B\left(\widehat{\beta}\right)=\frac{1}{N}Q\left[\left(\frac{1}{2}-\alpha\right)E\left(x_ix_i'Qx_i\right)f_u(0)-\frac{\alpha(1-\alpha)}{2}E[\left(x_ix_i'\right)\otimes x_i']f_u^{(1)}(0)\left(Q\otimes Q\right)E\left(x_i\otimes x_i\right)\right]. \tag{16}$$

Remark: One important point is whether an expansion of the bias may be useful for inference. The classical first-order asymptotic result for regression quantiles in (7) shows that the bias tends to zero. The results described in Theorem 1 for the second-order bias may be used together with the second-order asymptotic variance for the second-order asymptotic inference. We will report the second-order mean-squared errors (MSE) comparable with the second-order bias, so that we can conduct the second-order asymptotic inference. The fact that the bias tends to be larger in the tails will make the second-order asymptotic inference more useful and it would be interesting to compare with the first-order results in (7). We thank a referee for pointing this out.

4 Monte Carlo Simulation

We present simulation results for the second-order bias that was derived in Section 3. In the quantile regression model $y_i=x_i'\beta+u_i$, the error term u_i satisfies $E\left[\alpha-\mathbf{1}\left(y_i<x_i'\beta\right)|x_i\right]=0$. The α conditional quantile of u_i given x_i is zero. The error term u_i is normally distributed with the CDF $F(\cdot)$ with standard deviation σ_u, then the mean equals to $-\Phi^{-1}(\alpha)\sigma_u$, with $\Phi\left(\cdot\right)$ denoting the standard normal CDF. Therefore, we generate the error term u_i following normal distribution $N\left(-\Phi^{-1}(\alpha)\sigma_u,\sigma_u^2\right)$. We generate x_i from an exponential distribution with its density being $\exp\left(-x\right)$. Finally, y_i is generated from $y_i=x_i'\beta+u_i$. In this setup, $k=1$, $\beta=0$, $\sigma_u=0.5$, $N=100$. We use the Matlab package by Roger Koenker to estimate the model. We repeat this 10,000 times.

For each level of α, the first column in Table 1 presents the Monte Carlo average values of $\widehat{\beta}$ from 10,000 simulations. The second column presents the second-order bias $B\left(\widehat{\beta}\right)$ derived in Theorem 1. The third column presents the second-order bias-corrected quantile estimators $\widetilde{\beta}=\widehat{\beta}-B\left(\widehat{\beta}\right)$.

The Monte Carlo results are summarized as follows: (i) $\widetilde{\beta}$ is numerically closer to the true value $\beta = 0$ than $\widehat{\beta}$, as the bias in $\widehat{\beta}$ has been substantially corrected; (ii) the magnitude of bias is larger at lower and upper quantiles; (iii) the bias is zero at the median for symmetric errors; and (iv) there are upward bias at lower quantiles and downward bias at upper quantiles. The benefit of the second-order bias correction is substantial especially towards the tails.

Table 1: Second-order bias correction with x generated from the exponential distribution

α	$\hat{\beta}$	$B(\hat{\beta})$	$\tilde{\beta}$
0.01	0.0210	0.0163	0.0047
0.05	0.0060	0.0052	0.0009
0.10	0.0032	0.0031	0.0002
0.20	0.0023	0.0016	0.0007
0.30	0.0017	0.0009	0.0008
0.40	0.0010	0.0004	0.0006
0.50	-0.0002	0.0000	-0.0002
0.60	0.0006	-0.0004	0.0010
0.70	-0.0002	-0.0009	0.0007
0.80	-0.0021	-0.0016	-0.0005
0.90	-0.0040	-0.0031	-0.0009
0.95	-0.0063	-0.0052	-0.0011
0.99	-0.0220	-0.0163	-0.0057

Notes: For each level of α, the first column presents the quantile estimators $\hat{\beta}$. The second column presents the second-order bias $B(\hat{\beta})$ derived in Theorem 1. The third column presents the second-order bias corrected quantile estimators $\tilde{\beta} = \hat{\beta} - B(\hat{\beta})$.

5 Conclusions

This paper derives the second-order bias of conditional quantile estimators, which enables an improved bias correction and thus improved quantile estimation. 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.

References

Angist, J., Chernozhukov, V., Fernandez-Val, I., 2006. Quantile regression under misspecification with an application to the US wage structure. *Econometrica 74*, 539–563.

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.

Chernozhukov, V., Fernandez-Val, I., Galichon, A., 2007. "Rearranging estimator of the value-at-risk and other measures," MIT.

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

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

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

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

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

Koenker, R., Bassett, G., 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.

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., 2113-2247, Elsevier Science, Amsterdam.

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

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

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.

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

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

6 Supplemental Appendix

Due to the page limitation of *Economics Letters,* the following items will be made available at this supplemental appendix from authors' website:

1. the proofs of Lemmas 1-5, and
2. additional Monte Carlo results in Section 4 when x follows a mixture normal distribution of Marron and Wand (1992).

6.1 Proof of Lemmas 1-5

Proof of Lemma 1: (a) According to Phillips (1991, p. 457) and Kim and Pollard (1990), the term $V_N \equiv \nabla\Psi_N - \overline{\nabla\Psi_N}$ is $O_p\left(N^{-1/3}\right)$. Scaling equation (10) by $\sqrt{N}$, we obtain that $\sqrt{N}a_{-1/2}$ is bounded and normally distributed with zero mean. Note that $\sqrt{N}A_{31}$ will contribute to $\sqrt{N}A_3$ through the variance of $\sqrt{N}a_{-1/2}$, that is $\sqrt{N}A_{31} = O_p\left(N^{-2/3}\right)$. See Phillips (1991, p. 457). Therefore, $A_{31} = O_p\left(N^{-7/6}\right)$. (b) Since R_N is not of zero mean, because $E\left(R_N\right)$ is the high-order bias of quantile estimators. Using (9), $A_{32} = V_N R_N = O_p\left(N^{-1/3-3/4+\varepsilon}\right)$ is smaller than $O_p(N^{-1})$, i.e. $A_{32} = o_p\left(N^{-1}\right)$. ■

Proof of Lemma 2: (a) Rewrite A_4 by adding and subtracting, where $\nabla^2\Psi_N \xrightarrow{p} \overline{\nabla^2\Psi_N}$, that is $\frac{1}{N}\sum_{i=1}^N\left(x_i x_i'\right)\otimes x_i'\delta^{(1)}(x_i'\beta - y_i) \xrightarrow{p} E\left[\left(x_i x_i'\right)\otimes x_i'\right] f_u^{(1)}(0)$. Then, A_{41} can be rewritten as

$$\begin{aligned} A_{41} &= \frac{1}{2}\overline{\nabla^2\Psi_N}\left\{\left[(\widehat{\beta}-\beta_0) - a_{-1/2} + a_{-1/2}\right]\otimes\left[(\widehat{\beta}-\beta_0) - a_{-1/2} + a_{-1/2}\right]\right\} \\ &= \frac{1}{2}\overline{\nabla^2\Psi_N}\left(a_{-1/2}\otimes a_{-1/2}\right) \\ &\quad +\frac{1}{2}\overline{\nabla^2\Psi_N}\left(a_{-1/2}\otimes\left[(\widehat{\beta}-\beta_0) - a_{-1/2}\right]\right) \\ &\quad +\frac{1}{2}\overline{\nabla^2\Psi_N}\left(\left[(\widehat{\beta}-\beta_0) - a_{-1/2}\right]\otimes a_{-1/2}\right) \\ &\quad +\frac{1}{2}\overline{\nabla^2\Psi_N}\left(\left[(\widehat{\beta}-\beta_0) - a_{-1/2}\right]\otimes\left[(\widehat{\beta}-\beta_0) - a_{-1/2}\right]\right), \end{aligned} \tag{17}$$

where only the first term in equation (17), $\frac{1}{2}\overline{\nabla^2\Psi_N}\left(a_{-1/2}\otimes a_{-1/2}\right)$, is $O_p\left(N^{-1}\right)$, and the other three terms in equation (17) are of order smaller than $O_p\left(N^{-1}\right)$. (b) Since $\left(\nabla^2\Psi_N - \overline{\nabla^2\Psi_N}\right)$ is of order smaller than $O_p(1)$, A_{42} is of order smaller than $O_p(N^{-1})$, i.e. $A_{42} = o_p\left(N^{-1}\right)$. ■

Proof of Lemma 3: (a) Since only the term B_1 in equation (13) is $O_p\left(N^{-1/2}\right)$, it should be that $a_{-1/2} = B_1$. (b) Since Ψ_N is the sample moment condition, $E\left(B_1\right) = -QE\left(\Psi_N\right) = 0$. ■

Proof of Lemma 4: (a) By Lemma 1, $A_{31} = O_p(N^{-7/6})$. Then $B_{21} = -QA_{31} = O_p(N^{-7/6})$, because Q is bounded. (b) We will show that the *expectation* of B_{21} is of order $O\left(N^{-1}\right)$ even if B_{21} is of order $O_p\left(N^{-7/6}\right)$. To examine $E\left(B_{21}\right)$, we obtain the following results

$$H_1 = \nabla_\beta^1 \Psi_N = \frac{1}{N}\sum_{i=1}^{N} x_i x_i' \delta(x_i'\beta - y_i), \tag{18}$$

$$\overline{H_1} = E\left[\frac{1}{N}\sum_{i=1}^{N} x_i x_i' \delta(x_i'\beta - y_i)\right] = E\left(x_i x_i'\right) f_u(0), \tag{19}$$

$$Q = \overline{H_1}^{-1} = \left[E\left(x_i x_i'\right) f_u(0)\right]^{-1}, \tag{20}$$

where Ψ_N is a $k \times 1$ vector, H_1, $\overline{H_1}$, and Q are $k \times k$ matrixes. Recalling the properties of the delta function summarized in Section 1, we have

$$\begin{aligned}
E\left(B_{21}\right) &= E\left(QH_1Q\Psi_N\right) \qquad (21)\\
&= QE\left[\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\delta(x_i'\beta - y_i)\right) Q \left(\frac{1}{N}\sum_{j=1}^{N} s_j\right)\right]\\
&= Q\frac{1}{N^2}\sum_{i=1}^{N} E\left[x_i x_i'\delta(x_i'\beta - y_i)Qs_i\right]\\
&= Q\frac{1}{N}E\left[x_i x_i' E\left(\delta(x_i'\beta - y_i)Qs_i|x_i\right)\right]\\
&= \frac{1}{N}QE\left[x_i x_i' \int_{-\infty}^{+\infty} \delta(x_i'\beta - y_i)Q(\alpha - \mathbf{1}(y_i < x_i'\beta))(-x_i) f_{y|x}(y_i)\mathrm{d}y_i\right]\\
&= \frac{1}{N}QE\left[-x_i x_i' Q x_i \alpha \int_{-\infty}^{+\infty} \delta(x_i'\beta - y_i) f_{y|x}(y_i)\mathrm{d}y_i\right]\\
&\quad + \frac{1}{N}QE\left[x_i x_i' Q x_i \int_{-\infty}^{+\infty} \delta(x_i'\beta - y_i)\phi(x_i'\beta - y_i) f_{y|x}(y_i)\mathrm{d}y_i\right]\\
&= \frac{1}{N}QE\left[-x_i x_i' Q x_i \alpha f_{y|x}(x_i'\beta) + \frac{1}{2}x_i x_i' Q x_i f_{y|x}(x_i'\beta)\right]\\
&= \left(\frac{1}{2} - \alpha\right)\frac{1}{N}QE\left(x_i x_i' Q x_i\right) f_u(0)\\
&= O\left(N^{-1}\right).
\end{aligned}$$

Therefore, the cube-root asymptotic behavior in B_{21} arising from $V_N = H_1 - \overline{H_1} = O_p\left(N^{-1/3}\right)$ (see 14) disappears in $E\left(B_{21}\right)$. (c) $E\left(B_{22}\right) = -E\left(Q\overline{H_1}Q\Psi_N\right) = -QE\left(\Psi_N\right) = 0$ since $Q = \overline{H_1}^{-1}$. (d) $E\left(B_2\right) = E\left(B_{21} + B_{22}\right) = E\left(B_{21}\right)$ is $O\left(N^{-1}\right)$. Hence, only $E\left(B_{21}\right)$ is a part of the second-order bias. ■

Proof of Lemma 5: (a) By Lemma 2, $A_{41} = O_p(N^{-1})$. Thus $B_3 = -QA_{41} = O_p(N^{-1})$, because Q is bounded. (b) To examine $E(B_3)$, we obtain the following results

$$H_2 = \nabla^2_\beta \Psi_N = \frac{1}{N}\sum_{i=1}^{N}\left(x_i x_i'\right)\otimes x_i'\delta^{(1)}(x_i'\beta - y_i), \tag{22}$$

$$\overline{H_2} = E\left[\frac{1}{N}\sum_{i=1}^{N}\left(x_i x_i'\right)\otimes x_i'\delta^{(1)}(x_i'\beta - y_i)\right] = E\left[\left(x_i x_i'\right)\otimes x_i'\right] f_u^{(1)}(0), \tag{23}$$

where H_2 and $\overline{H_2}$ are $k \times k^2$ matrixes. Then,

$$\begin{aligned}
E(B_3) &= E\left(-\frac{1}{2}Q\overline{H_2}\left(a_{-1/2}\otimes a_{-1/2}\right)\right) \qquad (24)\\
&= -\frac{1}{2}Q\overline{H_2}E\left[(Q\Psi_N)\otimes(Q\Psi_N)\right]\\
&= -\frac{1}{2}Q\overline{H_2}\left(Q\otimes Q\right)E\left[\left(\frac{1}{N}\sum_{i=1}^{N}s_i\right)\otimes\left(\frac{1}{N}\sum_{j=1}^{N}s_j\right)\right]\\
&= -\frac{1}{2}Q\overline{H_2}\left(Q\otimes Q\right)\frac{1}{N^2}E\left[\sum_{i=1}^{N}\left(x_i\otimes x_i\right)\left(\alpha - \mathbf{1}(y_i < x_i'\beta)\right)^2\right]\\
&= -\frac{1}{2}Q\overline{H_2}\left(Q\otimes Q\right)\frac{1}{N}E\left[\left(x_i\otimes x_i\right)E\left((\alpha - \mathbf{1}(y_i < x_i'\beta))^2|x_i\right)\right]\\
&= -\frac{1}{2}Q\overline{H_2}\left(Q\otimes Q\right)\frac{1}{N}\alpha(1-\alpha)E\left(x_i\otimes x_i\right)\\
&= -\frac{1}{N}\frac{\alpha(1-\alpha)}{2}QE[\left(x_i x_i'\right)\otimes x_i']f_u^{(1)}(0)\left(Q\otimes Q\right)E\left(x_i\otimes x_i\right)\\
&= O\left(N^{-1}\right).
\end{aligned}$$

■

6.2 Monte Carlo Results when x is generated from a mixture of normal

In addition to the result reported in Table 1 for which we generate x_i from an exponential distribution with its density being $\exp(-x)$, we also generate x_i from a mixture-normal distribution $\frac{1}{5}N(0,1) + \frac{1}{5}N\left(\frac{1}{2},(\frac{2}{3})^2\right) + \frac{3}{5}N\left(\frac{13}{12},(\frac{5}{9})^2\right)$ which is a skewed unimodal density considered in Marron and Wand (1992). The result is reported here in Table 2. The findings are similar whether x follows the exponential distribution (Table 1) or the Marron-Wand's mixture-normal distribution (Table 2). In both cases, the benefits of the second-order bias correction are very substantial especially towards the tails.

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

Table 2: Second-order bias correction with x generated from a Marron-Wand mixture normal distribution

α	$\hat{\beta}$	$B(\hat{\beta})$	$\tilde{\beta}$
0.01	0.0089	0.0110	-0.0020
0.05	0.0027	0.0035	-0.0007
0.10	0.0014	0.0021	-0.0006
0.20	0.0017	0.0011	0.0006
0.30	0.0000	0.0006	-0.0006
0.40	0.0010	0.0003	0.0007
0.50	-0.0006	0.0000	-0.0006
0.60	-0.0002	-0.0003	0.0001
0.70	-0.0004	-0.0006	0.0002
0.80	-0.0016	-0.0011	-0.0005
0.90	-0.0018	-0.0021	0.0002
0.95	-0.0029	-0.0035	0.0006
0.99	-0.0090	-0.0110	0.0020

Notes: For each level of α, the first column presents the quantile estimators $\hat{\beta}$. The second column presents the second-order bias $B(\hat{\beta})$ derived in Theorem 1. The third column presents the second-order bias corrected quantile estimators $\tilde{\beta} = \hat{\beta} - B(\hat{\beta})$.