Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 10, No. 1, February, 2026

Probability Density Function Adjustment for Estimating Quantile Regression Coefficients

Pimpan Amphanthong [1*], Warangkhana Riansut [2]

[1] Deparment of Mathematics, Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi, Suphanburi, 13000, Thailand.

[2] Mathematics and Data Management Program, Faculty of Science and Digital Innovation, Thaksin University, Phatthalung, 93210, Thailand.

Abstract

This study aims to improve the estimation of quantile regression coefficients by adjusting probability density functions using a selected τ-function that exhibits symmetric properties. The research focuses on five quantile levels Q(20)th, Q(25)th, Q(50)th, Q(75)th, and Q(80)th and compares the proposed method with conventional multiple regression through simulation experiments under varying sample sizes and distributional conditions. Performance is evaluated using the mean absolute error (MAE) as the primary metric. The findings indicate that for small sample sizes (n=8, n=15), both multiple and quantile regression methods perform well, especially at lower quantiles (Q(20)th to Q(50)th). However, as sample sizes increase (n=50, n=100), quantile regression at higher quantiles (Q(50)th, Q(75)th, Q(80)th) demonstrates superior estimation accuracy. In relation to kurtosis and skewness, the Q(50)th and Q(80)th quantiles are sensitive to distributional changes, effectively capturing transitions from high to normal kurtosis and central shifts in skewed distributions. The novelty of this research lies in the integration of the τ-function into the quantile regression framework, enhancing robustness and accuracy in coefficient estimation under non-normal conditions. This approach contributes to methodological advancements in regression analysis, particularly in applications involving non-standard data distributions.

Keywords:

Multiple Regression;
Regression Coefficient;
Quartile Regression; Kernel Function;
Kurtosis Value; Skewed Value.

Article History:

Received:	27	July	2025
Revised:	19	November	2025
Accepted:	08	December	2025
Published:	01	February	2026

1- Introduction

Regression analysis is a basic statistical technique frequently employed in many fields to study the relationship between an independent variable and one or more independent variables. Ordinary Least Squares (OLS) is the most widely used technique, yielding parameter estimates under standard assumptions, including normally distributed errors, homoscedasticity, independence of error terms, and absence of influential outliers [1]. Although OLS works best under these assumptions, available data in the real world usually violate such assumptions. Skewness (Sk), kurtosis (Ku), heteroscedasticity, and outliers are some common deviations that may cause parameter estimate biasing and violate statistical conclusions [2]. These data anomalies can also be seen using simple graphical tools like box plots. Here, outliers are identified on the basis of the interquartile range: IQR = Q3 − Q1, and the distributions deviate from the canons of symmetry and central tendency. To span the gaps of OLS, Koenker & Bassett [3] developed quantile regression that allows us to make estimates of the conditional quantiles of the response variable. The technique finds special use in the estimation of non-normal error structures, capturing heterogeneous relationships along the distribution, and reducing the influence of outliers or non-constant variance. Its universality suits it for use in a wide range of practical problems, especially where OLS assumptions do not apply.

Nevertheless, one of the classic problems of quantile regression is parameter estimation when the true probability density function (PDF) is unknown. The precision and reliability of estimation and hypothesis testing are enhanced when

* **CONTACT**: pimpan.a@rmutsb.ac.th

DOI: http://dx.doi.org/10.28991/ESJ-2026-010-01-023

the underlying density function is known or is well estimated [4, 5]. Nonparametric methods—basically Kernel Density Estimation (KDE)—have become extremely popular tools for approximating complex, unknown distributions over recent years. KDE is not a particular distributional form and can quite suitably reflect a large variety of shapes of density by having a proper choice of bandwidth. Some studies have considered using KDE along with quantile regression, where the τ-function is used for rescaling the derived density function to enhance quantile estimation [6–8]. This provides the possibility of comparing quantile regression performance with ordinary multiple regression under different conditions of errors, such as non-normality and tail heaviness, generally measured with simulation [9].

In continuation of this line of research, several recent studies have extended quantile regression methods by adding advanced density adjustment methods. Zhang et al. [10], for example, introduced a nonparametric density-adjusted model to improve nonlinear data analysis. Lee & Kim [2] utilized kernel functions to improve robust estimation with heavy-tailed distribution. He et al. [11] created a new density adjustment procedure to minimize bias, while Zhou et al. [12] were constructing a nonparametric solution specifically aimed at skewed and heteroscedastic data. Koenker & Xiao [13] extended these solutions further to high-dimensional and time-correlated data structures. These articles demonstrate continued work in refining quantile regression in the face of increasingly sophisticated data issues.

Despite such advances, there is a lack of thorough comparative studies wherein quantile regression performance is compared with that of multiple regression under various distributions of errors—especially regarding varying quantile levels and nonparametric density adjustments. Filling this gap in research, the present study examines the estimation of quantile regression coefficients [14] from adjusted nonparametric probability density functions [15, 16] based on simulated data. Particularly, it measures estimation efficiency at five quantile levels (Q(20)th, Q(25)th, Q(50)th, Q(75)th, Q(80)th) [17, 18], under different assumptions of skewness, kurtosis, and non-constant variance. The outcomes are contrasted with those of conventional multiple regression, with mean absolute error (MAE), skewness (Sk), and kurtosis (Ku) being the criteria for evaluation. The results strive to enlighten researchers on the appropriateness of regression techniques depending on the distributional characteristics of their data and direct the construction of more reliable estimation procedures for non-standard statistical settings.

2- Materials and Methods

2-1- Symbols and Meanings

In presenting the results of the data analysis, the researcher employed the following symbols to interpret the data:

Y	Dependent variable vector of size $n \times 1$	X	Independent variable matrix of size $n \times p$
β	Regression coefficient vector of size p ×1	n	Sample size
p	Number of parameters	Q(r th)	Quantile at the percentile position r^{th} of Y given X
β_r	Coefficients of quantile vector at position of r^{th}	$f(\cdot)$	Probability density function
$K(\cdot)$	Kernel function	h	Window width or bandwidth
sk	Skewness	Ku	Kurtosis
r	Quantile regression	$\rho_r(u)$	Loss function of quantile regression

2-2- Multiple Regression Coefficients Estimation (MRE)

The multiple regression estimation [19] is expressed through the following matrix equations:

$$y = X\underline{\beta} + \underline{\epsilon} \tag{1}$$

where as, vector $\underline{y} = [y_1, y_2, ..., y_n]'_{1\times n}$. Instead of the dependent variable with size $n \times 1$, the matrix$X = [\underline{x}'_{i1}, \underline{x}'_{i2}, ..., \underline{x}'_{ip}]_{n\times p}$ represents the independent variable with size n ×p, which is represented by the regression coefficient number p (consists of constant values or called the y-intercept), and regression coefficient vector $\underline{\beta} = [\beta_1, \beta_2, ..., \beta_p]'_{1\times p}$ of size $p \times 1$ and vector $\underline{\varepsilon} = [\varepsilon_1, \varepsilon_2, ..., \varepsilon_n]'_{1\times p}$, instead of the error value that has a size $n \times 1$. When the number $\underline{\varepsilon} \sim N(0, \sigma^2)$ of forces of the matrix X is given $(rank(X) = p < n)$, the method of least squares is used to find the regression coefficient $\underline{\hat{\beta}}$. In estimation, $\underline{\beta}$ under the least square error is:

$$\underline{\hat{\beta}} = (X'X)^{-1}\left(X'\underline{y}\right) \tag{2}$$

2-3- Estimation of the Quantile Regression Coefficient (Q(r^{th}))

Quantile regression coefficient estimation is conceptually analogous to multiple regression in the way that it takes into account the dependent and independent variable relationship. Nevertheless, they are not too similar in terms of assumptions and calculation methods. Multiple regression is conditional based on the conditional mean, and therefore it is not appropriate in scenarios where data contain extreme values or outliers at both ends of distributions, or heteroskedasticity due to large-scale variation in the independent variable. In such a situation, the dependent variable contains too much variation or values far from the range of expectation, and multiple regression is inappropriate for the depiction of such variations. To this, quantile regression offers an increasingly superior alternative, especially if the dependent variable has a skew—left or right. The procedure allows researchers to estimate conditional quantiles, by which they are able to define the response of the dependent variable at any position along the distribution, particularly the tails. That makes quantile regression qualified to be applied on data in which values are divergent from the mean. Particularly, it predicts the q-th quantile of the distribution of the dependent variable [20], that is,

$$\mathrm{F_y}(\mu_q) = \mathrm{P}(y \leq \mu_q) = q \tag{3}$$

If specified $P(y \leq 0) = F_y(0) = 0.5$. It shows that the probability that a value of y is less than or equal to 0 is 0.5. It shows that the probability that y is greater than a given constant. $P(Y > y) = 1 - F(y)$ is given by q. It shows the meaning of estimating the quantile regression coefficient using the conditional probability distribution $F_y(y)$ when x the qth quantile condition is given by q [19, 21], as follows:

$$Q_{y|x}(q) = \text{infimum}\{y: F_y(y) \geq q\} \tag{4}$$

Where the distribution of the dependent variable $E(Y|X) = X\underline{\hat{\beta}}^{(r)}$; r instead of the percentile position that you want to estimate, you will get the estimated regression coefficient at the percentile position r^{th}.

$$Q_{y|x}(q) = X\underline{\beta}^{(r)} + \underline{\epsilon}^{(r)} \tag{5}$$

where, as regression coefficient vector $\underline{\beta}^{(r)} = \left[\beta_1^{(r)}, \beta_2^{(r)}, \ldots, \beta_p^{(r)}\right]'_{1\times p}$. Instead the quartile value is the position that r^{th} has the size $p \times 1$ and vector $\underline{\varepsilon}^{(r)} = \left[\varepsilon_1^{(r)}, \varepsilon_2^{(r)}, \ldots, \varepsilon_n^{(r)}\right]'_{1\times p}$. Substituting the error values at the position r^{th} with magnitude $n \times 1$ under the weighted sum (q) of the error terms defined by [6, 19], find the regression coefficients $\underline{\hat{\beta}}^{(r)}$; in the estimation $\underline{\beta}^{(r)}$; under the weighted sum, that is,

$$\underline{\hat{\beta}}^{(r)} = \min_{\underline{\hat{\beta}}_1}\left(q\sum_{\substack{y_i \geq \underline{\hat{\beta}}_1^{(r)} + \underline{\hat{\beta}}_2^{(r)}\underline{x'}_{i1} + \ldots + \underline{\hat{\beta}}_p^{(r)}\underline{x'}_{ip} \\ i=1}}\left|y_i - \underline{\hat{\beta}}_1^{(r)} - \underline{\hat{\beta}}_2^{(r)}\underline{x'}_{i1} - \ldots - \underline{\hat{\beta}}_p^{(r)}\underline{x'}_{ip}\right| + (1-q)\sum_{\substack{y_i \geq \underline{\hat{\beta}}_1^{(r)} + \underline{\hat{\beta}}_2^{(r)}\underline{x'}_{i1} + \ldots + \underline{\hat{\beta}}_p^{(r)}\underline{x'}_{ip} \\ i=1}}\left|y_i - \underline{\hat{\beta}}_1^{(r)} - \underline{\hat{\beta}}_2^{(r)}\underline{x'}_{i1} - \ldots - \underline{\hat{\beta}}_p^{(r)}\underline{x'}_{ip}\right|\right) \tag{6}$$

2-4- Density Estimation and Fitting of Probability Density Function for Estimation

2-4-1-Density Estimation

Theory 1 Let $\underline{X} = (X_1, X_2, \ldots, X_n)$ be a population size with an unknown probability density function of n and $f(X, x_i)$. Let $\underline{x} = (x_1, x_2, \ldots, x_n)$be a random sample of size n from the population and let $\hat{f}(X, x_i)$ be the density estimator, a method that uses the principle of random variables with density functions. $f(X, x_i)$ [1, 22] as follows:

$$f(X, x_i) = \lim_{h\to 0}\frac{1}{2h}P(x_i - h < X < x_i + h) \tag{7}$$

For values h specified using proportions in the range $(x_i - h < X < x_i + h)$and with a weighting function. $w(x)$ Therefore, the density estimator $\hat{f}(X, x_i)$of $f(X, x_i)$. will be:

$$\hat{f}(X, x_i) = \frac{1}{n}\sum_{i=0}^{n}\frac{1}{h}w\left(\frac{X - x_i}{h}\right) \tag{8}$$

Theory 2 Let $\underline{X} = (X_1, X_2, \ldots, X_n)$ be the population size with an unknown probability density function of $f(X, x_i)$. Let $\underline{x} = (x_1, x_2, \ldots, x_n)$ be a random sample of size n from the population, and let $\underline{x}$ be the sample size of n. The kernel function and $K\left(\frac{X - x_i}{h}\right)$ are the window widths corresponding to the symmetric function $\int_{-\infty}^{\infty} K\left(\frac{X - x_i}{h}\right)dx = 1$. So the density estimator $\hat{f}(X, x_i)$of $f(X, x_i)$ will be:

$$\hat{f}(X, x_i) = \frac{1}{nh}\sum_{i=0}^{n}K\left(\frac{X - x_i}{h}\right). \tag{9}$$

Theory 3 Let $\underline{X} = (X_1, X_2, \ldots, X_n)$ be the population size for which the probability density function is unknown $f(X, x_i)$, $Let\ \underline{x} = (x_1, x_2, \ldots, x_n)$. It is a random sample of size n from the population and is defined $\underline{x}$ as kernel function $K(t)$ with features $\int K(t)\,dt = 1$, $\int tK(t)\,dt = 0$ and $\int t^2K(t)\,dt = k_2 \neq 0$ when k_2 is a constant and h is the window width with a value $h = h(0) \to 0$ of $n \to \infty$, the bias of $\hat{f}(X, x_i), f(X, x_i)$ [23, 24] will get bias:

$$\left(\hat{f}(X\ x_i)\right) = \frac{1}{2}h^2 f^{(2)}(x)k_2 +,,, + O(h^2) \tag{10}$$

Fox [25] found that the optimal window width for a population with a normal distribution and a Gaussian Kernel function is the window width equal to $h = \frac{2\sigma}{n^{\frac{1}{5}}}$.

2-4-2-Probability Density Function Fitting for Estimation

Lemma states that $q \in (0,1)$ under the estimation of the regression coefficient quantile $\underline{\hat{\beta}}^{(r)}$ (Equation 6) has the following symmetric properties [26].

1. Scale equivariance: for every constant $c > 0$ and $q \in (0,1)$ will get;

   1.1. $\underline{\hat{\beta}}^{(r)}\left(c\underline{y}, X\right) = c\underline{\hat{\beta}}^{(r)}(y, X)$

   1.2. $\underline{\hat{\beta}}^{(r)}\left(-c\underline{y}, X\right) = -c\underline{\hat{\beta}}^{(r)}(y, X)$

2. Shift equivariance: for every value $d \in R^k$ and $q \in (0,1)$ will get;

$$\underline{\hat{\beta}}^{(r)}\left(\underline{y} + Xd, X\right) = \underline{\hat{\beta}}^{(r)}\left(\underline{y}, X\right) + d$$

3. Equivariance to reparameterization of design to make the matrix A have size $p \times p$ and $q \in (0,1)$ will get;

$$\underline{\hat{\beta}}^{(r)}\left(\underline{y}, XA\right) = A^{-1}\underline{\hat{\beta}}^{(r)}\left(\underline{y}, X\right)$$

Finding the regression quantifier (Q(r)th) at percentile position Q(20)th, Q(25)th , Q(50)th, Q(75)th and Q(80)th. In the simulation study, the model's probability density function (PDF) is calculated with the help of the quantile regression equation formula and properties of the random variable, with examples of symmetric distributions given where the variation of regression coefficients, percentage of outliers, and extent of deviation from the mean are taken into consideration. A careful choice of parameter values is practiced to cover a wide variety of situations. These simulations are conducted with the τ-function to approximate the kernel-based density function under the error term (ε), which is constructed to fulfill some statistical properties appropriate to the modeling conditions:

Step 1. Selecting the regression estimate from the function $\tau - function$. It is obtained by fitting the probability density function for the estimation by $\sum_{i=1}^{n} \tau\left(y_i - \hat{\beta}_1 - \hat{\beta}_2\underline{x'}_{i1} - \ldots - \hat{\beta}_p\underline{x'}_{ip}\right)$ the corresponding functions 1) $\tau(\varepsilon) \geq 0$, 2) $\tau(0) = 0$, 3) $\tau(\underline{\varepsilon}) = \tau(-\underline{\varepsilon})$, 4) $\tau(\varepsilon_i) \geq \tau(\varepsilon_j)$ for all values of $\lceil\varepsilon_i\rceil \geq \lceil\varepsilon_j\rceil$, represents $\underline{\boldsymbol{\varepsilon}}$ the error vector and 5) $\tau(\underline{\varepsilon}) = min(q\underline{\varepsilon}, (1-q)\underline{\varepsilon})$ (Conforming to Equation 4).

Step 2. Adjusting the probability density function from the outliers in equation (9) to $IQR = Q_3 - Q_1$ obtain $h = min\left(\sigma_\varepsilon, \frac{IQR}{1.34}\right) * n^{\left(\frac{-1}{5}\right)}$ in Equation 8, where σ_ε is the standard deviation of error.

Step 3. Calculating the Equation 6 to estimate the regression coefficient [27] from the distribution of the dependent variable at $E(Y|X) = X\underline{\hat{\beta}}^{(r)}$ the quartile values. Q(20)th, Q(25)th, Q(50)th, Q(75)th, Q(80)th and calculate the skewness (Sk) and kurtosis (Ku) values.

2-5-Simulation

Simulate the data 1,000 times from the population with the following distribution and show the process of the methodology in Figure1.

1) The independent variables are set to have a uniform distribution $\varepsilon_i \sim U(0,2)$. From the model $y_i = \beta_1 + \beta_1 x_{i1} + \ldots + \varepsilon_i$, $i = 1, 2, 3, 4, 5$ by setting the parameter value $\underline{\beta} = (10, 0.5, 2, 3, 4, 5)'$ when the number of independent variables $p = 2, 3, 4, 5$.

2) The error terms were specified to follow a normal distribution, $\varepsilon_i \sim N(0,3)$, and data were simulated accordingly. Three levels of sample size were considered to reflect different analytical contexts:

- Small sample size, typically used for preliminary analysis, was defined as fewer than 20 observations. Therefore, sample sizes of n = 8 and n = 15were selected.
- Moderate sample size, aimed at achieving more accurate estimation, was defined within the range 20 < n < 50. Accordingly, sample sizes of n = 20 and n = 30 were chosen.
- Large sample size, used to produce highly reliable analytical results, was defined as n > 50, with sample sizes of n = 50 and n = 100 included in the simulation.

3) To find the appropriate estimates of the multiple regression coefficient and the quantile regression coefficient. By using the absolute value of the mean error (MAE) in measuring the error when y_i substituted for the actual value, $\hat{y}_i$ substituted for the predicted value, and n substituted for the total number of sample units. The formula is as follows;

$$MAE = \frac{\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right|}{n} \qquad (11)$$

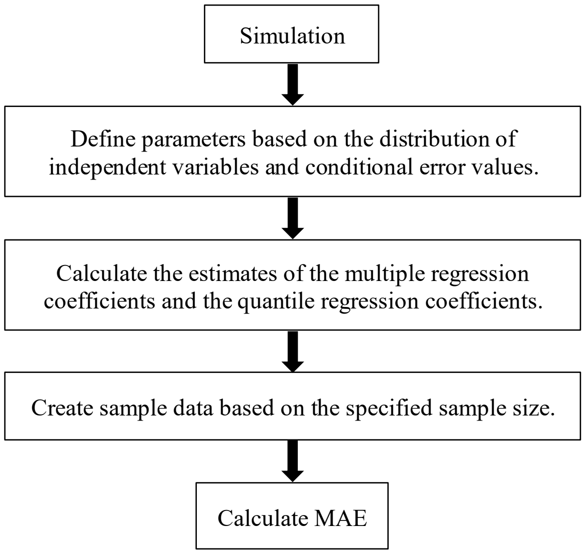


Figure 1. Flowchart of the process of the methodology

3- Results and Discussion

3-1-Simulation Results

Based on the simulation results, the data are divided into two main parts. The first part presents the estimated averages from the Minimum Risk Estimation (MRE) method and the quantile regression at the median position ($Q(r^{th})$), comparing their performance across different numbers of parameters and sample sizes. The second part compares the skewness and kurtosis values obtained from MRE and various quantile positions at different percentiles, also under varying sample sizes, as detailed below:

- The results show a comparison of the mean estimates from MRE and median quantile estimates, along with their standard deviations (standard errors) and mean absolute error (MAE) values, categorized by parameter values, as follows:

(1) Estimation of mean MRE and $Q(50)^{th}$ under the parameter p = 2

Table 1. Comparison of Std. Errors and MAE values under the parameter p = 2 for sample size 8, 15 and 20

Sample sizes	**8**		**15**		**20**	
Methods	**MRE**	**$Q(50)^{th}$**	**MRE**	**$Q(50)^{th}$**	**MRE**	**$Q(50)^{th}$**
Std. Errors	216.701	238.978	227.994	275,900	340.128	357.381
MAE	1.422	1.162	2.950	3.279	5.609	5.892
			Averages estimate			
$\hat{\beta}_1$	5.857 (-0.558)	0.033 (-0.003)	7.155 (-0.715)	0.060 (-0.006)	8.744 (-0.874)	0.107 (-0.010)
$\hat{\beta}_2$	3.282 (0.0164)	0.035 (0.000)	1.430 (0.007)	0.036 (0.000)	1.748 (0.008)	0.022 (0.000)
$\hat{\beta}_3$	3.484 (0.069)	0.048 (-0.001)	3.393 (0.067)	0.031 (-0.001)	3.438 (0.068)	0.019 (0.000)
Change from mean	0.157	0.001	0.213	0.002	0.265	0.003

Table 2. Comparison of Std. Errors and MAE values under the parameter p = 2 for sample size 30, 50 and 100

Sample sizes	30		50		100	
Methods	**MRE**	**Q(50)th**	**MRE**	**Q(50)th**	**MRE**	**Q(50)th**
Std. Errors	339.910	343.83	349.096	351.880	317.362	321.407
MAE	7.628	7.654	11.915	11.786	20.542	19.363
			Averages estimate			
$\hat{\beta}_1$	10.156 (1.015)	0.108 (0.010)	10.478 (1.047)	0.112 (-0.011)	11.637 (1.163)	0.112 (-0.011)
$\hat{\beta}_2$	1.328 (0.006)	0.016 (0.001)	0.893 (0.0044	0.003 (0.000)	0.990 (0.004)	0.015 (-0.001)
$\hat{\beta}_3$	2.593 (0.051)	0.018 (0.000)	2.842 (0.056)	0.023 (-0.001)	1.647 (-0.032)	0.010 (0.000)
Change from mean	0.357	0.003	0.369	0.003	0.378	0.004

As seen from Tables 1 and 2 (p = 2), the MAE values are to be compared in such a way that MRE yields lower MAE in small sample sizes (1.422 and 2.950). Similarly, $Q(50)^{th}$ also performs better in smaller samples with similar MAE values (1.162 and 3.279). Increasing the sample size, however, $Q(50)^{th}$ also performs better than MRE with lower MAE values (11.786 and 19.363). The MRE method exhibits variability around the mean for regression coefficients—left tail for smaller samples and right tail for larger samples. $Q(50)^{th}$ exhibits very minimal variability around the mean for all sample sizes and demonstrates higher stability and strength in the estimation of coefficients.

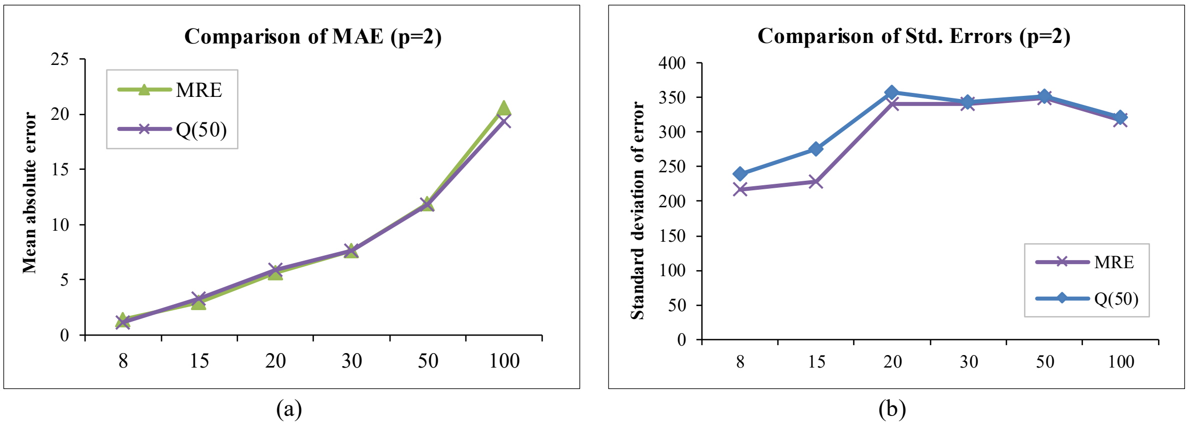


Figure 2. Comparison of MRE and $Q(50)^{th}$, under p = 2: (a) MAE, (b) Std. Errors

Figure 2 (p = 2) is a plot of MAE values and shows both procedures—MRE and $Q(50)^{th}$—giving the same results. The MAE values decrease, though, as the sample sizes become large, demonstrating better estimation precision with larger data sets. The MRE procedure provides smaller standard error values for small sample sizes compared to the $Q(50)^{th}$ procedure, but as the sample size increases, the standard errors of both procedures converge. The result implies that although MRE might do slightly better when confronted with small data sets, quantile regression is competitive and even stronger with larger sample data.

(2) Estimation of mean MRE and median $Q(50)^{th}$ under parameter p =3.

Table 3. Comparison of Std. Errors and MAE values under the parameter p=3 for sample size 8, 15 and 20

Sample sizes	8		15		20	
Methods	**MRE**	**Q(50)th**	**MRE**	**Q(50)th**	**MRE**	**Q(50)th**
Std. Errors	197.956	254.377	182.451	185.540	296.448	301.793
MAE	0.925	0.739	1.622	1.566	3.486	3.051
			Averages estimate			
$\hat{\beta}_1$	5.991 (-0.599)	0.039 (-0.003)	5.220 (-0.522)	0.062 (-0.006)	7.911 (-0.791)	0.049 (-0.004)
$\hat{\beta}_2$	4.088 (0.020)	0.045 (0.000)	3.409 (0.017)	0.033 (0.000)	4.013 (0.200)	0.047 (0.000)
$\hat{\beta}_3$	2.132 (0.042)	0.022 (0.000)	3.146 (0.062)	0.021 (0.000)	2.907 (0.058)	0.033 (-0.000)
$\hat{\beta}_4$	2.369 (-0.071)	0.037 (-0.001)	3.197 (0.095)	0.026 (-0.001)	1.805 (-0.054)	0.031 (-0.001)
Change from mean	0.151	0.001	0.086	0.001	0.146	0.001

Table 4. Comparison of Std. Errors and MAE values under the parameter p=3 for sample size 30, 50 and 100.

Sample sizes	30		50		100	
Methods	**MRE**	**Q(50)th**	**MRE**	**Q(50)th**	**MRE**	**Q(50)th**
Std. Errors	321.432	337.935	327.902	332.349	307.314	309.949
MAE	5.884	4.790	9.602	9.437	16.532	15.867
Averages estimate						
$\hat{\beta}_1$	8.270 (-0.827)	0.031 (-0.003)	8.602 (-0.860)	0.113 (-0.011)	9.847 (-0.984)	0.100 (-0.100)
$\hat{\beta}_2$	2.527 (0.012)	0.043 (0.000)	2.305 (0.011)	0.004 (0.000)	1.788 (0.008)	0.024 (0.000)
$\hat{\beta}_3$	2.934 (0.058)	0.045 (-0.001)	2.597 (0.051)	0.006 (-0.001	2.050 (0.041)	0.012 (0.000)
$\hat{\beta}_4$	3.337 (0.100)	0.046 (-0.001)	3.875 (0.116)	0.047 (-0.001)	3.604 (0.108)	0.032 (-0.001)
Change from mean	0.163	0.001	0.170	0.003	0196	0.025

Tables 3 and 4 (p = 3) provide a comparison of the variation in MAE values. The MRE procedure has smaller MAE for small samples (0.925 and 1.622). However, Q $(50)^{th}$ is always the smaller MAE for all samples, i.e., smaller samples (0.739 and 1.566), and it even dominates MRE in larger samples (9.437 and 15.867). On regression coefficient levels, the MRE method measures large departures from the mean in the direction of the left tail for any sample size. But $Q(50)^{th}$ measures the least variation away from the mean with the best stability of estimation and robustness.

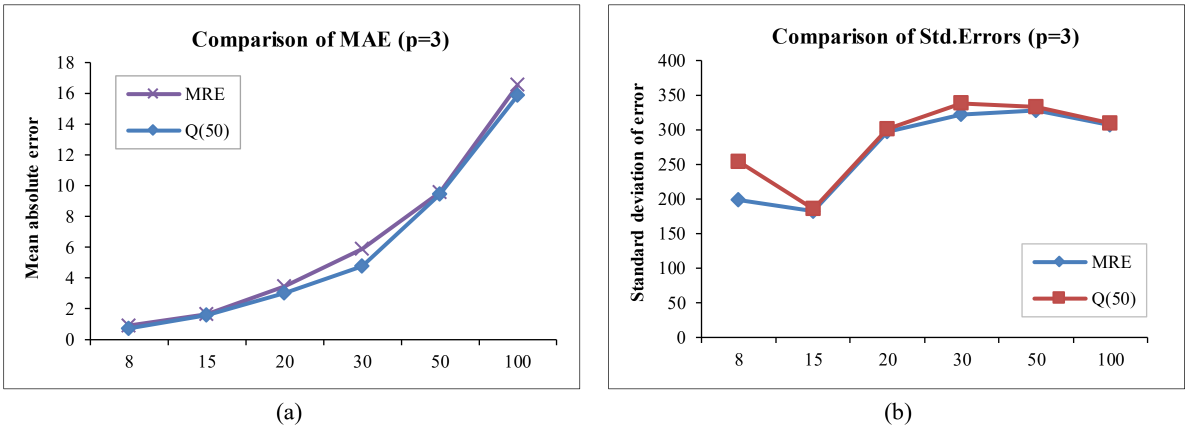


Figure 3. Comparison of MRE and Q $(50)^{th}$ under p = 3: (a) MAE, (b) Std. Errors

Figure 3 (p = 3) is a plot of MAE values and demonstrates that $Q(50)^{th}$ consistently has lower MAE irrespective of the sample size. In the case of standard errors, the MRE estimate has slightly lower standard errors for small sample sizes, but for large sample sizes, the standard errors of both methods are nearly indistinguishable from each other. This result is further evidence of the robustness of $Q(50)^{th}$ under different sample conditions.

(3) Estimation of mean MRE and median $Q(50)^{th}$ under parameter p = 4.

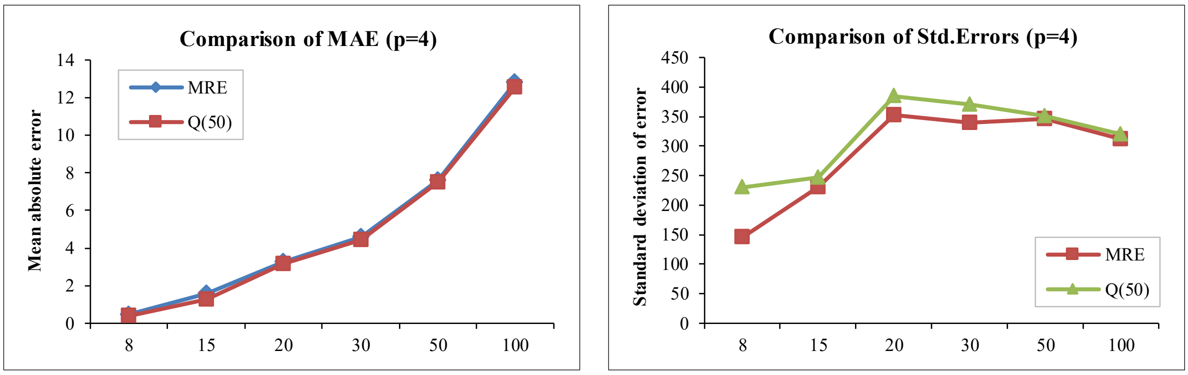
Table 5. Comparison of Std. Errors and MAE values under the parameter p = 4 for sample size 8, 15 and 20

Sample size	8		15		20	
Methods	**MRE**	**Q(50)th**	**MRE**	**Q(50)th**	**MRE**	**Q(50)th**
Std. Errors	145.442	230.003	230.173	247.290	352.566	384.290
MAE	0.478	0.389	1.566	1.296	3.284	3.157
Averages estimate						
$\hat{\beta}_1$	-3.510 (-1.351)	0.012 (-0.001)	9.045 (-0.904)	0.096 (-0.009)	10.616 (1.061)	0.161 (-0.016)
$\hat{\beta}_2$	4.729 (0.023)	0.044 (0.000)	1.719 (0.008)	0.022 (0.000)	1.030 (0.005)	-0.001 (-1.000)
$\hat{\beta}_3$	3.601 (0.072)	0.042 (0.000)	1.521 (0.030)	0.015 (0.000)	2.665 (0.053)	0.016 (0.000)
$\hat{\beta}_4$	7.540 (0.226)	0.043 (-0.001)	2.359 (-0.070)	0.011 (0.000)	2.609 (-0.078)	0.005 (0.000)
$\hat{\beta}_5$	7.438 (0.297)	0.050 (-0.002)	4.126 (0.165)	0.042 (-0.001)	3.625 (-0.145)	0.015 (0.000)
Change from mean	0.146	0.001	0.166	0.002	0.179	0.203

Table 6. Comparison of Std. Errors and MAE values under the parameter p = 4 for sample size 30, 50 and 100

Sample size	30		50		100	
Methods	**MRE**	**Q(50)th**	**MRE**	**Q(50)th**	**MRE**	**Q(50)th**
Std. Errors	340.148	370.788	345.658	351.677	312.663	319.316
MAE	4.617	4.430	7.588	7.489	12.803	12.551
Averages estimate						
$\hat{\beta}_1$	9.990 (-0.999)	0.097 (-0.009)	10.038 (1.003)	0.111 (-0.011)	10.602 (1.060)	0.104 (-0.010)
$\hat{\beta}_2$	1.146 (0.005)	0.003 (0.000)	1.090 (0.005)	0.001 (0.000)	0.272 (-0.001)	-0.001 (-0.002)
$\hat{\beta}_3$	2.629 (0.052)	0.028 (0.000)	1.764 (0.035)	0.018 (0.000)	1.791 (0.035)	0.020 (0.000)
$\hat{\beta}_4$	3.181 (0.095)	0.050 (-0.001)	4,000 (0.1200)	0.036 (-0.001)	3.916 (0.117)	0.046 (-0.001)
$\hat{\beta}_5$	4.151 (0.166)	0.032 (-0.001)	4.370 (0.174)	0.046 (-0.001)	4.693 (0.187)	0.042 (-0.001)
Change from mean	0.135	0.002	0.253	0.002	0.265	0.003

From Tables 5 and 6 (p = 4), the comparison of MAE values shows that the MRE method gives low MAE values in small sample sizes (0.478, 1.566), while Q(50)th gives low MAE values in all sizes, especially small sizes (0.389, 1.296), and it is also much lower in large sizes (7.489, 12.551). For the regression coefficients, MRE is different from the mean on the left side when the sample size is small and on the right side when the sample size is large, while Q(50)th remains close to the mean.



(a) (b)

Figure 4. Comparison of MRE and Q(50)th under p=4: (a) MAE, (b) Std. Errors

From Figure 4 (p = 4), it shows the comparison of MAE values. It was found that the Q(50)th method gives lower MAE values in all sample sizes as well as in the case of p=3. For the Std. Errors value, the MRE method gives lower values in both small and medium sample sizes. However, when the sample size is larger, the values of both methods are close to each other.

(4) Estimation of mean MRE and median Q(50)th under parameter p =5

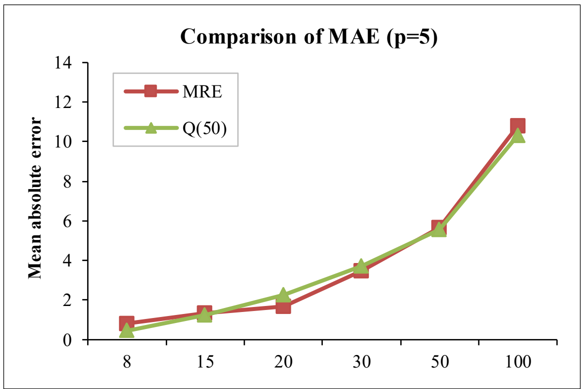
Table 7. Comparison of Std. Errors and MAE values under the parameter p=5 for sample size 8, 15 and 20

Sample sizes	8		15		20	
Methods	**MRE**	**Q(50)th**	**MRE**	**MRE**	**Q(50)th**	**MRE**
Std. Errors	250.352	364.377	236.654	250.352	364.377	236.654
MAE	0.819	0.448	1.332	0.819	0.448	1.332
Averages estimate						
$\hat{\beta}_1$	8.660 (-0.866)	0.142 (-0.014)	9.436 (-0.943)	8.660 (-0.866)	0.142 (-0.014)	9.436 (-0.943)
$\hat{\beta}_2$	1.939 (0.009)	0.048 (0.000)	1.626 (0.008)	1.939 (0.009)	0.048 (0.000)	1.626 (0.008)
$\hat{\beta}_3$	5.254 (0.105)	0.055 (-0.001)	1.828 (-0.036)	5.254 (0.105)	0.055 (-0.001)	1.828 (-0.036)
$\hat{\beta}_4$	9.581 (0.287)	0.077 (-0.002)	2.060 (-0.061)	9.581 (0.287)	0.077 (-0.002)	2.060 (-0.061)
$\hat{\beta}_5$	2.581 (-0.103)	-0.018 (-0.160)	4.031 (0.161)	2.581 (-0.103)	-0.018 (-0.160)	4.031 (0.161)
$\hat{\beta}_6$	-3.783 (-0.439)	-0.082 (-0.254)	4.822 (-0.241)	-3.783 (-0.439)	-0.082 (-0.254)	4.822 (-0.241)
Change from mean	0.167	0.072	0.185	0.167	0.072	0.185

Table 8. Comparison of Std. Errors and MAE values under the parameter p=5 for sample size 30, 50 and 100

Sample sizes	30		50		100	
Methods	**MRE**	**Q(50)th**	**MRE**	**Q(50)th**	**MRE**	**Q(50)th**
Std. Errors	323.898	387.169	335.313	347.430	315.310	319.930
MAE	3.475	3.728	5.664	5.536	10.804	10.325
Averages estimate						
$\hat{\beta}_1$	9.652 (-0.965)	0.116 (-0.011)	11.904 (1.190)	0.128 (-0.012)	11.837 (1.183)	0.106 (-0.010)
$\hat{\beta}_2$	1.600 (0.008)	-0.001 (-0.002)	1.361 (0.006)	0.024 (0.000)	1.160 (0.005)	0.008 (0.000)
$\hat{\beta}_3$	3.271 (0.065)	0.030 (-0.001)	2.742 (0.054)	0.023 (0.000)	2.323 (0.046)	0.019 (0.000)
$\hat{\beta}_4$	2.135 (-0.064)	0.033 (-0.001)	1.576 (0.047)	0.003 (0.000)	2.565 (-0.076)	0.027 (-0.001)
$\hat{\beta}_5$	4.907 (0.196)	0.059 (-0.002)	4.394 (0.175)	0.042 (-0.001)	3.892 (-0.155)	0.042 (-0.001)
$\hat{\beta}_6$	4.543 (-0.227)	0.021 (-0.001)	4.174 (-0.208)	0.041 (-0.002)	4.464 (-0.223)	0.052 (-0.002)
Change from mean	0.164	0.003	0.211	0.002	0.130	0.002

From Tables 7 and 8 (p = 5), the comparison of MAE values shows that the MRE method gives lower MAE than Q(50)th at small sample sizes (0.819, 1.332), but Q(50)th is superior at large sample sizes (5.536, 10.325). For regression coefficients, MRE is different from the left-tailed mean at all sample sizes, while Q(50)th remains the least changed.



(a)

(b)

Figure 5. Comparison of MRE and Q(50)th under p=5: (a) MAE, (b) Std. Errors

From Figure 5 (p = 5), it shows the comparison of MAE values. It was found that both methods gave similar MAE values at all sample sizes (similar to the case of p = 2). For the Std. Errors value, the MRE method gave lower values at small sample sizes, but when the sample size increased, the values of both methods became similar.

Based on the simulation results under parameter conditions p=2,3,4,5, the following can be concluded: First, in comparison to the mean absolute error (MAE), the MRE method tends to have a smaller MAE for small sample sizes. Yet, the Q(50)th quantile regression in all cases has low MAE with any sample size, and it performs very well with small samples and has much lower MAE in large samples as well. Secondly, based on the variability of estimates of regression coefficients, the MRE method produces estimates to be distant from the mean, the left side of the distribution in small samples and the right side in large samples. Contrarily, the Q(50)th approach has very little fluctuation in estimating coefficients, which is always very close to the mean in all circumstances.

- The results show a comparison of the skewness (Sk) and kurtosis (Ku) values of the quantile regression coefficient estimates at different percentile positions—Q(20)th, Q(25)th, Q(50)th, Q(75)th, Q(80)th—and the MRE method, categorized by parameter values. The findings are summarized as follows:

(1) Comparison of kurtosis (Ku) and skewness (Sk) values under the parameter p = 2

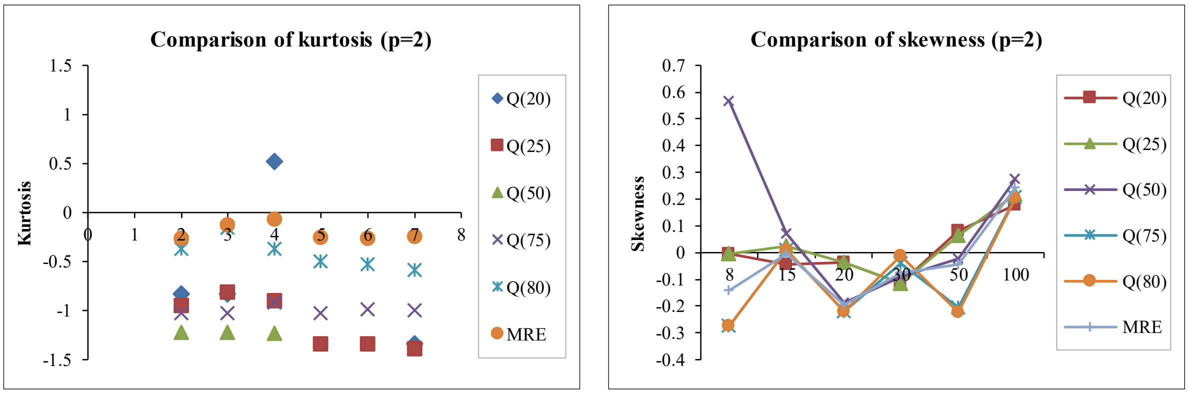
Table 9. Comparison of Ku and Sk values under the parameter p=2 for methods Q(20)th, Q(25)th and Q(50)th

Methods	Q(20)th		Q(25)th		Q(50)th,	
Sample sizes	Ku	Sk	Ku	Sk	Ku	Sk
8	-0.828	-0.004	-0.828	-0.004	-0.515	0.564
15	-0.953	-0.044	-0.813	-0.023	-0.896	0.071
20	-1.225	-0.035	-1.225	-0.035	-1.234	-0.187
30	-1.030	-0.114	-1.024	-0.114	-0.916	-0.089
50	-0.367	0.080	-0.158	0.063	-0.376	-0.023
100	-0.264	0.178	-0.132	0.219	-0.071	0.276

Table 10. Comparison of Ku and Sk values under the parameter p=2 for methods Q(75)th and Q(80)th

Methods	Q(75)th		Q(80)th		MRE	
Sample sizes	Ku	Sk	Ku	Sk	Ku	Sk
8	-2.201	-0.273	-2.201	-0.273	-1.339	-0.140
15	-1.341	0.009	-1.341	0.009	-1.393	-0.042
20	-1.529	-0.221	-1.530	-0.221	-1.528	-0.194
30	-1.025	-0.040	-0.993	-0.013	-1.001	-0.076
50	-0.501	-0.204	-0.531	-0.222	-0.591	-0.045
100	-0.256	0.207	-0.261	0.204	-0.249	0.243

From Tables 9 and 10, under the parameter value p=2p = 2p=2, the comparison of kurtosis values reveals that both the quantile regression estimates and the MRE method generally show decreasing kurtosis as the sample size increases. This trend is particularly noticeable at Q(20)th and Q(25)th. However, at Q(50)th, Q(75)th, Q(80)th, and with the MRE method, kurtosis values remain relatively high even as the sample size grows, indicating heavier tails. Regarding skewness, both methods display values lower than the standard level, and skewness decreases further with larger sample sizes. This suggests that the distributions become more centered around the mean, especially at the tails, as sample size increases.



(a) (b)

Figure 6. Comparison of MRE and Q(r th) under p=2: (a) Kurtosis, (b) Skewness

Figure 6, under the parameter value p = 2, presents a comparison of kurtosis values, showing that Q(20)th and Q(25)th exhibit the greatest concentration around the center, indicating lower peakedness and more balanced distributions. In the comparison of skewness, the MRE method and Q(80)th display the most pronounced skewness near the center, suggesting a stronger shift of the distribution toward one side while still centering around the mean.

(2) Comparison of kurtosis (Ku) and skewness (Sk) values under the parameter p = 3

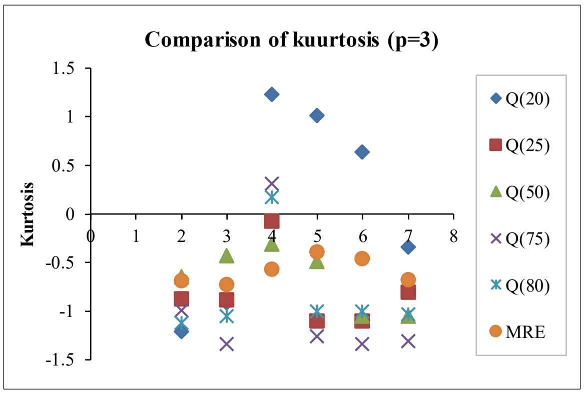
Table 11. Comparison of Ku and Sk values under the parameter p = 3 for methods Q(20)th, Q(25)th and Q(50)th

Methods	Q(20)th		Q(25)th		Q(50)th	
Sample sizes	Ku	Sk	Ku	Sk	Ku	Sk
8	-1.206	0.755	-0.915	0.8470	1.231	-0.753
15	-0.871	0.607	-0.884	0.6026	-0.070	0.712
20	-0.648	-0.115	-0.426	-0.0992	-0.310	-0.044
30	-0.991	0.159	-1.333	0.1196	0.312	0.055
50	-1.121	0.126	-1.047	0.1244	0.178	0.138
100	-0.689	0.264	-0.721	0.2445	-0.565	0.278

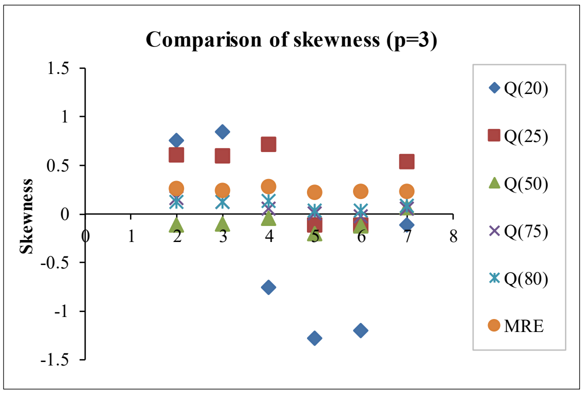
Table 12. Comparison of Ku and Sk values under the parameter p=3 for methods Q(75)th and Q(80)th

Methods	Q(75)th		Q(80)th		MRE	
Sample sizes	Ku	Sk	Ku	Sk	Ku	Sk
8	1.015	-1.277	0.639	-1.199	-0.337	-0.115
15	-1.101	-0.113	-1.100	-0.114	-0.802	0.543
20	-0.486	-0.198	-1.055	-0.125	-1.054	0.064
30	-1.260	0.001	-1.341	-0.021	-1.310	0.057
50	-1.003	0.038	-1.003	0.035	-1.035	0.086
100	-0.386	0.223	-0.453	0.234	-0.679	0.233

From Tables 11 and 12, under the parameter value p = 3, the comparison of kurtosis values shows that the regression coefficient estimates at Q(25)th exhibit a clear decrease in kurtosis as the sample size increases, indicating a move toward a more normal distribution. In contrast, the estimates at other quantile positions and the MRE method display fluctuating kurtosis values—alternating between high and low depending on the sample size. Regarding skewness, both the quantile regression methods and MRE show a decreasing trend as the sample size increases, suggesting that the distributions become more symmetric and centered around the mean with larger datasets.



(a)



(b)

Figure 7. Comparison of MRE and Q(r th) under p=3: (6.a) Kurtosis, (6.b) Skewness

Figure 7, under the parameter value p=3, presents a comparison of kurtosis values, showing that Q(50)th and MRE have the greatest spread around the center, indicating more balanced distributions with reduced peakedness. In terms of skewness, both MRE and Q(50)th exhibit the highest concentration of values near the center, suggesting that the data distributions are shifting closer to the mean and becoming more symmetric as sample size increases.

(3) Comparison of kurtosis (Ku) and skewness (Sk) values under the parameter p = 4

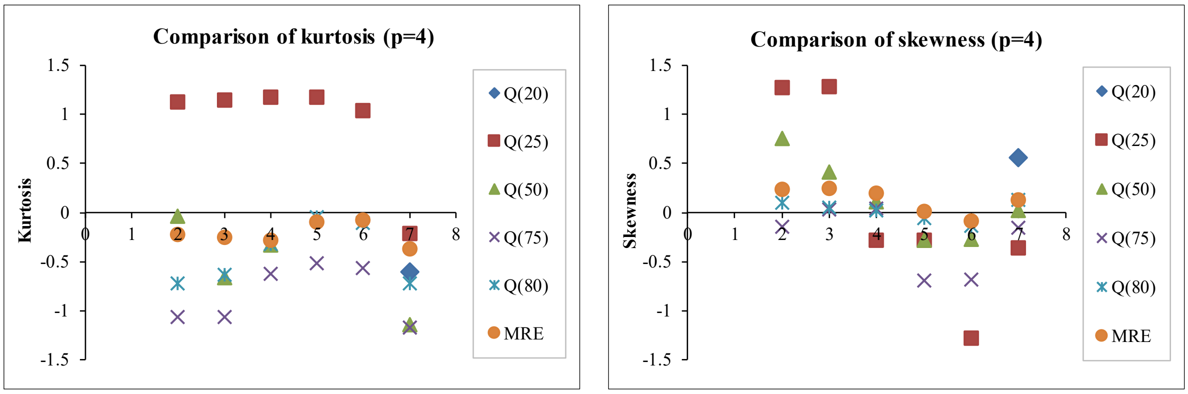
Table 13. Comparison of Ku and Sk values under the parameter p=4 for methods Q(20)th, Q(25)th and Q(50)th

Methods	Q(20)th		Q(25)th		Q(50)th,	
Sample sizes	Ku	Sk	Ku	Sk	Ku	Sk
8	6.811	2.544	4.631	2.098	6.730	2.491
15	1.135	1.281	1.146	1.284	1.181	-0.283
20	-0.036	0.754	-0.661	0.417	-0.325	0.109
30	-1.057	-0.137	-1.059	0.032	-0.618	0.043
50	-0.715	0.107	-0.628	0.056	-0.316	0.020
100	-0.221	0.242	-0.246	0.254	-0.274	0.196

Table 14. Comparison of Ku and Sk values under the parameter p = 4 for methods Q(75)th and Q(80)th

Methods	Q(75)th		Q(80)th		MRE	
Sample sizes	Ku	Sk	Ku	Sk	Ku	Sk
8	6.725	2.489	4.166	-1.980	-0.601	0.564
15	1.181	-0.283	1.043	-1.272	-0.207	-0.353
20	-1.713	-0.276	-1.722	-0.269	-1.135	0.023
30	-0.517	-0.693	-0.562	-0.676	-1.166	-0.149
50	-0.042	-0.049	-0.098	-0.136	-0.714	0.133
100	-0.095	0.012	-0.072	-0.085	-0.364	0.132

Tables 13 and 14, under the parameter value p = 4, present a comparison of kurtosis values. The results show that the regression coefficient estimates across different quantile positions, as well as the MRE method, initially exhibit very high kurtosis, which decreases as the sample size increases. This indicates a transition toward more normally distributed data with larger sample sizes. Regarding skewness, both methods demonstrate a consistent reduction in skewness as the sample size grows, suggesting that the data distributions become more symmetric and centered around the mean.



(a) (b)

Figure 8. Comparison of MRE and Q(r th) under p = 4: (a) Kurtosis, (b) Skewness

Figure 8, under the parameter value p = 4, presents a comparison of kurtosis values, showing that Q(75)th and the MRE method exhibit the greatest spread around the center, indicating lower peakedness and more balanced distributions. In the comparison of skewness, both MRE and Q(75)th show the most pronounced concentration around the center, suggesting that the data distributions are shifting closer to the mean and becoming more symmetric with increasing sample size.

(4) Comparison of kurtosis (Ku) and skewness (Sk) values under the parameter p = 5

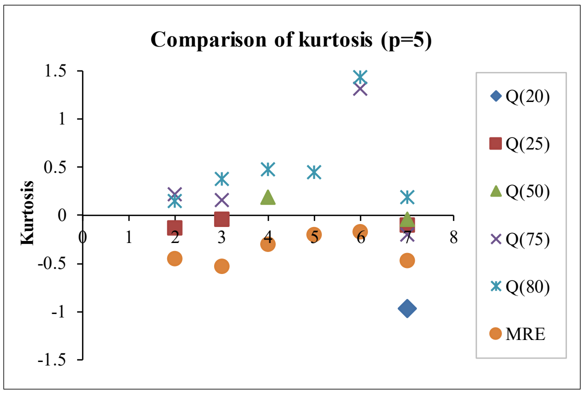
Table 15. Comparison of Ku and Sk values under the parameter p=5 for methods Q(20)th, Q(25)th and Q(50)th

Methods	Q(20)th		Q(25)th		Q(50)th,	
Sample sizes	Ku	Sk	Ku	Sk	Ku	Sk
8	7.975	2.822	7.970	2.821	7.970	2.821
15	-0.127	0.750	-0.035	0.065	1.547	0.056
20	-1.612	0.369	-1.608	0.372	0.197	-0.457
30	0.224	-0.303	0.160	-0.200	1.960	-0.754
50	0.153	0.034	0.380	-0.023	0.480	0.058
100	-0.442	0.184	-0.523	0.190	-0.296	0.228

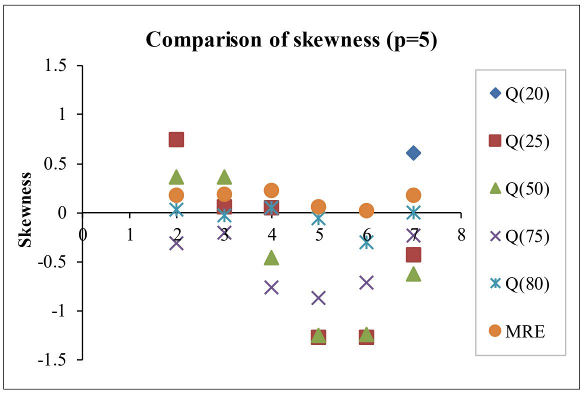
Table 16. Comparison of Ku and Sk values under the parameter p=5 for methods Q(75)th and Q(80)th

Methods	Q(75)th		Q(80)th		MRE	
Sample sizes	Ku	Sk	Ku	Sk	Ku	Sk
8	7.945	2.815	7.945	2.815	-0.958	0.616
15	2.007	-1.267	2.007	-1.267	-0.094	-0.426
20	1.954	-1.244	1.917	-1.237	-0.039	-0.618
30	1.751	-0.860	1.316	-0.707	-0.192	-0.226
50	0.451	-0.056	1.440	-0.299	0.194	0.001
100	-0.196	0.066	-0.160	0.029	-0.466	0.182

From Tables 15 and 16 under the parameter value p = 5, the results of the comparison of the kurtosis values are shown. It was found that the regression coefficient estimates at different positions and the MRE method were very high and decreased with increasing body size. As for the consideration of skewness, it was found that both methods had decreased skewness as the sample size increased.



(a)



(b)

Figure 9. **Comparison of MRE and Q(r)th under p = 5: (a) Kurtosis, (b) Skewness**

Figure 9, under the parameter value p = 5, presents a comparison of kurtosis values, showing that Q(25)th and the MRE method exhibit the widest spread around the center, indicating less peaked and more evenly distributed data. In terms of skewness, the MRE method, Q(25)th, and Q(75)th display the most noticeable distortion near the center, suggesting that the data are increasingly skewed toward the mean in the tails, resulting in a more centered distribution.

- An example: Export data to Russia from January 2021 to December 2022.

The worldwide 2019 coronavirus (COVID-19) pandemic [28], caused by a new strain of coronavirus called Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), was initially reported on 30 December 2019 by the Hubei Province Wuhan Health Office of China. The epidemic resulted in more than 8 million cases worldwide and lasted for more than two years, greatly influencing global and Thai economic output. Aside from the pandemic, there has been no progress shown in the Russia-Ukraine war since 24 February 2022. As Thailand has always had traditional trade transactions with Russia, it is the aim of this study to analyze the monthly volumes of major Thai products during those times. The chosen independent variables—rubber; mango, guava, and mangosteen; processed sweet corn; fresh fish; processed fish; and pet food—can hold six of the largest export categories from January 2021 to December 2022, with 24 months of information. No multicollinearity problem was detected between the variables. Data was split into two intervals: the first 20 months (January 2021 to August 2022) for model development and the last 4 months (September to December 2022) for testing prediction validity. Preliminary visualizations showed extreme product variability. Interestingly, demand for products moved in the opposite direction of these crises, i.e., pet food and natural rubber exports rose and orders of live and processed fish fell, as shown in Figure 10.

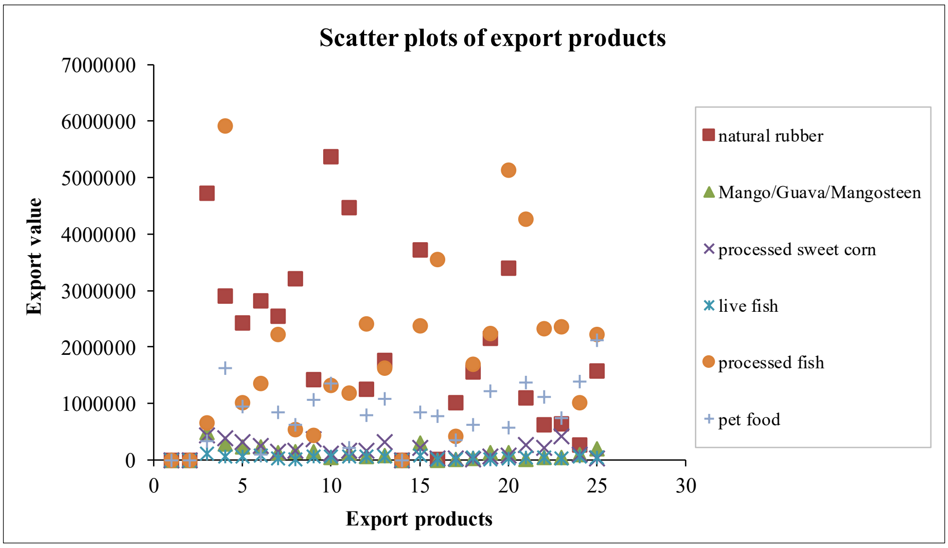


Figure 10. **Scatter plots of export products.**

Using Thailand's export value to Russia as the dependent variable and six key export products—natural rubber, mango/guava/mangosteen, processed sweet corn, live fish, processed fish, and pet food—as independent variables (treated as fixed parameters), the average estimates from both the MRE method and Q(r)th at the median position were evaluated. The comparative results of skewness and kurtosis values based on these estimations are summarized as follows:

Table 17. Comparison of Ku and Sk values Under the parameter p=6 for methods

Methods	Q(20)th		Q(25)th		Q(50)th,		Q(75)th		Q(80)th		MRE	
Distribution	Ku	Sk	Ku	Sk	Ku	Sk	Ku	Sk	Ku	Sk	Ku	Sk
	-0.318	0.749	-0.318	0.749	0.805	0.971	1.329	0.778	1.329	0.778	-0.413	0.014
Std. Errors	99388.037		99388.037		111,775.437		139184.970		139184.970		77091.199	
MAE	7.398		7.398		7.091		9.914		9.914		8.466	
$\hat{\beta}_1$	-0.021		-0.021		-0.021		-0.021		-0.021		1934.435	
$\hat{\beta}_2$	3.038×10^{-6}		3.038×10^{-6}		-2.356×10^{-5}		8.079×10^{-6}		8.079×10^{-6}		0.003	
$\hat{\beta}_3$	-2.356×10^{-5}		-2.356×10^{-5}		-2.348×10^{-5}		-2.348×10^{-5}		-2.348×10^{-5}		-0.005	
$\hat{\beta}_4$	2.790×10^{-5}		2.790×10^{-5}		2.920×10^{-5}		2.920×10^{-5}		2.920×10^{-5}		0.005	
$\hat{\beta}_5$	-2.600×10^{-5}		-2.600×10^{-5}		-2.600×10^{-5}		-2.600×10^{-5}		-2.600×10^{-5}		-0.001	
$\hat{\beta}_6$	9.070×10^{-7}		9.070×10^{-7}		-2.900×10^{-6}		-0.001×10^{-6}		-0.001×10^{-6}		-3.753×10^{-5}	
$\hat{\beta}_7$	4.836×10^{-6}		4.836×10^{-6}		7.522×10^{-6}		8.815×10^{-6}		8.815×10^{-6}		-0.006	

From Table 17, the comparison of MAE values indicates that Q(50)th yields the lowest MAE at 7.091. In terms of kurtosis, values closest to zero—indicating a more normal distribution—are observed for Q(20)th, Q(25)th, Q(50)th, and MRE. Regarding skewness, the MRE method produces a value closest to zero, suggesting a more symmetric distribution. Additionally, the estimated parameter values were applied to evaluate their predictive accuracy on a separate validation dataset covering four months (September to December 2022), allowing for assessment of deviation from actual values during this period.

Table 18. Percentage difference from actual value, September - December 2022

Methods	Q(20)th	Q(25)th	Q(50)th	Q(75)th	Q(80)th	MRE
September	0.310	0.310	0.201	0.055	0.055	0.193
October	0.153	0.153	0.244	0.402	0.402	1.432
November	0.583	0.583	0.373	0.223	0.223	0.429
December	0.303	0.303	0.100	0.476	0.476	0.976
MAE	0.013	0.013	0.009	0.011	0.011	0.031

Table 18 presents the differences between the estimated and actual values for both methods. The results show that the regression coefficient estimate from Q(50)th provides the closest match to the actual data, followed by Q(75)th and Q(25)th. In contrast, the MRE method yields the largest deviations from the real data. These findings suggest that when the dataset contains irregularities or deviates from typical patterns, quantile regression—particularly at the median—offers a more accurate and reliable predictive model compared to traditional methods.

3-2- Discussion

This research examined the estimation of regression coefficients by altering the probability density function (PDF) using a chosen τ-function with symmetric characteristics. The overall objective was to improve the precision and reliability of quantile regression estimates and compare their performance against usual multiple regression under both simulated and actual data scenarios [21, 22]. The findings confirm that the method is highly in line with recent literary trends calling for kernel-based and non-parametric methods to better quantify quantile regression estimation. For instance, Huang & Nguyen [29] proved how the use of kernel methods significantly decreases estimation error due to more precise modeling of difficult PDFs, a finding similar to the current study. Moreover, quantile regression provides a broader set of information about the data by estimating different points on the conditional distribution, in contrast to the usual least squares (OLS) regression, which only estimates the conditional mean. The multi-dimensionality can capture more refined knowledge of the relationship between the dependent and independent variables along the distribution, especially where data are not normally shaped [30]. Our results also validate that quantile regression works particularly well in dealing with skewed data or heavy-tailed distributions, as also found in earlier research by Chen et al. [31] and Santos et al. [32], who noted its outlier resistance and use in high-dimensional data settings.

Additional evidence that quantitative regression's good points are in the real application of data, considering the export figures, is shown at the median level, where it shows lower mean absolute error values than with the MRE method. This study is corroborating findings in Tang et al. [33] and Lee & Park [34], which hold that in terms of forecasting and decision-making, quantile regression is ahead. Furthermore, for big or heterogeneous samples, quantile regression will, in all circumstances, be better than OLS. As noted by Khaothong et al. [35], median-based quantile regression yields more robust conditional distribution estimates when data are highly variable or asymmetric. Perhaps the strongest factor for quantile regression reliability in this research was the symmetric PDF adjustment. Koenker [36] explicated how such an adjustment, while usually performed via kernel functions, is more versatile in approximating the actual shape of the data distribution. Our findings are congruent with Wang et al. [37], who highlight model specification and variable selection as key elements in quantile regression for addressing big data at an optimal level. The findings of this research indicate the integration of density corrections with robust estimation methods not only enhances fit but also sustains stability under changing sample conditions.

In addition to kernel-based estimation, several other parameter estimation methods were also investigated. The research proves the adequacy of methods like maximum likelihood estimation derived from the Fisher information matrix [38] and the Minimum Risk Estimation (MRE) method for certain situations. The MRE method, in particular, was found to be more suitable for small models or samples with few parameters, according to Nikitina et al. [39] and Ferrari & Paterlini [40]. Although MRE yields robust estimates under limited conditions, quantile regression is more versatile and precise in a broader range of data conditions, especially when the parameters or data size increases. Overall, the results of this current research support and supplement existing research using empirical data to back up the usability of quantile regression in dealing with asymmetric, skewed, and high-variance data. By applying symmetric τ-function-based PDF adjustments and performance assessment under a variety of conditions, this current research adds value to literature that calls for the use of quantile regression as a superior supplement to OLS and as a versatile response to existing data issues.

4- Conclusions

4-1-Simulation Results and Comparison of MRE and Quantile Regression

The simulation results indicate very large differences between the Minimum Risk Estimation (MRE) procedure and quantile regression according to estimation performance under different conditions. The increase in the number of parameters and sample size impacts considerably estimation accuracy first. MRE produced the lowest mean absolute error (MAE) for small and medium sample sizes, but quantile regression was superior to MRE for small and large sample sizes. This indicates that both methods are better adapted to other scales and the data conditions. Secondly, as regards the heteroscedasticity of the regression coefficients, quantile regression proved to be more stable since its estimates were closer to the mean and not very sensitive to data fluctuations. Third, in terms of distributional properties such as skewness and kurtosis, quantile regression reduced skewness by drawing outlier observations to the center and effectively eliminated kurtosis in bigger samples. Such results support that quantile regression produces more symmetrically distributed and less dispersed estimates and therefore is a safe alternative for heterogeneous data regimes.

4-2-Application to Real Data

Using real data from the two years of monthly export records of the six major Thai commodities to Russia exemplifies the usefulness of quantile regression. The approach gave the smallest MAE and best estimates between MRE and other estimators of quantiles. The results support the relevance of the quantile regression in analyzing asymmetrically or non-normally distributed data, especially when confronted with disturbance such as the COVID-19 crisis. The case study continued with the way external shocks such as the COVID-19 pandemic and the Russia–Ukraine war brought monumental volatility to export amounts. For example, the buying of rubber and pet food went up, and that of fresh and processed fish went down at some times. Volatility led to the effects of extremely dispersed, skewed, and heteroscedastic data.

Quantile regression in this case was superior to MRE, as it provided more information than the conditional mean. Estimating the relationship at different points in the distribution, involving the Q(25)th, Q(50)th, and Q(75)th, provides a better insight into independent variables' impact on different segments of the response variable. It is particularly helpful during times of crisis, with outliers and extreme values being the norm. On balance, evidence from simulation and real-data analysis verifies that quantile regression is a better and more stable method for modeling data in non-normal form under uncertain conditions.

5- Declarations

5-1-Author Contributions

Conceptualization, P.A. and W.R.; methodology, P.A.; software, P.A.; validation, P.A. and W.R.; formal analysis, P.A.; investigation, P.A.; resources, P.A.; data curation, P.A.; writing—original draft preparation, P.A.; writing—review and editing, P.A.; visualization, P.A.; supervision, P.A.; project administration, P.A.; funding acquisition, W.R. All authors have read and agreed to the published version of the manuscript.

5-2- Data Availability Statement

The data presented in this study are available in the article.

5-3- Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

5-4- Institutional Review Board Statement

Not applicable.

5-5- Informed Consent Statement

Not applicable.

5-6- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

6- References

[1] Draper, N. R., & Smith, H. (1998). Applied Regression Analysis. John Wiley & Sons, New Jersey, United States.

[2] Lee, J., & Kim, D. Robust quantile regression under heavy-tailed distributions with kernel-based density estimation. Statistics and Its Interface, 17(2), 210–225.

[3] Koenker, R., & Bassett, G. (1978). Regression Quantiles. Econometrica, 46(1), 33. doi:10.2307/1913643.

[4] Khaothong, K. (2019). Analysis of failing load and optimization of hot air welding parameters on PVC-acrylic coated polyester fabric by Taguchi and ANOVA technique. Engineering Journal, 23(6), 331–344. doi:10.4186/ej.2019.23.6.331.

[5] Bhattacharyya, H. T., Kleinbaum, D. G., & Kupper, L. L. (1979). Applied Regression Analysis and Other Multivariable Methods. Journal of the American Statistical Association, 74(367), 732. doi:10.2307/2287012.

[6] Robinson, A., Cook, R. D., & Weisberg, S. (1984). Residuals and Influence in Regression. Journal of the Royal Statistical Society. Series A (General), Chapman and Hall, Florid, United States. doi:10.2307/2981746.

[7] Lu, H., Dong, C., & Zhou, J. (2021). A Sequential Shrinkage Estimating Method for Tobit Regression Model. Open Journal of Modelling and Simulation, 09(03), 275–280. doi:10.4236/ojmsi.2021.93018.

[8] Moujahid, A., & Vadillo, F. (2021). A Comparison of Deterministic and Stochastic Susceptible-Infected-Susceptible (SIS) and Susceptible-Infected-Recovered (SIR) Models. Open Journal of Modelling and Simulation, 09(03), 246–258. doi:10.4236/ojmsi.2021.93016.

[9] Yu, K., & Moyeed, R. A. (2001). Bayesian quantile regression. Statistics & Probability Letters, 54(4), 437-447. doi:10.1016/S0167-7152(01)00124-9.

[10] Zhang, X., Wang, D., Lian, H., & Li, G. (2023). Nonparametric quantile regression for homogeneity pursuit in panel data models. Journal of Business & Economic Statistics, 41(4), 1238-1250. doi:10.1080/07350015.2022.2118125.

[11] He, X., Ng, P., & Portnoy, S. (1998). Bivariate quantile smoothing splines. Journal of the Royal Statistical Society Series B: Statistical Methodology, 60(3), 537-550. doi:10.1111/1467-9868.00138.

[12] Zhou, X. H., Lin, H., & Johnson, E. (2008). Non-parametric heteroscedastic transformation regression models for skewed data with an application to health care costs. Journal of the Royal Statistical Society Series B: Statistical Methodology, 70(5), 1029-1047. doi:10.1111/j.1467-9868.2008.00669.x.

[13] Koenker, R., & Xiao, Z. (2006). Quantile autoregression. Journal of the American Statistical Association, 101(475), 980–990. doi:10.1198/016214506000000672.

[14] Furno, M. (2014). Predictions in Quantile Regressions. Open Journal of Statistics, 4(7), 504–517. doi:10.4236/ojs.2014.47048.

[15] Boz, Ç. (2013). Estimating the New Keynesian Phillips Curve by Quantile Regression Method for Turkey. Modern Economy, 04(09), 627–632. doi:10.4236/me.2013.49067.

[16] Horowitz, J. L., & Lee, S. (2005). Nonparametric estimation of an additive quantile regression model. Journal of the American Statistical Association, 100(472), 1238–1249. doi:10.1198/016214505000000583.

[17] Hardle, W., Hall, P., & Ichimura, H. (2007). Optimal Smoothing in Single-Index Models. The Annals of Statistics, 21(1), 157–178. doi:10.1214/aos/1176349020.

[18] Hao, L., & Naiman, D. Q. (2007). Quantile Regression. Sage Publication, New Jersey, United States.

[19] Cui, W., & Wei, M. (2013). Strong Consistency of Kernel Regression Estimate. Open Journal of Statistics, 3(3), 179–182. doi:10.4236/ojs.2013.33020.

[20] Khalifa, E. H., Ramadan, D. A., & El-Desouky, B. S. (2021). Statistical Inference of Truncated Weibull-Rayleigh Distribution: Its Properties and Applications. Open Journal of Modelling and Simulation, 9(3), 281–298. doi:10.4236/ojmsi.2021.93019.

[21] Devroye, L., & Györfi, L. (1985). Density Estimation: The L1 View. John Wiley & Sons, New Jersey, United States.

[22] Efimov, V. (2023). Quantile loss & quantile regression. Towards Data Science, San Francisco, United States. Available online: https://towardsdatascience.com/quantile-loss-and-quantile-regression-b0689c13f54d/ (accessed on December 2025).

[23] Koenker, R., & d'Orey, V. (1987). Computing Regression Quantiles. Applied Statistics, 36, 383–393.

[24] Nilbai, T. (2020). Quantile regression. Economics, Ramkhamhaeng University, Bangkok, Thailand. Available online: http://www.eco.ru.ac.th/images/gallery/km/KMecon.pdf (accessed on December 2025).

[25] Fox, J. (2008). Applied Regression Analysis and Generalized Linear Models: Bootstraping Regression Model. Sage Publications, New Jersey, United States.

[26] Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. The Annals of Mathematical Statistics, 27(3), 832–837. doi:10.1214/aoms/1177728190.

[27] Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. The Annals of Mathematical Statistics, 33(3), 1065–1076. doi:10.1214/aoms/1177704472.

[28] Kanagarathinam, K., Ponkumar, G., & Sendil Kumar, S. (2024). Performance Analysis of Autoregressive Integrated Moving Average (ARIMA) and 'earlyR' Statistical Models for Predicting Epidemic Outbreaks: A Case Study on COVID-19 Data in India. Trends in Sciences, 21(1), 72–46. doi:10.48048/tis.2024.7246.

[29] Huang, M. L., & Nguyen, C. (2018). A nonparametric approach for quantile regression. Journal of Statistical Distributions and Applications, 5(1), 3. doi:10.1186/s40488-018-0084-9.

[30] Bühlmann, P. (2020). Rejoinder: Invariance, Causality and Robustness. Statistical Science, 35(3), 434–436. doi:10.1214/20-STS797.

[31] Chen, J., Wang, T., & Li, Z. Robust quantile regression for high-dimensional data. Statistica Sinica, 31(3), 1459–1482. doi:10.5705/ss.202019.0325.

[32] Santos, R., Pereira, J., & Oliveira, L. Skewed distributions and quantile regression: A simulation study. Statistics & Probability Letters, 179, 109348. doi:10.1016/j.spl.2021.109348.

[33] Tang, Y., Zhu, S., Luo, Y., & Duan, W. (2022). Input servitization, global value chain, and carbon mitigation: An input-output perspective of global manufacturing industry. Economic Modelling, 117, 106069. doi:10.1016/j.econmod.2022.106069.

[34] Lee, Y., & Park, C. Mean regression vs quantile regression: A comparative study. Computational Statistics, 35(4), 1671–1688. doi:10.1007/s00180-020-00967-2.

[35] Khaothong, K., Priyadumkol, J., Chaiworapuek, W., & Kaisinburasak, T. (2022). Optimization of High Frequency Welding Parameters of PVC Coating on Polyester Fabric. Trends in Sciences, 19(8), 34–63. doi:10.48048/tis.2022.3463.

[36] Koenker, R. (2017). Quantile regression. Annual Review of Economics, 9, 155-176.

[37] Wang, S., Li, F., & Chen, X. Model selection and variable screening for quantile regression in big data. Journal of Big Data, 11(1), 24. doi:10.1186/s40537-024-00512-y.

[38] Silverman, B. W. (2018). Density estimation: For statistics and data analysis. Density Estimation: For Statistics and Data Analysis. Chapman and Hall, Florid, United States. doi:10.1201/9781315140919.

[39] Nikitina, L., Paidi, R., & Furuoka, F. (2019). Using bootstrapped quantile regression analysis for small sample research in applied linguistics: Some methodological considerations. PloS one, 14(1), e0210668. doi:10.1371/journal.pone.0210668.

[40] Ferrari, D., & Paterlini, S. (2009). The Maximum L q-Likelihood Method: An Application to Extreme Quantile Estimation in Finance. Methodology and Computing in Applied Probability, 11(1), 3-19. doi:10.1007/s11009-007-9063-1.