Journal of Statistical Planning and Inference 142 (2012) 1586–1598

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

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

Estimating the conditional tail index by integrating a kernel conditional quantile estimator

L. Gardes [a], A. Guillou [b,*], A. Schorgen [b]

[a] INRIA Rhône-Alpes, projet Mistis, Inovallée, 655 av. de l'Europe, 38334 Montbonnot, Saint-Ismier cedex, France
[b] Université de Strasbourg et CNRS, IRMA, UMR 7501, 7 rue René Descartes, 67084 Strasbourg cedex, France

ARTICLE INFO

Article history:
Received 18 March 2011
Received in revised form
4 January 2012
Accepted 5 January 2012
Available online 17 January 2012

Keywords:
Heavy-tailed distribution
Covariates
Kernel estimator
Asymptotic normality

ABSTRACT

This paper deals with the estimation of the tail index of a heavy-tailed distribution in the presence of covariates. A class of estimators is proposed in this context and its asymptotic normality established under mild regularity conditions. These estimators are functions of a kernel conditional quantile estimator depending on some tuning parameters. The finite sample properties of our estimators are illustrated on a small simulation study.

1. Introduction

Tail index estimation has been extensively studied in the literature and several estimators proposed. The most popular semi-parametric estimator was proposed by Hill (1975) in the context of heavy-tailed distributions. Diebolt et al. (2008) considered the case of Weibull-tail distributions and the general case was studied by Dekkers et al. (1989). The aim of this paper is to extend this framework to the case where some covariate x is recorded simultaneously with the variable of interest Y. In this case, the tail index depends on the covariate x and is referred in the following to as the conditional tail index.

Such a context has been already considered by Smith (1989) or Davison and Smith (1990) who proposed parametric models, while semi-parametric approaches have been used in Hall and Tajvidi (2000) or Beirlant and Goegebeur (2003). Also, fully non-parametric methods can be used for instance based on splines (see Chavez-Demoulin and Davison, 2005) or local polynomials (see Davison and Ramesh, 2000). These latter estimators have been then extended by Beirlant and Goegebeur (2003) in case of multidimensional covariates. More recently, Gardes and Girard (2010) have addressed the problem of the estimation of a conditional extreme quantile via a nearest neighbor approach.

We propose here new conditional tail index estimators based on a class of functions satisfying some mild conditions. We only assume that the conditional distribution of Y given x is heavy-tailed whereas no parametric assumption on the covariates x is made. The conditional distribution of Y given x is then of polynomial-type, with a rate of convergence driven by the conditional tail index.

* Corresponding author. Tel.: +33 3 68 85 01 99; fax: +33 3 68 85 03 28.
E-mail address: armelle.guillou@math.unistra.fr (A. Guillou).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2012.01.011

The remainder of this paper is organized as follows. Our class of estimators is given in Section 2. Their asymptotic normality is established in Section 3. In Section 4, a sub-class of estimators is studied and their finite sample properties are illustrated in Section 5. All the proofs are postponed to Appendix A.

2. A class of conditional tail index estimators

Let E be a metric space associated to a distance d. For $x \in E$, denote by $F(.,x)$ (resp. $q(.,x)$) the conditional distribution function (resp. the conditional quantile function) of Y given x. We assume that for all $x \in E$,

$$\overline{F}(y,x) = y^{-1/\gamma(x)} L(y,x) \tag{1}$$

or, equivalently, for $\alpha \in (0,1)$,

$$q(\alpha,x) = \overline{F}^{\leftarrow}(\alpha,x) = \alpha^{-\gamma(x)} \ell(\alpha^{-1},x),$$

where $\overline{F}^{\leftarrow}(\alpha,x) = \inf\{t, \overline{F}(t,x) \leq \alpha\}$ denotes the generalized inverse of the conditional survival function and, for x fixed, $L(.,x)$ and $\ell(.,x)$ are slowly varying functions, that is for all $\lambda > 0$,

$$\lim_{y\to\infty} \frac{L(\lambda y,x)}{L(y,x)} = \lim_{y\to\infty} \frac{\ell(\lambda y,x)}{\ell(y,x)} = 1.$$

Here, $\gamma(.)$ is a unknown positive function of the covariate referred in the following to as the conditional tail index. For a given $x \in E$, our aim is to propose an estimator of $\gamma(x)$. Suppose that we have at our disposal pairs $(Y_1,x_1),\ldots,(Y_n,x_n)$ of independent observations from model (1) where the design points $x_1,\ldots,x_n$ are assumed to be non-random. We propose to estimate the conditional tail index as a function of the kernel conditional quantile estimator $\widehat{q}_n(\alpha,x) := \widehat{\overline{F}}_n^{\leftarrow}(\alpha,x)$, with

$$\widehat{\overline{F}}_n(y,x) = \sum_{i=1}^{n} H\left(\frac{d(x,x_i)}{h_{1,n}}\right) K\left(\frac{y-Y_i}{h_{2,n}}\right) \Bigg/ \sum_{i=1}^{n} H\left(\frac{d(x,x_i)}{h_{1,n}}\right), \tag{2}$$

where $h_{1,n}$ and $h_{2,n}$ are non-random positive sequences. This kernel estimator was defined for instance in Ferraty and Vieu (2006). The function $H(.)$ is an asymmetrical kernel with support in [0,1] and the function $K(.)$ is defined by

$$K(v) = \int_v^{\infty} g(s)\, ds,$$

where $g(.)$ is a bounded probability function with support included in $[-1,1]$. Our class of conditional tail index estimators is given by

$$\widehat{\gamma}(u_{n,x},x) = \int_0^{u_{n,x}} \Psi(\alpha,u_{n,x},x) \log \widehat{q}_n(\alpha,x)\, d\alpha, \tag{3}$$

where for all $u \in (0,1)$, $\Psi(.,u,x)$ is a non-null continuous function in $L_1(0,1)$ such that $\int_0^u \Psi(\alpha,u,x)\, d\alpha = 0$ and $u_{n,x}$ is a positive sequence. The sequence $u_{n,x}$ is introduced in order to select only the largest observation to estimate $\gamma(x)$. Its choice is thus very important in practice. Note that the estimator $\widehat{\gamma}(u_{n,x},x)$ depends only on the Y_i's for which the corresponding x_i's belong to the ball $B(x,h_{1,n}) = \{t \in E;\ d(t,x) \leq h_{1,n}\}$. The smoothness of the function $\widehat{\overline{F}}_n(.,x)$ is also controlled by the bandwidth $h_{2,n}$. Finally, note that $\widehat{\overline{F}}_n(y,x) \sim \overline{\mathbb{F}}_n(y,x)$ as $h_{2,n} \to 0$, where

$$\overline{\mathbb{F}}_n(y,x) = \sum_{i=1}^{n} H\left(\frac{d(x,x_i)}{h_{1,n}}\right) \mathbb{1}_{\{Y_i > y\}} \Bigg/ \sum_{i=1}^{n} H\left(\frac{d(x,x_i)}{h_{1,n}}\right)$$

is the empirical estimator of the conditional survival function. The same result is established for the associated conditional quantile in Lemma 1.

3. Main result

In this section, we give the useful notations and assumptions in order to establish the asymptotic normality of our estimators.

(A.1) The slowly varying function $\ell(.,x)$ is normalized.

Assumption (A.1) is equivalent to suppose that for all $y > 1$,

$$\ell(y,x) = c(x) \exp \int_1^y \frac{\Delta(v,x)}{v} dv,$$

where $c(x) > 0$ and $\Delta(v,x) \to 0$ as $v \to \infty$.

(A.2) The function $|\Delta(.,x)|$ is ultimately decreasing.

The largest oscillation of the log-quantile function with respect to its second variable is defined for all $a \in (0,1/2)$ as

$$\omega_n(a) = \sup\left\{ \left|\log \frac{q(\alpha,x)}{q(\alpha,x')}\right|,\ \alpha \in (a,1-a),\ (x,x') \in B(x,h_{1,n}) \times B(x,h_{1,n}) \right\}.$$