Journal of Statistical Planning and Inference 176 (2016) 64–77

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Estimation of high conditional quantiles using the Hill estimator of the tail index

Fengyang He [a], Yebin Cheng [b,*], Tiejun Tong [c]

[a] School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China
[b] Glorious Sun School of Business and Management, Donghua University, Shanghai, China
[c] Department of Mathematics, Hong Kong Baptist University, Hong Kong

ARTICLE INFO

Article history:
Received 10 March 2016
Received in revised form 16 March 2016
Accepted 21 March 2016
Available online 7 April 2016

Keywords:
Extreme value theory
High conditional quantiles
Hill estimator
Pickands estimator
Quantile regression
Tail index

ABSTRACT

To implement the extremal quantile regression, one needs to have an accurate estimate of the tail index that is involved in the limit distributions of extremal regression quantiles. However, the existing quantile estimation methods are often unstable owing to data sparsity in the tails. In this paper, we propose the Hill estimator for the tail index based on regression quantiles, and construct a new estimator for high conditional quantiles through an extrapolation of the intermediate regression quantiles. In both theory and simulation, we demonstrate that the proposed estimators are more efficient than those based on the refined Pickands estimator of the tail index. The applicability of the new method is also illustrated on the Occidental Petroleum daily stock return data.

1. Introduction

In the influential paper of Chernozhukov (2005), the author developed a theory of quantile regression in the tails by joining the linear quantile regression model together with the extreme value theory (EVT) and obtained the asymptotic properties of extremal (extreme order and intermediate order) quantile regression estimators. Specifically, it was shown that the intermediate order regression quantiles and their functionals converge in distribution to normal vectors with covariance matrices dependent on the tail index γ, and the extreme order regression quantiles converge weakly to argmin functionals of stochastic integrals of Poisson processes that are also dependent on the tail index. To implement the extremal quantile regression, we hence need to have an accurate estimate of the tail index. Chernozhukov (2005) provided the Pickands estimator of γ based on regression quantiles. However, the variance of the Pickands estimator, as implied in Theorem 3.3.5 in de Haan and Ferreira (2006), can be huge for large values of the tail index, and consequently, the estimated extreme regression quantiles may not be reliable. On the other hand, it is known that the conventional quantile estimation methods are often unstable owing to data sparseness in the tails. In view of this, in this paper we propose the Hill estimator for the tail index based on regression quantiles, and meanwhile, examine the accuracy of the estimation for high conditional quantiles through an extrapolation of the intermediate regression quantiles.

Needless to say, estimation of high unconditional quantiles is very important and has been investigated extensively in the literature, for which we refer to de Haan and Ferreira (2006) for an overview of the available methodology. Wang et al. (2012) estimated high conditional quantiles for heavy-tailed distribution through an extrapolation in the conventional quantile

* Corresponding author.
E-mail address: chengyebin@hotmail.com (Y. Cheng).

http://dx.doi.org/10.1016/j.jspi.2016.03.003
0378-3758/

regression framework. Wang and Li (2013) estimated extreme conditional quantiles through power transformation and also established the asymptotic normality for the proposed tail index and extreme conditional quantiles estimators. However, we note that their estimators are asymptotically biased under the strong second order condition. More details on the second order condition can be seen, for instance, in de Haan and Ferreira (2006). In the content of nonlinear quantile regression, Daouia et al. (2011) introduced nonparametric kernel methods to estimate arbitrary high order conditional quantiles, in particular for the setting of heavy-tailed conditional distributions. Daouia et al. (2013) extended the methodology in Daouia et al. (2011) to adapt to more general settings. Furthermore, Gardes et al. (2010) and Gardes and Girard (2012) have devoted to the estimation of functional extreme quantiles based on nonparametric methods when functional covariate information is available.

The remainder of this paper is organized as follows. In Section 2, we specify the distribution form to be dealt with, introduce the common assumptions, and present our new methodology for estimating high conditional quantiles and the tail index. Theoretical results of the proposed estimators, together with some specific results for certain heavy-tailed distributions, are presented in Section 3. In Section 4, we conduct simulation studies to evaluate the finite sample performance of the proposed estimators and compare them with existing methods. We then illustrate the applicability of the new method on the Occidental Petroleum daily stock return in Section 5, and provide the technical proofs in Appendices.

2. Methodology

In this section, we first present the required assumptions and then propose our new estimator for high conditional quantiles and also the Hill estimator of the tail index based on regression quantiles.

2.1. Assumptions

Let $\{(X_i, Y_i), i = 1, \ldots, n\}$ be a sequence of independent random vectors from the population of (X, Y), where $X = (1, X'_{-1})'$ is a d-dimensional covariate and Y is a 1-dimensional response. Let $\mathbf{X}$ be the support set of X and $F_Y(y|x)$ be the continuous conditional distribution function of Y given $X = x$. Let also $\bar{F}_Y(y|x) = 1 - F_Y(y|x)$ and $q_Y(\tau|x) = \bar{F}_Y^{-1}(y|x) = \inf\left\{y : \bar{F}_Y(y|x) \le \tau\right\}$ be the related $(1-\tau)$th conditional quantile. In this paper, we consider the following linear quantile regression model:

$$q_Y(\tau|x) = x'\beta(\tau), \quad \forall \tau \in [0, \tau_u],\ x \in \mathbf{X}, \tag{2.1}$$

with some real $0 < \tau_u \le 1$. The assumptions in this paper are listed as follows.

(A1) There exist a bounded vector $\beta_r \in \mathbb{R}^d$ and some distribution function F_u such that

$$U \equiv Y - X'\beta_r \quad \text{with } x_U^* = \infty, \tag{2.2}$$

$$\bar{F}_U(t|x) \sim K(x) \cdot \bar{F}_u(t) \quad \text{uniformly in } x \in \mathbf{X},\ \text{as } t \uparrow x_U^*, \tag{2.3}$$

where $x_U^* = \inf\left\{y : \bar{F}_U(y|x) \le 0\right\}$ is the upper endpoint. $\bar{F}_U(\cdot|x)$ is the conditional survival function of U given $X = x$, and $K(\cdot) > 0$ is a continuous bounded function on $\mathbf{X}$. For simplicity, let $K(x) = 1$ at $x = \mu_X = \mathrm{E}X$, and $\bar{F}_u(t) \equiv \bar{F}_U(t|\mu_X)$.

(A2) For the random variable u in (A1), we assume that $\bar{F}_u(t) = \mathrm{P}(u > t)$ satisfies

$$\bar{F}_u(t) = c \exp\left\{-\int_1^t \left(\frac{1}{\gamma} - \varepsilon(v)\right)\frac{dv}{v}\right\}, \tag{2.4}$$

where $c > 0$, $\gamma > 0$, $\varepsilon : (0, +\infty) \to \mathbb{R}$, and $\varepsilon(t) \to 0$ as $t \to \infty$.

(A3) $\mathbf{X}$ is a compact set in $\mathbb{R}^d$ and $\mathrm{E}(XX')$ is a positive definite matrix. For simplicity, let $\mu_X = \mathrm{E}X = (1, 0, \ldots, 0)'$.

(A4) Assume that

$$\begin{aligned} &\text{(i)}\ \frac{\partial \bar{F}_U^{-1}(\tau|x)}{\partial \tau} \sim \frac{\partial \bar{F}_u^{-1}(\tau/K(x))}{\partial \tau} \quad \text{uniformly in } x \in \mathbf{X}, \\ &\text{(ii)}\ \frac{-\partial \bar{F}_u^{-1}(\tau)}{\partial \tau} \text{ is regularly varying at 0 with exponent } -\gamma - 1. \end{aligned} \tag{2.5}$$

(A5) $|\varepsilon(t)|$ is a continuous function and is ultimately non-increasing.

For assumption (A2), it is a special case of Karamata's representation (see corollary of Theorem 0.6 in Resnick, 1987) of a regularly varying function. (A1), (A3) and (A4) are essentially the same as conditions R1–R3 in Chernozhukov (2005), respectively. The last assumption is also mild in the regularly varying framework. It is noteworthy that many important heavy-tailed distributions are of form (2.4), for which we have provided four examples in Table 1. Moreover, (2.4) implies that $\bar{F}_u(\cdot)$ is regularly varying at infinity with index $-1/\gamma$, i.e., $\bar{F}_u(\cdot) \in RV(-1/\gamma)$. That is to say for all $\lambda > 0$,

$$\lim_{t\to\infty} \frac{\bar{F}_u(\lambda t)}{\bar{F}_u(t)} = \lambda^{-\frac{1}{\gamma}}. \tag{2.6}$$