A Direct Quantile Regression

Rachel Morris and Mei Ling Huang*

Department of Mathematics & Statistics, Brock University,
St. Catharines, Ontario, Canada L2S 3A1

September 14, 2018

Abstract

Estimating high conditional quantiles is an important problem. Many studies on this problem use a quantile regression (QR) method. The regular quantile regression method often sets a linear or non-linear model which estimates the coefficients in the model to obtain the estimated conditional quantile. The real-world applications may be restricted by this approach's model setting. This paper proposes a direct nonparametric quantile regression method to overcome this restriction. Monte Carlo simulations show good efficiency for the proposed nonparametric QR estimator relative to the regular QR estimator. The paper also investigates a real-world example of applications by using the proposed method, and gives comparisons of the proposed method and existing methods.

Keywords: *Conditional quantile, goodness of fit, extreme value distribution, Gumbel's second kind of bivariate exponential distribution, nonparametric regression.*

AMS 2010 Subject Classifications: primary: 62G32; secondary: 62J05

1. Introduction

It is important to study quantile regression (Koenker, 2005) to estimate high conditional quantiles in real-world extreme events. Some extreme events can cause damages to society: stock market crashes, heavy rain falls, forest fires, pollution, earth quakes and hurricanes. We are interested in random variables with extreme events which usually are heavy tailed distributed. Such that estimation of high conditional quantiles of a heavy tailed distributed random variable y is an important problem.

The traditional mean linear regression estimates the conditional expectation $E(y|\mathbf{x})$, where $\mathbf{x} = (1, x_1, x_2, ..., x_k)^T,\ \mathbf{x} \in R^p, p = k + 1$. The linear mean regression model assumes

$$\mu_{y|\mathbf{x}} = E\left(y|x_1, x_2, ..., x_k\right) = \mathbf{x}^T\boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k. \tag{1}$$

*This research is supported by *the Natural Science and Engineering Research Council of Canada (NSERC) grant MLH, RGPIN-2014-04621.*

We estimate $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_k)^T$, where $\boldsymbol{\beta} \in R^p$, from a random sample $\{(y_i, \mathbf{x}_i), i = 1, ..., n\}$, where $\mathbf{x}_i$ is the p-dimensional design vector and y_i is the univariate response variable from a continuous distribution with a cumulative distribution function (c.d.f.) $F(y)$. The least square (LS) estimator $\widehat{\boldsymbol{\beta}}_{LS}$ is a solution to the following equation

$$\widehat{\boldsymbol{\beta}}_{LS} = \arg\min_{\boldsymbol{\beta} \in R^p} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2, \tag{2}$$

where $\widehat{\boldsymbol{\beta}}_{LS}$ is obtained by minimizing the L_2-distance.

Traditional mean regression is limited on its abilities to predict extreme events. When dealing with heavy tailed distributions, the traditional method simply cannot reach extreme events in a dataset. We now investigate a real-world example in extreme events to show the limitations of traditional mean regression, and later in Section 4 we will express why quantile regression is needed for these heavy tailed data sets.

Example: Amazon Stock Market Example

Amazon's (AMZN) stock had a great year in 2017. The stock surpassed a closing price of $1,000 on June 2, 2017. Two variables of interest for stock traders are closing price and volume. For this example we focus on the latter. Daily volume for a stock is the total amount of trading occurrences for the stock each day, therefore whether a stock is purchased or sold the volume increases. Volume is a great indicator for how liquid the stock is, i.e. how fast it can be sold if bought. Predicting the volume of a stock can be just as important as predicting the price of the stock for certain traders. We will investigate the effect of the opening price of AMZN on total daily volume. Therefore our response variable is the total daily volume of Amazon's stock and the regressor will be the opening price for that day. The data was collected (Yahoo Canadian Finance, 2018) for the last 10 years.

Table 1 lists the top 10 volumes. Investing websites (Investopedia, 2010) have mentioned that a volume is high if it is above 2 times the average. For 2008-2017 Amazon stock, 2,518 data points, the average volume is 5,427,194. In this paper, we are only interested in high volumes, so we set a threshold of 10,854,389, which is double the average volume. There are 182 trading days with volumes above the threshold.

Table 1. Top 10 Data of Amazon Total Daily Volume

Opening Price of AMZN (Canadian $)	Total Daily Volume
111.05	58,305,800
68.91	42,885,900
105.93	42,421,100
57.36	39,783,100
123.18	37,774,400
43.37	32,601,900
119.21	32,271,200
76.30	30,996,600
129.77	29,471,300
124.43	27,293,100

We apply a quadratic mean regression model to the data,

$$\mu_{y|x} = E(y|x) = \beta_0 + \beta_1 x + \beta_2 x^2, \quad (3)$$

where y is the square root of the daily volume of AMZN and x is the square root of the opening price that day. Note that we do data transformation for fitting a heavy tailed distribution.

Figure 1 shows the scatter plot and the LS mean regression line $\widehat{\mu}_{y|x}$ of the square root of the daily volume of AMZN related to the square root of the opening price that day ($n = 182$), by using formula *(2)*. Where

$$\widehat{\mu}_{LS} = \widehat{\mu}_{y|x} = 3806.81 + 48.810x - 1.544x^2.$$

There seems to be a small decrease in the LS mean square root volume when the square root opening price is increased, however the relationship looks likely not strong, and the mean regression estimate was not able to reach the higher quantile volumes.

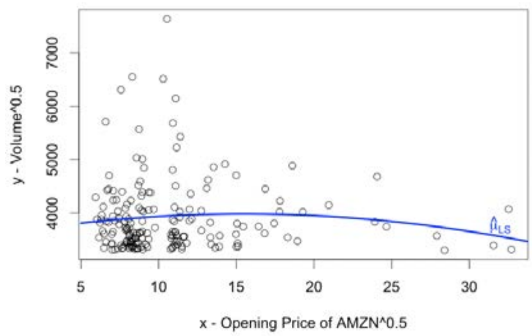


Figure 1. The scatter plot and LS mean regression line, $\widehat{\mu}_{LS}$, of the square root of the daily volume of AMZN related to the square root of the opening price that day ($n = 182$).

In this paper, we use quantile regression method, we wish to estimate high conditional quantiles of a random variable y with a c.d.f. $F(y)$ given a variable vector, $\mathbf{x} = (x_1, x_2, ..., x_d)$, and $\mathbf{x}_p = (1, x_1, x_2, ..., x_d)^T \in R^p$ where $p = d+1$. The τth conditional linear quantile is defined by

$$Q_y(\tau|\mathbf{x}) = Q_y(\tau|x_1, x_2, ..., x_d) = F^{-1}(\tau|\mathbf{x}), \ 0 < \tau < 1. \quad (4)$$

The traditional quantile regression is concerned with the estimation of the τth conditional quantile regression (QR) of y for given $\mathbf{x}$ which often sets a linear model as

$$Q_y(\tau|\mathbf{x}) = \mathbf{x}_p^T \boldsymbol{\beta}(\tau) = \beta_0(\tau) + \beta_1(\tau)x_1 + \cdots + \beta_d(\tau)x_d, \ 0 < \tau < 1. \quad (5)$$

For linear model *(5)*, we estimate the coefficient $\boldsymbol{\beta}(\tau) = (\beta_0(\tau), \beta_1(\tau), \beta_2(\tau), ..., \beta_d(\tau))^T \in R^p$ from a random sample $\{(y_i, \mathbf{x}_i), \ i = 1, ..., n\}$, where $\mathbf{x}_{pi} = (1, x_{i1}, x_{i2}, ..., x_{id})^T$ is the p-dimensional design vector and y_i is the univariate response variable from a continuous distribution with a c.d.f. $F(y)$. Koenker and Bassett (1978) proposed an L_1-weighted loss function to obtain estimator $\widehat{\boldsymbol{\beta}}(\tau)$ by solving

$$\widehat{\boldsymbol{\beta}}(\tau) = \arg \min_{\boldsymbol{\beta}(\tau) \in R^p} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_{pi}^T \boldsymbol{\beta}(\tau)), \ 0 < \tau < 1, \quad (6)$$

where ρ_τ is a loss function, namely

$$\rho_\tau(u) = u(\tau - I(u < 0)) = \begin{cases} u(\tau - 1), u < 0; \\ u\tau, \ u \geq 0. \end{cases}$$

The linear quantile regression problem can be formulated as a linear program

$$\min_{(\boldsymbol{\beta}(\tau),\mathbf{u},\mathbf{v}) \in R^p \times R_+^{2n}} \{\tau \mathbf{1}_n^T \mathbf{u} + (1-\tau)\mathbf{1}_n^T \mathbf{v} | \mathbf{X}\boldsymbol{\beta}(\tau) + \mathbf{u} - \mathbf{v} = \mathbf{y}\},$$

where $\mathbf{1}_n^T$ is an n-vector of 1s, $\mathbf{X}$ denotes the $n \times p$ design matrix, and $\mathbf{u}, \mathbf{v}$ are $n \times 1$ vectors with elements of u_i, v_i respectively (Koenker, 2005).

In recent years, studies are looking for efficiency improvements of estimator *(6)* (Wang and Li, 2013; Huang *et al.* 2015, Huang and Nguyen, 2017). The regular linear quantile regression *(5)* needs the estimator $\widehat{\boldsymbol{\beta}}(\tau)$ in *(6)* for the estimated conditional quantile curves. But this estimated conditional quantile curve may be restricted under the model setting.

To overcome the limitation of the model setting in *(5),* this paper proposes a direct nonparametric quantile regression method which uses the ideas of nonparametric kernel density estimation and nonparametric kernel regression to obtain a direct estimator of conditional quantile curves. In order to test if the new method is an improvement relative to the regular linear quantile regression, we will do two studies in this paper:

1. Monte Carlo simulations will be performed to confirm the efficiency of the new direct QR estimator relative to the regular QR estimator.
2. The new proposed method will be applied to the Amazon Stock Market example and compared with the regular QR method.

In Section 2, we propose a direct nonparametric quantile regression estimator. In Section 3, the results of Monte Carlo simulations generated from Gumbel's second kind of bivariate exponential distribution (Gumbel, 1960) show that the proposed direct method produces high efficiencies relative to existing linear QR methods. In Section 4, the regular linear quantile regression and the proposed direct quantile regression are applied to the Amazon Stock Market example from this Section. The study of this example illustrates that the proposed direct quantile regression model fits the data better than the existing linear quantile regression method.

2. Proposed Direct Nonparametric Quantile Regression

In this paper, for generality, we ignore the idea of the linear model *(5).* We obtain a direct estimator for the true conditional quantile in *(4):*

$$\widehat{Q}_y(\tau|\mathbf{x}) = \widehat{Q}_y(\tau|x_1, x_2, ..., x_d) = \widehat{F}^{-1}(\tau|\mathbf{x}),$$

by using local conditional quantile estimator $\xi_i(\tau|\mathbf{x}_i) = Q_y(\tau|\mathbf{x}_i)$ based on the ith point of given random sample, $\{(y_i, \mathbf{x}_i), i = 1, ..., n\}$, for $\mathbf{x}_i = (x_{1i}, x_{2i}, ..., x_{di})^T$.

We construct the following five-step algorithm of a direct nonparametric quantile regression:

Step 1: Estimate the conditional density of y for given $\mathbf{x} = (x_1,x_2, ...,x_d)$ using a kernel density estimation method (Silverman, 1986; Scott, 2015):

$$\widehat{f}(y|\mathbf{x}) = \frac{\widehat{f}(y, \mathbf{x})}{\widehat{g}(\mathbf{x})}, \tag{7}$$

where $\widehat{f}(y,\mathbf{x})$ is an estimator of the joint density of y and $\mathbf{x}$, and $\widehat{g}(\mathbf{x})$ is an estimator of the marginal density of $\mathbf{x}$.

A d-dimensional kernel density estimator from a random sample $\mathbf{X}_i = (X_{1i}, X_{2i}, ..., X_{di}),\ i = 1, 2, ..., n$, from a population $\mathbf{x} = (x_1, x_2, ..., x_d)$ for joint density $g(\mathbf{x})$, is given by

$$\widehat{g}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left\{\frac{\mathbf{x} - \mathbf{X}_i}{h}\right\},$$

where $h > 0$ is the bandwidth and the kernel function $K(\mathbf{x})$ is a function defined for d-dimensional $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_d)$ which satisfies $\int\limits_{R^d} K(\mathbf{x})d\mathbf{x} = 1$.

If a multivariate normal kernel is used for smoothing the normal distribution data with unit variance,

$$h_{opt} = \left\{\frac{4}{d+2}\right\}^{1/(d+4)} n^{-1/(d+4)}.$$

Step 2: Estimate the conditional c.d.f. of y given $\mathbf{x}$:

$$\widehat{F}(y|\mathbf{x}) = \int_{-\infty}^{y} \widehat{f}(y|\mathbf{x})dy.$$

Step 3: Estimate the local conditional quantile function $\xi(\tau|\mathbf{x})$ of y given $\mathbf{x}$ by inverting an estimated conditional c.d.f. $\widehat{F}(y|\mathbf{x})$.

$$\widehat{\xi}(\tau|\mathbf{x}) = \widehat{Q_y}(\tau|\mathbf{x}) = \inf\{y : \widehat{F}(y|\mathbf{x}) \geq \tau\} = \widehat{F}^{-1}(\tau|\mathbf{x}).$$

To avoid the computational difficulties of $\widehat{\xi}(\tau|\mathbf{x})$, we estimate the local conditional quantile function $\xi_i(\tau|\mathbf{x}_i)$ of y given $\mathbf{x}_i$ by inverting $\widehat{F}(y|\mathbf{x}_i)$ in Step 2 at the ith data point $(y_i, \mathbf{x}_i)$:

$$\widehat{\xi_i}(\tau|\mathbf{x}_i) = \widehat{Q_y}(\tau|\mathbf{x}_i) = \inf\{y : \widehat{F}(y|\mathbf{x}_i) \geq \tau\} = \widehat{F}^{-1}(\tau|\mathbf{x}_i), \quad i = 1, 2, ..., n. \qquad (8)$$

Step 4: We propose a direct nonparametric quantile regression estimator for the τth conditional quantile curve of $\mathbf{x}$ by using Nadaraya-Watson (NW) nonparametric regression estimator on $\left(\mathbf{x}_i, \widehat{\xi_i}(\tau|\mathbf{x}_i)\right),\ i = 1, 2, ..., n$:

$$Q_D(\tau|\mathbf{x}) = \widehat{\xi}(\tau|\mathbf{x}) = \frac{\sum\limits_{i=1}^{n} K\left\{\frac{\mathbf{x}-\mathbf{X}_i}{\mathbf{h}}\right\} \widehat{\xi_i}(\tau|\mathbf{x}_i)}{\sum\limits_{j=1}^{n} K\left\{\frac{\mathbf{x}-\mathbf{X}_j}{\mathbf{h}}\right\}} = \sum_{i=1}^{n} W_{h_{\mathbf{x}}}(\mathbf{x}, \mathbf{X}_i)\widehat{\xi_i}(\tau|\mathbf{x}_i), \quad 0 < \tau < 1, \qquad (9)$$

where $W_{h_x}(\mathbf{x}, \mathbf{X}_i)$ is called an equivalent kernel, and $\mathbf{h} = (h_1, ..., h_d)$,

$$W_{h_{\mathbf{x}}}(\mathbf{x}, \mathbf{X}_i) = \frac{K\left\{\frac{\mathbf{x}-\mathbf{X}_i}{\mathbf{h}}\right\}}{\sum\limits_{j=1}^{n} K\left\{\frac{\mathbf{x}-\mathbf{X}_j}{\mathbf{h}}\right\}}, \quad i = 1, 2, ..., n,$$

where

$$K\left\{\frac{\mathbf{x}-\mathbf{X}_i}{\mathbf{h}}\right\} = \frac{1}{nh_1...h_d}\prod_{j=1}^{d} K\left(\frac{x-x_{ij}}{h_j}\right), \quad i=1,...,n,$$

where K is the kernel function, and $h_j > 0$ is the bandwidth for the jth dimension.
Step 5: Check all procedures, and make any necessary adjustments.

3. Simulations

For investigating the proposed direct nonparametric quantile regression estimator in *(9),* in this Section, Monte Carlo simulations are performed. We generate m random samples with size n each from the second kind of Gumbel's bivariate exponential distribution (Gumbel, 1960) with a non-linear conditional quantile function of y given x in *(11).* It has c.d.f. $F(x,y)$:

$$F(x,y) = (1-e^{-x})(1-e^{-y})(1+\alpha e^{-(x+y)}),\ x \geq 0,\ y \geq 0,\ \alpha > 0, \tag{10}$$

The true τth conditional quantile function of y given x of *(10)* is

$$\xi(\tau|x) = Q_y(\tau|x) = \ln\left(\frac{2\alpha(2e^{-x}-1)}{\alpha(2e^{-x}-1)-1+\sqrt{(\alpha(2e^{-x}-1)+1)^2-4\alpha\tau(2e^{-x}-1)}}\right), \tag{11}$$
$$x \geq 0,\ \alpha > 0,\ 0 < \tau < 1.$$

We use two quantile regression methods:

1. The regular quantile regression $Q_R(\tau|x)$ estimation based on *(5):*

$$Q_R(\tau|x) = \widehat{\beta}_0(\tau) + \widehat{\beta}_1(\tau)x. \tag{12}$$

2. The direct nonparametric quantile regression $Q_D(\tau|x)$ estimation based on *(9)*

$$Q_D(\tau|x) = \sum_{i=1}^{n} W_{h_{\mathbf{x}}}(\mathbf{x},\mathbf{X}_i)\widehat{\xi}_i(\tau|x_i), \quad 0<\tau<1,\ i=1,2,...,n, \tag{13}$$

where $\widehat{\xi}_i(\tau|x_i)$ is obtained by *(8).*

For each method, we generate sample size $n = 100$, for $m = 100$ samples. $Q_{R,i}(\tau|x)$ and $Q_{D,i}(\tau|x),\ i = 1,2,...,m$, are estimated in the ith sample. Let $\alpha = 1$ in *(11).* Then the true τth conditional quantile is

$$\xi(\tau|x) = Q_y(\tau|x) = \ln\left(\frac{2e^{-x}-1}{e^{-x}-1+\sqrt{e^{-2x}-\tau(2e^{-x}-1)}}\right),\ x \geq 0,\ 0<\tau<1. \tag{14}$$

The simulation mean squared errors (SMSEs) of the estimators *(12)* and *(13)* are:

$$SMSE(Q_R(\tau|x)) = \frac{1}{m}\sum_{i=1}^{m}\int_0^N (Q_{R,i}(\tau|x)-Q_y(\tau|x))^2dx; \tag{15}$$

$$SMSE(Q_D(\tau|x)) = \frac{1}{m}\sum_{i=1}^{m}\int_0^N (Q_{D,i}(\tau|x)-Q_y(\tau|x))^2dx, \tag{16}$$

where the true τth conditional quantile $Q_y(\tau|x)$ is defined in *(14)*. N is a finite x value such that the c.d.f. in *(10)* $F(N, N) \approx 1$. We take $N = 6$ and the simulation efficiencies (SEFFs) are given by

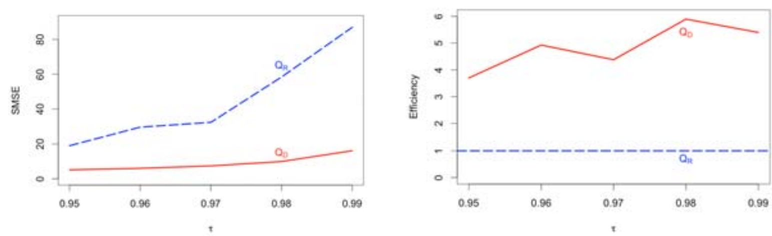
$$SEFF(Q_D(\tau|x)) = \frac{SMSE(Q_R(\tau|x))}{SMSE(Q_D(\tau|x))},$$

where $SMSE(Q_R(\tau|x))$ and $SMSE(Q_D(\tau|x))$ are defined in *(15)* and *(16)*, respectively.

Table 2. Simulation Mean Square Errors (SMSEs) and Efficiencies (SEFFs) of Estimating $Q_y(\tau|x), m = 100, n = 100, N = 6$.

τ	0.95	0.96	0.97	0.98	0.99
$SMSE(Q_R(\tau\|x))$	18.97	29.63	32.37	58.36	86.84
$SMSE(Q_D(\tau\|x))$	5.13	6.02	7.39	9.90	16.09
$SEFF(Q_D(\tau\|x))$	**3.70**	**4.93**	**4.38**	**5.89**	**5.39**

Table 2 and Figure 2 shows that all of the $SEFF(Q_D(\tau|x))$ are larger than 1 when $\tau =$ 0.95,..., 0.99. Overall, we can conclude that for $\tau = 0.95, ..., 0.99$, the proposed direct estimator $Q_D(\tau|x)$ in *(13)* is more efficient relative to the regular quantile regression $Q_R(\tau|x)$ in *(12)*.



(a) (b)

Figure 2. (a) $SMSE(Q_D(\tau|x))$ is the red solid line, $SMSE(Q_R(\tau|x))$ is the blue dash line. (b) $SEFF(Q_D(\tau|x))$ is the red solid line, $SEFF(Q_R(\tau|x)) \equiv 1$ is blue dash line.

Figure 3 shows the average curves of 100 estimated $\tau = 0.95$th quantile curves of the $Q_R(\tau|x)$ and $Q_D(\tau|x)$. The average $Q_D(\tau|x)$ is much closer than $Q_R(\tau|x)$ to the true quantile curve *(14)*.

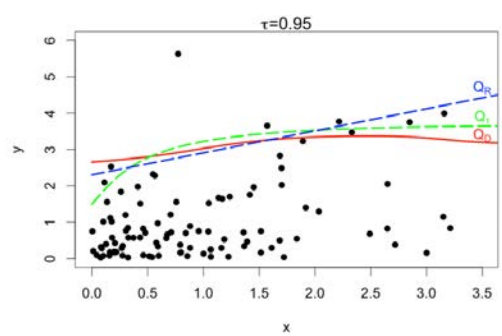


Figure 3. In $n = 100, m = 100, \tau = 0.95$ simulations, the true Quantile-green dash, the average regular Q_R-blue dash and the average direct Q_D-red solid.

4. Amazon Stock Market Example

We apply two quantile regression models to the Amazon Stock Market example in Section 1:

1. The regular quantile regression $Q_R(\tau|\mathbf{x})$ in model *(5)* using estimator $\widehat{\beta}(\tau)$ in *(6)*;
2. The direct nonparametric quantile regression $Q_D(\tau|\mathbf{x})$ in *(9)*.

The estimated quantiles in this Section are for the population of the AMZN daily volumes greater than the threshold 10,854,389. We also compare these models with the LS mean regression applied in Section 1.

At first, we use the following linear polynomial quantile regression model for this example:

$$Q_y(\tau|x) = \beta_0(\tau) + \beta_1(\tau)x + \beta_2(\tau)x^2,$$

where y is the square root of the daily volume of AMZN and x is the square root of the opening price that day.

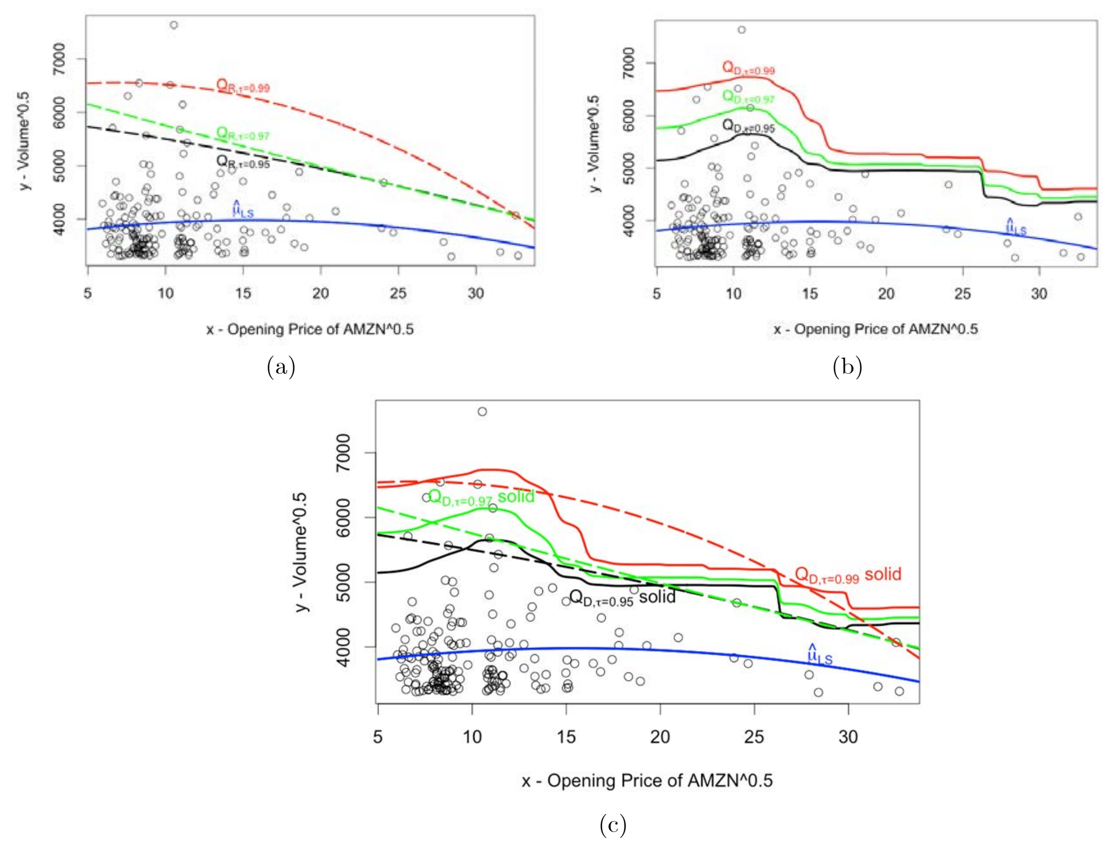


Figure 4. For Amazon Stock Market example, data − black, $n = 182$, the LS mean regression line $\widehat{\mu}_{LS}$ is blue solid. (a) Regular Q_R− dash; (b) Direct Q_D− solid; (c) Both of the Regular Q_R and Direct Q_D in a plot at $\tau = 0.95$ in black, $\tau = 0.97$ in green and $\tau = 0.99$ in red.

In this paper we use the proposed five-step algorithm in Section 2 to obtain the new direct nonparametric quantile estimator $Q_D(\tau|\mathbf{x})$ in *(9),* then compare it with the regular quantile estimator $Q_R(\tau|\mathbf{x})$ by using *(6).* Figures 4(a), (b) and (c) show the scatter plot of the AMZN daily volumes vs. opening price with mean LS curve $\widehat{\mu}_{LS}$, the fitted Q_R and Q_D quantile curves at $\tau = 0{,}95$, 0.97 and 0.99. It is interesting to see that the Q_D curves appear to follow the data patterns closer than the Q_R curves.

Table 3 lists the estimated AMZN daily volume quantile values at a given opening price for $\tau = 0.97$ and 0.99. It demonstrates that when quantiles are at high τ, the Q_D gives greater variety of AMZN daily volumes predictions than the Q_R. The relationship of AMZN daily volumes and opening price is not necessarily linear.

Table 3. Predicted High Quantiles of AMZN Daily Volumes Using Q_R and Q_D (Population of the AMZN daily volumes greater than the threshold 10,854,389)

	$\tau = 0.97$		$\tau = 0.99$	
Opening Price (Canadian $)	Q_R	Q_D	Q_R	Q_D
$25	37,843,205	33,185,929	42,789,842	41,843,321
$100	33,097,066	37,258,031	42,513,740	44,863,380
$225	28,766,778	27,880,550	39,806,775	34,989,182
$400	24,830,980	25,752,650	34,909,299	27,805,877
$625	21,268,810	25,329,384	28,277,940	27,007,946
$900	18,059,907	19,885,667	20,585,579	22,018,153

In order to compare the regular QR estimator based on *(6)* and the direct nonparametric QR estimator in *(9),* we extend the idea of measuring goodness-of-fit by Koenker and Machado (1999) and suggest using a Relative $R(\tau)$ (Huang and Nguyen, 2017), $0 < \tau < 1$, Figure 5 and Table 4 show the values of the Relative $R(\tau)$ for given $\tau = 0.95, ..., 0.99$. We note that $R(\tau) > 0$ which means that Q_D is a better fit to the data than Q_R.

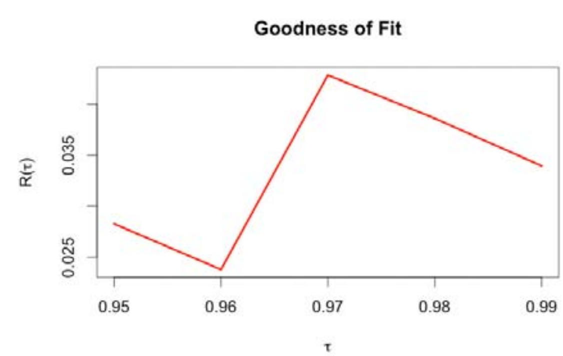


Figure 5. Relative $R(\tau)$ by using the direct quantile regression Q_D for Amazon volumes.

Table 4. Relative $R(\tau)$ Values for the Amazon Volumes.

	$\tau = 0.95$	$\tau = 0.96$	$\tau = 0.97$	$\tau = 0.98$	$\tau = 0.99$
Relative $R(\tau)$	0.0283	0.0238	0.0428	0.0386	0.0339

5. Conclusions and Suggestions

After the above studies, we can conclude:

1. This paper proposes a new direct nonparametric quantile regression method which is model free. It uses nonparametric density estimation and nonparametric regression techniques to estimate high conditional quantiles. The paper provides a computational five-step algorithm which overcomes the limitations of the estimation in the linear quantile regression model.

2. The Monte Carlo simulation works on the second kind of Gumbel's bivariate exponential distribution which has a non-linear conditional quantile function. The simulation results confirm that the proposed new method is more efficient relative to the regular quantile regression.

3. The proposed new direct nonparametric quantile regression can be used to predict extreme values of AMZN daily volumes for a given opening price. The proposed direct quantile regression estimator, Q_D, gives a variety of predictions which fits the data very well. The prediction of relationships are not simply just linear. We expect that the predictions from Q_D may be more reasonable than Q_R predictions. The new estimator may benefit investors understanding the stock market.

4. The proposed direct nonparametric quantile regression provides an alternative way for quantile regression. Further studies on the details of this method are suggested.

References

[1] Gumbel, E. J. (1960). Bivariate exponential distributions, *Journal of the American Statistical Association,* 55, 698-707.

[2] Huang, M. L. and Nguyen, C. (2017). High quantile regression for extreme events, *Journal of Statistical Distributions and Applications*, 4(4), 1-20.

[3] Huang, M. L., Xu, X. and Tashnev, D. (2015). A weighted linear quantile regression. *Journal of Statistical Computation and Simulation,* 85(13), 2596-2618.

[4] Investopedia (2010), https://www.investopedia.com

[5] Koenker, R. (2005). *Quantile regression.* Cambridge University Press: New York.

[6] Koenker, R. and Bassett, W.G. 1978. Regression Quantiles. *Econometrica*, 46. pp. 33-50.

[7] Koenker, R. and Machado, J.A.F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association,* 96(454), 1296-1311.

[8] Scott, D. W. (2015). *Multivariate Density Estimation, Theory, Practice and Visualization, second edition.* John Wiley & Sons: New York.

[9] Silverman, B.W. (1986). *Density estimation for statistics and data analysis.* Chapman & Hall: London, UK.

[10] Wang, H. J. and Li, D. (2013). "Estimation of extreme conditional quantile through power transformation", *Journal of the American Statistical Association,* 108(503), 1062-1074.

[11] Yahoo Canadian Finance (2018), https://ca.finance.yahoo.com