The kth power expectile estimation and testing

*Fuming Lin

School of Mathematics and Statistics, Sichuan University of Science & Engineering, Sichuan Zigong, China

Yingying Jiang

School of Mathematics and Statistics, Sichuan University of Science & Engineering, Sichuan Zigong, China

Yong Zhou

Institute of Statistics and Interdisciplinary Sciences and School of Statistics, Faculty of Economics and Management, East China Normal University, Shanghai, China

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

February 16, 2021

Abstract

This paper develops the theory of the kth power expectile estimation and considers its relevant hypothesis tests for coefficients of linear regression models. We prove that the asymptotic covariance matrix of kth power expectile regression converges to that of quantile regression as k converges to one, and hence provide a moment estimator of asymptotic matrix of quantile regression. The kth power expectile regression is then utilized to test for homoskedasticity and conditional symmetry of the data. Detailed

* corresponding author. Email-address: linfuming20062015@163.com (F. Lin).

comparisons of the local power among the kth power expectile regression tests, the quantile regression test, and the expectile regression test have been provided. When the underlying distribution is not standard normal, results show that the optimal k are often larger than 1 and smaller than 2, which suggests the general kth power expectile regression is necessary.

AMS subject classifications. 62G08, 62H15, 62F35

Keywords: The kth power expectiles; Expectiles; Quantiles; Testing for homoskedasticity; Testing for conditional symmetry; Estimating asymptotic matrix of quantile regression.

1 Introduction

Homoskedasticity and conditional symmetry are important to evaluate regression coefficient estimates. So tests of these conditions have been attracting many researchers in econometrics and statistics. In the beginning, for the linear regression models, the majority of tests for homoskedasticity are constructed recognizing whether the residuals (or their functions) from an interesting preliminary regression are dependent on the covariate, see Anscombe (1961), Glejser (1969), Goldfeld and Quandt (1972), Harvey (1976), Godfrey (1978), Breusch and Pagan (1979), and White (1980). Tests of symmetry of the error distributions have not been focused on enough, albeit Antille, Kersting, and Zucchini (1982), and Boos (1982) considered these tests for the i.i.d. data.

Another completely different group of methods are based on the quantlie regression or the expectile regression. Koenker and Bassett (1982)/Newey and Powell (1987) proposed to reject the null hypothesis of homoskedasticity provided the slope coefficients of the quantile/expectile regression at different weights of the loss function are significantly different. Newey and Powell (1987) also found that the quantile regression and the expectile regression can be utilized to test symmetry. Newey and Powell (1987) compared in terms of relative efficiency the expectile-based approach with the quantile-based one and other commonly-used approaches in linear regression models when the error term obeys a contaminated normal distribution. They reported that: the expectile-based approach is the same efficient as the method of absolute residual regression, and is generally better than the quantile-based approach when testing heteroskedasticity; the expectile-based approach is almost as good as the method of median versus mean, and is mostly better than the quantile-based approach when testing conditional symmetry. For quantile-based methods and expectile-based methods, each of course has its advantages and disadvantages. The former is comprehensible and robust in the non-normal cases, but computationally difficult, especially in calculating the asymptotic variance. The later is easy for calculation and more efficient in normal cases, but less intuitive. Anyway, as two extremely important families of approaches, they receives great attention from other areas far more than test of homoskedasticity and symmetry. For organized introduction and discussions to quantile regression methods, see Koenker (2005), Engle and Manganelli (2004), Kim (2007), Cai and Xu (2008), Cai and Xiao (2012), Andriyana et al. (2016), Koenker (2017) and Wang et al. (2020), among others. For the literature on expectile-based methods, see Efron (1991), Yao and Tong (1996), Granger and Sin (1997), Taylor (2008), Kuan et al. (2009), Gu and Zou (2016), Farooq and Steinwart (2017), Daouia et al. (2018) and Daouia et al. (2020), among others. Zhao et al. (2018) analyzed heteroscedasticity in high dimension using the expectile regression.

Methodologies based on quantiles regression or expectiles regression appeal to many researchers. But there is few attention put on the relation between quantiles and expectiles. Jiang et al. (2019) considered a general check function $Q_\tau(r) = (\tau - I\{r < 0\})|r|^k$ $(1 < k \leq 2)$ and proposed the kth power expectile regression. They prove the asymptotic normality of the estimators related to the regression and provide the consistent estimator of the asymptotic covariance matrix. They unexpectedly found that the asymptotic efficiencies of the kth power expectile estimation $(1 < k < 2)$ in both scale-location models and location shift models are much higher than those of the quantile estimation or the expectile estimation in many cases. Cabrera and Schulz (2017) used the same check function, but they only supposed $k = 1$ or 2. Efron (1991) studied the same check function but only setting $k = 1.5$ and asserted its corresponding regression method may balance the robustness (the merit of the quantile regression) and the effectiveness (the merit of the expectile regression). Efron (1991) did not provide theoretical analysis for his assertion. Returning to Jiang et al.(2019), they gave a proof of the asymptotic normality by restricting k to the interval $(1.5, 2]$ when the data are i.i.d. It is imperfect to overlook the case $k \in (1, 1.5]$ from either a theoretical point of view or a perspective of statistically empirical analysis. We ultimately found that the problem comes down to their condition (6) in Assumption 3, i.e.

$$\int_{-\infty}^{+\infty} |y - x'b|^{k-2} f(y|x)dy/|x|^k < c_2, \tag{1}$$

for any b, as $|x| \to +\infty$, which is the primary cause why Jiang et al. (2019) did not complete the proof of their asymptotic normality for $k \in (1, 1.5]$. For common distributions, such as those with bounded probability densities $f(y|x)$ (the boundedness is uniform in x), we have a more strong result, for $1 < k \leq 2$, i.e.,

$$\int_{-\infty}^{+\infty} |y - x'b|^{k-2} f(y|x)dy < c_1 \tag{2}$$

since $\int_{-\infty}^{+\infty} f(y|x)dy = 1$ and $\int_0^{c_3} s^{k-2}ds < c_4$, where c_i are positive constants independent on x. Condition (1) is too weak for a great many of distributions while condition (2) is suitable.

Under the condition similar to (2), we prove asymptotic normality of the kth power expectile regression estimators for $1 < k \leq 2$ by a method totally different from Jiang et al.(2019) when the data is either i.i.d. or local alternative to the i.i.d. setting. The a.s. uniform consistency of the estimators are also investigated. When $k = 2$, the kth power expectile regression theory is just the expectile regression theory in Newey and Powell (1987). Although the condition (2) is not satisfied for $k = 1$ and hence the quantile regression theory (Koenker and Bassett (1978)) could not be deduced from that of kth power expectile regression, we show that the mean vector and covariance matrix of the asymptotic distribution of the later converges to those of the asymptotic distribution of the quantile regression estimators in Koenker and Bassett (1978) when k converges to 1. This is a new discovery and interesting result, see Remark 3.4 below for a potential application, which theoretically makes sure we can get an optimal k in the whole $[1, 2]$ by minimizing the trace or the sum of eigenvalues of the estimated variance matrix of the kth power expectile estimation. This paper then constructs a general linear relation test based upon the kth power expectile regression following the train of thought of Newey and Powell (1987), which relaxes the requirement in Newey and Powell (1987) (which supposed the data mean exists) since the kth power expectile only requires that the $(k-1)$th order moment of the data exists, and hence is very friendly to the financial data. As its special cases, tests of homoskedasticity and conditional symmetry are mainly investigated and the limit distributions of the test statistics are provided. We compare the local power of the proposed test to that of the test based on the quantile regression/the expectile regression when the error comes from the contaminated normal distribution or the Student t distribution. Although on past results, the method based on the expectile regression will outweigh that based on the quantile regression under normal error cases, we find the kth $(1 < k < 2)$ power expectile regression miraculously performs better than the quantile regression and the expectile regression for the majority of the combination values of contamination percentage and relative scale of the contaminated normal distribution. Especially, the kth $(1 < k < 2)$ power expectile regression dominates the quantile regression for testing of both homoskedasticity and symmetry. It seems that using the kth $(1 < k < 2)$ power expectile regression instead of the expectile/quantile regression can produce more satisfying results in many issues. In practice, an optimal $\tilde{k}$ $(1 \leq \tilde{k} \leq 2)$ can easily be found such that the estimated variance matrix of the corresponding $\tilde{k}$th power expectile estimation has the smallest eigenvalues in the sense of norm.

The kth power expectile estimators are a special type of M-estimators. But the conditions of asymptotic results for the M-estimators are difficult to check, see He and Shao (2000), and our conditions of the limit theorems are easily verified. In this sense, the kth power expectile regression theory enriches the literature on M-estimators.

The paper has four main contributions. The first one is putting the quantile regression and the expectile regression in a unified framework, i.e., the proposed kth power expectile regression. The unified framework makes it possible that we can find an optimal k in $[1, 2]$ using cross validation or the aforementioned method. The second one is providing some easily-checked moment conditions under which the whole theories of the kth power expectile regression hold. Finding these conditions is the challenging work. The third one is proposing a completely new moment estimator for asymptotic matrix of quantile regression. In past, estimating asymptotic matrix of quantile regression undertakes the difficulty of estimating the density function of error, which can now be gotten round by our method. The last one is to present new methods to test homoskedasticity and symmetry, and a new view point about the efficiencies of the quantile regression test, the expectile regression test, and ours. Concretely, we find the efficiencies of the three tests all decrease as the tail of the data becomes thicker, but the descending rates are different from each other.

The paper proceeds as follows. In Section 2, the notation and a basic property of the kth power expectile regression estimators have been given. Section 2 also contains some comparisons between the kth power expectile estimation and the MLE in the skewed power exponential distribution. Section 3 contains the main asymptotic results and assumptions needed to derive these results. The test setting, test statistics, their asymptotic distributions are provided in Section 4, and Section 5 compares in details three tests in terms of relative efficiency, in which we consider both the normal error and the Student t distribution errors. Section 6 concludes the paper. All proofs are postponed to Section 7. Throughout the paper, the c_i, c, C_i and C are positive and finite constants which may vary from line to line.

2 Notation, properties and relationship with MLE

Firstly, notation and properties for the kth power expectile regression are presented. Secondly, we summarize some interesting links between the kth power expectile estimation and the MLE in the skewed power exponential distribution.

2.1 Notation and some properties

Suppose the data $(y_t, x_t')'$, $(t = 1, \cdots, T)$ are copies of $(y, x')'$, which satisfies

$$y = x'\beta_0 + u, \tag{3}$$

where x' is a p-dimension vector with first component being one, β_0 is a vector of parameters, and u is an error terms. Consider the kth power loss function:

$$Q_{\tau,k}(s) = |\tau - I(s < 0)||s|^k, \tag{4}$$

with weight $\tau \in (0, 1)$ and $1 \leq k \leq 2$. Letting Y be a random variable, $\mu(k, \tau, Y)$, the kth power expectile of Y for weight τ, is

$$\mu(k, \tau, Y) = \text{argmin}_l E(Q_{\tau,k}(Y - l) - Q_{\tau,k}(Y)). \tag{5}$$

So $\mu(1, \tau, Y)$ and $\mu(2, \tau, Y)$ are exactly the quantile and expectile of Y, respectively, see Koenker and Bassett (1978) and Newey and Powell (1987). The kth power expectile of Y contains information about the full distribution of Y.

The regression coefficient of kth power expectile regression of y on x' for weight τ is

$$\tilde{\beta}(k, \tau, y, x) = \text{argmin}_{\beta} E(Q_{\tau,k}(y - x'\beta) - Q_{\tau,k}(y)).$$

If $(y, x')'$ satisfies (3) and u is independent of x , i.e., a homoskedasticity hypothesis,

$$\mu(k, \tau, y) = x'\tilde{\beta}(k, \tau, y, x) = x'\beta_0 + \mu(k, \tau, u).$$

In other words, only the first component of $\tilde{\beta}(k, \tau, y, x)$ varies as τ changes, which motivates a test of homoskedasticity. Suppose $(y_t, x_t')'$ are i.i.d. copies of $(y, x')'$, we can construct the estimator $\hat{\beta}$ through minimizing the sample counterpart of $E(Q_{\tau,k}(y - x'\beta) - Q_{\tau,k}(y))$, i.e.,

$$\hat{\beta}(k, \tau, y_t, x_t) = \text{argmin}_{\beta \in R^p} \sum_{t=1}^{T} (Q_{\tau,k}(y_t - x_t'\beta) - Q_{\tau,k}(y_t))/T.$$

Remark 2.1. *Theorems 3.1 and 3.2 below discuss the existence a.s. and with probability one, respectively.*

In the sequel, we mainly focus on $k \in (1, 2]$ and often omit k, y, x, y_t, and x_t and write $\mu(k, \tau, Y)$, $\tilde{\beta}(k, \tau, y, x)$, and $\hat{\beta}(k, \tau, y_t, x_t)$ as $\mu(k, \tau)$, $\tilde{\beta}(k, \tau)$, and $\hat{\beta}(k, \tau)$ whenever it does not cause confusion, and even as $\mu(\tau)$, $\tilde{\beta}(\tau)$, and $\hat{\beta}(\tau)$ when we do not care about k. The existence and uniqueness of $\mu(\tau)$ were proved by Theorem 1 in Jiang et al. (2019). Some other basic properties about $\hat{\beta}(\tau)$ can be found in their Theorem 2. The following is a simple but important result about $\tilde{\beta}(\tau)$ which implies a test of conditional symmetry.

Theorem 2.1 *Suppose y satisfies* (3) *and fix $k \in (1, 2]$. The u is distributed symmetrically around zero with $E|u|^{k-1} < \infty$ and the c.d.f F_u. Then we*

$$\tilde{\beta}(k, \tau) + \tilde{\beta}(k, 1 - \tau) = 2\beta_0.$$

2.2 Relationship with MLE

Nelson (1991) put forward the PED (power exponential distribution), also called the generalized error distribution, with the following density

$$f(x) = f(x; \mu, \sigma, \eta) = \frac{\eta \exp(-\frac{1}{2}|\frac{x-\mu}{\sigma c}|^{\eta})}{2^{1+1/\eta} c\sigma\Gamma(\frac{1}{\eta})}, \tag{6}$$

where $\mu \in R$, $\sigma > 0$, $\eta > 0$, and $c = c(\eta) = \sqrt{\Gamma(1/\eta)(2^{2/\eta}\Gamma(3/\eta))^{-1}}$. Due to its tractability and usefulness in practice, generalized forms of the PED have been provided, such as the multidimension PED and various skewed PED. A natural extension of the PED, by replacing "$\frac{1}{2}$" in the exponent part of the density with "1-τ" or "τ" in (6), has the density

$$g(x) = g(y; \mu, \sigma, \eta, \tau) = \frac{\eta \exp(-|\tau - I(y < \mu)||\frac{x-\mu}{\sigma c}|^{\eta})}{((\frac{1}{1-\tau})^{1/\eta} + (\frac{1}{\tau})^{1/\eta}) c\sigma\Gamma(\frac{1}{\eta})}, \tag{7}$$

which is a skewed version of (6) and contains the PED as a special case, $\tau = 1/2$. It is obvious that the (skewed) PED deduces the (skewed) Gaussian distribution when $\eta = 2$, and the (skewed) Laplace distribution when $\eta = 1$.

We suppose the data y_i are generated by $y_i = x_i'\beta + \varepsilon_i$, where x_i' are covariates and ε_i are i.i.d. from the skewed PED. Some simple calculation shows that the kth power expectile estimation of β is equivalent to its MLE. Especially, the quantile regression estimation of β is exactly the same as the MLE of β in the skewed Laplace distribution, while the expectile regression estimation of β is exactly the same as the MLE of β in the skewed Gaussian distribution. Nevertheless, the kth power expectiles regression methods are completely different from and cannot be replaced by the MLE in skewed PED, as quantile regression and expectile regression have not been respectively replaced by the MLE in the skewed Laplace distribution and the MEL in the skewed Gaussian distribution. First and foremost, the kth power expectiles regression does not set the specific type of error distributions and has inherent robustness. Secondly, the kth power expectiles regression cares for the kth power expectiles of variables but not the whole distribution density of variables. Roughly speaking, the kth power expectiles regression theory in the present paper may be treated roughly as the quasi-likelihood theory related to the (skewed) PED distributions.

There are many MLE related to the PED or the generalized PED in the regression setting, such as latest Prataviera et al. (2019) and Prataviera et al. (2020). They suppose that the parameters of the generalized PED are the parametric or semiparametric expressions of covariates. We mainly focus on how the kth power expectiles of y_i is explained by the x_i'.

3 Asymptotic properties of the kth power expectile estimators

The asymptotic theory for the kth power expectile estimators will be proved under more general assumptions, which contain linear models (3) as special cases. Let l denote the Lebesgue measure on the real line and let $z \equiv (y, x')$, where x is a $p \times 1$ vector. For a matrix $A = [a_{ij}]$, let $|A| \equiv \max_{i,j} |a_{ij}|$. The l_2-norm is denoted by $\| \cdot \|$.

3.1 a. s. uniform consistency

For considering the consistency, we generally let the data be independent without requirement of identical distribution. The results are established on the following conditions.

Condition A: *Suppose* $z_t = (y_t, x_t')'$ $(t = 1, \ldots, T)$ *are independent but may not be identically distributed. Their probability density functions are* $f_t(y|x)g_t(x)$ *with respect to a measure* $M_{z,t} = l \times M_{x,t}$*, where* l *is the Lebesgue measure and* $M_{x,t}$ *is the measure related to* $g_t(x)$*. Additionally, the* $f_t(y|x)$ *is continuous in* y *for almost all* x*.*

Condition B: $\int |z_t|^k f_t(y|x)g_t(x)dM_{z,t} < c$*, where* c *is dependent from* t*.*

Condition C: *The following inequalities hold,*

$$\inf_{x_t \in R^p} \inf_{b \in \mathcal{B}} \int_{-\infty}^{+\infty} |y - x_t'b|^{k-2} f_t(y|x_t)dy = c_1 > 0,$$
$$\int_{-\infty}^{+\infty} |y - x_t'b|^{k-2} f_t(y|x_t)dy < c, \tag{8}$$

where c_1 *and* c *are independent from* x_t *and* t*,* $\mathcal{B}$ *is an arbitrary compact subset of* R^p*.*

Condition D: *(i) For fixed* $U = [\tau_l, \tau_h] \subset (0, 1)$ *and* $k \in (1, 2]$*, there exists a constant* $C_1 > 0$ *such that, for all* $\tau, \tau' \in U$*, we have*

$$|\tilde{\beta}(k, \tau) - \tilde{\beta}(k, \tau')\| \leq C_1|\tau - \tau'|.$$

(ii) For fixed $K = [k_l, k_h] \subset (1, 2]$ *and* $\tau \in (0, 1)$*, there exists a constant* $C_2 > 0$ *such that, for all* $k, k' \in K$*, we have*

$$|\tilde{\beta}(k, \tau) - \tilde{\beta}(k', \tau)\| \leq C_2|k - k'|.$$

Condition E: *The* $(1/T)\sum_{t=1}^{T} E(x_t x_t')$ *is nonsingular.*

Remark 3.1. *In fact, Condition C is not strong since it can be satisfied for common data generating processes, such as* $y_t = g(x_t, \beta) + \varepsilon_t$*, where* $g(x_t, \beta)$ *is the function of* x_t *and parameter* β*, and* ε_t *obeys the exponential distribution family or the Student t distribution..*

We have the following uniform consistency result for $\hat{\beta}(k, \tau)$.

Theorem 3.1 *If Conditions A-C and E are satisfied, then there exists unique global minimum* $\tilde{\beta}(k, \tau)$ *of* $\frac{1}{T}\sum_{t=1}^{T} E(Q_{\tau,k}(y_t - x_t' b))$*. Let* $\mathbf{\Theta}$ *(with* $\tilde{\beta}(k, \tau)$ *as its interior point) be a compact subset of the parameter space. There is a measurable function* $\hat{\beta}(k, \tau)$ *such that*

$$\hat{\beta}(k, \tau) = argmin_{b \in \mathbf{\Theta}} \frac{1}{T} \sum_{t=1}^{T} Q_{\tau,k}(y_t - x_t' b)$$

almost surely. In addition, if (i) of Condition D holds, then

$$\sup_{\tau \in U} \|\hat{\beta}(k, \tau) - \tilde{\beta}(k, \tau)\| \xrightarrow{a.s.} 0 \quad as\ T \to \infty\ for\ fixed\ k;$$

and if (ii) of Condition D holds, then

$$\sup_{k \in K} \|\hat{\beta}(k, \tau) - \tilde{\beta}(k, \tau)\| \xrightarrow{a.s.} 0 \quad as\ T \to \infty\ for\ fixed\ \tau.$$

3.2 Asymptotic normality

For analyzing asymptotic normality, we need the following conditions.

Assumption 1. *Let* $z_t = (y_t, x_t')'$ $(t = 1, \ldots, T)$ *be i.i.d. copies of* $z = (y, x')'$ *and* z *has a probability density function* $f(y|x, \xi_T)g(x)$ *with respect to a measure* $M_z = l \times M_x$*, where* $\xi_T = \xi_0 + \zeta/\sqrt{T}$ *(*ξ_0 *and* ζ *in* R^m*), with* T *being sample size and* M_x *being the measure related to* $g(x)$*. Additionally, the* $f(y|x, \xi_0)$ *is continuous in* y *for almost all* x*.*

Let $E(\cdot|\xi) = \int \cdot f(y|x,\xi)g(x)dM_z$, and $E(\cdot) = E(\cdot|\xi_0)$. Also, Define $\varphi_\tau(r) := (-1)^{I(r<0)}k|\tau - I(r<0)||r|^{k-1}$.

Assumption 2. *There exists an open set Ξ containing ξ_0 such that for almost all z, the conditional density $f(y|x,\xi)$ is continuous in ξ on Ξ. Also, $E(x\varphi_\tau(y - x'\beta(\tau))|\xi)$ is continuously differentiable in ξ on Ξ.*

Assumption 3. *There is a constant $c > 0$ and a measure function $\theta(z)$ that satisfies $\sup_{\xi\in\Xi} f(y|x,\xi) \leq \theta(z)$ and*

$$\int \theta(z)g(x)dM_z < +\infty,$$
$$\int |z|^{k+2+c}\theta(z)g(x)dM_z < +\infty, \text{ for } 1 < k \leq 2.$$

Assumption 4. *For $1 < k \leq 2$,*

$$\inf_{x\in R^p} \inf_{b\in\mathcal{B}} \int_{-\infty}^{+\infty} |y - x'b|^{k-2} f(y|x,\xi)dy = c_1 > 0,$$
$$\int_{-\infty}^{+\infty} |y - x'b|^{k-2}\theta(z)dy < c_2,$$

where c_1 and c_2 are independent on x, and $\mathcal{B}$ is an arbitrary compact subset of R^p.

Assumption 5. *The $E(xx')$ is nonsingular.*

Remark 3.2. *Assumptions 1, 2, and 5 are similar to Assumptions 1, 2, and 4 in Newey and Powell (1987). Assumption 3 is more weaker than Assumption 3 in Newey and Powell (1987). Assumption 4 is a typical condition for the kth power expectile regression, which can be satisfied for common data generating processes, such as $y_t = g(x_t,\beta)+\varepsilon_t$, where $g(x_t,\beta)$ is the function of x_t and parameter β, and ε_t obeys the exponential distribution family or the Student t distribution.*

For a vector of weights $(\tau_1, \cdots, \tau_n)'$, let $\hat{\eta} := \text{vec}(\hat{\beta}(\tau_1), \cdots, \hat{\beta}(\tau_n))$ denote the vector of ALS estimators and let $\tilde{\eta} := \text{vec}(\tilde{\beta}(\tau_1), \cdots, \tilde{\beta}(\tau_n))$ be the population counterpart (the distribution corresponds to the density $f(y|x, \xi_0)g(x)$ with respect to a measure $M_z = l \times M_x$). For $u(\tau) := y - x'\tilde{\beta}(\tau)$, $w(\tau) := |\tau - I(u(\tau) < 0)|$, define

$$
\begin{aligned}
D_j &:= E(k(k-1)w(\tau_j)|u(\tau_j)|^{k-2}xx'),\ D = \text{diag}(D_1, \cdots, D_n);\\
V_{ji} &:= E(\varphi_{\tau_j}(u(\tau_j))\varphi_{\tau_i}(u(\tau_i))xx'),\ V = (V_{ji})(j, i = 1, \cdots, n);\\
K_j &:= \partial E(-\varphi_{\tau_j}(u(\tau_j))x|\xi_0)/\partial\xi,\ K := (K_1', \cdots, K_n')'.
\end{aligned}
$$

Theorem 3.2. *If Assumptions 1-5 are satisfied, then for each τ in $(0,1)$, a unique solution $\tilde{\beta}(\tau)$ to the equation*

$$E((-1)^{1-I(y-x'\beta<0)}|\tau - I(y - x'\beta < 0)||y - x'\beta|^{k-1}x) = 0$$

exists, where the expectation is calculated with respect to $f(y|x, \xi_0)g(x)$. Also,

$$\sqrt{T}(\hat{\eta} - \tilde{\eta})) \xrightarrow{d} N(D^{-1}K\zeta, D^{-1}VD^{-1}).$$

In order to make the result feasible, the asymptotic variance matrix need to be estimated. Let $\hat{u}_t(\tau) := y_t - x_t'\hat{\beta}(\tau)$, $\hat{w}_t(\tau) := |\tau - I(\hat{u}_t(\tau) < 0)|$, define

$$
\begin{aligned}
\hat{D}_j &:= k(k-1)\sum_{t=1}^{T}\hat{w}_t(\tau_j)|\hat{u}_t(\tau_j)|^{k-2}x_t x_t'/T,\ \hat{D} = \text{diag}(\hat{D}_1, \cdots, \hat{D}_n);\\
\hat{V}_{ji} &:= \sum_{t=1}^{T}\varphi_{\tau_j}(\hat{u}_t(\tau_j))\varphi_{\tau_i}(\hat{u}_t(\tau_i))x_t x_t')/T,\ \hat{V} = (\widehat{V}_{ji})(j, i = 1, \cdots, n).
\end{aligned}
$$

Theorem 3.3. *Based on Assumptions 1-5, we have the following result*

$$\hat{D}^{-1}\hat{V}\hat{D}^{-1} \xrightarrow{P} D^{-1}VD^{-1}.$$

Remark 3.3. *Obviously, Theorem 3.2 and Theorem 3.3 are generalizations of Theorem 3 and Theorem 4 of Newey and Powell (1987), respectively. The main results in Jiang et al. (2019) are only very special cases of our Theorem 3.2.*

Combining our setting and the context of Koenker and Bassett (1982), we below prove the mean and covariance of the limit distribution in Theorem 3.2 converges to those of the limit distribution in Theorem 3.1 in Koenker and Bassett (1982), which displays that there is an inherent relation among the quantile regression, the expectile regression, and the kth power expectile regression. We need the related natation and conditions as follows.

Assumption 6. *We consider the i.i.d. data $z_t = (y_t, x_t')'(t = 1, 2, \dots, T)$. The data generating process is*

$$y = x'\beta + \Big(1 + x'\frac{\gamma_0}{\sqrt{T}}\Big)\varepsilon,$$

where γ_0 is a constant vector in R^p the error ε has the distribution function F_ε with the density f_ε being continuous and strictly positive. For all υ, $0 < F_\varepsilon(\upsilon) < 1$. x_t' is independent of ε with the density $g(x)$ with respect to μ_x. The $q_\varepsilon(\tau)$ denotes the τ quantile of ε.

Assumption 7.

$$f_\varepsilon(x)x \to 0, \text{as } x \to \pm\infty,\ (f_\varepsilon(x)|x|^{k-1} \to 0, \text{as } x \to \pm\infty).$$

Assumption 8.

$$\int_{-\infty}^{+\infty} |x|^{k-1} f_\varepsilon'(x)dx < c.$$

In Theorem 3.4 and its proof, let $E_T(\cdot)$, $E(\cdot)$, and $E_T(\cdot|\gamma)$ denote the expectations calculated with respect to the densities $f_\varepsilon((y - x'\beta)/(1 + x\frac{\gamma_0}{\sqrt{T}}))(g(x)/(1 + x\frac{\gamma_0}{\sqrt{T}}))$, $f_\varepsilon(y -$

$x'\beta)g(x)$, and $f_\varepsilon((y - x'\beta)/(1 + x\gamma))(g(x)/(1 + x\gamma))$, respectively. The other notation is the same as Theorem 3.1. Denote

$$\begin{aligned}
\tilde{D}_j &:= E(k(k-1)w(\tau_j)|u(\tau_j)|^{k-2}xx'),\ \tilde{D} = \mathrm{diag}(\tilde{D}_1, \cdots, \tilde{D}_n);\\
\tilde{V}_{ji} &:= E(\varphi_{\tau_j}(u(\tau_j))\varphi_{\tau_i}(u(\tau_i))xx'),\ \tilde{V} = (\tilde{V}_{ji})(j, i = 1, \cdots, n);\\
\tilde{K}_j &:= \partial E(-\varphi_{\tau_j}(u(\tau_j))x|0)/\partial\gamma,\ \tilde{K} := (\tilde{K}_1', \cdots, \tilde{K}_n')'.
\end{aligned}$$

Theorem 3.4. *Suppose Assumption 6 and Assumptions 2-5 are satisfied. then for each τ in $(0, 1)$ and k in $(1, 2]$, a unique solution $\tilde{\beta}(\tau)$ to the equation*

$$E((-1)^{1-I(y-x'\beta<0)}|\tau - I(y - x'\beta < 0)||y - x'\beta|^{k-1}x) = 0$$

exists. Then (i) a similar result as in Theorem 3.2 holds, i.e.,

$$\sqrt{T}(\hat{\eta} - \tilde{\eta})) \xrightarrow{d} N(\tilde{D}^{-1}\tilde{K}\gamma_0, \tilde{D}^{-1}\tilde{V}\tilde{D}^{-1}).$$

(ii) furthermore, if Assumptions 7 and 8 hold, when $k \to 1$, we have element wise convergence, $\tilde{D}^{-1}\tilde{K}\gamma_0 \to \bar{K}$, $\tilde{D} \to \bar{D}$, and $\tilde{V} \to \bar{V}$, where

$$\begin{aligned}
\bar{D}_j &:= f_\varepsilon(q_\varepsilon(\tau_j))E(xx'),\ \bar{D} = diag(\bar{D}_1, \cdots, \bar{D}_n);\\
\bar{V}_{jk} &:= (\min(\tau_j, \tau_k) - \tau_j\tau_k)E(xx'),\ \bar{V} = (\bar{V}_{jk})(j, k = 1, \cdots, n);\\
\bar{K}_j &:= q_\varepsilon(\tau_j)\gamma_0,\ \bar{K} := (\bar{K}_1', \cdots, \bar{K}_n')'.
\end{aligned}$$

Remark 3.4. *Theorem 3.4 demonstrates that, for $k \to 1$, the mean and covariance of the asymptotic distribution of the estimators for $1 < k \leq 2$ can converges to those of the asymptotic distribution in Theorem 3.1 in Koenker and Bassett (1982). This illuminating result implies that we could estimate the covariance matrix $\bar{D}^{-1}\bar{V}\bar{D}^{-1}$ of quantile regression by estimator $\hat{D}^{-1}\hat{V}\hat{D}^{-1}$ in Theorem 3.3 letting k as small as possible.*

Remark 3.5. *The kth power expectile regression could be also of interest in other problems except testing for homoskedasticity and conditional symmetry. For example, we can estimate regression percentiles by the kth power expectile regression through the same procedure as Efron (1991) estimated regression percentiles by regression expectiles. Concretely, suppose the data is $\{y_t, x_t'\}$, $t = 1, \ldots, T$, x_t' are the $1 \times p$ covariate vectors, and y_t are scaler responses. We first obtain enough kth power expectile regression lines, like the expectile regression lines in Figure 2 of Efron (1991). We then select a line in them as the α quantile regression line such that the ratio of the number of data below the line to T is α. We are motivated by two reasons: One is that the asymptotic variances of the kth power expectile regression is smaller in many distribution cases, see Fig. 2 in Jiang et al. (2019); the other is that the kth power expectile regression is computationally easy, especially estimating the variances.*

Remark 3.6. *In practice, we can adopt different approaches to obtain the optimal k. If we are interesting in parameters estimation, we reach the optimal k, for fixed τ, in a grid research such that the eigenvalues of the corresponding asymptotic variance matrix estimate in Theorem 3.3 are as small as possible. If we focus on forecast, we can use the cross validation method. Testing is closely related to parameters estimating, we can use the first approach to select the optimal k.*

4 Test statistics and the asymptotic theory

We examine testing for homoskedasticity and conditional symmetry of the error distribution using the kth power expectile regressions. In order to conform to the setting of Koenker and Bassett (1982) and Newey and Powell (1987), we give the following assumptions.

Assumption 9. *Let the data generating process be*

$$y_t = x_t'\beta_0 + u_t \tag{9}$$

with

$$u_t = \sigma_t\varepsilon_t,\ \ \sigma_t = 1 + x_t'\xi_{Th} + I(\varepsilon_t > 0)x_t'\xi_{Ts}, \tag{10}$$

where $\xi_{Th} = \zeta_h/\sqrt{T}$, $\xi_{Ts} = \zeta_s/\sqrt{T}$, ζ_h *and* ζ_s *are p-dimensional constant vectors,* T *is sample size and* ε_t *is i.i.d. and symmetrically distributed around zero. Furthermore,* ε_t *is independent of* x_t *and has the c.d.f.* $F_\varepsilon(r)$ *with a continuous density* $f_\varepsilon(r)$.

Assumption 10. *The* x_t *has a compact support set. Moreover, there are positive and finite constants* C *and* c *such that*

$$f_\varepsilon(r) \leq \frac{C}{1+|r|^{k+3+c}}.$$

Consider the general linear hypothesis

$$H_0 : H\tilde{\eta} = m. \tag{11}$$

When testing heteroscedasticity, we let $m = 0$ and

$$H = \nabla^h \otimes \Upsilon,$$

where '$\otimes$' denotes the Kronecker product, ∇^h is an $(n-1)\times n$ matrix with representative element $\nabla^h_{ij} = \delta_{ij} - \delta_{i(j-1)}$ (δ_{ij} is Kronecker delta), and $\Upsilon = [0, I_{p-1}]$ (here and below I denotes a unit matrix). Suppose weights $(\tau_1,\ldots,\tau_n)$ satisfy $\tau_1 < \ldots < \tau_n$. According to Remark 4 in Jiang et. al (2019), $H\tilde{\eta} = 0$ implies homoskedasticity in the linear model (9).

When testing conditional symmetry, we suppose i^* is the median of $(1, \cdots, n)$ where n is odd, and

$$\tau_{i^*} = \frac{1}{2}, \tau_i = 1 - \tau_{2i^*-i},\ 0 < i < i^*. \tag{12}$$

At the moment, letting $m = 0$ and

$$H = \nabla^s \otimes \Upsilon,$$

where $\nabla^s = [I_{(m-1)/2}, -2e_{(m-1)/2}, I_{(m-1)/2}]$, and $\Upsilon = I_p$, $H\tilde{\eta} = 0$ suggests the conditional symmetry of y_t on x_t in the linear model (9) according to Theorem 2.1.

In order to test (11), using estimators in Theorem 3.2 and Theorem 3.3, we construct the test statistics TS as follows.

$$TS = T(H\hat{\eta} - m)'(H\hat{D}^{-1}\hat{V}\hat{D}^{-1}H')^{-1}(H\hat{\eta} - m). \tag{13}$$

In order to present the limit distribution of TS, we introduce some new pieces of notation, where known notation is explained in section 3.2. Let Ω denote the matrix with representative element ϖ_{jk}, where

$$\begin{aligned}
\varpi_{jk} &:= c(\tau_j, \tau_k)/(l(\tau_j)l(\tau_k)),\ (j, k = 1, \ldots, n), \\
c(\tau_j, \tau_k) &:= E(\varphi_{\tau_j}(u(\tau_j))\varphi_{\tau_k}(u(\tau_k))), \\
l(\tau_j) &:= E(k(k-1)\omega(\tau_j)|u(\tau_j)|^{k-2}).
\end{aligned}$$

Write $L := E(x_t x_t')$ and

$$\upsilon 1^* = (\upsilon 1(\tau_1)/l(\tau_1), \ldots, \upsilon 1(\tau_n)/l(\tau_n)),\ \upsilon 2^* = (\upsilon 2(\tau_1)/l(\tau_1), \ldots, \upsilon 2(\tau_n)/l(\tau_n)),$$

where

$$
\begin{aligned}
\upsilon 1(\tau) \quad &:= \quad -k(k-1)\Bigg((1-\tau)\int_{-\infty}^{\mu(k,\tau)}(\mu(k,\tau)-r)^{k-2} r f_{\varepsilon}(r)dr \\
&\qquad\qquad + \tau\int_{\mu(k,\tau)}^{\infty}(r-\mu(k,\tau))^{k-2} r f_{\varepsilon}(r)dr\Bigg), \\
\upsilon 2(\tau) \quad &:= \quad -k(k-1)\Bigg((1-\tau)\int_{0}^{\max(\mu(k,\tau),0)}(\mu(k,\tau)-r)^{k-2} r f_{\varepsilon}(r)dr \\
&\qquad\qquad + \tau\int_{\max(\mu(k,\tau),0)}^{\infty}(r-\mu(k,\tau))^{k-2} r f_{\varepsilon}(r)dr\Bigg),
\end{aligned}
$$

and $\mu(k,\tau)$ is the kth power expectile of ε_t.

Theorem 4.1. *Let Assumptions 1, 5, 9, and 10 be satisfied. If H_0 is satisfied when $\xi := (\xi_{Th}', \xi_{Ts}')' = 0$, Ω is nonsingular, and H is of full row rank with rank being $r(H)$, TS converges in distribution to a noncentral chi-squared random variable with $r(H)$ degrees of freedom and noncentrality parameter*

$$
(\upsilon 1^* \otimes \zeta_h + \upsilon 2^* \otimes \zeta_s)' H' (H(\Omega \otimes L^{-1})H')^{-1} H (\upsilon 1^* \otimes \zeta_h + \upsilon 2^* \otimes \zeta_s). \tag{14}
$$

Corollary 4.1. *For $H = \nabla^h \otimes \Upsilon$ and $\zeta_s = 0$, (14) reduces to $\kappa_{kp}^h \cdot (\Upsilon\zeta_h)'(\Upsilon L^{-1}\Upsilon')^{-1}(\Upsilon\zeta_h)$ with*

$$
\kappa_{kp}^h := (\nabla^h \upsilon 1^*)'(\nabla^h \Omega (\nabla^h)')^{-1}(\nabla^h \upsilon 1^*).
$$

For $H = \nabla^s \otimes \Upsilon$ and $\zeta_h = 0$, (14) reduces to $\kappa_{kp}^s \cdot (\Upsilon\zeta_s)'(\Upsilon L^{-1}\Upsilon')^{-1}(\Upsilon\zeta_s)$ with

$$
\kappa_{kp}^s := (\nabla^s \upsilon 2^*)'(\nabla^s \Omega (\nabla^s)')^{-1}(\nabla^s \upsilon 2^*).
$$

5 Comparisons of the local efficiencies

We will compare in detail our proposed tests based on the kth power expectile regression (test III) with the tests based on the quantile regression by Koenker and Bassett (1982) (test I) and the tests based on expectile regression by Newey and Powell (1987) (test II) in term of local efficiencies. The data for simulation study are generated by letting the error term obey the contaminated normal distribution and the Student t distribution.

5.1 Testing under the contaminated normal error

We set ε_t in (10) obeying the contaminated normal distribution with the density:

$$f_\varepsilon(r) = (1-\vartheta)\varphi(r) + (\vartheta/\varrho)\varphi(r/\varrho), \tag{15}$$

where ϑ is the proportion of contamination and ϱ is the scale. Its kurtosis function and related partial derivatives are

$$K(\vartheta,\varrho) = \frac{3(1-\vartheta)+3\vartheta\varrho^4}{((1-\vartheta)+\vartheta\varrho^2)^2},$$

$$\frac{\partial K(\vartheta,\varrho)}{\partial\vartheta} = \frac{3(\varrho^2-1)(\varrho^2-1+\vartheta-\vartheta\varrho^4)}{((1-\vartheta)+\vartheta\varrho^2)^3}, \tag{16}$$

$$\frac{\partial K(\vartheta,\varrho)}{\partial\varrho} = \frac{12\vartheta\varrho(1-\vartheta)(\varrho^2-1)}{((1-\vartheta)+\vartheta\varrho^2)^3}. \tag{17}$$

It is well-known that the bigger the kurtosis value of a distribution is, the heavier its tail becomes. The family of contaminated normal distributions are often selected as heavier-tailed alternatives to strictly normal errors, used by Tukey (1960), Koenker and Bassett (1982), Newey and Powell (1987), among others. We can show easily that Assumption 10 is satisfied for the contaminated normal distributions. Under this error setting, the local efficiencies of tests of homoskedasticity and symmetry are exhibited by computing Pitman's asymptotic relative efficiencies (AREs).

5.1.1 Testing heteroskedasticity

We suppose $\zeta_s = 0$ when testing heteroskedasticity. Under the linear restriction (11), Corollary 4.1 demonstrates that the statistics of test III converges to the noncentral chi-squared random variable in distribution with $r(H)$ degrees of freedom and noncentrality parameter $\kappa_{kp}^h \cdot (\Upsilon\zeta_h)'(\Upsilon L^{-1}\Upsilon')^{-1}(\Upsilon\zeta_h)$ where

$$\kappa_{kp}^h := (\nabla^h \upsilon 1^*)'(\nabla^h \Omega(\nabla^h)')^{-1}(\nabla^h \upsilon 1^*). \tag{18}$$

According to Theorem 4.1 in Koenker and Bassett (1982), Corollary 1 in Newey and Powell (1987), the asymptotic distributions for tests I and II are the same as test III, except for different noncentrality parameters. The noncentrality parameter for the quantile regression is $\kappa_{QR}^h \cdot (\Upsilon\zeta_h)'(\Upsilon L^{-1}\Upsilon')^{-1}(\Upsilon\zeta_h)$ where

$$\kappa_{QR}^h := (\nabla^h \mu(1,\tau))'(\nabla^h \Sigma_{QR}(\nabla^h)')^{-1}(\nabla^h \mu(1,\tau)), \tag{19}$$

$\mu(1,\tau)$ is the quantile vector, Σ_{QR} is the asymptotic covariance matrix of the vectors of quantile regression estimators. The noncentrality parameter for the expectile regression is $\kappa_{ER}^h \cdot (\Upsilon\zeta_h)'(\Upsilon L^{-1}\Upsilon')^{-1}(\Upsilon\zeta_h)$, where

$$\kappa_{ER}^h := (\nabla^h \mu(2,\tau))'(\nabla^h \Sigma_{ER}(\nabla^h)')^{-1}(\nabla^h \mu(2,\tau)), \tag{20}$$

$\mu(2,\tau)$ is the expectile vector, and Σ_{ER} is just the Ω in (18) for $k = 2$.

We examine the efficiency of test III through considering the case of $n = 2$, $\tau = (\tau_1, \tau_2)'$ and $\tau_1 = 1 - \tau_2$, $0.5 < \tau_2 < 1$. Tests I and II employ the same setting. Since the three tests have the similar limit distributions with the same degrees of freedom, the desired AREs reduce to the ratios of noncentrality parameters, which turn out to be $\kappa^h_{kp}/\kappa^h_{QR}$ and $\kappa^h_{kp}/\kappa^h_{ER}$. For different proportion of contamination ϑ and scale ϱ that take values in sets $\{0.05, 0.1, \cdots, 0.5\}$ and $\{2, 3, 4, 5\}$, respectively, we obtain in a grid research the optimal combination of k and τ, on which the κ^h_{kp} reaches its maximum, see the heteroskedasticity panel in Table 1. Under the same setting of proportion of contamination and scale as above, Newey and Powell (1987) gave the optimal weights for tests I and II, see their Table 1. As can be seen in our Table 1, the optimal τ values are between 0.54 and 0.97; the optimal k values are between 1.2 and 1.9. At each fixed contamination level, the optimal k values decrease as scale ϱ increases; but the optimal weight τ values vary little except at contamination levels of 0.45 and 0.5. Both optimal τ values and optimal k values decrease firstly and then increase as contamination level ϑ increases for each fixed scale. The averages of the optimal τ and k values are 1.60 and 0.75, respectively.

In our paper, we use the optimal weights (in our Table 1 for test III and Newey and Powell's Table 1 for tests I and II) across different combinations of ϱ and ϑ to calculate values of κ^h_{kp} κ^h_{QR} and κ^h_{ER} using R codes, see columns 3-5 in Table 2. Table 2 shows that for the fixed ϑ, κ^h_{kp}, κ^h_{QR} and κ^h_{ER} all decrease as ϱ increases. In other words, as the tail thickness of the error distribution is a strictly monotone increasing function of ϱ by (17), κ^h_{kp} κ^h_{QR} and κ^h_{ER} all decrease as the tail thickness of the data distribution increases (i.e., the data set which contain more and more fairly large outliers). Since the tail thickness of the error distribution is a concave function of ϑ according to (16), for fixed ϱ, κ^h_{kp} κ^h_{QR} and κ^h_{ER} almost decrease firstly and then increase as ϑ increases, except for cases of $\varrho = 2, 3$ in which κ^h_{QR} is the strictly decreasing function of ϑ.

The third column in Table 3 lists the AREs of test III relative to test II, which show

31/41 of the ARE values larger than 1. By and large, test III is superior to test II. For $\varrho = 1$, i.e., the standard normal error, while it is well known test II is the most efficient method, there is an efficiency loss of only one percent when using instead test III with $k = 1.9$. Given ϑ, the AREs increase as ϱ increases, except at the level of 0.5. At the fixed ϑ, the superiority intensity, if test III is superior to test II, increases with increasing ϱ. The phenomenon is mainly due to the fact that both κ_{kp}^h and κ_{ER}^h decrease as ϱ increases, but the rate of the former is smaller than the latter, see columns 3 and 4 in Table 2. Another phenomenon is that, under fixed ϱ, the AREs increase firstly and then decrease as ϑ increases due to the different rates at which the κ_{kp}^h and the κ_{ER}^h change over ϑ.

The fourth column in Table 3 lists the AREs of test III relative to test I, which proclaim test III dominates overall test I. At the fixed ϑ, the superiority intensity decreases with increasing ϱ in $0.05 \leq \vartheta \leq 0.40$, and increases firstly and then decreases in $0.45 \leq \vartheta \leq 0.50$.

5.1.2 Testing symmetry

We suppose $\zeta_h = 0$ when testing symmetry. Under the linear restriction (11), Corollary 4.1 demonstrates the test statistics of test III asymptotically obeys the noncentral chi-squared distribution with $r(H)$ degrees of freedom and noncentrality parameter $\kappa_{kp}^s \cdot (\Upsilon\zeta_s)'(\Upsilon L^{-1}\Upsilon')^{-1}(\Upsilon\zeta_s)$, where

$$\kappa_{kp}^s = (\nabla^s \upsilon 2^*)'(\nabla^s \Omega (\nabla^s)')^{-1}(\nabla^s \upsilon 2^*). \tag{21}$$

Similarly, tests I and II have the same asymptotic distribution as above except for different noncentrality parameter. The noncentrality parameter for test I is $\kappa_{QR}^s \cdot (\Upsilon\zeta_s)'(\Upsilon L^{-1}\Upsilon')^{-1}(\Upsilon\zeta_s)$ where

$$\kappa_{QR}^s = (\nabla^s \eta^+)'(\nabla^s \Sigma_{QR} (\nabla^s)')^{-1}(\nabla^s \eta^+), \tag{22}$$

with $\eta^+ = \max\{\mathbf{0}, \mu(1,\theta)\}$ where "$\mathbf{0}$" is a zero vector with the same dimension as the quantile vector $\mu(1,\theta)$, Σ_{QR} is the same as in (19). The noncentrality parameter for test II is $\kappa_{ER}^s \cdot (\Upsilon\zeta_s)'(\Upsilon L^{-1}\Upsilon')^{-1}(\Upsilon\zeta_s)$ where

$$\kappa_{ER}^s = (\nabla^s\mu^*)'(\nabla^s\Sigma_{ER}(\nabla^s)')^{-1}(\nabla^s\mu^*), \tag{23}$$

Σ_{ER} is the same as in (20), and μ^* is a vector with typical element

$$\begin{aligned} \mu^*(\theta) \quad = \quad & \left(\theta\int_0^\infty sf_\varepsilon(s)ds + (1-2\theta)\int_0^{\max\{0,\mu(2,\theta)\}} sf_\varepsilon(s)ds\right) \\ & /(\theta(1-F_\varepsilon(\mu(2,\theta))) + (1-\theta)F_\varepsilon(\mu(2,\theta))). \end{aligned}$$

In the above expression θ is the weight in the expectile loss function and $\mu(2,\theta)$ is the θ expectile of the error distribution.

We use the same setting of the proportion of contamination ϑ and the scale ϱ as in the above subsection. According to (12), for convenience, we choose only three weights: $1-\tau$, $1/2$, and τ, i.e., $\theta = (1-\tau, 1/2, \tau)$ to construct test statistics. Let $\nabla^s \equiv (1,-2,1)'$. Similarly, to obtain AREs, we only need to calculate $\kappa_{kp}^s/\kappa_{QR}^s$ and $\kappa_{kp}^s/\kappa_{ER}^s$. Newey and Powell (1987) listed in their Table 2 the optimal weights for tests I and II. We still let ϑ and ϱ take its values in sets $\{0.05, 0.1, \cdots, 0.5\}$ and $\{2,3,4,5\}$, respectively, and under each couple (ϑ,ϱ) pick out the optimal combination of k and τ such that the κ_{kp}^s achieves its maximum at this combination, see the symmetry part in Table 1. Table 1 shows that these optimal τ values are between 0.54 and 0.90; the optimal k values are between 1.2 and 1.9. At each fixed contamination level, the optimal k values decrease as ϱ increases when $0.05 \le \vartheta \le 0.25$; the optimal k values converge to 1.9 when ϑ converges to 0.5 or ϱ converges to 2. Both optimal τ values and optimal k values decrease firstly and then increase as ϑ increases for each fixed scale value. The averages of the optimal τ and k values are 1.74 and 0.65, respectively.

Similarly, we use the optimal weights (in our Table 1 for test III and Newey and Powell's Table 2 for tests I and II) for different combinations of ϱ and ϑ to obtain values of κ^s_{kp} κ^s_{QR} and κ^s_{ER} and collect the results in columns 6-8 in Table 2. Results state for the fixed ϑ value, κ^s_{kp}, κ^s_{QR}, and κ^s_{ER} all decrease as ϱ increases in most of the cases. That is to say that κ^s_{kp} κ^s_{QR} and κ^s_{ER} decrease as the tail thickness of the data distribution increases. For fixed ϱ, κ^h_{kp} κ^h_{QR} and κ^h_{ER} approximately decrease firstly and then increase as ϑ increases, but the variations are very slight.

The fifth column in Table 3 contains the AREs of test III relative to test II, which exhibit that there are 25/41 of the values larger than 1. The maximum and minimum of these AREs are 1.4601 and 0.9576. In general, test III performs better than test II. Similarly to the test of homoskedasticity, the AREs depend on ϑ and ϱ. Specially, when $\varrho = 1$, i.e., the standard normal error, test II is still the most efficient, but there is an efficiency loss of less than two percent when employing instead test III with $k = 1.9$. The proportion of AREs, which are lager than 1 increases as ϱ increases. The main reason of the phenomenon is that both the κ^s_{kp} and the κ^s_{ER} decrease as ϱ increases, but the former does at a slower rate, see columns 6 and 7 in Table 2.

The last column in Table 3 contains the AREs of test III relative to test I, which suggest test III dominates test I. The superiority extent decreases with increasing ϱ.

Overall, we reveal two interesting phenomena: One is that test III outweighs test I, which does not depend on the parameter values of the contaminated normal distribution and the particular null hypothesis; another is that test III becomes more powerful than test II when the tail of the error distribution becomes heavier, which does not depend on the particular null hypothesis, too.

5.2 Testing under the Student t distribution error

In the section, let ε_t in (10) obey the Student t distribution, which is the frequently-used heavy-tailed distribution, and let the degrees of freedom df of the Student t distribution vary in the set $\{3, 4, 5, 6, 7, 8, 9, 10\}$. We examine how the local efficiencies of three test methods change with df. We use formulas (18)-(23) to calculate noncentrality parameters. The results are contained in Table 4, which provides the optimal k-values and weights for test III as well the optimal weights for tests I and II.

As seen, the optimal k increases with df; the change of optimal weights has not clear trend; κ_{kp}^h increases with df. About tests I and II, the optimal weight increases with df; both κ_{ER}^h and κ_{QR}^h increase with df. Figures 1 and 2 display the change of AREs of test III to tests I and II. The local efficiencies of test III is better than tests I and II. The AREs of test III to test II decreases with df while AREs of test III to test I do not decreases with df. Especially, when $df = 3$, i.e., the Student t distribution has much heavy tail, the proposed method has obvious advantages. The results in the Student t distribution error case are accordant with those in the contaminated normal error case.

Simulations explicitly exhibit that the kth power expectile method is more powerful than the quantile method in any commonly-used errors, more powerful than the expectile method in the heavy-tailed errors, and as well as the expectile method in the errors that are approximately standard normal. This is an interesting and important finding. In practice, we can adopt approaches in Remark 3.6 to select the optimal k-values and weights for the kth power expectile regression method.

Table 1. The optimum values of k and weights τ in the loss function of the kth power expectile when being used to test homoskedasticity and symmetry

	ϑ	ϱ=2		ϱ=3		ϱ=4		ϱ=5	
		k	weight τ	k	weight τ	k	weight τ	k	weight τ
homoskedasticity	0.05	1.9	0.83	1.5	0.87	1.4	0.86	1.3	0.87
	0.10	1.8	0.83	1.4	0.85	1.3	0.84	1.2	0.84
	0.15	1.8	0.81	1.5	0.81	1.3	0.82	1.2	0.82
	0.20	1.8	0.79	1.5	0.79	1.4	0.78	1.3	0.78
	0.25	1.8	0.78	1.6	0.75	1.4	0.76	1.3	0.76
	0.30	1.9	0.73	1.7	0.71	1.5	0.72	1.4	0.72
	0.35	2.0	0.66	1.8	0.66	1.7	0.65	1.6	0.65
	0.40	2.0	0.66	1.9	0.60	1.8	0.60	1.8	0.57
	0.45	2.0	0.67	2.0	0.51	2.0	0.51	2.0	0.97
	0.50	2.0	0.69	2.0	0.52	2.0	0.51	2.0	0.97
symmetry	0.05	1.9	0.77	1.4	0.89	1.3	0.89	1.2	0.90
	0.10	1.9	0.73	1.4	0.87	1.3	0.86	1.2	0.86
	0.15	1.9	0.68	1.6	0.79	1.3	0.84	1.2	0.84
	0.20	1.9	0.70	1.9	0.60	1.5	0.76	1.3	0.79
	0.25	1.8	0.78	1.9	0.59	1.7	0.67	1.5	0.72
	0.30	1.9	0.69	1.7	0.75	1.9	0.56	1.7	0.63
	0.35	1.9	0.68	1.6	0.66	1.9	0.56	1.9	0.54
	0.40	1.9	0.68	1.9	0.59	1.9	0.56	1.9	0.54
	0.45	1.9	0.69	1.9	0.60	1.9	0.56	1.9	0.54
	0.50	1.9	0.69	1.7	0.72	1.9	0.56	1.9	0.54

Table 2. κ^h_{kp} κ^h_{QR}, κ^h_{ER}, κ^s_{kp} κ^s_{QR}, and κ^s_{ER} in the noncentrality parameters for tests of homoskedasticity and symmetry,

ϱ	ϑ	κ^h_{kp}	κ^h_{ER}	κ^h_{QR}	κ^s_{kp}	κ^s_{ER}	κ^s_{QR}
2	0.05	1.6316	1.6259	1.2092	0.2693	0.2743	0.2376
	0.10	1.4987	1.4827	1.1340	0.2611	0.2669	0.2279
	0.15	1.4178	1.4012	1.0729	0.2962	0.2637	0.2198
	0.20	1.3655	1.3533	1.0219	0.2550	0.2627	0.2130
	0.25	1.3382	1.3263	0.9785	0.2548	0.2633	0.2071
	0.30	1.3161	1.3139	0.9413	0.2554	0.2649	0.2020
	0.35	1.3125	1.3125	0.9113	0.2571	0.2672	0.1980
	0.40	1.3196	1.3196	0.8867	0.2594	0.2702	0.1949
	0.45	1.3340	1.3340	0.8682	0.2623	0.2736	0.1928
	0.50	1.3549	1.3549	0.8566	0.2656	0.2773	0.1920
3	0.05	1.3788	1.2261	1.1395	0.2425	0.2388	0.2279
	0.10	1.2076	1.0379	1.0234	0.2265	0.2213	0.2117
	0.15	1.1039	0.9570	0.9327	0.2167	0.2141	0.1983
	0.20	1.0345	0.9227	0.8584	0.2131	0.2124	0.1871
	0.25	0.9883	0.9137	0.7921	0.2136	0.2142	0.1771
	0.30	0.9631	0.9206	0.7417	0.2460	0.2185	0.1681
	0.35	0.9575	0.9385	0.6950	0.3083	0.2245	0.1602
	0.40	0.9699	0.9650	0.6548	0.2247	0.2319	0.1531
	0.45	0.9984	0.9984	0.6203	0.2305	0.2405	0.1468
	0.50	1.0379	1.0379	0.6084	0.2914	0.2501	0.1515
4	0.05	1.2569	0.9271	1.1034	0.2316	0.1983	0.2223
	0.10	1.0793	0.7554	0.9694	0.2119	0.1750	0.2033
	0.15	0.9658	0.7001	0.8671	0.1979	0.1677	0.1878
	0.20	0.8851	0.6875	0.7841	0.1885	0.1675	0.1744
	0.25	0.8304	0.6961	0.7147	0.1844	0.1710	0.1625
	0.30	0.7961	0.7177	0.6554	0.1851	0.1771	0.1520
	0.35	0.7854	0.7483	0.6036	0.1899	0.1851	0.1424
	0.40	0.7970	0.7862	0.5578	0.1960	0.1947	0.1335
	0.45	0.8302	0.8302	0.5410	0.2031	0.2057	0.1352
	0.50	0.8800	0.8800	0.6081	0.2113	0.2182	0.1520
5	0.05	1.1916	0.7143	1.0818	0.2254	0.1618	0.2193
	0.10	1.0081	0.5796	0.9377	0.2036	0.1395	0.1983
	0.15	0.8907	0.5485	0.8301	0.1876	0.1345	0.1816
	0.20	0.8056	0.5518	0.7433	0.1754	0.1361	0.1672
	0.25	0.7441	0.5716	0.6705	0.1675	0.1411	0.1544
	0.30	0.7023	0.6018	0.6087	0.1653	0.1483	0.1429
	0.35	0.6847	0.6395	0.5550	0.1679	0.1571	0.1323
	0.40	0.6958	0.6833	0.5073	0.1744	0.1674	0.1228
	0.45	0.7641	0.7641	0.5410	0.1820	0.1791	0.1348
	0.50	0.8558	0.8558	0.6081	0.1905	0.1924	0.1518

Table 3. Ratio of local efficiencies of the proposed test to the quantule/expectile regression tests when using optimal k and optimal weights

ϱ	ϑ	$\kappa^h_{kp}/\kappa^h_{ER}$	$\kappa^h_{kp}/\kappa^h_{QR}$	$\kappa^s_{kp}/\kappa^s_{ER}$	$\kappa^s_{kp}/\kappa^s_{QR}$
1	······	0.9892	1.4665	0.9834	1.1476
2	0.05	1.0034	1.3492	0.9818	1.1331
	0.10	1.0108	1.3215	0.9782	1.1453
	0.15	1.0117	1.3214	1.1232	1.3475
	0.20	1.0090	1.3363	0.9705	1.1973
	0.25	1.0089	1.3675	0.9679	1.2303
	0.30	1.0016	1.3982	0.9643	1.2646
	0.35	1.0000	1.4402	0.9619	1.2981
	0.40	1.0000	1.4881	0.9598	1.3309
	0.45	1.0000	1.5365	0.9584	1.3603
	0.50	1.0000	1.5817	0.9576	1.3832
3	0.05	1.1245	1.2100	1.0152	1.0642
	0.10	1.1635	1.1800	1.0233	1.0700
	0.15	1.1534	1.1834	1.0122	1.0928
	0.20	1.1211	1.2051	1.0029	1.1387
	0.25	1.0817	1.2476	0.9970	1.2064
	0.30	1.0462	1.2985	1.1261	1.4636
	0.35	1.0201	1.3775	1.3737	1.9250
	0.40	1.0051	1.4811	0.9688	1.4676
	0.45	1.0000	1.6095	0.9583	1.5706
	0.50	1.0000	1.7059	1.1650	1.9230
4	0.05	1.3557	1.1390	1.1681	1.0421
	0.10	1.4288	1.1133	1.2106	1.0419
	0.15	1.3794	1.1137	1.1795	1.0538
	0.20	1.2874	1.1288	1.1253	1.0805
	0.25	1.1929	1.1619	1.0779	1.1341
	0.30	1.1092	1.2146	1.0451	1.2178
	0.35	1.0495	1.3011	1.0261	1.3341
	0.40	1.0138	1.4288	1.0067	1.4675
	0.45	1.0000	1.5344	0.9874	1.5023
	0.50	1.0000	1.4469	0.9685	1.3899
5	0.05	1.6682	1.1014	1.3931	1.0279
	0.10	1.7391	1.0751	1.4601	1.0270
	0.15	1.6235	1.0729	1.3941	1.0330
	0.20	1.4601	1.0839	1.2885	1.0493
	0.25	1.3015	1.1096	1.1871	1.0847
	0.30	1.1669	1.1536	1.1147	1.1563
	0.35	1.0707	1.2337	1.0686	1.2684
	0.40	1.0182	1.3716	1.0419	1.4207
	0.45	1.0000	1.4123	1.0157	1.3501
	0.50	1.0000	1.4072	0.9901	1.2551

Table 4. The optimum values of k and weights τ in the loss function of the kth power expectile when being used to test homoskedasticity and symmetry under the t distribution error

	df	the proposed method			expectile		quantile	
		k	weight	κ_{kp}	weight	κ_{ER}	weight	κ_{QR}
homoskedasticity	3	1.5	0.79	0.9643	0.51	0.6814	0.85	0.7516
	4	1.6	0.80	1.1133	0.51	0.9999	0.87	0.8411
	5	1.7	0.80	1.2257	0.59	1.1771	0.88	0.9056
	6	1.8	0.78	1.3131	0.65	1.2925	0.89	0.9543
	7	1.9	0.76	1.3824	0.69	1.3751	0.89	0.9919
	8	1.9	0.78	1.4401	0.72	1.4378	0.90	1.0231
	9	2.0	0.74	1.4874	0.74	1.4874	0.90	1.0481
	10	2.0	0.76	1.5276	0.76	1.5276	0.91	1.0685
symmetry	3	1.6	0.76	0.2051	0.51	0.1627	0.87	0.1750
	4	1.8	0.70	0.2271	0.51	0.2240	0.89	0.1901
	5	2.0	0.51	0.2493	0.51	0.2493	0.90	0.2002
	6	2.0	0.51	0.2614	0.51	0.2614	0.91	0.2073
	7	2.0	0.53	0.2680	0.53	0.2680	0.91	0.2127
	8	2.0	0.55	0.2723	0.55	0.2723	0.92	0.2169
	9	2.0	0.58	0.2754	0.58	0.2754	0.92	0.2203
	10	2.0	0.59	0.2776	0.59	0.2776	0.92	0.2229

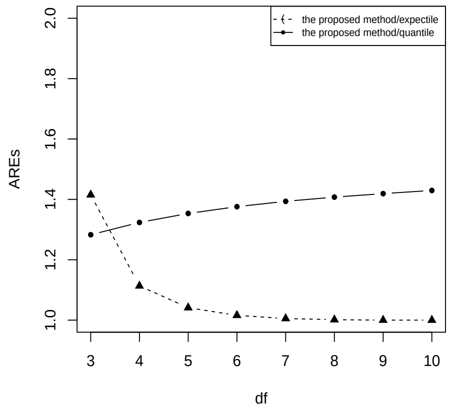

Figure 1: AREs of the proposed method to the expectile/quantile method when testing heteroscedasticity.

6 Conclusion

This paper develops the theory of the kth power expectile regression and proposes tests of homoskedasticity and conditional symmetry based on this regression method. Results suggest that the proposed tests perform more efficiently than the expectile/quantile regression test, especially in testing heteroskedasticity. No matter what the null hypothesis is, the priority should be given to the kth power expectile regression when the data distribution have a very heavy tail. These merits are attributed to the fact that the asymptotic variances of the kth power expectile regression are very small under heavy-tailed distributions. Although the theory of the kth power expectile regression is provided for the i.i.d. data case, it can be readily extended to the dependent data case because it is built on the easily-checked comments conditions. We can also consider other models, such as

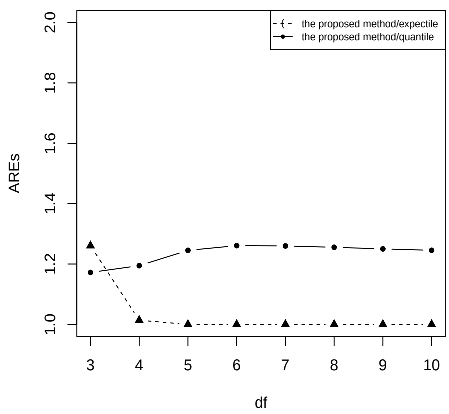

Figure 2: AREs of the proposed method to the expectile/quantile method when testing symmetry.

generalized linear models and high-dimensional sparse models using the kth power expectile regression or its non-parameter statistics version. Belloni and Chernozhukov (2011) and Gu and Zou (2016) considered the applications in high-dimensional sparse models of the quantile regression and the expectile regression, respectively. We believe that the kth power expectile regression could exhibit more merits in high-dimensional sparse models, which may be a future study.

7 Appendix A: The proofs of main results

Proof of Theorem 2.1. Denote the kth power expectile of u by $\delta(\tau)$, i.e., $\delta(\tau) = \text{argmin}_l E(Q_\tau(u-l) - Q_\tau(u))$. The first order condition for this minimization problem shows that $\delta(\tau)$ is the solution to

$$\frac{1-\tau}{\tau}=\frac{\int_{\delta(\tau)}^{\infty}(x-\delta(\tau))^{k-1}dF_u(x)}{\int_{-\infty}^{\delta(\tau)}(\delta(\tau)-x)^{k-1}dF_u(x)}. \tag{24}$$

The equation (24) has a unique solution according to Theorem 1 in Jiang et al. (2019), and can be rewritten as

$$\frac{1-(1-\tau)}{1-\tau}=\frac{\int_{-\delta(\tau)}^{\infty}(x-(-\delta(\tau)))^{k-1}dF_u(x)}{\int_{-\infty}^{-\delta(\tau)}(-\delta(\tau)-x)^{k-1}dF_u(x)}.$$

So the solution uniqueness deduces $\delta(\tau)=-\delta(1-\tau)$. Similarly, for the regression case, $\tilde{\beta}(k,\tau)$ satisfies equation

$$E((-1)^{1-I(y-x'\tilde{\beta}(k,\tau)<0)}x|y-x'\tilde{\beta}(k,\tau)|^{k-1}|\tau-I(y-x'\tilde{\beta}(k,\tau)<0)|)=0. \tag{25}$$

The item (iii) of Theorem 1 in Jiang et al. (2019) implies that

$$\begin{aligned} y-x'\tilde{\beta}(k,\tau) &= u+x'\beta_0-x'\tilde{\beta}(k,\tau)=u-\delta(\tau) \\ &= u+\delta(1-\tau)=u+x'\tilde{\beta}(k,1-\tau)-x'\beta_0 \\ &= y-x'(2\beta_0-\tilde{\beta}(k,1-\tau)). \end{aligned}$$

So, the left-hand side of (25) is equal to

$$\begin{aligned} &E((-1)^{1-I(y-x'(2\beta_0-\tilde{\beta}(k,1-\tau))<0)}x|y-x'(2\beta_0-\tilde{\beta}(k,1-\tau))|^{k-1} \\ &\qquad \cdot|\tau-I(y-x'(2\beta_0-\tilde{\beta}(k,1-\tau))<0)|). \end{aligned}$$

The solution uniqueness makes sure $\tilde{\beta}(k,\tau)+\tilde{\beta}(k,1-\tau)=2\beta_0$. $\square$

Proof of Theorem 3.1. We first need two lemmas.

Lemma 1. *If Conditions A-B are satisfied, for a compact set* $\mathbf{\Theta}$, *we have that*

$$\sup_{b\in\mathbf{\Theta}}\left|\frac{1}{T}\sum_{t=1}^{T}Q_{\tau,k}(y_t-x_t'b)-\frac{1}{T}\sum_{t=1}^{T}E(Q_{\tau,k}(y_t-x_t'b))\right|\xrightarrow{a.s.}0, \textit{ as } n\to\infty.$$

Lemma 2. *If Conditions A-C and E are satisfied,* $\frac{1}{T}\sum_{t=1}^{T} E(Q_{\tau,k}(y_t - x_t' b))$ *has unique global minimum* $\tilde{\beta}(k,\tau)$.

The existence and uniqueness of $\tilde{\beta}(k,\tau)$ in Theorem 3.1 is obtained using Lemma 2. As Condition A implies Assumption B. 1 and Assumption B.1.i in Bates and White (1985), using their Theorem 2.2 we obtain the existence of $\hat{\beta}(k,\tau)$. Furthermore, Lemma 1 and Lemma 2 show that Assumption B.1.ii and Assumption B.1.iii in Theorem 2.2 of Bates and White (1985) are satisfied hence

$$\hat{\beta}(k,\tau) \xrightarrow{a.s.} \tilde{\beta}(k,\tau) \quad \text{as } T \to \infty. \tag{26}$$

The following is based on the classical Glivenco-Cantelli argument. Let ε be any positive constant, and $n \in N$, and pick out $n+1$ τ-values by continuity $\tau_l = \tau_0 \leq \tau_1 \leq \ldots \leq \tau_n = \tau_h$, which makes sure that $\max_{1\leq i\leq n}\{|\tau_i - \tau_{i-1}|\} < \varepsilon$. By the continuity of $\tilde{\beta}(k,\tau)$, we can find τ_i^* and τ_{*i} such that $\tilde{\beta}(k,\tau_i^*) = \sup_{\tau\in[\tau_{i-1},\tau_i]}\{\tilde{\beta}(k,\tau)\}$ and $\tilde{\beta}(k,\tau_{*i}) = \inf_{\tau\in[\tau_{i-1},\tau_i]}\{\tilde{\beta}(k,\tau)\}$. For $\tau \in [\tau_{i-1},\tau_i]$, we have that

$$\begin{aligned} |\hat{\beta}(k,\tau) - \tilde{\beta}(k,\tau)| &\leq |\hat{\beta}(k,\tau_i^*) - \tilde{\beta}(k,\tau_{*i})| \\ &\leq |\hat{\beta}(k,\tau_i^*) - \tilde{\beta}(k,\tau_i^*)| + |\tilde{\beta}(k,\tau_i^*) - \tilde{\beta}(k,\tau_{*i})|. \end{aligned}$$

So

$$\sup_{\tau\in[\tau_l,\tau_h]} |\hat{\beta}(k,\tau) - \tilde{\beta}(k,\tau)| \leq \max_{1\leq i\leq n} |\hat{\beta}(k,\tau_i^*) - \tilde{\beta}(k,\tau_i^*)| + C_1\varepsilon,$$

where the second term in the right-hand side is due to the uniform continuity of $\tilde{\beta}(k,\tau)$ with respect to τ for fixed k. By (26), for any $n \in N$,

$$\begin{aligned} \limsup_{T\to\infty} \sup_{\tau\in[\tau_l,\tau_h]} \|\hat{\beta}(k,\tau) - \tilde{\beta}(k,\tau)\| &\leq \limsup_{T\to\infty} \max_{1\leq i\leq n} |\hat{\beta}(k,\tau_i^*) - \tilde{\beta}(k,\tau_i^*)| + C_1\varepsilon \\ &= C_1\varepsilon \quad \text{a.s.} \end{aligned}$$

The arbitrariness of ε deduces that

$$\sup_{\tau\in[\tau_l,\tau_h]} \|\hat{\beta}(k,\tau)-\tilde{\beta}(k,\tau)\| \xrightarrow{a.s.} 0 \quad \text{as } T\to\infty \text{ for fixed } k.$$

Using the same argument as the above, we can also prove that

$$\sup_{k\in K} \|\hat{\beta}(k,\tau)-\tilde{\beta}(k,\tau)\| \xrightarrow{a.s.} 0 \quad \text{as } T\to\infty \text{ for fixed } \tau.$$

The proof is completed. $\square$

Proof of Theorem 3.2. We only present the proof for the case $n=1$, as the argument for the case $n>1$ is similar. We first focus on the case $\zeta=0$ and consider the minimum $\tilde{\beta}(\tau)$ of $E(Q_{\tau,k}(y_t-x_t'b))$. Noting, when $\zeta=0$, Assumptions 1, 3, 4, and 5 imply Assumptions A, B, C, and E, the existence and uniqueness of $\tilde{\beta}(\tau)$ is established by Lemma 2. There exist constants c_1 and c_2 such that $Q_{\tau,k}(y_t-x_t'b)\le |z_t|^k(c_1+c_2|b|^k)$. Combining this and Assumptions 2 and 3, using Lemma A1 of Newey (1985) yields that

$$\sup_{b\in\Upsilon} |(1/T)\sum_{t=1}^{T} Q_{\tau,k}(y_t-x_t'b)-E(Q_{\tau,k}(y_t-x_t'b))| \xrightarrow{P} 0,$$

where Υ is a bounded open set containing $\tilde{\beta}(\tau)$. So Lemma A in Newey and Powell (1987) makes sure $\hat{\beta}(\tau)=\operatorname{argmin}_{R^p}((1/T)\sum_{t=1}^{T} Q_{\tau,k}(y_t-x_t'b))$ exists with probability approaching one and $\hat{\beta}(\tau)\xrightarrow{P}\tilde{\beta}(\tau)$.

We then provide the proof of the asymptotic normality. Let $E_T(\cdot):=E(\cdot|\xi_T)$. By the arguments similar to those in the proof of Theorem 3.1, we write $Q_{\tau,k}(\cdot)$ as $Q_\tau(\cdot)$, and have that $E_T(Q_\tau(y-x'\beta))$ is twice continuously differentiable in β for large enough T. Moreover,

$$\begin{aligned}
\lambda_T(\beta) &= \partial E_T(Q_\tau(y-x'\beta))/\partial\beta = E_T(g(\beta)),\ \ g(\beta)=-\varphi_\tau(y-x'\beta)x \\
\partial\lambda_T(\beta)/\partial\beta &= \partial^2 E_T(Q_\tau(y-x'\beta))/\partial\beta\partial\beta' = k(k-1)E_T(xx'|\tau-I(y<x'\beta)||y-x'\beta|^{k-2}).
\end{aligned}$$

By the continuity of $Q_\tau(y-x'\beta)$ in β, the continuity of $f(y|x,\xi)$ in ξ, and Assumption 3, the dominated convergence theorem makes sure that $E_T(Q_\tau(y-x'\beta))$ converges uniformly to $E(Q_\tau(y-x'\beta))$ on any compact neighborhood M of $\tilde{\beta}(\tau)$. We, therefore, show that there is a sequence $\tilde{\beta}_T(\tau)$ that minimizes $E_T(Q_\tau(y-x'\beta))$ on M such that $\lim_{T\to\infty}\tilde{\beta}_T(\tau)=\tilde{\beta}(\tau)$, and that for large enough T,

$$0=\lambda_T(\tilde{\beta}_T(\tau))=E_T(g(\tilde{\beta}_T(\tau))). \tag{27}$$

Using the continuity of $f(y|x,\xi)$ in ξ, Assumptions 3 and 4, the dominated convergence theorem also implies that $\partial\lambda_T(\beta)/\partial\beta$ converges uniformly on M to $\partial G(k,\beta,\tau)/\partial\beta$, where

$$\begin{aligned}\partial G(k,\beta,\tau)/\partial\beta \;=\; & k(k-1)E\Bigg(xx'\Bigg(\tau\int_{x'\beta}^{+\infty}(y-x'\beta)^{k-2}f(y|x,\xi_0)dy\\ & +(1-\tau)\int_{-\infty}^{x'\beta}(x'\beta-y)^{k-2}f(y|x,\xi_0)dy\Bigg)\Bigg).\end{aligned}$$

Noting that $\lim_{T\to\infty}\tilde{\beta}_T(\tau)=\tilde{\beta}(\tau)$ and $\partial G(k,\beta,\tau)/\partial\beta$ is nonsingular with respect to β in a compact set (see the argument in the proof of Lemma 2), there exist positive constants c and c' such that for T large enough

$$|\beta-\tilde{\beta}_T(\tau)|<c\Rightarrow|\lambda_T(\beta)|>c'|\beta-\tilde{\beta}_T(\tau)|. \tag{28}$$

The (i) item of (N-3) of Huber (1967) is satisfied. Now let $\eta(\beta,d)=\sup_{|\alpha-\beta|\le d}|g(\alpha)-g(\beta)|$.

Write

$$
\begin{aligned}
\eta(\beta,d) &= \sup_{|\alpha-\beta|\le d} |(-1)^{1-I(y-x'\alpha<0)}k|\tau - I(y-x'\alpha<0)||y-x'\alpha|^{k-1}x \\
&\qquad\qquad -(-1)^{1-I(y-x'\beta<0)}k|\tau - I(y-x'\beta<0)||y-x'\beta|^{k-1}x| \\
&= k|x| \sup_{|\alpha-\beta|\le d} |(-1)^{1-I(y-x'\alpha<0)}|\tau - I(y-x'\alpha<0)||y-x'\alpha|^{k-1} \\
&\qquad\qquad -(-1)^{1-I(y-x'\beta<0)}|\tau - I(y-x'\beta<0)||y-x'\beta|^{k-1}| \\
&= k|x|(\sup_{|\alpha-\beta|\le d} |I(y-x'\alpha<0, y-x'\beta<0)(1-\tau)((x'\alpha-y)^{k-1}-(x'\beta-y)^{k-1})| \\
&\quad + \sup_{|\alpha-\beta|\le d} |I(y-x'\alpha\ge 0, y-x'\beta\ge 0)\tau((y-x'\alpha)^{k-1}-(y-x'\beta)^{k-1})| \\
&\quad + \sup_{|\alpha-\beta|\le d} |I(y-x'\alpha<0, y-x'\beta\ge 0)((1-\tau)(x'\alpha-y)^{k-1}+\tau(y-x'\beta)^{k-1})| \\
&\quad + \sup_{|\alpha-\beta|\le d} |I(y-x'\alpha\ge 0, y-x'\beta<0)(\tau(y-x'\alpha)^{k-1}+(1-\tau)(x'\beta-y)^{k-1})|) \\
&=: I_1+I_2+I_3+I_4.
\end{aligned}
$$

We have that, $\tilde{\alpha}$ being between α and β,

$$
I_1 \le c_2|x| \sup_{|\alpha-\beta|\le d} |I(y-x'\alpha<0, y-x'\beta<0)(x'\tilde{\alpha}-y)^{k-2}x'(\alpha-\beta)|.
$$

So,

$$
\begin{aligned}
E_T(I_1) &\le c_3 E\left(|x|^2 \int_{-\infty}^{x'\tilde{\alpha}} (x'\tilde{\alpha}-y)^{k-2} f(y|x,\xi_T)dy\right) d \\
&\le c_4 E\left(|x|^2 \int_{-\infty}^{+\infty} |x'\tilde{\alpha}-y|^{k-2}\theta(z)dy\right) d = O(d),
\end{aligned}
$$

where the last equality is due to Assumptions 3 and 4. Using the same argument as for I_1, we have $I_2 = O(d)$. Furthermore

$$
\begin{aligned}
I_3 &\le c_5|x| \sup_{|\alpha-\beta|\le d} |I(y-x'\alpha<0, y-x'\beta\ge 0)((1-\tau)(x'\alpha-y)^{k-1}+\tau(y-x'\beta)^{k-1})| \\
&\le c_6|x| \sup_{|\alpha-\beta|\le d} |x'\alpha-x'\beta| \le c_7|x|^2 d.
\end{aligned}
$$

The second inequality comes from the fact that $a^r + b^r \le a + b$, for $a,\ b > 0$ and $0 < r < 1$. Thus Assumption 3 implies $E_T(I_3) = O(d)$. Using the same argument as for I_3, $E_T(I_4) = O(d)$. Additionally,

$$\begin{aligned}
\eta^2(\beta, d) &= \Big(\sup_{|\alpha-\beta|\le d} |g(\alpha) - g(\beta)| \Big)^2 \\
&\le k^2|x|^2 \sup_{|\alpha-\beta|\le d} (((-1)^{1-I(y-x'\alpha)}|\tau - I(y - x'\alpha < 0)||y - x'\alpha|^{k-1})^2 \\
&\quad +((-1)^{1-I(y-x'\beta)}|\tau - I(y - x'\beta < 0)||y - x'\beta|^{k-1})^2 - 2(-1)^{I(y-x'\alpha<0)+I(y-x'\beta<0)} \\
&\quad |\tau - I(y - x'\alpha < 0)||\tau - I(y - x'\beta < 0)||y - x'\alpha|^{k-1}|y - x'\beta|^{k-1}) \\
&\le k^2|x|^2 \sup_{|\alpha-\beta|\le d} (I(y - x'\alpha < 0, y - x'\beta < 0)((1-\tau)^2(x'\alpha - y)^{2(k-1)} \\
&\qquad +(1-\tau)^2(x'\beta - y)^{2(k-1)} - 2(1-\tau)^2(x'\alpha - y)^{k-1}(x'\beta - y)^{k-1})) \\
&\quad +k^2|x|^2 \sup_{|\alpha-\beta|\le d} (I(y - x'\alpha \ge 0, y - x'\beta \ge 0)(\tau^2(y - x'\alpha)^{2(k-1)} \\
&\qquad +\tau^2(y - x'\beta)^{2(k-1)} - 2\tau^2(y - x'\alpha)^{k-1}(y - x'\beta)^{k-1})) \\
&\quad +k^2|x|^2 \sup_{|\alpha-\beta|\le d} (I(y - x'\alpha < 0, y - x'\beta \ge 0)((1-\tau)(x'\alpha - y)^{k-1} + \tau(y - x'\beta)^{k-1})^2) \\
&\quad +k^2|x|^2 \sup_{|\alpha-\beta|\le d} (I(y - x'\alpha \ge 0, y - x'\beta < 0)(\tau(y - x'\alpha)^{k-1} + (1-\tau)(x'\beta - y)^{k-1})^2) \\
&=: J_1 + J_2 + J_3 + J_4.
\end{aligned}$$

$$\begin{aligned}
E_T(J_1) &= E_T\Big(k^2|x|^2 \sup_{|\alpha-\beta|\le d} (I(y - x'\alpha < 0, y - x'\beta < 0)((1-\tau)^2((x'\alpha - y)^{k-1} \\
&\quad -(x'\beta - y)^{k-1})(x'\alpha - y)^{k-1} + (1-\tau)^2((x'\beta - y)^{k-1} - (x'\alpha - y)^{k-1})(x'\beta - y)^{k-1}))\Big) \\
&= E_T(k^2(k-1)|x|^2 \sup_{|\alpha-\beta|\le d} |I(y - x'\alpha < 0, y - x'\beta < 0)((1-\tau)^2(x'\tilde{\alpha}_1 - y)^{k-2}x'(\alpha - \beta) \\
&\quad (x'\alpha - y)^{k-1} + (1-\tau)^2(x'\tilde{\alpha}_2 - y)^{k-2}x'(\beta - \alpha)(x'\beta - y)^{k-1}))| \\
&\le c_8 E\Big(|z|^{k+2} \int_{-\infty}^{+\infty} (|x'\tilde{\alpha}_1 - y|^{k-2} + |x'\tilde{\alpha}_2 - y|^{k-2})\theta(z)dz\Big)d \\
&= O(d),
\end{aligned}$$

where $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ are between α and β, the second equality is based on the mean value theorem, and the last one is due to Assumptions 3 and 4. The same argument deduces $E_T(J_2) = O(d)$. For $1 < k < 1.5$,

$$\begin{aligned} E_T(J_3) &= E_T\Big(k^2|x|^2 \sup_{|\alpha-\beta|\le d} |I(y - x'\alpha < 0, y - x'\beta \ge 0)|((1-\tau)(x'\alpha - y)^{k-1} \\ &\quad +\tau(y - x'\beta)^{k-1})^2\Big) \\ &\le c_9 E_T\Big(k^2|x|^2 \sup_{|\alpha-\beta|\le d} |I(y - x'\alpha < 0, y - x'\beta \ge 0)|((x'\alpha - y)^{k-1} \\ &\quad +(y - x'\beta)^{k-1})^2\Big) \\ &\le c_{10}E(|x|^2|x'(\alpha-\beta)|) \le c_{11}d = O(d), \end{aligned}$$

where the first inequality is based on the fact $a^r + b^r \le c_{12}(a+b)^{1/2}$, for $0 < r < 0.5$, $a, b > 0$ and Assumption 3. For $1.5 \le k \le 2$,

$$\begin{aligned} E_T(J_3) &\le c_{14}E_T\Big(|x|^2 \sup_{|\alpha-\beta|\le d} |I(y - x'\alpha < 0, y - x'\beta \ge 0)||x'(\alpha-\beta)|^{2k-2}\Big) \\ &\le c_{15}E(|x|^{2k})d^{2k-2} \le c_{15}E(|x|^{k+2})d^{2k-2} = O(d), \end{aligned}$$

where the first inequality is due to the concavity of the function x^{k-1}, $1 < k \le 2$. Using the same argument, we have $E_T(J_4) = O(d)$. Combining the bounds of $I_i, J_i, i = 1, 2, 3, 4$, we obtain

$$E_T(\eta(\beta, d)) = O(d),\ E_T(\eta^2(\beta, d)) = O(d). \tag{29}$$

Combining (27), (28), and (29), Assumptions (N-1)-(N-4) of Huber (1967) are satisfied uniformly in T. Furthermore, $\hat{\beta}(\tau) \xrightarrow{P} \tilde{\beta}(\tau)$ and $\beta_T(\tau) \to \tilde{\beta}(\tau)$ imply $\hat{\beta}(\tau) - \beta_T(\tau) \xrightarrow{P} 0$. Theorem 3 in Huber (1967) makes sure that

$$\sum_{t=1}^{T} g_t(\beta_T(\tau))/T + \sqrt{T}\lambda_T(\hat{\beta}(\tau)) = o_P(1),$$

where $g_t(\beta_T(\tau)) = -\varphi_\tau(y_t - x_t'\beta_T(\tau))x_t$. A mean value expansion of $\lambda_T(\hat{\beta}(\tau))$ around $\tilde{\beta}(\tau)$ provides

$$(\partial\lambda_T(\dot{\beta}(\tau))/\partial\beta)\sqrt{T}(\hat{\beta}(\tau) - \tilde{\beta}(\tau)) = -\sqrt{T}\lambda_T(\tilde{\beta}(\tau)) - \sum_{t=1}^{T} g_t(\beta_T(\tau))/\sqrt{T} + o_P(1),$$

where $\dot{\beta}(\tau)$ between $\hat{\beta}(\tau)$ and $\tilde{\beta}(\tau)$ is the mean value. Combining continuity of $\partial G(k,\beta,\tau)/\partial\beta$ and uniform convergence of $\partial\lambda_T(\beta)/\partial\beta$ ensures $\partial\lambda_T(\dot{\beta}(\tau))/\partial\beta \xrightarrow{P} D$. We can show that results similar to (A.5), (A.6), and (A.7) in Newey (1985) hold with $g_t(\beta_T(\tau))$ in place of $g_T(\theta_0)$, so using the Lindberg-Feller central limit theorem and the Cramer-Wold device yields that $\sum_{t=1}^{T} g_t(\beta_T(\tau))/\sqrt{T}$ converges to $N(0,V)$ in distribution. Using the argument in the proof of Theorem 3 in Newey and Powell (1987), we have $\lim_{T\to\infty} -\sqrt{T}\lambda_T(\tilde{\beta}(\tau)) = K\zeta$. So using Slutsky's Theorem can complete the proof. □

Proof of Theorem 3.3. By Slutsky's Theorem, it is sufficient to show $\hat{D}_j \xrightarrow{P} D_j$, $j = 1,2,\ldots,n$ and $\hat{V}_{ji} \xrightarrow{P} V_{ji}$, $j,i = 1,2,\ldots,n$. Hiding the j subscript, write

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} k(k-1)x_t x_t'\hat{\omega}_t(\tau)|\hat{u}_t(\tau)|^{k-2} - \frac{1}{T}\sum_{t=1}^{T} k(k-1)x_t x_t'\omega_t(\tau)|u_t(\tau)|^{k-2} \\
= \;& \frac{1}{T}\sum_{t=1}^{T} k(k-1)x_t x_t' I(y_t - x_t'\hat{\beta}(\tau) < 0, y_t - x_t'\tilde{\beta}(\tau) < 0) \\
& \cdot(1-\tau)((x_t'\hat{\beta}(\tau) - y_t)^{k-2} - (x_t'\tilde{\beta}(\tau) - y_t)^{k-2}) \\
& + \frac{1}{T}\sum_{t=1}^{T} k(k-1)x_t x_t' I(y_t - x_t'\hat{\beta}(\tau) > 0, y_t - x_t'\tilde{\beta}(\tau) > 0) \\
& \quad \cdot\tau((y_t - x_t'\hat{\beta}(\tau))^{k-2} - (y_t - x_t'\tilde{\beta}(\tau))^{k-2}) \\
& + \frac{1}{T}\sum_{t=1}^{T} k(k-1)x_t x_t' I(y_t - x_t'\hat{\beta}(\tau) \le 0, y_t - x_t'\tilde{\beta}(\tau) \ge 0) \\
& \quad \cdot((1-\tau)(x_t'\hat{\beta}(\tau) - y_t)^{k-2} - \tau(y_t - x_t'\tilde{\beta}(\tau))^{k-2})
\end{aligned}$$

$$
\begin{aligned}
&+ \quad \frac{1}{T}\sum_{t=1}^{T} k(k-1)x_t x_t' I(y_t - x_t'\hat{\beta}(\tau) \geq 0, y_t - x_t'\tilde{\beta}(\tau) \leq 0) \\
&\quad\;\; \cdot(\tau(y_t - x_t'\hat{\beta}(\tau))^{k-2} - (1-\tau)(x_t'\tilde{\beta}(\tau) - y_t)^{k-2}) \\
&=: \quad I_1 + I_2 + I_3 + I_4.
\end{aligned}
$$

Letting $\varepsilon = \{\text{sign}(x_i)\varepsilon_i\}, i = 1, 2, \ldots, p$, we have, with probability approaching one,

$$
\begin{aligned}
I_{1,1}(+\varepsilon) \;&:=\; \frac{1}{T}\sum_{t=1}^{T} k(k-1)x_t x_t' I(y_t - x_t'(\tilde{\beta}(\tau) + \varepsilon) < 0, y_t - x_t'\tilde{\beta}(\tau) < 0) \\
&\quad\; \cdot(1-\tau)(x_t'(\tilde{\beta}(\tau) + \varepsilon) - y_t)^{k-2} \\
&\leq\; \frac{1}{T}\sum_{t=1}^{T} k(k-1)x_t x_t' I(y_t - x_t'\hat{\beta}(\tau) < 0, y_t - x_t'\tilde{\beta}(\tau) < 0) \\
&\quad\; \cdot(1-\tau)(x_t'\hat{\beta}(\tau) - y_t)^{k-2} =: I_{1,1} \\
&\leq\; \frac{1}{T}\sum_{t=1}^{T} k(k-1)x_t x_t' I(y_t - x_t'(\tilde{\beta}(\tau) - \varepsilon) < 0, y_t - x_t'\tilde{\beta}(\tau) < 0) \\
&\quad\; \cdot(1-\tau)(x_t'(\tilde{\beta}(\tau) - \varepsilon) - y_t)^{k-2} =: I_{1,1}(-\varepsilon)
\end{aligned}
$$

Noting $E(x_t x_t' I(y_t - x_t'(\tilde{\beta}(\tau) \pm \varepsilon) < 0, y_t - x_t'\tilde{\beta}(\tau) < 0)(x_t'(\tilde{\beta}(\tau) \pm \varepsilon) - y_t)^{k-2}) < \infty$, Khintchine law of large numbers yields that

$$
\begin{aligned}
I_{1,1}(\pm\varepsilon) \;&\xrightarrow{P}\; k(k-1)(1-\tau) \\
&\quad \cdot E(x_t x_t' I(y_t - x_t'(\tilde{\beta}(\tau) \pm \varepsilon) < 0, y_t - x_t'\tilde{\beta}(\tau) < 0)(x_t'(\tilde{\beta}(\tau) \pm \varepsilon) - y_t)^{k-2}).
\end{aligned}
$$

Similarly,

$$\begin{aligned}
I_{1,2}(+\varepsilon) &:= \frac{1}{T}\sum_{t=1}^{T} k(k-1)x_t x_t' I(y_t - x_t'(\tilde{\beta}(\tau)+\varepsilon) < 0, y_t - x_t'\tilde{\beta}(\tau) < 0) \\
&\quad \cdot(1-\tau)(x_t'\tilde{\beta}(\tau) - y_t)^{k-2} \\
&\leq \frac{1}{T}\sum_{t=1}^{T} k(k-1)x_t x_t' I(y_t - x_t'\hat{\beta}(\tau) < 0, y_t - x_t'\tilde{\beta}(\tau) < 0) \\
&\quad \cdot(1-\tau)(x_t'\tilde{\beta}(\tau) - y_t)^{k-2} =: I_{1,2} \\
&\leq \frac{1}{T}\sum_{t=1}^{T} k(k-1)x_t x_t' I(y_t - x_t'(\tilde{\beta}(\tau)-\varepsilon) < 0, y_t - x_t'\tilde{\beta}(\tau) < 0) \\
&\quad \cdot(1-\tau)(x_t'\tilde{\beta}(\tau) - y_t)^{k-2} =: I_{1,2}(-\varepsilon),
\end{aligned}$$

and similarly,

$$\begin{aligned}
I_{1,2}(\pm\varepsilon) &\xrightarrow{P} k(k-1)(1-\tau) \\
&\quad \cdot E(x_t x_t' I(y_t - x_t'(\tilde{\beta}(\tau)\pm\varepsilon) < 0, y_t - x_t'\tilde{\beta}(\tau) < 0)(x_t'\tilde{\beta}(\tau) - y_t)^{k-2}).
\end{aligned}$$

So, letting $\varepsilon \to 0$, $I_1 \xrightarrow{P} 0$. Using the same argument as the above, $I_2 \xrightarrow{P} 0$.

For any positive δ, write $J(x,\delta) \equiv [x'\tilde{\beta} - \delta|x|, x'\tilde{\beta} + \delta|x|]$ and

$$E_T[I(|u_j(\tau)| \leq \delta|x|)|x] = \int_{J(x,\delta)} f(y|x,\xi_T)dy \leq \int_{J(x,\delta)} \theta(z)dy \equiv \theta_\delta(x).$$

Noting $\theta(z)$ is integrable in y with probability one, we have $\theta_\delta(x)$ converges to zero monotonically with δ by the monotone convergence theorem. According to the continuity of $f(y|x,\xi)$ in Assumption 3, I_3 can be written almost surely as

$$\begin{aligned}
I_3 &= \frac{1}{T}\sum_{t=1}^{T} k(k-1)x_t x_t' I(y_t - x_t'\hat{\beta}(\tau) < 0, y_t - x_t'\tilde{\beta}(\tau) > 0) \\
&\quad \cdot((1-\tau)(x_t'\hat{\beta}(\tau) - y_t)^{k-2} - \tau(y_t - x_t'\tilde{\beta}(\tau))^{k-2}).
\end{aligned}$$

Furthermore, using the preceding ε, with probability approaching one, we have

$$
\begin{aligned}
|I_3| \;\leq\; & \frac{1}{T}\sum_{t=1}^{T} k(k-1)|x_t|^2 I(y_t - x_t'\hat{\beta}(\tau) < 0, y_t - x_t'\tilde{\beta}(\tau) > 0) \\
& \cdot((1-\tau)(x_t'(\tilde{\beta}(\tau) + \varepsilon - y_t))^{k-2} + \tau(y_t - x_t'\tilde{\beta}(\tau))^{k-2}).
\end{aligned}
$$

$$
\begin{aligned}
\leq\; & \frac{1}{T}\sum_{t=1}^{T} k(k-1)|x_t|^2((1-\tau)|x_t'(\tilde{\beta}(\tau) + \varepsilon) - y_t|^{k-2} + \tau|y_t - x_t'\tilde{\beta}(\tau)|^{k-2}) \\
& \cdot I(|u_t(\tau)| \leq |\varepsilon||x_t|) \\
\leq\; & E(k(k-1)|x_t|^2\theta_{|\varepsilon|}(x_t)((1-\tau)|x_t'(\tilde{\beta}(\tau) + \varepsilon) - y_t|^{k-2} + \tau|y_t - x_t'\tilde{\beta}(\tau)|^{k-2})) \\
& + |\varepsilon| \\
\leq\; & cE(|x_t|^2\theta_{|\varepsilon|}(x_t)) + |\varepsilon|,
\end{aligned}
$$

where the third inequality follow from Khintchine law of large numbers and the last inequality is based on the Assumption 4. Since $E(|x_t|^2\theta_{|\varepsilon|}(x_t)) + |\varepsilon|$ converges to zero with $|\varepsilon|$ by the monotone convergence theorem, we have $I_3 \xrightarrow{P} 0$. Using the same argument, we also have $I_4 \xrightarrow{P} 0$. The triangle inequality yields

$$
\begin{aligned}
|\hat{D} - D| \;\leq\; & |\frac{1}{T}\sum_{t=1}^{T} k(k-1)x_t x_t'\hat{\omega}_t(\tau)|\hat{u}_t(\tau)|^{k-2} - \frac{1}{T}\sum_{t=1}^{T} k(k-1)x_t x_t'\omega_t(\tau)|u_t(\tau)|^{k-2}| \\
& + |\frac{1}{T}\sum_{t=1}^{T} k(k-1)x_t x_t'\omega_t(\tau)|u_t(\tau)|^{k-2}| - D|. \qquad (30)
\end{aligned}
$$

The first term in the right-hand side of inequality (30) converges to zero in probability by combining $I_i \xrightarrow{P} 0$, $i = 1, 2, 3, 4$, and the second term converges to zero according to Khintchine law of large numbers. So we have $\hat{D}_j \xrightarrow{P} D_j$. Focusing on $\hat{V}_{ji}$, it is easy to show there are positive constants c_{16}, c_{17}, and c_{18} such that

$$
|\varphi_\tau(y_t - x_t'\beta_1)\varphi_\theta(y_t - x_t'\beta_2)x_t x_t')| \leq |z_t|^{2k}(c_{16} + c_{17}|\beta_1|^{k-1} + c_{18}|\beta_2|^{k-1}).
$$

Using the method in the proof of Theorem 2.2 in Newey (1985) can prove $\hat{V}_{ji} \xrightarrow{P} V_{ji}$. $\square$

Proof of Theorem 3.4. The proof of (i) is the same as Theorem 3.2, for Assumption 6 easily induces Assumption 1. To prove (ii), it is sufficient to prove the case $n = 1$. It is sufficient to prove that $\tilde{D} \to \bar{D}$, $\tilde{V} \to \bar{V}$, and $\tilde{D}^{-1}\tilde{K}\gamma_0 \to \bar{K}$, as $k \to 1$. We mainly focus on

$$(\partial\lambda_T(\dot{\beta}(\tau))/\partial\beta)\sqrt{T}(\hat{\beta}(\tau) - \tilde{\beta}(\tau)) = -\sqrt{T}\lambda_T(\tilde{\beta}(\tau)) - \sum_{t=1}^{T} g_t(\beta_T(\tau))/\sqrt{T} + o_P(1), \quad (31)$$

where $\dot{\beta}(\tau)$ between $\hat{\beta}(\tau)$ and $\tilde{\beta}(\tau)$ is the mean value. Write, f_y being the density of y,

$$\begin{aligned}
\tilde{D} &= E(k(k-1)\omega(\tau)|u(\tau)|^{k-2}xx') \\
&= k(k-1)E\left(xx'\int_R |\tau - I(y - x'\tilde{\beta}(\tau))||y - x'\tilde{\beta}(\tau)|^{k-2} f_y(y)dy\right) \\
&= kE\left(xx'\left(\int_0^\infty (1-\tau)f_y(x'\tilde{\beta}(\tau) - z)dz^{k-1} + \int_0^\infty \tau f_y(x'\tilde{\beta}(\tau) + z)dz^{k-1}\right)\right) \\
&= kE\left(xx'\left(\int_0^\infty (1-\tau)f_\varepsilon\left(\frac{x'(\tilde{\beta}(\tau) - \beta) - z}{1 + x'\frac{\gamma_0}{\sqrt{T}}}\right)\frac{1}{1 + x\frac{\gamma_0}{\sqrt{T}}}dz^{k-1}\right.\right. \\
&\quad \left.\left. + \int_0^\infty \tau f_\varepsilon\left(\frac{x'(\tilde{\beta}(\tau) - \beta) + z}{1 + x'\frac{\gamma_0}{\sqrt{T}}}\right)\frac{1}{1 + x'\frac{\gamma_0}{\sqrt{T}}}dz^{k-1}\right)\right) \\
&\sim kE\left(xx'\left(\int_0^\infty (1-\tau)f_\varepsilon(x'(\tilde{\beta}(\tau) - \beta) - z)dz^{k-1} + \int_0^\infty \tau f_\varepsilon(x'(\tilde{\beta}(\tau) - \beta) + z)dz^{k-1}\right)\right) \\
&= kE\left(xx'\left(\int_0^\infty (1-\tau)f_\varepsilon(q_\varepsilon(\tau) - z)dz^{k-1} + \int_0^\infty \tau f_\varepsilon(q_\varepsilon(\tau) + z)dz^{k-1}\right)\right),
\end{aligned}$$

where the '$\sim$' above is obtained by the dominated convergence theorem as $T \to \infty$. Note that, as $k \to 1$,

$$
\begin{aligned}
& \int_0^\infty ((1-\tau) f_\varepsilon(q_\varepsilon(\tau) - z) + \tau f_\varepsilon(q_\varepsilon(\tau) + z)) dz^{k-1} \\
= \ & ((1-\tau) f_\varepsilon(q_\varepsilon(\tau) - z) + \tau f_\varepsilon(q_\varepsilon(\tau) + z)) z^{k-1}|_0^\infty \\
& - \int_0^\infty z^{k-1} d((1-\tau) f_\varepsilon(q_\varepsilon(\tau) - z) + \tau f_\varepsilon(q_\varepsilon(\tau) + z)) \\
= \ & - \int_0^\infty z^{k-1} d((1-\tau) f_\varepsilon(q_\varepsilon(\tau) - z) + \tau f_\varepsilon(q_\varepsilon(\tau) + z)) \\
\to \ & - \int_0^\infty d((1-\tau) f_\varepsilon(q_\varepsilon(\tau) - z) + \tau f_\varepsilon(q_\varepsilon(\tau) + z)) \\
= \ & f_\varepsilon(q_\varepsilon(\tau)),
\end{aligned}
$$

where the second equality is due to Assumption 7, and the '$\to$' is due to Assumption 8 and the dominated convergence theorem. So we have that $\tilde{D} \to \bar{D}$ as $k \to 1$ hence $\partial \lambda_T(\dot{\beta}(\tau))/\partial\beta \to \bar{D}$. As $k \to 1$,

$$
\begin{aligned}
\tilde{V}_{jk} \ \to \ & E(\varphi_{\tau_j}(u(\tau_j))\varphi_{\tau_k}(u(\tau_k)) x x') \\
= \ & E(xx' E((-1)^{I(y - x'\tilde{\beta}(\tau_j) < 0)} |\tau_j - I(y - x'\tilde{\beta}(\tau_j) < 0)| \\
& \qquad (-1)^{I(y - x'\tilde{\beta}(\tau_k) < 0)} |\tau_k - I(y - x'\tilde{\beta}(\tau_k) < 0)||x)) \\
= \ & E(xx' E((-1)^{I(\varepsilon < q_\varepsilon(\tau_j))} |\tau_j - I(\varepsilon < q_\varepsilon(\tau_j))| \\
& \qquad (-1)^{I(\varepsilon < q_\varepsilon(\tau_k))} |\tau_k - I(\varepsilon < q_\varepsilon(\tau_k))|)) \\
= \ & (\min(\tau_j, \tau_k) - \tau_j \tau_k) E(xx') = \bar{V}_{jk}.
\end{aligned}
$$

Assumption 3 and the dominated convergence theorem yield that

$$
\lim_{k \to 1} -\sqrt{T} \lambda_T(\tilde{\beta}(\tau)) = \sqrt{T} E_T((-1)^{I(y - x'\tilde{\beta}(\tau) < 0)} |\tau - I(y - x'\tilde{\beta}(\tau) < 0)| x). \tag{32}
$$

Letting $\gamma := \frac{\gamma_0}{\sqrt{T}}$, a mean value expansion of the right-hand side of (32) around zero shows that it can be written as, $\tilde{\gamma}$ is the mean value,

$$
\begin{aligned}
& \sqrt{T} E((-1)^{I(y-x'\tilde{\beta}(\tau)<0)}|\tau - I(y - x'\tilde{\beta}(\tau) < 0)|x) \\
+ \quad & \sqrt{T} \frac{\partial(\int_{X\times Y}(-1)^{I(y-x'\tilde{\beta}(\tau)<0)}|\tau - I(y - x'\tilde{\beta}(\tau) < 0)|f_\varepsilon(\frac{y-x'\beta}{1+x'\gamma})\frac{g(x)}{1+x'\gamma}dyd\mu_x)}{\partial\gamma}\bigg|_{\gamma=\tilde{\gamma}} \Big(\frac{\gamma_0}{\sqrt{T}}\Big) \\
= \quad & \frac{\partial(\int_{X\times Y}(-1)^{I(y-x'\tilde{\beta}(\tau)<0)}|\tau - I(y - x'\tilde{\beta}(\tau) < 0)|f_\varepsilon(\frac{y-x'\beta}{1+x'\gamma})\frac{g(x)}{1+x'\gamma}dyd\mu_x)}{\partial\gamma}\bigg|_{\gamma=\tilde{\gamma}} \gamma_0. \qquad (33)
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
& \frac{\partial(\int_{X\times Y}(-1)^{I(y-x'\tilde{\beta}(\tau)<0)}|\tau - I(y - x'\tilde{\beta}(\tau) < 0)|f_\varepsilon(\frac{y-x'\beta}{1+x'\gamma})\frac{g(x)}{1+x'\gamma}dyd\mu_x)}{\partial\gamma}\bigg|_{\gamma=\tilde{\gamma}} \\
= \quad & \int_{X\times Y}(-1)^{I(y-x'\tilde{\beta}(\tau)<0)}|\tau - I(y - x'\tilde{\beta}(\tau) < 0)| \\
& \Big(-f'_\varepsilon\Big(\frac{y-x'\beta}{1+x'\tilde{\gamma}}\Big)\frac{y-x'\beta}{(1+x'\tilde{\gamma})^3}xx' - f_\varepsilon\Big(\frac{y-x'\beta}{1+x'\tilde{\gamma}}\Big)\frac{1}{(1+x'\tilde{\gamma})^2}xx'\Big)g(x)dyd\mu_x \\
= \quad & E\Bigg(xx'\int_{-\infty}^{\infty}(-1)^{I(y-x'\tilde{\beta}(\tau)<0)}|\tau - I(y - x'\tilde{\beta}(\tau) < 0)| \\
& \Bigg(-f'_\varepsilon\Big(\frac{y-x'\beta}{1+x'\tilde{\gamma}}\Big)\frac{y-x'\beta}{(1+x'\tilde{\gamma})^3} - f_\varepsilon\Big(\frac{y-x'\beta}{1+x'\tilde{\gamma}}\Big)\frac{1}{(1+x'\tilde{\gamma})^2}\Bigg)dy\Bigg) \\
=: \quad & E(xx'(I_1 + I_2)). \qquad (34)
\end{aligned}
$$

We have that

$$
\begin{aligned}
\lim_{T\to\infty} I_1 &= \int_{-\infty}^{\infty}(-1)^{1+I(y-x'\tilde{\beta}(\tau)<0)}|\tau - I(y-x'\tilde{\beta}(\tau)<0)|f_\varepsilon'(y-x'\beta)(y-x'\beta)dy \\
&= \int_{-\infty}^{x'\tilde{\beta}(\tau)}(1-\tau)f_\varepsilon'(y-x'\beta)(y-x'\beta)dy - \int_{x'\tilde{\beta}(\tau)}^{\infty}\tau f_\varepsilon'(y-x'\beta)(y-x'\beta)dy \\
&= \int_{-\infty}^{q_\varepsilon(\tau)}(1-\tau)f_\varepsilon'(y)ydy - \int_{q_\varepsilon(\tau)}^{\infty}\tau f_\varepsilon'(y)ydy \\
&= \int_{-\infty}^{q_\varepsilon(\tau)}(1-\tau)ydf_\varepsilon(y) - \int_{q_\varepsilon(\tau)}^{\infty}\tau ydf_\varepsilon(y) \\
&= (1-\tau)yf_\varepsilon(y)|_{-\infty}^{q_\varepsilon(\tau)} - (1-\tau)\int_{-\infty}^{q_\varepsilon(\tau)}f_\varepsilon(y)dy - \tau yf_\varepsilon(y)|_{q_\varepsilon(\tau)}^{\infty} + \int_{q_\varepsilon(\tau)}^{\infty}\tau f_\varepsilon(y)dy \\
&= q_\varepsilon(\tau)f_\varepsilon(q_\varepsilon(\tau)) - \int_{-\infty}^{q_\varepsilon(\tau)}f_\varepsilon(y)dy + \tau, \qquad (35)
\end{aligned}
$$

and

$$
\begin{aligned}
\lim_{T\to\infty} I_2 &= \int_{-\infty}^{\infty}(-1)^{1+I(y-x'\tilde{\beta}(\tau)<0)}|\tau - I(y-x'\tilde{\beta}(\tau)<0)|f_\varepsilon(y-x'\beta)dy \\
&= \int_{-\infty}^{q_\varepsilon(\tau)}(1-\tau)f_\varepsilon(y)dy - \int_{q_\varepsilon(\tau)}^{\infty}\tau f_\varepsilon(y)dy = \int_{-\infty}^{q_\varepsilon(\tau)}f_\varepsilon(y)dy - \tau. \qquad (36)
\end{aligned}
$$

According the dominated convergence theorem, (32)-(36) deduce

$$
\lim_{T\to\infty}\lim_{k\to 1} -\sqrt{T}\lambda(\tilde{\beta}(\tau)) = \gamma_0 q_\varepsilon(\tau)f_\varepsilon(q_\varepsilon(\tau))E(xx').
$$

According to (31) and $\partial\lambda_T(\dot{\beta}(\tau))/\partial\beta \to \bar{D}$, as $k \to 1$, we have that the expectation of $\sqrt{T}(\hat{\beta}(\tau) - \tilde{\beta}(\tau))$ converges to $\gamma_0 q_\varepsilon(\tau)$, i.e., $\tilde{D}^{-1}\tilde{K}\gamma_0 \to \bar{K}$. □

Proof of Theorem 4.1. First we need to prove the asymptotic normality of $\hat{\eta}$, and the consistency of the related covariance matrix estimator like in Theorem 3.3. Then the noncentral chi-square asymptotic distribution of TS follows naturally. It is sufficient to verify that conditions in Theorem 3.2 and Theorem 3.3 are satisfied. Denote $I \equiv (-1/2, 1/2)$, and write, for ς in I,

$$\frac{1}{1+\varsigma} f_\varepsilon\Big(\frac{u}{1+\varsigma}\Big) \leq 2 f_\varepsilon\Big(\frac{u}{1+\varsigma}\Big) \leq 2C/(1+|\frac{2u}{3}|^{k+3+c}).$$

There exists a $2p$-dimension open neighborhood of zero, U_0, such that, for ξ in U_0, $x_t'\xi_{Th} + I(\varepsilon_t > 0)x_t'\xi_{Ts}$ is an element of I with probability one. Noting that

$$f(y|x,\xi) = \frac{1}{1 + x_t'\xi_{Th} + I(\varepsilon_t > 0)x_t'\xi_{Ts}} f_\varepsilon\left(\frac{y - x'\beta_0}{1 + x_t'\xi_{Th} + I(\varepsilon_t > 0)x_t'\xi_{Ts}}\right),$$

the continuity of $f(y|x,\xi)$ in y and ξ holds based the continuity of f_ε. When taking $\theta(z) = 2C/(1+(2|y-x'\beta_0|/3)^{k+3+c})$, we can show that domination conditions of Assumption 3 hold. The first inequality in Assumption 4 is satisfied if ε_t is not equal to ∞ almost surely; the last inequality is satisfied using the simple calculation. In the following, we prove the continuous differentiability of $E(x\varphi_\tau(y - x'\beta(\tau))|\xi)$ and calculate noncentrality parameter of the noncentral chi-squared distribution. The parameter in general can be written as

$$(HD^{-1}K\zeta)'(HD^{-1}VD^{-1}H')^{-1}HD^{-1}K\zeta. \tag{37}$$

Note that, for $\xi := (\xi_{Th}', \xi_{Ts}')' = 0$, $u(\tau) = y_t - x_t'\tilde{\beta}(\tau) = u_t - \mu(\tau)$ is dependent of x_t. We have $D_j = l(\tau_j)L$, hence $D = \mathrm{diag}(l(\tau_1), \ldots, l(\tau_n)) \otimes L$, and $V = (c(\tau_j, \tau_k))_{n\times n} \otimes L$. Using the matrix inversion law of Kronecker products, we have $D^{-1}VD^{-1} = \Omega \otimes L^{-1}$.

Furthermore, $E(-\varphi(u(\tau))x_t|\xi) = E(x_t E(-\varphi(u(\tau))|x_t,\xi))$, and, for $\mu(\tau) > 0$,

$$\begin{aligned}
E(-\varphi(u(\tau))|x_t,\xi) &= E((-1)^{1+I(u_t<\mu(\tau))}k|\tau - I(u_t < \mu(\tau))||u_t - \mu(\tau)|^{k-1}|x_t,\xi) \\
&= (1-\tau)\int_{-\infty}^{0} k\frac{(\mu(\tau)-r)^{k-1}}{\varsigma_{th}} f_\varepsilon\Big(\frac{r}{\varsigma_{th}}\Big)dr \\
&\quad +(1-\tau)\int_{0}^{\mu(\tau)} k\frac{(\mu(\tau)-r)^{k-1}}{\varsigma_{tp}} f_\varepsilon\Big(\frac{r}{\varsigma_{tp}}\Big)dr \\
&\quad -\tau\int_{\mu(\tau)}^{\infty} k\frac{(r-\mu(\tau))^{k-1}}{\varsigma_{tp}} f_\varepsilon\Big(\frac{r}{\varsigma_{tp}}\Big)dr \\
&= (1-\tau)\int_{-\infty}^{0} k(\mu(\tau) - \varsigma_{th}r)^{k-1} f_\varepsilon(r)dr \\
&\quad +(1-\tau)\int_{0}^{\mu(\tau)/\varsigma_{tp}} k(\mu(\tau) - \varsigma_{tp}r)^{k-1} f_\varepsilon(r)dr \\
&\quad -\tau\int_{\mu(\tau)/\varsigma_{tp}}^{\infty} k(\varsigma_{tp}r - \mu(\tau))^{k-1} f_\varepsilon(r)dr,
\end{aligned}$$

where $\varsigma_{th} = 1 + x_t'\xi_{Th}$, and $\varsigma_{tp} = 1 + x_t'\xi_{Th} + x_t'\xi_{Ts}$. We have

$$\partial(E(-\varphi(u(\tau))|x_t,0))/\partial\xi = (\upsilon 1(\tau), \upsilon 2(\tau))' \otimes x_t,$$

which can be dominated by an integrable function, thus

$$\partial(E(-\varphi(u(\tau))x_t|0))/\partial\xi = (\upsilon 1(\tau), \upsilon 2(\tau))' \otimes L. \tag{38}$$

For $\mu(\tau) < 0$, we still have the result of (38) by the similar argument. So the continuous differentiability in Assumption 2 has been examined. We have $D_j^{-1}K_j = (\upsilon 1(\tau_j)/l(\tau_j), \upsilon 2(\tau_j)/l(\tau_j)) \otimes I_p$, and obtain (14) in Theorem 4.1 according to (37). $\square$

Proof of Lemma 1. By Condition A, we have that $Q_{\tau,k}(y_t - x_t'b)f_t(y_t|x_t)g_t(x_t)$ is continuous in $b \in \mathbf{\Theta}$ uniformly in t almost surely. The definition of $Q_{\tau,k}(y_t - x_t'b)$ makes sure that it is measurable for each t and each $b \in \mathbf{\Theta}$. Condition B yields that

$$\begin{aligned}
&\int \sup_{t\geq 1, b\in\mathbf{\Theta}} |Q_{\tau,k}(y_t - x_t'b)| f_t(y_t|x_t) g_t(x_t) dM_{z,t} \\
= &\int \sup_{t\geq 1, b\in\mathbf{\Theta}} |\tau - I(y_t - x_t'b)||y_t - x_t'b|^k f_t(y_t|x_t) g_t(x_t) dM_{z,t} \\
\leq &\; c\int \sup_{t\geq 1, b\in\mathbf{\Theta}} (1 + \|b\|)|z_t|^k f_t(y_t|x_t) g_t(x_t) dM_{z,t} < \infty.
\end{aligned}$$

Define $Q^*_{\tau,k}(z_t, b, r) := \sup\{Q_{\tau,k}(y_t - x_t'\tilde{b}), \tilde{b} \in \delta(b, r)\}$ and $Q_{*\tau,k}(z_t, b, r) := \inf\{Q_{\tau,k}(y_t - x_t'\tilde{b}), \tilde{b} \in \delta(b, r)\}$, with $\delta(b, r) = \{\tilde{b} \in \mathbf{\Theta}, \|\tilde{b} - b\| < \delta\}$. We have $\{Q^*_{\tau,k}(z_t, b, r) \leq y\} = \{\max\{Q_{\tau,k}(y_t - x_t'\tilde{b}), \tilde{b} \in \delta(b, r) \cap Q^p\} \leq y\}$, Q^p being the space of p-dimension rational numbers, as $Q_{\tau,k}(y_t - x_t'\tilde{b})$ is the continuous function of b for any z_t. Thus $Q^*_{\tau,k}(z_t, b, r)$ is a random variable and so do $Q_{*\tau,k}(z_t, b, r)$ using the same argument. These show that Assumptions A1, A2 and A6 in Andrews (1987) are satisfied, and then using his Corollary 3 can complete our proof. □

Proof of Lemma 2. Write $M(b, \tau, T) := \frac{1}{T}\sum_{t=1}^T E(Q_{\tau,k}(y_t - x_t'b))$ and $g_t(b) := \partial E(Q_{\tau,k}(y_t - x_t'b))\partial b$. Obviously, there are positive constants c and d such that $g_t(b) \leq (c + d|b|)|z_t|$. Hence $g_t(b)$ is controlled by an integrable function on a neighborhood of any b according to Condition B. So we can calculate the derivative of $g_t(b)$ as follows.

$$\begin{aligned}
\partial M(b, \tau, T)/\partial b = \;&= \frac{k}{T}\sum_{t=1}^T E\Bigg(x_t\Bigg(-\tau\int_{x_t'b}^{\infty}(y - x_t'b)^{k-1} f_t(y|x_t) dy \\
&\quad +(1-\tau)\int_{-\infty}^{x_t'b}(x_t'b - y)^{k-1} f_t(y|x_t) dy\Bigg)\Bigg) \\
&=: G_T(b).
\end{aligned}$$

Functions $\int_{x_t'b}^{\infty}(y - x_t'b)^{k-1} f_t(y|x_t) dy$ and $\int_{-\infty}^{x_t'b}(x_t'b - y)^{k-1} f_t(y|x_t) dy$ are continuously differentiable in b and their derivatives are controlled uniformly in x_t and in b by integrable functions. Thus the derivative of $G_T(b)$ is

$$\begin{aligned}\partial G_T(b)/\partial b = &= \frac{k(k-1)}{T}\sum_{t=1}^{T} E\bigg(x_t x_t'\bigg(\tau\int_{x_t'b}^{\infty}(y-x_t'b)^{k-2}f_t(y|x_t)dy \\ &\quad +(1-\tau)\int_{-\infty}^{x_t'b}(x_t'b-y)^{k-2}f_t(y|x_t)dy\bigg)\bigg). \qquad (39)\end{aligned}$$

Condition C and (39) make sure that there is a positive constant c such that

$$\partial G_T(b)/\partial b - ck(k-1)\frac{1}{T}\sum_{t=1}^{T} E(x_t x_t')$$

is positive semi-definite for $b \in \mathcal{B}_1$ (a compact subset of R^p) and thus $\partial G_T(b)/\partial b$ is positive definite by Condition E for the same b. So, for $b, \tilde{b} \in \mathcal{B}_1$,

$$\begin{aligned}M(b,\tau,T) - M(\tilde{b},\tau,T) &= G_T(\tilde{b})'(b-\tilde{b}) + (b-\tilde{b})'(\partial G_T(\dot{b})/\partial b)(b-\tilde{b}) \\ &\geq G_T(\tilde{b})'(b-\tilde{b}) + m_e p|b-\tilde{b}|^2, \qquad (40)\end{aligned}$$

where $\dot{b}$ is the mean value and m_e the minimum eigenvalue of $\partial G_T(\dot{b})/\partial b$. The function $M(b,\tau,T)$ is convex as $E(Q_{\tau,k}(y_t - x_t'b))$ is convex with respect to b, and converges to infinity as $|b| \to \infty$. So there is a global minimum $\tilde{\beta}(k,\tau)$ and $G_T(\tilde{\beta}(k,\tau)) = 0$. Letting $\tilde{b} = \tilde{\beta}(k,\tau)$ and using reduction to absurdity (40) shows that the global minimum is also unique one over R^p. $\square$

Acknowledgements

The authors gratefully acknowledge the Opening Project of Sichuan Province University Key Laboratory of Bridge Non-destruction Detecting and Engineering Computing (2018QZJ01) and the talent introduction project of Sichuan University of Science & Engineering (2019RC10).

References

[1] Andrews, Donald W. K. (1987). Consistency in nonlinear econometric models: a generic uniform law of large numbers. Econometrica 55 1465-1471.

[2] Andriyana, Y., Gijbels, I. and Verhasselt, A. (2016). Quantile regression in varying-coefficient models: non-crossing quantile curves and heteroscedasticity. Stat. Pap. 57 1-33.

[3] Anscombe, F. J. (1961). Examination of Residuals. Proceedings of the Fifth Berkeley Symposium. 1 1-36.

[4] Antille, A., Kersting, G. and Zuccini, W. (1982). Testing Symmetry. J. Amer. Statist. Assoc. 77. 639-646.

[5] Bates, C. and White, H. (1985). A unified theory of consistent estimation for parametric models. Economet. Theor. 1 151-178.

[6] Belloni, A. and Chernozhukov, V. (2011). l_1-penalized quantile regression in high-dimensional sparse models. Ann. Stat. 39 82-130.

[7] Boos, D. D. (1982). A Test for Asymmetry Associated With the Hodges-Lehmann Estimator. J. Amer. Statist. Assoc. 77. 647-651.

[8] Breusch, T. S. and Pagan, A. R.(1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. Econometrica 47 1287-1294.

[9] Cabrera, B. L. and Schulz, F. (2017). Forecasting generalized quantiles of electricity demand: A functional data approach. J. Amer. Statist. Assoc. 112, 127-136.

[10] Cai, Z. and Xiao, Z. (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. J. Econmetrics 167 413-425.

[11] Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. J. Amer. Statist. Assoc. 103 1595-1608.

[12] Daouia, A., Girard, S. and Stupfler G. (2018). Estimation of tail risk based on extreme expectiles, J. R. Stat. Soc. Ser. B 80 263-292.

[13] Daouia, A., Girard, S. and Stupfler, G. (2020). Tail expectile process and risk assessment. Bernoulli 26 531-556.

[14] Efron, B. (1991). Regression percentiles using asymmetric squared error loss. Stat. Sinica. 1 93-125.

[15] Engle, R. F. and Manganelli, S. (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. J. Bus. Econ. Stat. 22 367-381.

[16] Farooq, M. and Steinwart, I. (2017). An SVM-like approach for expectile regression. Commun. Statist. Theor. Meth. 109 159-181.

[17] Glesjer, H. (1969). A New Test for Heteroscedasticity. J. Amer. Statist. Assoc. 64 316-323.

[18] Godfrey, L. G. (1978). Testing for Multiplicative Heteroskedasticity. J. Econmetrics 8 227-236.

[19] Goldfeld, S. M. and Quandt, R. E. (1972). Nonlinear Methods in Econometrics. Amsterdam. North-Holland.

[20] Granger, C. W. J. and Sin, C. Y. (1997). Estimating and forecasting quantiles with asymmetric least squares. Working Paper, University of California, San Diego.

[21] Gu, Y. and Zou, H. (2016). High-dimensional generalizations of asymmetric least squares regression and their applications. Ann. Stat. 44 2661-2694.

[22] Harvey, A. C. (1976). Estimating Regression Models with Multiplicative Heteroscedasticity. Econometrica 44 461-466.

[23] He, X. and Shao, Q. M. (2000) On parameters of increasing dimenions. J. Multivariate Anal., 73, 120-135.

[24] Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability 1 221-233.

[25] Jiang, Y., Lin, F. and Zhou, Y. (2019). The kth power expectile regression. Ann. Inst. Stat. Math. https://xs.scihub.ltd/https://doi.org/10.1007/s10463-019-00738-y

[26] Kim, M. O. (2007). Quantile regression with varying coefficients. Ann. Stat. 35 92-108.

[27] Koenker, R. and Jr. Bassett, G. (1978). Regression quantiles. Econometrica 46 33-50.

[28] Koenker, R. and Jr. Bassett, G. (1982). Robust Tests for Heteroscedasticity Based on Regression Quantiles. Econometrica 50 43-61.

[29] Koenker, R. (2005). Quantile regression. No. 38. Cambridge university press, Cambridge.

[30] Koenker, R. (2017). Quantile Regression: 40 Years On. Annu. Rev. Econom. 9 155-176.

[31] Kuan, C. M., Yeh, J. H. and Hsu, Y. C. (2009). Assessing value at risk with care, the conditional autoregressive expectile models. J. Econmetrics 150 261-270.

[32] Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59. 347-370.

[33] Newey, W. K. (1985). Maximum likelihood specification testing and conditional moment tests. Econometrica. 53. 1047-1070.

[34] Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. Econometrica. 55. 819-847.

[35] Prataviera, F., Ortega, E. M. M., Cordeiro, G. M. and Cancho, V. G. (2020). The exponentiated power exponential semiparametric regression model. Commun. Stat.-Simul. C. https://doi.org/10.1080/03610918.2020.1788585.

[36] Prataviera, F., Vasconcelos, J. C. S., Cordeiro, G. M., Hashimoto, E. M. and Ortega, E. M. M.(2019). The exponentiated power exponential regression model with different regression structures: Application in nursing data. J. Appl. Stat. 46. 1792-821.

[37] Taylor, J. W. (2008). Estimating value at risk and expected shortfall using expectiles. J. Financ. Econ. 6 231-252.

[38] Tukey, J. W. (1960). A Survey of Sampling from Contaminated Distributions, in Contributions to Probability and Statistics. ed. by I. Olkin. Stanford: Stanford University Press.

[39] Wang, Z., Liu, X., Tang, W. and Lin, Y. (2020). Incorporating graphical structure of predictors in sparse quantile regression. J. Bus. Econ. Stat. DOI: 10.1080/07350015.2020.1730859.

[40] White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. Econometrica 48 817-838.

[41] Yao, Q. and Tong, H. (1996). Asymmetric least squares regression estimation: A nonparametric approach. J. Nonparametr. Stat. 6 273-292.

[42] Zhao, J., Chen, Y. and Zhang, Y. (2018). Expectile regression for analyzing heteroscedasticity in high dimension. Stat. Probabil. Lett. 137 304-311.