Fin Mkts Portfolio Mgmt (2006) 20:103–118
DOI 10.1007/s11408-006-0002-x

Extremes and robustness: a contradiction?

Rosario Dell'Aquila · Paul Embrechts

Published online: 23 March 2006

Abstract Stochastic models play an important role in the analysis of data in many different fields, including finance and insurance. Many models are estimated by procedures that lose their good statistical properties when the underlying model slightly deviates from the assumed one. Robust statistical methods can improve the data analysis process of the skilled analyst and provide him with useful additional information. For this anniversary issue, we discuss some aspects related to robust estimation in the context of extreme value theory (EVT). Using real data and simulations, we show how robust methods can improve the quality of EVT data analysis by providing information on influential observations, deviating substructures and possible mis-specification of a model while guaranteeing good statistical properties over a whole set of underlying distributions around the assumed one.

R. Dell'Aquila
RiskLab, ETH Zurich, Rämistrasse 101, 8092 Zurich, Switzerland
E-mail: rosario@math.ethz.ch

P. Embrechts (✉)
Department of Mathematics, ETH Zurich, Rämistrasse 101,
8092 Zürich, Switzerland
E-mail: embrechts@math.ethz.ch
Tel.: +41-44-6326741
Fax: +41-44-6323419

Keywords Robust statistics · Robust estimation · M-estimator · Extreme value theory · Extreme value distributions · Generalized Pareto distribution

JEL classification: G10, G40

1 Introduction

Stochastic models play an important role in the analysis of data in many different fields, including finance and insurance. In parametric statistics these models are typically estimated by estimators such as maximum likelihood or OLS. However, these methods are generally optimal for an assumed reference model, but slight deviations from the assumed model may quickly destroy the good statistical properties of the estimator. Since we can assume that deviation from the model assumptions almost always occurs in finance and insurance data, it is useful to complement the analysis with procedures that are still reliable and reasonably efficient under small deviations from the assumed parametric model and highlight which observations (e.g. outliers) or deviating substructures have most influence on the statistical quantity under observation. 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.

In this paper, we discuss some methodological aspects related to robust estimation in the context of extreme value theory (EVT). Using real data and simulations, we show how robust methods can improve the quality of EVT data analysis by providing information on influential observations, deviating substructures and possible mis-specification of a model, while guaranteeing good statistical properties over a whole set of underlying distributions around the assumed one. We have chosen this example as a sort of 'provocation', after all it seems that one performs EVT just to consider and 'emphasize' the role of extremes and also because the choice of a distribution is often of heuristic nature. Moreover EVT is of increasing importance in risk management, see Embrechts et al. (1997) and McNeil et al. (2005). Specific examples include VaR estimation and quantitative modelling of

operational risk, see Moscadelli (2004). Throughout the paper, we follow a rather non-technical style and leave details to the books cited above.

Overall we find that robust statistical methods can improve the data analysis process of the skilled analyst and provide him with useful additional information. However, robust statistics is not a framework to apply blindly. We also at least mention some methodological issues that deserve more attention and that we cannot fully treat here.

In section 2 we will shortly review some key concepts from EVT and robust statistics. Then in section 3 we will consider some methodological issues and show the added value of robust statistics and give a first impression on how the whole data analysis process can benefit from additionally using robust statistical procedures. We will use insurance data which are close to i.i.d. observations to make a case with real data without the complication of truly depending observations. Part of the analyses presented in this paper were done in parallel with the writing of Dell'Aquila and Ronchetti (2006).

2 Extreme value theory and robust statistics

2.1 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 McNeil et al. (2005), Chapter 7. We also refer to these books for further references.

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_θ^{GEV} where $\theta = [\mu, \beta, \xi]^{\mathrm{T}}$ given by

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

where $1 + \xi(x - \mu)/\beta > 0$ and $\beta > 0$. The parameters μ and β are the location and scale and ξ is the shape parameter. The latter determines which extreme value distribution is represented: Fisher – Tippett types I, II and III (Gumbel, Fréchet and Weibull) correspond to $\xi = 0$, $\xi > 0$ and

$\xi < 0$ respectively. A special case is the Gumbel distribution when $\xi = 0$, i.e., (taking the limit $\xi \to 0$)

$$F^{\text{Gum}}_{\mu,\beta,0}(x) = \exp\left(-\exp\left(-(x-\mu)/\beta\right)\right).$$

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

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

$$F^{\text{GPD}}_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \frac{\xi x}{\beta})^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp(-x/\beta) & \text{if } \xi = 0, \end{cases}$$

where $\beta > 0$, and the support is $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ when $\xi < 0$. For $\xi = 0$ the limiting distribution is exponential. The GPD distribution is motivated by the so-called Pickands–Balkema–de Haan Theorem in EVT. The class of distributions for which this result applies contains essentially all the common continuous distributions of statistics. These may be further subdivided into three groups according to the value of the parameter ξ in the limiting GPD approximation to the excess distribution. 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 log-normal, 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.

The parameter θ of a GEV and GPD are typically estimated by maximum likelihood, i.e. by $\widehat{\theta} = \arg\min_\theta \Sigma_{i=1}^n \log f_\theta(x_i)$, or finding the zeros of the estimating equations $\Sigma_{i=1}^n s(x_i;\theta) = 0$, where $s(x;\theta) = \frac{\partial \log f_\theta(x)}{\partial \theta}$ is the score function.

In particular, for the GEV class, the density function is given by[1]

$$f^{\text{GEV}}_{\mu,\beta,\xi}(x) = \frac{1}{\beta}\left(1 + \xi\frac{x-\mu}{\beta}\right)^{-(\frac{1}{\xi}+1)} \exp\left(-\left(1 + \xi\frac{x-\mu}{\beta}\right)^{-1/\xi}\right),$$

where $1 + \frac{\xi(x-\mu)}{\beta} > 0$. In the Gumbel case the density function is given by

$$f^{\text{Gum}}_{\mu,\beta}(x) = \frac{1}{\beta}\exp(-(x-\mu)/\beta)\exp(-\exp(-(x-\mu)/\beta)).$$

[1] For notational convenience, we do not limit the domain and assume that score functions are used where ML in the standard case can be applied.

In the case of a GPD, the density function is given by ($\xi \neq 0$):

$$f_{\xi,\beta}^{\mathrm{GPD}}(x) = \frac{1}{\beta}\left(1 + \xi\frac{x}{\beta}\right)^{-(\frac{1}{\xi}+1)}.$$

2.2 Some key facts in robust statistics

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.

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. In this sense it is a generalization of the classical statistical procedures.
- At the same time these procedures should highlight which observations (e.g. outliers) or deviating substructures have most influence on the statistical quantity under observation. Thus robust statistics can also be seen as a diagnostic tool describing the bulk of the data and offering an alternative analysis to the researcher.

To fulfil this aim several related statistical frameworks have been developed, indeed more than we can discuss here. We will consider only the most general framework, the M-estimation framework and we refer to Dell'Aquila and Ronchetti (2006) for a comprehensive discussion.

M-estimators can be seen as a generalization of the maximum likelihood approach and allow to analyze the robustness properties of estimators and tests in a *unified framework*. An M-estimator is defined as the solution to the minimization problem

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

for some objective function ρ. If ρ has a derivative $\Psi(x;\theta) = \frac{\partial\rho(x;\theta)}{\partial\theta}$, then the M -estimator satisfies the first order conditions

$$\sum_{i=1}^{n} \Psi(x_i;\theta) = 0. \tag{2}$$

An M-estimator can be more generally defined through the estimating equations (2). In general we restrict to estimators which are Fisher consistent, i.e. we require a Ψ such that $E_{F_\theta}[\Psi(X;\theta)] = 0$. Under weak conditions on Ψ it can be shown that the resulting estimator is normally distributed with variance–covariance matrix given by

$$V = M^{-1} \cdot E[\Psi(X;\theta)\Psi(X;\theta)^{\mathrm{T}}] \cdot M^{-\mathrm{T}}, \tag{3}$$

where $M = E[-\frac{\partial \Psi(X;\theta)}{\partial \theta}]$.

The classical maximum likelihood estimator corresponds to (1) with $\rho(x;\theta) = -\log f_\theta(x)$ or to (2) with $\Psi(x;\theta) = s(x;\theta) = \frac{\partial \log f_\theta(x)}{\partial \theta}$, where $s(x;\theta)$ is the score function.

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 neighbourhood can, for example, be formalized by $\mathcal{A}_\varepsilon(F_\theta) = \{G_\varepsilon | G_\varepsilon = (1-\varepsilon)F_\theta + \varepsilon G$, G arbitrary$\}$ and ε is between 0 and 1 , thought of as a measure for *contamination*.

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* Ψ function [2] (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.

As a simple example consider the location model $x_i = \mu + e_i$. Assuming normal errors e_i, the maximum likelihood estimate is the sample mean $\widehat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i$, which is non-robust, i.e. small deviations from the assumed model (e.g. a t_9 or a slight mixture of normals) can considerably lower the efficiency and a single outlying point can highly bias the outcome. Notice that the mean can be rewritten as an M-estimator by writing $\sum_{i=1}^n (x_i - \mu) = 0$. A simple robust alternative can be constructed by choosing a bounded Ψ function, e.g. the so-called Huber function[3]

$$\psi_c(r) = \begin{cases} r & \text{if } |r| \le c \\ c \cdot \operatorname{sign}(r) & \text{otherwise} \end{cases}, \tag{4}$$

[2] More precisely, this is generally the case for small deviations from the assumed reference model. As a side comment notice that there are estimators with a non-bounded Ψ function but a bounded asymptotic bias. This situation can typically arise when exogenous variables are present (e.g. a linear regression model). For details see Dell'Aquila and Ronchetti (2006). In any case robust estimators can be *constructed* by choosing a bounded Ψ function.

[3] We use lower case ψ for one-dimensional functions.

where c is a constant which determines the degree of robustness and efficiency. This ψ function leaves the score nearly unchanged where most of the data lie (and thus delivers a highly efficient estimator), but is still bounded (which ensures that the maximal bias in a neighbourhood of the model is bounded). Notice that the estimator can be rewritten as a weighted estimator, i.e. $\sum_{i=1}^{n} w_c(r_i) r_i = 0$, where $w_c(r) = \frac{\psi_c(r)}{r}$ can be seen as a weight that measures the "outlyingness" of each observation. The tuning constant c is typically chosen by requiring a given efficiency at the assumed reference model. Typically a lower c will lead to a model which has a lower maximal bias in a neighbourhood of the reference model and lower efficiency at the reference model and vice versa.

There is an obvious 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 (3), 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

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

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 simultaneously by solving $E_{F_\theta}[h_c(A(\theta)(X - a(\theta)))] = 0$ and $E_{F_\theta}[\Psi_c^{A,a}(X;\theta)\Psi_c^{A,a}(X;\theta)^{\mathrm{T}}] = I$, which, respectively, 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. For an introduction in a broader context, details, interpretation of the different estimators and a step-by-step explanation of the computation along with numerous examples, we refer to Dell'Aquila and Ronchetti (2006).

Notice that the multivariate Huber function can be rewritten as

$$\Psi_c^{A,a}(x;\theta) = A(\theta)\,[s(x;\theta) - a(\theta)]\, w_c(A(\theta)\,[s(x;\theta) - a(\theta)]),$$

where $w_c(r,\theta) = \min\left(1; \frac{c}{\|r\|}\right)$ are weights attached to each observation. These weights can be used to trace the "outlyingness" of each observation and deliver additional information on the observations.

We would like to stress that different Ψ functions can be meaningful for different types of data and the kind of question asked on the data. Hence this procedure does not yield a framework to apply blindly.

For this paper, we will pick out one specific robustness issue and discuss some aspects of the estimation in the context of EVT. As an illustration, we take the case of the generalized Pareto model. The score functions are then given by

$$s_\xi(x;\theta) = \frac{x}{(\beta\xi+\xi^2 x)}(\xi+1) + \log\left(1+\frac{\xi x}{\beta}\right)\left(\frac{1}{\xi^2}\right),$$
$$s_\beta(x;\theta) = -\frac{1}{\beta} + \left(1+\frac{\xi x}{\beta}\right)^{-1} x\left(\frac{1+\xi}{\beta^2}\right).$$

It is easy to verify that these functions are not bounded in x. A robust version can be constructed by means of an estimator with a suitably chosen Ψ function, e.g. the optimal one described above.

In the next section, we focus on four key messages to give an idea of the possible added value of robust methods, even in the case of EVT. However, we also mention some methodological issues that deserve more attention and that we cannot fully treat here.

3 Applying robust methods to EVT: some key messages

3.1 Message 1: Robust methods do not downweight 'extreme' observations if they conform to the majority of the data

As a first example we consider the estimation of a Gumbel model, one of the GEV distributions. It is easily verified that the score function is unbounded in x. A robust estimator for estimating (μ, β), respectively, can be found by applying directly the optimal solution described above, but other robust estimators could be considered. In this example 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

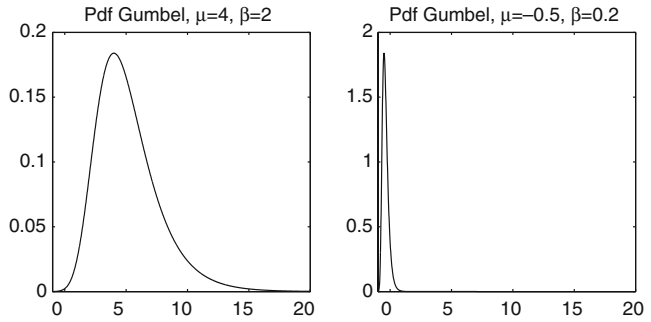

Fig. 1 Densities of a Gumbel distribution with $\mu = 4$ and $\beta = 2$ (on the $left$) and $\mu = -0.5$ and $\beta = 0.2$ on the $right$

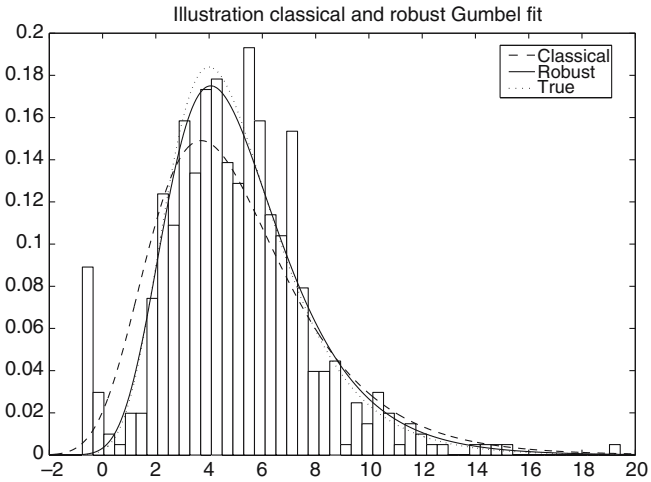

Fig. 2 Illustration of the classical and robust estimations of the parameters of a Gumbel distribution. The classical estimator is attracted by the outlying observations around 0

the model) fits the distribution much better where most data are located. Additionally, the robust estimator explicitly downweights the observations on the left and thus signals to the analyst that some observations do not seem to conform to the majority of the other observations when using this model. An important aspect to notice is that the robust estimator does not downweight the large observations on the right because they conform to the majority of the data.

Similar robustness issues apply for other EVT distributions such as the Weibull and GPD distribution and for other distributions such as gamma, beta, etc.

3.2 Message 2: Robust methods can guarantee a stable efficiency, MSE and a bounded bias over a whole neighbourhood of the assumed distribution

Consider for example the Pareto model defined by the density $f_\alpha(x)$ given by

$$f_\alpha(x) = \alpha x^{-(\alpha+1)} x_0^\alpha, \quad 0 \le x_0 \le x < \infty \text{ and } \alpha > 0.$$

It is easily verified that in this case too, the score function is unbounded. The maximum likelihood estimator $\hat{\alpha}_{MLE} = (\frac{1}{n}\sum_{i=1}^{n} \log(\frac{x_i}{x_0}))^{-1}$ is not robust. In this example we would like to highlight the differences of the robust and classical estimates on the mean squared error (MSE) when the underlying data vary over a whole set of underlying distributions.

As an illustration, we plot the MSE $E[(\theta - \widehat{\theta})^2]$ of the classical and the (optimal) robust estimator of the parameters of a Pareto model for different c and different levels of contamination. To illustrate the point we generate 300 observations from a Pareto distribution with $\alpha = 5$ (and $x_0 = 1$). We contaminate the Pareto distribution by $100\varepsilon\%$ of observations from $\alpha = 5$ and $x_0 = 10$. Figure 3 plots the MSE for the classical and two robust estimators (with $c = 1.3$ and $c = 2.0$, respectively) for various values of ε. These choices of c correspond to estimators which have about 85 and 75% of efficiency at the model.

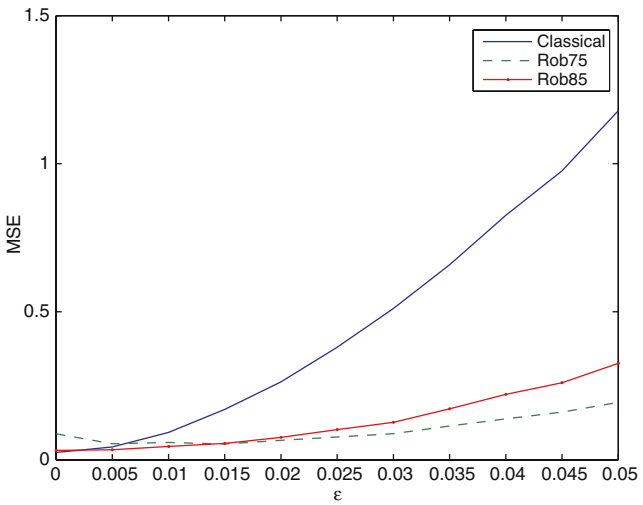

Fig. 3 MSE of the parameter estimates in a Pareto model

Figure 3 shows the MSE in the uncontaminated and contaminated case averaged over 1,000 simulation runs. When the data are uncontaminated ($\varepsilon = 0$), the bias and MSE are similar. However, a slight contamination (less than 1%) is sufficient to make some alternative robust estimator more appealing in terms of low MSE. Furthermore when contamination increases, the bias and MSE of the classical ML estimator explodes, while the robust estimators remain stable. Similar results are obtained for bias and efficiency and for different distributions such as the exponential, Gamma, GPD, Gumbel, Weibull and other distributions. Notice that in the case of outlying points, the variance of the estimator tends to shrink, i.e. it seems that the estimation is more precise (which is particularly dramatic when the outlying point distorts the estimates).

In this context, we would like to briefly mention that there is a rather involved methodological issue that arises when the underlying 'real' distribution has clearly fatter tails than the assumed one. In particular, we have to understand clearly which parameter we would like to estimate. This is a very important issue in a risk management context, especially for volatility and variance–covariance estimation. We refer to Dell'Aquila and Ronchetti (2006) for further discussion of this issue.

3.3 Message 3: Robust methods can identify influential points in real data

As an empirical illustration of robust methodology to insurance and finance data, we apply the robust estimation procedures to the Danish Fire Loss dataset, which is analysed in Embrechts et al. (1997) and McNeil et al. (2005), p. 275. We have chosen this example because of its importance and because we would like to make the issue in a case with i.i.d. data and avoid the complications that arise for dependent data.

The data consist of 2156 fire insurance losses over one million Danish Krone from the years 1980 to 1990 inclusive. The loss figure represents a combined loss for a building and its contents, as well as in some cases a loss of business earnings; the numbers are inflation adjusted to reflect 1985 values. Looking at a mean excess plot, Embrechts et al. (1997) and McNeil et al. (2005) choose a threshold of $u = 10$ and fit a GPD to the 109 excess losses. Replicating the classical analysis we obtain the estimates $(\widehat{\xi}, \widehat{\beta}) = (0.50, 6.98)$ with standard errors $(0.13, 1.17)$. The robust estimates, tuned to have about 95% efficiency [4] at the model, are given

[4] From a methodological point of view we suggest to use a high tuning constant, especially in the presence of very heavy tailed models.

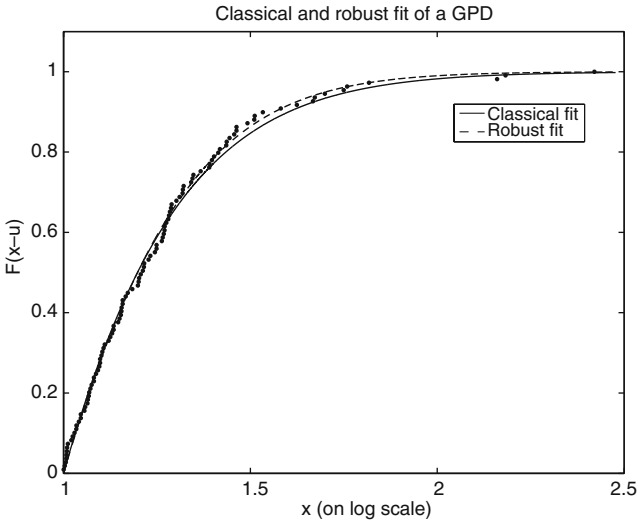

Fig. 4 Empirical distribution of excesses and the classical (*solidline*) and robust (*dashedline*) GPD fits

by $(\widehat{\xi}, \widehat{\beta}) = (0.37, 7.28)$ with standard errors $(0.11, 1.16)$. Figure 4 plots the empirical distribution of excesses and the classical- and robust-fitted GPD. The curves at a first glance look similar, however, major differences occur for larger values; it seems that the classical estimator does not fit well some proportion of the data. The robust fit performs well in this respect and models accurately the majority of the data. In particular it seems that some of the most extreme observations 'attract' the classical estimator. In the robust case these are heavily downweighted. Because these points are additionally leverage points, our interpretation is that the analyst has to decide how to cope with these observations or whether to change the model, based on his/her knowledge of the data and being aware that this is a judgemental decision. In any case, the analyst should be aware of this issue and make this choice explicit.

A common misconception that we have often encountered when discussing about EVT and robustness, particularly in this example, is that after all in EVT you want to model extremes, so if it turns out in the analysis that just one of the most extreme observations is downweighted, it seems that one leaves away important information. It is fair to say that we are particularly interested in the most extreme observations, especially from a business point of view. However, from a model and statistical point of view all the observations are equal.

We have seen in the preceding examples that observations are downweighted only if, given a specific model, they do not conform to the majority of the data. (And indeed in the case of the AT&T weekly loss data in McNeil et al. (2005), p. 280, we get comparable results between classical and robust estimates.) It is perfectly legitimate to think that the most extreme observations should be fitted well, but the analyst should know that he/she has to state this preference explicitly at the level of model formulation and in the final wording.

3.4 Message 4: Not all Ψ functions are adequate

Although in the M-estimation framework the analyst is free to choose the Ψ function, the choice should be made with care. Reiss and Thomas (2001) briefly mention M-estimators for estimating the parameters of distributions. For example, for the exponential model given by $f_\beta(z) = \frac{1}{\beta}\mathrm{e}^{-z/\beta}$ with $\beta > 0$, the maximum likelihood estimator for β is the sample mean. Reiss and Thomas (2001) propose to use the function $\psi_b^{\mathrm{RT}}(x) := b\,(-\exp(-x/b) + b/(1+b))$, where b is a tuning constant. While this leads to a robust and consistent estimator, we can question whether the proposed estimator is satisfactory. At first sight the function seems quite arbitrary. Indeed Figure 5 (on the left) plots the $\psi_b^{\mathrm{RT}}(x)$ function for the exponential model with $\beta = 10$ and for different values of b. The straight line represents the classical score function for the exponential model, while others correspond to $\psi_b^{\mathrm{RT}}(x)$ for different values of b. It is apparent from these shapes that they can substantially deviate from the classical score function, already for the central values, and the analyst may have difficulties to justify the choice of the ψ function. The optimal solution (for an efficiency of about 95%) is plotted on the right. As is apparent, it is similar to the classical score function where most of the data lie, but bounded for larger values of x. (And for lower tuning constants or higher levels of β the score function is also truncated on the left, avoiding that small values bias the estimator.) Compared to the estimator defined by $\psi_b^{\mathrm{RT}}(x)$, the optimal solution has a lower maximal bias when choosing the tuning constant such that both estimators have the same efficiency at the reference model and a better efficiency at the reference model when the tuning constants are chosen such that both estimators have the same maximal bias. (However, we are not dogmatic, every analyst is obviously free to choose the ψ function of his/her choice as long as the implications are clear. We have made this point just to look a bit deeper into the construction of a particular estimator.) In general, different ψ functions may be useful in different situations.

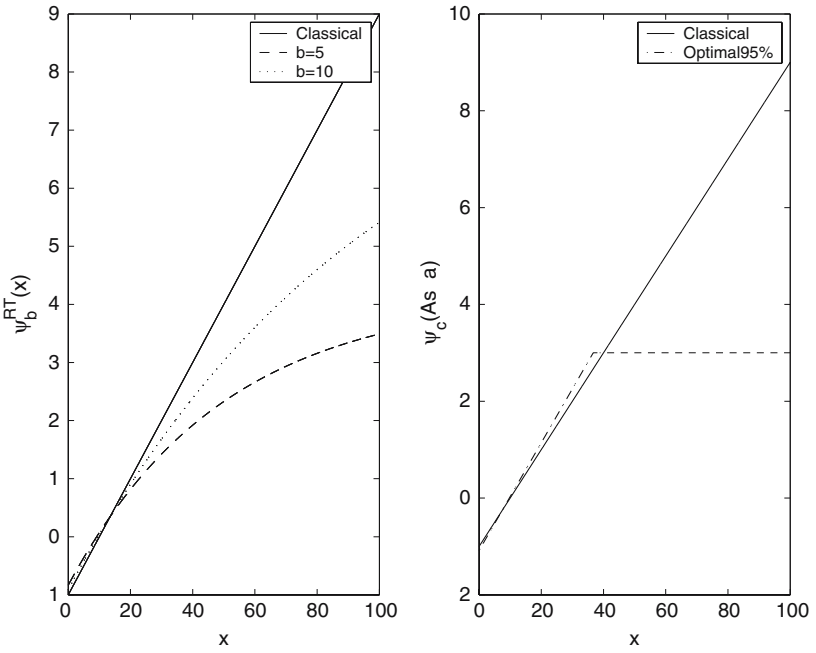

Fig. 5 Comparision of the classical score function of the exponential model with $\psi_b^{\mathrm{RT}}(x)$ as chosen in Reiss and Thomas (2001) (*graph on the left*) and with the optimal solution (*graph on the right*)

For example ψ functions that completely downweight outlying points (and correct for obtaining consistency) can be useful in the presence of a high contamination.

4 Conclusions

In this contribution we have sketched 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.

As a final note, we would like to stress that robustness issues are not *specific* to EVT. Indeed similar robustness issues arise for many other models such as linear regression, generalized linear models, multivariate

models and virtually all time series models. The robustness issues are even more severe in these cases because of the presence of exogenous variables and higher dimensions. For example an application for the estimation and testing of short rate interest rate models can be found in Dell'Aquila et al. (2003); further applications and extensions to many models used in risk management and asset allocation can be found in Dell'Aquila and Ronchetti (2006).

Acknowledgements Rosario Dell'Aquila would like to thank RiskLab for financial support.

References

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

Dell'Aquila, R., Ronchetti, E., Trojani, F.: Robust GMM analysis of models for the short rate process. J Empir Finance **10**, 373–397 (2003)

Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling extremal events for insurance and finance. Berlin Heidelberg New York: Springer 1997

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

Huber, P.: Robust statistics. New York: Wiley 1981

Moscadelli, M.: The modelling of operational risk: experience with the analysis of the data collected by the Basel Committee. Banca d'Italia, Temi di discussione del Servizio Studi, no. 517 (2004)

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

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

About the Authors

Rosario Dell'Aquila is a senior researcher at RiskLab, ETH Zürich. Previously he has worked as a senior quantitative analyst at the Zürcher Kantonalbank, as researcher at the University of Lugano and the Centro di Studi Bancari in Vezia and as a financial application developer at Swiss Bank Corporation. He has graduated in physics at ETH Zürich, has completed the first year Ph.D. program in economics at the Studienzentrum Gerzensee and holds a Ph.D. in Economics (summa cum laude) from the University of Lugano. His research interests range from applied research in risk management, asset allocation and return forecasting to methodological aspects of data analysis. He has co-authored a book with Elvezio Ronchetti on "Robust Statistics and Econometrics with Economic and Financial Applications", (Wiley, 2006, forthcoming). Besides, he teaches courses

for practitioners, including portfolio theory and asset pricing at the Executive MBA of the SDA Bocconi in Milan.

Paul Embrechts is Professor of Mathematics at the ETH Zurich. Previous academic positions include the Universities of Leuven, Limburg and London (Imperial College). Dr. Embrechts has held visiting appointments at the University of Strasbourg, ESSEC Paris, the Scuola Normale in Pisa and the London School of Economics (Centennial Professor of Finance). His research interests include insurance risk theory, quantitative risk management, the interplay between insurance and finance, and the modeling of rare events. Together with C. Klüppelberg and T. Mikosch he is a co-author of "Modelling of Extremal Events for Insurance and Finance", Springer, 1997 and with A.J. McNeil and R. Frey of "Quantitative Risk Management: Concepts, Techniques and Tools", Princeton University Press, 2005.