HEAVY-TAILED DISTRIBUTIONS AND RATING

BY

J. BEIRLANT, G. MATTHYS AND G. DIERCKX

University Center of Statistics, Katholieke Universiteit Leuven

ABSTRACT

In this paper we consider the problem raised in the Astin Bulletin (1999) by Prof. Benktander at the occasion of his 80th birthday concerning the choice of an appropriate claim size distribution in connection with reinsurance rating problems. Appropriate models for large claim distributions play a central role in this matter. We review the literature on extreme value methodology and consider its use in reinsurance. Whereas the models in extreme-value methods are non-parametric or semi-parametric of nature, practitioners often need a fully parametric model for assessing a portfolio risk both in the tails and in more central portions of the claim distribution. To this end we propose a parametric model, termed the generalised Burr-gamma distribution, which possesses such flexibility. Throughout we consider a Norwegian fire insurance portfolio data set in order to illustrate the concepts. A small sample simulation study is performed to validate the different methods for estimating excess-of-loss reinsurance premiums.

1. INTRODUCTION

The topic raised by Professor Benktander on the occasion of his 80th birthday concerning the choice of an appropriate claim size distribution in connection with a (multi-layer) rating problem is indeed a very fundamental area of discussion, both in the academic as in the practical (re-)insurance world.

On the one hand, modelling extreme events through Pareto-type and other heavy-tailed distributions attracts more and more attention. The number of statisticians working in extreme value methodology and the number of publications in this area is systematically growing; see the reference list for some recent books and papers with special emphasis on actuarial applications. Several important methods in this area were influenced by methods developed in the actuarial literature, not in the least by the paper by Benktander and Segerdahl *On the analytical representation of claim distributions with special reference to excess-of-loss distributions* (the XVIth International Congress of Actuaries, Brussels, 1960). Indeed, in that contribution the concept of the mean excess (or mean residual life) function was illuminated, which turned out to be quite a useful tool in extreme value statistics. Professor Benktander

ASTIN BULLETIN, Vol. 31, No. 1, 2001, pp. 41-62

was also one of the first to introduce the concept of probability and quantile plotting in actuarial statistical practice, which, in our opinion, is the right way to view the data with the aim of tail modelling.

On the other hand, actuaries working in a reinsurance context are sometimes feeling uneasy with this material. One of the main problems is that the statistical extreme value models concern only the *ultimate tail* section of the distribution while a practitioner faced with reinsurance rating will need to model also more *central* areas of the distribution in order to handle the different layers in a flexible way. This, we believe, leads to another important merit of the abovementioned paper by Professor Benktander: the Benktander I and II distributions offer a nice compromise between statistical flexibility and efficiency, and computational simplicity with regard to premium rates. These classes contain all popular heavy-tailed models ranging from the Pareto distributions, over lognormal-type models to Weibull-type tails. At the same time the elegant expressions of their mean excess functions makes them especially attractive for the actuarial practitioner.

In this text we present a personal view on the link between statistical extreme value methods and the selection of appropriate statistical claim size models on the one side, and actuarial concepts, in particular the mean excess function, on the other. Proposals for statistical models that are able to capture both central and tail characteristics of the distribution will be presented. Finally, recent new directions in extreme value statistics, again motivated mainly by actuarial applications, will also be discussed. In Section 2 the relation between quantile plotting and the mean excess function is explained. In Section 3 we add the connection with extreme value methods. We order the presentation of the different approaches from non-parametric techniques over semi-parametric ones to a final fully parametric model in Section 4. The implications to premium rating are clarified along the way.

Throughout the text we use the fire claim data from a Norwegian portfolio in 1990 (taken from [1]) to illustrate the different methods and to give an idea of the typical problems with claim data modelling.

2. Quantile plotting and the mean excess function

Let $x_1, x_2, ..., x_n$ be claim data that come from a random sample $X_1, X_2, ..., X_n$ with distribution function F and survival function $\bar{F}(x) = P(X > x)$, denoting the probability to obtain a claim larger than x. The ordered data will be denoted by

$$x_{1,n} \leq x_{2,n} \leq ... \leq x_{n,n},$$

which are the sample values of the order statistics $X_{1,n} \leq ... \leq X_{n,n}$.
In case the expected value of X exists, i.e. $E(X) < \infty$, the *mean excess function* is given by

$$m(x) = E(X - x \mid X > x),$$

the expected excess claim size.

This function plays a central role in the rating of an excess-of-loss reinsurance in excess of a retention or priority level R, as the corresponding risk premium $\Pi(R)$ for the layer from R to infinity is given by (a multiple of)

$$\Pi(R) = \bar{F}(R)m(R) = E((X-R)_+).$$

It is a well-known fact that the most efficient way to derive m from $\bar{F}$ is by using the expression

$$m(x) = \frac{\int_x^\infty \bar{F}(u)du}{\bar{F}(x)},$$

while the inverse operation is given by

$$\bar{F}(x) = \frac{m(B)}{m(x)} \exp\left(-\int_B^x \frac{du}{m(u)}\right), \quad \text{for } x > B,$$

where B denotes the left limit of the support of F.

In practice, the mean excess function m is easily estimated at $x = X_{n-k,n}$ for some $k = 1, \ldots, n-1$ by the (empirical) average excess of the k data points higher than $X_{n-k,n}$:

$$\hat{m}_{k,n} = \frac{1}{k}\sum_{j=1}^{k} X_{n-j+1,n} - X_{n-k,n}.$$

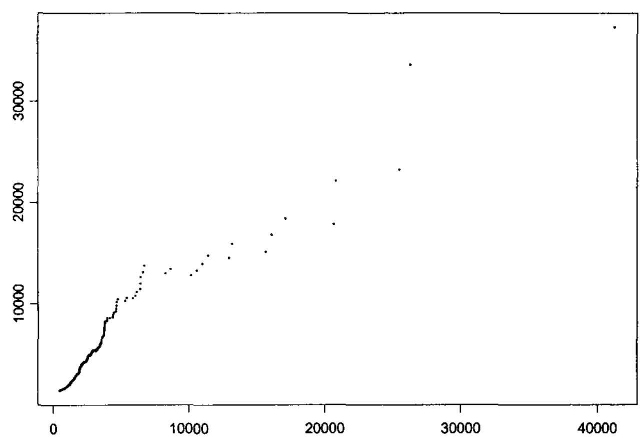

FIGURE 1: Plot of $\hat{m}_{k,n}$ as a function of $x_{n-k,n}$ for the Norwegian fire insurance data.

The *hazard rate* $\mu(x)$ defined by

$$\mu(x) = \frac{F'(x)}{\bar{F}(x)},$$

is closely linked to the mean excess function through the expression

$$\mu(x) = \frac{1 + m'(x)}{m(x)}.$$

So far for the recapitulation of the basic notions from (re)insurance mathematics. On the statistical side the potential of *quantile plotting* through quantile-quantile or QQ plots (or, alternatively, of probability plotting) for the graphical description and for the analysis of claim data has been stressed by several authors considering extreme value methods, see for instance [1], [17]. What may look innocuous or only somewhat suspect in a density comparison may become quite glaring in a QQ plot. Starting from the point of view that a heavy-tailed distribution is a distribution for which the tail is heavier than any exponential tail, i.e.

$$\lim_{x\to\infty} \frac{\exp(-\lambda x)}{\bar{F}(x)} = 0, \text{ for any } \lambda > 0,$$

the degree of deviation can be depicted through visual inspection of an *exponential quantile plot* of points with coordinates

$$\left(-\log(\frac{j}{n+1}), X_{n-j+1,n}\right).$$

Here the empirical quantiles $X_{n-j+1,n}$ appear as estimates of the unknown quantiles $Q(1-\frac{j}{n+1})$, defined as the claim levels that are surpassed in $\frac{j}{n+1}$ 100% of the cases. Hence, a straight line pattern in the exponential quantile plot will direct the practitioner to a model of the type

$$Q(1-p) = a + \frac{1}{\lambda}(-\log p)$$

for some a and $\lambda > 0$, and hence to

$$\bar{F}(x) = \exp(-\lambda(x-a)), x > a$$

i.e. an exponential model, perhaps shifted over a distance a.

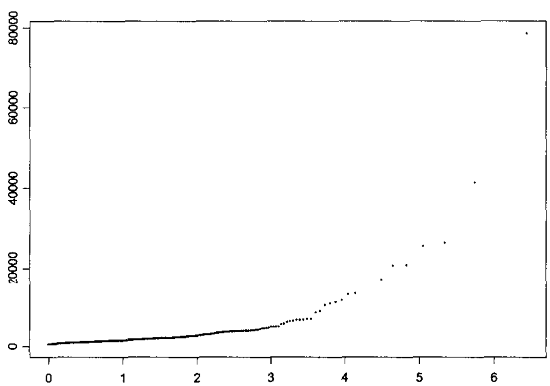

FIGURE 2: Exponential quantile plot for the Norwegian fire insurance data.

https://doi.org/10.2143/AST.31.1.993 Published online by Cambridge University Press

By the definition of the exponential quantile plot itself, namely that the vertical coordinates of the plotted positions are given by the data themselves, it follows that in case of a distribution with a tail heavier than any exponential, the plot will bend upwards away from a linear fit which is 'in line' with the exponential model. Rephrasing the expression 'bend upwards' more rigorously, we are led to stating that for such 'sub-exponential'[1] distributions the slope or the derivative of the exponential quantile plot increases as we increase the claim level.

One very naive way to estimate the slope of the exponential plot to the right of a point, say the position $\left(-\log(\frac{k+1}{n+1}), X_{n-k,n}\right)$, is to use the quotient of the average vertical and horizontal excesses over this position:

$$\frac{\frac{1}{k}\sum_{j=1}^{k} X_{n-j+1,n} - X_{n-k,n}}{\frac{1}{k}\sum_{j=1}^{k} \log\left(\frac{n+1}{j}\right) - \log\left(\frac{n+1}{k+1}\right)},$$

or, even simpler,

$$\frac{1}{k}\sum_{j=1}^{k} X_{n-j+1,n} - X_{n-k,n},$$

since the denominator of the first expression is very closely approximated by 1 (as it is an approximation of the mean excess function of the unit exponential distribution, which is constantly equal to 1).

Hence we conclude that the empirical mean excess function $\hat{m}$ defined above is a naive derivative function for the exponential quantile plot. It also follows easily that the mean excess function of distributions with tails heavier than the exponential model all have an increasing mean excess function. The strongest increase is found for Pareto distributions for which the increase is linear.

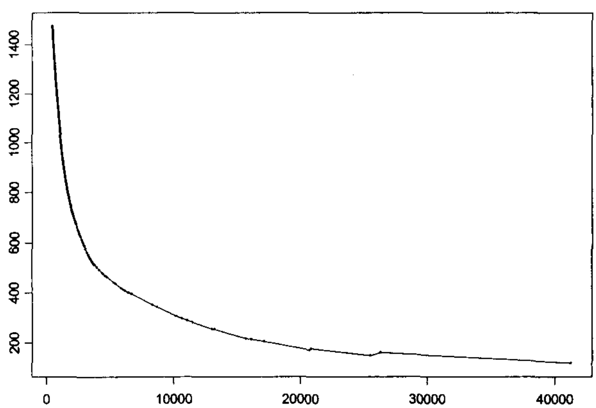

FIGURE 3: Non-parametric estimator of the excess-of-loss net premium as a function of the retention level R for the Norwegian fire insurance data.

[1] The concept of sub-exponentiality is only used here in an intuitive way, for a rigorous treatment see [13]

Further, this relationship shows that the exponential quantile plot is not only useful in the statistical validation of a claim model, but also in the calculation of a risk premium for a layer from $R = x_{n-k,n}$ to ∞, which can be estimated by $\frac{k+1}{n+1}\hat{m}_{k,n}$ in a purely non-parametric way, i.e. without assuming any parametric part in the statistical claim model.

Finally, to clarify the relationship between quantile plotting and the hazard rate μ, observe that the exponential quantile plot is the graph of the function $Q(1-e^{-x})$, which has derivative $e^{-x}/F'(Q(1-e^{-x}))$ or $(1/\mu)\,(Q(1-e^{-x}))$. Hence the reciprocal of the hazard rate follows exactly from the derivative of the exponential quantile plot at a plotting position.

3. Quantile plotting, mean excess and extreme value methods

In contrast to the previous fully *non-parametric* approach for premium calculation for an upper layer, extreme value methods typically use a *semi-parametric* approach, containing one or two parameters next to a functional part which is not specified. This seems reasonable from the fact that these methods are designed to make extrapolations outside the sample, for instance to estimate an extremely large quantile $Q(1-p)$ with $p < \frac{1}{n}$. Using a fully parametric model would then induce a second extrapolation from the sample towards the statistical population, and hence bias risk would only become larger.

3.1. Pareto-type distributions

The most famous example of such a semi-parametric extreme value model is *the Pareto-type model,* which is deduced from limit theory for the maximum $X_{n,n}$ of a sample:

$$\bar{F}(x) = x^{-\alpha}\ell(x),$$

where ℓ is a slowly-varying function (at infinity), i.e. which satisfies,

$$\frac{\ell(tx)}{\ell(x)} \to 1, \text{ as } x \to \infty, \text{ for every } t > 0.$$

Here the tail index α is the important, decisive parameter, while ℓ is a nuisance function. Working under this model amounts to assuming that the survival function behaves in first order as a power law. Examples of popular claim size models which belong to this class are, of course, the (strict) Pareto model itself (and hence the Benktander distribution with the parameter B equal to 0), next to the Burr, the generalised Pareto, the loggamma, the log-logistic and the Fréchet distribution, among others.

The estimator of α which has received by far the most attention (and still does attract a lot of research) was proposed by Hill (1975) [15] and was shown by Mason (1982) [18] to be consistent under the complete Pareto-type model, and in that sense appears to be a perfect semi-parametric estimator at first sight:

$$1/\hat{a}_{k,n} = \frac{1}{k}\sum_{j=1}^{k}\log X_{n-j+1,n} - \log X_{n-k,n}.$$

The Hill statistic is nothing else than the mean excess estimate of the log-transformed data at $X_{n-k,n}$, and hence can be deduced from a Pareto quantile plot, which is the exponential quantile plot of the log-transformed data. Indeed, under the Pareto-type model such a Pareto quantile plot can be shown to be ultimately linear with slope approaching $1/a$ above some high threshold $X_{n-k,n}$, i.e. for small enough k and large enough n.

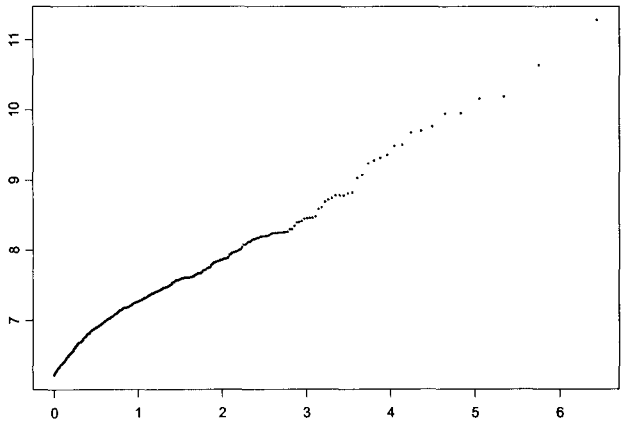

FIGURE 4: Pareto quantile plot for the Norwegian fire insurance data.

Remark that we have in fact as many estimates of a as we have data points; for each value of k we obtain a new estimate of a. Plots of $\hat{a}_{k,n}$ as a function of k are often quite volatile. In [20] it is mentioned that it is helpful to plot the Hill estimates as a function of $\log k$ (in fact, this is equivalent to using the same horizontal scale as in the Pareto quantile plot). For the Norwegian fire claim data, however, there is no apparent gain with this approach.

Several authors have tried to guide the practitioner in choosing k, leading to an adaptive choice $\hat{k}$ such that an estimate of the mean squared error of the Hill estimator is minimised at $\hat{k}$. This was done by bootstrap methods (see, for instance, [9]) or by regression diagnostics on a Pareto quantile plot in [2]. A somewhat different solution was proposed in [12]. In case of the Norwegian fire insurance portfolio the method indicated in [3] yields the value $k = 290$, which results in the estimate $1/\hat{a} = 0.62$.

Other problems are for instance the non-invariance of the Hill estimator with respect to shifts that could be applied to the data, and most importantly, the bias that the Hill estimator exhibits in certain cases. This can be understood from the fact that for certain Pareto-type distributions (as it is the case for the loggamma distribution, for instance) the influence of the slowly-varying

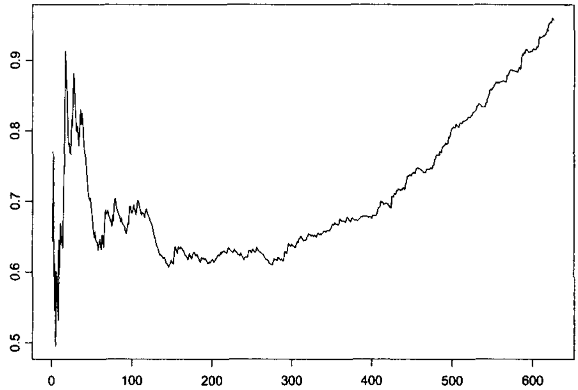

FIGURE 5: Plot of $1/\hat{a}_{k,n}$ as a function of k for the Norwegian fire insurance data.

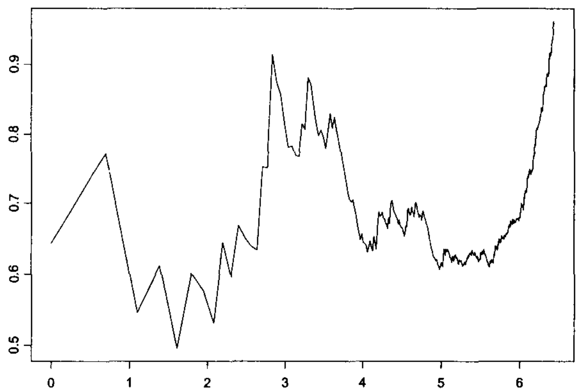

FIGURE 6: Plot of $1/\hat{a}_{k,n}$ as a function of $\log k$ for the Norwegian fire insurance data.

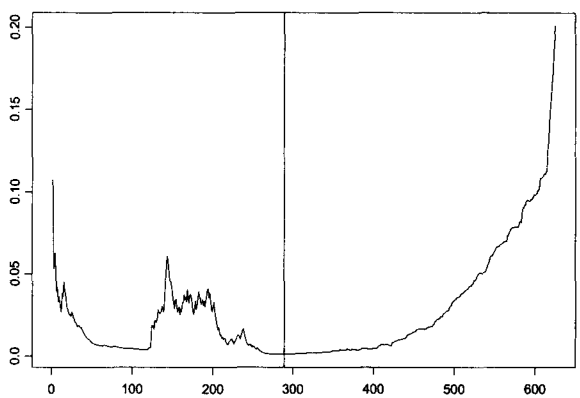

FIGURE 7: Plot of an estimator (see [3]) of the asymptotic mean squared error of the Hill estimator as a function of k for the Norwegian fire insurance data. A minimum is found at $k = 290$.

https://doi.org/10.2143/AST.31.1.993 Published online by Cambridge University Press

part ℓ is still imminent near the top end of the Pareto quantile plot. As a consequence, confidence intervals for a will not show the required coverage probability in such cases. This puts a serious restriction on the reliability of these methods.

Next to methods based on high order statistics, such as the Hill estimator, an alternative is offered by the *peaks-over-threshold approach* (POT). This method consists of fitting the *generalised Pareto distribution* (GPD) to the distribution of the excesses $Y = X - u$ (if $X > u$) over a high threshold u, for instance by maximum likelihood methods [23], the method of moments [16], or modern Bayesian estimation methods [8]. By its nature this approach has a natural link with excess-of-loss reinsurance replacing the retention level R by the statistical threshold u ; for a discussion, see [19], [22]. This approach is based on a limit result of Pickands (1975) [20] stating that as $u \to \infty$, the survival function of the excesses tends to the survival function of the GPD given by $(1 + \frac{x}{\alpha\sigma})$ with the scale parameter $\sigma = \sigma_u$ depending on u. Again, every choice of u leads to another estimator of a and of course σ_u. Smith has advised to choose $u = X_{n-k,n}$ at the smallest value of $X_{n-k,n}$ to the right of which the mean excess plot remains approximately linear as a function of the ordered data. Pure adaptive algorithmic choices have not yet been explored systematically, however. The POT method possesses some advantages over the methods based on extreme order statistics such as the ones derived from the Hill estimates: it is invariant with respect to shifts and the plots of the estimates of a as a function of k are often more stable, apparently because of the use of the second parameter σ. However, also here the asymptotic result of Pickands can set in really slowly, leading to biased estimates of a for this method too. In case of the Norwegian fire insurance data, the POT method does not lead to a more stable graph when the estimates are plotted as a function of k; see Figure 8.

Let us now consider again the estimation of the risk premium for a layer from R to ∞ with the semi-parametric approach. Using the concept of the

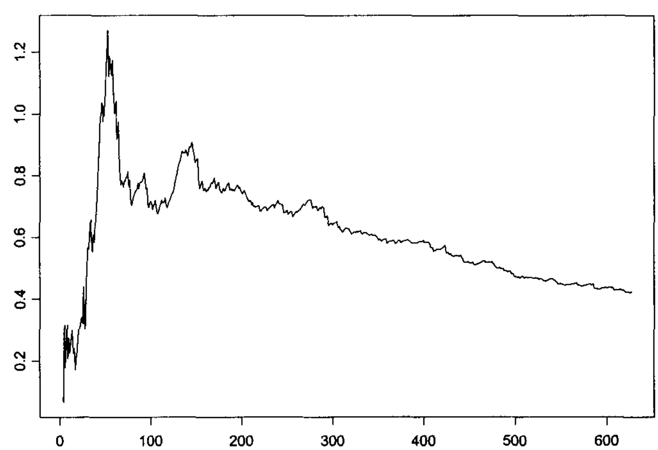

FIGURE 8: Plot of the POT estimates (*ML*) as a function of k for the Norwegian fire insurance data.

https://doi.org/10.2143/AST.31.1.993 Published online by Cambridge University Press

Hill estimator, we arrive at the following approximation based on the famous Karamata theorem (see for instance [7]) for $a > 1$:

$$\begin{aligned} \Pi(R) &= \int_R^\infty u^{-a} \ell(u) du \\ &\sim \frac{1}{a-1} R^{1-a} \ell(R) \\ &= \frac{1}{a-1} R\bar{F}(R). \end{aligned}$$

When the priority R is situated within the sample, i.e. when claims at the magnitude of R have previously been observed and R is taken equal to $x_{n-k,n}$ for some k, this leads to the estimate $\hat{\Pi}(x_{n-k,n}) = \frac{1}{\hat{a}_k - 1} x_{n-k,n} \left(\frac{k+1}{n+1}\right)$. Figure 9 presents these estimates for the fire insurance data.

If R is not fixed at one of the sample points, extreme value formulas for estimation of $\bar{F}(R)$ in the expression for $\Pi(R)$ can be applied (see for instance [1] or [12]): $\hat{\bar{F}}(R) = \left(\frac{\hat{k}+1}{n+1}\right)\left(\frac{R}{X_{n-\hat{k},n}}\right)^{-\hat{a}}$ with $\hat{k}$ denoting an appropriate adaptive choice for the number of extreme order statistics used in the procedure, which can be obtained with the methods mentioned above. This is shown in Figure 10 for our example.

Alternatively, the POT approach suggests substituting the conditional expected value of the GPD for the mean excess function at a *high* priority R (for $R > u$) : $m(R) = \frac{a_u}{a_u - 1} \sigma_u \left(1 - \frac{R-u}{a_u \sigma_u}\right)$, while $\bar{F}(R)$ will be estimated with the formula $\frac{k+1}{n+1}\left(1 - \frac{R-u}{a\sigma_R}\right)^{-a}$ when k observations exceed the threshold u. Replacing a_u and σ_u by their estimates then leads to an estimate of the risk premium as in Figure 11.

When estimating the premium at a retention level within the sample, i.e. $R = x_{n-k,n}$, one can fix the threshold u at R and then the POT approach leads to an estimate $\frac{\hat{a}_R}{\hat{a}_R - 1} \hat{\sigma}_R \frac{k+1}{n+1}$. This is of the same form as the first estimation

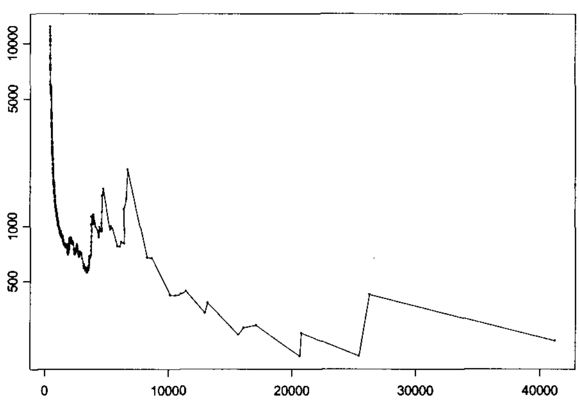

FIGURE 9: Plot of $\frac{1}{\hat{a}_k - 1} x_{n-k,n} \left(\frac{k+1}{n+1}\right)$ as a function of $R = x_{n-k,n}$ for the Norwegian fire insurance data. (log-scale on Y-axis).

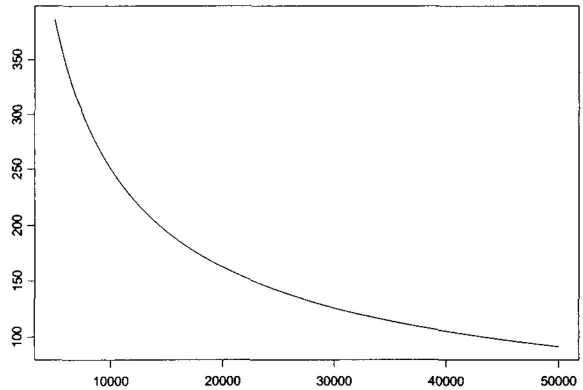

FIGURE 10: Plot of $\frac{1}{\hat{a}_k-1} R\left(\frac{\hat{k}+1}{n+1}\right)\left(\frac{R}{X_{n-\hat{k},n}}\right)^{-\hat{a}_{\hat{k}}}$ as a function of R for the Norwegian fire insurance data.

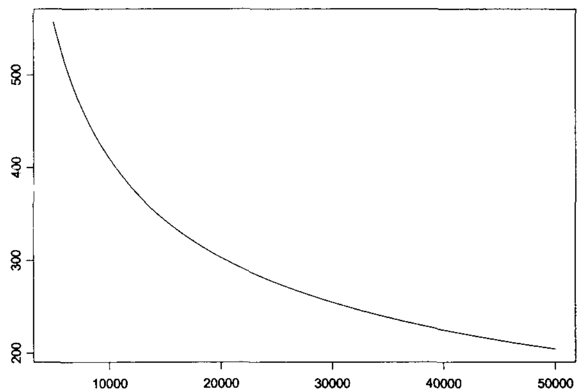

FIGURE 11: Plot of POT-based premium estimates as a function of retentions R situated beyond the threshold $u = x_{n-290,n}$ for the Norwegian fire insurance data.

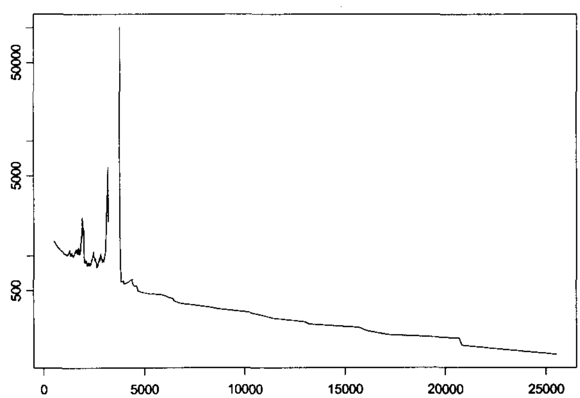

FIGURE 12: Plot of the POT based premium estimates $\frac{\hat{a}_R}{\hat{a}_R-1}\hat{\sigma}_R\frac{k+1}{n+1}$ as a function of $R = x_{n-k,n}$ for the Norwegian fire insurance data. (log-scale on Y-axis).

https://doi.org/10.2143/AST.31.1.993 Published online by Cambridge University Press

method based on the Hill estimator, replacing $x_{n-k,n}$ by $\hat{a}_R\hat{\sigma}_R$. A plot of these estimates for the Norwegian fire insurance data is shown in Figure 12.
A summary of all above estimation methods can be found in Table 1.

3.2. Bias reduction in estimating the Pareto index

The abovementioned problems with systematic biases appearing in the 'classical' extreme value methods have only recently led some authors [3], [14] to look in more detail at important (parametrised) subclasses of the set of all slowly-varying functions. The following class was first indicated by Hall (1984):

$$\ell(x) = C\left(1 + Dx^{-\beta}(1 + o(1))\right),$$

(with C, D and β denoting positive constants) to which belong for instance the Burr, the generalised Pareto and the Fréchet distribution. Another helpful subclass is given by

$$\ell(x) = C(\log x)^{\beta}(1 + o(1))$$

to which belongs for instance the loggamma distribution.

It is then shown that for k not too large the scaled logarithmic spacings $Z_j := j(\log X_{n-j+1,n} - \log X_{n-j,n})$, $j = 1, \ldots, k$, can be modelled by the following generalised regression models:

a power regression model

$$Z_j = \left(\frac{1}{a} + b_{n,k}\left(\frac{j}{k+1}\right)^{\rho}\right) f_j, 1 \le j \le k, k < n$$

with $b_{n,k}$ and $\rho(>0)$ depending on C, D and β, and $f_1, f_2, \ldots$ denoting independent and identically distributed unit exponential random variables; respectively,

a logarithmic regression model

$$Z_j = \frac{1}{a}\left(j\log\left(\frac{j+1}{j}\right) + \beta j\log\frac{\log\frac{n+1}{j}}{\log\frac{n+1}{j+1}}\right) + \varepsilon_j,\ 1 \le j \le k, k < n,$$

where ε_j denote centered exchangeable error random variables. The latter model is to be used when the parameter ρ in the power regression model is close to 0.

Again for every k another estimate of a is obtained, e.g. by joint maximum likelihood estimation of a, ρ and $b_{n,k}$, or β, but typically the plots of the estimates as a function of k are much more stable. Also, the covariate terms $b_{n,k}\left(\frac{j}{k+1}\right)^{\rho}$, respectively $\beta j\log\frac{\log\frac{n+1}{j}}{\log\frac{n+1}{j+1}}$, remove the bias of the original Hill-type estimators to a high extent. Finally, the problem concerning the non-invariance of the original estimators with respect to shifts has also been lifted up, i.e. one

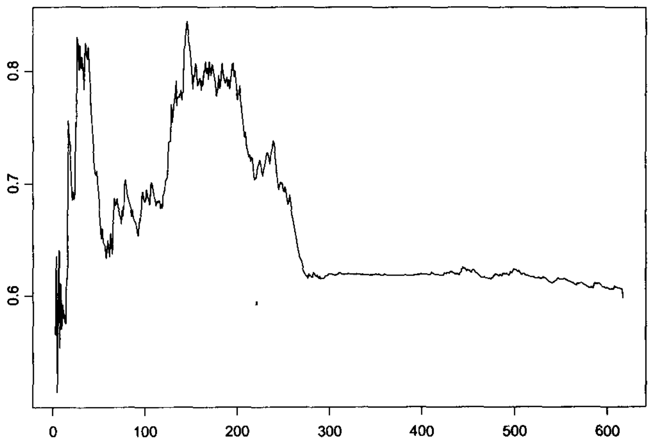

FIGURE 13: Plot of estimates of $\frac{1}{\alpha}$ based on the power regression model as a function of k for the Norwegian fire insurance data.

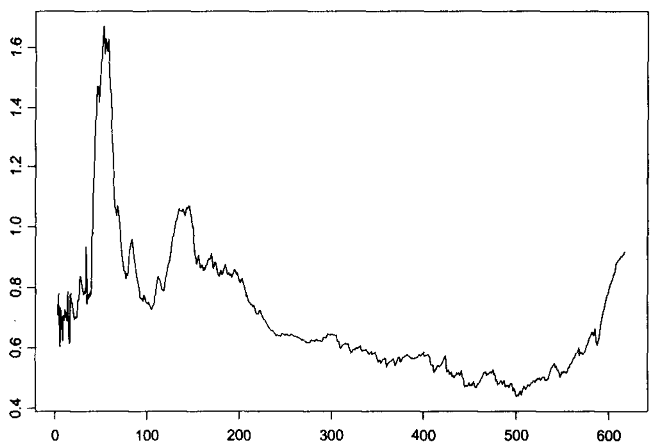

FIGURE 14: Plot of estimates of $\frac{1}{\alpha}$ based on the logarithmic regression model as a function of k for the Norwegian fire insurance data.

can add or subtract values up to the third quartile of the underlying distribution while the bias-corrected estimates remain stable. On the other hand the standard deviation has inflated in comparison with the simpler estimators but it stays of order $1/\sqrt{k}$.

Of course, a practitioner has to choose between the two estimates of a obtained by each of these two generalised regression models. In the Norwegian fire insurance example, the estimates obtained from the power regression model seem to be more stable than those from the logarithmic model. Here again the value around 0.6 appears as an estimate of $1/a$. The estimates corresponding to $k < 290$ indicate the possibility of a mixture with even a heavier

https://doi.org/10.2143/AST.31.1.993 Published online by Cambridge University Press

tail at the extreme right end of the distribution. In the whole, the logarithmic model does not appear to fit well in this case, which gives rise to a larger variability in the estimates over the range of k-values.

The different ways to estimate a premium for an excess-of loss reinsurance contract with retention R covered above, namely $\hat{\Pi}(x_{n-k,n}) = \frac{1}{\hat{a}_k - 1} x_{n-k,n}\left(\frac{k+1}{n+1}\right)$ for a retention $R = x_{n-k,n}$, respectively $\frac{1}{\hat{a}_{\hat{k}}-1} R\left(\frac{\hat{k}+1}{n+1}\right)\left(\frac{R}{X_{n-\hat{k},n}}\right)^{-\hat{a}_{\hat{k}}}$ when $R > x_{n-\hat{k},n}$, can now be recomputed replacing the Hill estimate $\hat{a}$ of a by the new estimates based on the power or logarithmic regression model. The results for the Norwegian fire claim data are given in Figures 15 through 18.

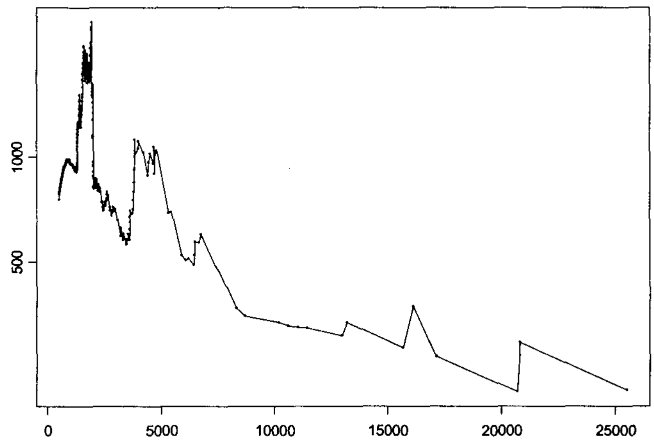

FIGURE 15: Plot of $\frac{1}{\hat{a}_k-1} x_{n-k,n}\left(\frac{k+1}{n+1}\right)$ as a function of $R = x_{n-k,n}$ with $\hat{a}$ the ML estimator from the power regression model for the Norwegian fire insurance data. (log-scale on Y-axis).

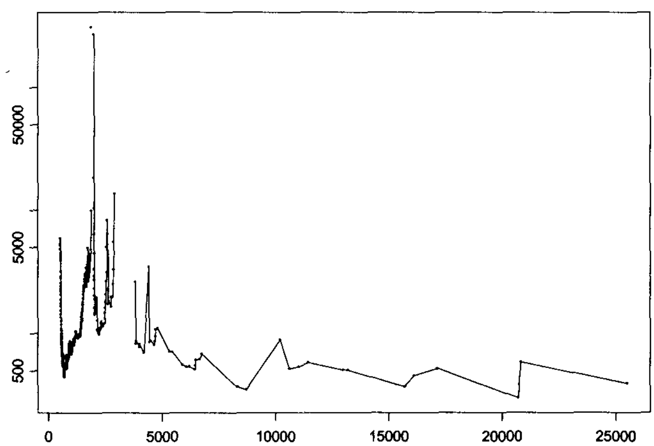

FIGURE 16: Plot of $\frac{1}{\hat{a}_k-1} x_{n-k,n}\left(\frac{k+1}{n+1}\right)$ as a function of $R = x_{n-k,n}$ with $\hat{a}$ the ML estimator from the logarithmic regression model for the Norwegian fire insurance data. (log-scale on Y-axis).

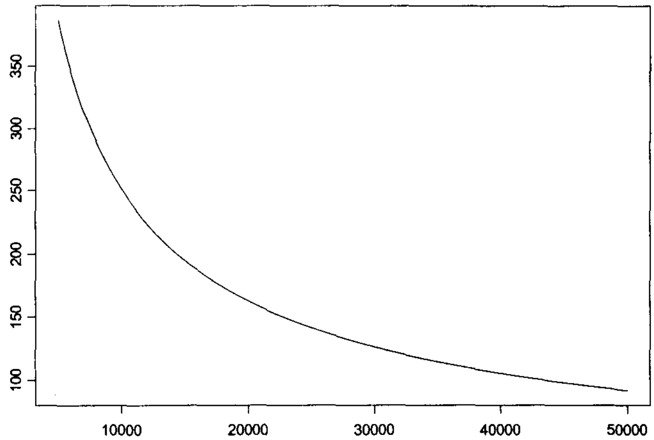

FIGURE 17: The Norwegian fire insurance data: plot of $\frac{1}{\hat{a}_k - 1} R \left(\frac{k+1}{n+1}\right)\left(\frac{R}{X_{n-k,n}}\right)^{-\hat{a}_k}$ as a function of R with $\hat{a}_k$ obtained from the power regression model.

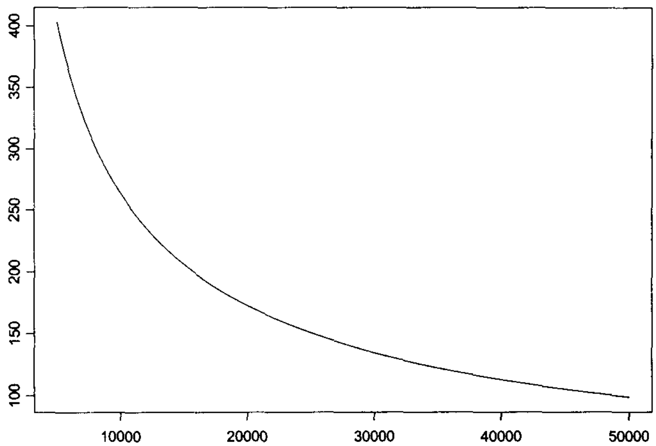

FIGURE 18: The Norwegian fire insurance data: plot of $\frac{1}{\hat{a}_k - 1} R \left(\frac{k+1}{n+1}\right)\left(\frac{R}{X_{n-k,n}}\right)^{-\hat{a}_k}$ as a function of R with $\hat{a}_k$ obtained from the logarithmic regression model.

3.3. The Gumbel maximum domain of attraction

Next to the Pareto-type models, important claim distributions such as the lognormal and the Weibull distributions (which are included in the framework of the Benktander I and II classes of distributions) have to be available in a practitioner's toolbox. Formally, this class is defined as the set of distributions for which maxima are attracted in distribution to the Gumbel distribution with distribution function $\exp(-\exp(-x))$ for large sample sizes. In extreme value methodology this group of distributions is modelled with an extension of the Pareto-type distributions through the *extreme value index* $\gamma = 1/\alpha$, defining the extreme value index γ to be 0 for this large class of distributions with exponentially fast decreasing tails. Remark that the lognormal distribution is then really on the borderline between the Pareto-type distributions and the $\gamma = 0$ class, as the first order approximation (for $x \to \infty$) of the survival function of the lognormal distribution is given by $\bar{F}(x) \sim C_1 \exp(-C_2(\log x)^2)$ for some positive constants C_1, C_2.

The difficulties encountered by the extreme value methods can be illustrated by the POT approach, for which the Gumbel class approximation is obtained formally by letting $\alpha \to \infty$ in the definition of the GPD, leading to an exponential fit $\exp(-x/\sigma_u)$. In general the goodness of fit of an exponential distribution to the excess distribution over a high threshold u will only appear to be accurate for extremely large thresholds u, which are only useful in practice for very high sample sizes.

Extensions of the Hill estimator are also available. Here we mention the *moment estimator* [10] of φ given by

$$\hat{\gamma}_{k,n}^{M} = M_{k,n}^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{\left(M_{k,n}^{(1)}\right)^2}{M_{k,n}^{(2)}}\right)^{-1},$$

where

$$M_{k,n}^{(i)} = \frac{1}{k}\sum_{j=1}^{k}(\log X_{n-j+1,n} - \log X_{n-k,n})^i, \quad i = 1,2$$

with $M_{k,n}^{(1)}$ being the Hill statistic.

An extension to the case $\gamma \geq 0$ of the graphical support that was offered by the Pareto quantile plot for the Pareto-type distributions appears to be a natural question. In [4] it was shown that in this general case, the mean residual life function m satisfies $m(Q(1-p)) = p^{-\gamma}\ell(1/p)$ for some slowly-varying function ℓ (in case $0 \leq \gamma < 1$). Hence the *the quantile – mean excess plot*, or QM plot,

$$\left(-\log\frac{k+1}{n+1}, \log\hat{m}_{k,n}\right), \quad 2 \leq k \leq n$$

will be ultimately linear with slope γ. This then leads to an estimator of γ as it was done on the basis of the Pareto quantile plot in case of $\gamma > 0$ which entailed the Hill estimator and other bias reduced estimators. So, here the message is to plot the log-transformed empirical mean excess values $\log\hat{m}_{k,n}$ against the log-scale k in order to estimate the value of γ and to capture the Pareto versus non-Pareto behaviour of the tail of the distribution: ultimately horizontal QM plots point in the direction of an exponentially decreasing tail.

This technique can be adapted so as to work without the restriction $\gamma < 1$ by replacing $\log\hat{m}_{k,n}$ in the QM plot by $\log\left(X_{n-k,n}\frac{1}{k}\sum_{j=1}^{k}(\log X_{n-j+1,n} - \log X_{n-k,n})\right)$.

3.4. Comparing the different premium calculation techniques

The different semi-parametric ways to estimate excess-of-loss premiums that are covered above are summarised in Table 1. They do yield quite different results in our case study. In order to inspect this in more detail, a small sample simulation study was performed using a Burr distribution with

$$F(x) = 1 - \left(\frac{1}{1+\sqrt{x}}\right)^2.$$

We focus on the methods developed for estimating $\Pi(R)$ when the retention satisfies $R > X_{n-\hat{k},n}$ with $\hat{k}$ chosen to minimise the mean squared error of the Hill estimator, i.e. (2), (4) and (6). The results based on the POT method (4)

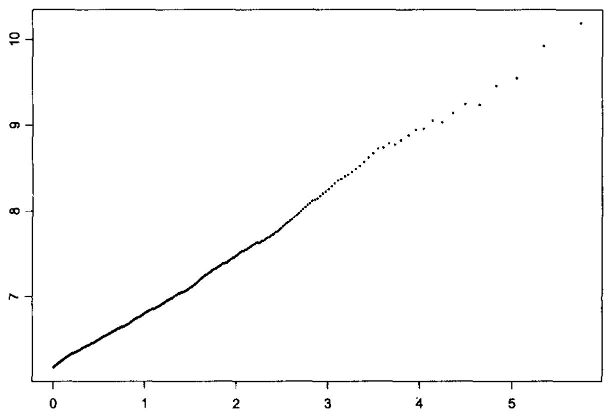

FIGURE 19: Adapted QM plot for the Norwegian fire insurance data.

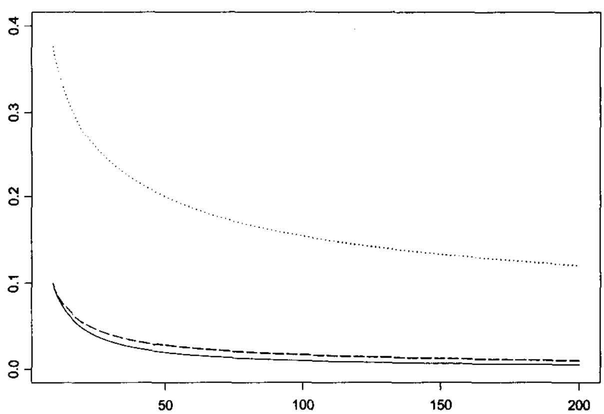

FIGURE 20: Simulation study based on 100 simulated data sets of size 500 from a Burr distribution: exact $\Pi(R)$ (solid line); median $\hat{\Pi}$(R) for methods (4) and (6) (dashed line); median $\hat{\Pi}$(R) for method (2) (dotted line).

TABLE 1

OVERVIEW OF THE DIFFERENT NON- AND SEMI-PARAMETRIC ESTIMATION METHODS FOR AN EXCESS-OF-LOSS REINSURANCE

Estimation method $\hat{a}$	Retention within sample $R = x_{n-k,n} (= u)$	Retention beyond threshold $R > x_{n-\hat{k},n} (= u)$ with $\hat{k}$ obtained by minimizing $AMSE(1/\hat{a})$
Hill estimator	$\hat{\Pi} = \frac{1}{\hat{a}_k - 1} x_{n-k,n} \left(\frac{k+1}{n+1}\right)$ (1)	$\hat{\Pi} = \frac{1}{\hat{a}_{\hat{k}} - 1} R \left(\frac{\hat{k}+1}{n+1}\right) \left(\frac{R}{X_{n-\hat{k},n}}\right)^{-\hat{a}_{\hat{k}}}$ (2)
POT ML	$\hat{\Pi} = \frac{\hat{a}_R}{\hat{a}_R - 1} \hat{\sigma}_R \left(\frac{k+1}{n+1}\right)$ (3)	$\hat{\Pi} = \frac{\hat{a}_u}{\hat{a}_u - 1} \hat{\sigma}_u \left(\frac{\hat{k}+1}{n+1}\right) \left(1 + \frac{R-u}{\hat{a}_u \hat{\sigma}_u}\right)^{1-\hat{a}_u}$ (4)
Regression model ML	$\hat{\Pi} = \frac{1}{\hat{a}_k - 1} x_{n-k,n} \left(\frac{k+1}{n+1}\right)$ (5)	$\hat{\Pi} = \frac{1}{\hat{a}_{\hat{k}} - 1} R \left(\frac{\hat{k}+1}{n+1}\right) \left(\frac{R}{X_{n-\hat{k},n}}\right)^{-\hat{a}_{\hat{k}}}$ (6)

and the regression model estimates (6) are almost identical and are in fact quite satisfactory. The simplest method based on the Hill estimator typically overestimates the correct Π function and entails a strong positive bias. In Figure 20 the median curves $\hat{\Pi}$ as a function of R are given, based on 100 simulated data sets of size $n = 500$.

4. CAPTURING CENTRAL ÁND TAIL CHARACTERISTICS

Having explained the difficulties and merits with nowadays' methods from extreme value statistics, we clearly recognise the need for completely *parametric claim models* that are capable to fit well both the tail and more central parts of the claim domain. However, fitting any such model, if existing, cannot be performed in a classical statistical way, e.g. by the use of χ^2 goodness-of-fit techniques. The parameters linked with the tail behaviour need to be estimated by methods from extreme value statistics as described above.

One such class of distributions was recently proposed in [5], termed the *generalised Burr-gamma distribution*. The distribution function is given by

$$F(x) = \frac{1}{\Gamma(p)} \int_0^{u_\xi(x)} e^{-u} u^{p-1} du$$

where

$$u_\xi(x) = \frac{1}{\xi} \log(1 + \xi u(x))$$

with $u(x) = x^{\frac{1}{b}} / \tau$.[2]

[2] In fact, in [5] $u(x)$ is modelled by $\exp\left\{\psi(p) + \frac{\log(x)+\mu}{\sigma} \sqrt{\psi'(p)}\right\}$, with ψ, resp. ψ', denoting the digamma, resp. the trigamma function. For simplicity we introduce the parameters b and τ here.

It can be seen that the parameter $b\xi$ equals the extreme value index for this parametric model. Several sub-models have appeared in the discussion above and show the flexibility of this model:

- If $\xi = 0$ then X is distributed as a generalised gamma distribution. Remark that in this case u_ξ is to be read as u and hence this model provides a generalization of the Weibull distribution

$$\bar{F}(x) = \frac{1}{\Gamma(p)} \int_{\frac{x^{1/b}}{\tau}}^{\infty} e^{-u} u^{p-1} du.$$

  The Weibull distribution is obtained choosing $p = 1$.

- If $\xi = 0$ and $p \to \infty$ this model approximates a lognormal distribution (see [5]).
- In case $\xi > 0$ we find that for p a positive integer

$$\bar{F}(x) = \left(1 + \xi \frac{x^{1/b}}{\tau}\right)^{-1/\xi} \left(\sum_{j=0}^{p-1} \frac{1}{j!} \frac{1}{\xi^j} \log^j \left(1 + \xi \frac{x^{1/b}}{\tau}\right)\right).$$

  Hence important actuarial claim models such as the Burr model (which includes the GPD) and the loggamma distribution are special cases of, or can be mimicked by this model.

How can one proceed to estimate the different parameters p, ξ, b and τ in this model? First, as $b\xi$ is the extreme value index for this model, it can be estimated with the methods discussed in the preceding section. This part of the estimation procedure is then based on a number k of extreme order statistics, i.e. the number k of highest claims in the sample, which is to be chosen adaptively as discussed above. In fact, supposing for instance that $\gamma > 0$, one finds that for this model the extreme value regression model

$$Z_j = \frac{1}{a} \left(j \log\left(\frac{j+1}{j}\right) + \beta j \log \frac{\log \frac{n+1}{j}}{\log \frac{n+1}{j+1}} \right) + \varepsilon_j, \quad 1 \le j \le k$$

holds for $k/n \to 0$ with $a = 1/(b\xi)$ and $\beta = p - 1$. This allows for estimation of $b\xi$ and p, for instance by a least-squares method based on the k highest claim data.

In Figure 21 we show the result for $\hat{p}$ for the fire claim data, which indicates the choice $p = 1$ and confirms the validity of a model without logarithmic factors. Hence, in this case the generalised Burr-gamma model reduces to

$$\bar{F}(x) = \left(1 + \xi \frac{x^{1/b}}{\tau}\right)^{-1/\xi}$$

which is in fact a Burr model. The method of moments yields the following estimates for b and τ

$$\hat{b} = 0.195,$$
$$\hat{\tau} = 8.94 \cdot 10^{14},$$

leading to an estimate $\hat{\xi} = 3.197$ for ξ.

The goodness of fit of this model is analyzed in Figure 22 using a QQ plot that shows the empirical quantiles versus the corresponding theoretical quantiles from the fitted Burr distribution. A point of inflection appears, which confirms our previous supposition of a mixture of distributions in the tail. This of course complicates the analysis.

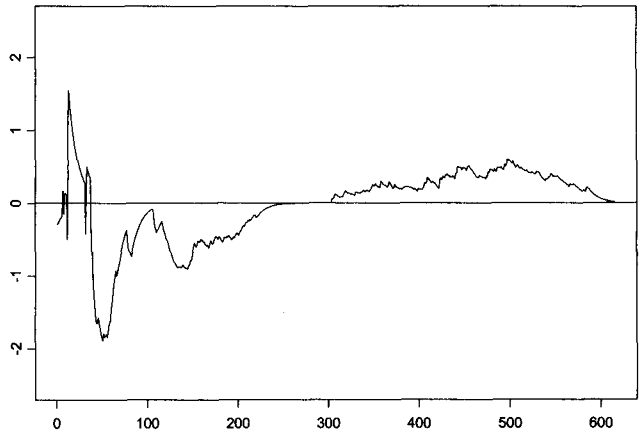

FIGURE 21: The Norwegian fire insurance data: plot of $\hat{\rho}-1$ as a function of k.

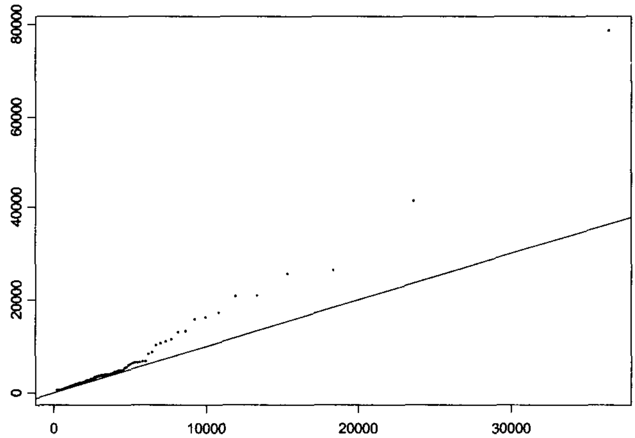

FIGURE 22: The Norwegian fire insurance data: QQ plot of empirical quantiles versus fitted Burr quantiles.

Finally, the premium $\Pi(R)$ for the fitted Burr model is easily computed numerically for different values of R. The result, given in Figure 23, is situated a bit lower than the results obtained in Figures 17 and 18. This can be understood from the fact that – partly due to the complication of the tail mixture – less weight is given to the tail section in this fully parametric analysis.

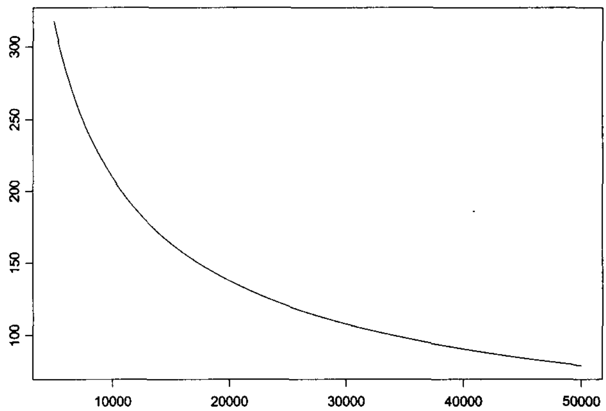

FIGURE 23: The Norwegian fire insurance data: plot of Π as a function of R based on the fitted Burr model.

5. CONCLUSION

In this paper we have tried to overview the different stages in a claim modelling process and risk premium calculation, starting with a completely non-parametric, over a semi-parametric, towards a completely parametric approach. A constant theme throughout this approach is the inspection of the tail behaviour, which is a prerequisite for accurate premium calculations, especially with reinsurance layers which cover the highest risks. Of course this discussion is certainly not the final answer but a description of the state-of-the-art in an active field of research.

REFERENCES

1. BEIRLANT, J., TEUGELS, J.L. and VYNCKIER, P. (1996) *Practical Analysis of Extreme Values.* Leuven University Press.
2. BEIRLANT, J., TEUGELS, J.L. and VYNCKIER, P. (1996) Tail index estimation, Pareto quantile plots and regression diagnostics. *J. Am. Statist. Assoc.*, **91**, 1659-1667.
3. BEIRLANT, J., DIERCKX, G., GOEGEBEUR, Y. and MATTHYS, G. (1999) Tail index estimation and an exponential regression model. *Extremes*, **2**, 177-200.
4. BEIRLANT, J., TEUGELS, J.L. and VYNCKIER, P. (1996) Excess functions and the estimation of the extreme value index. *Bernoulli*, **2**, 293-318.
5. BEIRLANT, J., DE WAAL, D.J. and TEUGELS, J.L. (2000) The generalized Burr-gamma family of distributions with applications in extreme value analysis. To appear in: *Proceedings of the 4th Conference on Limit Theorems in Probability and Statistics of the J. Bolyai Soc.*
6. BENKTANDER, G. and SEGERDAHL, C.O. (1960) On the analytical representation of claim distributions with special reference to excess-of-loss reinsurance. *Trans. 16-th Intern. Congress Actuaries*, 626-646.
7. BINGHAM, N.H., GOLDIE, C.M. and TEUGELS, J.L. (1987) *Regular Variation.* Cambridge University Press.
8. COLES, S.G. and POWELL, E.A. (1996) Bayesian methods in extreme value modelling: a review and new developments. *Int. Statist. Rev.*, **64**, 119-136.

9. DANIELSSON, J., DE HAAN, L., PENG, L. and DE VRIES, C.G. (1997) Using a bootstrap method to choose the sample fraction in tail index estimation. Technical report, Erasmus University, Rotterdam.
10. DEKKERS, A.L.M., EINMAHL, J.H.J. and DE HAAN, L. (1989) A moment estimator for the index of an extreme value distribution. *Ann. Statist.*, **17**, 1833-1855.
11. DIERCKX, G. (2000) *Estimation of the extreme value index*. Ph.D. Thesis, Dept. of Mathematics, K.U. Leuven.
12. DREES, H. and KAUFMANN, E. (1998) Selecting the optimal sample fraction in univariate extreme value estimation. *Stoch. Proc. Appl.*, **75**, 149-172.
13. EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, Th. (1997) *Modelling Extremal Events for Insurance and Finance*. Springer Verlag, Berlin Heidelberg New York.
14. FEUERVERGER, A. and HALL, P. (1999) Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.*, **27**, 760-781.
15. HILL, B.M. (1975) A simple approach to inference about the tail of a dsitribution. *Ann. Statist.*, **3**, 1163-1174.
16. HOSKING, J.R.M. and WALLIS, J.R. (1987) Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, **29**, 339-349.
17. KRATZ, M. and RESNICK, S. (1996) The qq-estimator of the index of regular variation. *Commun. Statist., Stoch. Models*, **12**, 699-724.
18. MASON, D.M. (1982) Laws of large numbers for sums of extreme values. *Ann. Prob.*, 10, 756-764.
19. MCNEIL, A. (1997) Estimating the tails of severity distributions using extreme value theory. *Astin Bulletin*.
20. PICKANDS, J. III (1975) Statistical inference using extreme order statistics. *Ann. Statist.*, **3**, 119-131.
21. RESNICK, S.I. and STARICA, C. (1997) Smoothing the Hill estimator. *Adv. Appl. Probab.*, **29**, 271-293.
22. ROOTZÉN, H. and TAJVIDI, N. (1997) Extreme value statistics and wind storm losses: a case study. *Scand. Act. J.*, 70-94.
23, SMITH, R.L. (1987) Estimating tails of probability distributions. *Ann. Statist.*, **15**, 1174-1207.