Statistics and Probability Letters 84 (2014) 149–157

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Empirical likelihood based confidence intervals for the tail index when $\gamma < -1/2$

Haoze Sun [a,*], Yuexiang Jiang [b]

[a] *Department of Mathematics, Zhejiang University, Hangzhou, 310027, China*
[b] *College of Economics, Zhejiang University, Hangzhou, 310027, China*

ARTICLE INFO

Article history:
Received 7 June 2013
Received in revised form 26 September 2013
Accepted 3 October 2013
Available online 11 October 2013

Keywords:
Negative tail index
Empirical likelihood
Confidence interval

ABSTRACT

Empirical mortality data reveals that there is a finite age limit in the life span of humans, which means that it has a negative tail index. So far, there is a little literature on the confidence intervals for the tail index, especially for the negative tail index. In this paper, we construct its empirical likelihood based confidence intervals when $\gamma < -1/2$, which is known as the irregular case and derive the asymptotic $\chi^2(1)$ distribution. At last a limited simulation study is conducted, which indicates that our method is better than the normal approximation in the sense of coverage probability and less sensitive to the selection of k.

1. Introduction

Assume that $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables with a distribution function F, satisfying

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right),$$

for all $x > 0$ with $1+\gamma x > 0$, with real constants $a_n > 0$ and b_n, $\gamma \in \mathbb{R}$, and for $\gamma = 0$ the right-hand side is interpreted as $\exp(-e^{-x})$. The distribution function $G_\gamma(x)$ is called the generalized extreme value distribution and γ is called the extreme value index or tail index; see De Haan and Ferreira (2006).

Nowadays, there is increasing evidence suggesting that the financial return data is nonnormally distributed and its distribution tends to have heavy-tailed phenomena ($\gamma > 0$). In order to make some statistical inferences on a population with a heavy-tailed distribution, for instance, estimating its high quantiles or tail probability, one has to estimate its tail fatness, which is characterized by the tail index. So far, there are various estimators proposed for the tail index. Most literature is focused on the positive tail index. The existing estimating methods may be divided into following categories: based on extreme order statistics (Hill, 1975), the moment method (Dekkers et al., 1989), the maximum likelihood method (Smith, 1987), the smooth functional method (Drees, 1998), kernel estimators (Csörgö et al., 1995), least-squares estimators (Csörgö and Viharos, 1997) and the regression method (Beirlant et al., 1999; Huisman et al., 2001).

A distribution with a negative tail index has a finite endpoint. For example, empirical mortality data reveals that there is a finite age limit in the life span of humans and its distribution has a negative tail index (Aarssen and de Haan, 1994). Estimation of the endpoint usually needs to estimate the tail index first. Thus, estimation of the negative tail index is also of practical importance. In the literature, estimating the negative tail index may be divided into two cases: $-1/2 \le \gamma < 0$

* Corresponding author.
E-mail address: sunhaoze@zju.edu.cn (H. Sun).

0167-7152/$ – see front matter
http://dx.doi.org/10.1016/j.spl.2013.10.001

and $\gamma < -1/2$. For $-1/2 \leq \gamma < 0$, Smith (1987) obtained the limit of the above maximum likelihood estimator (MLE), so one can easily construct the confidence intervals based on the MLE; Drees et al. (2004) reconsidered the MLE by taking the threshold as the $(k+1)$-th largest order statistic and obtained its limit by employing a weighted approximation of tail quantile processes when $\gamma > -1/2$. For $\gamma < -1/2$, Smith (1987) considered a different estimator from MLE for the case $(-1, -1/2)$; Peng and Qi (2009) derived the limit and the rate of convergence of MLE when the tail index is in $(-1, -1/2]$ by adding an additional inequality constraint. However, there is a little literature on $\gamma < -1/2$ specifically, which we are concerned in this paper. Falk (1995) proposed his uniformly minimum variance unbiased (UMVU) estimator for $\gamma < -1/2$ and proved its asymptotic normality. Of course, it is a natural way to construct confidence intervals based on the normal approximation. But, this estimator is based on the upper k order statistics. When k is small, the variance of the estimator is large, and a large k results in a large bias. So, the selection of k is very important.

The empirical likelihood (EL) method is a powerful nonparametric method for constructing confidence intervals and testing, which was first proposed by Owen (1988, 1990, 1991). Based on the EL method, constructing confidence intervals for a heavy-tailed distribution becomes popular recently; see Lu and Peng (2002), Peng and Qi (2006), Worms and Worms (2011), Li et al. (2011) and Chan et al. (2012). The advantages of EL method are that we do not need to estimate the asymptotic variance explicitly, for the shape of confidence region is determined by data and the confidence regions are Bartlett correctable in some regular cases. And in finite samples, the EL method is also very accurate.

In this paper, we apply the EL method to construct the confidence intervals for $\gamma < -1/2$ based on Falk's UMVU estimator, which supplements the results by Li et al. (2011). And our statistical methodology is from Lu and Peng (2002). We have proved that the log empirical likelihood ratio function is asymptotic $\chi^2(1)$ distributed and the simulation study indicates that the EL method is better than the normal approximation proposed by Falk (1995) in the sense of coverage probability. Moreover, we find that the EL method is less sensitive to the selection of k.

This paper is organized as follows. In Section 2, we present our methodology and main results. Then, in Section 3, a simulation study is conducted, which compares the two different methods for constructing confidence intervals for the tail index, in the sense of coverage probability and finally, Appendix gives the proofs of our results.

2. Methodology and main results

Here, we focus on the estimation of the tail index under a special class of distributions. Assume that $X_1, \ldots, X_n$ is a sequence of i.i.d. random variables, with the distribution function F, which possesses a density f in a left neighborhood of θ, satisfying

$$f(x) = -\exp(-b/\gamma)/\gamma(\theta - x)^{-1/\gamma-1}(1 + O((\theta - x)^{-\delta/\gamma})), \tag{2.1}$$

as $x \to \theta < \infty$ for some $\delta > 0$, $\gamma < 0$ and $b \in \mathbb{R}$, where θ is the right endpoint of F ($\theta := \sup\{x \in \mathbb{R} : F(x) < 1\} < \infty$). The estimation of the tail index γ with this class of distribution functions was discussed in Falk (1995).

Let $\widetilde{X}_i := \frac{1}{\theta - X_i}$, then $\widetilde{X}_i \in RV_{1/\gamma}$, $i = 1, \ldots, n$, which means the survival function of $\widetilde{X}_i$ is regular varying at infinity with index $1/\gamma$, i.e.,

$$\lim_{t \to +\infty} \frac{1 - \widetilde{F}(tx)}{1 - \widetilde{F}(t)} = x^{1/\gamma},$$

for all $x > 0$, where $\widetilde{F}$ is the distribution function of $\widetilde{X}_i$. According to Hill's estimator, we have

$$-\widetilde{\gamma}_H = \frac{1}{k}\sum_{i=1}^{k} \log \widetilde{X}_{n,n-i+1} - \log \widetilde{X}_{n,n-k}, \tag{2.2}$$

where $\widetilde{X}_{n,1} \leq \cdots \leq \widetilde{X}_{n,n}$ are the order statistics, (2.2) can be written as

$$\widetilde{\gamma}_H = \frac{1}{k}\sum_{i=1}^{k} \log(\theta - X_{n,n-i+1}) - \log(\theta - X_{n,n-k}). \tag{2.3}$$

Since the endpoint θ is unknown, Falk (1995) replaced θ with the maximal order statistic $X_{n,n}$. When $\gamma < -1/2$, it is a good approximation of θ (Aarssen and de Haan, 1994); thus, $\widetilde{\gamma}_H$ is to be

$$\widetilde{\gamma}_F = \frac{1}{k-1}\sum_{i=2}^{k} \log(X_{n,n} - X_{n,n-i+1}) - \log(X_{n,n} - X_{n,n-k}). \tag{2.4}$$

In order to obtain its limiting distribution, k should satisfy

$$k \to \infty, \qquad k/n \to 0, \qquad \sqrt{k}(k/n)^{\delta} \quad \text{and} \quad \log n/k^{1/2} \to 0,$$

as $n \to \infty$, then

$$\sqrt{k}(\widetilde{\gamma}_F - \gamma) \xrightarrow{d} N(0, \gamma^2)$$