ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 228, 2009

*Grażyna Trzpiot**

EXTREME VALUE DISTRIBUTIONS AND ROBUST ESTIMATION

Abstract. In parametric statistics estimators such as maximum likelihood or OLS typically estimate stochastic models, which play an important role in finance and insurance. These methods are generally optimal for an assumed reference model. Slight deviations from the assumed model may easy destroy the good statistical properties of the estimator. We present some aspects related to robust estimation in the context of extreme value theory (ETV). We discuss some methodological aspects how robust methods can improve the quality of extreme value theory data analysis by providing information on influential observations.

Key words: Extreme value theory, Extreme value distributions, Robust estimation, *M*-estimator.

I. INTRODUCTION

Robust statistics achieves this by a set of different statistical frameworks that generalize classical statistical procedures such as maximum likelihood or OLS. Seminal contributions are Huber (1981) and Hampel et al. (1986). Since then many different and related approaches have emerged. Dell'Aquila and Ronchetti (2006) give a comprehensive introduction to the principles of robust statistics estimation, testing and model selection and apply and extend the theory to different models used in risk management, asset allocation and insurance. We discuss how robust methods can improve the EVT data analysis by providing information on influential observations, deviating substructures and possible wrong specification of a model. Basically EVT is using in risk management, see Embrechts et al. (1997) and McNeil et al. (2005). We find that robust statistical methods can improve the data analysis process of the skilled analyst and provide him with useful additional information. We will shortly review some key concepts from EVT and robust statistics and next we will consider how the whole data analysis process can be improved by additionally using robust statistical procedures.

* Professor, Department of Statistics, University of Economics in Katowice.

II. EXTREME VALUE THEORY

Extreme value theory is more and more used in recent years to model extremes of financial and economic data or natural phenomena. The EVT framework provides on the one hand asymptotic distributions for the description of (normalized) maxima or minima and on the other hand the asymptotic distribution of extremes over a high threshold. Basic references with a focus on finance and insurance are Embrechts et al. (1997) and Malevergne and Sornette (2006).

2.1. Generalized extreme value distributions

The EVT analyses the asymptotic distribution of (normalized) *maxima* or *minima* of i.i.d. samples, i.e. $M_n = \max(X_1, \ldots, X_n)$. It turns out that under weak conditions, the normalized maximum of n i.i.d. random variables is distributed as Gumbel, Weibull or Fréchet, depending on the data generation process.

The generalized extreme value (GEV) distributions can be combined into a single form

$$F_\theta^{GEV}(x) = F_{\mu,\beta,\xi}^{GEV}(x) = \exp\left(-\left(1+\frac{\xi(x-\mu)}{\beta}\right)^{-1/\xi}\right) \tag{1}$$

where

$$1+\frac{\xi(x-\mu)}{\beta} > 0 \text{ and } \beta > 0.$$

The parameters μ and β are the location and scale and, ξ is the shape parameter of limiting distribution. Its sign determines the three possible limiting forms of the GEV of distribution of maxima:

- If $\xi = 0$ the limit distribution is the Gumbel distribution (double- exponential),
- If $\xi > 0$ the limit distribution is the (shifted) Fréchet power-like distribution,
- If $\xi < 0$ the limit distribution is the Weibull distribution.

A special case is the Gumbel distribution (taking the limit $\xi \to 0$)

$$F_\theta^{Gum}(x) = F_{\mu,\beta,0}^{Gum}(x) = \exp\left(-\exp\left(\frac{-(x-\mu)}{\beta}\right)\right) \tag{2}$$

This distribution is widely used as it is the appropriate limit of maxima from many common distributions, e.g. normal, lognormal, Weibull and gamma.

2.2. Generalized Pareto distribution

Another important result in EVT is related to the distribution function for exceedances over a given threshold. It turns out that excesses over a high threshold u have a generalized Pareto distribution (GPD) with distribution function:

$$F_{\theta}^{GPD}(x) = F_{\beta,\xi}^{GPD}(x) = \begin{cases} 1-(1+\dfrac{\xi x}{\beta})^{-1/\xi}, & \text{dla} \quad \xi \neq 0 \\ 1-\exp(-x/\beta), & \text{dla} \quad \xi = 0 \end{cases} \tag{3}$$

where
$\beta>0$ and $x \geq 0$ when $\xi \geq 0$,
$0 \leq x \leq -\beta/\xi$ when $\xi < 0$,
for $\xi = 0$ the limiting distribution is exponential.

The case $\xi > 0$ corresponds to heavy-tailed distributions whose tails decay like power functions, such as the Pareto, Student *t*, Cauchy, Burr, log-gamma and Fréchet distributions. The case $\xi = 0$ corresponds to distributions like the normal, exponential, gamma and lognormal, with tails essentially decaying exponentially. The final group of distributions ($\xi < 0$) are short-tailed distributions with a finite right endpoint, such as the uniform and beta distributions.

III. ESTIMATION METHODS

Consider a parametric model given by a distribution F_θ with density f_θ . In classical statistics, one often chooses an estimation framework that is optimal at the *assumed model distribution* (e.g. the maximum likelihood framework delivers the asymptotically most efficient estimator at the model distribution). However, as soon as the real underlying model deviates from the assumed one, the estimator may lose its good statistical properties and many alternative estimators may perform better.

In robust statistics we want to construct estimators and tests that have good statistical properties (high efficiency, low bias) for a *whole neighbourhood of the assumed model distribution* F_θ . Such a neighborhood can, for example, be formalized by:

$$A_\varepsilon(F_\theta) = \{G_\varepsilon | G_\varepsilon = (1-\varepsilon)F_\theta + \varepsilon G, \ G \text{ arbitrary}\}$$

and $0 < \varepsilon < 1$, thought of as a measure for *contamination*.

The aim of robust statistics is to provide statistical procedures:

– which are still reliable and reasonably efficient under small deviations from the assumed parametric model and to quantify the maximal bias on the statistical quantity of interest when the underlying distribution lies in a neighborhood of the reference model[1].

– these procedures should highlight which observations (e.g. outliers) or deviating substructures have most influence on the statistical quantity under observation.

3.1. Maximum likelihood estimators

The parameter θ for distributions such as GEV ($\theta = [\mu, \beta, \xi]^T$) or GDP ($\theta = [\beta, \xi]^T$) are typically estimated by maximum likelihood, i.e.:

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} \log f_{\theta}(x_i) \tag{4}$$

or finding the zeros of the estimating equations

$$\sum_{i=1}^{n} s(x_i, \theta) = 0 \tag{5}$$

where

$$s(x_i, \theta) = \frac{\partial \log f_{\theta}(x)}{\partial \theta} \tag{6}$$

is the score function.

3.2. *M*-estimators

We will consider only the most general robust framework, the *M*-estimation framework. *M*-estimators can be seen as a generalization of the maximum likelihood approach and allow to analyses the robustness properties of estimators and tests.

An *M*-estimator is defined as the solution to the minimization problem

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} \rho(x_i; \theta) \tag{7}$$

[1] In this sense it is a generalization of the classical statistical procedures.

for some objective function ρ. If ρ has a derivative

$$\psi(x;\theta) = \frac{\partial \rho(x;\theta)}{\partial \theta} \tag{8}$$

then the M-estimator satisfies the first order conditions

$$\sum_{i=1}^{n} \psi(x_i, \theta) = 0 . \tag{9}$$

In general we restrict to estimators, which are Fisher consistent, we require a ψ such that

$$E_{F_\theta}[\psi(x;\theta)] = 0 . \tag{10}$$

Under weak conditions on ψ it can be shown that the resulting estimator is normally distributed with variance–covariance matrix given by

$$V = M^{-1} \cdot E[\psi(X;\theta)\psi(X;\theta)^T] \cdot M^{-T} \tag{11}$$

where

$$M = E[-\frac{\partial \psi(X;\theta)}{\partial \theta}] . \tag{12}$$

The classical maximum likelihood estimator corresponds to

$$\rho(x;\theta) = -\log f_\theta(x) \tag{13}$$

or

$$\psi(x;\theta) = s(x;\theta) = \frac{\partial \log f_\theta(x)}{\partial \theta} \tag{14}$$

where $s(x;\theta)$ is the score function.

3.3. Robust estimators - properties

General results in robust statistics imply that an estimator with

– a *bounded asymptotic bias* in a neighbourhood of the reference model can be constructed by choosing a *bounded* function ψ (in x);

– a high *asymptotic efficiency* can be achieved by choosing a ψ function which is *similar* to the score function $s(x;\theta)$ in the range where most of the observations lie.

There is an trade off between maximal asymptotic bias in a neighbourhood of the model distribution F_θ and the asymptotic efficiency of the estimator at the reference model. Because ψ enters in the linear approximation of the asymptotic bias as well as in the asymptotic variance of the estimator (11), it is possible to solve a general optimality problem to find the estimator that is the most efficient given a bound on the maximal bias of the estimator in a neighbourhood of the model.

The solution to this problem is the *M*-estimator defined by

$$\psi_c^{A,a}(x;\theta) = h_c(A(\theta)(s(x,\theta) - a(\theta)) \tag{15}$$

where

$$h_c(r) = r\min(1, \frac{c}{\|r\|})$$

is a multivariate version of the Huber function seen above and the matrix A and the vector a are determined by solving:

$$E_{F_\theta}[h_c(A(\theta)(X - a(\theta))] = 0 \text{ and } E_{F_\theta}[\psi_c^{A,a}(X;\theta)\psi_c^{A,a}(X;\theta)]^T = I \tag{16}$$

which ensure that the estimator is consistent and the asymptotic bias remains below the chosen bound. In the one-dimensional location case presented above, the optimal solution reduces to using the ψ_c function, in particular in this symmetric case $a = 0$. For asymmetric reference models, $a(\theta)$ must be typically found numerically to ensure consistency of the estimator. The computation of the estimator can typically be performed by a slightly adapted Newton-Raphson type procedure.

IV. APPLYING *M*-ESTIMATORS TO EVT

We consider the estimation of a Gumbel model, one of the GEV distributions. It can be verified that the score function is unbounded in x. In this example (Dell'Aquila, Ronchetti, 2006) we would like to highlight that the robust (in this case the optimal robust) estimator is able to detect observations that do not conform to the bulk of the data, even in the case of a very asymmetric model. These observations must not necessarily be far away. To illustrate this robustness issue, we generate 300 observations from a contaminated model given by 95% from a Gumbel with $\mu = 4$ and $\beta = 2$ and 5% from a Gumbel with $\mu = -0.5$ and $\beta = 0.2$; notice that the contaminated model puts more mass on the left of the 'true' Gumbel distribution as can be seen in Figure 1, which plots the two densities.

Figure 2 shows that the classical estimator for (μ, β) is clearly attracted by the contaminating structure and fails to model part of the majority of the data. The robust estimator (tuned to have approximately 90% efficiency at the model) fits the distribution much better where most data are located.

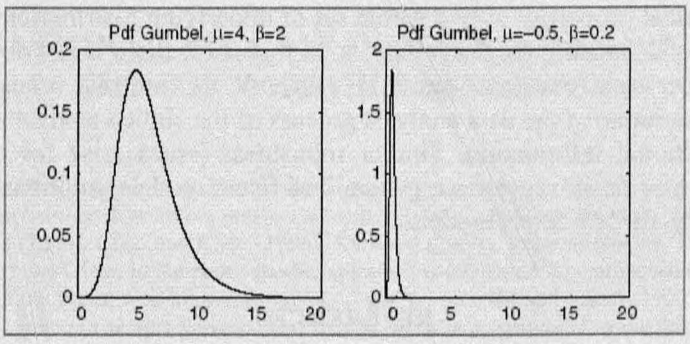

Fig. 1. Gumbel model – the GEV distributions

Source: Dell'Aquila, Ronchetti (2006).

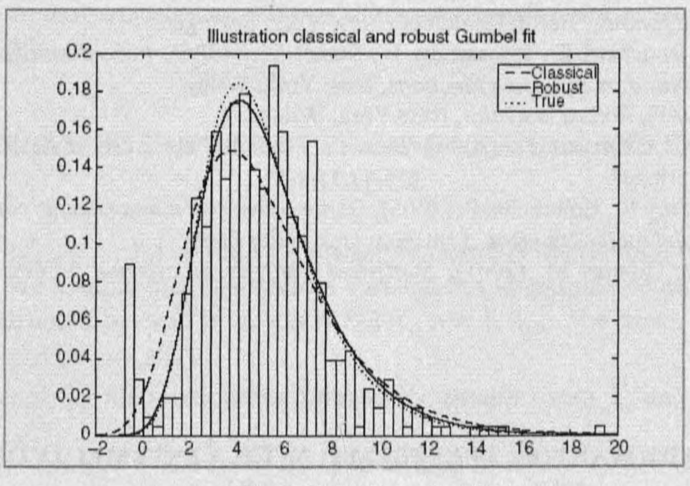

Fig. 2. Contaminated model

Source: Dell'Aquila, Ronchetti (2006).

Similar robustness issues apply for other EVT distributions such as the Weibull and GPD distribution and for other distributions such as gamma, beta, etc. We can observe that robust methods do not down weight 'extreme' observations if they conform to the majority of the data. Additionally robust methods can guarantee a stable efficiency, MSE and a bounded bias over a whole neighborhood of the assumed distribution.

V. CONCLUSIONS

In this paper we have discussed how robust statistics may improve the data analysis process in the specific case of EVT. We have seen that robust methods can help to identify deviating structure, influential observations and guarantee good statistical properties over a whole set of underlying distributions, therefore considerably enhancing the data analysis. In this sense there is no 'obvious' contradiction between robustness and EVT. Overall we find that robust statistical methods can improve the data analysis process of the skilled analyst and provide useful additional information. Similar robustness issues arise for many other models such as linear regression, generalized linear models, multivariate models and virtually all time series models.

REFERENCES

Dell'Aquila R., Ronchetti E. (2006), *Robust statistics and econometrics with economic and financial applications*, New York, Wiley.

Embrechts P., Klüppelberg C., Mikosch T. (1997), *Modeling extremal events for insurance and finance*, Berlin Heidelberg New York, Springer.

Hampel F., Ronchetti E., Rousseeuw P., Stahel W. (1986), *Robust statistics: The approach based on influence functions*, New York, Wiley.

Huber P. (1981), *Robust statistics*, New York, Wiley.

Malevergne Y., Sornette D. (2006), *Extreme Financial Risk*, Berlin Heidelberg New York, Springer.

McNeil A., Frey R., Embrechts P. (2005), *Quantitative risk management: concepts, techniques and tools*, Princeton, Princeton University Press.

Reiss R.-D., Thomas M. (2001), *Statistical analysis of extreme values*, Basel, Birkhäuser.

Grażyna Trzpiot

ROZKŁADY WARTOŚCI EKSTREMALNYCH A ESTYMACJA ODPORNA

Modele stochastyczne są istotne dla zastosowań w finansach czy w ubezpieczeniach. Statystyczne metody estymacji parametrycznej wykorzystywane najczęściej do wyznaczania parametrów modeli to metoda największej wiarygodności lub MNK. Metody te dają optymalne oszacowania modeli, jednakże odchylenia obserwowanych wartości w kalibrowanym modelu mogą zachwiać dobre własności estymatorów. Przedstawimy pewne aspekty estymacji odpornej w kontekście rozkładów wartości ekstremalnych. Podejmiemy dyskusję metodologicznych aspektów zagadnienia pokazując, jak estymatory odporne wpływają na jakość analiz z wykorzystaniem rozkładów wartości ekstremalnych poprzez informacje o obserwacjach wpływowych.