Annals of the Institute of Statistical Mathematics (2021) 73:83–113
https://doi.org/10.1007/s10463-019-00738-y

The *k*th power expectile regression

Yingying Jiang[1,2] · Fuming Lin[1,3] · Yong Zhou[4,5]

Received: 25 April 2019 / Revised: 1 October 2019 / Published online: 12 December 2019

Abstract
Check functions of least absolute deviation make sure quantile regression methods are robust, while squared check functions make expectiles more sensitive to the tails of distributions and more effective for the normal case than quantiles. In order to balance robustness and effectiveness, we adopt a loss function, which falls in between the above two loss functions, to introduce a new kind of expectiles and develop an asymmetric least kth power estimation method that we call the kth power expectile regression, k larger than 1 and not larger than 2. The asymptotic properties of the corresponding estimators are provided. Simulation results show that the asymptotic efficiency of the kth power expectile regression is higher than those of the common quantile regression and expectile regression in some data cases. A primary procedure of choosing satisfactory k is presented. We finally apply our method to the real data.

Keywords Asymptotic variance · The kth power expectile · Expectiles · Quantiles

✉ Fuming Lin
linfuming20062015@163.com

Yingying Jiang
jyy@suse.edu.cn

Yong Zhou
yzhou@amss.ac.cn

1 School of Mathematics and Statistics, Sichuan University of Science and Engineering, Zigong 643000, Sichuan, China

2 School of Statistics, Southwestern University of Finance and Economics, Chengdu 611130, Sichuan, China

3 School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China

4 Academy of Statistics and Interdisciplinary Sciences and School of Statistics, Faculty of Economics and Management, East China Normal University, Shanghai 200241, China

5 Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

1 Introduction

Consider the following linear model

$$y_i = x_i'\beta + \varepsilon_i, \ i = 1, 2, \ldots, n,$$

where β is an unknown $p \times 1$ parameter vector and ε_i is the error term. In order to estimate β, we use the asymmetric kth power loss functions

$$Q_\tau(r) = \begin{cases} (1-\tau)|r|^k & r < 0 \\ \tau|r|^k & r \geq 0 \end{cases} \tag{1}$$

where $\tau \in (0, 1)$. For $k = 1$ and $k = 2$, these loss functions are ones used by the quantile regression and by the expectile regression proposed in Koenker and Bassett (1978) and Newey and Powell (1987), respectively. According to the extreme function estimation theory, minimizing the total asymmetric power error loss

$$S_\tau(b) = \sum_{i=1}^{n}(Q_\tau(y_i - x_i'\beta) - Q_\tau(y_i))$$

obtains an estimator $\widehat{\beta}(\tau, k)$, which provides an estimation of the scalar parameter that minimizes the function $E(Q_\tau(Y - X'\beta) - Q_\tau(Y))$ over β. For convenience, we hide k and write $\widehat{\beta}(\tau, k)$ as $\widehat{\beta}(\tau)$ or $\widehat{\beta}$ without confusion. Hereafter, in other notation that relates to k, we often suppress the symbol k.

Returning to $k = 1$ and $k = 2$, the consistency and asymptotic normality of $\widehat{\beta}$ were proved by Koenker and Bassett (1978) and Newey and Powell (1987). Since then, due to their great advantages, the quantile regression, the expectile regression, and their derivatives have been using by researchers all over the fields of science. For a detailed and systematic introduction to quantile regression and some interesting extensions of basic quantile-based models, we refer to Koenker (2005), Engle and Manganelli (2004), Kim (2007), Cai and Xu (2008), Cai and Xiao (2012), Andriyana et al. (2016) and Koenker (2017), among others. More expectile-based models can be found in Efron (1991), Yao and Tong (1996), Granger and Sin (1997), Taylor (2008), Kuan et al. (2009), Gu and Zou (2016), Farooq and Steinwart (2017), among others. Ehm et al. (2016) considered the problems involving prediction for expectiles. Holzmann and Klar (2016) dealt with the asymptotic distribution of sample expectiles in detail. Chen (1996) considered L_k quantiles and mainly investigated its application in testing symmetry. Arcones (1996) investigated Lp-regression, which is only related to a symmetric check function.

Efron (1991) pointed out that the power loss function with $k = 1.5$ is appealing as a compromise between the robustness of $k = 1$ and the high-normal-theory efficiency of $k = 2$. In contrast to quantile regression, the common expectile regression imposes a condition that the mean of the true distribution underlying data exists. The condition is sometimes strong, especially for some financial data; the kth power expectile regression relaxes the requirement to existence of the $(k-1)$th order moment. Quantile regression requires no moment condition, but the computation involved, for instance calculating the variance, is not so easy as that of the kth power expectile regression. As a trade-off between these aspects, the kth $(1 < k < 2)$ power expectile regression may be a better choice. Moreover, the relevant asymptotic variance of kth power expectile regression is smaller than those of quantile and expectile regression under some data cases. But about the basic properties of the kth power expectile regression, there are yet no consideration. Obviously, it is important to fill the theory gap.

This paper focuses on the loss function in (1) with $1 < k \leq 2$. Our results partially contain those in Newey and Powell (1987) as special cases. We give an explicit definition of the kth power expectile and prove its existence and uniqueness under the assumption that the distribution has the $(k-1)$th order moment. The proof is strongly technical and easy to extend to cases $k > 2$. Some basic properties of kth power expectiles are also studied. Furthermore, we turn our attention to the issues concerned with the kth power expectile regression, such as the asymptotic properties of the kth power expectile regression estimator $\widehat{\beta}(\tau)$. The proofs are different from those for quantile regression. Koenker and Bassett (1978) considered the properties related to regression quantiles. The linear programming formulation and the property of polyhedra were used to prove the uniqueness of the minimum point of $E(Q_\tau(Y-m) - Q_\tau(Y))$, and the algorithm of the estimates also came from the linear programming. The asymptotic properties of the estimators were proved by the basic event probability and Scheffé's theorem (Scheffé 1947). However, for $1 < k \leq 2$, the object function $S_\tau(b)$ is not of the linear and polyhedra, and we mainly use the result of Hjort and Pollard (1993) and the spirit of the methods in Newey and Powell (1987) to prove our theorems. Newey and Powell (1987) used the theory of Huber (1967) to prove their asymptotic normality result, while we use argument in Hjort and Pollard (1993). Because the expressions related to the kth power expectiles are more complex, we have to confront two challenges: How to conceive the proof of the existence and uniqueness of kth power expectiles and how to estimate $E(|y - x'b|^{k-1}|x)$. We use more mathematical techniques in the proof of the first issue and obtain some easy-checking conditions for the second issue. Some comparisons with the quantile regression and the expectile regression have been made in detail, which illustrate the advantage of the general kth power expectile regression. A way of determining the proper value of k is provided. An application, analyzing the data of incomes of migrant workers, is implemented and exemplifies the merits of our proposed method although the method is applicable in much more fields. It is worth mentioning that the empirical results show the kth (especially $k \neq 2$) power expectile regression outweighs the common expectile regression and quantile regression in terms of variances in many cases.

The remainder of the paper is organized as follows: In Sect. 2, basic properties of kth power expectiles and kth power expectile regression have been provided. Main results for the kth power expectile regression estimators and some remarks are presented in Sect. 3. A basic algorithm is provided in Sect. 4. Some comparisons with the quantile regression and the expectile regression are given in Sect. 5. Section 6 contains the method of choosing the value of k, and Sect. 7 involves an application example. Section 8 concludes the paper. All proofs are postponed to Sect. 9. The c_i, c and C are positive and finite constants which may vary from line to line.

2 Basic properties of *k*th power expectiles

Suppose the observable data $\{(y_i, x_i')', i = 1, 2, \ldots, n\}$ come from the linear model

$$y_i = x_i'\beta_0 + \varepsilon_i, \tag{2}$$

where $\{x_i\}$ is a sequence of regression vectors of dimension p with the first component $x_{i1} = 1$, $\beta_0 (\in R^p)$ is a vector of unknown parameters, and $\{\varepsilon_i\}$ is a sequence of scalar error terms. Define the kth power loss function as follows:

$$Q_\tau(r) = \begin{cases} (1-\tau)|r|^k & r < 0 \\ \tau|r|^k & r \geq 0 \end{cases}$$

where $\tau \in (0, 1)$ and $1 < k \leq 2$. We can get an estimator $\widehat{\beta}(\tau)$ via minimizing the total asymmetric power error loss function

$$S_\tau(b) = \sum_{i=1}^{n} (Q_\tau(y_i - x_i'b) - Q_\tau(y_i)). \tag{3}$$

By the extreme function estimation theory, the estimator can be used to estimate the scalar parameter yielded by minimizing the function $E(Q_\tau(Y - X'\beta) - Q_\tau(Y))$ over β if $\{(y_i, x_i')', i = 1, 2, \ldots, n\}$ are independently identically distributed (i.i.d. for short) variables from a population (Y, X'). For the quantile regression or the expectile regression, under homoscedasticity, the probability limits of $\widehat{\beta}(\tau)$ deviate from β_0 only in their intercept terms. Under homoscedasticity and for different choices of τ, we have

$$\text{plim}_{n\to\infty}\widehat{\beta}(\tau) = \beta_0 + \eta(\tau)e_1,$$

where e_j denotes the jth unit vector and $\eta(\tau) \equiv F^{-1}(\tau)$, the quantile or expectile function for the error term ε_i. Under heteroscedasticity, the probability limits for the slope coefficients will generally vary with τ and rely on the joint distribution of ε_i and x_i. The regression quantile estimators are thus a class of empirical "location" measures for the dependent variable whose sampling behavior involves the true regression coefficients and the randomness of the error terms. The general regression of the case $1 < k < 2$ shares the properties above. Below, we give the definition of "expectiles" in the present paper.

We consider the simple scalar parameter $\mu(\tau)$ which minimizes the function $E(Q_\tau(Y - m) - Q_\tau(Y))$ over m, where the expectation is taken with respect to F, the cumulative distribution function of the random variable Y. It is easy to show that the parameter $\mu(\tau)$ is the solution to the equation

$$k(1-\tau)\int_{-\infty}^{\mu(\tau)} (\mu(\tau) - x)^{k-1} \mathrm{d}F(x) = k\tau \int_{\mu(\tau)}^{\infty} (x - \mu(\tau))^{k-1} \mathrm{d}F(x). \tag{4}$$

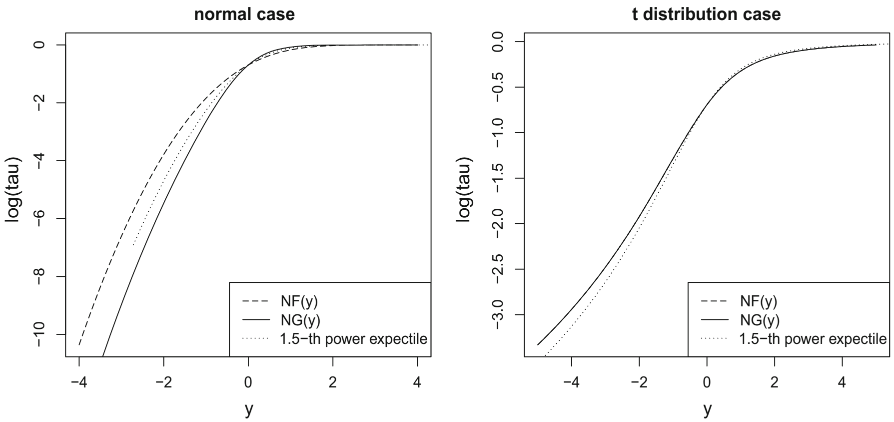

Fig. 1 The cumulative distribution functions, the inverses of expectile functions and the inverses of 1.5 power expectile functions for the normal distribution and the scaled t(2) distribution, which are denoted by NF, NG and 1.5th power expectile, respectively. In order to exhibit the curves well, we set the ordinate scale as $\log(\tau)$

Following Newey and Powell (1987), we refer to $\mu(\tau)$ as the τ kth power expectile. The $\mu(\cdot)$ is also called the kth power expectile function, whose properties are summarized in Theorem 1. The first graph in Fig. 1 depicts the cumulative distribution function, the inverse of the expectile function and the inverse of the 1.5 power expectile function for the normal distribution. The values of the inverse of the 1.5 power expectile function are exactly between those of the cumulative distribution function and the inverse of the expectile function. An appropriate scaled $t(2)$ (a t distribution with 2 degrees of freedom), with $f(x) = (1 + y^2/4)^{-3/2}/4$ as its density distribution, has expectiles equal to its quantiles. So someone will double whether kth power expectiles in this case are still sandwiched between quantiles and expectiles, like the normal case. The second graph in Fig. 1, which contains the same result as the first graph but for the scaled $t(2)$, gives a negative answer. Graphically, the cumulative distribution function is exactly the same as the inverse of the expectile function, as expected, but they are widely different from the inverse of the 1.5 power expectile function. (When k taking other values in the interval (1, 2), the results are similar, so we omitted them.) When considering $k = 1$, the common quantiles, we use α instead of τ to stand for the corresponding probability. Table 1 collects some couples of τ and α such that τ 1.5 power expectiles equal α quantiles for some common distributions. As mentioned above, it can be concluded that kth power expectiles are totally different from quantiles and expectiles.

The kth power expectile $\mu(\tau)$ can be adopted to define a distribution in much the same way as the quantile function or the expectile function does. One-to-one correspondence between these three functions, described in Fig. 1, is the intrinsic property of a distribution. In other words, we can describe a distribution by either its cumulative distribution function or its kth power expectile function. Let I_F denote the set $\{y|0 < F(y) < 1\}$. The following are the basic properties of kth power expectiles and their regression estimators.

Table 1 Implied α values under different distributions

τ	$U(-1,1)$	$N(0,1)$	$t(30)$	$t(10)$	$t(5)$	$t(3)$	$Exp(2)$
0.01	0.044	0.024	0.023	0.021	0.019	0.016	0.053
0.03	0.089	0.061	0.059	0.055	0.050	0.044	0.107
0.05	0.123	0.090	0.088	0.083	0.078	0.069	0.147
0.10	0.187	0.154	0.152	0.147	0.140	0.128	0.223
0.25	0.324	0.302	0.298	0.295	0.289	0.280	0.379

$U(-1,1)$, $N(0,1)$, $t(i)$, and Exp(2) stand for the uniform distribution, the standard normal distribution, the t distribution with i degrees of freedom, and the exponential distribution with parameter 2, respectively

Theorem 1 *Suppose that* $E|Y|^{k-1} < \infty$. *Then for each* τ, $0 < \tau < 1$, *a unique solution* $\mu(\tau)$ *to* (4) *exists and has the following properties:*

(i) *As a function* $\mu(\tau) : (0,1) \to R$, $\mu(\tau)$ *is strictly monotonic increasing.*
(ii) *The range of* $\mu(\tau)$ *is* I_F, *and* $\mu(\tau)$ *maps* $(0,1)$ *onto* I_F.
(iii) *For* $X = sY + t$, *where* $s > 0$, $\mu(\tau)$ *is the* τ *kth power expectile of* Y, *and the* τ *kth power expectile* $\widetilde{\mu}(\tau)$ *of* X *satisfies* $\widetilde{\mu}(\tau) = s\mu(\tau) + t$.

In the regression case, the vector $\widetilde{\beta}_0(\tau)$ that minimizes $R(\beta, \tau) \equiv E(Q_\tau(y_i - x_i'\beta) - Q_\tau(y_i))$ will be determined by the condition distribution of y_i given x_i. Here, the signs $\mathcal{Y}$ and $\mathcal{X}$ denote $(y_1, y_2, \dots, y_n)'$ and $(x_1, x_2, \dots, x_n)$. Return to (3) and define

$$\widehat{\beta}(\tau, \mathcal{Y}, \mathcal{X}) \equiv \operatorname{argmin}_{b \in R^p} S_\tau(b)$$

as an estimator of $\widetilde{\beta}_0(\tau)$. Sometimes, suppress $\mathcal{Y}$ and $\mathcal{X}$ for notational convenience, i.e., write $\widehat{\beta}(\tau, \mathcal{Y}, \mathcal{X})$ as $\widehat{\beta}(\tau)$. The following result about the estimator can be proved.

Theorem 2 *For the solution* $\widehat{\beta}(\tau, \mathcal{Y}, \mathcal{X})$, *we have*

(i) $\widehat{\beta}(\tau, \lambda\mathcal{Y}, \mathcal{X}) = \lambda\widehat{\beta}(\tau, \mathcal{Y}, \mathcal{X}),\ \lambda \in [0, \infty)$;
(ii) $\widehat{\beta}(1-\tau, \lambda\mathcal{Y}, \mathcal{X}) = \lambda\widehat{\beta}(\tau, \mathcal{Y}, \mathcal{X}),\ \lambda \in (-\infty, 0)$;
(iii) $\widehat{\beta}(\tau, \mathcal{Y} + \mathcal{X}'\gamma, \mathcal{X}) = \widehat{\beta}(\tau, \mathcal{Y}, \mathcal{X}) + \gamma,\ \gamma \in R^p$;
(iv) $\widehat{\beta}(\tau, \mathcal{Y}, \mathcal{X}'A) = A^{-1}\widehat{\beta}(\tau, \mathcal{Y}, \mathcal{X}),\ A_{p\times p}$ *is nonsingular.*

3 Large sample properties of estimators

The asymptotic theory for the estimators will be considered under the below assumptions. 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 $|A| \equiv \max_{i,j} |a_{ij}|$.

Assumption 1 For each sample size n, $z_i = (y_i, x_i')'$ $(i = 1, \dots, n)$ are i.i.d. copies of Z and Z has a probability density function $f(y|x)g(x)$ with respect to a measure $\mu_z = l \times \mu_x$ with μ_x being the measure related to $g(x)$.

Assumption 2 $E|Z|^{2+k} < c_0$.

Assumption 3 For $1 < k \leq 2$,

$$0 < c_1 = \inf_{x \in R^p} \inf_{b \in R^p} \int_{-\infty}^{+\infty} |y - x'b|^{k-2} f(y|x)\mathrm{d}y \leq \int_{-\infty}^{+\infty} |y - x'b|^{k-2} f(y|x)\mathrm{d}y < \infty \tag{5}$$

and, for any b,

$$\int_{-\infty}^{+\infty} |y - x'b|^{k-2} f(y|x)\mathrm{d}y/|x|^k < c_2, \text{ as } |x| \to +\infty. \tag{6}$$

Assumption 4 $1.5 < k \leq 2$.

Assumption 5 $E(x_i x_i')$ is nonsingular.

Remark 1 Assumption 3 in fact implies, when $1.5 < k \leq 2$,

$$\int_{-\infty}^{+\infty} |y - x'b|^{2(k-2)} f(y|x)\mathrm{d}y < \infty, \tag{7}$$

and, for any b,

$$\int_{-\infty}^{+\infty} |y - x'b|^{2(k-2)} f(y|x)\mathrm{d}y/|x|^k < c_3, \text{ as } |x| \to +\infty. \tag{8}$$

Remark 2 Assumption 1 is similar to Assumption 1 in Newey and Powell (1987), and we do not consider γ varying in a small neighborhood of γ_0. Assumption 2 is weaker than Assumption 3 in Newey and Powell (1987). Assumptions 3 and 4 are not very restrict, for it must be satisfied when $k = 2$. Assumption 5 is a common restriction in dealing with regression issues.

Remark 3 We suppose that the $f(y|x)$ is bounded. In fact, the first inequality in (5) holds, provided Y is not equal to ∞ almost surely. The last inequality in (5) and the inequality in (7) can be established if the support of Y is finite or $f(y|x) = o(y^s)$, $s < 1 - k$, as $y \to \infty$. The tail properties of (6) and (8) hold for some common distributions. Take the normal distribution for example. We consider the case of $p = 2$, $b = (0, 1)'$, suppose $Y \sim N(\beta_0 x, 1)$, and thus write

$$\begin{aligned}
\int_{-\infty}^{+\infty} |y - x'b|^{k-2} f(y|x)\mathrm{d}y &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} |y - x|^{k-2} \exp(-(y - \beta_0 x)^2/2)\mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} |v|^{k-2} \exp(-(v + (1 - \beta_0)x)^2/2)\mathrm{d}v \\
&= \frac{1}{\sqrt{2\pi}(k-1)} \int_{-\infty}^{+\infty} |v|^{k} \exp(-(v + (1 - \beta_0)x)^2/2)\mathrm{d}v \\
&\quad + \frac{1 - \beta_0}{\sqrt{2\pi}(k-1)} \left(x \int_{0}^{+\infty} v^{k-1} \exp(-(v + (1 - \beta_0)x)^2/2)\mathrm{d}v \right.
\end{aligned}$$

$$-x\int_{-\infty}^{0}(-v)^{k-1}\exp(-(v+(1-\beta_0)x)^2/2)\mathrm{d}v\Big).$$

Hence,

$$\int_{-\infty}^{+\infty}|y-x'b|^{k-2}f(y|x)\mathrm{d}y \leq \frac{1}{k-1}(1+(1-\beta_0)^2x^2)^{k/2}$$
$$+\frac{|(1-\beta_0)x|}{k-1}\left(\sqrt{\frac{2}{\pi}}+\sqrt{\frac{2}{\pi}}|e^{-(1-\beta_0)^2x^2/2}-1|+|(1-\beta_0)x|\right)^{k-1},$$

thus (6) holding. Using the same argument can check the validity of (8).

Theorem 3 *Assume that the data come from* (2). *Write*

$$\widehat{\beta}(\tau)=argmin_{b\in R^p}S_\tau(b),$$

$$\widetilde{\beta}_0(\tau)=argmin_{\beta\in R^p}E(Q_\tau(Y-X'\beta)-Q_\tau(Y)).$$

We have, with ε being homoscedastic,

$$\widetilde{\beta}_0(\tau)=\beta_0+\eta(\tau)e_1,$$

where e_j denotes the jth unit vector and $\eta(\tau)$ is the kth power expectile of ε, i.e., $\eta(\tau)=argmin_{m\in R}E(Q_\tau(\varepsilon-m)-Q_\tau(\varepsilon))$. If Assumptions 1, 2, 3 and 5 hold, then, for each τ in $(0,1)$, a unique solution $\widetilde{\beta}_0(\tau)$ exists and $\widehat{\beta}(\tau)\xrightarrow{P}\widetilde{\beta}_0(\tau)$ as $n\to\infty$.

Remark 4 In (2) and Theorem 3, the homoscedasticity means that ε_i is independent of x_i and only the location of y_i depends on x_i. The property (iii) in Theorem 1 implies that $\eta(\tau,x_i)=x_i'\beta_0+\eta(\tau)$, where $\eta(\tau,x_i)$ is the τ kth power expectile of y_i. The linearity of kth power expectile yields $\eta(\tau,x_i)=x_i'\widetilde{\beta}_0(\tau)$, where $\widetilde{\beta}_0(\tau)=\beta_0+\eta(\tau)e_1$. Only the intercept coefficient in the expression of $\eta(\tau,x_i)$ varies with τ. We set a simple heteroscedasticity case, $\varepsilon_i=x_i'\zeta_i$ with ζ_i independent from x_i'. We have $\eta(\tau,x_i)=x_i'\beta_0+x_i'\eta_1(\tau)=x_i'(\beta_0+\eta_1(\tau))$, where $\eta_1(\tau)$ is the τ kth power expectile of ζ_i. Hence, the slope coefficients in the expression are also related to τ. This property implies that we can develop a test for heteroscedasticity using kth power expectile estimators.

Theorem 4 *If Assumptions 1–5 hold, then $\sqrt{n}(\widehat{\beta}(\tau)-\widetilde{\beta}_0(\tau))$ (hereafter $\widetilde{\beta}_0$ stands for $\widetilde{\beta}_0(\tau)$) is asymptotically normal with mean 0 and covariance matrix $J^{-1}K(J')^{-1}$, where K stands for the covariance matrix of $\Upsilon(X,Y,\widetilde{\beta}_0)=(-1)^{1-I(Y<X'\widetilde{\beta}_0)}X|\tau-I(Y-X'\widetilde{\beta}_0<0)||Y-X'\widetilde{\beta}_0|^{k-1}$, J is the derivative of $E((-1)^{1-I(Y<X'\beta)}X|\tau-I(Y-X'\beta<0)||Y-X'\beta|^{k-1})$ with respect to β at $\widetilde{\beta}_0$, and J' is the transpose of J.*

Writing $u_i(\tau):=y_i-x_i'\widetilde{\beta}_0$ and $w_i(\tau):=|\tau-I(u_i(\tau)<0)|$, we have $J=(k-1)E(w_i(\tau)|u_i(\tau)|^{k-2}x_ix_i')$ and $K=E(x_ix_i'(w_i(\tau))^2(u_i(\tau))^{2(k-1)})$. Writing $\widehat{u}_i(\tau)=y_i-x_i'\widehat{\beta}(\tau)$ and $\widehat{w}_i(\tau)=|\tau-I(\widehat{u}_i(\tau)<0)|$, we can construct estimators $\widehat{J}=$

$(k-1)\sum_{i=1}^{n}\widehat{w}_i(\tau)|\widehat{u}_i(\tau)|^{k-2}x_i x_i'/n$ and $\widehat{K} = \sum_{i=1}^{n} x_i x_i'(\widehat{w}_i(\tau))^2(\widehat{u}_i(\tau))^{2(k-1)}/n$. Using the notation, we have the following result, which makes Theorem 4 feasible.

Theorem 5 *If Assumptions 1–3 and Assumption 5 hold, then*

$$\widehat{J}^{-1}\widehat{K}(\widehat{J}')^{-1} \xrightarrow{P} J^{-1}K(J')^{-1}.$$

Remark 5 Obviously, the results in the two theorems will deduce the corresponding ones of Newey and Powell (1987) when $k = 2$. In fact, Theorem 4 holds for $1.5 < k \leq 2$, and Theorem 5 and the consistency of $\widehat{\beta}(\tau)$ hold for $1 < k \leq 2$.

Remark 6 Chen (1996) used a stronger condition, i.e., $k > 1.5$, in proving his Theorem 3, a result similar to Theorem 5. But he focused on the kernel estimation of the kth power expectile.

4 Algorithm

Consider the objective function (3) with $1 < k \leq 2$. Write $r_i(b) = y_i - x_i'b$. Under the assumption that none of the residuals is such that $r_i(b) = 0$, we can calculate the first and second derivatives of $S_\tau(b)$ as follows.

$$\dot{S}_\tau(b) \equiv k\sum_{i=1}^{n}(-1)^{1-I(y_i<x_i'b)}x_i'|\tau - I(y_i < x_i'b)||y_i - x_i'b|^{k-1},$$

$$\ddot{S}_\tau(b) \equiv \left(\frac{\partial^2 S_\tau(b)}{\partial b_j \partial b_h}\right)_{j,h=1,2,\ldots,p} = k(k-1)\sum_{i=1}^{n} x_i x_i'|\tau - I(y_i < x_i'b)||y_i - x_i'b|^{k-2}.$$

So we reach to a Newton–Raphson updating formula

$$\widehat{\beta}_j(\tau) = \widehat{\beta}_{j-1}(\tau) - (\ddot{S}_\tau(\widehat{\beta}_{j-1}(\tau)))^{-1}\dot{S}_\tau(\widehat{\beta}_{j-1}(\tau)). \tag{9}$$

For any τ $(0 < \tau < 1)$, we choose the estimate of the corresponding least square regression or least absolute regression as the iterative initial value, $\widehat{\beta}_0(\tau)$.

Consider that $\widehat{\beta}(\tau)$ is a function of τ and let τ vary, and another method can be given. When $\widehat{\beta}(\tau)$ has been found, we can get an asymptotic solution for a nearby value $\tau + \Delta\tau$ by

$$\widehat{\beta}(\tau + \Delta\tau) \approx \widehat{\beta}(\tau) + \frac{\mathrm{d}\widehat{\beta}(\tau)}{\mathrm{d}\tau}\Delta\tau, \tag{10}$$

where

$$\frac{\mathrm{d}\widehat{\beta}(\tau)}{\mathrm{d}\tau} \equiv (\ddot{S}_\tau(\widehat{\beta}(\tau)))^{-1}k\sum_{i=1}^{n}(-1)x_i'|\tau - I(y_i < x_i'\widehat{\beta}(\tau))||y_i - x_i'\widehat{\beta}(\tau)|^{k-1}.$$

In fact, we often utilize an algorithm that combines steps (9) and (10). Because we use the condition that none of the residuals is such that $r_i(b) = 0$, there may be some problems in the algorithm in practice, but this happens at almost zero probability if the real data arise from a continuous distribution. The simulation results indicate the algorithm works very well for $k \geq 1.5$, but some calculation difficulty may emerge when k ($1 < k < 1.5$) is very small. Some similar remark can be found in Efron (1991). The "optimize" package in R can be used when $1 < k < 1.5$. A more stable algorithm, such as the MM [the majorize–minimize algorithm proposed by Hunter and Lange (2000)], can also be further developed.

5 Comparisons with the quantile and the expectile regression

In this section, some efficiency comparisons of the kth power expectile regression with the quantile regression and the expectile regression are carried out using simulation data.

5.1 Scale-location models

The expectile regression method is, as we all know, a reasonably efficient way of estimating true regression percentiles in a normal-theory model. In this subsection, we investigate the change of asymptotic efficiencies of estimates in the kth power expectile regression relative to the maximum likelihood estimates (MLE for short), as k takes various values.

We consider a simple model where there are no covariates. The data of n observations are generated by a scale-location family

$$z_i = \mu + \sigma Y_i, \; i = 1, 2, \ldots, n, \tag{11}$$

where $Y_1, Y_2, \ldots, Y_n$ are i.i.d. variables with $E(Y_i) = 0$ and $\text{Var}(Y_i) = 1$ and are drawn from a known probability density function $f(y)$ on the real line. Let $\beta_k^0(\tau)$ stand for true kth power expectiles, i.e., the minimizer of $E(Q_\tau(Y_i - b) - Q_\tau(Y_i))$ over the b. The kth power expectile $\beta_k(\tau)$ of $z = \mu + \sigma Y$, the minimizer over b of $E(Q_\tau(z - b) - Q_\tau(z))$, can be calculated as

$$\beta_k(\tau) = \mu + \sigma \beta_k^0(\tau).$$

The estimator of $\beta_k(\tau)$ is denoted by $\widehat{\beta}_k(\tau)$. We compare the asymptotic variance of $\widehat{\beta}_k(\tau)$ with that of the MLE estimator of $\beta_k(\tau)$. It is useful to give some notation about the MLE estimator. The Fisher information matrix for estimating $\mu, \; \sigma$ in (11) is

$$\mathbf{I}(\mu, \sigma) = \frac{1}{\sigma^2} \begin{pmatrix} i_{11} & i_{12} \\ i_{12} & i_{22} \end{pmatrix},$$

where $i_{11} = Eh(Y)^2$, $i_{12} = Eh(Y)(h(Y)Y)$, $i_{22} = E(h(Y)Y)^2$, and $h(y) = \partial \log f(y)/\partial y$. Let $\widetilde{\mu}$ and $\widetilde{\sigma}$ be the MLEs of μ and σ, respectively. Then, the asymptotic variance $\text{AVAR}(\widetilde{\beta}_k(\tau_k)) \equiv \lim_{n\to\infty} n \cdot \text{var}(\widetilde{\beta}_k(\tau_k))$ of the MLE $\widetilde{\beta}_k(\tau_k) = \widetilde{\mu} + \widetilde{\sigma}\beta_k^0(\tau_k)$ is

$$\frac{\text{AVAR}(\widetilde{\beta}_k(\tau_k))}{\sigma^2} = \frac{i_{22} - 2i_{12}\beta_k^0(\tau_k) + i_{11}(\beta_k^0(\tau_k))^2}{i_{11}i_{22} - i_{12}^2}. \tag{12}$$

According to Newey and Powell (1987), the asymptotic variance of the expectile regression estimator $\widehat{\beta}_2(\tau_2)$ is

$$\frac{\text{AVAR}(\widehat{\beta}_2(\tau_2))}{\sigma^2} = \frac{E(W(Y - \beta_2^0(\tau_2))(Y - \beta_2^0(\tau_2)))^2}{(1 - \tau_2 + (2\tau_2 - 1)P(Y > \beta_2^0(\tau_2)))^2}.$$

By Theorem 4, we deduce the asymptotic variance of the kth power expectile regression estimator $\widehat{\beta}_k(\tau_k)$ as follows.

$$\frac{\text{AVAR}(\widehat{\beta}_k(\tau_k))}{\sigma^2} = \frac{E(W(Y - \beta_k^0(\tau_k))|Y - \beta_k^0(\tau_k)|^{k-1})^2}{(E(\Upsilon'(Y, \beta_k^0(\tau_k))))^2}$$

where

$$\Upsilon'(y, c) = \begin{cases} (k-1)(1-\tau_k)(c-y)^{k-2} & y - c < 0 \\ (k-1)\tau_k(y-c)^{k-2} & y - c \geq 0. \end{cases}$$

$$W\left(Y - \beta_k^0(\tau_k)\right) = |\tau_k - I\left(Y - \beta_k^0(\tau_k) < 0\right)|.$$

The signs $\beta_k^0(\tau_k)$ and $\beta_2^0(\tau_2)$ denote the kth power expectile and the 2 power expectile (the common expectile) of Y_i, respectively. Due to Koenker and Bassett (1982), the asymptotic relative efficiency of $\widehat{\beta}_1(\alpha)$, i.e., the 100$\alpha$th sample percentile of $z_1, z_2, \ldots, z_n$, is

$$\frac{\text{AVAR}(\widehat{\beta}_1(\alpha))}{\sigma^2} = \frac{\alpha(1-\alpha)}{f(y^{(\alpha)})^2},$$

where $y^{(\alpha)}$ is the α quantile of Y_i. We suppose that Y_i in (11) comes from one of three types of distributions: the standard normal distribution with the density function $f(y) = \phi(y)$, the student distribution with 3 degrees of freedom ($t(3)$ for short) with $f(y) = 6\sqrt{3}/(\pi(3 + y^2)^2)$ and the Chi-square distribution with 6 degrees of freedom ($\chi^2(6)$ for short) with $f(y) = y^2 \exp(-y/2)/2^4$. Motivation for choosing these distributions is that the normal distribution is one of the most frequently used distributions, the t distribution is a typical heavy-tail distribution wildly used in many fields, such as in finance, and the Chi-square distribution is taken as a representative of skew distributions. For each case, we will give the efficiency change of our method in pace with k. There is a relation among $\beta_k^0(\tau_k)$, $\beta_2^0(\tau_2)$ and $y^{(\alpha)}$ in the term: for every α, we set suitable τ_k and τ_2 such that $\beta_k^0(\tau_k) = y^{(\alpha)}$ and $\beta_2^0(\tau_2) = y^{(\alpha)}$ using (4). For the normal distribution, the t distribution and the $\chi^2(6)$ distribution, (12) can be further written as

$$\frac{\mathrm{AVAR}(\widetilde{\beta}_k(\tau_k))}{\sigma^2} = 1 + (y^{(\alpha)})^2/3,$$
$$\frac{\mathrm{AVAR}(\widetilde{\beta}_k(\tau_k))}{\sigma^2} = (3 + (y^{(\alpha)})^2)/2$$

and

$$\frac{\mathrm{AVAR}(\widetilde{\beta}_k(\tau_k))}{\sigma^2} = ((y^{(\alpha)})^2 - 4y^{(\alpha)} + 16)/3,$$

respectively. Divide the asymptotic variances of MLE by the asymptotic variances of expectiles, kth power expectiles and quantiles, respectively, and write

$$\mathrm{ARE}_{\mathrm{SL}}(\widehat{\beta}_2(\tau_2)) := \frac{\mathrm{AVAR}(\widetilde{\beta}_2(\tau_2))}{\mathrm{AVAR}(\widehat{\beta}_2(\tau_2))}$$
$$\mathrm{ARE}_{\mathrm{SL}}(\widehat{\beta}_k(\tau_k)) := \frac{\mathrm{AVAR}(\widetilde{\beta}_2(\tau_2))}{\mathrm{AVAR}(\widehat{\beta}_k(\tau_k))}$$
$$\mathrm{ARE}_{\mathrm{SL}}(\widehat{\beta}_1(\alpha)) := \frac{\mathrm{AVAR}(\widetilde{\beta}_2(\tau_2))}{\mathrm{AVAR}(\widehat{\beta}_1(\alpha))}.$$

Tables 2, 3, 4 include values of $\mathrm{ARE}_{\mathrm{SL}}(\widehat{\beta}_2(\tau_2))$, $\mathrm{ARE}_{\mathrm{SL}}(\widehat{\beta}_k(\tau_k))$ and $\mathrm{ARE}_{\mathrm{SL}}(\widehat{\beta}_1(\alpha))$ for three distributions under various values of α. For the normal distribution, the efficiency of $\widehat{\beta}_k(\tau_k)$ is higher than $\widehat{\beta}_1(\alpha)$ and increases as k varies from 1 to 2. And $\widehat{\beta}_{1.9}(\tau_{1.9})$ almost shares the same efficiency with $\widehat{\beta}_2(\tau_2)$, i.e., the common expectile regression estimator. The efficiency change goes much faster when α approaches 0.5 more closely. There is a completely different picture for the $t(3)$ distribution. The efficiency of $\widehat{\beta}_{1.3}(\tau_{1.3})$ is highest for $\alpha = 0.5$ and $\alpha = 0.67/0.33$; the efficiency of $\widehat{\beta}_{1.2}(\tau_{1.2})$ is highest for $\alpha = 0.75/0.25$, $\alpha = 0.84/0.16$, $\alpha = 0.90/0.10$ and $\alpha = 0.95/0.05$. Significantly, the efficiency of $\widehat{\beta}_2(\tau_2)$ is lowest and keeps pace with almost half of that of $\widehat{\beta}_{1.3}(\tau_{1.3})$. In the Chi-square distribution case, for $\alpha = 0.5$, the efficiency decreases with k varying from 2 to 1; for other value of α, the efficiency first increases and then decreases when k takes its value from 2 to 1. Figure 2 further depicts these results.

Generally, for scale-location models, the kth power expectile regression ($1 < k < 2$) may produce more efficient estimators for non-normal distributions in some quantile cases.

Table 2 Asymptotic relative efficiency of $\widehat{\beta}_2(\tau_2)$, $\widehat{\beta}_k(\tau_k)$ and $\widehat{\beta}_1(\alpha)$ for the normal case

α	.50	.67 or .33	.75 or .25	.84 or .16	.90 or .10	0.95 or 0.05
$\mathrm{ARE_{SL}}(\widehat{\beta}_2(\tau_2))$	1.000	0.980	0.947	0.865	0.752	0.568
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.9}(\tau_{1.9}))$	0.997	0.954	0.896	0.779	0.646	0.463
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.8}(\tau_{1.8}))$	0.989	0.950	0.897	0.786	0.657	0.476
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.7}(\tau_{1.7}))$	0.975	0.941	0.891	0.788	0.664	0.486
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.6}(\tau_{1.6}))$	0.955	0.924	0.880	0.784	0.667	0.494
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.5}(\tau_{1.5}))$	0.927	0.900	0.861	0.774	0.664	0.498
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.4}(\tau_{1.4}))$	0.890	0.868	0.834	0.757	0.655	0.497
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.3}(\tau_{1.3}))$	0.843	0.826	0.798	0.731	0.639	0.491
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.2}(\tau_{1.2}))$	0.786	0.773	0.751	0.694	0.614	0.478
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.1}(\tau_{1.1}))$	0.718	0.709	0.692	0.647	0.578	0.457
$\mathrm{ARE_{SL}}(\widehat{\beta}_1(\alpha))$	0.636	0.631	0.620	0.585	0.529	0.425

Table 3 Asymptotic relative efficiency of $\widehat{\beta}_2(\tau_2)$, $\widehat{\beta}_k(\tau_k)$ and $\widehat{\beta}_1(\alpha)$ for the $t(3)$ distribution

α	.50	.67 or .33	.75 or .25	.84 or .16	.90 or .10	0.95 or 0.05
$\mathrm{ARE_{SL}}(\widehat{\beta}_2(\tau_2))$	0.500	0.416	0.329	0.212	0.129	0.062
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.9}(\tau_{1.9}))$	0.597	0.503	0.404	0.265	0.164	0.080
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.8}(\tau_{1.8}))$	0.686	0.585	0.475	0.318	0.200	0.099
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.7}(\tau_{1.7}))$	0.766	0.659	0.541	0.368	0.235	0.118
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.6}(\tau_{1.6}))$	0.832	0.722	0.598	0.413	0.267	0.136
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.5}(\tau_{1.5}))$	0.883	0.772	0.645	0.452	0.296	0.153
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.4}(\tau_{1.4}))$	0.915	0.805	0.679	0.482	0.320	0.167
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.3}(\tau_{1.3}))$	0.927	0.820	0.696	0.501	0.337	0.179
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.2}(\tau_{1.2}))$	0.915	0.814	0.696	0.507	0.346	0.187
$\mathrm{ARE_{SL}}(\widehat{\beta}_{1.1}(\tau_{1.1}))$	0.876	0.785	0.676	0.499	0.344	0.189
$\mathrm{ARE_{SL}}(\widehat{\beta}_1(\alpha))$	0.810	0.729	0.633	0.473	0.331	0.185

Table 4 Asymptotic relative efficiency of $\widehat{\beta}_2(\tau_2)$, $\widehat{\beta}_k(\tau_k)$ and $\widehat{\beta}_1(\alpha)$ for the $\chi^2(6)$ distribution

α	.50	.67 or .33	.75 or .25	.84 or .16	.90 or .10	0.95 or 0.05
$\text{ARE}_{\text{SL}}(\widehat{\beta}_2(\tau_2))$	0.795	0.733	0.666	0.548	0.429	0.285
$\text{ARE}_{\text{SL}}(\widehat{\beta}_{1.9}(\tau_{1.9}))$	0.789	0.736	0.675	0.561	0.443	0.297
$\text{ARE}_{\text{SL}}(\widehat{\beta}_{1.8}(\tau_{1.8}))$	0.778	0.734	0.678	0.570	0.454	0.308
$\text{ARE}_{\text{SL}}(\widehat{\beta}_{1.7}(\tau_{1.7}))$	0.762	0.728	0.678	0.575	0.463	0.318
$\text{ARE}_{\text{SL}}(\widehat{\beta}_{1.6}(\tau_{1.6}))$	0.740	0.716	0.671	0.577	0.469	0.325
$\text{ARE}_{\text{SL}}(\widehat{\beta}_{1.5}(\tau_{1.5}))$	0.713	0.697	0.659	0.572	0.470	0.330
$\text{ARE}_{\text{SL}}(\widehat{\beta}_{1.4}(\tau_{1.4}))$	0.679	0.671	0.639	0.562	0.467	0.332
$\text{ARE}_{\text{SL}}(\widehat{\beta}_{1.3}(\tau_{1.3}))$	0.639	0.638	0.612	0.545	0.458	0.331
$\text{ARE}_{\text{SL}}(\widehat{\beta}_{1.2}(\tau_{1.2}))$	0.591	0.596	0.577	0.519	0.441	0.324
$\text{ARE}_{\text{SL}}(\widehat{\beta}_{1.1}(\tau_{1.1}))$	0.535	0.545	0.531	0.485	0.417	0.311
$\text{ARE}_{\text{SL}}(\widehat{\beta}_1(\alpha))$	0.470	0.484	0.476	0.440	0.383	0.290

5.2 Location shift models

In the subsection, we study a simple stochastic linear model

$$y_i = a + bx_i + \varepsilon_i,$$

where x_i, ε_i are i.i.d. copies of variables X and ε, and a, b are constants. The explained variable and explanatory variables are y_i and x_i. The model is designed to investigate the impact of k on the asymptotic variances of $\hat{a}$ and $\hat{b}$, the estimators of a and b. Let X obey the uniform distribution, $U[-1, 1]$, and ε be either the normal or the t(3) random variable. We assume that $a = 20$ and $b = 100$. The assumption just suits the convenience of the simulation, for the values of a, b have no influence on the asymptotic variances if ε is distributed symmetrically. For a given value of α ($\in (0, 1)$), the following steps are used to calculate the asymptotic variances. For the quantile regression, we adopt the formula $\alpha_0(1-\alpha_0)(f_\varepsilon(q_\varepsilon(\alpha_0)))^{-2}E((1, X)'(1, X))$ with f_ε being the density function of ε. For the kth power expectile regression, we firstly obtain the kth power expectile of ε, denoted by $\mathcal{X}_\varepsilon(k, \alpha_0)$, and secondly use the expression of the asymptotic variance in Theorem 4 to complete the calculation.

The asymptotic variances are summarized in Tables 5 and 6, Table 5 for the t distribution and Table 6 for the normal. In the first columns of two tables, a_j, b_j, $j = 1.1, \ldots, 2$ indicate that the elements following them are obtained using the jth power expectile regression and a_1, b_1 show the elements after they come from the quantile regression. As well known in the literature, Table 6 shows the expectile regression, i.e., $k = 2$, has the least variances no matter which value α takes. But for

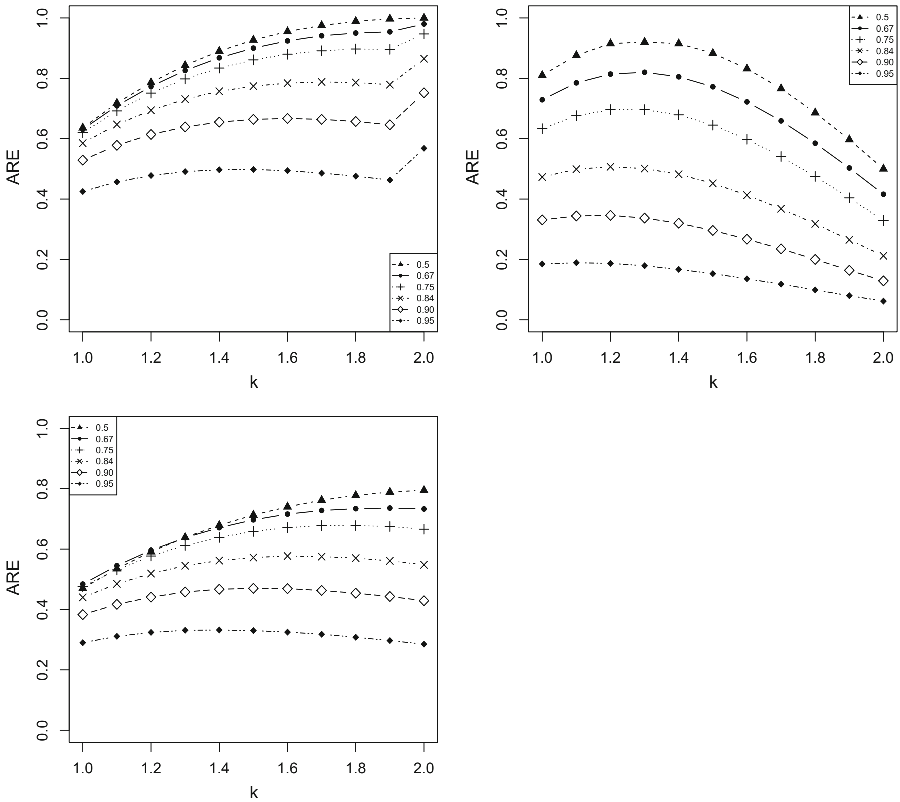

Fig. 2 Asymptotic relative efficiency of $\widehat{\beta_2}(\tau_2)$, $\widehat{\beta_k}(\tau_k)$ and $\widehat{\beta_1}(\alpha)$ for three distributions, the top left graph for the normal, the top right one for the t distribution and the bottom one for the χ^2

the t distribution, there is a new and thought-provoking phenomenon. The variances of the kth power expectile regression are smaller than those of quantile and expectile regressions when $k = 1.4, 1.5, 1.6$ and α takes any values. Taking $k = 1.4, \alpha = 0.90$, we find the variance of the kth power expectile regression estimator of b is 12.54, less than half of the corresponding variance of the quantile regression estimator. So, this implies that the wildly used quantile regression does not always outweigh the kth power expectile regression in terms of asymptotic variance.

6 The method of choosing suitable *k*

We investigate the way of choosing suitable k to gain more efficient regression methods for real data. When coming to the problem, too many issues need to be dealt with rigorously, such as the specification test of the relation between the kth power expectile and some covariates (parametric or nonparametric and linear or nonlinear) and variable selection. But for easy accessibility, we suppose that

Table 5 Asymptotic variances of estimators in linear model when the error term obeys the *t* distribution with freedom degree 3

α	0.05	0.10	0.16	0.25	0.33	0.5	0.67	0.75	0.84	0.90	0.95
a_2	22.84	9.68	5.63	3.71	3.02	2.88	3.45	4.39	6.88	11.88	31.66
b_2	59.04	26.57	15.76	10.51	8.71	8.75	10.26	12.90	20.14	34.76	95.04
$a_{1.9}$	24.17	9.88	5.72	3.64	2.90	2.45	2.92	3.72	5.74	9.82	26.07
$b_{1.9}$	77.25	31.28	18.31	11.45	9.00	7.42	8.68	10.93	16.85	28.62	78.36
$a_{1.8}$	21.13	8.54	4.97	3.17	2.52	2.14	2.54	3.26	4.98	8.28	21.94
$b_{1.8}$	67.15	26.82	15.91	9.89	7.77	6.46	7.54	9.56	14.64	24.21	64.80
$a_{1.7}$	19.30	7.57	4.46	2.84	2.24	1.92	2.25	2.94	4.42	7.24	18.77
$b_{1.7}$	59.66	23.62	14.03	8.79	6.83	5.77	6.74	8.62	12.99	20.76	59.09
$\underline{a_{1.6}}$	17.06	6.87	4.12	2.65	2.09	1.75	2.08	2.74	4.10	6.89	17.29
$\underline{b_{1.6}}$	51.70	20.82	12.89	8.13	6.33	5.26	6.30	7.99	12.06	19.54	49.21
$\underline{a_{1.5}}$	15.75	6.61	3.82	2.55	2.01	1.63	1.95	2.49	3.89	6.25	17.41
$\underline{b_{1.5}}$	54.14	20.02	11.61	7.95	6.09	4.88	5.81	7.29	11.74	17.89	50.10
$\underline{a_{1.4}}$	18.71	3.43	3.77	2.35	1.90	1.55	1.73	2.54	4.36	3.88	17.18
$\underline{b_{1.4}}$	64.20	14.92	12.35	6.43	5.15	4.58	5.89	7.81	12.08	12.54	49.94
$a_{1.3}$	14.80	8.08	4.21	2.02	2.23	1.52	1.95	2.89	4.73	5.78	21.95
$b_{1.3}$	44.52	20.57	14.96	4.55	6.81	4.37	5.60	9.03	11.40	14.42	60.12
$a_{1.2}$	31.94	8.58	4.95	3.54	2.99	1.64	2.72	1.40	6.82	10.23	23.69
$b_{1.2}$	79.59	42.34	11.65	9.78	7.92	4.50	9.12	9.03	19.45	34.31	80.11
$a_{1.1}$	64.57	9.19	9.99	5.51	5.60	2.66	4.20	6.85	3.94	30.75	46.63
$b_{1.1}$	316.79	42.67	50.53	18.73	14.16	6.81	14.86	15.83	43.20	88.76	274.46
a_1	23.07	8.57	4.66	2.83	2.21	1.85	2.21	2.83	4.66	8.57	23.07
b_1	69.40	25.79	14.01	8.51	6.66	5.56	6.66	8.51	14.01	25.79	69.40

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \varepsilon,$$

where the covariates $(1, x_1, \ldots, x_p)'$ have been recognized. And we only consider the value choosing of k in the kth power expectile regression under this setting. At present, we can calculate $\widehat{J}^{-1}\widehat{K}(\widehat{J}')^{-1}$ in Theorem 5 for each value of the couple of k and α (the value range of k being in $\{1 + i/n, i = 1, 2, \ldots, n-1\}$, for n large enough) and find a favorable k value, for example k_0, such that the sum of diagonal elements of

Table 6 Asymptotic variances of estimators in linear model when the error term obeys the normal distribution

α	0.05	0.10	0.16	0.25	0.33	0.5	0.67	0.75	0.84	0.90	0.95
$\underline{a_2}$	2.15	1.61	1.34	1.16	1.08	1.01	1.05	1.11	1.26	1.48	1.99
$\underline{b_2}$	5.97	4.67	3.87	3.39	3.16	3.00	3.13	3.31	3.81	4.52	6.51
$a_{1.9}$	2.19	1.63	1.34	1.17	1.08	1.01	1.05	1.12	1.27	1.50	2.05
$b_{1.9}$	6.12	4.75	3.91	3.44	3.18	2.99	3.13	3.34	3.87	4.54	6.63
$a_{1.8}$	2.23	1.68	1.37	1.19	1.09	1.01	1.06	1.13	1.29	1.55	2.16
$b_{1.8}$	6.23	4.91	3.99	3.49	3.23	3.00	3.13	3.35	3.88	4.63	6.91
$a_{1.7}$	2.31	1.73	1.41	1.21	1.10	1.02	1.07	1.14	1.32	1.57	2.14
$b_{1.7}$	6.29	5.01	4.11	3.56	3.28	3.02	3.19	3.39	3.93	4.61	7.10
$a_{1.6}$	2.26	1.85	1.43	1.24	1.12	1.03	1.10	1.18	1.39	1.64	2.28
$b_{1.6}$	6.28	5.36	4.18	3.59	3.37	3.07	3.24	3.39	4.21	5.06	7.34
$a_{1.5}$	2.91	1.96	1.44	1.29	1.02	1.04	1.12	1.17	1.51	1.81	2.68
$b_{1.5}$	7.78	5.72	4.12	3.76	3.58	3.16	3.33	3.46	4.37	5.74	8.23
$a_{1.4}$	3.36	2.24	1.72	1.44	1.26	1.09	1.26	1.24	1.57	2.13	3.16
$b_{1.4}$	9.07	6.74	4.69	4.29	3.91	3.34	3.76	3.61	4.57	6.51	10.64
$a_{1.3}$	3.96	2.43	1.93	1.35	1.42	1.20	1.47	1.56	2.02	1.83	4.17
$b_{1.3}$	12.87	9.05	5.58	3.32	4.16	3.74	4.47	4.50	5.59	5.19	14.33
$a_{1.2}$	6.47	2.20	1.74	1.93	2.11	1.53	1.79	1.72	1.81	3.41	6.72
$b_{1.2}$	17.78	8.53	8.56	6.34	6.63	4.89	4.70	5.30	1.97	13.38	19.82
$a_{1.1}$	11.60	5.48	3.31	4.54	1.25	3.19	2.20	4.42	7.33	10.72	8.82
$b_{1.1}$	17.54	8.53	20.06	17.42	5.57	10.37	3.40	13.44	21.39	34.36	13.81
a_1	4.60	3.82	3.10	2.44	2.10	1.85	2.10	2.44	3.10	3.82	4.60
b_1	13.83	11.49	9.35	7.34	6.32	5.56	6.32	7.34	9.35	11.49	13.83

$\widehat{J}^{-1}\widehat{K}(\widehat{J}')^{-1}$ of the corresponding k_0th power expectile regression is smallest. Hence, the k_0th power expectile regression is our desirable method. The skeleton of procedures is given as follows.

(i) For a given α, run a quantile regression to get estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$, which estimate $\beta_0 + q_\varepsilon(\alpha), \beta_1, \ldots, \beta_p$ consistently. We use the R package "quantreg" to carry out the regressions and obtain the standard deviations of estimators. For the comparison, we multiply these standard deviations by $\sqrt{n}$, then square them, finally calculate the sum of diagonal elements of the matrices.

(ii) For a given k, using the equation in (4) and the data y, we can obtain θ_α such that the θ_α kth power expectile of y is equal to the α quantile of y.

(iii) We use the R function "optim" to complete the θ_α kth power expectile regression and then utilize the expression in Theorem 5 to calculate variances and then the sums of diagonal elements of them.

(iv) Compare the sums in (i) and those in (iii) for the different values of k and determine the proper value of k.

We take advantage of the simulated data in Sect. 5.2 to demonstrate the method. The sample size is 10,000, partition intervals (0, 1), (1, 2) by sequences $j/20$, $j = 1, 2, \ldots, 19$ and $1 + i/10$, $i = 1, 2, \ldots, 9$, respectively. Using every split points as the values of α and k, we repeat the first three steps and acquire the sums of corresponding variances. For the (iii) step, we adopt the OLS estimates as initial values to run "optim". Results are contained in Tables 7 and 8, where the values in parenthesis are the asymptotic variances. The estimators of b with smaller variances are highlighted by underlines. The smallest variances do not always appear in the quantile regression. We find the suitable values of k: $k = 1.3$ for $\alpha = 0.4$, $k = 1.5$ for $\alpha = 0.45$, $k = 1.4$ for $\alpha = 0.5$, $k = 1.3$ for $\alpha = 0.55, 0.6$, and $k = 1.2$ for $\alpha = 0.75$.

7 Real data example

For the empirical study, we utilize the method in Sect. 6 and analysis the data of incomes of migrant workers. The data come from a survey of incomes of migrant workers in China conducted by State Statistical Bureau of China at 2011. Ultimately, we get 3372 effective observations. There are three variables: annual incomes y, years of education x_1 and working years x_2. According to Mincer (1958), in our setting, the explained variable is $\log(y)$ and three explanatory variables are x_1, x_2 and x_2^2. The model we used is

$$\log(y) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_2^2 + \varepsilon,$$

where ε is the error term. We suppose α takes values in the range: $\{\alpha = 0.05 + i * 0.05, i = 0, 1, \ldots, 18\}$. Consider the following values of k: $1 + i/10$, $i = 1, 2, \ldots, 10$ in $(1, 2]$. By detailed comparisons, the suitable k is obtained for each α value using the method in Sect. 6, and relevant results are gathered in Table 9. Letting k take its values in $[1, 2]$ densely enough, we can obtained an optimal value of k for each α.

The table contains estimates, variances and the significant test results as well. Estimates and standard deviations are put in columns, but the first column from left and values enclosed in parentheses are the standard deviations of the estimates over them. The real values are ones in columns multiplied by the corresponding powers in the first column. From the results, we find the working years and their squares have significant impact on the incomes at the level 0.01 when α takes the value: 0.15, 0.25, 0.30, 0.35, 0.40, 0.45 or 0.50, but education has little influence on the incomes except $\alpha = 0.4$. When α gets large value: 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80,0.85, 0.90 or 0.95, the

Table 7 The suitable k for simulation data where the error term obeys the t distribution with freedom degree 3

α	k																					
	1		1.1		1.2		1.3		1.4		1.5		1.6		1.7		1.8		1.9		2	
	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$
0.05	17.61	100.03	17.41	100.08	17.17	100.14	16.91	100.21	16.62	100.34	16.29	100.46	15.96	100.53	15.59	100.59	15.22	100.64	14.79	100.71	14.25	100.83
	(24.10)	(72.31)	(96.11)	(323.41)	(93.93)	(152.48)	(77.03)	(233.20)	(118.11)	(464.76)	(183.91)	(558.24)	(264.70)	(669.39)	(313.96)	(898.30)	(576.13)	(2230.16)	(1041.28)	(3529.09)	(2320.68)	(7346.52)
0.1	18.34	100.06	18.23	100.07	18.09	100.08	17.95	100.08	17.80	100.10	17.63	100.12	17.45	100.17	17.24	100.23	17.02	100.30	16.78	100.39	16.52	100.47
	(8.01)	(23.93)	(20.26)	(51.62)	(17.37)	(62.73)	(14.59)	(45.69)	(21.72)	(78.13)	(31.00)	(97.22)	(40.31)	(112.77)	(66.77)	(201.18)	(92.38)	(291.70)	(153.75)	(472.29)	(233.17)	(731.92)
0.15	18.73	100.01	18.65	100.03	18.57	100.04	18.47	100.06	18.38	100.07	18.28	100.08	18.16	100.09	18.04	100.11	17.90	100.14	17.76	100.18	17.60	100.22
	(5.36)	(16.12)	(11.00)	(32.35)	(9.85)	(26.25)	(8.30)	(27.09)	(8.75)	(26.21)	(11.40)	(36.26)	(14.39)	(43.43)	(19.37)	(62.78)	(28.01)	(90.11)	(40.88)	(133.45)	(62.39)	(193.33)
0.2	19.01	100.00	18.95	100.02	18.88	100.02	18.82	100.02	18.74	100.03	18.67	100.04	18.59	100.05	18.50	100.07	18.41	100.09	18.31	100.11	18.20	100.13
	(3.76)	(11.27)	(7.49)	(26.01)	(6.99)	(18.00)	(4.49)	(13.52)	(5.32)	(17.58)	(4.63)	(19.46)	(7.29)	(22.67)	(9.36)	(29.68)	(12.24)	(38.52)	(17.42)	(55.11)	(24.95)	(78.89)
0.25	19.23	99.99	19.18	100.00	19.14	100.00	19.09	100.01	19.04	100.01	18.99	100.02	18.93	100.03	18.87	100.04	18.80	100.05	18.73	100.06	18.66	100.08
	(2.82)	(8.49)	(4.07)	(16.10)	(4.25)	(15.71)	(3.08)	(8.43)	(3.68)	(12.22)	(3.42)	(11.24)	(4.47)	(14.11)	(5.44)	(16.82)	(6.96)	(22.23)	(8.89)	(28.35)	(11.84)	(37.94)
0.3	19.41	99.98	19.38	99.98	19.35	99.99	19.31	99.99	19.28	100.00	19.24	100.00	19.20	100.01	19.16	100.02	19.11	100.02	19.07	100.03	19.02	100.05
	(2.31)	(6.95)	(3.74)	(6.94)	(3.50)	(9.34)	(2.14)	(7.08)	(2.24)	(8.22)	(2.73)	(9.14)	(3.13)	(9.93)	(3.62)	(11.53)	(4.35)	(13.73)	(5.41)	(17.26)	(6.88)	(21.98)
0.35	19.57	99.97	19.54	99.98	19.52	99.98	19.49	99.98	19.47	99.98	19.44	99.99	19.41	99.99	19.38	100.00	19.35	100.01	19.31	100.01	19.28	100.03
	(2.05)	(6.18)	(4.58)	(12.77)	(2.97)	(9.55)	(2.09)	(6.76)	(2.13)	(6.24)	(2.20)	(6.53)	(2.47)	(7.56)	(2.77)	(8.63)	(3.23)	(10.12)	(3.86)	(12.23)	(4.80)	(15.36)
0.4	19.71	99.98	19.69	99.97	19.68	99.97	19.66	99.97	19.65	99.98	19.63	99.98	19.61	99.98	19.59	99.99	19.57	99.99	19.55	100.00	19.53	100.01
	(1.93)	(5.80)	(1.95)	(7.77)	(2.64)	(7.75)	(1.05)	(3.60)	(1.78)	(5.08)	(1.90)	(5.40)	(2.00)	(6.03)	(2.26)	(6.90)	(2.58)	(7.95)	(3.01)	(9.34)	(3.61)	(11.32)
0.45	19.85	99.95	19.85	99.96	19.84	99.96	19.83	99.96	19.82	99.96	19.82	99.97	19.81	99.97	19.80	99.97	19.79	99.98	19.78	99.98	19.77	99.99

Table 7 continued

α	k																					
	1		1.1		1.2		1.3		1.4		1.5		1.6		1.7		1.8		1.9		2	
	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$
	(1.86)	(5.58)	(5.26)	(16.56)	(1.41)	(6.82)	(1.75)	(5.83)	(1.84)	(5.78)	<u>(1.70)</u>	<u>(5.28)</u>	(1.85)	(5.57)	(2.02)	(6.11)	(2.25)	(6.83)	(2.57)	(7.86)	(3.03)	(9.30)
0.5	19.98	99.96	19.98	99.95	19.98	99.95	19.98	99.95	19.98	99.96	19.98	99.96	19.98	99.96	19.98	99.96	19.98	99.97	19.98	99.97	19.98	99.98
	(1.75)	(5.26)	(1.98)	(7.41)	(1.70)	(5.37)	(1.59)	(4.69)	<u>(1.46)</u>	<u>(4.18)</u>	(1.55)	(4.68)	(1.69)	(4.94)	(1.90)	(5.73)	(2.14)	(6.44)	(2.45)	(7.42)	(2.88)	(8.75)

Table 8 The suitable k for simulation data where the error term obeys the t distribution with freedom degree 3

α	k 1		1.1		1.2		1.3		1.4		1.5		1.6		1.7		1.8		1.9		2	
	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$
0.55	20.12	99.94	20.12	99.94	20.13	99.94	20.14	99.94	20.14	99.94	20.15	99.95	20.16	99.95	20.17	99.95	20.18	99.96	20.19	99.96	20.20	99.97
	(1.80)	(5.41)	(3.80)	(9.25)	(1.84)	(6.33)	(1.65)	(4.80)	(1.55)	(5.13)	(1.63)	(4.86)	(1.79)	(5.33)	(1.98)	(5.91)	(2.23)	(6.65)	(2.57)	(7.69)	(3.05)	(9.13)
0.6	20.25	99.92	20.26	99.93	20.28	99.93	20.30	99.93	20.32	99.93	20.33	99.94	20.35	99.94	20.37	99.94	20.40	99.95	20.42	99.96	20.44	99.96
	(1.71)	(5.14)	(2.82)	(5.81)	(2.63)	(7.71)	(1.58)	(4.84)	(1.89)	(5.48)	(1.86)	(5.68)	(2.03)	(6.04)	(2.26)	(6.69)	(2.58)	(7.64)	(3.03)	(8.98)	(3.68)	(10.86)
0.65	20.39	99.92	20.41	99.92	20.44	99.92	20.46	99.93	20.49	99.93	20.52	99.93	20.55	99.93	20.58	99.94	20.61	99.94	20.65	99.95	20.68	99.96
	(1.97)	(5.92)	(4.14)	(16.77)	(1.92)	(6.02)	(2.02)	(6.21)	(2.32)	(6.48)	(2.23)	(6.43)	(2.45)	(7.15)	(2.81)	(8.24)	(3.26)	(9.58)	(3.89)	(11.45)	(4.82)	(14.11)
0.70	20.54	99.91	20.58	99.91	20.62	99.91	20.66	99.91	20.70	99.92	20.74	99.92	20.78	99.93	20.82	99.93	20.87	99.94	20.92	99.95	20.97	99.96
	(2.47)	(7.43)	(4.52)	(10.43)	(2.69)	(9.12)	(2.95)	(8.62)	(2.71)	(7.49)	(2.88)	(8.28)	(3.25)	(9.33)	(3.72)	(10.76)	(4.49)	(13.23)	(5.55)	(16.32)	(7.16)	(20.99)
0.75	20.73	99.90	20.78	99.90	20.82	99.91	20.88	99.91	20.93	99.91	20.98	99.92	21.04	99.93	21.10	99.93	21.16	99.94	21.23	99.95	21.30	99.97
	(3.18)	(9.59)	(7.76)	(20.50)	(1.62)	(6.90)	(3.00)	(9.26)	(3.77)	(11.27)	(3.96)	(11.48)	(4.32)	(11.83)	(5.33)	(15.73)	(6.76)	(19.92)	(8.56)	(25.04)	(11.82)	(34.83)
0.80	20.96	99.89	21.02	99.90	21.08	99.90	21.15	99.91	21.22	99.92	21.28	99.93	21.36	99.94	21.44	99.94	21.52	99.94	21.62	99.95	21.72	99.97
	(4.09)	(12.30)	(5.91)	(21.10)	(6.17)	(17.45)	(5.12)	(16.20)	(5.16)	(14.48)	(5.48)	(15.38)	(6.97)	(20.48)	(8.59)	(25.35)	(11.70)	(33.85)	(16.16)	(48.01)	(23.77)	(69.71)
0.85	21.25	99.90	21.32	99.92	21.40	99.94	21.48	99.94	21.57	99.93	21.67	99.92	21.77	99.92	21.88	99.92	22.01	99.93	22.15	99.96	22.31	100.02
	(5.60)	(16.52)	(16.18)	(55.87)	(8.33)	(24.02)	(8.20)	(25.62)	(8.09)	(23.14)	(10.26)	(29.64)	(13.02)	(36.81)	(18.87)	(54.58)	(25.95)	(77.66)	(38.20)	(120.64)	(61.58)	(190.55)
0.90	21.63	99.93	21.73	99.94	21.84	99.91	21.96	99.89	22.10	99.89	22.24	99.88	22.40	99.90	22.57	99.92	22.77	99.97	22.99	100.05	23.25	100.17
	(8.32)	(25.03)	(23.07)	(51.21)	(12.90)	(40.60)	(20.43)	(59.90)	(18.00)	(49.70)	(23.14)	(79.81)	(29.62)	(77.31)	(52.78)	(180.31)	(82.42)	(269.06)	(144.99)	(455.84)	(242.04)	(732.92)
0.95	22.32	99.84	22.51	99.84	22.71	99.81	22.93	99.84	23.18	99.83	23.49	99.91	23.81	100.04	24.18	100.22	24.60	100.43	25.05	100.68	25.59	101.05
	(21.46)	(63.55)	(74.06)	(148.35)	(33.74)	(78.96)	(54.77)	(193.85)	(106.91)	(400.99)	(126.29)	(395.87)	(207.58)	(555.76)	(410.40)	(1373.84)	(639.06)	(1994.47)	(1193.45)	(3701.39)	(2590.54)	(6622.53)

Table 9 Suitable k for various α, estimates and variances

α	0.05	0.1	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
k	1.2	1.9	1.5	1.0	1.1	1.4	1.4	1.1	1.2	1.0
Variables	est.	est.	est.	est.	est.	est.	est.	est.	est.	est.
cons	4.08***	4.08***	4.17***	4.25***	4.20***	4.19***	4.21***	4.27***	4.26***	4.28***
$(\times 10^{-2})$	(1.33)	(5.45)	(3.07)	(2.84)	(.224)	(2.97)	(2.28)	(1.84)	(2.45)	(2.26)
$x_1(\times 10^{-3})$	−5.51	−1.41	−1.17	0.00	−1.03	.799	.966	1.60**	1.59	2.49
$(\times 10^{-3})$	(6.23)	(2.90)	(1.51)	(2.01)	(1.23)	(1.42)	(1.37)	(.726)	(1.50)	(1.89)
$x_2(\times 10^{-3})$	−.0196	5.03*	5.22***	0.00	5.28***	8.37***	8.68***	6.13***	6.69***	6.22***
$(\times 10^{-3})$	(.624)	(2.94)	(1.83)	(1.41)	(.811)	(1.57)	(1.21)	(.750)	(1.31)	(1.13)
$x_2^2(\times 10^{-5})$	.0191	−7.65*	−7.67***	0.00	−7.29***	−11.3***	−11.8***	−8.48***	−9.19***	−8.00***
$(\times 10^{-5})$	(.780)	(4.06)	(2.70)	(2.00)	(1.30)	(2.08)	(1.69)	(.762)	(1.82)	(1.00)

α	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
k	2	1.5	1.1	1.2	1.2	1.1	1.0	1.7	1.3
Variables	est.	est.	est.	est.	est.	est.	est.	est.	est.
cons	4.26***	4.26***	4.28***	4.31***	4.31***	4.32***	4.37***	4.35***	4.38***
$(\times 10^{-2})$	(2.55)	(2.36)	(1.66)	(2.05)	(2.12)	(1.80)	(2.62)	(2.79)	(2.42)
$x_1(\times 10^{-3})$	1.24	2.26	2.03	2.75***	2.73***	1.58***	.450	2.91*	1.61
$(\times 10^{-3})$	(1.69)	(1.56)	(1.43)	(.764)	(.826)	(.406)	(1.51)	(1.72)	(1.69)
$x_2(\times 10^{-3})$	8.83***	9.45***	9.93***	8.98***	8.98***	11.8***	9.91***	11.4***	14.0***
$(\times 10^{-3})$	(1.27)	(1.17)	(.715)	(.863)	(.882)	(.717)	(1.31)	(1.40)	(1.22)
$x_2^2(\times 10^{-4})$	−1.18	−1.23	−1.30	−1.14	−1.14	−1.52**	−1.30	−1.48	-1.85
$(\times 10^{-5})$	(1.66)	(1.55)	(.829)	(.809)	(.822)	(.684)	(2.00)	(1.75)	(1.30)

*,** and *** label significance at 0.1, 0.05 and 0.01 levels, respectively

incomes strongly depend on working years, but do not their squares except $\alpha = 0.80$. Education has no significant contribution to incomes until $\alpha = 0.70, 0.75$ or 0.80.

The result illustrates the contribution rate of education is higher for migrant workers with upper income, which is in agreement with the stylized fact in labor economics.

8 Conclusion

In this paper, we consider the kth power expectiles and the kth power expectile regression method as well, mainly for $1 < k \leq 2$; the latter can be seen as an important special type of M estimation method.[1] This work is a partial extension of that of Newey and Powell (1987). We attempt to construct a bridge between quantiles and expectiles. The existence and uniqueness of the kth power expectiles have been proved under mild conditions. Furthermore, we discuss the consistency and asymptotic normality of the estimators of the kth power expectile regression. Some comparisons of kth power expectile estimators with estimators of the common quantile and expectile indicate the advantage of the kth power expectile regression. Another tentative conclusion is that the property of the kth power expectile regression is not close to that of the quantile regression as k approaches 1, while the property gets close to that of the expectile regression as k tends to 2. Researchers can choose the suitable k to run a satisfying kth power expectile regression by the method in Sect. 6 according to the specific problem and their preference. In the real data analysis, we fit the Mincerian earnings function model to the data of incomes of migrant workers in China at 2011. Results show that the kth power expectile regression delivers smaller variances for the majority of the values of α.

In the present paper, we only focus on the case of i.i.d. data, and it is interesting to extend our results to some more general cases, such as the dependent data. Another important issue is to apply the kth power expectile regression to testing heteroscedasticity, and we believe there will be some promising merits. In consideration of the space, these problems, specification test and variable selection for the kth power expectile regression mentioned above will be involved in the future study.

9 Proofs

The proof of Theorem 1 Write (4) as

$$\frac{1}{1-\tau}\int_{\mu}^{\infty}(x-\mu)^{k-1}\mathrm{d}F(x) = E|X-\mu|^{k-1}. \tag{13}$$

We will first prove, for any $0 < \tau < 1$, there exists a μ such that (13) holds. Let $S(\mu) = \int_{\mu}^{\infty}(x-\mu)^{k-1}\mathrm{d}F(x)$, $\widetilde{S}(\mu) = \int_{-\infty}^{\mu}(\mu-x)^{k-1}\mathrm{d}F(x)$ and thus $S(\mu)+\widetilde{S}(\mu) = E|X-\mu|^{k-1}$. The improper integral theorem taken into consideration; it follows that $S(\mu)$ and $-\widetilde{S}(\mu)$ are strictly decreasing. The following results are easy to deduce.

$$S(\mu) < E|X-\mu|^{k-1}, \ \lim_{\mu\to\infty} S(\mu) = 0, \ \lim_{\mu\to-\infty}(S(\mu) - E|X-\mu|^{k-1}) = 0.$$

[1] For M estimation methods, see Portnoy (1985), Breckling and Chambers (1988), Welsh (1989), Welsh (1990), Bai and Wu (1994), He and Shao (1996), He and Shao (2000), and Arcones (2001).

For $\tau(0 < \tau < 1)$ fixed, we have

$$\frac{1}{1-\tau}S(\mu) > E|X-\mu|^{k-1} \text{ for } \mu \text{ small enough,} \qquad (14)$$
$$\frac{1}{1-\tau}S(\mu) < E|X-\mu|^{k-1} \text{ for } \mu \text{ large enough.}$$

The intermediate value theorem therefore makes sure that the solution of (13) exists. For fixed τ, let $\mu_s(\tau)$ be the smallest solution to equation (13); it must exist according to (14), i.e., we have $\frac{1}{1-\tau}S(\mu_s(\tau)) = E|X-\mu_s(\tau)|^{k-1}$. Noting the derivation of $\frac{1}{1-\tau}S(\mu)$ is strictly smaller than that of $E|X-\mu|^{k-1}$; thus, there is no value of μ any more such that $\frac{\tau}{1-\tau}S(\mu)$ equals $E|X-\mu|^{k-1}$ when $\mu > \mu_s(\tau)$. As a result, (13) has a unique solution. The strictly monotonic property of $\mu(\tau)$ can be proved by the equation

$$\frac{1}{1-\tau} = 1 + \frac{\widetilde{S}(\mu)}{S(\mu)}$$

and the monotonic property of $S(\mu)$, $\widetilde{S}(\mu)$ and $1/1-\tau$. We show that $\mu(\tau)$ must lie in I_F. When μ is greater than any element of I_F, we have $S(\mu) = 0$. When μ is less than any element of I_F, we have $S(\mu) = E|X-\mu|^{k-1}$ and thus $\frac{1}{1-\tau}S(\mu) > E|X-\mu|^{k-1}$. Hence, no solution to (13) lies outside of I_F. To prove that $\mu(\tau)$ is onto I_F, we suppose μ is an element of I_F. Then, we have $0 < S(\mu) < E|X-\mu|^{k-1}$, and there is a τ such that μ satisfies (13). According to the definition of the τ kth power expectile $\widetilde{\mu}(\tau)$ of X, we have

$$\frac{1}{1-\tau}\int_{\widetilde{\mu}}^{\infty}(x-\widetilde{\mu})^{k-1}\mathrm{d}F_X(x) = E|X-\widetilde{\mu}|^{k-1}.$$

Noting $X = sY + t$, the above equation can be written as

$$\frac{1}{1-\tau}\int_{\frac{\widetilde{\mu}-t}{s}}^{\infty}\left(x-\frac{\widetilde{\mu}-t}{s}\right)^{k-1}\mathrm{d}F_Y(x) = E\left|Y-\frac{\widetilde{\mu}-t}{s}\right|^{k-1},$$

so we get $\widetilde{\mu}(\tau) = s\mu(\tau) + t$. □

The proof of Theorem 2 Let

$$\Upsilon(b;\tau,\mathcal{Y},\mathcal{X}) := \sum_{\{t:y_t \geq x_t'b\}} \tau(y_t - x_t'b)^k + \sum_{\{t:y_t < x_t'b\}} (1-\tau)(x_t'b - y_t)^k.$$

We have

$$\Upsilon(\lambda b;\tau,\lambda\mathcal{Y},\mathcal{X}) = \lambda^k \Upsilon(b;\tau,\mathcal{Y},\mathcal{X}).$$

So, $\widehat{\beta}(\tau, \lambda\mathcal{Y}, \mathcal{X}) = \lambda\widehat{\beta}(\tau, \mathcal{Y}, \mathcal{X})$. For $\lambda < 0$,

$$\begin{aligned}
&\Upsilon(\lambda b; 1-\tau, \lambda\mathcal{Y}, \mathcal{X})\\
&= \sum_{\{t:\lambda y_t \ge x_t'\lambda b\}} (1-\tau)(\lambda y_t - x_t'\lambda b)^k + \sum_{\{t:\lambda y_t < x_t'\lambda b\}} (1-(1-\tau))(x_t'\lambda b - \lambda y_t)^k\\
&= (-\lambda)^k \sum_{\{t:y_t < x_t' b\}} (1-\tau)(x_t' b - y_t)^k + (-\lambda)^k \sum_{\{t:y_t \ge x_t' b\}} \tau(y_t - x_t' b)^k\\
&= (-\lambda)^k \Upsilon(\lambda b; \tau, \lambda\mathcal{Y}, \mathcal{X}).
\end{aligned}$$

So (ii) follows. Further,

$$\begin{aligned}
&\Upsilon(b+\gamma; \tau, \mathcal{Y} + \mathcal{X}'\gamma, \mathcal{X})\\
&= \sum_{\{t:y_t + x_t'\gamma \ge x_t'(b+\gamma)\}} \tau(y_t + x_t'\gamma - x_t'(b+\gamma))^k\\
&\quad + \sum_{\{t:y_t + x_t'\gamma < x_t'(b+\gamma)\}} (1-\tau)(x_t'(b+\gamma) - y_t - x_t'\gamma)^k\\
&= \Upsilon(b; \tau, \mathcal{Y}, \mathcal{X}).
\end{aligned}$$

So,

$$\widehat{\beta}(\tau, \mathcal{Y} + \mathcal{X}'\gamma, \mathcal{X}) = \widehat{\beta}(\tau, \mathcal{Y}, \mathcal{X}) + \gamma.$$

Finally,

$$\begin{aligned}
&\Upsilon(A^{-1}b; \tau, \mathcal{Y}, \mathcal{X}'A)\\
&= \sum_{\{t:y_t \ge x_t'AA^{-1}b\}} \tau(y_t - x_t'AA^{-1}b)^k + \sum_{\{t:y_t < x_t'AA^{-1}b\}} (1-\tau)(x_t'AA^{-1}b - y_t)^k\\
&= \Upsilon(b; \tau, \mathcal{Y}, \mathcal{X}).
\end{aligned}$$

So (iv) follows. □

The proof of Theorem 3 The proofs of main results are based on the following lemma that comes from Newey and Powell (1987).

Lemma 1 *Let θ_0 be a point in $\mathbf{R}^q$ and $\mathbf{\Theta}$ an open set containing θ_0. If*

(A) *$Q_n(\theta)$ converges to $Q(\theta)$ in probability uniformly on $\mathbf{\Theta}$,*
(B) *$Q(\theta)$ has a unique minimum on $\mathbf{\Theta}$ at θ_0*
(C) *$Q_n(\theta)$ is convex in θ; then for $\theta = argmin_{R^q} Q_n(\theta)$,*
(i) *$\hat{\theta}$ exists with probability approaching one,*
(ii) *$\hat{\theta}$ converges in probability to θ_0.*

We mainly verify (A), (B) and (C) under Assumptions 1–3. Write $R(\beta, \tau) \equiv E(Q_\tau(Y - X'\beta) - Q_\tau(Y))$ and $g(\beta) \equiv \partial Q_\tau(Y - X'\beta)/\partial\beta$. There exist some positive constants c_1 and c_2 such that $g(\beta) \le |Z|^k(c_1 + c_2|\beta|)$. On a neighborhood of any β, the $g(\beta)$ is uniformly dominated by an integrable function using Assumption 2. We have

$$\begin{aligned}\partial R(\beta,\tau)/\partial\beta &= kE\Big(X\Big(-\tau\int_{X'\beta}^{\infty}(y-X'\beta)^{k-1}f(y|X)\mathrm{d}y\\&\quad+(1-\tau)\int_{-\infty}^{X'\beta}(X'\beta-y)^{k-1}f(y|X)\mathrm{d}y\Big)\Big)\\&=: G(k,\beta,\tau).\end{aligned}$$

According to Assumptions 2 and 3, using an argument similar to that in the proof of (17), we can prove

$$\begin{aligned}\partial G(k,\beta,\tau)/\partial\beta = k(k-1)E\Big(XX'\Big(\tau\int_{X'\beta}^{\infty}(y-X'\beta)^{k-2}f(y|X)\mathrm{d}y\\+(1-\tau)\int_{-\infty}^{X'\beta}(X'\beta-y)^{k-2}f(y|X)\mathrm{d}y\Big)\Big),\end{aligned} \tag{15}$$

and the expectation in (15) is even bounded locally uniformly with respect to β. Using the improper integral theorem and Assumption 3, there exists positive constant c_1 such that

$$\partial G(k,\beta,\tau)/\partial\beta - c_1k(k-1)E(XX')$$

is positive semi-definite. Let $\delta = c_1k(k-1)\min\{\tau, 1-\tau\}$. We have $\partial G(k,\beta,\tau)/\partial\beta - \delta E(XX')$ is positive semi-definite. Here, the equation, $G(k,\beta,\tau) = 0$, is analogous to that of (2.9) in Newey and Powell (1987). We can obtain an expansion like (A.12) in Newey and Powell (1987). Under Assumption 5, using the same argument in the proof of Theorem 3 in Newey and Powell (1987), the existence and uniqueness of $\widetilde{\beta}_0$ can be proved. Thus, (B) holds for $R(\beta,\tau)$. Note

$$S_\tau(b) = \sum_{i=1}^{n}(|\tau - I(y_i - x_i'b < 0)||y_i - x_i'b|^k - |\tau - I(y_i < 0)||y_i|^k).$$

Without loss of generality, let $\boldsymbol{\Theta}$ be any bounded open set containing $\widetilde{\beta}_0$, and we can get a compact set $\boldsymbol{\Theta}_1$ such that $\boldsymbol{\Theta}_1 \supset \boldsymbol{\Theta}$. Using Assumption 2 and the compactness of $\boldsymbol{\Theta}_1$, we have $\max\{1, \sup_{b\in\boldsymbol{\Theta}_1}|b|\}E|Z|^k < \infty$ and $E|Y|^k < c_0$. So

$$\begin{aligned}&E\Big(\sup_{b\in\boldsymbol{\Theta}_1}(|\tau - I(Y - X'b < 0)||Y - X'b|^k - |\tau - I(Y < 0)||Y|^k)\Big)\\&\le \max\{\tau, (1-\tau)\}\Big(\max\big\{1, \sup_{b\in\boldsymbol{\Theta}_1}|b|\big\}E|Z|^k + E|Y|^k\Big) < \infty.\end{aligned}$$

The same argument in the proof of Lemma 5.2.2 in van de Geer (2006) can make sure that

$$\sup_{b\in\boldsymbol{\Theta}_1} |S_\tau(b)/n - R(b,\tau)| \xrightarrow{P} 0.$$

Noting $\boldsymbol{\Theta}_1 \supset \boldsymbol{\Theta}$, we have $\sup_{b\in\boldsymbol{\Theta}} |S_\tau(b)/n - R(b,\tau)| \xrightarrow{P} 0$, and (A) is satisfied for $S_\tau(b)/n$. The convexity of $S_\tau(b)$ is palpable. So using lemma 9.1 can complete the proof. □

Remark 7 A reviewer suggested proving (A) using Lemma 1 in Hjort and Pollard (1993). In our paper, the convexity of $S_\tau(b)$ is obvious and the convergence $S_\tau(b)/n$ to $R(b,\tau)$ in probability can be obtained according to the i.i.d. setting of z_i and Assumption 2. So a straight application of Lemma 1 in Hjort and Pollard (1993) can complete the proof.

The proof of Theorem 4 We utilize Theorem 2.1 in Hjort and Pollard (1993) to complete the proof. To this end, we need to verify that the conditions in that theorem are satisfied when the relevant objects are replaced with ours. Using the Taylor formula, we have

$$\begin{aligned} &Q_\tau(y_i - x_i'(\tilde{\beta}_0(\tau)+t)) - Q_\tau(y_i - x_i'\tilde{\beta}_0(\tau)) \\ &= -x_i'\varphi_\tau(y_i - x_i'\tilde{\beta}_0(\tau))t + \frac{1}{2}t'x_ix_i'\psi_\tau(y_i - x_i'\xi(\tau))t \\ &=: D(x_i,y_i)'t + R(x_i,y_i,t), \end{aligned} \tag{16}$$

with $\xi(\tau)$ being some vector between $\tilde{\beta}_0(\tau)+t$ and $\tilde{\beta}_0(\tau)$,

$$\varphi_\tau(r) = (-1)^{I\{r<0\}}k|\tau - I\{r<0\}||r|^{k-1}$$

and

$$\psi_\tau(r) = k(k-1)|\tau - I\{r<0\}||r|^{k-2}.$$

The proof of Theorem 3 ensures that the matrix $E(x_ix_i'\psi_\tau(y_i - x_i'\tilde{\beta}_0(\tau))) - c_1k(k-1)E(x_ix_i')$ is positive semi-definite. The matrix $E(x_ix_i')$ is positive definite matrix, and so do $J := E(x_ix_i'\psi_\tau(y_i - x_i'\tilde{\beta}_0(\tau)))$. The expectation of the second term in (16) can be written as

$$\begin{aligned} E(R(x_i,y_i,t)) &= \frac{1}{2}t'Jt + \frac{1}{2}t'E\left(x_ix_i'\left(\int_{-\infty}^{+\infty}\psi_\tau(y - x_i'\xi(\tau)) - \psi_\tau(y - x_i'\tilde{\beta}_0(\tau))f(y|x_i)\mathrm{d}y\right)\right)t \\ &= \frac{1}{2}t'Jt + o(|t|^2). \end{aligned}$$

In fact, in order to prove the second equality, it is enough to show each element of the matrix $E(x_ix_i'(\int_{-\infty}^{+\infty}\psi_\tau(y - x_i'\xi(\tau)) - \psi_\tau(y - x_i'\tilde{\beta}_0(\tau))f(y|x_i)\mathrm{d}y))$ converges to zero as $|t| \to 0$. It is further sufficient to prove the element $E(|x_i|^2(\int_{-\infty}^{+\infty} \psi_\tau(y - x_i'\xi(\tau)) - \psi_\tau(y - x_i'\tilde{\beta}_0(\tau))f(y|x_i)\mathrm{d}y))$ tends to zero as $|t| \to 0$. According to (5), for any $\varepsilon_1 > 0$, there exists a positive constant M_1 enough large such that

$$
\begin{aligned}
&|\int_{-\infty}^{+\infty} \psi_\tau(y - x_i'\xi(\tau)) - \psi_\tau(y - x_i'\tilde{\beta}_0(\tau))f(y|x_i)\mathrm{d}y| \\
&\leq |\int_{-M_1}^{M_1} \psi_\tau(y - x_i'\xi(\tau)) - \psi_\tau(y - x_i'\tilde{\beta}_0(\tau))f(y|x_i)\mathrm{d}y| + 2\varepsilon_1. \qquad (17)
\end{aligned}
$$

The limit of the first term in (17) is zero by the dominated convergence theorem as $|t| \to 0$. Combining this and the arbitrariness of ε_1 implies $\int_{-\infty}^{+\infty} \psi_\tau(y - x_i'\xi(\tau)) - \psi_\tau(y - x_i'\tilde{\beta}_0(\tau))f(y|x_i)\mathrm{d}y \to 0$ as $|t| \to 0$. On the basis of (6), for any $\varepsilon_2 > 0$, we can find a positive M_2 that makes sure that, for $|x_i| > M_2$

$$
\int_{-\infty}^{+\infty} \psi_\tau(y - x_i'\xi(\tau))f(y|x_i)\mathrm{d}y \leq (1 + \varepsilon_2)c_2|x_i|^k
$$

and

$$
\int_{-\infty}^{+\infty} \psi_\tau(y - x_i'\tilde{\beta}_0(\tau))f(y|x_i)\mathrm{d}y \leq (1 + \varepsilon_2)c_2|x_i|^k.
$$

In light of Assumption 2, for any $\varepsilon_3 > 0$, there is $M_3 > 0$ such that

$$
E(|x_i|^{2+k}) \leq E(|x_i|^{2+k}I\{|x_i| \leq M_3\}) + 2\varepsilon_3.
$$

Let $M_4 = \max\{M_2, M_3\}$ and we have, as $|t| \to 0$,

$$
\begin{aligned}
& E\left(x_i x_i'\left(\int_{-\infty}^{+\infty} \psi_\tau(y - x_i'\xi(\tau)) - \psi_\tau(y - x_i'\tilde{\beta}_0(\tau))f(y|x_i)\mathrm{d}y\right)\right) \\
\leq\ & E\left(x_i x_i' I\{|x_i| \leq M_4\}\left(\left|\int_{-\infty}^{+\infty} \psi_\tau(y - x_i'\xi(\tau)) - \psi_\tau(y - x_i'\tilde{\beta}_0(\tau))f(y|x_i)\mathrm{d}y\right|\right)\right) \\
& + 4c_2\varepsilon_3(1 + \varepsilon_2) \\
\longrightarrow\ & 0.
\end{aligned}
$$

The last limit is due to the dominated convergence theorem and the arbitrariness of ε_2 and ε_3. Using the argument similar to the above, Assumptions 2, 4 and Remark 1 can deduce $Var(R(x_i, y_i, t)) = o(|t|^2)$. The definition of $\tilde{\beta}_0(\tau)$ implies $E(D(x, y)) = 0$. Using Assumption 2, it easy to examine that

$$
E(Q_\tau(y_i - x_i'(\tilde{\beta}_0(\tau) + t)) - Q_\tau(y_i - x_i'\tilde{\beta}_0(\tau))) = \frac{1}{2}t'Jt + o(|t|^2), \text{ as } t \to 0.
$$

Here, $D(x_i, y_i)$ has a finite covariance matrix $K = E(D(x_i, y_i)D(x_i, y_i)')$ based on Assumption 2. So the conditions of Theorem 2.1 in Hjort and Pollard (1993) are satisfied, and the proof is completed. □

The proof of Theorem 5 Write

$$|\widehat{J} - J| \leq \left| \sum_{i=1}^{n} (\widehat{w}_i(\tau)|\widehat{u}_i(\tau)|^{k-2} x_i x_i'/n - w_i(\tau)|u_i(\tau)|^{k-2}) x_i x_i'/n \right| + \left| \sum_{i=1}^{n} w_i(\tau)|u_i(\tau)|^{k-2} x_i x_i'/n - J \right| =: I_1 + I_2.$$

Firstly, it follows that $E(w_i(\tau)|u_i(\tau)|^{k-2}x_ix_i') < \infty$ due to Assumptions 2 and 3; thus, we have $I_2 \xrightarrow{P} 0$. Secondly, let M_i stand for $\widehat{w}_i(\tau)|\widehat{u}_i(\tau)|^{k-2}x_ix_i' - w_i(\tau)|u_i(\tau)|^{k-2}x_ix_i'$ and it can be yielded that $M_i \xrightarrow{P} 0$ uniformly for $i = 1, 2, \ldots, n$. For any positive constant $\varepsilon/2$, there exists a n_0 such that when $n > n_0$ $P(|M_i| > \varepsilon/2) < \varepsilon/2$. So

$$E\left(\frac{|M_i|}{1 + |M_i|/n}\right) = E\left(I\{|M_i| > \varepsilon/2\}\frac{|M_i|}{1 + |M_i|/n}\right) + E\left(I\{|M_i| \leq \varepsilon/2\}\frac{|M_i|}{1 + |M_i|/n}\right) \leq \varepsilon.$$

Furthermore, $E(I_1/(1 + I_1)) \leq (1/n)\sum_{i=1}^{n} E(|M_i|/(1 + |M_i|/n)) \leq \varepsilon$, which deduces $E(I_1/(1 + I_1)) \longrightarrow 0$. The latter implies $I_1 \xrightarrow{P} 0$ thanks to Theorem 4.1.5 in Chung (1974). We therefore complete the proof of $\widehat{J} \xrightarrow{P} J$. Similarly to the proof of Theorem 4 in Newey and Powell (1987), there exist constants d and d' such that

$$|x_ix_i'(w_i(\tau))^2(u_i(\tau))^{2(k-1)}| \leq |z_i|^{2k}(d + d'|\widetilde{\beta}_0|^2).$$

Then using the same argument as in the proof of Theorem 2.2 of Newey (1985) can produce

$$\widehat{K} \xrightarrow{P} K.$$

The proof is completed by Slutsky's theorem. □

Acknowledgements The authors would like to thank the Associate Editor whose constructive comments helped to improve a previous version of this paper greatly. The authors would also like to thank two referees for careful reading of the paper and a number of perceptive and beneficial comments which improved it greatly. Any remaining errors are the authors' responsibility. We are very grateful to professor Kaizhi Yu for generously sharing of the incomes data. The research of the first author was supported by the talent introduction project of Sichuan University of Science & Engineering (2019RC10). The research of the second author was supported by the Scientific research cultivation project of Sichuan University of Science & Engineering (2013PY07), the Scientific Research Fund of Shanghai University of Finance and Economics (2017110080), and the Opening Project of Sichuan Province University Key Laboratory of Bridge Non-destruction Detecting and Engineering Computing (2018QZJ01). Zhou's work was supported by the State

Key Program of National Natural Science Foundation of China (71331006), the State Key Program in the Major Research Plan of National Natural Science Foundation of China (91546202).

References

Andriyana, Y., Gijbels, I., Verhasselt, A. (2016). Quantile regression in varying-coefficient models: Non-crossing quantile curves and heteroscedasticity. *Statistical Papers*, *57*, 1–33.

Arcones, M. A. (1996). The Bahadur–Kiefer representation of L_p regression estimators. *Econometric Theory*, *12*, 257–283.

Arcones, M. A. (2001). Asymptotic distribution of regression M-estimators. *Journal of Statistical Planning and Inference*, *97*, 235–261.

Bai, Z., Wu, Y. (1994). Limiting behavior of M-estimators of regression coefficients in high dimensional linear models, I. Scale-dependent case. *Journal of Multivariate Analysis*, *51*, 211–239.

Breckling, J., Chambers, R. (1988). M-quantiles. *Biometrika*, *75*, 761–771.

Cai, Z., Xiao, Z. (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics*, *167*, 413–425.

Cai, Z., Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, *103*, 1595–1608.

Chen, Z. (1996). Conditional L_p-quantiles and their application to testing of symmetry in non-parametric regression. *Statistics & Probability Letters*, *29*, 107–115.

Chung, K. L. (1974). *A course in probability theory*, 2nd ed. New York: Academic Press.

Efron, B. (1991). Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, *1*, 93–125.

Ehm, W., Gneiting, T., Jordan, A., Krüger, F. (2016). Of quantiles and expectiles: Consistent scoring functions, choquet representations and forecast rankings. *Journal of the Royal Statistical Society*: Series B, *78*, 505–562.

Engle, R. F., Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, *22*, 367–381.

Farooq, M., Steinwart, I. (2017). An SVM-like approach for expectile regression. *Communications in Statistics-Theory and Methods*, *109*, 159–181.

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

Gu, Y., Zou, H. (2016). High-dimensional generalizations of asymmetric least squares regression and their applications. *Annals of Statistics*, *44*, 2661–2694.

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

He, X., Shao, Q. M. (2000). On parameters of increasing dimenions. *Journal of Multivariate Analysis*, *73*, 120–135.

Hjort, N. L., Pollard, D. B. (1993). Asymptotics for minimizers of convex processes (**unpublished manuscript**).

Holzmann, H., Klar, B. (2016). Expectile asymptotics. *Electronic Journal of Statistics*, *10*, 2355–2371.

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

Hunter, D. R., Lange, K. (2000). Quantile regression via an MM algorithm. *Journal of Computational and Graphical Statistics*, *9*, 60–77.

Kim, M. O. (2007). Quantile regression with varying coefficients. *Annals of Statistics*, *35*, 92–108.

Koenker, R. (2005). *Quantile regression, No. 38*. Cambridge: Cambridge University Press.

Koenker, R. (2017). Quantile regression: 40 years on. *Annual Review of Economics*, *9*, 155–176.

Koenker, R, Jr., Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*, 33–50.

Koenker, R, Jr., Bassett, G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, *50*, 43–61.

Kuan, C. M., Yeh, J. H., Hsu, Y. C. (2009). Assessing value at risk with care, the conditional autoregressive expectile models. *Journal of Econometrics*, *150*, 261–270.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, *66*, 281–302.

Newey, W. K. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica*, *53*, 1047–1070.
Newey, W. K., Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, *55*, 819–847.
Portnoy, S. (1985). Asymptotic behavior of M estimators of p regression parameters when p^2/n is large: II. Normal approximation. *Annals of Statistics*, *13*, 1403–1417.
Scheffé, H. (1947). A useful convergence theorem for probability distributions. *Annals of the Institute of Statistical Mathematics*, *18*, 434–438.
Taylor, J. W. (2008). Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Economics*, *6*, 231–252.
van de Geer, S. A. (2006). Empirical process theory and applications. *Handout WS 2006*. Zürich: ETH.
Welsh, A. H. (1989). On M-processes and M-estimation. *Annals of Statistics*, *17*, 337–361.
Welsh, A. H. (1990). Correction: "On M-processes and M-estimation". *Annals of Statistics*, *18*, 1500.
Yao, Q., Tong, H. (1996). Asymmetric least squares regression estimation: A nonparametric approach. *Journal of Nonparametric Statistics*, *6*, 273–292.