A NEW CLASS OF BAYESIAN ESTIMATORS IN PARETIAN EXCESS-OF-LOSS REINSURANCE

BY

R.-D. REISS AND M. THOMAS

ABSTRACT

For estimating the shape parameter of Paretian excess claims, certain Bayesian estimators, which are closely related to the Hill estimator, have been suggested in the insurance literature. It turns out that these estimators may have a poor performance – just as the Hill estimator – if a certain location parameter is unequal to zero in the Paretian modeling. In an alternative formulation this means that a scale parameter is unequal to 1. Thus, it suggests itself to add the scale parameter in the modeling and to deal with Bayesian estimators of the shape and scale parameters in a full Paretian model. These estimators will be applied to fire and motor reinsurance data. The performance of these estimators will be illustrated by means of Monte Carlo simulations.

KEYWORDS

Excess claims; Fire and motor reinsurance; Full Pareto model; Generalized Pareto model; Bayesian estimators; Hill estimator; Maximum likelihood estimator.

1. INTRODUCTION

We deal with an excess-of-loss reinsurance cover for the excess claims z over a certain priority u. First, let us assume that the normalized excess claims $y_i = z/u$ and their arrival times t_i up to time T can be distributionally described by a Paretian distribution function

$$F_\alpha(y) = 1 - (1 + y)^{-\alpha}, \; y \geq 0, \tag{1}$$

with unknown shape parameter $\alpha > 0$, and by a homogeneous Poisson process with unknown intensity $\lambda > 0$. Keep in mind that the mean and, respectively, the variance of a Paretian distribution does not exist if $\alpha \leq 1$ and $\alpha \leq 2$.

Given the data (t_1, y_1), ..., (t_k, y_k), the maximum likelihood estimate (MLE) of (λ, α) is

$$(\hat{\lambda}, \hat{\alpha}) = \left(\frac{k}{T}, \frac{k}{\sum_{i \leq k} \log(1 + y_i)} \right). \qquad (2)$$

Hereby, $\hat{\alpha}$ is the Hill estimator (in the original paper by Hill (1975) a random threshold, namely the kth largest order statistics, is utilized instead of a fixed threshold u). One obtains an unbiased estimator of α (cf. Rytgaard, 1990) with minimal variance (cf. Schnieper, 1993) if $\hat{\alpha}$ is replaced by

$$\frac{k - 1}{\sum_{i \leq k} \log(1 + y_i)}.$$

Another early reference to the Hill estimator, applied to excess claims, is Reiss (1989).

Next, Bayesian estimators of λ and α are defined for independent gamma priors. We merely deal with Bayesian estimators with respect to the quadratic loss function. Therefore, these estimators are expectations of posterior distributions. For the general background of Bayesian estimation we refer to the books by Hartigan (1983) and Klugman (1992).

The gamma densities $\psi_{r,c}$ and $\psi_{s,d}$ with shape parameters r, s and reciprocal scale parameters c, d are taken as priors for λ and α. Recall that

$$\psi_{r,c}(x) = \frac{c}{\Gamma(r)} (cx)^{r-1} \exp(-cx), \ x > 0.$$

The mean, variance and coefficient of variation (that is the standard deviation divided by the mean) are $r/c, r/c^2$ and $r^{-1/2}$. The mode is equal to $(r - 1)/c$ for $r \geq 1$ and equal to zero, otherwise.

The Bayesian estimates of λ and α are

$$(\hat{\lambda}, \hat{\alpha}) = \left(\frac{r + k}{c + k/\hat{\lambda}}, \frac{s + k}{d + k/\hat{\alpha}} \right) \qquad (3)$$

(cf. Hill (1975), Rytgaard (1990), Schnieper (1993) and Hesselager (1993) and the literature cited therein). It is apparent that the estimators in (2) and (3) are asymptotically equivalent as k tends to infinity. Also, these estimators are identical if we formally take $r = s = c = d = 0$ which corresponds to the use of improper priors with densities $1/x$.

Next, we study fire and motor reinsurance data in that context. These case studies will be continued in Section 4 after the preparations made in the Sections 2 and 3.

Case Study 1 (Fire Reinsurance Data). The claim sizes in Table 1 were extensively investigated by Schnieper (1993) and further discussed in Reiss and Thomas (1997). The original claim sizes are detrended, indexed and

made anonymous. Our attention will be focused again on the estimation of the shape parameter α for a Paretian modeling.

TABLE 1

CLAIMS SIZES OVER A PRIORITY OF $u = 22.0$ MILLION NKr FROM 1983 TO 1992.

Year	Claim size	Year	Claim size
1983	42.719		23.208
1984	105.860	1990	37.772
1986	29.172		34.126
	22.654		27.990
1987	61.992	1992	53.472
	35.000		36.269
1988	26.891		31.088
1989	25.590		25.907
	24.130		

Applying the Hill estimator, the value $\alpha = 2.219$ is obtained by Schnieper. As noted in Reiss and Thomas (1997), an application of the MLE, implemented in Xtremes, for the generalized Pareto model (see Section 2) yields an estimate $\alpha = 3.9$. We remark that the MLE in the Xtremes program is numerically evaluated with the moment estimator taken as the initial value. The visual insight, gained from mean excess functions, speaks in favor of the later estimate. Thus, the estimated upper tail of the claim size distribution is less heavy than found by Schnieper. A related remark holds with respect to the collective claim degree distribution with a gamma prior as presented in Schnieper (1993). This should also have some impact on the choice of the prior for the shape parameter α.

The estimates of the shape parameter α given in the next example will be confirmed to some extent by the investigations in the subsequent sections.

Case Study 2 (Motor Reinsurance Data). These claim size data are taken from Rytgaard (1990). The reinsurer has access to claim sizes exceeding $u = 1.5$ millions over a period of five years. This data set was further investigated by Hesselager (1993).

CLAIMS SIZES OVER $u = 1.5$ MILLIONS DURING FIVE YEARS.

Year	Claim size	Year	Claim size	Year	Claim size	Year	Claim size
1	2.495		1.650	3	3.215	5	19.180
	2.120	2	1.985		2.105		1.915
	2.095		1.810		1.765		1.790
	1.700		1.625		1.715		1.755

It is mentioned by Rytgaard that "For fire losses, we will usually expect an α near 1.5 – for motor liability an α perhaps near 2.5". The Hill estimate and a certain linear credibility estimate of α are 2.314 and 2.496. Hesselager specifies a gamma prior for α with parameters $s = 11.1$ and $d = 5.6$ and obtains a Bayesian estimate 2.24 of α.

In the subsequent section, we suggest to include a scale parameter $\sigma > 0$ in (1) and to compute Bayesian estimates of λ, α and σ.

It is well known that the Hill estimator is still consistent in the extended model, yet this estimator is inaccurate for larger parameters α (see, e.g., Reiss and Thomas (1997), pages 120 and 149). Because of the close relationship between the estimators in (2) and (3) one may conjecture that this remark is also relevant for the Bayesian estimator $\hat{\alpha}$ of α.

2. INCLUDING A SCALE PARAMETER IN THE PARETIAN MODELING OF EXCESSES

Our statistical model for normalized excess claims $y = z/u$ over the priority u consists of Pareto dfs

$$F_{\alpha,\sigma}(y) = 1 - (1 + y/\sigma)^{-\alpha}, \ y \geq 0, \tag{4}$$

with shape and scale parameters $\alpha, \sigma > 0$.

Next, we describe in detail the relationship between the scale parameter σ and the location and scale parameters μ and η of an initial Paretian tail. Assume that the actual claim size df $F(x)$ is close to a Pareto df of the form

$$W_{\alpha,\mu,\eta}(x) = 1 - \left(\frac{x-\mu}{\eta}\right)^{-\alpha}, \ x > \mu + \eta, \tag{5}$$

in the upper tail (that is, for all values x exceeding a higher priority $u \geq \mu + \eta$). Then, the exceedance df

$$F^{[u]}(x) = \frac{F(x) - F(u)}{1 - F(u)}, \ x \geq u,$$

is close to $W^{[u]}_{\alpha,\mu,\eta}(x) = W_{\alpha,\mu,u-\mu}(x)$ for $x \geq u$. If one deals with normalized excesses $y = z/u = (x-u)/u$ instead of exceedances x, one arrives at a df $F_{\alpha,\sigma}$ with a scale parameter

$$\sigma = 1 - \mu/u$$

as mentioned in (4). This argument also reveals why the Hill estimator – being the MLE for the submodel with $\sigma = 1$ – is consistent within the full model defined in (4) when the priority u goes to infinity as the sample size increases.

The parametrization in (4) is similar to the parametrization taken for the generalized Pareto family. The standard generalized Pareto dfs are given by

$$W_\gamma(x) = 1 - (1 + \gamma x)^{-1/\gamma}, \ 1 + \gamma x > 0, \tag{6}$$

with Pareto and certain beta dfs when $\gamma > 0$ and $\gamma < 0$, and the standard exponential df as a limiting case when γ goes to zero. In (4) we have $\alpha = 1/\gamma > 0$ and the factor γ is omitted. We refer to Falk et al. (1993), McNeil (1997), Embrechts et al. (1997) and Reiss and Thomas (1997) for more details about the generalized Pareto family.

Case Studies 1 and 2 (Continued). For the data sets presented in Section 1, the MLE for the generalized Pareto model attains a value within the present submodel of Pareto distributions. The MLE attains the values $\alpha = 3.9$, $\sigma = 2.13$ in Study 1 and $\alpha = 1.6$, $\sigma = 0.48$ in Study 2.

It becomes apparent why there can be a greater discrepancy between the Hill estimate and the MLE in the generalized Pareto model: the larger actual scale parameter puts more weight on the upper tail of the Pareto df. If such a Pareto df is estimated, yet $\sigma = 1$ is kept fixed in the modeling, then the heavier weight in the tail due to the larger actual σ must be compensated by a smaller estimated shape parameter.

The performance of the Hill estimator and the MLE in the generalized Pareto model may be characterized in the following manner: whereas the Hill estimator and related estimators have an excellent performance if $\sigma = 1$, they may have a large bias otherwise. On the other hand, the MLE in the generalized Pareto model has a smaller bias if $\sigma \neq 1$, yet always a larger variance. Therefore, it is desirable to make the Bayesian estimation principle applicable to the full Paretian model for excesses.

We make some further remarks about the Bayesian modeling in the Paretian framework. Arnold and Press (1989) consider Paretian dfs of the form $G_{\alpha,\sigma} = 1 - (x/\sigma)^{-\alpha}$, $x > \sigma$. Thus, the additional parameter σ is the lower endpoint of the support of the df and the nature of the modeling becomes completely different from ours. Bayesian inference within the generalized Pareto model is also discussed in Pickands (1994).

3. Bayesian Estimators in the Full Paretian Model

Given the pairs of excess times and excesses (t_1, y_1), ..., (t_k, y_k), one may deduce (e.g., from Theorem 3.1.1 in Reiss (1993), where densities are established in a point process framework) that the likelihood function for our joint model of Poisson times with intensity λ and Paretian marks under parameters α and σ is determined by

$$L(\lambda, \alpha, \tau) \propto (\lambda \alpha / \sigma)^k \exp\left(-(1+\alpha) \sum_{i \le k} \log(1 + y_i/\sigma)\right) \exp(-\lambda T).$$

Here, $\propto$ expresses that both sides are proportional.

The Bayesian estimators of λ, α and σ are defined for independent priors with gamma densities $\psi_{r,c}$ and $\psi_{s,d}$ as priors for λ and α and another density p as prior for σ. The posterior density is

$$h(\lambda, \alpha, \sigma) \propto L(\lambda, \alpha, \sigma) \psi_{r,c}(\lambda) \psi_{s,d}(\alpha) p(\sigma).$$

Let $r' = r + k$, $c' = c + T$, $s' = s + k$, and $d'(\sigma) = d + \sum_{i \le k} \log(1 + y_i/\sigma)$. In addition, let $\tilde{p}$ be the probability density such that

$$\tilde{p}(\sigma) \propto \sigma^{-k}(d'(\sigma))^{-(s+k)} \exp\left(-\sum_{i \le k} \log(1 + y_i/\sigma)\right) p(\sigma).$$

Check that

$$h(\lambda, \alpha, \sigma) = \psi_{r',c'}(\lambda) \psi_{s',d'(\sigma)}(\alpha) \tilde{p}(\sigma).$$

One recognizes again the well-known fact that the posterior for λ is independent of those pertaining to α and σ, yet there is a dependence between the posteriors of α and σ. The joint density $\psi_{s',d'(\sigma)}(\alpha)\tilde{p}(\sigma)$ of α and σ is represented by the marginal density $\tilde{p}$ with respect to σ and the conditional density $\psi_{s',d'(\sigma)}$ of α given σ. The conditional density is again a gamma density. This representation considerably simplifies the further computations and, also, gives some insight in the relation between the shape and scale parameters. We see no possibility to define a conjugate prior for α and σ as it was done for λ and α.

Check that the Bayesian estimates of λ, α and σ are

$$\lambda^* = \frac{r+k}{c+T},$$

$$\alpha^* = \int \frac{s'}{d'(\sigma)} \tilde{p}(\sigma) d\sigma, \tag{7}$$

$$\sigma^* = \int \sigma \tilde{p}(\sigma) d\sigma.$$

We see that the Bayesian estimator of α is just the estimator in (3), if the prior distribution – and, thus, also the posterior – is a point measure with mass equal to one at $\sigma = 1$.

Gamma priors were chosen for the parameters λ and α in the restricted model because they possess the nice property of being conjugate priors. This property still holds in the full model in so far that the conditional posterior for α is again a gamma distribution. Such a natural choice seems not to exist for the scale parameter σ.

After some Monte Carlo simulations for small sample sizes, we decided to take reciprocal gamma distributions instead of gamma distributions as priors p for σ. We remark that this is identical to taking gamma priors for the reciprocal scale parameter $\tau = 1/\sigma$ and estimating the functional parameter $T(\tau) = 1/\tau$.

Notice that a random variable X has a reciprocal gamma density

$$\tilde{\psi}_{a,b}(x) = \frac{1}{b\Gamma(a)}(x/b)^{-(1+a)} \exp(-b/x), \; x > 0, \tag{8}$$

with shape and scale parameters $a, b > 0$, if $1/X$ has the gamma density $\psi_{a,b}$. A prominent example of a reciprocal gamma density is the sum-stable Lévy density with index $a = 1/2$. The mean, variance and coefficient of variation are $b/(a-1)$, $b^2/((a-1)^2(a-2))$ and $(a-2)^{-1/2}$, if $a > 1$ and $a > 2$, respectively. The mode is equal to $b/(a+1)$.

If the hyperparameters a and b are not estimated from the data, then it can be plausible to choose these parameters such that $\tilde{\psi}_{a,b}$ is centered around $\sigma = 1$. In our applications, we take a different centering because there is a greater deviation between the value $\sigma = 1$, suggested by asymptotic considerations, and the value obtained by the MLE in the generalized Pareto model. The choice of the centering influences the Bayesian estimation to some greater extent if this method is applied to data sets of smaller sizes.

4. APPLICATIONS AND SIMULATIONS

To give a first insight, we simulate the Hill estimator under Paretian samples of size $k = 20$. The distributions of the Hill estimator are represented by kernel densities based on 4000 simulated data. The Paretian data are generated according to the distribution $F_{\alpha,\sigma}$ in (4) with parameters $\alpha = 4$ and $\sigma = 0.5, \; 1, \; 2$.

One recognizes the unfavorable performance of the Hill estimator if σ deviates from 1. This property also becomes apparent from the fact that the Hill estimator is not invariant under a scale parameter. We remark that the performance of the Hill estimator improves for smaller shape parameters α.
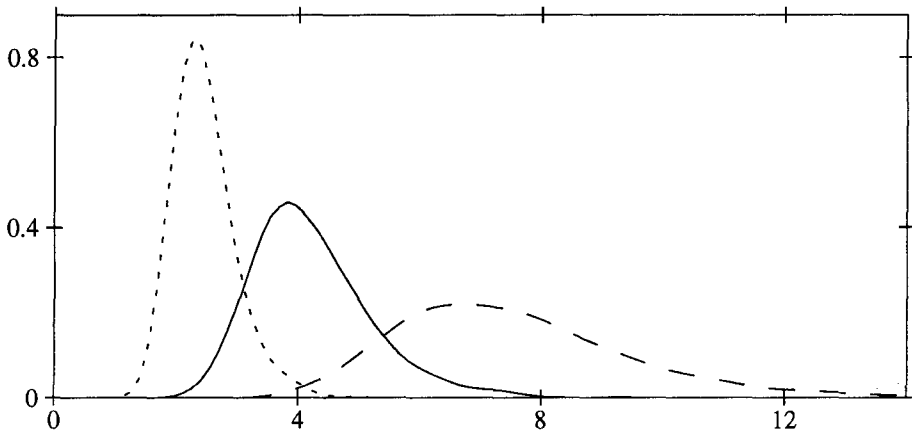
Fig. 1. Simulated densities of Hill estimators under Paretian data with $k = 20$, $\alpha = 4$, $\sigma = 2$ (dotted), $\sigma = 1$ (solid) and $\sigma = 0.5$ (dashed).

If $\sigma = 1$, then the distribution of the Hill estimator is nicely centered around the true shape parameter $\alpha = 4$. Otherwise, there is a stronger deviation from this parameter. There exist numerous research papers about the Hill estimator presumably because of the simple analytical structure of this estimator. Yet, one should carefully control the statistical modeling when this estimator is applied to real problems.

In the following case study, we also estimate the net premium

$$\frac{\lambda \sigma u}{\alpha - 1} \tag{9}$$

for the excesses over the priority u for the next period. For that purpose, we plug in the empirical estimator – also being the MLE – for the density λ, and the Bayesian estimators for α and σ. Other important functional parameters of this kind would be, e.g., the T-year initial reserve (as introduced in Reiss and Thomas (1997), Section 9.4) for a reserve (risk) process.

Case Study 1 (Continued). Corresponding to the choice of hyperparameters in Schnieper (1993), we firstly take $s = 30$ and $c = 16$. Then, the Bayesian estimator for α in the restricted model (1) has the value $\tilde{\alpha} = 1.99$. Yet as mentioned before, the collective distribution of claim degrees, as given in Schnieper (1993), indicates shape parameters centered around 4, if the evaluation is done in the full Paretian model (4). Therefore, we choose a gamma prior for α with parameters $s = 4$ and $d = 1$. Thus, the prior has the

mean 4. We remark that the median of this prior distribution is about 3.7. The variance and coefficient of variation of magnitudes 4 and 1/2 reflect the greater uncertainty in the choice of the prior for α.

There is a greater deviation between the mean, median and mode of the chosen priors for σ and, therefore, we provide all three values in the following two tables. The medians are numerically evaluated. The estimated intensity is $\lambda = 1.7$ which is the average number of excesses in each year.

In the first table, the hyperparameters a and b of the reciprocal gamma prior for σ are chosen such that the mean is equal to 2 which is close to the value 2.13 of the MLE of σ.

TABLE 3

PARAMETERS a, b, MEAN, MEDIAN AND VARIANCE OF RECIPROCAL GAMMA PRIOR; BAYESIAN ESTIMATES α^* AND σ^*, ESTIMATED NET PREMIUM (IN MILLIONS) BASED ON α^* AND σ^*.

	Mean	Median	Mode	Variance	α^*	σ^*	Net premium
$a = 2, b = 2$	2	1.2	2/3	∞	3.47	1.84	27.8
$a = 3, b = 4$	2	1.5	1	4	3.59	1.92	27.7
$a = 4, b = 6$	2	1.6	6/5	2	3.65	1.96	27.6

In the second table, the centering of the prior for σ is done by means of the median. We take hyperparameters a and b so that the median of the prior is equal to 2.

TABLE 4

PARAMETERS a, b, MEAN, MEDIAN AND VARIANCE OF RECIPROCAL GAMMA PRIOR; BAYESIAN ESTIMATES α^* AND σ^*, ESTIMATED NET PREMIUM (IN MILLIONS) BASED ON α^* AND σ^*.

	Mean	Median	Mode	Variance	α^*	σ^*	Net premium
$a = 2, b = 3.35$	3.4	2	1.1	∞	3.93	2.22	28.4
$a = 3, b = 5.34$	2.7	2	1.3	7.2	3.94	2.21	28.1
$a = 4, b = 7.32$	2.4	2	1.5	3.0	3.93	2.20	27.9

There seems to be a greater stability in the estimated values if the centering of the prior is done by the median. Particularly, the values α^* in Table 4 correspond to those of the MLE in the generalized Pareto model.

The estimated net premiums correspond to the estimated net premium 27.23 based on the MLE in the generalized Pareto model. If one takes the Hill estimate $\alpha = 2.219$ and $\sigma = 1$, the estimated net premium is 30.7. Thus,

one gets a reasonable estimate of the net premium although the estimated parameters are inaccurate (the smaller estimate of α is compensated to some extent by the smaller value $\sigma = 1$).

Our Bayesian analysis confirms that of Hesselager (1993) to some extent. This is related to the fact that, as indicated before, the Hill estimator becomes more accurate for smaller actual shape parameters.

Case Study 2 (Continued). Hesselager (1993) takes a gamma prior for the shape parameter α with parameters $s = 11.1$ and $d = 5.6$. Thus, the gamma distribution is centered around 2. The Bayes estimate in (3) is $\tilde{\alpha} = 2.24$. The values obtained from the Bayes estimators in (7) are $\alpha^* = 1.73$ and $\sigma^* = 0.53$ if the same prior is taken for α, and the reciprocal gamma prior with parameters $a = 3$ and $b = 1$ is taken for the scale parameter σ.

We conclude this paper with simulations of distributions of Bayesian estimators $\tilde{\alpha}$ and α^* under Paretian data of size $k = 20$ for a shape parameter $\alpha = 4$ and scale parameters $\sigma = 1$, 2. The hyperparameters of the gamma and reciprocal gamma priors for α and σ are $s = 4$, $d = 1$, $a = 3$
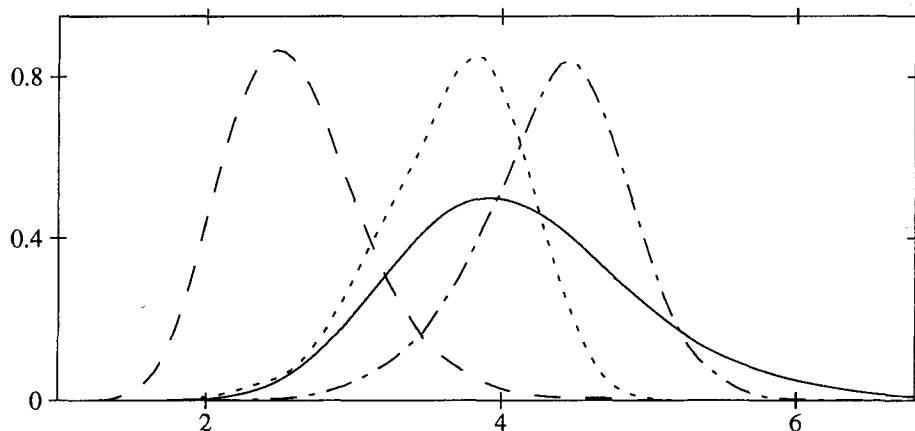

Fig. 2. Simulated densities of Bayesian estimator $\tilde{\alpha}$ for Paretian data under $\sigma = 1$ (solid), $\sigma = 2$ (dashed), and of Bayesian estimator α^* under $\sigma = 1$ (dashed-dotted), $\sigma = 2$ (dotted).

and $b = 4$. The distributions are represented by means of kernel densities based on 4000 simulated estimates.

The performance of the Bayesian estimator $\tilde{\alpha}$ in the restricted model resembles that of the Hill estimator (cf. Fig. 1). The Bayesian estimators α^* in the full model reveals a greater invariance under the variation of σ.

The Bayesian estimates in (7) are computed numerically. We use an adaptive Gaussian integration as described in Klugman (1992). All the computations are carried out with the help of the statistical software

package Xtremes. The performance of the Bayesian estimators can be extensively simulated with the help of the program bayessim.sp stored in http://www.xtremes.math.uni-siegen.de/spprograms.htm.

REFERENCES

ARNOLD, B.C. and PRESS, S.J. (1989). Bayesian estimation and prediction for Pareto data. JASA **84**, 1079-1084.

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events*. Springer, New York.

FALK, M., HÜSLER, J. and REISS, R.-D. (1993). *Laws of Small Numbers: Extremes and Rare Events*. DMV Seminar 23, Birkhäuser, Basel.

HARTIGAN, J.A. (1983). *Bayes Theory*. Springer, New York.

HESSELAGER, O. (1993). A class of conjugate priors with applications to excess-of-loss reinsurance. ASTIN Bulletin **23**, 77-90.

HILL, B.M. (1975). A simple general approach to inference about the tail of a distribution. Ann. Statist. **3**, 1163-1174.

KLUGMAN, S.A. (1992). *Bayesian Statistics in Actuarial Science*. Kluwer, Boston.

McNEIL, A.J. (1997). Estimating the tails of loss severity distributions using extreme value theory. ASTIN Bulletin **27**, 117-137.

PICKANDS, J. III (1994). Bayes quantile estimation and threshold selection for the generalized Pareto family. In: Extreme Value Theory and Applications, 123-138, J. GALAMBOS et al. (eds.), Kluwer, Dordrecht.

REISS, R.-D. (1989). Statistical inference based on large claims via Poisson approximation. Part II: Poisson process approach. Blätter DGVM **XIX**, 123-128.

REISS, R.-D. (1993). *A Course on Point Processes*. Springer, New York.

REISS, R.-D. and THOMAS, M. (1997). *Statistical Analysis of Extreme Values*. Birkhäuser, Basel (with Xtremes on CD-ROM).

RYTGAARD, M. (1990). Estimation in the Pareto distribution. ASTIN Bulletin **20**, 201-216.

SCHNIEPER, R. (1993). Praktische Erfahrungen mit Grossschadenverteilungen. Mitteil. Schweiz. Verein Versicherungsmath., 149-165.

ROLF-DIETER REISS and MICHAEL THOMAS
FB 6, Universität Siegen
D-57068 Siegen
Germany