Computational Statistics and Data Analysis 98 (2016) 91–104

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Estimating extreme tail risk measures with generalized Pareto distribution

Myung Hyun Park, Joseph H.T. Kim *

Department of Applied Statistics, College of Business and Economics, Yonsei University, Seoul, 120-749, Republic of Korea

HIGHLIGHTS

- A new GPD parameter estimator is proposed.
- It is based on a nonlinear weighted least squares method.
- Under the POT framework, we estimate tail risk measures. Extensive simulation studies show the new method works well.

ARTICLE INFO

Article history:
Received 18 May 2015
Received in revised form 16 November 2015
Accepted 15 December 2015
Available online 4 January 2016

Keywords:
Generalized Pareto distribution
Value-at-Risk (VaR)
Conditional Tail Expectation (CTE)
Tail VaR
Peaks over threshold (POT)
Weighted nonlinear least squares

ABSTRACT

The generalized Pareto distribution (GPD) has been widely used in modelling heavy tail phenomena in many applications. The standard practice is to fit the tail region of the dataset to the GPD separately, a framework known as the peaks-over-threshold (POT) in the extreme value literature. In this paper we propose a new GPD parameter estimator, under the POT framework, to estimate common tail risk measures, the Value-at-Risk (VaR) and Conditional Tail Expectation (also known as Tail-VaR) for heavy-tailed losses. The proposed estimator is based on a nonlinear weighted least squares method that minimizes the sum of squared deviations between the empirical distribution function and the theoretical GPD for the data exceeding the tail threshold. The proposed method properly addresses a caveat of a similar estimator previously advocated, and further improves the performance by introducing appropriate weights in the optimization procedure. Using various simulation studies and a realistic heavy-tailed model, we compare alternative estimators and show that the new estimator is highly competitive, especially when the tail risk measures are concerned with extreme confidence levels.

1. Introduction

The last few decades have witnessed an unprecedented increase in the size of datasets available, a phenomenon of massive data, in various applications such as finance, insurance, computer science and communications. Such large datasets now allow various quantitative risk analyses which were once thought infeasible to implement, including the investigation on rare but huge loss events occurring in the tail of the distribution. For instance, for financial institutions, extreme quantiles of the loss distribution are of great interest for both internal and regulatory purposes. Accurate estimation of such tail-related quantities however generally requires generating considerably large loss samples (see Section 5 for details), which may be very time-consuming for large complicated financial portfolios, and calculating those quantities is also known to be

* Corresponding author.
E-mail address: jhtkim@yonsei.ac.kr (J.H.T. Kim).

http://dx.doi.org/10.1016/j.csda.2015.12.008
0167-9473/

expensive because of the computing time and memory storage, as observed in the computer science literature, e.g., Liechty et al. (2003), Chen et al. (2000) and Munro and Paterson (1980); see Song and Song (2012) for a more detailed account on the issue of estimating extreme quantiles from massive datasets.

In view of these difficulties, Extreme value theory (EVT) has received much attention as a modern tool to study tail quantities of the distribution, including the extreme quantiles. In the centre of EVT framework, the generalized Pareto distribution (GPD) emerged as the distribution of the exceedances above a sufficiently high threshold for arbitrary heavy-tailed loss data (Pickands, 1975). The standard procedure of fitting heavy-tailed data calls for a separate modelling of the tail region of the dataset using the GPD, a procedure commonly known as the peaks over threshold (POT); see, e.g., Embrechts et al. (1997). In finance and insurance applications, a main purpose of applying EVT under the POT is to determine common tail risk measures such as the Value-at-Risk (VaR) or Conditional Tail Expectation[1] (a.k.a. Tail-VaR) from the fitted GPD. However, residing in the tail region, these risk measure estimates are highly sensitive to the estimated GPD parameters, and their volatility becomes larger as the required quantile level gets extreme. For example, according to Basel II (BCBS, 2006), the operational risk capital a bank should hold must cover unexpected losses with at least a 99.9% probability under the Advanced Measurement Approach (AMA), equivalent to VaR at a confidence level of 99.9%. Similarly, the credit risk capital under the Internal Ratings Based (IRB) approach also requires VaR 99.9%; furthermore, VaR 99.99% is used in the regulator's backtesting analysis, assessing rare events occurring with probability 0.0001. Apparently, such extreme quantiles are highly sensitive to the estimation methods and small differences in the estimates could lead to a considerable impact on the financial position of banks, underscoring the importance of accurate GPD estimation.

Estimating the GPD parameters is a long-standing problem and various approaches have been investigated in the literature. For example, the traditional maximum likelihood estimation (MLE) is discussed in Grimshaw (1993), Davison (1984) and Smith (1985). Pickands (1975) proposed a method using the sample order statistics; Hosking and Wallis (1987) used the method of moment (MME) and probability weighted moments (PWM). A generalized version of the MME was discussed by Ashkar and Ouarda (1996) and generalized probability weighted moments (GPWM) was proposed by Rasmussen (2001); Dupuis and Tsao (1998) developed a hybrid PWM. Juárez and Schucany (2004) introduced a minimum density power divergence method, and Zhang (2007) proposed a likelihood moment estimator (LME). Zhang and Stephens (2009) and Zhang (2010) surveyed some relevant contributions to the literature of the GPD parameter estimation by means of the Bayesian methodology. The reader is also referred to de Zea Bermudez and Kotz (2010) for a survey of various GPD estimators. More recently, Song and Song (2012) introduced a new – yet computationally simple and fast – GPD parameter estimator for large samples based on a nonlinear least square method that minimizes the sum of squared deviations between the empirical distribution function (EDF) of the sample and the theoretical GPD, which was reported to outperform other existing methods they considered.

In this paper, we propose a new GPD parameter estimator under the POT framework to estimate tail risk measures at extreme quantiles. Our method is adapted from that of Song and Song (2012) and uses a nonlinear least squares method that minimizes the sum of squared deviations between the empirical distribution function and the theoretical GPD for the data exceeding the tail threshold. However, the proposed method uses a different object function and is better in its performance, and these are our contributions in the current paper. In particular, we first examine the method of Song and Song (2012) and point out its caveat, to show that their method is only applicable for the case where the tail of the loss sample is GPD distributed unconditionally. We address this issue and present a revised procedure. Second, in order to further improve the estimation, we introduce suitable weights in the revised optimization procedure using the weighted regression setup. Using the proposed procedure we estimate the VaR and CTE, the two popular tail risk measures, at extreme quantile levels. Using various simulation studies and a realistic heavy-tailed model, we compare alternative estimators and show that the performance of the proposed estimator is highly competitive compared to other existing estimators, especially for the risk measures with extreme confidence levels.

This paper is organized as follows. The POT approach in EVT is briefly reviewed in Section 2. In Section 3, after reviewing existing methods, we describe a new GPD parameter estimator under the POT. Section 4 is devoted to numerical exercises, where we compare the performances of different methods in estimating tail risk measures under various heavy-tailed common parametric distributions. In Section 5 a more realistic loss model is considered for a similar exercise. Section 6 concludes the paper.

2. EVT for extreme tail risk measures

2.1. Peaks over threshold

We start with a brief account for Extreme Value Theory (EVT) focused on the peaks over Threshold (POT) framework; comprehensive treatments of book length on EVT can be found in, e.g., Embrechts et al. (1997) and Beirlant et al. (2006). Let us denote the tail or survival function of a continuous random variable X by $\bar{F}(x) = 1 - F(x)$, $0 < x < \infty$. Then we say that

[1] In the literature, this is also known as the Conditional Value-at-Risk (CVaR) or Expected Shortfall (ES). The CTE will be formally defined later; see, e.g., McNeil et al. (2005) for a comprehensive discussion on risk measures.