

The Second-order Asymptotic Properties of Asymmetric Least Squares Estimation*

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

First version: May 2018
This version: December 2018

Abstract

The higher-order asymptotic properties provide better approximation of the bias for a class of estimators. The first-order asymptotic properties of the asymmetric least squares (ALS) estimator have been investigated by Newey and Powell (1987). This paper develops the second-order asymptotic properties (bias and mean squared error) of the ALS estimator, extending the second-order asymptotic results for the symmetric least squares (LS) estimators of Rilstone, Srivastava and Ullah (1996). The LS gives the mean regression function while the ALS gives the “expectile” regression function, a generalization of the usual regression function. The second-order bias result enables an improved bias correction and thus an improved ALS estimation in finite sample. In particular, we show that the second-order bias is much larger as the asymmetry is stronger, and therefore the benefit of the second-order bias correction is greater when we are interested in extreme expectiles which are used as a risk measure in financial economics. The higher-order MSE result for the ALS estimation also enables us to better understand the sources of estimation uncertainty. The Monte Carlo simulation confirms the benefits of the second-order asymptotic theory and indicates that the second-order bias is larger at the extreme low and high expectiles.

Key Words: asymmetric least squares, expectile, delta function, second-order bias, Monte Carlo.
JEL Classification: C13, C33, C52.

*We are thankful to the editors and two anonymous referees for many valuable comments that have helped improving the paper.

[†]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: hwang041@ucr.edu

1 Introduction

The higher-order asymptotic properties permit us to obtain better approximation of the bias of estimators, and allow us to find an approach to improve the behavior of estimators and test statistics. 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 results for time series dependent observations. In this paper, we extend the second-order asymptotic results for the symmetric least squares (LS) estimators to asymmetric least squares (ALS) estimators.

The ALS estimation was first interpreted as a maximum likelihood estimator when the disturbances arise from a normal distribution with unequal weight placed on positive and negative disturbances by Aigner, Amemiya and Poirier (1976). Newey and Powell (1987) proposed the term, ALS, and investigated the estimation and hypothesis tests for coefficients of linear ALS models. The symmetric LS gives the mean regression function while the ALS gives the "expectile" regression function, a generalization of the usual regression function. The ALS model has been used in many economic applications. A lot of recent research in financial economics uses the large sample theory to study the properties of ALS models in financial risk management. Kuan, Yeh and Hsu (2009), proposed an expectile based value-at-risk and extended asymptotic results to allow for stationary and weakly dependent data using a parametric method. Xie, Zhou and Wan (2014) developed a nonparametric varying-coefficient approach for modeling the expectile-based value-at-risk. However, the literature on the ALS model has been entirely the first-order asymptotic properties. The first-order asymptotic properties of the ALS model can be improved by considering the higher-order asymptotic approximations which are better approximations. In this paper, we try to fill this unexplored area by developing the analytical results of the second-order bias and mean squared error (MSE) for the ALS models. We show that the second-order bias is much larger as the asymmetry is stronger, and therefore the benefit of the second-order bias correction is greater when we are interested in extreme expectiles. The higher-order MSE result for the ALS estimation enables us to better understand the sources of estimation uncertainty. The Monte Carlo simulations results present that the second-order bias corrected estimator has better behavior than the uncorrected one.

The paper is organized as follows. In Section 2, we review Newey and Powell (1987) to introduce the ALS estimator, and present the moment condition of the ALS regression and the assumptions used in this paper. In Section 3, we derive the second-order bias and MSE of the conditional ALS regression estimators. In Section 4, a special case of the ALS regression model without a covariate is considered, which gives the unconditional ALS estimator. In Section 5, we present Monte Carlo simulations. Section 6 concludes.

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

2 Asymmetric Least Squares Estimation

2.1 Loss Functions

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

$$y_i = x_i' \beta + u_i, \quad (1)$$

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

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

$$Q_N(\beta; \theta) = \sum_{i=1}^N r_{\theta}(y_i - x_i' \beta), \quad (2)$$

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

$$r_{\theta}(\lambda) \equiv |\theta - \mathbf{1}(\lambda < 0)| \cdot |\lambda|. \quad (3)$$

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

$$R_N(\beta; \tau) = \sum_{i=1}^N \rho_{\tau}(y_i - x_i' \beta), \quad (4)$$

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

$$\rho_\tau(\lambda) \equiv |\tau - \mathbf{1}(\lambda < 0)| \cdot \lambda^2. \quad (5)$$

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

$$\widehat{\beta}(\tau) = \left[\sum_{i=1}^N \left| \tau - \mathbf{1}(y_i < x_i' \widehat{\beta}(\tau)) \right| x_i x_i' \right]^{-1} \sum_{i=1}^N \left| \tau - \mathbf{1}(y_i < x_i' \widehat{\beta}(\tau)) \right| x_i y_i. \quad (6)$$

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

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

$$\min_{\beta} E[\rho_\tau(y_i - x_i' \beta)] = E \left[\left| \tau - \mathbf{1}(y_i < x_i' \beta) \right| \cdot (y_i - x_i' \beta)^2 \right]. \quad (7)$$

Equation (7) reduces to the standard least squares objective function when $\tau = 0.5$.

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

the population moment condition is

$$\begin{aligned}
& \nabla_{\beta}^1 E [\rho_{\tau}(y_i - x'_i \beta)] \\
&= E [\nabla_{\beta}^1 \rho_{\tau}(y_i - x'_i \beta)] \\
&= E [\nabla_{\beta}^1 |\tau - \mathbf{1}(y_i < x'_i \beta)| \cdot (y_i - x'_i \beta)^2] + 2E[|\tau - \mathbf{1}(y_i < x'_i \beta)|(y_i - x'_i \beta)(-x_i)]. \quad (8)
\end{aligned}$$

By the definition of the delta function in Appendix (subsection 7.2, Property B.1), we have $\mathbf{1}(y_i - x'_i \beta < 0) = \mathbf{1}(x'_i \beta - y_i \geq 0) = \phi(x'_i \beta - y_i)$. See Gelfand and Shilov (1964). Then

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

The first term of the Equation (8) can be written as $E[x'_i \delta(x'_i \beta - y_i)(y_i - x'_i \beta)^2]$, which equals zero, because according to Properties B.3 and B.4 of the Dirac delta function (see subsection 7.2), we have

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

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

$$\begin{aligned}
\nabla_{\beta}^1 E [\rho_{\tau}(y_i - x'_i \beta)] &= 2E[|\tau - \mathbf{1}(y_i < x'_i \beta)|(y_i - x'_i \beta)(-x_i)] \\
&\equiv E [s_i(\beta)], \quad (9)
\end{aligned}$$

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

To get rid of the absolute value in (9), we first rewrite the score function as

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

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

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

Thus, the score function can be rewritten as

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

The sample moment condition for (9) is denoted as

$$\Psi_N(\beta) = \frac{1}{N} \sum_{i=1}^N s_i(\beta).^1 \quad (10)$$

2.2 Assumptions

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

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

Assumption B. For some neighborhood of β_0 , $(\nabla \Psi_N(\beta))^{-1} = O_p(1)$.

¹Expressing this score function in matrix notation, $s(\beta) = 2(2\tau - 1)X'u^- - 2\tau X'u$, where $X' = (x_1, \dots, x_N)$, $u = (u_1, \dots, u_N)'$, and $u^- = (u_1 \mathbf{1}(u_1 < 0), \dots, u_N \mathbf{1}(u_N < 0))'$, it may be possible to rewrite the bias expression in the next section. However, due to some difficulty in taking derivatives of the score and using the properties of the delta functions in matrix form, we leave this direction for future endeavor.

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

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

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

$$\begin{aligned} \nabla_{\beta}^1 s_i(\beta) &= \nabla_{\beta}^1 [2x_i (y_i - x'_i\beta) [(2\tau - 1) \mathbf{1}(y_i < x'_i\beta) - \tau]] \\ &= -2x_i x'_i [(2\tau - 1) \mathbf{1}(y_i < x'_i\beta) - \tau] + 2(2\tau - 1) x_i x'_i (y_i - x'_i\beta) \delta(x'_i\beta - y_i) \\ &= -2(2\tau - 1) x_i x'_i \mathbf{1}(y_i < x'_i\beta) + 2\tau x_i x'_i + 2(2\tau - 1) x_i x'_i (y_i - x'_i\beta) \delta(x'_i\beta - y_i). \end{aligned}$$

Using Properties A.2, B.3 and B.4 in Appendix, we obtain

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Next, we discuss Assumption B. For ALS models, Assumption B requires $\text{p} \lim_{N \rightarrow \infty} (\nabla_{\beta}^1 \Psi_N(\beta))^{-1} = \left(\lim_{N \rightarrow \infty} E \nabla_{\beta}^1 \Psi_N(\beta) \right)^{-1} = \left(\lim_{N \rightarrow \infty} 4\tau(1 - \tau) E(x_i x'_i) \right)^{-1} = O(1)$, that implies $E(x_i x'_i)$ is nonsingular, which is the same as Assumption 4 of Newey and Powell (1987).

3 Second-order Bias and MSE of the ALS Estimators

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

3.1 Bias

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

$$\begin{aligned}
B\left(\widehat{\beta}(\tau)\right) &= \frac{1}{N^2} \sum_{i=1}^N 4Q \left\{ (2\tau - 1) E \left[x_i x_i' Q x_i u_i \mathbf{1}(u_i < 0) \right] - \tau^2 E \left[x_i x_i' Q x_i u_i \right] \right\} \\
&\quad + \frac{1}{N} \sum_{i=1}^N (2\tau - 1) Q E \left[(x_i x_i') \otimes x_i' f_i(0|x_i) \right] \\
&\quad \times \frac{1}{N^2} \sum_{i=1}^N 4(Q \otimes Q) \left\{ \begin{array}{l} -(2\tau - 1) E \left[(x_i \otimes x_i) u_i^2 \mathbf{1}(u_i < 0) \right] \\ + \tau^2 E \left[(x_i \otimes x_i) u_i^2 \right] \end{array} \right\}, \quad (11)
\end{aligned}$$

where $Q = \left(4\tau(1 - \tau) \frac{1}{N} \sum_{i=1}^N E \left[x_i x_i' \right] \right)^{-1}$, and $f_i(0|x_i)$ is the conditional density of u_i given x_i evaluated at $u_i = 0$. \square

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

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

We have

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

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

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

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

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

$$\overline{H}_1 = E \nabla_\beta^1 \Psi_N = 4\tau(1 - \tau) \frac{1}{N} \sum_{i=1}^N E(x_i x'_i),$$

$$\overline{H}_2 = E \nabla_\beta^2 \Psi_N = -2(2\tau - 1) \frac{1}{N} \sum_{i=1}^N E[(x_i x'_i) \otimes x'_i f_i(x'_i \beta)],$$

$$\overline{H}_3 = E \nabla_\beta^3 \Psi_N = -2(2\tau - 1) \frac{1}{N} \sum_{i=1}^N E[(x_i x'_i) \otimes x'_i \otimes x'_i f_i^{(1)}(x'_i \beta)],$$

$$Q = (\overline{H}_1)^{-1} = \left(4\tau(1 - \tau) \frac{1}{N} \sum_{i=1}^N E[x_i x'_i] \right)^{-1},$$

$$V = H_1 - \overline{H}_1,$$

$$W = H_2 - \overline{H}_2,$$

and

$$d = Q \Psi_N,$$

where Ψ_N , s_i and d are all $k \times 1$ vectors. H_1 , \overline{H}_1 , Q , and V are all $k \times k$ matrixes, H_2 , \overline{H}_2 and W

are all $k \times k^2$ matrixes. H_3 and $\overline{H_3}$ are $k \times k^3$ matrixes. Using Properties B.8 in Appendix,

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

and

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

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

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

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

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

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

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

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

Remark 1.1. When x_i and u_i are i.i.d., and $k = 1$, we observe that $x_i, \Psi_N, s_i, d, H_1, \overline{H}_1, Q, V, H_2, \overline{H}_2, W, H_3, \overline{H}_3$ are all scalars, and the second-order bias of $\widehat{\beta}(\tau)$ up to $O(N^{-1})$ can be rewritten as

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

Remark 1.2. The second-order bias of $\widehat{\beta}(\tau)$ goes to zero as the sample size goes to infinity. The second-order bias of $\widehat{\beta}(\tau)$ is larger at the extreme expectiles of a distribution, because (i) as τ goes up toward 1 or down toward 0, Q also goes up; and (ii) $(2\tau-1)/[\tau(1-\tau)]$ is monotonic in τ , and equals to zero at $\tau = 0.5$, then as τ goes down to 0.5 from 1 or up to 0.5 from 0, in equation (11), $\frac{1}{N^2}\sum_{i=1}^N 4Q\{(2\tau-1)E[x_i x_i' Q x_i u_i \mathbf{1}(u_i < 0)]\}$ goes to zero, and $\frac{1}{N}\sum_{i=1}^N (2\tau-1)QE[(x_i x_i') \otimes x_i' f_i(0|x_i)] \times \frac{1}{N^2}\sum_{i=1}^N 4(Q \otimes Q)\left\{\begin{array}{l} -(2\tau-1)E[(x_i \otimes x_i) u_i^2 \mathbf{1}(u_i < 0)] \\ + \tau^2 E[(x_i \otimes x_i) u_i^2] \end{array}\right\}$

also goes to zero, therefore, the only term left is $\frac{1}{N^2} \sum_{i=1}^N 4Q \{-\tau^2 E[x_i x_i' Q x_i u_i]\}$, Therefore, the second-order bias of $\widehat{\beta}(\tau)$ is larger at the extreme expectiles of a distribution.

Remark 1.3. The objective function of ALS model reduces to the standard least-squares objective function when $\tau = 0.5$. In this case, the second-order bias of $\widehat{\beta}(\tau)$ up to $O(N^{-1})$ equals the second-order bias of the OLS estimator. The OLS estimator is unbiased because $E(u_i|x_i) = 0$.

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

3.2 MSE

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

$$\begin{aligned}
M(\widehat{\beta}(\tau)) &= \frac{1}{N} 4Q^2 C_1 - \frac{1}{N^2} 16Q^3 C_3 + \frac{1}{N^2} 8Q^4 C_1 - \frac{1}{N^2} Q^4 16(2\tau - 1) E(x_i^3) f(0) C_4 \\
&+ \frac{1}{N^2} 96Q^4 C_2^2 + \frac{1}{N^2} 48Q^4 \left\{ \tau(1 - \tau) E(x_i^4) - 4\tau^2(1 - \tau)^2 [E(x_i^2)]^2 \right\} C_1 \\
&+ \frac{1}{N^2} 384Q^5 (2\tau - 1) E(x_i^3) f(0) C_1 C_2 + \frac{1}{N^2} 240Q^6 (2\tau - 1)^2 [E(x_i^3)]^2 [f(0)]^2 C_1^2 \\
&+ \frac{1}{N^2} 32Q^5 (2\tau - 1) E(x_i^4) f^{(1)}(0) C_1^2, \tag{12}
\end{aligned}$$

where

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

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

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

$$M(\widehat{\beta}) = \frac{1}{N}A_1 + \frac{1}{N^2}(A_2 + A'_2) + \frac{1}{N^2}(A_3 + A_4 + A'_4)$$

where $\frac{1}{N}A_1 = E(a_{-1/2}a'_{-1/2})$, $\frac{1}{N^2}(A_2 + A'_2) = E(a_{-1}a'_{-1/2} + a_{-1/2}a'_{-1})$, $\frac{1}{N^2}(A_3 + A_4 + A'_4) = E(a_{-1}a'_{-1} + a_{-3/2}a'_{-1/2} + a_{-1/2}a'_{-3/2})$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

where we have

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

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

We also observe

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

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

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

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

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

Since the conditional density of y_i given x_i evaluated at $y_i = x'_i\beta$ is the same as the conditional

density of u_i given x_i evaluated at $u_i = 0$. Then the MSE up to $O(N^{-2})$ can be written as

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

This is as stated in Theorem 2. ■

Remark 2. The asymptotic variance of the expectile estimator $\widehat{\beta}(\tau)$ equals the first-order term of $M\left(\widehat{\beta}(\tau)\right)$ in (12) in Theorem 2. Newey and Powell (1987) derived the first-order asymptotic distribution of the ALS estimator as follows

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

where $w_i(\tau) = |\tau - \mathbf{1}(u_i < 0)|$, $W = E\left[w_i(\tau)x_i^2\right]$, $V = E\left[w_i^2 x_i^2 u_i^2\right]$. It can be shown that the asymptotic variance, $\frac{1}{N}W^{-1}VW^{-1}$, equals to $\frac{1}{N}4Q^2 C_1$, which is the first-order term in (12).

4 Special Case: Unconditional ALS Model

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

$$y_i = \beta + u_i, \quad (13)$$

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

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

where the asymmetric least squares loss function is

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

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

4.1 Bias

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

$$\begin{aligned} B\left(\hat{\beta}(\tau)\right) &= \frac{1}{N^2} \sum_{i=1}^N 4Q^2 \left\{ (2\tau - 1) E[u_i \mathbf{1}(u_i < 0)] - \tau^2 E(u_i) \right\} \\ &\quad + \frac{1}{N} \sum_{i=1}^N 4(2\tau - 1) Q^3 f_i(0) \\ &\quad \times \frac{1}{N^2} \sum_{i=1}^N \left\{ -(2\tau - 1) E[u_i^2 \mathbf{1}(u_i < 0)] + \tau^2 E(u_i^2) \right\}, \end{aligned} \quad (14)$$

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

Proof: Consider the linear ALS regression model $y_i = \beta + u_i$, where y_i is a scalar, u_i is the error defined to be the difference between y_i and its τ -expectile β , we call $\hat{\beta}(\tau)$ as the unconditional

ALS estimator. Given the asymmetric least squares loss function, the ALS estimators $\widehat{\beta}(\tau)$ can be obtained by solving

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

Then the population moment condition is

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

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

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

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

According to Property B.3 of the Dirac delta function, we have

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

Thus, the moment condition can be written as

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

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

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

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

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

Then, the score function can be written as

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

Therefore, the sample moment condition can be written as

$$\Psi_N(\beta) = \frac{1}{N} \sum_{i=1}^N s_i(\beta), \quad (15)$$

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

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

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

where

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

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

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

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

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

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

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

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

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

and

$$d = Q\Psi_N.$$

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

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

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

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

written as

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

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

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

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

\square

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

Remark 3.1. The second-order bias of $\widehat{\beta}(\tau)$ goes to zero as the sample size goes to infinity. The second-order bias of $\widehat{\beta}(\tau)$ is larger at the extreme expectiles of a distribution, because (i) as τ goes up toward 1 or down toward 0, Q also goes up; and (ii) $(2\tau - 1) / [\tau(1 - \tau)]$ is monotonic in τ , and equals to zero at $\tau = 0.5$, then as τ goes down to 0.5 from 1 or up to 0.5 from 0, in equation (14), $\frac{1}{N^2} \sum_{i=1}^N 4Q^2 \{(2\tau - 1) E[u_i \mathbf{1}(u_i < 0)]\}$ goes to zero, and $\frac{1}{N} \sum_{i=1}^N 4(2\tau - 1) Q^3 f_i(0) \times \frac{1}{N^2} \sum_{i=1}^N \{-(2\tau - 1) E[u_i^2 \mathbf{1}(u_i < 0)] + \tau^2 E(u_i^2)\}$ also goes to zero, therefore, the only term left is $\frac{1}{N^2} \sum_{i=1}^N 4Q^2 \{-\tau^2 E(u_i)\}$. Therefore the second-order bias of $\widehat{\beta}(\tau)$ is larger at the extreme extreme expectiles of a distribution.

Remark 3.2. The objective function of the ALS model reduces to the standard least-squares objective function when $\tau = 0.5$. In this case, the second-order bias of $\widehat{\beta}(\tau)$ up to $O(N^{-1})$ equals the second-order bias of the OLS estimator. The OLS estimator is unbiased because $E(u_i) = 0$.

4.2 MSE

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

$$\begin{aligned}
M\left(\widehat{\beta}(\tau)\right) &= \frac{1}{N}4Q^2C_1 - \frac{1}{N^2}16Q^3C_3 + \frac{1}{N^2}8Q^4C_1 - \frac{1}{N^2}Q^416(2\tau - 1)f(0)C_4 \\
&+ \frac{1}{N^2}96Q^4C_2^2 + \frac{1}{N^2}48Q^4\left\{\tau(1 - \tau) - 4\tau^2(1 - \tau)^2\right\}C_1 \\
&+ \frac{1}{N^2}384Q^5(2\tau - 1)f(0)C_1C_2 + \frac{1}{N^2}240Q^6(2\tau - 1)^2[f(0)]^2C_1^2 \\
&+ \frac{1}{N^2}32Q^5(2\tau - 1)f^{(1)}(0)C_1^2.
\end{aligned} \tag{16}$$

where

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

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

Proof: By Theorem 2, when $x_i = 1$, the MSE of the unconditional ALS estimator $\widehat{\beta}(\tau)$ up to

$O(N^{-2})$ is

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

This is as stated in Theorem 4. ■

5 Monte Carlo Simulation

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

$$\tau \int_{\mu(\tau)}^{\infty} (y - \mu(\tau)) dF(y) + (\tau - 1) \int_{-\infty}^{\mu(\tau)} (\mu(\tau) - y) dF(y) = 0, \quad (17)$$

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

$$\frac{\tau}{1-\tau} = \frac{\int_{-\infty}^{\mu(\tau)} (\mu(\tau) - y) dF(y)}{\int_{\mu(\tau)}^{\infty} (y - \mu(\tau)) dF(y)}. \quad (18)$$

If we set the true β to be zero, then y_i have the same distribution as u_i . To generate u_i from uniform distribution on $[a, b]$ and $\mu(\tau) = 0$, we have $\int_{-\infty}^{\mu(\tau)} (\mu(\tau) - y) dF(y) = \int_{-\infty}^0 (-y) dF(y) =$

$\int_a^0 (-y) \frac{1}{b-a} dy = \frac{a^2}{2(b-a)}$, and $\int_{\mu(\tau)}^{\infty} (y - \mu(\tau)) dF(y) = \int_0^{\infty} y dF(y) = \int_0^b y \frac{1}{b-a} dy = \frac{b^2}{2(b-a)}$. Then we can get the relationship between a and b , i.e. $a = -\sqrt{\frac{\tau}{1-\tau}} b$. In the DGP, we generate the error term u_i from uniform distribution on $[a, b]$, where $a = -\frac{\sqrt{\frac{\tau}{1-\tau}} R}{1 + \sqrt{\frac{\tau}{1-\tau}}}$, $b = \frac{R}{1 + \sqrt{\frac{\tau}{1-\tau}}}$, and the range $R = b - a$. For example, $R = 4, \tau = 0.1$, implies that u_i is generated from $U[-1, 3]$, the mean of u_i is 1, variance of u_i is $\frac{4}{3}$, and $\mu(0.1) = 0$. The DGP of u_i guarantees that the 0.1 conditional expectile of u_i given x_i is zero. In addition, we can verify the relationship of quantile and expectile, that is if u_i follows $U[-1, 3]$, then $\mu(0.1) = q(0.25) = 0$. We simulate x_i from exponential distribution, $f(x_i) = \exp(-x_i)$. Then, y_i is simulated from $y_i = x_i' \beta + u_i$. In this setup, $k = 1, \beta = 0, R = 4, N \in \{100, 300\}$. In this DGP, with knowledge of the distribution of $\{x_i, u_i\}$, we know the various expectations in the theorems. See Appendix in subsection 7.3, where we explain how those expectations are computed in our Monte Carlo.

Following Newey and Powell (1987) and Kuan et al. (2009), we use the iterated weighted least squares algorithm to compute the ALS estimator in equation (6). We use the OLS estimates as the initial value of $\hat{\beta}$ for the iterated weighted least squares estimates and iterate until the estimates converge. The convergence was quick and did not depend on the choice of initial value of $\hat{\beta}$. We repeat the Monte Carlo simulations 10,000 times and take the average.

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

6 Conclusions

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

7 Appendix

7.1 Properties of a norm

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

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

A.2

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

A.3

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

A.4

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

7.2 Properties of the Dirac delta function

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

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

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

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

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

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

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

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

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

7.3 Calculation in Monte Carlo Simulation

In the Monte Carlo simulation in Section 5, with knowledge of the distribution of $\{x_i, u_i\}$ under the data generating process (DGP), we know the various expectations in the theorems.

The DGP of u_i guarantees that the τ -conditional expectile of u_i given x_i is zero. We generate the error term u_i from the uniform distribution $[a, b]$, where $a = -\frac{\sqrt{\frac{\tau}{1-\tau}}R}{1+\sqrt{\frac{\tau}{1-\tau}}}$, $b = \frac{R}{1+\sqrt{\frac{\tau}{1-\tau}}}$, and we set the range $R = b - a = 4$, then for each τ , we are able to calculate

$$\begin{aligned} E(u_i) &= (a+b)/2 = \left(1 - \sqrt{\frac{\tau}{1-\tau}}\right) b/2, \\ E(u_i^2) &= [E(u_i)]^2 + V(u_i) = \left(1 - \sqrt{\frac{\tau}{1-\tau}}\right)^2 b^2/4 + 4/3, \\ E[u_i \mathbf{1}(u_i < 0)] &= \int_a^0 \frac{1}{b-a} u_i du_i = -\frac{a^2}{2(b-a)}, \\ E[u_i^2 \mathbf{1}(u_i < 0)] &= \int_a^0 \frac{1}{b-a} u_i^2 du_i = -\frac{a^3}{3(b-a)}, \end{aligned}$$

and $f(0) = \frac{1}{b-a}$, which is the density of u_i evaluated at $u_i = 0$.

For the simulation results in Table 1, we generate x_i from the exponential distribution with the density $f(x_i) = \exp(-x_i)$. Thus we have $E(x_i^2) = 2$, $E(x_i^3) = 6$, and $E(x_i^4) = 24$. We simulate y_i from $y_i = x_i' \beta + u_i$, and set $k = 1$, $\beta = 0$, $R = 4$. Since in this setup x_i and u_i are both i.i.d., for each τ , we have

$$Q = \left(4\tau(1-\tau) \frac{1}{N} \sum_{i=1}^N E[x_i x_i']\right)^{-1} = [4\tau(1-\tau) E(x_i^2)]^{-1} = [8\tau(1-\tau)]^{-1},$$

$$\begin{aligned}
\frac{1}{N^2} \sum_{i=1}^N E [x_i x_i' Q x_i u_i \mathbf{1}(u_i < 0)] &= \frac{1}{N} Q E(x_i^3) E[u_i \mathbf{1}(u_i < 0)] = -\frac{1}{2N} Q \frac{a^2}{(b-a)}, \\
\frac{1}{N^2} \sum_{i=1}^N E [x_i x_i' Q x_i u_i] &= \frac{1}{N} Q E(x_i^3) E(u_i) = \frac{3}{N} Q \left(1 - \sqrt{\frac{\tau}{1-\tau}}\right) b, \\
\frac{1}{N^2} \sum_{i=1}^N E [(x_i x_i') \otimes x_i' f_i(0 | x_i)] &= \frac{1}{N} E(x_i^3) f(0) = \frac{6}{N} \frac{1}{b-a}, \\
\frac{1}{N^2} \sum_{i=1}^N E [(x_i \otimes x_i) u_i^2 \mathbf{1}(u_i < 0)] &= \frac{1}{N} E(x_i^2) E[u_i^2 \mathbf{1}(u_i < 0)] = -\frac{2}{N} \frac{a^3}{3(b-a)}, \\
\frac{1}{N^2} \sum_{i=1}^N E [(x_i \otimes x_i) u_i^2] &= \frac{1}{N} E(x_i^2) E(u_i^2) = \frac{1}{N} E \left(1 - \sqrt{\frac{\tau}{1-\tau}}\right) b + \frac{8}{3N}.
\end{aligned}$$

For the simulation results in Table 2, we generate $x_i = 1$. We simulate y_i from $y_i = x_i' \beta + u_i$, and set $k = 1$, $\beta = 0$, $R = 4$. Since in this setup x_i and u_i are both i.i.d., then for each τ , we have

$$\begin{aligned}
Q &= [4\tau(1-\tau)]^{-1}, \\
\frac{1}{N^2} \sum_{i=1}^N Q E[u_i \mathbf{1}(u_i < 0)] &= -\frac{1}{N} Q \frac{a^2}{2(b-a)}, \\
\frac{1}{N^2} \sum_{i=1}^N Q E(u_i) &= \frac{1}{2N} Q \left(1 - \sqrt{\frac{\tau}{1-\tau}}\right) b, \\
\frac{1}{N^2} \sum_{i=1}^N f_i(0) &= \frac{1}{N} \frac{1}{b-a}, \\
\frac{1}{N^2} \sum_{i=1}^N E[u_i^2 \mathbf{1}(u_i < 0)] &= \frac{1}{N} E[u_i^2 \mathbf{1}(u_i < 0)] = -\frac{1}{N} \frac{a^3}{3(b-a)}, \\
\frac{1}{N^2} \sum_{i=1}^N E(u_i^2) &= \frac{1}{N} E(u_i^2) = \frac{1}{2N} \left(1 - \sqrt{\frac{\tau}{1-\tau}}\right) b + \frac{4}{3N}.
\end{aligned}$$

In empirical applications without knowledge of the distribution of $\{x_i, u_i\}$, the expectations can be evaluated by a bootstrap method. The density and derivatives of density can be evaluated by non-parametric kernel methods.

References

- Aigner, D.J., Amemiya, T., and Poirier, D., 1976. On the estimation of production frontiers: maximum likelihood estimation of the parameters of a discontinuous density function. *International Economic Review* 17, 377-396.
- Bao, Y., Ullah, A., 2007. The second-order bias and mean squared error of estimators in time-series models. *Journal of Econometrics* 140, 650-669.
- Gelfand, I.M., Shilov, G.E., 1964. *Generalized Functions* Vol. 1. New York: Academic Press.
- Koenker, R., Bassett, G.S., 1978. Regression quantiles. *Econometrica* 46, 33-50.
- Kuan, C., Yeh, J., and Hsu, Y., 2009. Assessing value at risk with CARE, the conditional autoregressive expectile models. *Journal of Econometrics* 150, 261-270.
- Newey, W.K., Powell, J.L., 1987. Asymmetric Least Squares Estimation and Testing. *Econometrica* Vol. 55, No. 4, 819-847.
- Pagan, A., Ullah, A., 1999. *Nonparametric Econometrics*. Cambridge university press.
- Raju, C. K., 1982. Products and compositions with the Dirac delta function. *Journal of Physics A: Mathematical and General* 15, 381-396.
- Rilstone, P., Srivastava, V. K., Ullah, A., 1996. The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics* 75, 369-395.
- Xie, S., Zhou, Y., and Wan, A., 2014, A varying-coefficient expectile model for estimating value at risk. *Journal of Business and Economic Statistics* 32:4, 576-592.
- Yao, Q., Tong, H., 1996. Asymmetric least squares regression estimation: a nonparametric approach. *Nonparametric Statistics* 6, 273-292.

Table 1: Conditional ALS regression

		$N = 100$			$N = 300$		
τ		$\hat{\beta}_{MC}$	$\hat{\beta}_{Thm}$	$\tilde{\beta}_{MC}$	$\hat{\beta}_{MC}$	$\hat{\beta}_{Thm}$	$\tilde{\beta}_{MC}$
0.1	bias	0.0110	0.0185	-0.0075	0.0039	0.0062	-0.0023
	MSE	5.8157	3.6989	5.7503	1.7702	3.4974	1.7608
0.2	bias	0.0063	0.0091	-0.0028	0.0026	0.0030	-0.0004
	MSE	6.4746	8.9006	6.4433	2.0444	3.0466	2.0376
0.3	bias	0.0045	0.0047	-0.0003	0.0014	0.0016	-0.0002
	MSE	6.9401	7.7755	6.9204	2.1407	2.5508	2.1388
0.4	bias	0.0029	0.0021	0.0008	-0.0004	0.0007	-0.0011
	MSE	6.9121	7.1058	6.9045	2.2059	2.3170	2.2070
0.5	bias	-0.0006	0.0000	-0.0006	0.0002	0.0000	0.0002
	MSE	7.0851	6.9000	7.0851	2.2253	2.2481	2.2253
0.6	bias	-0.0005	-0.0021	0.0016	-0.0011	-0.0007	-0.0004
	MSE	6.9674	7.1058	6.9697	2.2591	2.3170	2.2580
0.7	bias	-0.0037	-0.0047	0.0011	-0.0025	-0.0016	-0.0009
	MSE	6.7513	7.7755	6.7391	2.2169	2.5508	2.2115
0.8	bias	-0.0051	-0.0091	0.0040	-0.0014	-0.0030	0.0017
	MSE	6.3461	8.9006	6.3360	1.9873	3.0466	1.9883
0.9	bias	-0.0102	-0.0185	0.0083	-0.0033	-0.0062	0.0029
	MSE	5.8627	3.6989	5.8274	1.7144	3.4974	1.7121

Notes: Table 1 presents the simulation results when x_i is generated from the exponential distribution. Table 2 presents the simulation results when $x_i = 1$. For each τ , the first row is for bias and the second row is for the mean squared error ($\times 10^3$) of the ALS estimator. The results are presented in the following manner in each corresponding cell,

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

where the true value $\beta(\tau) = 0$ for all τ . For each panel, the first and third columns present the bias and MSE of the expectile estimator before and after bias correction, the second column presents the second-order bias and MSE calculated by Theorems. The bias-corrected estimate is $\tilde{\beta}(\tau) = \hat{\beta}(\tau) - B(\hat{\beta}(\tau))$. The subscript j denotes the j th Monte Carlo replication ($j = 1, \dots, J$). We replicate $J = 10,000$ times in the Monte Carlo.

Table 2: Unconditional ALS regression

		$N = 100$			$N = 300$		
τ		$\hat{\beta}_{MC}$	$\hat{\beta}_{Thm}$	$\tilde{\beta}_{MC}$	$\hat{\beta}_{MC}$	$\hat{\beta}_{Thm}$	$\tilde{\beta}_{MC}$
0.1	bias	0.0079	0.0123	-0.0045	0.0026	0.0041	-0.0015
	MSE	10.0440	30.0709	10.0024	3.2959	9.5140	3.2911
0.2	bias	0.0036	0.0061	-0.0025	0.0009	0.0020	-0.0011
	MSE	11.7827	18.9978	11.7761	3.9036	6.2261	3.9040
0.3	bias	0.0036	0.0032	0.0004	0.0003	0.0011	-0.0007
	MSE	12.8124	15.3133	12.7997	4.2727	5.0752	4.2731
0.4	bias	0.0004	0.0014	-0.0010	0.0009	0.0005	0.0004
	MSE	13.1874	13.7730	13.1882	4.3606	4.5853	4.3600
0.5	bias	0.0007	0.0000	0.0007	0.0006	0.0000	0.0006
	MSE	13.1945	13.3333	13.1945	4.4281	4.4444	4.4281
0.6	bias	-0.0035	-0.0014	-0.0021	-0.0015	-0.0005	-0.0011
	MSE	13.2806	13.7730	13.2727	4.5048	4.5853	4.5036
0.7	bias	-0.0037	-0.0032	-0.0006	0.0007	-0.0011	0.0017
	MSE	13.0086	15.3133	12.9951	4.2295	5.0752	4.2320
0.8	bias	-0.0049	-0.0061	0.0012	-0.0029	-0.0020	-0.0009
	MSE	11.9406	18.9978	11.9185	4.0243	6.2261	4.0165
0.9	bias	-0.0060	-0.0123	0.0063	-0.0016	-0.0041	0.0025
	MSE	10.1822	30.0709	10.1857	3.3284	9.5140	3.3322

Notes: See Table 1.