15

Estimation of Extreme Conditional Quantiles

Huixia Judy Wang

George Washington University, United States

Deyuan Li

Fudan University, China

CONTENTS

Abstract 319
15.1 Introduction 320
15.2 Estimation of Extreme Unconditional Quantiles 320
15.3 Estimation of Extreme Conditional Quantiles 322
15.3.1 Parametric Methods 322
15.3.2 Semiparametric Methods 323
15.3.3 Nonparametric Methods 325
15.3.3.1 Likelihood-Based Methods 325
15.3.3.2 Two-Step Local Estimation 326
15.3.3.3 Kernel Estimation 327
15.3.4 Quantile Regression Methods 328
15.3.4.1 Estimation Based on the Common-Slope Assumption 329
15.3.4.2 Estimation without the Common-Slope Assumption 329
15.3.4.3 Three-Stage Estimation 330
15.4 Numerical Comparison 332
15.5 Final Remarks 334
Acknowledgements 334
References 334

Abstract

The estimation of extreme quantiles of the response distribution is of great interest in many areas. Extreme value theory provides a useful tool for estimating extreme quantiles. However, current extreme value literature focuses primarily on the ex-

treme quantiles of a univariate variable. In this chapter, we provide a survey of available methods, including parametric, nonparametric, semiparametric and quantile-regression-based approaches, for estimating the extreme conditional quantiles of the quantity of interest when some covariates are recorded simulataneously. A simulation study is carried out to assess the performance of various methods.

15.1 Introduction

Estimation of tail quantiles is of great interest in many studies of rare events that happen infrequently but have heavy consequences. Extreme value theory provides a useful tool for modeling rare events and estimating extreme quantiles. The current extreme value literature focuses primarily on the tail quantiles of a univariate variable. However, in many applications, the conditional extreme quantiles of the response variable Y given some covariates $\mathbf{X}$ are of interest, for instance, high quantiles of tropical cyclone intensity given time or certain climate variables (Jagger and Elsner, 2008), localized high precipitation conditional on global climate model projections (Friederichs, 2010), low conditional quantiles of a portfolio's future return given the past or assumptions on future interest rate changes (Engle and Manganelli, 2004), low quantiles of birth weight given maternal behavior (Abrevaya, 2001), and so on. In this chapter, we provide a survey of methods for estimating extreme conditional quantiles. Without loss of generality, we focus on the estimation of conditional high quantiles, because a low quantile of Y can be viewed as a high quantile of $-Y$.

Throughout, let Y denote the univariate response of interest, and $\mathbf{X}$ be the p-dimensional covariate vector. In addition, let $\xi(\mathbf{x})$ denote the conditional extreme value index of Y given $\mathbf{X} = \mathbf{x}$, which determines the rate of tail decay of the conditional distribution of Y. Suppose that we observe a random sample $\{(y_i, \mathbf{x}_i), i = 1, \ldots, n\}$ of $(Y, \mathbf{X})$. Our main interest is in estimating the τ_n-th conditional quantile of Y given $\mathbf{X} = \mathbf{x}$, $Q_Y(\tau_n|\mathbf{x})$, which satisfies $P\{Y > Q_Y(\tau_n|\mathbf{x})|\mathbf{x}\} = \tau_n$, where $\tau_n \to 1$ as $n \to \infty$. The conditional quantile $Q_Y(\tau_n|\mathbf{x})$ can also be interpreted as the $1/(1-\tau_n)$ return level of Y given that the covariate $\mathbf{X} = \mathbf{x}$.

The rest of this chapter is organized as follows. In Section 15.2, we review the commonly used methods for estimating unconditional extreme quantiles. In Section 15.3, we discuss four classes of approaches for estimating conditional extreme quantiles: (1) parametric methods; (2) semiparametric methods; (3) nonparametric methods and (4) quantile-regression-based methods. We present some numerical comparison of different estimation methods in Section 15.4 and some final remarks in Section 15.5.

15.2 Estimation of Extreme Unconditional Quantiles

We first review some classic methods for estimating extreme quantiles of a univariate response distribution without considering the covariate information. Let $\{y_1, \ldots, y_n\}$ be a random sample of Y with cumulative distribution function F, and $y_{1,n} \leqslant y_{2,n} \leqslant \cdots \leqslant y_{n,n}$ be the order statistics. Denote $Q(\tau) = F^{\leftarrow}(\tau) = \inf\{y : F(y) \geqslant \tau\}$ as the τ-th quantile of Y. We are interested in estimating the high quantile $Q(\tau_n)$ when $\tau_n \to 1$ as $n \to \infty$.

For a general distribution F, we assume that F belongs to the maximum domain of attraction of an extreme value distribution G_ξ with the extreme value index (EVI) $\xi \in \mathbb{R}$ that measures the heaviness of the tail of F, denoted by $F \in D(G_\xi)$. This means there exist $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$\lim_{n\to\infty} \{F(a_n y + b_n)\}^n = G_\xi(y) = \exp\{-(1+\xi y)^{-1/\xi}\}, \quad 1 + \xi y > 0.$$

This condition is equivalent to

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\xi - 1}{\xi}, \quad x > 0,$$

where $U(t) = F^{\leftarrow}(1 - 1/t) = Q(1 - 1/t)$ and $a(\cdot)$ is some positive function. There are some other equivalent conditions for $F \in D(G_\xi)$, for example see Theorems 1.1.6 and 1.1.8 in de Haan and Ferreira (2006). Based on the above relation, $Q(\tau_n)$ can be estimated by

$$\widehat{Q}(\tau_n) = y_{n-k,n} + \widehat{a}(n/k) \left\{ \left(\frac{k}{np_n} \right)^{\widehat{\xi}} - 1 \right\} \widehat{\xi}^{-1},$$

where $p_n = 1 - \tau_n$, $k = k_n$ is a positive integer such that $k \to \infty$ and $k/n \to 0$, $\widehat{\xi}$ and $\widehat{a}(\cdot)$ are some estimators of ξ and $a(\cdot)$, respectively. The asymptotic normality of $\widehat{Q}(\tau_n)$ can be obtained under some second order conditions. For example, Dekkers et al. (1989) and de Haan and Rootzén (1993) established the asymptotical properties of $\widehat{Q}(\tau_n)$ based on the moment estimator of (ξ, a). In general, if the estimator $(\widehat{\xi}, \widehat{a})$ is asymptotically normal (for example, the maximum likelihood estimator, see Drees et al. (2004)), then the asymptotic properties of $\widehat{Q}(\tau_n)$ can be obtained by applying Theorem 4.3.1 in de Haan and Ferreira (2006).

For a heavy-tailed F, one common assumption is that for some $\xi > 0$,

$$1 - F(y) = y^{-1/\xi} l(y), \quad \text{as } y \to \infty, \tag{15.1}$$

where $l(\cdot)$ is a slowly varying function that satisfies the condition $l(ty)/l(y) \to 1$ as $y \to \infty$ for all $t > 0$. The condition (15.1) is equivalent to the following condition on the quantile function:

$$Q(1 - 1/y) = y^\xi L(y), \tag{15.2}$$

where $L(\cdot)$ is also a slowly varying function and is related to $l(\cdot)$. Consequently, as $\tau \to 1$ and $\tau_n \to 1$, $Q(\tau_n)/Q(\tau) \sim \{(1-\tau)/(1-\tau_n)\}^{\xi}$. This is the basis of the popular Weissman estimator (Weissman, 1978),

$$\widehat{Q}(\tau_n) = y_{n-k,n}\,[k/\{n(1-\tau_n)\}]^{\widehat{\xi}},$$

where $\widehat{\xi}$ is some estimator of ξ, for instance, the Hill estimator (Hill, 1975) $\widehat{\xi} = k^{-1}\sum_{i=1}^{k}\log(y_{n-i+1,n}/y_{n-k,n})$. Under some second order conditions, the asymptotic normality of the Weissman estimator $\widehat{Q}(\tau_n)$ based on some asymptotically normal estimator $\widehat{\xi}$ is presented in Theorem 4.3.8 of de Haan and Ferreira (2006).

The Weismman estimator of extreme quantiles was also adapted to Weibull-tail distributions in Diebolt et al. (2008) and Gardes and Girard (2005). Recently, some bias-reduced extreme quantile estimation methods for heavy-tailed distributions have been developed; see for instance Gomes and Figueiredo (2006), Gomes and Pestana (2007), Li et al. (2010) and references therein. In addition, Drees (2003) discussed the estimation of extreme quantiles for dependent random variables.

15.3 Estimation of Extreme Conditional Quantiles

In this section, we focus on the estimation of extreme high conditional quantile of Y given covariate $\mathbf{x}$, $Q_Y(\tau_n|\mathbf{x})$, where $\tau_n \to 1$ as $n \to \infty$. We discuss four different classes of approaches: parametric, semiparametric, nonparametric and quantile-regression-based methods. The focus of this chapter differs from that in Smith (1994), Portnoy and Jurečková (1999), which studied extreme quantile regression with quantile level $\tau = 0$ or 1.

15.3.1 Parametric Methods

To incorporate the covariate information in modeling extremes, the first class of work fit parametric models such as the generalized extreme value (GEV) distribution based on block maximum data or the generalized Pareto distribution (GPD) based on exceedances over high thresholds, where the location, shape and scale parameters are assumed to depend on covariates parametrically.

One parametric model is based on block maximum data, for example, the annual maximum of daily precipitation. Suppose that Y is the block maximum variable. The basic model assumes that the conditional distribution $F_Y(\cdot|\mathbf{x})$ can be approximated by the GEV distribution, that is,

$$F_Y(y|\mathbf{x}) \approx H\{y;\mu(\mathbf{x}),\sigma(\mathbf{x}),\xi(\mathbf{x})\} = \exp\left[-\left\{1+\xi(\mathbf{x})\frac{y-\mu(\mathbf{x})}{\sigma(\mathbf{x})}\right\}_{+}^{-1/\xi(\mathbf{x})}\right], \quad (15.3)$$

where $\mu(\mathbf{x})$, $\sigma(\mathbf{x})$ and $\xi(\mathbf{x})$ are the location, scale and shape parameters, respectively, and $1+\xi(\mathbf{x})\{y-\mu(\mathbf{x})\}/\sigma(\mathbf{x}) > 0$. The GEV approximation is based on the result of Fisher and Tippett (1928). Under this model assumption, the τ-th conditional quantile of Y given $\mathbf{x}$ is

$$Q_Y(\tau|\mathbf{x}) = \begin{cases} \mu(\mathbf{x}) + \frac{\sigma(\mathbf{x})}{\xi(\mathbf{x})}\{(-\log\tau)^{-\xi(\mathbf{x})} - 1\}, & \xi(\mathbf{x}) \neq 0, \\ \mu(\mathbf{x}) - \sigma(\mathbf{x})\log(-\log\tau), & \xi(\mathbf{x}) = 0. \end{cases} \tag{15.4}$$

To capture the dependence of the distribution of Y on $\mathbf{x}$, one common practice is to model $\mu(\mathbf{x})$, $\sigma(\mathbf{x})$ and $\xi(\mathbf{x})$ as some linear functions of $\mathbf{x}$ after known link transformations, that is, assume

$$\mu(\mathbf{x}) = \Lambda_\mu(\mathbf{x}^T\boldsymbol{\gamma}), \quad \sigma(\mathbf{x}) = \Lambda_\sigma(\mathbf{x}^T\boldsymbol{\beta}), \quad \xi(\mathbf{x}) = \Lambda_\xi(\mathbf{x}^T\boldsymbol{\theta}),$$

where Λ_μ, Λ_σ and Λ_ξ are some known link functions. We can then estimate $\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\theta}$ and the extreme conditional quantile $Q_Y(\tau_n|\mathbf{x})$ by using existing estimation methods such as maximum likelihood estimation. This GEV modeling approach has been considered in Sang and Gelfand (2009), Coles (2001, Chapter 6), Friederichs and Thorarinsdottir (2012), to name a few.

One limitation of the GEV modeling approach based on maximum data is its inefficient use of the available data. This problem can be remedied by using the observations exceeding a high threshold (Davison and Smith, 1990; Smith, 1989). Let u be some high threshold, and $Z = Y - u|Y > u$ be the positive exceedance. Motivated by the GPD approximation result from Pickands (1975), the method assumes that for $z > 0$,

$$F_Z(z|\mathbf{x}) = \frac{F_Y(u+z|\mathbf{x}) - F_Y(u|\mathbf{x})}{1 - F_Y(u|\mathbf{x})} \approx G\{z; \sigma(\mathbf{x}), \xi(\mathbf{x})\}, \tag{15.5}$$

where $G(z;\sigma,\xi) = 1 - (1 + \xi z/\sigma)^{-1/\xi}$ is the cumulative distribution function of GPD, $\sigma(\mathbf{x}) > 0$ and $\xi(\mathbf{x})$ are the scale and shape parameters satisfying $1 + z\xi(\mathbf{x})/\sigma(\mathbf{x}) > 0$. Similar to the method based on maximum data, $\sigma(\mathbf{x})$ and $\xi(\mathbf{x})$ can be modeled parametrically by

$$\sigma(\mathbf{x}) = \Lambda_\sigma(\mathbf{x}^T\boldsymbol{\beta}), \quad \xi(\mathbf{x}) = \Lambda_\xi(\mathbf{x}^T\boldsymbol{\theta}),$$

where Λ_ξ and Λ_σ are some known functions. Let $(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\beta}})$ be the estimator of $(\boldsymbol{\theta}, \boldsymbol{\beta})$ based on the sample $\{(z_i, \mathbf{x}_i), i = 1, \ldots, n\}$ with $z_i = y_i - u > 0$, for instance, the maximum likelihood estimator (Smith, 1985) or the method of moments estimator (Hosking and Wallis, 1987). Consequently, $Q_Y(\tau_n|\mathbf{x})$ can be estimated by

$$\widehat{Q}_Y(\tau_n|\mathbf{x}) = u + \frac{\Lambda_\sigma(\mathbf{x}^T\widehat{\boldsymbol{\beta}})}{\Lambda_\xi(\mathbf{x}^T\widehat{\boldsymbol{\theta}})}\left[\left\{\frac{1 - F_Y(u|\mathbf{x})}{1 - \tau_n}\right\}^{\Lambda_\xi(\mathbf{x}^T\widehat{\boldsymbol{\theta}})} - 1\right].$$

15.3.2 Semiparametric Methods

Instead of assuming an exact distribution form for $Y|\mathbf{x}$ as in the parametric methods discussed in Section 15.3.1, some researchers (Beirlant and Goegebeur, 2003; Wang and Tsai, 2009) considered semiparametric approaches that model the tail of $Y|\mathbf{x}$ as a Pareto-type distribution with parameters depending on $\mathbf{x}$ in a parametric way.

The basic assumption is that the conditional distribution of Y given $\mathbf{x}$ is heavy-tailed or Pareto-type, that is, there exists a $\xi(\mathbf{x}) > 0$ such that

$$1 - F_Y(y|\mathbf{x}) = y^{-1/\xi(\mathbf{x})} l(y; \mathbf{x}), \quad y > 0, \tag{15.6}$$

where $l(\cdot; \mathbf{x})$ is an unknown slowly varying function at infinity, which means that for any $y > 0$, $l(ty; \mathbf{x})/l(t; \mathbf{x}) \to 1$ as $t \to \infty$. The extreme value index $\xi(\mathbf{x})$ is modeled parametrically. For instance, Beirlant and Goegebeur (2003) and Wang and Tsai (2009) assumed that $\xi(\mathbf{x}) = \exp(\mathbf{x}^T \boldsymbol{\beta})$ for some unknown parameter $\boldsymbol{\beta}$.

Suppose that the first element of $\mathbf{x}$ is 1. Write $\mathbf{x} = (1, \tilde{\mathbf{x}}^T)^T$, and $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1^T)^T$ with β_0 denoting the coefficient corresponding to the intercept. Beirlant and Goegebeur (2003) assumed that the transformation $R = R(\boldsymbol{\beta}_1) = Y^{\exp(-\tilde{\mathbf{x}}^T \boldsymbol{\beta}_1)}$ removes the dependence of ξ and l on $\mathbf{x}$ completely, so that $1 - F_R(r|\mathbf{x}) = r^{-1/\xi_0} l(r)$, where $\xi_0 = \exp(\beta_0)$.

Define $Z_j = j(\log R_{n-j+1,n} - \log R_{n-j,n})$, $j = 1, \ldots, n$, where $R_{1,n} \leqslant \cdots \leqslant R_{n,n}$ are the order statistics of the so-called generalized residuals $\{R_1, \ldots, R_n\}$. Under a so-called slow variation with remainder condition on the slowly varying function $l(\cdot)$, Beirlant and Goegebeur (2003) proposed the following exponential regression model:

$$Z_j = \left\{ \xi_0 + b_{n,k} \left(\frac{j}{k+1} \right)^{-\rho} \right\} F_j, \quad j = 1, \ldots, k,$$

where $F_1, \ldots, F_k$ denote independent standard exponential random variables, $\rho < 0$, and $b_{n,k} = b\{(n+1)/(k+1)\}$ with $b(\cdot)$ a rate function satisfying $b(t) \to 0$ as $t \to \infty$. The authors then proposed a maximum likelihood estimation procedure to estimate $\xi_0, \rho, b_{n,k}$ and consequently $\boldsymbol{\beta}$.

Wang and Tsai (2009) proposed an alternative approximate maximum likelihood estimator for $\boldsymbol{\beta}$. They assumed that as $y \to \infty$, the slowly varying function $l(y; \mathbf{x})$ converges to a constant $c(\mathbf{x})$ with a reasonably fast speed. Under this assumption, the distribution of Y given $\mathbf{x}$ can be approximated by an exponential distribution, that is, for sufficiently large y, $f_Y(y|\mathbf{x}) \approx c(\mathbf{x})/\xi(\mathbf{x}) y^{-1/\xi(\mathbf{x})-1}$. Therefore, the approximate maximum likelihood estimator of $\boldsymbol{\beta}$ is defined as

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} \Big\{ \exp(-\mathbf{x}_i^T \boldsymbol{\beta}) \log(y_i/\omega_n) + \mathbf{x}_i^T \boldsymbol{\beta} \Big\} I(y_i > \omega_n),$$

where ω_n is the threshold. Under some second order conditions, Wang and Tsai (2009) established the asymptotic normality of $\widehat{\boldsymbol{\beta}}$.

Once $\boldsymbol{\beta}$ is estimated, by adapting the Weissman estimator, the extreme conditional quantile $Q_Y(\tau_n|\mathbf{x})$ can be estimated by

$$\widehat{Q}_Y(\tau_n|\mathbf{x}) = \widehat{Q}_Y(1 - k/n|\mathbf{x}) \left\{ \frac{k}{n(1-\tau_n)} \right\}^{\exp(\mathbf{x}^T \widehat{\boldsymbol{\beta}})},$$

where $\widehat{Q}_Y(1-k/n|\mathbf{x})$ is some estimation of the $(1-k/n)$-th conditional quantile of Y given $\mathbf{x}$, for instance, $\left\{\widehat{R}_{n-k,n}\right\}^{\exp(\tilde{\mathbf{x}}^T\widehat{\boldsymbol{\beta}}_1)}$ with $\widehat{R}_{n-k,n}$ representing the $(k+1)$-th largest order statistic of the generalized residuals based on $\widehat{\boldsymbol{\beta}}_1$.

15.3.3 Nonparametric Methods

The parametric and semiparamtric methods all model the dependence of the distributional parameters (location, scale and shape) on covariates parametrically, which are often restrictive and may not describe the data well. As an alternative, researchers have considered nonparametric modeling of the distributional parameters, which are more flexible and can be used for exploratory data analysis or for checking the adequacy of a parametric model.

In the current literature, there exist three main classes of nonparametric methods. By focusing on either maximum data or exceedances, the first class of work is based on a likelihood assumption of either GEV distribution or GPD, which allow the parameters to depend on covariates in a nonparametric way. The second class of work is based on a local two-step estimation, where in the first step a subset of data within a neighborhood of $\mathbf{x}$ of interest is selected and then univariate extreme value theory is applied to y_i in the neighborhood to estimate $\xi(\mathbf{x})$ and $Q_Y(\tau_n|\mathbf{x})$ in the second step. In the third class of work, the intermediate conditional quantiles are first obtained by inverting the kernel estimation of the conditional distribution function and then extrapolated to the high tails to estimate extreme conditional quantiles.

15.3.3.1 Likelihood-Based Methods

In Section 15.3.1, we discussed parametric methods that assume either the GEV distribution for block maximum data or the GPD for exceedances over high thresholds, where the form of the dependence of the distributional parameters on $\mathbf{x}$ is fully specified. In many applications, however, the dependence on $\mathbf{x}$ is more complex than what a simple parametric model could accommodate; see Hall and Tajvidi (2000) for examples. To allow more flexibility, we can model the parameters in the GEV distribution or GPD to be nonparametric functions of $\mathbf{x}$. For instance, Davison and Ramesh (2000) assumed the GEV distribution (15.3) for block maximum data, and proposed a local polynomial estimator of $\mu(t)$, $\sigma(t)$ and $\xi(t)$, where t is the univariate time variable. Beirlant and Goegebeur (2004) proposed a local polynomial estimator by fitting the GPD to exceedances over high thresholds. To estimate $Q_Y(\tau_n|\mathbf{x})$ at a given $\mathbf{x}$, the method uses covariate-dependent thresholds $u_{\mathbf{x}}$ and assumes that the positive exceedances $z_i = y_i - u_{\mathbf{x}}$ are independent following the GPD as in (15.5). Focusing on the case with a univariate covariate x, the authors established the consistency and asymptotic normality of the proposed local polynomial estimator, and also suggested a leave-one-out cross validation procedure for choosing the bandwidth h and threshold u_x. Using a similar GPD approximation to exceedances over high thresholds, Chavez-Demoulin and Davison (2005) proposed an alternative smoothing spline estimator obtained by maximizing the penalized GPD likelihood, and studied the finite sample properties of the estimator.

15.3.3.2 Two-Step Local Estimation

Gardes and Girard (2010) developed a nearest-neighbor method and Gardes et al. (2010) developed a moving window approach for estimating the extreme conditional quantiles of heavy-tailed distributions. The main idea of the two methods is to first select observations in a neighborhood of $\mathbf{x}$ of interest, and then apply the univariate extreme value methods to the neighborhood data to estimate the conditional quantiles of Y given $\mathbf{x}$.

Suppose that the design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are nonrandom. Let E be a metric space associated to a metric d. Assume that for all $\mathbf{x} \in E$, $F_Y(\cdot|\mathbf{x})$ is a heavy-tailed distribution with EVI $\xi(\mathbf{x}) > 0$. In the first step of the estimation, Gardes and Girard (2010) proposed to first select $m_{n,\mathbf{x}} = m_{\mathbf{x}}$ nearest covariates of $\mathbf{x}$ (with respect to the distance d), where $m_{\mathbf{x}}$ is a sequence of integers such that $1 < m_{\mathbf{x}} < n$. On the other hand, to accommodate functional covariates, Gardes et al. (2010) proposed to form the neighborhood covariates by including the $m_{n,\mathbf{x}}$ covariates that belong to the ball $B(\mathbf{x}, h_{\mathbf{x}}) = \{\mathbf{t} \in E, d(\mathbf{t}, \mathbf{x}) \leqslant h_{\mathbf{x}}\}$, where $h_{\mathbf{x}}$ is a positive sequence tending to zero as $n \to \infty$, and $m_{\mathbf{x}} = \sum_{i=1}^{n} I\{\mathbf{x}_i \in B(\mathbf{x}, h_{\mathbf{x}})\}$. Denote the covariates in the selected neighborhood by $\{\mathbf{x}_1^*, \ldots, \mathbf{x}_{m_{\mathbf{x}}}^*\}$, and the associated observations taken from $\{y_1, \ldots, y_n\}$ by $\{z_1, \ldots, z_{m_{\mathbf{x}}}\}$. In the second step, univariate extreme value methods are applied to the order statistics $z_{1,m_{\mathbf{x}}} \leqslant \ldots \leqslant z_{m_{\mathbf{x}},m_{\mathbf{x}}}$ to estimate the EVI $\xi(\mathbf{x})$ and $Q_Y(\tau_n|\mathbf{x})$. For instance, Gardes and Girard (2010) considered the following estimator of $\xi(\mathbf{x})$ based on weighted rescaled log-spacings:

$$\widehat{\xi}(\mathbf{x}; a, \lambda) = \sum_{i=1}^{k_{\mathbf{x}}} \{w(i/k_{\mathbf{x}}, a, \lambda) i (\log z_{m_{\mathbf{x}}-i+1,m_{\mathbf{x}}} - \log z_{m_{\mathbf{x}}-i,m_{\mathbf{x}}})\} / \sum_{i=1}^{k_{\mathbf{x}}} w(i/k_{\mathbf{x}}, a, \lambda),$$

where $k_{\mathbf{x}} = k_{n,\mathbf{x}}$ is a sequence of integers such that $1 \leqslant k_{\mathbf{x}} \leqslant m_{\mathbf{x}}$, and

$$w(s, a, \lambda) = \frac{\lambda^{-a}}{\Gamma(a)} s^{1/\lambda - 1} (-\log s)^{a-1}, \text{ for } s \in (0,1), a \leqslant 1, 0 \leqslant \lambda \leqslant 1$$

is the density of log-gamma distribution defined in Consul and Jain (1971). The parameters (a, λ) in the weighting function determine the weights assigned to different extreme order statistics. A special case $a = \lambda = 1$ leads to the Hill estimator (Hill, 1975), and $(a, \lambda) = (2, 1)$ leads to the Zipf estimator (Kratz and Resnick, 1996; Schultze and Steinebach, 1996). Adopting the unconditional quantile estimator proposed by Weissman (1978), based on the local EVI estimator of $\xi(\mathbf{x})$, $Q_Y(\tau_n|\mathbf{x})$ can be estimated by

$$\widehat{Q}_Y(\tau_n|\mathbf{x}) = z_{m_{\mathbf{x}}-k_{\mathbf{x}}+1,m_{\mathbf{x}}} \left\{ \frac{k_{\mathbf{x}}}{m_{\mathbf{x}}(1-\tau_n)} \right\}^{\widehat{\xi}(\mathbf{x})},$$

which can be viewed as an extrapolation from the $(1 - k_{\mathbf{x}}/m_{\mathbf{x}})$-th conditional quantile of Y. Suppose that $k_{\mathbf{x}}$ is an intermediate sequence such that $k_{\mathbf{x}} \to \infty$ and $k_{\mathbf{x}}/m_{\mathbf{x}} \to 0$ as $n \to \infty$. Under the second order condition and some regularity conditions, Gardes and Girard (2010) and Gardes et al. (2010) have established the asymptotic distribution of $\widehat{Q}_Y(\tau_n|\mathbf{x})$.

15.3.3.3 Kernel Estimation

Based on the kernel estimation of $F_Y(\cdot|\mathbf{x})$, Daouia et al. (2011) and Gardes and Girard (2011) proposed kernel-type estimators of the extreme conditional quantile $Q_Y(\tau_n|\mathbf{x})$ for heavy-tailed distributions. Assume that $F_Y(\cdot|\mathbf{x})$ belongs to the Fréchet maximum domain of attraction with EVI $\xi(\mathbf{x})$. For any $(\mathbf{x}, y) \in \mathbb{R}^p \times \mathbb{R}$, Daouia et al. (2011) defined the kernel estimator of $F_Y(y|\mathbf{x})$ as

$$\widehat{F}_Y(y|\mathbf{x}) = 1 - \left\{ \sum_{i=1}^{n} K_h(\mathbf{x} - \mathbf{x}_i) I(y_i > y) \right\} / \sum_{i=1}^{n} K_h(\mathbf{x} - \mathbf{x}_i),$$

where h is the bandwidth such that $h \to 0$ as $n \to \infty$, and $K_h(t) = K(t/h)/h^p$ with K being a p-dimensional kernel function. For any $\tau \in (0, 1)$, the kernel estimator of $Q_Y(\tau|\mathbf{x})$ is defined via the generalized inverse of $\widehat{F}_Y(\cdot|\mathbf{x})$:

$$\widehat{Q}_Y(\tau|\mathbf{x}) = \inf\{t : \widehat{F}_Y(t|\mathbf{x}) \geqslant \tau\}.$$

Daouia et al. (2011) showed that the kernel estimator $\widehat{Q}_Y(\tau_n|\mathbf{x})$ still has the asymptotic normality for intermediate quantiles such that $n(1 - \tau_n) > \{\log(n)\}^p$. However, for extreme order of quantiles, for instance $\tau_n \to 1$ at a rate faster than $1/n$, the kernel estimation is not feasible as it cannot extrapolate beyond the maximum observation in the ball centered at $\mathbf{x}$ with radius h. To overcome this difficulty, Daouia et al. (2011) proposed a Weissman-type estimator of the extreme conditional quantile $Q_Y(\tau_n|\mathbf{x})$:

$$\widehat{Q}_Y(\tau_n|\mathbf{x}) = \widehat{Q}_Y(\alpha_n|\mathbf{x}) \left(\frac{1 - \alpha_n}{1 - \tau_n} \right)^{\widehat{\xi}(\mathbf{x})},$$

where α_n is an intermediate quantile level, $(1 - \tau_n)/(1 - \alpha_n) \to 0$ as $n \to \infty$ and $\widehat{\xi}(\mathbf{x})$ is an estimator of the conditional EVI $\xi(\mathbf{x})$, for instance, a kernel version of the Hill estimator (Hill, 1975)

$$\widehat{\xi}(\mathbf{x}) = \sum_{j=1}^{J} \left(\log \left[\widehat{Q}_Y\{1 - w_j(1 - \alpha_n)|\mathbf{x}\} \right] - \log \left\{ \widehat{Q}_Y(\alpha_n|\mathbf{x}) \right\} \right) / \sum_{j=1}^{J} \log(1/w_j),$$

where $w_1 > w_2 > \ldots > w_J > 0$ is a decreasing sequence of weights and J is a positive integer. This extrapolation allows the estimation of extreme conditional quantile with $\tau_n \to 1$ arbitrarily fast.

The estimation procedure in Gardes and Girard (2011) is similar but the authors considered a different double-kernel estimator of $F_Y(y|\mathbf{x})$:

$$\widehat{F}_Y(y|\mathbf{x}) = 1 - \left[\sum_{i=1}^{n} K_h(\mathbf{x} - \mathbf{x}_i) G\{(y_i - y)/\lambda\} \right] / \sum_{i=1}^{n} K_h(\mathbf{x} - \mathbf{x}_i),$$

where $G(t) = \int_{-\infty}^{t} g(s)ds$ with $g(\cdot)$ being a univariate kernel function, and λ is the bandwidth parameter associated with $G(\cdot)$.

15.3.4 Quantile Regression Methods

Quantile regression, first introduced by Koenker and Bassett (1978), focuses on studying the impact of covariates on the quantiles of the response variable and thus provides a natural alternative to estimating conditional tail quantiles. Researchers have applied quantile regression for estimating tail quantiles in different areas of studies. For instance, Bremnes (2004a) and Bremnes (2004b) used a local quantile regression method to predict the conditional quantiles of precipitation and wind power given outputs from numerical weather prediction models. To account for zero precipitation, Friederichs and Hense (2007) applied a censored linear quantile regression method to estimate the high quantiles of precipitation conditional on the NCEP (National Centers for Environmental Prediction) reanalysis variables. Jagger and Elsner (2008) applied linear quantile regression to study the conditional quantiles of tropical cyclone wind speeds given climate variables. Taylor (2008) proposed an exponentially weighted quantile regression method to estimate the value at risk, which corresponds to the tail quantile of financial returns conditional on the current information. In the above work, conventional parametric or nonparametric quantile regression was directly applied even when the interests are at the extreme tails. However, due to data sparsity, direct estimation from quantile regression is often unstable or infeasible at the extreme tails.

To estimate extreme conditional quantiles in the very far tails with few or no observations available, additional conditions or models for the tails are needed. Chernozhukov and Du (2008), Wang et al. (2012) and Wang and Li (2013) proposed new estimating methods for extreme conditional quantiles that combine linear quantile regression and extreme value theory.

Let $0 < \tau_L < 1$ be a fixed constant that is close to one. Consider the following linear quantile regression model:

$$Q_Y(\tau|\mathbf{x}) = \alpha(\tau) + \mathbf{x}^T \boldsymbol{\beta}(\tau),\ \tau \in [\tau_L, 1], \tag{15.7}$$

where $\alpha(\tau) \in \mathbb{R}$ and $\boldsymbol{\beta}(\tau) \in \mathbb{R}^p$ are the unknown quantile coefficients. Given the random sample $\{(y_i, \mathbf{x}_i), i = 1, \ldots, n\}$, the quantile coefficients can be estimated by

$$(\widehat{\alpha}(\tau), \widehat{\boldsymbol{\beta}}(\tau)) = \underset{\alpha,\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau(y_i - \alpha - \mathbf{x}_i^T \boldsymbol{\beta}), \tag{15.8}$$

where $\rho_\tau(u) = \{\tau - I(u < 0)\}u$ is the quantile loss function.

At the extreme quantiles such that $\tau_n \to 1$ as $n \to \infty$, the conventional quantile regression estimators $\widehat{\alpha}(\tau)$ and $\widehat{\boldsymbol{\beta}}(\tau)$ are often not precise due to data sparsity. The basic idea of the estimation methods in Chernozhukov and Du (2008), Wang et al. (2012) and Wang and Li (2013) is to first estimate less extreme quantiles through conventional quantile regression, and then extrapolate these quantile estimates to the high end based on different assumptions on the tail behavior of the conditional response distribution. We will focus on the estimation for heavy-tailed distributions.

15.3.4.1 Estimation Based on the Common-Slope Assumption

We first consider a common-slope assumption, which assumes that the quantile slope coefficient $\boldsymbol{\beta}(\tau)$ in model (15.7) is constant in the upper quantiles, that is, $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta}$ for $\tau \in [\tau_L, 1]$. In addition, assume that $F_Y(\cdot|\mathbf{x} = 0)$ belongs to the maximum domain of attraction with extreme value index $\xi > 0$. Let $\widehat{e}_i = y_i - \mathbf{x}_i^T\widehat{\boldsymbol{\beta}}$, $i = 1, \ldots, n$, where $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$. For instance, we can take $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\tau_L)$ or the composite estimator proposed in Koenker (1984) and Zou and Yuan (2008), which is obtained by pooling information across a sequence of quantiles $\tau_L = \tau_1 < \ldots < \tau_l = \tau_U$ with $\tau_L < \tau_U < 1$ and $l \geqslant 1$. Let $\widehat{e}_{1,n} \leqslant \cdots \leqslant \widehat{e}_{n,n}$ be the order statistics of $\{\widehat{e}_1, \ldots, \widehat{e}_n\}$. Wang et al. (2012) showed that the upper order statistics of $\widehat{e}_i$ are asymptotically equivalent to those of $Q_Y(u_i|\mathbf{x} = 0)$, where $\{u_1, \ldots, u_n\}$ is a random sample from $U(0, 1)$. Therefore, the order statistics of $\{\widehat{e}_1, \ldots, \widehat{e}_n\}$ can be used to estimate the EVI ξ by existing estimating methods, for instance, the Hill estimator (Hill, 1975),

$$\widehat{\xi} = \frac{1}{k}\sum_{j=1}^{k} \log \frac{\widehat{e}_{n-j+1,n}}{\widehat{e}_{n-k,n}},$$

where k is an integer such that $k = k_n \to \infty$ and $k/n \to 0$ as $n \to \infty$. A Weissman-type extrapolation estimator for $Q_Y(\tau_n|\mathbf{x})$ can be constructed by

$$\widehat{Q}_Y(\tau_n|\mathbf{x}) = \mathbf{x}^T\widehat{\boldsymbol{\beta}} + \left(\frac{k/n}{1-\tau_n}\right)^{\widehat{\xi}} \widehat{e}_{n-k,n},$$

where $1 - \tau_n = o(k/n)$.

15.3.4.2 Estimation without the Common-Slope Assumption

We next discuss an estimation method proposed by Chernozhukov and Du (2008) based on a more relaxed assumption that allows the quantile slope coefficient $\boldsymbol{\beta}(\tau)$ in model (15.7) to vary across τ. In addition to model (15.7), assume that after being transformed by some auxiliary regression line, the response variable Y has regularly varying tails with EVI $\xi > 0$. More specifically, suppose that there exists an auxiliary slope $\boldsymbol{\beta}_e$ such that the following tail-equivalence relationship holds as $\tau \to 1$,

$$Q_Y(\tau|\mathbf{x}) - \mathbf{x}^T\boldsymbol{\beta}_e \sim F_0^{\leftarrow}(\tau), \text{ uniformly in } \mathbf{x}, \tag{15.9}$$

where $F_0(\cdot)$ is a distribution that belongs to the maximum domain of attraction with EVI $\xi > 0$. The tail-equivalence condition (15.9) implies that the covariate $\mathbf{x}$ affects the extreme quantiles of Y through $\boldsymbol{\beta}_e$ approximately.

Under model (15.7) and the tail-equivalence condition (15.9), Chernozhukov (2005) showed that for intermediate order sequences $\tau_n \to 1$ and $n(1 - \tau_n) \to \infty$, $a_n\{\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau)\}$ converges to a normal distribution with mean zero, where $\boldsymbol{\theta}(\tau) = (\alpha(\tau), \boldsymbol{\beta}(\tau)^T)^T$, $\widehat{\boldsymbol{\theta}}(\tau) = (\widehat{\alpha}(\tau), \widehat{\boldsymbol{\beta}}(\tau)^T)^T$ and $a_n = \{(1 - \tau_n)n\}/\left[(1, E(\mathbf{X})^T)^T\{\boldsymbol{\theta}(\tau_n) - \boldsymbol{\theta}(1 - m(1 - \tau_n))\}\right]$ with $m > 1$. This suggests that we can estimate the intermediate conditional quantiles by conventional quantile regression, and then extrapolate these estimates to the high tail to estimate extreme conditional quantiles. With this idea, Chernozhukov and Du (2008) proposed to estimate the EVI ξ by the Hill estimator

$$\widehat{\xi} = \{n(1-\tau_{0n})\}^{-1} \sum_{i=1}^{n} \log \left(\frac{y_i}{\widehat{\alpha}(\tau_{0n}) + \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}(\tau_{0n})} \right)_+ ,$$

where $\log(u)_+ = \log(u) I(u > 0)$, $\tau_{0n} \to 1$ and $n(1-\tau_{0n}) \to \infty$. For $1-\tau_n = o(1-\tau_{0n})$, the Weissman-type extrapolation estimator of $Q_Y(\tau_n|\mathbf{x})$ thus can be constructed as

$$\widehat{Q}_Y(\tau_n|\mathbf{x}) = \{\widehat{\alpha}(\tau_{0n}) + \mathbf{x}^T \widehat{\boldsymbol{\beta}}(\tau_{0n})\} \left(\frac{1-\tau_{0n}}{1-\tau_n} \right)^{\widehat{\xi}} . \tag{15.10}$$

15.3.4.3 Three-Stage Estimation

The methods in Chernozhukov and Du (2008) and Wang et al. (2012) are based on two main assumptions: (1) the conditional quantiles of Y are linear in $\mathbf{x}$ at the upper quantiles; (2) the conditional distribution $F_Y(\cdot|\mathbf{x})$ is tail equivalent across $\mathbf{x}$ with a common EVI ξ. In many applications, the covariate may affect the heaviness of the tail distribution of Y and thus the EVI $\xi(\mathbf{x})$ is dependent on $\mathbf{x}$. It would be interesting to construct a covariate-dependent EVI estimator while still being able to adopt linear quantile regression to borrow information across multi-dimensional covariates. However, Proposition 2.1 in Wang and Li (2013) suggests that in situations where the EVI $\xi(\mathbf{x})$ varies with $\mathbf{x}$, it is rarely the case that the conditional high quantiles of Y are still linear in $\mathbf{x}$. This result suggests that to accommodate covariate-dependent EVI, we have to consider nonparametric quantile regression, which, however, is known to be unstable at tails in finite samples especially when the dimension of $\mathbf{x}$ is high. Wang and Li (2013) showed that in some cases with covariate-dependent EVI, the quantiles of Y may still be linear in $\mathbf{x}$ after some appropriate transformation such as log transformation. Motivated by this, Wang and Li (2013) considered a power-transformed quantile regression model:

$$Q_{\Lambda_\lambda(Y)}(\tau|\mathbf{x}) = \mathbf{z}^T \boldsymbol{\theta}(\tau), \tau \in [\tau_L, 1], \tag{15.11}$$

where $\mathbf{z} = (1, \mathbf{x}^T)^T$ and

$\Lambda_\lambda(y) = \{(y^\lambda - 1)/\lambda\} I(\lambda \neq 0) + \log(y) I(\lambda = 0)$ denotes the family of power transformations (Box and Cox, 1964).

For estimating the extreme conditional quantiles of Y, Wang and Li (2013) proposed a three-stage estimating procedure. In the first stage, the power transformation parameter λ is estimated by

$$\widehat{\lambda} = \underset{\lambda \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} \{R_n(\mathbf{x}_i, \lambda; \tau_L)\}^2 , \tag{15.12}$$

where $R_n(\mathbf{t}, \lambda; \tau) = n^{-1} \sum_{j=1}^{n} I(\mathbf{x}_j \leqslant \mathbf{t}) \left[\tau - I\{\Lambda_\lambda(y_j) - \mathbf{z}_j^T \widehat{\boldsymbol{\theta}}(\tau; \lambda) \leqslant 0\}\right]$ is a residual cusum process that is often used in lack-of-fit tests, $\mathbf{x} \leqslant \mathbf{t}$ means that each component of $\mathbf{x}$ is less than or equal to the corresponding component of $\mathbf{t} \in \mathbb{R}^p$, and $\widehat{\boldsymbol{\theta}}(\tau; \lambda) = \text{argmin}_{\boldsymbol{\theta}} \sum_{i=1}^{n} \rho_\tau \left\{\Lambda_\lambda(y_i) - \mathbf{z}_i^T \boldsymbol{\theta}\right\}$.

In the second stage, the conditional quantiles of Y at a sequence of intermediate quantile levels are estimated by first fitting model (15.11) on the transformed scale and then transforming the estimates back to the original scale. Specifically, define a sequence of quantile levels $\tau_L < \tau_{n-k} < \ldots < \tau_m \in (0, 1)$, where $k = k_n \to \infty$ and $k/n \to 0$, $m = n - [n^\eta]$ with $\eta \in (0, 1)$ as some small constant satisfying $n^\eta < k$, and $\tau_j = j/(n + 1)$. The trimming parameter η is introduced for technical purposes, more specifically, to obtain the Bahadur representation of $\widehat{\boldsymbol{\theta}}(\tau_m; \lambda)$. In practice we can choose $\eta = 0.1$. For each $j = n - k, \ldots, m$, $Q_{\Lambda_\lambda(Y)}(\tau_j|\mathbf{x})$ can be estimated by $\mathbf{z}^T \widehat{\boldsymbol{\theta}}(\tau_j; \widehat{\lambda})$. By the equivariance property of quantiles to monotone transformations, we can estimate $Q_Y(\tau_j|\mathbf{x})$ by $\widehat{Q}_Y(\tau_j|\mathbf{x}) = \Lambda_{\widehat{\lambda}}^{-1} \left\{\mathbf{z}^T \widehat{\boldsymbol{\theta}}(\tau_j; \widehat{\lambda})\right\}$. For a given $\mathbf{x}$, $\{\widehat{Q}_Y(\tau_j|\mathbf{x}), j = n - k, \ldots, m\}$ can be roughly regarded as the extreme order statistics of a sample from $F_Y(\cdot|\mathbf{x})$.

In the third stage, extrapolation from the intermediate quantile estimates is performed to estimate $Q_Y(\tau_n|\mathbf{x})$ with $1 - \tau_n = o(k/n)$ as

$$\widehat{Q}_Y(\tau_n|\mathbf{x}) = \widehat{Q}_Y(\tau_{n-k}|\mathbf{x}) \left(\frac{1 - \tau_{n-k}}{1 - \tau_n}\right)^{\widehat{\xi}(\mathbf{x})}, \tag{15.13}$$

where $\widehat{\xi}(\mathbf{x}) = (k - [n^\eta])^{-1} \sum_{j=[n^\eta]}^{k} \log \left\{\widehat{Q}_Y(\tau_{n-j}|\mathbf{x}) / \widehat{Q}_Y(\tau_{n-k}|\mathbf{x})\right\}$ is the Hill estimator based on the pseudo order statistics of a sample from $F_Y(\cdot|\mathbf{x})$.

The method in Wang and Li (2013) allows the EVI $\xi(\mathbf{x})$ to depend on $\mathbf{x}$ and thus provides more flexibility. However, due to lack of information, the covariate-dependent EVI estimator could be unstable in regions where $\mathbf{x}$ is sparse. In situations where $\xi(\mathbf{x})$ is constant across $\mathbf{x}$ (or in some region of $\mathbf{x}$), we can estimate the common ξ by the pooled estimator $\widehat{\xi}_p = n^{-1} \sum_{i=1}^{n} \widehat{\xi}(\mathbf{x}_i)$. Numerical studies in Wang and Li (2013) showed that the pooled EVI estimator often leads to more stable and efficient estimation of the extreme conditional quantiles when $\xi(\mathbf{x})$ is indeed constant or varies little across $\mathbf{x}$. To identify the commonality of the EVI, Wang and Li (2013) proposed a test statistic $T_n = n^{-1} \sum_{i=1}^{n} \{\widehat{\xi}(\mathbf{x}_i) - \widehat{\xi}_p\}^2$ and established the asymptotic distribution of T_n under $H_0 : \xi(\mathbf{x}) = \xi$ for all $\mathbf{x}$ in its support. Suppose that $E(\mathbf{X}) = \mathbf{0}_p$. For two special cases: (1) homogenous case such as the location-shift model (Koenker, 2005); (2) the EVI of $\Lambda_\lambda(Y)$ is $\xi^* = 0$, it was shown that $kT_n \xrightarrow{d} \xi^2 \chi^2(p - 1)$ under H_0, so the test can be easily carried out.

The quantile-regression-based methods discussed in this section can be regarded as semiparametric methods since they make no parametric distributional assumptions but assume that the conditional upper quantiles of the response (or some transformation thereof) are linear in covariates. This quantile linearity assumption allows us to model the effect of covariates $\mathbf{x}$ across the entire range of $\mathbf{x}$ and thus borrow information across $\mathbf{x}$ to estimate the extreme conditional quantiles of the response, and to avoid the curse of dimensionality issue faced by the nonparametric methods.

15.4 Numerical Comparison

We carry out a simulation study to compare different methods for estimating extremely high conditional quantiles. The data are generated from the following four different models.

- Model 1: $y_i = 2 + 2x_{i1} + 2x_{i2} + (2 + 1.6x_{i1})\epsilon_i$, ϵ_i ~Pareto(0.5), $i = 1, \ldots, n$.
- Model 2: $y_i|x_i \sim$ Pareto with $\xi(\mathbf{x}_i) = \exp(-1 + x_i)$, $i = 1, \ldots, n$.
- Model 3: $\log(y_i) = 2 + x_{i1} + x_{i2} + (0.5 + 0.25x_{i1})\epsilon_i$, and ϵ_i are i.i.d. random variables with quantile function $Q(\tau) = \tau - 1 - \log(1 - \tau)$ for $\tau \in (0, 1)$, $i = 1, \ldots, n$. In this case, the conditional distribution of Y is in the domain of attraction with EVI $\xi(x_{i1}, x_{i2}) = 0.5 + 0.25x_{i1}$.
- Model 4: $y_i|x_i \sim$ Fréchet distribution with distribution function $F_Y(y|x_i) = \exp\{-y^{-1/\xi(x_i)}\}$, where $\xi(x) = 1/2[1/10 + \sin\{\pi(x + 1)/2\}]\{11/10 - 1/2\exp(-16x^2)\}$, $i = 1, \ldots, n$.

In the four models, $x_{i1}, x_{i2}, x_i, i = 1, \ldots, n$, are independent random variables from Uniform$(-1, 1)$. The sample size is set as $n = 2000$. For each model, the simulation is repeated 500 times.

We compare five estimators: (1) the parametric method assuming GPD for the exceedances with scale $\sigma(\mathbf{x}) = \exp(\mathbf{x}^T\boldsymbol{\beta})$ and shape $\xi(\mathbf{x}) = \exp(\mathbf{x}^T\boldsymbol{\theta})$; (2) the semiparametric tail index regression (TIR) method of Wang and Tsai (2009); (3) the nonparametric kernel method (KER) of Daouia et al. (2011); (4) the quantile-regression-based method of Chernozhukov and Du (2008), denoted by CD; (5) the three-stage estimator (3Stage) of Wang and Li (2013).

The extreme value index $\xi(\mathbf{x})$ is a constant in Model 1, while it depends on the covariates in different ways in Models 2–4. The CD method is based on the assumption of linear conditional quantiles of Y, which is satisfied in Model 1 but violated in Models 2–4. Since the conditional quantiles of Y are linear in $\mathbf{x}$ after log transformation in Model 3, the model assumption required by the 3Stage method is satisfied in Models 1 and 3 but violated in Models 2 and 4. Both the GPD and TIR methods assume that $\log\{\xi(\mathbf{x})\}$ is linear in $\mathbf{x}$, and this assumption is satisfied only in Models 1 and 2. The KER method is most flexible and it works in all four cases.

Table 15.1 summarizes the performance of different estimators for estimating $Q_Y(\tau_n|\mathbf{x})$ at $\tau_n = 0.99$ and 0.995. The IBias is the integrated bias defined as the average of $n^{-1}\sum_{i=1}^n\{\widehat{Q}_Y(\tau_n|\mathbf{x}_i) - Q_Y(\tau_n|\mathbf{x}_i)\}$, and RIMSE is the root integrated mean squared error defined as the square root of the average of $n^{-1}\sum_{i=1}^n\{\widehat{Q}_Y(\tau_n|\mathbf{x}_i) - Q_Y(\tau_n|\mathbf{x}_i)\}^2$ across 500 simulated data. The GPD and TIR methods rely on the correct specification of the EVI function; they perform competitively well in Model 2 when $\xi(\mathbf{x})$ is correctly specified but they are slightly less efficient than the 3Stage method in Model 3 when the function form is misspecified.

TABLE 15.1
The integrated bias (IBias) and root integrated mean squared error (RIMSE) of different estimators of $Q_Y(\tau_n|\mathbf{x}_i)$ at $\tau_n = 0.99$ and 0.995. Values in the parentheses are the standard errors. The results of KER and 3Stage are taken from Tables 1–2 of Wang and Li (2013).

	IBias		RIMSE	
Method	$\tau_n = 0.99$	$\tau_n = 0.995$	$\tau_n = 0.99$	$\tau_n = 0.995$
Model 1 ($p = 2$, constant EVI)				
GPD	0.27 (0.16)	0.75 (0.42)	3.53 (4.47)	9.33 (9.33)
TIR	-0.72 (0.07)	-1.87 (0.14)	1.78 (0.50)	3.63 (1.04)
KER	0.77 (0.55)	2.08 (2.74)	12.23 (15.06)	61.29 (161.39)
CD	0.04 (0.06)	-0.26 (0.10)	1.28 (0.37)	2.30 (0.66)
3Stage	-0.52 (0.07)	-1.62 (0.13)	1.75 (0.53)	3.26 (0.87)
Model 2 ($p = 1$, Pareto distribution)				
GPD	-2.63 (0.53)	-5.65 (1.02)	12.09 (5.19)	23.45 (10.38)
TIR	-2.60 (0.58)	-5.90 (1.09)	13.25 (5.17)	25.02 (10.13)
KER	1.66 (0.81)	3.97 (2.32)	18.19 (13.35)	52.04 (58.19)
CD	-3.66 (0.74)	-9.05 (1.52)	16.97 (6.37)	35.15 (13.48)
3Stage	-1.67 (0.42)	-3.45 (1.11)	9.58 (9.74)	24.96 (29.69)
Model 3 ($p = 2$, EVI linear in $\mathbf{x}$)				
GPD	1.52 (4.89)	16.29 (11.84)	109.34 (83.00)	265.27 (171.47)
TIR	6.13 (6.53)	17.07 (9.85)	146.16 (60.98)	221.01 (94.47)
KER	18.31 (15.20)	74.73 (85.72)	340.28 (483.88)	1918.17 (4441.34)
CD	-15.55 (8.09)	-43.81 (13.89)	181.47 (75.09)	313.72 (131.05)
3Stage	-14.01 (3.82)	-41.09 (7.90)	86.47 (90.71)	181.27 (203.67)
Model 4 ($p = 1$, Fréchet distribution)				
GPD	0.77 (0.16)	1.65 (0.31)	3.57 (1.61)	7.04 (3.63)
TIR	0.48 (0.12)	0.83 (0.18)	2.75 (0.61)	4.10 (0.90)
KER	0.26 (0.08)	0.67 (0.19)	1.90 (0.85)	4.22 (2.48)
CD	0.46 (0.12)	0.78 (0.18)	2.74 (0.60)	4.12 (0.90)
3Stage	0.13 (0.12)	0.07 (0.19)	2.79 (0.63)	4.27 (1.01)

In Model 1 with a constant EVI, the GPD and TIR methods underperform the CD and 3Stage methods due to the noise involved in estimating the zero parameters in $\xi(\mathbf{x}) = \mathbf{x}^T\boldsymbol{\theta}$. The CD estimator is slightly more efficient than 3Stage in Model 1 when the conditional quantiles of untransformed Y are linear in covariates but the latter is more flexible and performs competitively well or slightly better in the other three models even in Models 2 and 4 where the power-transformation model (15.11) is violated. As observed in Wang and Li (2013), the nonparametric method KER is the most flexible and can capture complicated dependence of EVI on the covariates for instance as in Model 4. However, the KER method gives unstable estimation especially in Models 1 and 3 with two predictors due to the data sparsity and curse of dimensionality.

15.5 Final Remarks

Estimation of conditional extreme quantiles has drawn much attention in recent years. We surveyed and compared various types of estimation methods based on different model assumptions. As for most statistical problems, there is a tradeoff between the model flexibility and stability. The nonparametric methods are more flexible but are subject to the curse of dimensionality and reduced effective sample size with local estimation. The parametric methods are sensitive to the misspecification of models. Semiparametric methods aim to achieve a better balance between model flexibility and parsimony. Regarded as also semiparametric, the quantile-regression-based methods discussed in Section 15.3.4 are relatively newer to the extreme value literature but they serve as useful alternative tools in cases where the conditional tail quantiles of the response after some transformation appear to be linear in covariates. In practice, we would suggest first use nonparametric methods as exploratory tools to examine the dependence of extreme value index and the conditional tail quantiles on the covariates, which may help identify a reasonable parametric or semiparametric model to carry out analysis with less variability.

Acknowledgements

The research of Wang is partially supported by the National Science Foundation CAREER award DMS-1149355. The research of Li is partially supported by the National Natural Science Foundation of China grant 11171074.

References

Abrevaya, J. (2001), "The effect of demographics and maternal behavior on the distribution of birth outcomes," *Empirical Economics*, 26, 247–259.

Beirlant, J. and Goegebeur, Y. (2003), "Regression with response distributions of Pareto-type," *Computational Statistics and Data Analysis*, 42, 595–619.

— (2004), "Local polynomial maximum likelihood estimation for Pareto-type distributions," *Journal of Multivariate Analysis*, 89, 97–118.

Box, G. E. P. and Cox, D. R. (1964), "An analysis of transformations," *Journal of the*

Royal Statistical Society, Series B, 26, 211–252.

Bremnes, J. B. (2004a), "Probabilistic forecasts of precipitation in terms of quantiles using NWP model output," *Monthly Weather Review*, 132, 338–347.

— (2004b), "Probabilistic wind power forecasts using local quantile regression," *Wind Energy*, 7, 47–54.

Chavez-Demoulin, V. and Davison, A. C. (2005), "Generalized additive modelling of sample extremes," *Journal of the Royal Statistical Society, Series C*, 54, 207–222.

Chernozhukov, V. (2005), "Extremal quantile regression," *Annals of Statistics*, 33, 806–839.

Chernozhukov, V. and Du, S. (2008), "Extremal quantiles and value-at-risk," in *The New Palgrave Dictionary of Economics*, eds. Durlauf, S. N. and Blume, L. E., Basingstoke: Palgrave Macmillan.

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer.

Consul, P. C. and Jain, G. C. (1971), "On the log-gamma distribution and its properties," *Statistische Hefte*, 12, 100–106.

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

Davison, A. C. and Ramesh, N. I. (2000), "Local likelihood smoothing of sample extremes," *Journal of the Royal Statistical Society Series B,*, 62, 191–208.

Davison, A. C. and Smith, R. L. (1990), "Models for exceedances over high thresholds," *Journal of the Royal Statistical Society. Series B*, 52, 393–442.

de Haan, L. and Ferreira, A. (2006), *Extreme Value Theory: An Introduction*, Springer.

de Haan, L. and Rootzén, H. (1993), "On the estimation of high quantiles," *Journal of Statistical Planning and Inference*, 35, 1–13.

Dekkers, A., Einmahl, J., and de Haan, L. (1989), "A moment estimator for the index of an extreme-value distribution," *Annals of Statistics*, 17, 1833–1855.

Diebolt, J., Gardes, L., Girard, S., and Guillou, A. (2008), "Bias-reduced extreme quantiles estimators of Weibull distributions," *Journal of Statistical Planning and Inference*, 138, 1389–1401.

Drees, H. (2003), "Extreme quantile estimation for dependent data with applications to finance," *Bernoulli*, 9, 617–657.

Drees, H., Ferreira, A., and de Haan, L. (2004), "On maximum likelihood estimation of the extreme value index," *Annals of Apllied Probability*, 14, 1179–1201.

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

Fisher, R. A. and Tippett, L. H. C. (1928), "Limiting forms of the frequency distribution in the largest particle size and smallest number of a sample," *Proceedings of the Cambridge Philosophical Society*, 24, 180–190.

Friederichs, P. (2010), "Statistical downscaling of extreme precipitation events using extreme value theory," *Extremes*, 13, 109–132.

Friederichs, P. and Hense, A. (2007), "Statistical downscaling of extreme precipitation events using censored quantile regression," *Monthly Weather review*, 135, 2365–2378.

Friederichs, P. and Thorarinsdottir, T. L. (2012), "Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction," *Environmetrics*, 23, 579–594.

Gardes, L. and Girard, S. (2005), "Estimating extreme quantiles of Weibull tail-distributions," *Communications in Statistics–Theory and Methods*, 34, 1065–1080.

— (2010), "Conditional extremes from heavy-tailed distributions: an application to the estimation of extreme rainfall return levels," *Extremes*, 13, 177–204.

— (2011), "Functional Kernel Estimators of Conditional Extreme Quantiles," in *Recent Advances in Functional Data Analysis and Related Topics Contributions to Statistics*, Springer, pp. 135–140.

Gardes, L., Girard, S., and Lekina, A. (2010), "Functional nonparametric estimation of conditional extreme quantiles," *Journal of Multivariate Analysis*, 101, 419–433.

Gomes, M. and Figueiredo, F. (2006), "Bias reduction in risk modelling: semiparametric quantile estimation," *Test*, 15, 375–396.

Gomes, M. I. and Pestana, D. (2007), "A sturdy reduced-bias extreme quantile (VaR) estimator," *Journal of the American Statistical Association*, 102, 280–292.

Hall, P. and Tajvidi, N. (2000), "Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data," *Statistical Science*, 15, 153–167.

Hill, B. M. (1975), "A simple general approach to inference about the tail of a distribution," *Annals of Statistics*, 3, 1163–1174.

Hosking, J. R. M. and Wallis, J. R. (1987), "Parameter and quantile estimation for the generalized Pareto distribution," *Technometrics*, 339, 339–349.

Jagger, T. H. and Elsner, J. B. (2008), "Modeling tropical cyclone intensity with quantile regression," *International Journal of Climatology*, 29, 1351–1361.

Koenker, R. (1984), "A note on L-estimates for linear models," *Statistics and Probability Letters*, 2, 323–325.

— (2005), *Quantile Regression*, Cambridge University Press, Cambridge.

Koenker, R. and Bassett, G. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.

Kratz, M. and Resnick, S. (1996), "The QQ-estimator and heavy tails," *Stochastic Models*, 12, 699–724.

Li, D., Peng, L., and Yang, J. (2010), "Bias reduction for high quantiles," *Journal of Statistical Planning and Inference*, 140, 2433–2441.

Pickands, J. (1975), "Statistical inference using extreme order statistics," *Annals of Statistics*, 3, 119–131.

Portnoy, S. and Jurečková (1999), "On extreme Regression Quantiles," *Extremes*, 2, 227–243.

Sang, H. and Gelfand, A. E. (2009), "Hierarchical modeling for extreme values observed over space and time," *Environmental and Ecological Statistics*, 16, 407–426.

Schultze, J. and Steinebach, J. (1996), "On least squares estimates of an exponential tail coefficient," *Statistics and Decisions*, 14, 353–372.

Smith, R. L. (1985), "Maximum likelihood estimation in a class of nonregular cases," *Biometrika*, 72, 67–92.

— (1989), "Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone," *Statistical Science*, 4, 367–377.

— (1994), "Nonregular Regression," *Biometrika*, 81, 173–183.

Taylor, J. W. (2008), "Using exponentially weighted quantile regression to estimate value at risk and expected shortfall," *Journal of Financial Econometrics*, 6, 382–406.

Wang, H. and Li, D. (2013), "Estimation of extreme conditional quantiles through power trnsformation," *Journal of the American Statistical Association*, 108, 1062–1074.

Wang, H., Li, D., and He, X. (2012), "Estimation of high conditional quantiles for heavy-tailed distributions," *Journal of the American Statistical Association*, 107, 1453–1464.

Wang, H. and Tsai, C. L. (2009), "Tail index regression," *Journal of the American Statistical Association*, 104, 1233–1240.

Weissman, I. (1978), "Estimation of parameters and large quantiles based on the k largest observations," *Journal of the American Statistical Association*, 73, 812–815.

Zou, H. and Yuan, M. (2008), "Composite quantile regression and the oracle model selection theory," *The Annals of Statistics*, 36, 1108–1126.