Journal of Multivariate Analysis 127 (2014) 19–35

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Self-consistent estimation of conditional multivariate extreme value distributions

Y. Liu [a,*], J.A. Tawn [b]

[a] JBA Risk Management Limited, South Barn, Broughton Hall, Skipton BD23 1QU, UK

[b] Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK

ARTICLE INFO

Article history:
Received 8 August 2012
Available online 17 February 2014

AMS subject classification:
62G32

Keywords:
Coefficient of tail dependence
Conditional analysis
The Heffernan–Tawn model
Multivariate extreme value theory
Quantile regression
Self-consistency

ABSTRACT

Analysing the extremes of multi-dimensional data is a difficult task for many reasons, e.g. the wide range of extremal dependence structures and the scarcity of the data. Some popular approaches that account for various extremal dependence types are based on asymptotically motivated models so that there is a probabilistic underpinning basis for extrapolating beyond observed levels. Among these efforts, Heffernan and Tawn developed a methodology for modelling the distribution of a d-dimensional variable when at least one of its components is extreme. Their approach is based on a series ($i = 1, \ldots, d$) of conditional distributions, in which the distribution of the rest of the vector is modelled given that the ith component is large. This model captures a wide range of dependence structures and is applicable to cases of large d. However their model suffers from a lack of self-consistency between these conditional distributions and so does not uniquely determine probabilities when more than one component is large. This paper looks at these unsolved issues and makes proposals which aim to improve the efficiency of the Heffernan–Tawn model in practice. Tests based on simulated and financial data suggest that the proposed estimation method increases the self-consistency and reduces the RMSE of the estimated coefficient of tail dependence.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Extreme value theory is a collection of theory and methodology for studying rare events characterized by small probabilities of occurrence. Its application ranges from environmental risk assessment, financial risk management, industrial safety regulation, to pharmaceutical studies. Multivariate extreme value theory is a generalization of the univariate case where multi-dimensional data are involved. Pioneering research efforts in this area include de Haan and Resnick [5], Pickands [17], see [19,9] for reviews. A popular method for studying multivariate extremes is to use threshold-based models, see for example [3,4,11,21]. These models however make implicit assumptions about the extremal dependence between components, a property termed asymptotic dependence which implies that there is a positive probability that the large values for each component can occur simultaneously.

More recently theory has been developed based on hidden regular variation, summarized in [20], which overcomes this problem of implicit assumptions, see for example [14,15,6,18,8]. But their focus was on the region of the joint tail where all components are extreme simultaneously. Later Heffernan and Tawn [10] proposed a conditional approach for modelling

* Corresponding author.
E-mail addresses: liuye1999@gmail.com, ye.liu@jbarisk.com (Y. Liu), j.tawn@lancaster.ac.uk (J.A. Tawn).

http://dx.doi.org/10.1016/j.jmva.2014.02.003
0047-259X/© 2014 Elsevier Inc. All rights reserved.

multivariate extremes where an extreme event is such that at least one component is large. Their model breaks the limitation to the joint tail area, captures a much broader range of dependence structures than had previously been studied, and provides a parsimonious model in higher dimensional problems. This paper extends the Heffernan–Tawn model and is motivated by some unresolved issues within the original Heffernan and Tawn [10] paper, namely the self-consistency issue and the form of the residual distribution.

Like in the original Heffernan and Tawn [10] paper, we assume the marginal distribution of the multivariate random variable is dealt with separately from the dependence structure. However we replace the standard Gumbel marginal distribution with the standard Laplace distribution as suggested in [13] in order to allow the flexibility for parsimonious modelling of negative dependence. In a bivariate case of the Heffernan–Tawn model, if Y_j and Y_i are margins of variables with standard Laplace distribution, then the pairwise dependence is modelled by assuming that the following limiting distribution holds,

$$\lim_{u\to\infty} \mathbb{P}\left(\frac{Y_j - \alpha_{j|i} Y_i}{Y_i^{\beta_{j|i}}} < z,\ Y_i - u > y \,\middle|\, Y_i > u \right) = G_{j|i}(z)\exp(-y), \tag{1.1}$$

for all $y > 0$, where $-1 \le \alpha_{j|i} \le 1$, $\beta_{j|i} \le 1$ are selected so that $G_{j|i}$ is a non-degenerate distribution. The original Heffernan–Tawn model assumes the limiting condition in Eq. (1.1) still holds when u is large but finite, but does not conclude a parametric form for the limiting distribution $G_{j|i}$. So an aim of this paper is to propose a general class of distribution $G_{j|i}$ which may be used for statistical inferences.

In the original paper the conditional structure of (1.1) is imposed on both directions ($Y_j|Y_i$ and $Y_i|Y_j$) separately, i.e. given $\alpha_{j|i}$ currently no constraints are imposed on $\alpha_{i|j}$. A problem arising naturally from this approach is which margin do we choose to condition on and if not selecting one form what additional structure is required on the model to impose self-consistency. The fundamental issue is the lack of self-consistent results when a statistic is estimated conditioning on different margins. For example when we consider the joint tail probability of Y_i and Y_j both being larger than a big value v, we have the following identities

$$\mathbb{P}\left(Y_i > v, Y_j > v\right) \equiv \mathbb{P}\left(Y_j > v|Y_i > v\right)\mathbb{P}\left(Y_i > v\right) \equiv \mathbb{P}\left(Y_i > v|Y_j > v\right)\mathbb{P}\left(Y_j > v\right).$$

However when estimating the two conditional probabilities using the original Heffernan–Tawn model separately, it is not guaranteed that the probabilities will be equal and thus we have different estimates for the joint survival probability.

To illustrate this problem, we consider data on daily returns of a pair of stock market indices, S&P 500 in the US and Nikkei 225 in Japan, over the period from January 2002 to March 2010. Let X_1 and X_2 denote the two returns respectively, we first remove the volatility effect by fitting a univariate GARCH model to either margin and taking the standardized returns $\tilde{X}_1$ and $\tilde{X}_2$ (see Appendix A for more details). Then we perform a probability integral transform[1] on $\tilde{X}_1, \tilde{X}_2$ and transform them into standard Laplace distributed random variables Y_1 and Y_2. The dashed lines in Fig. 1 are the chosen threshold $Y_1 = u$ and $Y_2 = u$ with u equal to the 90-th quantile of the standard Laplace distribution. We are interested in events with both margins exceeding the threshold u, i.e. $Y_1 > u$ and $Y_2 > u$, which are represented by the darker dots in Fig. 1. Since this region lies above (and to the right of) both threshold lines, inference can be made about this area using the standard Heffernan–Tawn model, by conditioning on $Y_1 > u$ or on $Y_2 > u$. For the model to be self-consistent the probability density for events in the joint tail must be identical regardless of the choice of conditioning margin. This therefore defines some implicit constraints on the parameters $(\alpha_{j|i}, \beta_{j|i})$, the conditional distribution of $Y_j|Y_i > u$ as well as their respective counterparts, i.e. with i and j interchanged.

We generalize the threshold choice u to the p-th quantile $Q_L(p)$ of the standard Laplace distribution and define the conditional tail probability $\chi(p) = \mathbb{P}(Y_1 > Q_L(p), Y_2 > Q_L(p))/(1-p)$. The self-consistency in this situation is therefore represented by

$$\chi(p) \equiv \mathbb{P}(Y_1 > Q_L(p)|Y_2 > Q_L(p)) \equiv \mathbb{P}(Y_2 > Q_L(p)|Y_1 > Q_L(p)) \quad \forall p \in (0,1).$$

The conditional tail probability $\chi(p)$ can be estimated over a range of quantiles p by fitting the Heffernan–Tawn model to (Y_1, Y_2) conditioning on either margin. Fig. 2 shows that the conditional tail probability $\chi(p) \to 0$ as $p \to 1^-$, which indicates that Y_1 and Y_2 are asymptotically independent, see [14]. Nonetheless, there is a clear difference between the estimation conditional on either Y_1 or Y_2 being large. In particular as we extrapolate further towards the upper tail of the marginal distribution, the difference widens. This problem of self-consistency is given a brief discussion by the authors in the original paper and we explore more about this issue here.

The rest of the paper is structured as follows. Section 2 introduces the Heffernan–Tawn model and its application in more details; Section 3 focuses on the conditions of self-consistency and its implication for the residual distribution; Section 4 proposes a self-consistent non-parametric estimation method as well as a diagnostic method; Section 5 includes some examples based on both simulated and real data and compares the performance of the new estimation method with the existing method; Section 6 concludes the main findings and discusses briefly the areas for future work.

[1] Without making any distributional assumption on $\tilde{X}_1$ and $\tilde{X}_2$, we model the body of the distribution by standard Gaussian kernel density and the upper and lower tail by generalized Pareto distribution.

Download English Version:

https://daneshyari.com/en/article/1145567

Download Persian Version:

https://daneshyari.com/article/1145567

Daneshyari.com