On Nonparametric Extreme Quantile Regression

Andrew Kenig[a], Mei Ling Huang[a,*] and Percy Brill[b,†]

[a]Department of Mathematics & Statistics, Brock University, Canada
[b]Department of Mathematics & Statistics, University of Windsor, Canada

August 31, 2020

Abstract

Quantile regression (QR) estimates conditional quantiles with wide applications in the real world. Estimating extreme conditional quantiles is an important and difficult problem. The regular quantile regression method often sets a linear model with estimating the coefficients to obtain the estimated conditional quantile. This approach may be restricted by the model setting, there are also computational difficulties. To overcome this problem, this paper proposes a two-stage nonparametric quantile regression method with a 6-step algorithm by using extrapolation. Monte Carlo simulations show good efficiency for the proposed nonparametric QR extrapolation estimator relative to the regular linear QR extrapolation estimator. The paper also investigates an Alberta Wildfire example by using the proposed method. Comparisons of the proposed method and existing methods are given.

Keywords: *Conditional quantile, extreme value distribution, Fréchet distribution, generalized Pareto distribution, Hill estimator, nonparametric regression.*

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

1. Introduction

Extreme events occur in many fields such as financial markets, weather, industrial engineering, actuarial science and other stochastic models. When statisticians are interested in estimating high quantiles of heavy-tailed distributions of extreme events, they often face theoretical difficulties in doing so. It is important to estimate extreme conditional quantiles of a random variable y given a variable vector $\mathbf{x} = (x_1, x_2, ..., x_d)^T$, and we let $\mathbf{x}_p = (1, x_1, x_2, ..., x_d)^T \in R^p$, $p = d+1$. First, we will review the mean regression and linear quantile regression models:

The mean linear regression model assumes

$$\mu_{y|\mathbf{x}} = E(y|x_1, x_2, ..., x_d) = \mathbf{x}_p^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_d x_d. \quad (1)$$

*Corresponding author. E-mail: mhuang@brocku.ca.

†The research is supporeted by *the Natural Science and Engineering Research Council of Canada (NSERC) grants RGPIN-2019-04206 and RGPIN-2014–05697..*

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

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

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

The mean linear regression provides the mean relationship between a response variable and explanatory variables. We are interested in estimating the conditional quantiles of y given $\mathbf{x}$.

$$Q_Y(\tau|\mathbf{x}) = \inf(t : F_Y(t|\mathbf{x}) \geq \tau\} = F_Y^{-1}(\tau|\mathbf{x}),\ 0 < \tau < 1. \tag{3}$$

Koenker and Bassett (1978) proposed a linear quantile regression model for estimating true conditional quantiles in *(3)*. It is defined as

$$Q_L(\tau|\mathbf{x}) = \mathbf{x}_p^T \boldsymbol{\beta}(\tau) = \beta_0(\tau) + \beta_1(\tau)x_1 + \cdots + \beta_d(\tau)x_d,\ 0 < \tau < 1, \tag{4}$$

where $\boldsymbol{\beta}(\tau) = (\beta_0(\tau), \beta_1(\tau), \beta_2(\tau), ..., \beta_d(\tau))^T$.

In model *(4)*, 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 $\{(y_i, \mathbf{x}_i),\ i = 1, ..., n\}$ by using an L_1- loss function to obtain estimator $\widehat{\boldsymbol{\beta}}(\tau)$

$$\widehat{\boldsymbol{\beta}}(\tau) = \arg\min_{\boldsymbol{\beta}(\tau) \in R^p} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_{pi}^T \boldsymbol{\beta}(\tau)),\ 0 < \tau < 1, \tag{5}$$

where ρ_τ is a loss function, namely

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

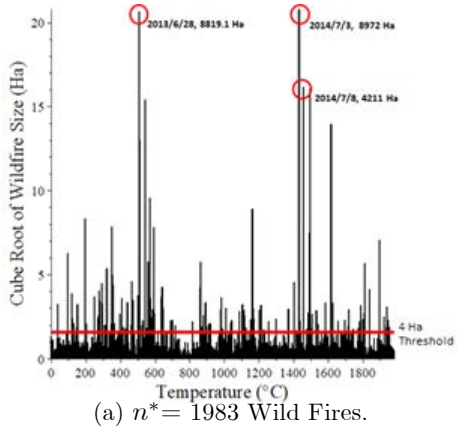
In recent years, many studies have focussed on improvements of estimator *(5)*. In this paper, we concentrate on estimating extremely high or low conditional quantiles, which may be restricted under the model setting. We are motivated by the following Canada Alberta wildfires example.

Example. Canada Alberta Wildfires (2013-2014)

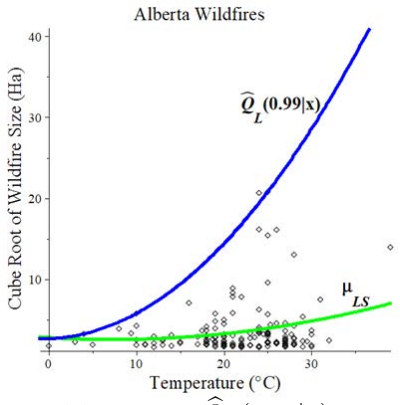
Wildfires can be hazardous to the safety of the general public and animals. They can also cause damage to houses and other infrastructure. Wildfires can occur from natural causes such as dry climate, volcanoes or lightning strikes; or from human activity such as arson, light cigarettes or electrical sparks. Depending on the weather conditions such as temperature, precipitation and wind speed, wildfires can quickly burn out of control which makes it very difficult for firefighters to prevent damage to homes and resources. It can be very expensive for the government to supply the resources needed to put out the wildfire as often thousands of gallons of water, multiple helicopters and many firefighters are needed.

The province of Alberta, Canada is highly affected by wildfires. The Alberta Ministry of Agriculture and Forestry (www.wildfire.alberta.ca/resources/historical-data) for the province of

Alberta recorded 1983 wildfires for the time period of January 1st 2013 to December 31st 2014. The total area covered by the wildfire, measured in hectares (Ha), as well as the temperature ^oC were recorded. Over the last decade, yearly expenditures for fire protection ranged between $500 million and $1 billion according to the National Resources of Canada (www.nrcan.gc.ca). They also predict this figure to double by the year 2040. Therefore, preventing wildfires is of great importance to the safety of the general public/wildlife and to avoid damage to infrastructure. A threshold of 4 Ha was implemented and $n = 149$ larger fires remained. Figure 1(a) shows a column graph of the cubed root of area burned by wildfires before implementing the threshold.



(a) n^*= 1983 Wild Fires.



(b) μ_{LS} and $\widehat{Q}_L(0.99|x)$.

Figure 1. (a) The Alberta Wildfires during January 1st, 2013 - December 31st, 2014 ordered by date, $n^* = 1983$, and a threshold cubed root of 4 Ha (red). (b) Scatter plot (black dots) of Alberta wildfire area covered (cubed root of Ha) y related to temperature ^oC, x; the least square mean regression curve μ_{LS} (red) and the 99% quantile regression curve $\widehat{Q}_L(0.99|x)$ (green), using $n = 149$ large fires over the 4 Ha threshold.

Based on the $n = 149$ wildfires greater than the 4 Ha threshold, we perform a cubed root transformation on the wildfire size and then implement a polynomial mean regression model using *(2)*, the least squares estimate is

$$\mu_{LS} = \widehat{\mu}_{y|\mathbf{x}} = 2.8023 - 0.0646x + 0.0044x^2,$$

where y represents the response variable (cubed root of area burned in Ha) and x represents the temperature ^oC.

The quantile regression estimate for $\tau = 0.99$ in *(5)* is

$$\widehat{Q}_L(0.99|x) = 2.6687 + 0.0319x + 0.0277x^2.$$

Figure 1(b) shows the mean regression μ_{LS} and the $\tau = 0.99$ quantile regression $\widehat{Q}_L$ curves. Note that both μ_{LS} and $\widehat{Q}_L$ do not catch the extreme wildfires data well. In Section 5, we will study this example by implementing the new proposed quantile regression method.

In recognition of these complications for estimating extreme conditional quantiles, the non-parametric quantile regression has been used (Yu, et al., 2003; Huang and Nguyen, 2018).

Furthermore, extrapolations have been used: one is based on linear models *(4)* and *(5)* (Chernozhukov and Du, 2008, Wang and Li, 2013); another extrapolation is based on Kernel estimation (Daouia, et. at, 2011; Gardes and Girard, 2011). The linear extrapolation is restricted by the model setting, and quantile curves are estimated one by one. The issue that arises is that they may intersect. The extrapolation based on the nonparametric kernel method avoids the quantile curve crossing problem, but it requires strong heavy tailed assumptions. This paper proposes a 6-step algorithm to estimate extreme conditional quantile curves. Our contributions are:

a) Improve the existing kernel estimation method;

b) Perform goodness of fit tests for the response y having a heavy tailed distribution before the extrapolation.

In Section 2, we propose an extrapolation nonparametric quantile regression estimator with a 6-step algorithm in two stages. In Section 3, the results of Monte Carlo simulations generated from Fréchet (Fréchet, 1927) show that the proposed nonparametric extrapolation method produces high efficiencies relative to existing linear extrapolation methods. A relative measure of comparing goodness of fit for these two models is given in Section 4. In Section 5, we study Alberta wildfires example by using two extrapolation methods: linear extrapolation quantile regression and proposed nonparametric extrapolation quantile regression. The simulations and the example illustrate that the proposed nonparametric extrapolation quantile regression model fits the data set better than the linear extrapolation quantile regression method.

2. Proposed Extrapolation for Extreme Conditional Quantiles

We ignore the idea of the linear model *(4)* to obtain a nonparametric kernel estimator for the true conditional quantile in *(3):*

$$\widehat{Q}_Y(\tau|\mathbf{x}) = \inf(t : \widehat{F}_Y(t|\mathbf{x}) \geq \tau\} = \widehat{F}_Y^{-1}(\tau|\mathbf{x}),\ 0 < \tau < 1,$$

by using local conditional quantile estimator $\xi_i(\tau|\mathbf{x}) = Q_Y(\tau|\mathbf{x}_i)$ based the ith point of given random sample, $\{(Y_i, \mathbf{X}_i), i = 1, ..., n\}$, for $\mathbf{X}_i = (x_{i1}, x_{i2}, ..., x_{id})^T$.

2.1. Extreme Value Distribution

Let $X_1, X_2, ..., X_n$ be independent and identically distributed (i.i.d.) continuous random variables with common cumulative distribution function (c.d.f.) $F(x)$. It is important to study the limiting behavior of the sample maxima and minima, defined as $\max(X_1, X_2, ..., X_n)$ and $\min(X_1, X_2, ..., X_n)$, respectively. The main interest of extreme value theory (EVT) is in finding possible limiting distributions of the sample maxima of i.i.d. random variables. Any nondegenerate distribution that can be derived as such a limit is called an *extreme value distribution* (Haan and Ferreira, 2006).

Definition 1. *(Fisher and Tippett, 1928, Gnedenko, 1943) The c.d.f. of any extreme value distribution is of the form $G_\gamma(ax + b)$ for some constants $a > 0$, $b \in R$, where*

$$G_\gamma(x) = \begin{cases} 1 - \exp(-(1+\gamma x)^{-1/\gamma}), & 1 + \gamma x > 0 \text{ and } \gamma \neq 0; \\ 1 - \exp(-e^{-x}), & \gamma = 0, \end{cases} \tag{6}$$

where the parameter γ is called the extreme value index (EVI).

Note that when $\gamma > 0$, the corresponding densities for both $G_\gamma(x)$ and $G_\gamma(ax+b)$ have heavier tails than the exponential distribution, which are referred to as *heavy tailed distributions*. In many applications, it is important to include observations that take extremely high or low values in the statistical analysis.

Definition 2. *(Hill Estimator, Hill, 1975) Consider a random sample $X_1, X_2, ..., X_n$ with sample size n from the distribution in (6), and let $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$ denote its order statistics. Hill (1975) introduced a maximum likelihood estimators (MLE) for the index parameter γ as*

$$\widehat{\gamma}_{Hill} = \frac{1}{k}\sum_{i=1}^{k}(\log\{X_{n-i,n}\} - \log\{X_{n-k,n}\}) = \frac{1}{k}\sum_{i=0}^{k-1}\log\{X_{n-i,n}\} - \ln X_{n-k,n}, \qquad (7)$$

where $X_{i,n}$ represents the ith order statistics, $i = 1, ..., n$. and note that $\widehat{\gamma}_{Hilll}$ uses the k largest order statistics.

A limit conditional extreme value distribution for exceeding a threshold has a *generalized Pareto distribution* (GPD).

Definition 3. *(Pickands, 1975) The c.d.f. $H_\gamma(x)$ and its corresponding probability density function (p.d.f.) $f(x)$ of the two-parameter GPD(γ, σ) with shape parameter $\gamma > 0$ and scale parameter σ of a random variable X are given by*

$$H_\gamma(x) = 1 - \left(1 + \gamma\frac{x}{\sigma}\right)^{1/\gamma}, \quad \gamma, \sigma > 0, \quad x > 0. \qquad (8)$$

2.2. Propose Extrapolation based on Kernel Method

We build the following steps to construct a direct nonparametric quantile regression estimator:

Stage 1: Kernel Quantile Regression (5 steps)

Step 1: Kernel Estimation for conditional c.d.f.: Estimate the conditional c.d.f. $F(y|\mathbf{x})$ of y for given $\mathbf{x} = (x_1,x_2,...,x_d)$ using kernel estimation method (Scott, 2015)

$$\widehat{F}_Y(y|\mathbf{x}) = \frac{\sum_{i=1}^{n} I(Y_i \leq y\}K_h(\mathbf{x} - \mathbf{X}_i)}{\sum_{i=1}^{n} K_h(\mathbf{x} - \mathbf{X}_i)},$$

where $I(Y_i \leq y\}$ is an indicator function, and

$$K_h(\mathbf{x} - \mathbf{X}_i) = \frac{1}{h^d}K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right),$$

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

Step 2. Normalization: Normalize the estimated conditional c.d.f. $\widehat{F}_Y(y|\mathbf{x}) \in [0,1]$ to a true c.d.f. function.
Step 3: Localization: Estimate the local conditional quantile function $\xi(\tau|\mathbf{x})$ of y given $\mathbf{x}$ by inverting an estimated conditional c.d.f. $\widehat{F}_Y(y|\mathbf{x})$.

$$\widehat{\xi}(\tau|\mathbf{x}) = \widehat{Q}_Y(\tau|\mathbf{x}) = \inf\{y : \widehat{F}_Y(y|\mathbf{x}) \geq \tau\} = \widehat{F}_Y^{-1}(\tau|\mathbf{x}). \tag{9}$$

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 an estimated conditional c.d.f. $\widehat{F}_Y(y|\mathbf{x}_i)$ at the ith data point:

$$\widehat{\xi}_i(\tau|\mathbf{x}_i) = \widehat{Q}_Y(\tau|\mathbf{x}_i) = \inf\{y : \widehat{F}_Y(y|\mathbf{x}_i) \geq \tau\} = \widehat{F}_Y^{-1}(\tau|\mathbf{x}_i), \quad i = 1, 2, ..., n.$$

Step 4: Nonparametric Regression based on local estimates in Step 3. We obtain a nonparametric quantile regression estimator for the τth conditional quantile curve of $\mathbf{x}$ by using Nadaraya-Watson (NW) nonparametric regression estimator on $\left(\mathbf{x}_i, \widehat{\xi}_i(\tau|\mathbf{x}_i)\right)$, $i = 1, 2, ..., n$,

$$\widehat{Q}_N(\tau|\mathbf{x}) = \widehat{\xi}(\tau|\mathbf{x}) = \frac{\sum_{i=1}^{n} K\left\{\frac{\mathbf{x}-\mathbf{X}_i}{\mathbf{h}}\right\} \widehat{\xi}_i(\tau|\mathbf{x}_i)}{\sum_{j=1}^{n} K\left\{\frac{\mathbf{x}-\mathbf{X}_j}{\mathbf{h}}\right\}} = \sum_{i=1}^{n} W_h(\mathbf{x}, \mathbf{X}_i)\widehat{\xi}_i(\tau|\mathbf{x}_i), \quad 0 < \tau < 1, \tag{10}$$

where $W_i(\mathbf{x})$ is called an equivalent kernel,

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

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: Goodness of fit test: Check that the response variable y is heavy tailed distributed.

Stage 2: Extrapolation (1-Step)

Step 6: Extrapolation for an extreme conditional quantile estimator:

$$\widehat{Q}_{NE}(\tau_n|\mathbf{x}) = \widehat{Q}_N(\alpha_n|\mathbf{x}) \left(\frac{1-\alpha_n}{1-\tau_n}\right)^{\widehat{\gamma}_N^H(\mathbf{x})}, \tag{11}$$

where α_n is an intermediate quantile level, and $\widehat{Q}_N(\alpha_n|\mathbf{x})$ are the intermediate quantiles defined in *(10)*, for extreme order of quantiles, as $\tau_n \to 1$ at a rate faster than $1/n$.

$$n(1-\tau_n) > (\log(n))^p, \quad 0 \leq \tau_n \leq 1, \quad \left(\frac{1-\tau_n}{1-\alpha_n}\right) \to 0, \quad as \quad n \to \infty.$$

which means $(1-\tau_n) \to 0$ faster than $(1-\alpha_n) \to 0$ as $n \to \infty$, and $\widehat{\gamma}_N^H(\mathbf{x})$ is a kernel version of Hill estimator in *(7)* for conditional EVI $\gamma(\mathbf{x})$.

$$\widehat{\gamma}_N^H(\mathbf{x}) = \frac{\sum_{j=1}^{J} \left[\log\left\{\widehat{Q}_N(1-w_j(1-\alpha_n^*)|\mathbf{x})\right\} - \log\left\{\widehat{Q}_N(\alpha_n^*)|\mathbf{x})\right\}\right]}{\sum_{j=1}^{J} \log\left(1/w_j\right)}, \quad 0 < w_j < 1, \quad j = 1, ..., J, \tag{12}$$

where $w_1 > w_2 > ... > w_J > 0$ is a decreasing sequence of weights, and J is a positive integer. $\widehat{Q}_N(\alpha_n^*|\mathbf{x})$ in *(10)* is the intermediate quantile curve starting at α_n^* level, then the kernel estimated quantile curves $\widehat{Q}_N(1-w_j(1-\alpha_n^*)|\mathbf{x}), \quad j = 1, ..., J$, are order statistics, which are non-crossing quantile curves.

Similarly, $\widehat{\gamma}_L^H(\mathbf{x})$ is a linear version of Hill estimator in *(7)* for conditional EVI $\gamma(\mathbf{x})$.

$$\widehat{\gamma}_L^H(\mathbf{x}) = \frac{\sum_{j=1}^{J} \left[\log\left\{\widehat{Q}_L(1-w_j(1-\alpha_n^*)|\mathbf{x})\right\} - \log\left\{\widehat{Q}_L(\alpha_n^*)|\mathbf{x})\right\}\right]}{\sum_{j=1}^{J} \log\left(1/w_j\right)}, \quad 0 < w_j < 1, \quad j = 1, ..., J, \tag{13}$$

where $\widehat{Q}_L(\alpha_n|\mathbf{x})$ is the linear estimated intermediate quantiles given by *(5)*.

Remark: To use extrapolation we must have that the conditional c.d.f. of Y given a variable vector $\mathbf{x}$, $F_Y(\bullet|\mathbf{x})$, belongs to the Fréchet maximum domain of attraction with conditional extreme value index (EVI) $\gamma(\mathbf{x})$ for any ($\mathbf{x}$,y) $\in R^p \times R$. The τth conditional quantile $Q_Y(\tau|\mathbf{x})$ is defined by *(3)*. The kernel estimator is not feasible as it can not extrapolate beyond the maximum observation in the ball centered at $\mathbf{x}$ with radius h. (Daonia et al., 2011). This extrapolation allows the estimation of extreme conditional quantile with $\tau_n \to 1$ arbitrarily fast.

In this paper, we compare the $\widehat{Q}_{NE}(\tau_n|\mathbf{x})$ in *(11)* with Chernozhukov and Du (2008) proposed method with an assumption that allows the quantile slope coefficient $\boldsymbol{\beta}(\tau)$ in model *(4)* to vary cross τ. Assume model *(4)* after being transformed by some auxiliary regression line, the response variable Y has regularly varying tails with EVI $\gamma > 0$. More specifically, suppose that there exists an auxiliary slope $\boldsymbol{\beta}(\tau)$ such that the following tail-equivalence relationship hold as $\tau_n \to 1$,

$$\widehat{Q}_{LE}(\tau_n|\mathbf{x}) = \widehat{Q}_L(\alpha_n|\mathbf{x}) \left(\frac{1-\alpha_n}{1-\tau_n}\right)^{\widehat{\gamma}_L^H(\mathbf{x})}. \tag{14}$$

where α_n is an intermediate quantile level, and $\widehat{Q}_L(\alpha_n|\mathbf{x})$ is the linear estimated intermediate quantiles given by *(5)*. $\widehat{\gamma}_L^H(\mathbf{x})$ is defined in *(13)*.

3. Monte Carlo Simulation

In this section, we will run Monte Carlo simulations to investigate the efficiency of the proposed nonparametric extrapolation conditional quantile estimator $\widehat{Q}_{NE}(\tau|x)$ in *(11)*, and extrapolation estimator based on linear quantile regression estimator $\widehat{Q}_{LE}(\tau|x)$ in *(14)*. We generate $m = 100$

random samples with size $n = 300$ each from one-dimensional random variables X $(p = 1)$ uniformly distributed on $E = [0, 1]$. Suppose Y given $X = x$ is Fréchet distributed (Haan and Ferreira, 2006), the conditional c.d.f. is

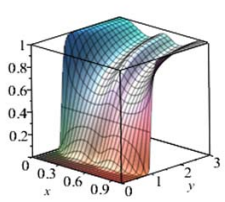
$$F(y|x) = e^{(-y^{-1/\gamma(x)})}, \quad 0 \le x \le 1, \quad \gamma(x) > 0, \tag{15}$$

with the conditional tail index

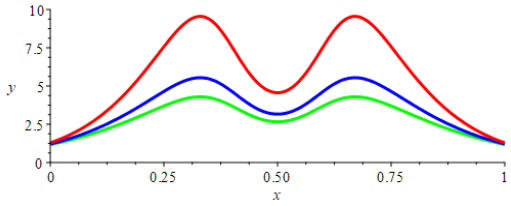
$$\gamma(x) = \frac{1}{2}\left(\frac{1}{10} + \sin(\pi x)\right)\left(\frac{11}{10} - \frac{1}{2}e^{(-64(x-1/2)^2}\right), \quad 0 \le x \le 1.$$

The true conditional quantile of *(15)* is

$$Q_Y(\tau|x) = (-\log(\tau))^{-\gamma(x)}, \quad 0 \le x \le 1, \quad 0 < \tau < 1. \tag{16}$$



(a) The joint distribution.



(b) The True Conditional Quantiles.

Figure 2. (a) The joint c.d.f. of Fréchet distribution of y and uniform distribution of x. (b) The true conditional quantiles at $\tau = 0.99$ (red), $\tau = 0.97$ (blue), $\tau = 0.95$ (green).

We use two conditional quantile extrapolation estimators $\widehat{Q}_{NE}(\tau|x)$ in *(11)* and $\widehat{Q}_{LE}(\tau|x)$ in *(14)* to estimate the true conditional quantile $Q_Y(\tau|x)$ in *(16)*:

For each method, we generate size $n = 300$, $m = 100$ samples. $\widehat{Q}_{NE,i}(\tau|x)$ and $\widehat{Q}_{LE,i}(\tau|x)$, $i = 1, 2, ..., m$, are estimated in the ith sample.

The simulation mean squared errors (SMSEs) of the estimators *(11)* and *(14)* are respectively:

$$SMSE\left(\widehat{Q}_{NE}(\tau)\right) = \frac{1}{m}\sum_{i=1}^{m}\int_0^1 \left(\widehat{Q}_{NE,i}(\tau|x) - Q_Y(\tau|x)\right)^2 dx; \tag{17}$$

$$SMSE\left(\widehat{Q}_{LE}(\tau)\right) = \frac{1}{m}\sum_{i=1}^{m}\int_0^1 \left(\widehat{Q}_{LE,i}(\tau|x) - Q_Y(\tau|x)\right)^2 dx, \tag{18}$$

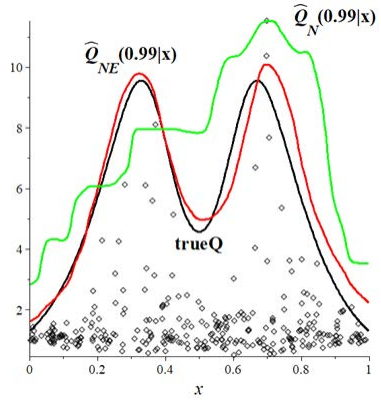
where the true τth conditional quantile $Q_Y(\tau|x)$ is defined in *(16)*. The simulation efficiencies (SEFFs) are given by

$$SFEF\left(\widehat{Q}_{NE}(\tau)\right) = \frac{SMSE\left(\widehat{Q}_{LE}(\tau)\right)}{SMSE\left(\widehat{Q}_{NE}(\tau)\right)},$$

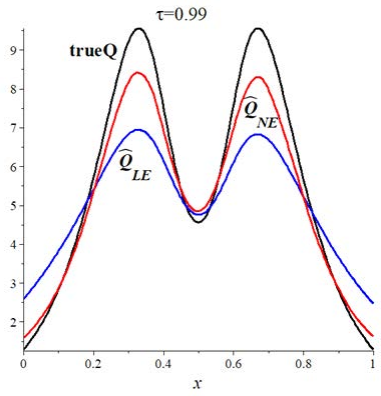
where $SMSE\left(\widehat{Q}_{NE}(\tau)\right)$ and $SMSE\left(\widehat{Q}_{LE}(\tau)\right)$ are defined in *(17)* and *(18)*.

Table 1. Simulation mean squared errors (SMSEs) and efficiencies (SEFFs) of $\widehat{Q}_{NE}(\tau|x)$ and $\widehat{Q}_{LE}(\tau|x)$ estimating $Q_Y(\tau|x), m = 100, n = 300$,

τ	0.95	0.96	0.97	0.98	0.99
$SMSE\left(\widehat{Q}_{LE}(\tau)\right)$	0.6393	0.7555	0.9424	1.2995	2.3027
$SMSE\left(\widehat{Q}_{NE}(\tau)\right)$	0.3428	0.4188	0.5420	0.7797	1.4538
$SEFF\left(\widehat{Q}_{NE}(\tau)\right)$	**1.8646**	**1.8042**	**1.7386**	**1.6668**	**1.5839**



(a) Stage 1 $\widehat{Q}_N$ and Stage 2 $\widehat{Q}_{NE}$.



(b) Stage 2: $\widehat{Q}_{NE}$ and $\widehat{Q}_{LE}$.

Figure 3. The true quantile curves (black), and for a simulation $n = 300$, data (black dots), (a) Stage 1 $\widehat{Q}_N(\tau|x)$ (green) and Stage 2 $\widehat{Q}_{NE}(\tau|x)$ (red) for one simulation; (b) Stage 2 Average $\widehat{Q}_{LE}(\tau|x)$ (blue) and $\widehat{Q}_{NE}(\tau|x)$ (red).

From the simulation results above, we conclude that

(1) Table 1 shows that all of the $SEFF(\widehat{Q}_{NE}(\tau)) > 1$ when $\tau = 0.95,..., 0.99$. Thus using the proposed nonparametric extrapolation estimator $\widehat{Q}_{NE}(\tau|x)$ in *(11)* is more efficient relative to the extrapolation based on linear extrapolation estimator $\widehat{Q}_{LE}(\tau|x)$ in *(14)*.

(2) Figure 3 shows that Stage 2 $\widehat{Q}_{NE}(\tau|x)$ estimated curves are much closer to the true conditional quantile curve than Stage 1 $\widehat{Q}_N(\tau|x)$ curves. Thus the Stage 2 extrapolations contribute more accurate estimates. Also in Stage 2, $\widehat{Q}_{NE}(\tau|x)$ is much accurate than $\widehat{Q}_{LE}(\tau|x)$.

4. Comparison of Goodness-of Fit on Quantile Regression Models

To compare the nonparametric extrapolation estimator $\widehat{Q}_{NE}(\tau|x)$ in *(11)* and the linear extrapolation estimator $\widehat{Q}_{LE}(\tau|x)$ in *(14)*, we extend the idea of measuring goodness-of-fit by Koenker and Machado (1999). We use a Relative $R_{NE}(\tau)$ of $\widehat{Q}_{NE}(\tau|x)$ to $\widehat{Q}_{LE}(\tau|x)$, $0 < \tau < 1$, which is defined as

$$Relative\ R_{NE}(\tau) = 1 - \frac{V_{NE}(\tau)}{V_{LE}(\tau)}, \quad -1 \le R_{NE}(\tau) \le 1, \quad \text{where} \qquad (19)$$

$$V_{NE}(\tau) = \sum_{y_i \geq \widehat{Q}_{NE}(\tau|\mathbf{x}_i)} \frac{\tau}{n} \left| y_i - \widehat{Q}_{NE}(\tau|\mathbf{x}_i) \right| + \sum_{y_i < \widehat{Q}_{NE}(\tau|\mathbf{x}_i)} \frac{(1-\tau)}{n} \left| y_i - \widehat{Q}_{NE}(\tau|\mathbf{x}_i) \right|,$$

where $\widehat{Q}_{NE}(\tau|\mathbf{x}_i)$ is obtained by *(11)*, and

$$V_{RE}(\tau) = \sum_{y_i \geq \widehat{Q}_{LE}(\tau|\mathbf{x}_i)} \frac{\tau}{n} \left| y_i - \widehat{Q}_{LE}(\tau|\mathbf{x}_i) \right| + \sum_{y_i < \widehat{Q}_{LE}(\tau|\mathbf{x}_i)} \frac{(1-\tau)}{n} \left| y_i - \widehat{Q}_{LE}(\tau|\mathbf{x}_i) \right|,$$

where $\widehat{Q}_{LE}(\tau|\mathbf{x}_i)$ is given by *(14)*.

5. Alberta Wildfires Example

Recall the Alberta Wildfires example in Section 1, we use the polynomial mean regression model in *(1)* and the polynomial linear quantile regression model in *(4)*,

$$\begin{aligned} E(y|x) &= \beta_0 + \beta_1 x + \beta_2 x^2, \\ Q_Y(\tau|x) &= \beta_0(\tau) + \beta_1(\tau)x + \beta_2(\tau)x^2, \quad 0 < \tau < 1. \end{aligned}$$

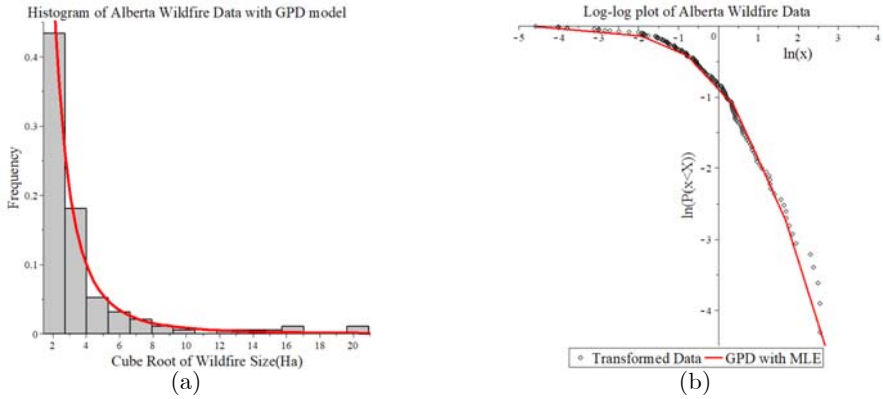
where y represents the cubed root of total fire area covered in Ha, and x represents the temperature in ^oC. The mean regression only estimates the average area covered by wildfires.

In this Section, we will compare the proposed nonparametric extrapolation method with the linear extrapolation methods, we use the intermediate level $\alpha_n^* = 0.92$ and $J = 9$.

1. The linear extrapolation estimator $\widehat{Q}_{LE}(\tau|x)$ in *(14)* uses the intermediate quantile $\widehat{Q}_L(\tau|x)$ based on *(4)* obtained by linear programming and $\widehat{\gamma}_L^H(\mathbf{x})$ is in *(13)* Section 2.

2. The nonparametric extrapolation estimator $\widehat{Q}_{NE}(\tau|x)$ in *(11)* uses the intermediate quantile $\widehat{Q}_N(\tau|x)$ obtained by the 6-Step algorithm *(10)* and $\widehat{\gamma}_N^H(\mathbf{x})$ is in *(12)* Section 2.

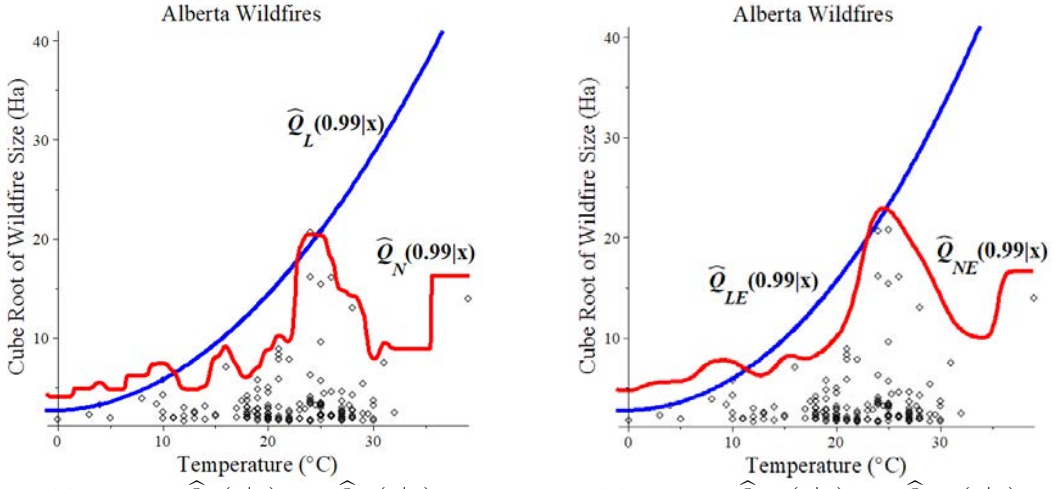
Figure 4(a), (b) show the histogram and log-log plot of the Alberta Wildfire data with GPD model in *(8)* with MLEs of the parameters. The theoretical GDP curve follows the shape of the Alberta Wildfire data very well.



(a) (b)

Figure 4. (a) Histogram of the Alberta Wildfire data, $n = 149$, greater than a threshold of 4 Ha with the GPD (red). (b) a Log-log plot of Alberta Wildfire cover area. The black dots are the data and the solid line is the GPD curve (red).

Figure 5(b) shows that the proposed Stage 2 estimator $\widehat{Q}_{NE}(\tau|x)$ predicts that for high temperatures, such as 25 ^oC, it is likely to have larger wildfire size but for very high temperatures it predicts a lower wildfire size, but the linear estimator $\widehat{Q}_{LE}(\tau|x)$ curve is concave up. Also, Stage 2 $\widehat{Q}_{NE}(\tau|x)$ fits extreme data smoother than Stage 1 $\widehat{Q}_{N}(\tau|x)$ in Figure 5(a). Predicting extreme wildfires is related to weather forecasts. The proposed extrapolation estimator $\widehat{Q}_{NE}(\tau|x)$ may be useful for predicting extreme wildfires in Alberta.



(a) Stage 1: $\widehat{Q}_L(\tau|x)$ and $\widehat{Q}_N(\tau|x)$ (b) Stage 2: $\widehat{Q}_{LE}(\tau|x)$ and $\widehat{Q}_{NE}(\tau|x)$.

Figure 5. For Alberta Wildfire data, $n = 149$, (a) $\widehat{Q}_L(\tau|x)$ regression curve (blue) and $\widehat{Q}_N(\tau|x)$ regression curve (red) at $\tau = 0.99$. (b) Stage 2: extrapolation estimate conditional quantile curves $\widehat{Q}_{LE}(\tau|x)$ (blue) and $\widehat{Q}_{NE}(\tau|x)$ (red) at $\tau = 0.99$.

Table 2 show the values of the Relative $R_{NE}(\tau)$ for given $\tau = 0.95$,..., 0.99 when comparing the linear extrapolation estimator $\widehat{Q}_{LE}(\tau|x)$ with the proposed nonparametric extrapolation estimator $\widehat{Q}_{NE}(\tau|x)$. We note that $R_{NE}(\tau) > 0$, which means that $V_{NE}(\tau) < V_{LE}(\tau)$, and $\widehat{Q}_{NE}(\tau|x)$ is better fit to the data than $\widehat{Q}_{LE}(\tau|x)$.

Table 2. Relative $R_{NE}(\tau)$ of $\widehat{Q}_{NE}(\tau|x)$ relative to $\widehat{Q}_{LE}(\tau|x)$ for the Alberta Wildfires example.

	$\tau = 0.95$	$\tau = 0.96$	$\tau = 0.97$	$\tau = 0.98$	$\tau = 0.99$
Relative $R_{NE}(\tau)$	0.1114	0.1946	0.2679	0.3068	0.3514

6. Conclusions

After the studies above, we can conclude:

1. Traditional mean regression estimates the conditional mean by using the L_2- loss function. The linear quantile regression uses a L_1- loss function. But both models have limitations for estimating extreme conditional quantiles for the analysis of extreme events. In a heavy tailed

population, the proposed two-stage nonparametric extrapolation quantile regression method has advantages to predicting extreme conditional quantiles compared to other existing methods.

2. The Monte Carlo computational simulation results show that the proposed nonparametric extrapolation quantile regression is more efficient relative to the linear extrapolation method using linear quantile regression.

3. The proposed nonparametric extrapolation quantile regression can be used to predict extreme values of the Canada Alberta Wildfire example.

References

[1] Alberta Ministry of Agriculture Forestry, www.wildfire.alberta.ca/resources/historical-data/historical-wildfire-database-aspx, Willdfires, 2013-2014, accessed May 2019.

[2] Chernozhukov, V. and Du, S. (2008). Extreme quantiles and value-at risk, *The new Palgrave Dictionary of Economics, eds.,* Durlauf, S. N. and Blume, L. E., Basingstoke: Plagrave Macmillian.

[3] Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves, *Test*, 20, 311-333.

[4] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction.* Springer, New York.

[5] Fréchet, M. (1927). Sur loi de probabilté de l'écart maximun, *Ann, Soc. Math. Polon*, 6, 93-116.

[6] Gardes, L. and Girard, S. (2011). Functional Kernel Estimators of Conditional Extreme Quantiles, in *Recent Advances in Functional Data Analysis and Related Topics Contributions to Statistics,* Springer, 135-140.

[7] Hill, B. M. (1975). A simple general approach to inference about the yail of a distribution, *The Annals of Statistics, Vol.3, No.5, 1163-1174.*

[8] Huang, M. L. and Nguyen, C. (2018). A nonparametric approach for quantile regression, *Journal of Statistical Distributions and Applications,* Springer, Vol. 5(1), 1-14.

[9] Koenker, R. (2005). *Quantile Regression.* Cambridge University Press, New York.

[10] National Resouces of Canada (2017). http://www.nrcan.gc.ca.

[11] Picklands, J. (1975). Statistical inference using extreme order statistics. *Ann. Stat.* 3, 119-131.

[12] Scott, D. W. (2015). *Multivariate Density Estimation, Theory, Practice and Visualization.* 2nd edition, John Wiley & Sons, New York.

[13] 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.

[14] Yu, K., Lu, Z. and Stander, J. (2003). Quantile regression: applications and current research areas. *Statistician,* 52(3), 331-350.