



School of Social Sciences
CITY UNIVERSITY LONDON

Department of Economics

**Extreme Value Theory Filtering Techniques for Outlier
Detection**

Jose Olmo¹
City University London

**Department of Economics
Discussion Paper Series
No. 09/09**

¹ Dept. Economics, City University London. Northampton Square, EC1 V0HB, London. E-mail: j.olmo@city.ac.uk

Extreme Value Theory Filtering Techniques for Outlier Detection

JOSE OLMO*, *City University London*.

June 2009

Abstract

We introduce asymptotic parameter-free hypothesis tests based on extreme value theory to detect outlying observations in finite samples. Our tests have nontrivial power for detecting outliers for general forms of the parent distribution and can be implemented when this is unknown and needs to be estimated. Using these techniques this article also develops an algorithm to uncover outliers masked by the presence of influential observations.

Keywords: Extreme value theory, Hypothesis tests, Outlier detection, Power function, Robust estimation.

*Corresponding Address: Dept. Economics, City University London. Northampton Square, EC1 V0HB, London. Jose Olmo, E-mail: j.olmo@city.ac.uk. The author thanks Peter Burrige for many useful discussions that helped to improve the focus of the article. He is also grateful to Angeles Carnero, Andrew Harvey and Robert Taylor for suggestions and comments on previous versions of the article. Any remaining errors are the author's own.

1 Introduction

All types of statistical data may contain particular observations that are in some way inconsistent with the rest of the sample. When such observations lie outside the range of the remaining data they are known as outliers. In general the presence of these aberrant observations has an impact on subsequent analyses. In time series applications, the strength of such effects depends on the type of time series, type(s) of outlier(s), its (their) location and the distribution of the rest of the data, as well as on the purpose of the analysis. Thus outlier detection is very important in applied econometrics and in empirical finance, as well as challenging.

When data are expected to be sampled independently and to be identically distributed (IID), outlying values are usually defined in terms of standard deviations apart from the mean. In this way Verhoeven and McAleer (2000) define extremes as observations between two and three standard deviations away from the mean, and outliers as those observations more than three standard deviations away. Basmann (2003) discusses the use of quantile-based definitions, which, for the Normal distribution, can be expressed in terms of the standard deviation, but notes that such definitions are inappropriate for non-Normal data, and argues in favour of a more flexible definition based on the tail probability.

The first papers in outlier detection for *IID* data concentrated on outliers in Normal distributions. Some of these important contributions are Grubbs (1950, 1969), Ferguson (1961), or Rosner (1975) focusing on multiple outlier detection. Other more recent papers are Hawkins (1980), or the surveys of Beckman and Cook (1983) and Barnett and Lewis (1994). Also, due to the development in recent years of extreme value theory and literature on stable laws and heavy tails, some articles based on these methodologies have pursued the topic of outlier detection. Some examples are Mittnick, Rachev and Samorodnitsky (2001), Doornik and Ooms (2005) or Schluter and Trede (2008) in risk management.

Alternatively, in fields as labor economics or robust statistics is standard practice to use ad-hoc trimming in the tails. The practice of discarding some fraction of the observations in the tails has the advantage of reduction in parameter estimation bias produced by the influence of outlying observations. Unfortunately, most of the methods introduced in the literature also face efficiency losses due to the trimming of useful information obtained in the tails. Some pioneering works are due to Bickel (1965) and Stigler (1973). More recent studies on robust statistics are

the monographs of Rousseeuw and Leroy (1987) and Maronna, Martin and Yohai (2006).

The contributions of this paper are twofold. First, using a definition of outlier based on extreme value theory we propose a hypothesis test that permits to disentangle the occurrence of outliers from extreme values in finite samples. We divide our analysis in two scenarios; one, in which observations are generated by an exponentially decreasing distribution, and a second case in which observations are drawn from distribution functions with heavy tails. Given that in practice, the knowledge of the parent distribution F or even the rate of tail decay, are not known, our second contribution is to propose a test statistic in this case. Here, we study the effects of outliers in the estimation of the tail index, the parameter determining the rate of tail decay and hence characterizing whether a distribution decays exponentially or polynomially, and propose a filtering technique to clean masking effects derived from the presence of outliers. Both contributions, the hypothesis test and the filtering technique, are combined in an algorithm that permits to detect and remove, iteratively, outliers in finite samples.

The paper is structured as follows. Section 2 introduces a hypothesis test to detect outliers in an *IID* environment, and that makes allowance for parent distributions that decrease exponentially and polynomially. Section 3 studies the statistical power of both versions of the test for F known and also when the distribution function is not known and the tail index needs to be estimated. We also introduce in the section an algorithm that allows to detect and filter, iteratively, the presence of outlying and influential observations. Finally Section 4 concludes and discusses extensions of this research.

2 Outlier Detection Tests Based on EVT

We define outliers as those observations with a negligible probability of belonging to the data generating process. For independent and identically distributed (*IID*) observations the data generating process is a common distribution function F . To make the concept of outlier operational we need some basic extreme value theory (EVT).

Let $M_n^0 = \max(X_1, \dots, X_n)$ be the sample maximum of an *IID* sequence of length n , and x_F be the right end point of F , defined as $x_F = \sup\{x | F(x) < 1\} \leq \infty$. The regularity condition

$$\lim_{x \uparrow x_F} \frac{1 - F(x)}{1 - F(x^-)} = 1, \quad (1)$$

guarantees that there are no jumps in the right tail of the distribution F of the data, see Embrechts, Klüppelberg, and Mikosch (1997) for details.

We will assume hereafter this regularity condition as our minimum set of assumptions on the distribution function F . Under these regularity conditions on the upper tail of F , the finite-sample distribution of the standardized version of M_n^0 satisfies that

$$P\{a_n^{-1}(M_n^0 - b_n) \leq x\} = F^n(a_n x + b_n) \longrightarrow G(x), \quad (2)$$

where a_n and b_n are normalizing sequences associated with F , and $G(x)$ is an extreme value distribution that can be one of only three types, see Fisher and Tippett (1928) and Gnedenko (1943). Further, given that our interest is in detecting outliers, and we identify these observations with elements of the random sample with a very low probability of being generated by F , below we will limit ourselves to cases where $x_F = \infty$. In this framework there are just two possible classes of limiting distributions for the standardized sample maximum:

$$\begin{aligned} \text{Type I: (Gumbel)} \quad & G(x) \equiv \Lambda = \exp(-\exp(-x)), \quad -\infty < x < \infty, \\ \text{Type II: (Fréchet)} \quad & G(x) \equiv \Phi_\xi = \begin{cases} 0 & x \leq 0, \\ \exp(-x^{-\frac{1}{\xi}}) & x > 0, \quad \xi > 0. \end{cases} \end{aligned}$$

The Fréchet class of distributions is governed by the so-called *tail index* ξ , that describes the heaviness of the tails of the parent distribution F if this decays polynomially. More formally,

$$1 - F(x) = x^{-1/\xi} L(x), \quad (3)$$

where $L(x)$ is a slowly varying function, that is, $L(tx)/L(x) \rightarrow 1$ for every $t > 0$ as $x \rightarrow x_F$. On the other hand, if the parent distribution, F , is exponentially decreasing in the upper tail, the limiting distribution of the maximum of a sample of size n properly standardized is Gumbel and $\xi = 0$. These two extreme value distributions along with the Weibull family of extreme value distributions can be gathered in the so-called generalized extreme value distribution:

$$G(x) = \exp\left(-\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}}\right). \quad (4)$$

Finally note that if the distribution of the standardized sample maximum converges to one of these extreme value distributions we will say that F belongs to the maximum domain of attraction of $G(x)$, $F \in MDA(G)$.

2.1 Hypothesis test

The question of interest is to determine if the maximum observation of a random sample of size n can be statistically the sample maximum of an *IID* sequence of same size drawn from F . If this claim is true we will retain the observation in our sample; if, however, it is proven to be wrong we will discard the observation and treat it as an outlier. This technique can be of much interest for improving the estimation of parameters in statistics and regression models. By discarding outlying observations we reduce the bias in the parameter estimates. On the other hand, by retaining the extreme values we improve the efficiency of estimators.

Definition: *Let M_n be the maximum of an IID sample of size n , with common distribution function F for at least $n - 1$ observations. Then, the hypothesis test to determine statistically if M_n is an outlier is the following:*

$$\begin{cases} H_{O,n} : M_n \text{ is an observation of } F, \\ H_{A,n} : M_n \text{ is an outlying observation.} \end{cases} \quad (5)$$

To make this hypothesis test operational we propose the following test statistic

$$T_{n,G} = a_n^{-1}(M_n - b_n), \quad (6)$$

with critical value, G_α , at an α significance level given by the $1 - \alpha$ cumulative quantile of the extreme value distribution $G(x)$, and satisfying

$$\lim_{n \rightarrow \infty} P_{H_0}\{T_{n,G} > G_\alpha\} = \alpha. \quad (7)$$

For illustration purposes we will divide our analysis in an *IID* framework between parent distributions F that are exponentially decreasing and parent distributions that decay to their right end point x_F , polynomially. Thus, for F exponentially decreasing in the upper tail the

corresponding limiting distribution $G(x)$ is a Gumbel extreme value distribution, and the sample maximum is an outlier if $G(T_{n,\Lambda}) > 1 - \alpha$. Note that we use $T_{n,\Lambda}$ to denote the test statistic in this case. The corresponding rejection region at an α significance level is $\{T_{n,\Lambda} > \Lambda_\alpha\}$, with $\Lambda_\alpha := -\ln(-\ln(1 - \alpha))$, obtained from the inverse of the Gumbel extreme value distribution. Note that the hypothesis test is not restricted to detect outliers in a Normal framework as in most of literature on outlier detection, but in the large family of exponentially decreasing distribution functions. Other paper using a similar methodology for outlier detection for exponentially decreasing distributions in time series models is Doornik and Ooms (2005).

For F heavy-tailed distributed the limiting distribution of the standardized maximum is not Gumbel but Frèchet, implying a test statistic $T_{n,G}$ that is no longer parameter-free, in fact the asymptotic critical value at an α significance level is $\Phi_{\xi,\alpha} := \frac{1}{(-\ln(1-\alpha))^\xi}$, with $\xi > 0$. Nevertheless, we can exploit the following relationship between the two extreme value distributions to obtain a test statistic that is parameter-free. Other papers studying alternative outlier detection methods for heavy-tailed distributions are for example Mittnick, Rachev and Samorodnitsky (2001) for stable law processes, or Schluter and Trede (2008) for detecting multiple outliers in financial time series.

Result 1: $Y \stackrel{d}{\sim} \Phi_\xi \Leftrightarrow \ln Y^{\frac{1}{\xi}} \stackrel{d}{\sim} \Lambda$, for Y a random variable, and where $\stackrel{d}{\sim}$ denotes equivalence in distribution, see Embrechts, Klüppelberg and Mikosch (p. 123, 1997).

Under $H_{O,n}$, if F is heavy tailed the transformed version of the test statistic $T_{n,G}$ is

$$T_{n,\Phi_\xi} := \ln (a_n^{-1}(M_n - b_n))^{1/\xi}, \quad (8)$$

and follows a Gumbel extreme value distribution. The advantage of this transformation is that the asymptotic critical value of the relevant test is, as before, Λ_α . Furthermore, in the polynomially decaying case the tail index is $\xi > 0$ and the normalizing sequences are universally defined as $a_n = F^{-1}(1 - \frac{1}{n})$ and $b_n = 0$, see Embrechts, Klüppelberg and Mikosch (p. 130, 1997).

The finite-sample properties of the two versions of this hypothesis test require further study. The next subsection studies the asymptotic power for both tests and Section 3 discusses estimation effects and the presence of influential observations.

2.2 Study of Power function of the Tests

For the study of the asymptotic power function of the outlier test we will start with the exponential scenario. Here, M_n denotes the sample maximum from a serially independent sequence of random variables and M_n^0 the maximum of a sample of same size with all observations following a common distribution function F . Let $\eta_{n,\Lambda} := M_n - M_n^0$; the relevant test statistic $T_{n,\Lambda}$ can be decomposed as

$$T_{n,\Lambda} = a_n^{-1}(M_n^0 - b_n) + a_n^{-1}\eta_{n,\Lambda} = T_{n,\Lambda}^O + a_n^{-1}\eta_{n,\Lambda}, \quad (9)$$

with $T_{n,\Lambda}^O$ the test statistic under $H_{O,n}$.

Although every distribution function F with tail exponentially decaying belongs to the maximum domain of attraction of the Gumbel extreme value distribution, the asymptotic convergence of the test statistic is characterized by the normalizing sequences a_n and b_n that are idiosyncratic to F . As a result, the power function of the test statistic $T_{n,\Lambda}$ will depend on the rate of convergence a_n , and on $\eta_{n,\Lambda}$.

Proposition 1: *Let F be a distribution function exponentially decaying in its tail, and satisfying the above regularity condition (1). Then, the hypothesis test (5) is consistent under the alternative hypothesis $H_{A,n}$ if and only if $a_n = o(\eta_{n,\Lambda})$.*

Proof. Consider a significance level $0 < \alpha < 1/2$. The proof immediately follows from observing that

$$\lim_{n \rightarrow \infty} P\{T_{n,\Lambda} > \Lambda_\alpha\} = 1 - \exp(-\exp(-(\Lambda_\alpha - a_n^{-1}\eta_{n,\Lambda}))) + o_P(1),$$

and

$$\exp(-\exp(-(\Lambda_\alpha - a_n^{-1}\eta_{n,\Lambda}))) = (1 - \alpha)^{\exp(a_n^{-1}\eta_{n,\Lambda})} \longrightarrow 0, \quad \text{as } n \rightarrow \infty, \quad (10)$$

if and only if $a_n = o(\eta_{n,\Lambda})$. □

Note from this result that the power of the test for distribution functions F in which $a_n = o(1)$ (Normal, Lognormal) is greater than for distribution functions with $a_n = O(1)$, as $\text{Exp}(\lambda)$ or $\text{Gamma}(a, b)$, see Embrechts, Klüppelberg and Mikosch (pp. 153-157, 1997) for rates of convergence of different exponentially decreasing distributions.

Consider now the heavy tailed case described by a parent distribution F belonging to the

maximum domain of attraction of a Frèchet distribution. In this scenario, and for ease of calculus, we define the ratio $\eta_{n,\Phi_\xi} := \frac{M_n}{M_n^0} - 1$, and decompose the relevant test statistic as

$$T_{n,\Phi_\xi} = \frac{1}{\xi} \ln M_n^0 + \frac{1}{\xi} \ln(1 + \eta_{n,\Phi_\xi}) - \frac{1}{\xi} \ln F^{-1}\left(1 - \frac{1}{n}\right). \quad (11)$$

This statistic can be expressed in terms of the test statistic under $H_{O,n}$, denoted hereafter T_{n,Φ_ξ}^O , as

$$T_{n,\Phi_\xi} = T_{n,\Phi_\xi}^O + \frac{1}{\xi} \ln(1 + \eta_{n,\Phi_\xi}). \quad (12)$$

Proposition 2: *Let F be a distribution function polynomially decaying in its tail, and satisfying the above regularity condition (1). Then, the hypothesis test (5) is consistent under the alternative hypothesis $H_{A,n}$ if and only if $\eta_{n,\Phi_\xi}^{-1} = o(1)$.*

Proof. Consider a significance level $0 < \alpha < 1/2$. The proof follows immediately from observing that

$$\lim_{n \rightarrow \infty} P\{T_{n,\Phi_\xi} > \Lambda_\alpha\} = \lim_{n \rightarrow \infty} P\{T_{n,\Phi_\xi}^O > \Lambda_\alpha - \frac{1}{\xi} \ln(1 + \eta_{n,\Phi_\xi})\}.$$

Now note that

$$\lim_{n \rightarrow \infty} P\{T_{n,\Phi_\xi}^O > \Lambda_\alpha - \frac{1}{\xi} \ln(1 + \eta_{n,\Phi_\xi})\} = \left(1 - (1 - \alpha)^{(1 + \eta_{n,\Phi_\xi})^{1/\xi}}\right) + o_P(1),$$

where $(1 - \alpha)^{(1 + \eta_{n,\Phi_\xi})^{1/\xi}} \rightarrow 0$ as $n \rightarrow \infty$, if and only if $\eta_{n,\Phi_\xi}^{-1} = o(1)$, and given that $\xi > 0$. \square

In contrast to the exponential case the power function for heavy tailed distributions depends on an extra parameter gauging the speed of decay in the right tail of the parent distribution. In particular, the heavier the tail of F less power of the test against outlying observations. This finding is consistent with our definition of outliers, since heavier tails can produce observations of bigger magnitude belonging to F and that can be confounded with outliers, therefore diminishing the power of the test to reject the latter observations as exogenous to the distribution of the data.

As a byproduct of these propositions we can determine the minimum sample size n necessary for the outlier candidate not to be statistically rejected. We only elaborate on the case of F polynomially decaying. For exponentially decaying distributions this result depends on the form of the distribution function F and on the normalizing sequences a_n and b_n .

Corollary 1: *Let M_n be a possible outlier for a sample of size n and generated from a distribution function F polynomially decaying characterized by a tail index ξ . Then, the minimum sample n^* necessary for M_n not to be an outlier of F , at an α significance level, is*

$$n^* = \frac{1}{1 - F(M_n \exp(-\xi \Lambda_\alpha))}, \quad (13)$$

with Λ_α the corresponding asymptotic critical value.

Proof. The critical point that characterizes an outlier, at an α significance level, for F heavy-tailed is

$$\frac{1}{\xi} \ln M_n - \frac{1}{\xi} \ln F^{-1}\left(1 - \frac{1}{n}\right) = \Lambda_\alpha.$$

After some algebra we have that

$$F^{-1}\left(1 - \frac{1}{n}\right) = \exp(\ln M_n - \xi \Lambda_\alpha),$$

and by applying F to both terms and further algebra the result follows. \square

The following section extends this study to the situation where F is not known and we need to estimate the tail index in order to find out the rate of decay of the parent distribution.

3 Estimation effects on the Hypothesis Tests

The hypothesis test introduced in the previous section is very convenient and shows strong power for different parent distribution functions F . This test depends on the knowledge of the normalizing sequences a_n and b_n that unfortunately are not usually known, and have to be estimated. In particular for the exponentially decaying case this has to be done on a case-by-case basis. Thus, whereas for the Normal distribution there are no nuisance parameters to be estimated since $a_n = (2 \ln n)^{-1/2}$ and $b_n = a_n^{-1} - \frac{\ln(4\pi) + \ln \ln n}{2(2 \ln n)^{1/2}}$, see Embrechts, Klüppelberg and Mikosch (p. 156, 1997); for an exponential distribution as $F(x) = 1 - K \exp^{-\lambda x}$, the relevant sequences are $a_n = \lambda^{-1}$ and $b_n = \lambda^{-1} \ln(Kn)$ that can be estimated using sample estimates of K and λ .

For the heavy-tailed case, given that $a_n = F^{-1}\left(1 - \frac{1}{n}\right)$ and $b_n = 0$, the estimation of these normalizing sequences boils down to estimating the inverse of the parent distribution F . There are nonparametric as well as parametric techniques in the literature to estimate F . Whereas the

former techniques are appealing by their generality and, in contrast to parametric methods lack of model risk, these methods are usually very intensive in the use of data, their convergence towards F is very slow and are rather inaccurate in the tails of the distribution. We choose instead a semi-parametric estimator of the parent distribution based on the following decomposition of F given by the conditional probability theorem;

$$F(x) = P\{X \leq x\} = P\{X \leq x | X \leq u\}P\{X \leq u\} + P\{X \leq x | X > u\}P\{X > u\}, \quad (14)$$

for every x in the domain of F and with u a threshold value.

We are concerned, in particular, with estimating the $(1 - \frac{1}{n})$ -quantile of the distribution function F . Further, as n increases, one can assume with no loss of generality that $(x = a_n \Rightarrow) F^{-1}(1 - \frac{1}{n}) > u$ and hence obtain the following equation

$$\frac{1}{n} = P\{X > a_n | X > u\}P\{X > u\}. \quad (15)$$

From here the quantile of interest a_n is

$$a_n = \inf_{x \in [u, \infty)} \left\{ x \mid P\{X > x \mid X > u\} \geq \frac{1}{n(1 - F(u))} \right\}. \quad (16)$$

A natural semi-parametric estimator for this sequence is obtained by approximating the conditional distribution function by a distribution of Pareto type (Pareto, Generalized Pareto) as proposed by Pickands (1975) and Balkema-de Haan (1974), and $F(u)$ by the nonparametric empirical distribution function. Balkema-de Haan (1974) and Pickands (1975) theorems (*BHP*) show that the conditional excess distribution function defined by F and u is approximated by a Generalized Pareto distribution when u converges to infinity at a certain rate. Therefore, we will assume hereafter $u := u_n$ being a threshold sequence that converges to infinity.

Result 2: *BHP theorem:*

$$\lim_{u_n \rightarrow \infty} [P\{X \leq a_n | X > u_n\} - GPD_{\xi, \sigma_{u_n}}(a_n - u_n)] \longrightarrow 0, \quad \text{as } n \rightarrow \infty, \quad (17)$$

and where

$$GPD_{\xi, \sigma_{u_n}}(a_n - u_n) = \begin{cases} 1 - (1 + \frac{(a_n - u_n)}{\xi \sigma_{u_n}})^{-\frac{1}{\xi}} & \text{if } \xi \neq 0, \\ 1 - \exp\left(\frac{-(a_n - u_n)}{\sigma_{u_n}}\right) & \text{if } \xi = 0, \end{cases} \quad (18)$$

is the so-called Generalized Pareto distribution (GPD), with F a distribution function satisfying the above regularity condition (1), and ξ and σ_{u_n} the location and scale parameters, respectively.

This result can be further refined if F decays polynomially since in this case there exists a reparametrization of the GPD distribution into a Pareto distribution.

Result 3: BHP theorem for F polynomially decaying:

$$\lim_{u_n \rightarrow \infty} \left[P\{X \leq a_n | X > u_n\} - \left(1 - \left(\frac{a_n}{u_n}\right)^{-\frac{1}{\xi}}\right) \right] \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (19)$$

We concentrate on the second result. Expression (19) along with the nonparametric estimator of $F(u_n)$ given by $(1 - \frac{n_{u_n}}{n})$, with n_{u_n} the number of exceedances of the threshold sequence u_n by the observations X_1, X_2, \dots, X_n , yield the following semi-parametric candidate to approximate the normalizing sequence a_n ;

$$\hat{a}_n = u_n(n_{u_n}^{\hat{\xi}_n}), \quad (20)$$

where $\hat{\xi}_n$ a consistent estimator of the tail index ξ . Pickands (1975) proposed an estimator for the tail index based on the Generalized Pareto distribution, other alternative estimators are found in Dekker, Einmahl and de Haan (1989) or Huisman, Koedijk, Kool, and Palm (2001) better suited for small sample sizes. Nevertheless, due to its simplicity and applicability the Hill (1975) estimator, defined by

$$\hat{\xi}_{Hill}(u_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} \ln M_{i:n} - \ln u_n, \quad (21)$$

with $\{M_{i:n}\}_{i=1}^{k_n}$ the sequence of decreasing order statistics and k_n denoting the number of observations exceeding u_n , is still considered as benchmark for estimating ξ . Clearly, this statistic is a function of the threshold sequence, whose choice introduces bias and inefficiencies into the estimation of ξ . To minimize these effects Pickands (1975) proposed a candidate for this sequence based on the distance of the supremum between the empirical version of $P\{X \leq a_n | X > u_n\}$ and the Generalized Pareto distribution. Other methods are proposed by Hall and Welsh (1984),

Guillou and Hall (2001) or Gonzalo and Olmo (2004). In practice this sequence can be chosen as an intermediate order statistic of the sequence of data, with $k_n \rightarrow \infty$ and $k_n = o(n)$.

The relevant test statistic is in this case

$$T_{n,\Phi_{\hat{\xi}_n}} = \frac{1}{\hat{\xi}_n} \ln M_n - \frac{1}{\hat{\xi}_n} \ln u_n - \ln n_{u_n}. \quad (22)$$

The following subsection studies the effects of outlying observations in the estimation of the tail index.

3.1 Detection of Influential observations

In this section we concentrate on parent distributions with $\xi > 0$. In this case when the distribution function F is not known the relevant test statistic is $T_{n,\Phi_{\hat{\xi}_n}}$, shown above. Under the null hypothesis of no outliers this estimator converges in distribution to a Gumbel extreme value distribution. Note, however, that under the alternative hypothesis $H_{A,n}$ the presence of outliers produces two opposite effects on the value of the test statistic. On the one hand, the sample maximum is larger than it should be under no outliers, and on the other hand the outlier also produces estimates of ξ that consistently overshoot the true parameter. If the second effect is stronger than the first effect the outlying observations are masked, and due to estimation effects, cannot be detected. Observations in this category are usually termed influential observations.

The next paragraphs investigate the effect of these observations on our hypothesis test for heavy tails. Further, in order to illustrate these effects more clearly and for ease of calculation, we will ignore the effect of estimating $F^{-1}(1 - \frac{1}{n})$ and write the relevant test statistic as in (11);

$$T_{n,\Phi_{\hat{\xi}_n}} = \frac{1}{\hat{\xi}_n} \ln M_n^0 - \frac{1}{\hat{\xi}_n} \ln F^{-1}(1 - \frac{1}{n}) + \frac{1}{\hat{\xi}_n} \ln(1 + \eta_{n,\Phi_\xi}), \quad (23)$$

where $\eta_{n,\Phi_\xi} = \frac{M_n}{M_n^0} - 1$.

Under the presence of the outlier M_n the tail index can be written as

$$\hat{\xi}_n = \hat{\xi}_n^0 + \nu_n, \quad (24)$$

with $\nu_n := \frac{1}{\hat{\xi}_n} \ln(1 + \eta_{n,\Phi_\xi})$, and $\hat{\xi}_n^0$ the corresponding Hill estimator when there are no outliers in

the sequence of length n . Now, replacing in the above expression we obtain

$$T_{n,\Phi_{\hat{\xi}_n}} = \left(\frac{1}{1 + \frac{\nu_n}{\hat{\xi}_n^0}} \right) \left(\frac{1}{\hat{\xi}_n^0} \ln M_n^0 - \frac{1}{\hat{\xi}_n^0} \ln F^{-1} \left(1 - \frac{1}{n} \right) + \frac{1}{\hat{\xi}_n^0} \ln(1 + \eta_{n,\Phi_\xi}) \right), \quad (25)$$

and the power function of the test with $\hat{\xi}_n^0$ a consistent estimator of ξ can be approximated by the asymptotic Gumbel extreme value distribution.

Proposition 3: *Let F be an unknown distribution function polynomially decaying in its tail, and satisfying the above regularity condition (1). Then, the asymptotic power function of (5) when the tail index is estimated is*

$$\lim_{n \rightarrow \infty} P\{T_{n,\Phi_{\hat{\xi}_n}} > \Lambda_\alpha\} = \left[1 - (1 - \alpha)^{(1 + \eta_{n,\Phi_\xi})^{(1 - \frac{\Lambda_\alpha}{k_n}) \frac{1}{\hat{\xi}_n^0}}} \right] + o_P(1), \quad as \quad n \rightarrow \infty. \quad (26)$$

Proof. From (25) we have that

$$\lim_{n \rightarrow \infty} P\{T_{n,\Phi_{\hat{\xi}_n}} > \Lambda_\alpha\} = \lim_{n \rightarrow \infty} P\{T_{n,\Phi_{\hat{\xi}_n}}^0 > \Lambda_\alpha \left(1 + \frac{\nu_n}{\hat{\xi}_n^0} \right) - \frac{1}{\hat{\xi}_n^0} \ln(1 + \eta_{n,\Phi_\xi})\}.$$

The asymptotic distribution of $T_{n,\Phi_{\hat{\xi}_n}}^0$ is a Gumbel extreme value distribution. Therefore, after some algebra we have that

$$\lim_{n \rightarrow \infty} P\{T_{n,\Phi_{\hat{\xi}_n}} > \Lambda_\alpha\} = \left[1 - (1 - \alpha)^{\exp(-\Lambda_\alpha \frac{\nu_n}{\hat{\xi}_n^0}) (1 + \eta_{n,\Phi_\xi})^{1/\hat{\xi}_n^0}} \right] + o_P(1), \quad as \quad n \rightarrow \infty.$$

Now, noting that $\exp(-\Lambda_\alpha \frac{\nu_n}{\hat{\xi}_n^0}) = \exp(\ln(1 + \eta_{n,\Phi_\xi}) \frac{-\Lambda_\alpha}{k_n \hat{\xi}_n^0})$, and after further simple algebra we obtain

$$\lim_{n \rightarrow \infty} P\{T_{n,\Phi_{\hat{\xi}_n}} > \Lambda_\alpha\} = \left[1 - (1 - \alpha)^{(1 + \eta_{n,\Phi_\xi})^{(1 - \frac{\Lambda_\alpha}{k_n}) \frac{1}{\hat{\xi}_n^0}}} \right] + o_P(1), \quad as \quad n \rightarrow \infty. \quad \square$$

This proposition shows that the effects of parameter estimation in the hypothesis test (5) vanish asymptotically as $k_n \rightarrow \infty$. In finite sample studies, on the other hand, this effect reduces the power of the test compared to the version given by ξ known, since the ratio of the approximation of the power with estimation effects by the power without these effects is less than one.

Corollary 2: Let F be an unknown distribution function polynomially decaying in its tail, and satisfying the above regularity condition (1). Then, the statistical power of the test statistic (25) is smaller than the power of the test statistic (11). More formally, under $H_{O,n}$,

$$\lim_{n \rightarrow \infty} \frac{P\{T_{n,\Phi_{\hat{\xi}_n}} > \Lambda_\alpha\}}{P\{T_{n,\Phi_\xi} > \Lambda_\alpha\}} < 1. \quad (27)$$

Proof. This result can be shown by computing the ratio $\lim_{n \rightarrow \infty} \frac{P\{T_{n,\Phi_{\hat{\xi}_n}} \leq \Lambda_\alpha\}}{P\{T_{n,\Phi_\xi} \leq \Lambda_\alpha\}}$ and observing that it is greater than one:

$$\lim_{n \rightarrow \infty} \frac{P\{T_{n,\Phi_{\hat{\xi}_n}} \leq \Lambda_\alpha\}}{P\{T_{n,\Phi_\xi} \leq \Lambda_\alpha\}} = (1 - \alpha) \left[(1 + \eta_{n,\Phi_\xi})^{(1 - \frac{\Lambda_\alpha}{k_n}) \frac{1}{\xi_n^0} - (1 + \eta_{n,\Phi_\xi})^{\frac{1}{\xi}}} \right] + o_P(1).$$

This expression is greater than one if the following exponent is negative;

$$(1 + \eta_{n,\Phi_\xi})^{(1 - \frac{\Lambda_\alpha}{k_n}) \frac{1}{\xi_n^0} - (1 + \eta_{n,\Phi_\xi})^{\frac{1}{\xi}}} = (1 + \eta_{n,\Phi_\xi})^{\frac{1}{\xi}} \left((1 + \eta_{n,\Phi_\xi})^{(1 - \frac{\Lambda_\alpha}{k_n}) \frac{1}{\xi_n^0} - \frac{1}{\xi}} - 1 \right). \quad (28)$$

For reasonably large n , this expression can be approximated by $(1 + \eta_{n,\Phi_\xi})^{\frac{1}{\xi}} \left((1 + \eta_{n,\Phi_\xi})^{-\frac{\Lambda_\alpha}{k_n \xi}} - 1 \right)$, that is clearly negative, provided that $\xi > 0$. \square

In what follows we propose an algorithm to detect iteratively the presence of multiple outliers, and filtering at the same time, the perturbations produced by the estimation of the tail index. To illustrate this procedure we show first the increase in power obtained from repeating twice the test. The first time we compute the test statistic with the parameters estimated using n observations, that is, $T_{n,\Phi_{\hat{\xi}_n}}$, and a second time with tail index estimated with $n - 1$ observations, $T_{n,\Phi_{\hat{\xi}_{n-1}}}$, in which stage we assume there are no estimation effects on ξ , that is, $\hat{\xi}_{n-1} = \hat{\xi}_{n-1}^0$.

The increase in statistical power is given by observing that $\lim_{n \rightarrow \infty} \frac{P\{T_{n,\Phi_{\hat{\xi}_n}} \leq \Lambda_\alpha\}}{P\{T_{n,\Phi_{\hat{\xi}_{n-1}}} \leq \Lambda_\alpha\}}$ is greater than one. This is shown in the following formulas;

$$\frac{(1 - \alpha)^{(1 + \eta_{n,\Phi_\xi})^{(1 - \frac{\Lambda_\alpha}{k_n}) \frac{1}{\xi_n^0}}}}{(1 - \alpha)^{(1 + \eta_{n,\Phi_\xi})^{\frac{1}{\xi_{n-1}^0}}}} = (1 - \alpha) \left[(1 + \eta_{n,\Phi_\xi})^{(1 - \frac{\Lambda_\alpha}{k_n}) \frac{1}{\xi_n^0} - (1 + \eta_{n,\Phi_\xi})^{\frac{1}{\xi_{n-1}^0}}} \right]. \quad (29)$$

Following the same argument as before it is easy to see that the limiting ratio of probabilities is greater than one if the exponent is negative. Further, for n sufficiently large, the two estimators

of the tail index converge to the true parameter ξ , and therefore

$$(1 + \eta_{n, \Phi_\xi})^{\frac{1}{\xi}} \left((1 + \eta_{n, \Phi_\xi})^{-\frac{\Lambda_\alpha}{k_n \xi}} - 1 \right) < 0.$$

In the algorithm we repeat this procedure tol times for each outlier candidate, with tol determined a priori, in order to filter the effects of estimating ξ with possible influential observations $M_{1:n}, M_{2:n}, \dots, M_{tol-1:n}$. Also note that m denotes the maximum number of order statistics tested as possible outlier candidates. The algorithm is as follows.

Algorithm: (Outlier detection under the presence of influential observations)

1. $j = k = n$; $iter = 0$; $tol_o = 0$.
2. Let $M_j := \max\{X_1, \dots, X_j\}$. Choose k_j as a fraction of j to estimate the tail index ξ .
 $iter = iter + 1$.
3. $tol_o = tol_o + 1$. Estimate $\hat{\xi}_{k_j}$ with $M_{j+1-k_j:j}, \dots, M_{j+k_j-k_j:j}$ and $u_n := M_{j+k_j-k+1:j}$.
4. Compute $T_{j, \Phi_{\hat{\xi}_{k_j}}} = \frac{1}{\hat{\xi}_{k_j}} \ln M_j - \frac{1}{\hat{\xi}_{k_j}} \ln u_n - \ln n_{u_n}$, with n_{u_n} the number of exceedances of the sequence $\{X_i\}_{i=1}^n$ by u_n .
5. If $T_{j, \Phi_{\hat{\xi}_{k_j}}} > \Lambda_\alpha$, reject $H_{O,j}$. Then
 - (a) report " M_j is an outlier."
 - (b) $j = j - k$; $k = j$; $tol_o = 0$.
 - (c) If $iter < m$, go to step 2.
 - (d) Otherwise report *END*.
6. Otherwise,
 - (a) If $tol_o \leq tol$, $k = j - k$. Go to step 3.
 - (b) Otherwise, if $iter < m$ go to step 2.
 - (c) Otherwise report *END*.

4 Conclusions

In applied statistical and econometric problems it is important to distinguish between observations produced by the process generating the data from observations that are obtained exogenously to

the process. Whereas observations in the first group have a positive probability of happening again, and due to their magnitude, can have a considerable impact on the process under study, observations in the second group should be neglected from the analysis. The effects of assuming that these observations are extreme values rather than outliers are important from various standpoints. Thus, standard estimation methods and statistical inference, as confidence intervals and hypothesis tests, can be heavily affected and yield misleading results. Also, the conclusions drawn from looking at the data can be very different from identifying these observations in one group or another.

We show that extreme value theory provides a very powerful device to detect and identify the presence of anomalies in the largest observations of a sample of a given length. In particular the mechanism introduced in this paper allows to exploit available information in a more efficient manner than robust trimming methods by discarding not a fixed fraction of observations in the tails, but only those that fail to pass our hypothesis test. At the same time, the method minimizes bias problems produced by including observations that do not correspond to the data generating process.

The implemented detection methods can be applied to general forms of the distribution function only satisfying very weak regularity conditions implying smoothness in the tails. Further, the hypothesis tests and algorithms developed in this paper make allowance for estimation effects, and permit to filter for influential observations masking the occurrence of outliers. As a byproduct of the study of the statistical power of the test we show that for tests based on extreme value theory the power depends on the magnitude of the outlier, but also on the rate of convergence of the sample maximum and on the degree of polynomial decay (heaviness) of the parent distribution.

Extensions of this research are manifold. These methods can be applied very easily to detecting outliers in regression models and time series models by simply filtering the corresponding residual sequences and re-estimating the model parameters. Note also that no comparison study with other existing methods for outlier detection has been carried out in the paper. Another interesting extension is, therefore, to do this for different detection methods and under different alternative hypotheses.

References

- [1] Balkema, A.A., and L. de Haan (1974). *Residual lifetime at great age*. Annals of Probability 2, 792-804.
- [2] Barnett, V., and T. Lewis (1994). *Outliers in Statistical Data*. Chichester: Wiley.
- [3] Basmann, R.L. (2003). *Statistical outlier analysis in litigation support: the case of Paul F. Engler and Cactus Feeders, Inc., v. Oprah Winfrey et al.* Journal of Econometrics 113, 159-200.
- [4] Beckman, R., and R. Cook (1983). *Outlier.....s*. Technometrics 25, 2, 119-149.
- [5] Bickler, P., (1965). *On Some Robust Estimates of Location*. The Annals of Mathematical Statistics 36, 3, 847-858.
- [6] Doornik, J.A., and M. Ooms (2005). *Outlier Detection in GARCH Models*. Tinbergen Institute Discussion Paper 05-092/4.
- [7] Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling extremal events for insurance and finance*. Springer, Berlin.
- [8] Ferguson, T.S., (1961). *On the rejection of outliers*. In Proceedings of the 4th Berkeley Symposium in Mathematical Statistics and Probability, 253-287.
- [9] Fisher, R.A., and L.H.C Tippet (1928). *Limiting forms of the frequency distribution of the largest and smallest member of a sample*. Proc. Cambridge Philos. Soc., 24, 180-190.
- [10] Gnedenko, B.V., (1943). *Sur la distribution limit du terme d'une série aléatoire*. Annals of Mathematics 44, 423-453.
- [11] Gonzalo, J., and J. Olmo (2004). *Which Extreme Values are Really Extremes?*. Journal of Financial Econometrics 2 (3), 349-369.
- [12] Grubbs, F.E., (1950). *Sample Criteria for Testing Outlying Observations*. The Annals of Mathematical Statistics 21, 27-58.
- [13] Grubbs, F.E., (1969). *Procedures for Detecting Outlying Observations in Samples*. Technometrics, 11, 1, 1-21.

- [14] Guillou, A., and P. Hall (2001). *A diagnostic for selecting the threshold in extreme value analysis*. Journal of Royal Statistical Society B, 63, 3, 293-305.
- [15] Hall, P., and A.H. Welsh (1984). *Best Attainable Rates of Convergence for Estimates of Parameters of Regular Variation*. The Annals of Statistics 12, 1079-1084.
- [16] Hawkins, D. (1980). Identification of Outliers. London: Chapman and Hall.
- [17] Hill, B.M., (1975). *A simple general approach to inference about the tail of a distribution*. The Annals of Statistics 3, 5, 1163-1174.
- [18] Maronna, R., D. Martin, and V. Yohai (2006). Robust Statistics - Theory and Methods. Wiley Publisher.
- [19] Mittnik, S., S.T., Rachev, and G. Samorodnitsky (2001). *The Distribution of Test Statistics for Outlier Detection in Heavy-tailed Samples*. Mathematical and Computer Modelling 34, 9, 1171-1183.
- [20] Pickands III, James, (1975). *Statistical Inference Using Extreme Order Statistics*. The Annals of Statistics 3, 1, 119-131.
- [21] Rosner, B., (1983). *On the detection of many outliers*. Technometrics 17, 221-227.
- [22] Rousseeuw, P.J., and A.M. Leroy (1987). Robust Regression and Outlier Detection. Wiley Publisher.
- [23] Schluter, C., and M. Trede (2002). *Identifying Multiple Outliers in Heavy-Tailed Distributions with an Application to Market Crashes*. Journal of Empirical Finance 15, 4, 700-713.
- [24] Stigler, S.M., (1973). *The Asymptotic Distribution of the Trimmed Mean*. The Annals of Statistics 1, 3, 472-477.
- [25] Verhoeven, P., and M. McAleer (2000). *Modelling Outliers and Extreme Observations for ARMA-GARCH Processes*. <http://fmwww.bc.edu/RePEc/es2000/1922.pdf>