ISSN 1440-771X

MONASH University

Australia

Department of Econometrics
and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

ADAPTIVE PREMIUMS FOR EVOLUTIONARY CLAIMS IN NON-LIFE INSURANCE

Roger Gay

November 2004

Working Paper 25/04

ADAPTIVE PREMIUMS FOR EVOLUTIONARY CLAIMS IN NON-L IFE INSURANCE

Roger Gay
Accounting and Finance
Monash University,
Wellington Rd. Clayton,
Australia, 3800
Email: roger.gay@buseco.monash.edu.au

ABSTRACT

Rapid growth in heavy-tailed claim severity in commercial liability insurance requires insurer response by way of flexible mechanisms to update premiums.
To this end in this paper a new premium principle is established for heavy-tailed claims, and its properties investigated.
Risk-neutral premiums for heavy-tailed claims are consistently and unbiasedly estimated by the ratio of the first two extremes of the claims distribution.
That is, the heavy-tailed risk-neutral premium has a Pareto distribution with the same tail-index as the claims distribution. Insurers must predicate premiums on larger tail-index values, if solvency is to be maintained.
Additionally, the structure of heavy-tailed premiums is shown to lead to a natural model for tail-index imprecision (demonstrably inescapable in the sample sizes with which we deal). Premiums which compensate for tail-index uncertainty preserve the ratio structure of risk-neutral premiums, but make a 'prudent' adjustment which reflects the insurer's risk-profile. An example using Swiss Re's (1999) major disaster data is used to illustrate application of the methodology to the largest claims in any insurance class.

KEYWORDS: INSURANCE CLAIMS, PREMIUMS, TAIL-INDEX, EXTREME VALUES

JEL classification G22, IM classification 30

Introduction: objectives and motivation

The economic imperative
In an important report and policy document released by Swiss Re "*The economics of liability losses – insuring a moving target*" Enz and Holzheu (2004) have drawn attention to the alarming rate of increase of commercial liability insurance claims. The problem is most acute in the US where long term estimates suggest claims are growing 1.5 to 2 times as fast as nominal GDP. The relative cost of commercial liability claims in the US is currently about 0.64% of GDP, about three times the level seen in Europe's largest economies.
The report signals the need for a flexible mechanism for rapid adjustment of liability premiums to enable insurers to remain solvent, and for commercial liability insurance systems to remain in place and affordable.

The pragmatic imperative
In the introduction to their excellent book, Embrechts *et al.* (1997) declare

" It is all too easy for the academic to hide constantly behind the screen of theoretical research: the actuary or finance expert facing the real problems has to take important decisions based on the data at hand."

In response to these two imperatives, this paper proposes a means of updating premiums for commercial liability insurance on an annual basis, to augment, not replace, existing statistical procedures.

Specifically, we seek to provide a convincing answer to the question 'For the most costly classes of commercial liability insurance, to what extent can insurance premiums for *next* year's k largest claims be determined on the basis of *this* year's k largest claims?' We have in mind a value of k no greater than about 20; but for the relevant classes of insurance, 20 claims will cover the most significant costs.

Claims are assumed to arise from heavy-tailed distributions with a mean but no variance, so not subject to conventional central limit theorems.
The objective is to provide an apparatus for independent assessment of the impact of the previous year's largest claims, held apart, so to speak from the ongoing trend in claim severity.

Premium-setting for the largest fat-tailed general insurance claims based on a small sample poses two distinct statistical problems.

<u>Problem 1</u> is determination of a suitable *premium principle* for fat-tailed claims; that is, formulation of a general procedure by which premiums with appropriate properties can be calculated.

<u>Problem 2</u> comprises the very considerable technical difficulties which beset estimation of the tail-index of the any claims distribution, especially when sample sizes are small. Standard analyses of tail-index error derived from asymptotic results are ruled out, so that any tail-index value used to set premiums is necessarily imprecise.

In respect of Problem 1, we show that there is a natural premium principle ('the power principle' for fat-tailed claims delivering a premium which:

- has the 'usual' properties of a good premium (outlined below),
- characterises the claims distribution in terms of its first two extremes,
- admits a consistent unbiased estimator (with, it transpires, a Pareto distribution with the same tail-index as claims).

The structure of the premium obtained from the power principle leads naturally to a particular model for tail-index uncertainty, which can then be used to construct risk-averse premiums which compensate for imprecision in knowledge of the tail-index.

- *By modelling imprecision in tail-index knowledge we are also able to clarify the nature of implicit judgements premium-setters make when resorting to pragmatic rules adopted in the face of tail-index uncertainty.*

- *In particular, it is shown that an insurer cannot remain risk-neutral in the face of tail-index uncertainty*

For Problem 2, a new methodology is developed for setting premiums for next year's k largest extreme losses, based only on this year's set of extremes.
This methodology is illustrated using Swiss Re's ten largest man-made insured losses of 1999 (major fires and explosions).
That premiums are themselves estimates (and so random variables) is frequently overlooked. The random nature of premiums for heavy-tailed extreme claims is made quite explicit; its exact distribution is determined.
Further information about the dangerousness of the tail generating a particular class of claim is provided by *the expected new record loss,* and its approximate confidence interval.

The paper is divided into the following sections.

1: Premium principles

Problem 1 We briefly survey premium principles for both thin and fat-tailed insurance claims; this serves to introduce in context, the power principle for fat-tailed claims. Properties of power principle premiums are outlined.

2: Estimation of the tail-index

Problem 2 We look briefly at models used for large claims, dedicated estimators of the tail-index, and maximum likelihood estimation. Results of a simulation example demonstrate the difficulties attaching to use of small samples to elicit information about heavy tails. Estimation and error assessment problems are discussed.

3: The nature of risk-neutral fat-tailed premiums

The power principle is applied to heavy-tailed claims and features of risk-neutral premiums are discussed. The premiums are shown to be ratios of the two largest expected extremes of the claim distribution, consistently and unbiasedly estimated by the ratio of actual extremes. The distribution of the ratio of extremes i.e. the premium, is shown to be Pareto with the same tail-index as the claims.
By implication, the long-term solvency of commercial liability system providers depends upon their setting premiums based on a larger tail-index than the claims generator.

4: Modelling tail-index imprecision; risk-averse premiums

The structure of the heavy-tailed premium is shown to lead to a natural model for tail-index uncertainty. Risk-averse premiums are constructed as weighted mean values or alternatively as Bayesian posterior means. Use of such risk-weighted premiums is shown to be essentially equivalent to making a 'prudent' adjustment (i.e. based on each insurer's risk profile) to tail-index used for risk-neutral premiums.
The consequence of an insurer remaining risk-neutral in the face of tail-index uncertainty is discussed.

5: Setting premiums for the largest extreme claims

The methodology is most transparent when the pattern of extreme claims is fairly pronounced. An illustration using Swiss Re (1999) data 'man-made disasters; 10 largest major fires and explosions' which exhibits a fairly strong pattern enables convincing premium determination and provision of approximate confidence intervals for the largest expected claims.

6: The expected new record disaster claim
The dangerousness of the tail is further revealed by the value of the expected new record loss and its approximate confidence interval.

7. Summary and conclusions

8: Appendix with mathematical derivations

9: References

1. Premium principles

1.1 Survey of premium principles

In general insurance markets, insurers are price setters. Markets are non-arbitrabeable (Albrechts, 1992), the competitive markets paradigm of Black and Scholes (1973), the starting point of so much of modern finance not being directly applicable.
Bladt and Rydberg (1998) provide an alternative methodology for setting insurance premiums as an option price, without market assumptions. Our focus however, is on the traditional approach.

In the traditional approach annual aggregate claims S_N for a given class of insurance derive from independent identically distributed random variables $\{X_i\}$, $(X_i \geq 0)$ with common distribution $F(\cdot)$, i.e.

$$S_N = X_1+X_2+\cdots+X_N \tag{1.1}$$

where N is the claims number, independent of $\{X_i\}$. The expected value $E[X_i]$ is assumed to exist, but in this paper, not necessarily $E[X_i^2]$

Part of the attraction of setting a premium as an option price is that the claims number distribution does not need to be formulated; the option premium determines the cost of bearing risk for each and every claim.

We adopt the traditional approach because no convincing general model of the evolution of a claim over time, essential to an option-based treatment is available. The most basic premium principle applying to any realized claim X_i $(\equiv X)$ is to set premium P as

$$P = E[X]+ W \tag{1.2}$$

where W is a loading which does some or all of the following:

(i) takes account of the *variability* (at least) of claims
(ii) in probability, enables a reserve to accumulate
(iii) reflects the insurer's risk preferences, and
(iv) compensates the insurer for bearing risk.

So called 'office premiums' take further account of costs of administration, marketing, management, government charges and so on.

In this paper we distinguish between thin- and fat-tailed claims as follows:

Thin-tailed claims: X has moments of all orders.

Fat-tailed claims: $E[X^r]$ only exists for $r < \delta$.

Even this dichotomy is not uncontroversial. Notwithstanding, it is usual to describe distributions $F(x)$ generating fat-tailed claims X as belonging to class $\mathbf{F}$ for which

$$1- F(x) = L(x)x^{-\delta} \; (x > 0, \delta > 0) \tag{1.3}$$

where $L(x)$ is 'slowly varying at infinity' ($L(ax)/L(x) \to 1$ as $x \to \infty$, $(a > 0)$). (See for instance, Feller, (1971), p.278, Embrechts et al. (1997), p.131).

For existence of $E[X]$, $\delta > 1$ is needed. If $\delta > 2$, $E[X^2]$ exists and annual aggregate claims are subject to central limit theorems for the usual claims number distributions. (See for instance, Mikosch (1997).

A classic result for aggregate claims (1.1) is that when constituent claims $\{X_i\}$ arise from (1.3), and when the sum is large, it is likely to be mainly attributable to the largest claim.
Because for large x

$$\Pr[\text{largest of } \{X_i \text{ of } N\} > x] = \underset{N}{E}\ [1-\{1-L(x)x^{-\delta}\}^N]$$

$$\approx \sum p_n \times nL(x)x^{-\delta} + o(x^{-\delta})$$

$$= E[N]\times L(x)x^{-\delta} + o(x^{-\delta})$$

On the other hand,

$$\Pr[S_N > x] = \Pr[\sum p_n F\star^{(n)} > x]$$

where $F\star^{(n)}$ represents the convolution of any n of the $\{X_i\}$, p_n the probability of the realization of n claims. Thus

$$\Pr[S_N > x] \sim \sum np_n L(x)x^{-\delta} \qquad \text{(using Feller, 1971, p.279)}$$

$$= E[N]\times L(x)x^{-\delta}$$

i.e.

$$\Pr[\text{largest of } \{X_i \text{ of } N\} > x] \sim \Pr[S_N > x] = E[N]\times L(x)x^{-\delta}$$

(c.f. Feller, 1971, p288, Problem 31, where the result is posed for Poisson N).

Premium (1.2) gives no clue as to how W is to be determined. Other general premium principles which result in premium form (1.2) *do* provide some direction in this matter.

One such general principle employs a *pricing function* $m_\alpha(x)$ and a *pricing rule* (Gay, 2004a) which determines premium P for claim X via

(1.4) $$m_\alpha(P) = E[m_\alpha(X)]$$

where α is a 'risk' parameter. Some special cases are:

(i) *Risk-neutral premiums*; $P = E[X]$ follows from (1.4) using $m_\alpha(x) = x$.

(ii) The important *exponential principle* (Rolski et al, (1999), p.80) for thin-tailed claims X possessing a moment generating function $M_X(\cdot)$ determines P as

(1.5) $$P = \{\ln M_X(s)\}/s$$

Premium (1.5) derives from pricing rule (1.4) using $m_s(x) = \exp(sx)$.
In this case risk parameter s is a measure of the (constant) absolute risk-aversion of the insurer (Pratt, (1964), Arrow (1971) Gay (2004b)).

Bowers et al (1986), p.7 infer that *insurers are risk-neutral or only mildly risk-averse*.

This means that s is small and (1.5) leads to

(1.6) $$P = \mu + \tfrac{1}{2}s\sigma^2 + o(s)$$

- *'the variance principle'* (Rolski et al, (1999)). Here $\mu = E[X]$, $\sigma^2 = Var[X]$

In fact if the claims number N of (1.1) also possesses a moment generating function, the exponential pricing function can be applied to annual aggregate claims (1.1) since the moment generating function of S_N is $M_N[\ln\{M_X(s)\}]$, and this leads to some satisfyingly transparent expressions for the aggregate premium in various special cases(e.g. $\{X_i\}$ negative exponential, N Poisson, full and stop-loss insurance)

Equation (1.6) is much more informative about loading W to be added to the mean for each realised claim. It depends on variability of claims through claims variance σ^2, and the insurer's risk preference via its measure s of constant absolute risk-aversion.

Insurer compensation for bearing risk increases therefore, with both s and σ^2. Because the premium is larger than the mean claim μ, and because the claims are subject to central limit theorems, reserves will, in probability accumulate.

Variants of (1.6) are:

the *'standard deviation principle'* whereby premium P is set as: $P = \mu + \kappa\sigma$,

and

the *'modified standard deviation principle'* where P is set as: $P = \mu + \kappa\sigma/\mu$

(both principles given in Rolski et al.(1999), p.80).

These principles for thin-tailed claims specify the *premium in terms of the first two moments of the distribution.*

(iii) The *quantile principle* (e.g. Rolski *et al*.(1999), p.82) sets premium P for claim X from distribution $F(\cdot)$ using a suitable quantile of the distribution satisfying for instance

(1.7) $$F(P) = (1+\theta)^{-1/\theta}$$

Equation (1.7) follows from (1.4) using the pricing function $m_\theta(x) = F^\theta(x)$, $(\theta \geq 0)$. Note that $(1+\theta)^{-1/\theta}$ is monotonically increasing with θ with $e^{-1} < (1+\theta)^{-1/\theta} < 1$, (see Gay (2004a)). The quantile principle applies both to thin- and fat-tailed risk.

(iv) *Fat-tailed claims*
Claims X arising from distributions from class **F** are characterised by tail-index δ (often ρ (or γ) = $1/\delta$). The very description (1.3) of the class **F** in terms of δ suggests that the importance of any other distributional parameters pales into insignificance

compared with that of δ. This is particularly so when $\delta \in (1,2]$ ($\rho \in [\frac{1}{2}, 1)$, the focus of this paper when claims have no second moment.
Distributions with tail-index in this range are used to model large claims arising in natural/man-made disaster, public liability, professional indemnity insurance and the like.
The variance principle, standard deviation principle, the modified standard deviation principle (and of course the exponential principle) are not applicable to claims with no second moment.

Indeed, application of the quantile principle is hazardous in these circumstances. Reasonably precise knowledge of the tail-index value is required to set premiums under the quantile principle. But estimation of δ is notoriously difficult, particularly with small sample sizes (simulation results and references are given below). One cannot ensure, using an estimate of δ, that chosen P exceeds E[X]. Premium P needs to be chosen using (1.7) with $\theta > \theta_\mu$ where

$$F(\mu) = (1+\theta_\mu)^{-1/\theta_\mu} \tag{1.8}$$

But the mean is 'a rare event' as $\delta \downarrow 1$ in that $\Pr(X > \mu) \to 0$, μ ultimately being larger than any quantile!
Use of a suitable pricing function $m_\alpha(x)$ and pricing rule (1.7) leads to a new pricing principle for fat-tailed claims which we now describe.

1.2 A new principle for fat-tailed claims: the power principle

The pricing function $m_\alpha(x) = x^{\alpha+1}$ with pricing rule (1.4) leads to premium P for claims arising from distributions $F(\cdot)$ of (1.3) as

$$P(\alpha) = \{\mu'_{\alpha+1}\}^{1/(\alpha+1)}, \quad (\alpha+1 < \delta) \tag{1.9}$$

Premium $P(\alpha)$ depends on what moments claim X *does* possess, the risk parameter α measuring *constant relative risk-aversion* in the insurer (Pratt, (1964), Arrow (1971)). A more complete rationale for (1.9) is provided in Gay (2004b).

1.3 Properties of $P(\alpha)$

(i) $P(\alpha)$ is risk-neutral for $\alpha = 0$, i.e. is the mean; $P(\alpha)$ increases as α increases (see for instance Puri and Sen, (1971), p.12) and reflects increasing risk aversion in the insurer.
Additionally in obvious notation, P has the following attributes of a good premium (see for instance, Rolski *et al*, (1999, p.79):

(ii) $P_{X+Y} \leq P_X + P_Y$ (premiums can't be reduced by splitting risks)

(iii) $P_{aX} = aP_X$ (proportionality)

(iv) $P_a = a$ (no unjustified safety loading)

(v) If X stochastically greater than Y, $P_X \geq P_Y$

Note that $\mu'_{\alpha+1}$ is the *Mellin Transform* of f(x) for which extensive tables exist (e.g. Oberhettinger, (1974)).

1.4 Application: Pareto premiums

The Pareto distribution with $F(\cdot)$ such that

(1.10) $$1 - F(x:\delta,\lambda) = (1+x/\lambda)^{-\delta} \quad (x > 0, \delta > 0)$$

is used to model large claims. Not only because it is one of the simplest fat-tailed models, but also because, for all claim distributions F from **F**, Balkema and De Haan (1974) and Pickands (1975) proved that it provides a good approximation to the conditional distribution $F_t(x)$ where

(1.11) $$F_t(x) = P(X \leq t+x|X>t) = \frac{F(t+x)-F(t)}{1-F(t)}$$

governs behaviour of 'exceedances over threshold t'.
Distribution (1.11) is the heavy-tailed version of the Generalised Pareto Distribution (GPD); (for a more precise statement of exact convergence results see Drees, Ferreira and de Haan (2004)).
In general δ is independent of t, but λ is not. Leadbetter (1991) provided conditions under which λ also is independent of level t.
For $\delta > (\alpha+1)$ application of the power principle (1.9) to Pareto claim $(1+X/\lambda)$ leads to premium

(1.12) $$P(\alpha) = \{1-(\alpha+1)/\delta\}^{-1/(\alpha+1)}$$

for insurer with constant relative risk-aversion α for the class of insurance generating claim X (dealing with a claim in the form $(1+X/\lambda)$ fulfils our purposes, and leads to results in a more convenient form than dealing directly with claim X).

It is convenient to re-parameterise premium (1.12) in the form

(1.13) $$P_\beta = (1-\rho/\beta)^{-\beta}$$

where now $\beta = 1/(\alpha+1)$ determines for each particular risk, the insurer's level *of risk-tolerance* as the heaviest tail ($\beta = \rho_{max}$) with which the insurer will deal.

2. Estimation of the tail-index

Mikosch (1997) stated "*Statistical analyses of large claim data are based on extreme value theory and related methods. These methods are known to be very sensitive with respect to the tails of the distribution, and therefore the existence of one very large claim may justify the fit of a Pareto instead of a lognormal distribution, say*."

While Teugels and Vanroelen (2004) declare "*The estimation of the extreme value index is the first and main statistical challenge"*

2.1 Estimators of the tail-index

Direct tail-index estimators due to Hill (1975), Pickands (1975), Hall (1982), Smith (1987), Dekkers, Einmahl and De Haan (1989), Feuerverger and Hall (1999) are popularly used, perhaps in tandem with bootstrap methods (Draisma *et al*, (1999)). *An inherent dilemma besets these methods*; large variability when the number of extremes is too low, large bias when it is too large. A survey of estimators is provided in Grimshaw (1990), of bias problems by Beirlant *et al*, (1996). An approach to bias reduction via Box-Cox transforms is outlined in Teugels and Vanroelen (2004). Choice of an appropriate *sample fraction* to reduce bias is discussed for example in Dekkers and de Haan (1993), Pickands (1994), Drees and Kaufmann (1998), Danielsson et al (2001) and many of the references cited by these authors.

2.2 A simulated estimation example

Imprecision in knowledge of the tail-index, especially when estimated from small samples is inescapable (Beirlant *et al,* (1994), Huisman *et al* (2002)). The simulation results below clearly illustrate the problem, and give an indication of the level of estimation performance to be expected from tail-index estimators using small samples (the exercise can easily be replicated and checked on a standard PC or laptop).

Simulation exercise
A large number (100,000) samples of size n = 100 were generated from Pareto distribution

(2.1) $$F(x) = 1-(1+x)^{-\delta}$$

with $\delta = 1.1$

Three tail-index estimators for δ were used:

(i) the maximum likelihood estimator (MLE)

(ii) a method of moments estimator in conjunction with a Box-Cox power transformation (see Teugels and Vanroelen, (2004), referred to as TVBC).

(iii) Hill's estimator

Starting with 100,000 samples of fixed size n = 100 from distribution (2.1) with $\delta = 1.1$ (so that $\mu = 10$) we calculated the proportion of estimates which lay in the range $\delta = 1.11111$ and $\delta = 1.09091$
That is, the proportion of estimates, expressed as a percentage, from which it would be concluded that μ is in the range (9,11).

These values appear in rows 1-3, "*%; Run 1*" etc.

The overall 'grand mean' tail-index of the 100,000 sample estimates is also provided for each estimator.

These values appear in "*Values: Run 1*" etc.

Less relevant in the context of the fat-tailed class (1.3) but included for comparison is the performance of the sample mean of (2.1).
Thus the first row of Table 1.1 shows that in the first run of 100,000 samples of size 100, only 2.41% of tail-index estimates from the MLE and TVBC were consonant with a mean in the range (9,11), while for Hill's estimator the figure was 2.40%. The sample mean fell in this range in 3.09% of samples.
Rows 2 and 3 have the results for Runs 2 and 3.

The fourth row (Values: Run 1) gives the grand average of tail-index estimates over 100,000 samples of size 100. For the MLE this was 1.205, 1.205 for TVBC, 1.207 for Hill's estimator. The mean of means was 7.749 (implying a tail-index of 1.129).

Sample size 100

Estimator	MLE	TVBC	Hill	Mean
% ; Run 1	2.41	2.41	2.40	3.09
% ; Run 2	2.45	2.43	2.44	3.08
% ; Run 3	2.45	2.45	2.56	3.09
Values: Run1	1.205	1.205	1.207	7.749
Values: Run 2	1.133	1.132	1.128	7.585
Values: Run 3	1.303	1.302	1.292	7.955

Table 1.1 The percentage of Pareto samples of size 100 with $\delta = 1.1$ ($\mu = 10$) for which tail-index estimates imply a mean in the range (9,11) in 3 runs of size 100,000 (rows 1-3). The grand average tail-index estimate is also given (rows 4-6). Column (5) contains corresponding values for the sample mean.

The simulation exercise was repeated for sample size n increased to 200. Results are given in Table 1.2

Sample size 200

Estimator	MLE	TVBC	Hill	Mean
% Run 1	3.44	3.44	3.46	3.32
% Run 2	3.34	3.44	3.42	3.33
% Run 3	3.40	3.39	3.41	3.24
Values: Run1	1.117	1.119	1.128	8.169
Values: Run 2	1.177	1.178	1.180	8.197
Values: Run 3	1.077	1.078	1.077	8.263

Table 1.2 The percentage of Pareto samples of size 200 with $\delta = 1.1$ ($\mu = 10$) for which tail-index estimates imply a mean in the range (9,11) in 3 runs of size 100,000 (rows 1-3). The grand average tail-index estimate is also given. Column (5) contains corresponding values for the sample mean.

The results are not particularly encouraging. *Perhaps the most unnerving aspect of the exercise is the variability in grand totals (means of 100,000 tail-index estimates) in the "Values" rows. While concordance between all three estimators is satisfactory, the results are less than reassuring for an insurer desperate to know the mean and tail-index of the claims generating process.*

Doubling the sample size improves matters, but not markedly.

(Hall (1982) investigated optimal achievable convergence rates of tail-index estimators).

Conditional maximum likelihood (ML) estimation of δ based on the largest extremes is attractive, because for all insurance losses governed by distributions from **F** 'exceedances over a threshold' have a generalized Pareto distribution equivalent to (1.10).

2.3 An extended Pareto model

Bierlant, Joossens and Segers (2004) have proposed an extension of the GPD devised to approximate the conditional distribution of (X-t) given $X > t$ for much lower thresholds t, than the GPD is capable of. Establishing their result involves a refinement of Pickands' original (1975) workings. The new model is:

(2.2) $$1-F_t(x:\delta,\lambda,\gamma) = \{1+x/\lambda+\nu[1-(x/t+1)^{1+\gamma}]\}^{-\delta}$$

where $\nu = t/\lambda-1$ of which (1.10) is a special case when $\gamma = -1$.

The extended model has the capacity to fit a much larger proportion of claims in large data sets. For the Society of Actuaries, (1991) Group Medical Insurance, Large Claims Database, the authors report good fits for 75,789 claims above $25,000 compared with Generalized Pareto which only provided a good fit for 7,860 claims above $100,000. The principal deficiency of the new 3-parameter distribution being its inability to handle *only the very largest claims.* Such claims are the focus of this paper (for which the (1.10) *can* be assumed to provide an adequate model, as does the Frechet distribution, which we now describe).

2.4 Frechet extreme value distribution

Another law which serves to characterise class **F** is the Frechet distribution,

(2.3) $$f_k(x) = \delta x^{-k\delta-1}\exp(-x^{-\delta})/\Gamma(k) \quad (x > 0, \delta > 0)$$

for which **F** is the maximum domain of attraction (Embrechts *et al*, (1997) p.131) The standardised order statistic $X_{(n-k+1)}/v_n$ from a large sample of n independent identically distributed claims with a common distribution F ε **F** tends under non-restrictive regularity conditions in law to $X^*_{(k)}$ - *the kth extreme value* with distribution (2.3). Precise results are to be found in Gnedenko (1943).

The sequence of normalizing constants $\{v_n\}$ – 'the tail-quantile function – derives from $n\{1-F(v_n)\}=1$ for large n and for all k not too large compared with n (Kendall and Stewart, (1969), p.331, David and Nagaraja, (2003), p.306). For example, for Pareto (1.10), $v_n = \lambda n^{\rho}$. Tables of tail-quantile functions are given in Embrechts *et al.* (1997), Section 3.4

A number of authors have cautioned about difficulties which arise in applying ML theory to tail-index estimation, (e.g. Smith (1985, 1987), Smith and Naylor (1987), Nagaraja, (2004)) which in any case only provides *asymptotic standard errors* ((Embrechts *et al*, (1997), Chapter 6), Drees, Ferreira and de Haan, (2004)) for situations (i.e. for particular classes of general insurance) where in fact sample sizes are likely to be small even with pooled data across entire national industries.

3. The nature of risk-neutral fat-tailed premiums

The insurer is risk-neutral or only mildly risk-averse.

From (1.13) for claims above a sufficiently high level, i.e. which are Pareto (1.10) premium P for a claim in the form $(1+X/\lambda)$ is given by $P = (1-\rho/\beta)^{-\beta}$.
Risk-neutral insurers have risk-aversion parameter $\alpha = 0$, i.e. $\beta = 1$, the premium being $E[1+X/\lambda] = (1-\rho)^{-1}$.

Less obvious is that for Pareto (1.10),

(3.1) $$(1-\rho)^{-1} = \frac{E[1+X_{(n)}/\lambda]}{E[1+X_{(n-1)}/\lambda]}$$

where $X_{(k)}$ is the kth order statistic from a sample of size n.

Thus the risk-neutral premium is the ratio of the two largest expected claims.

More generally, we have the following.

Theorem 1 For integer k ≥ 1, and Pareto (1.10) order statistics $X_{(\cdot)}$

(3.2) $$\frac{E[1+X_{(n)}/\lambda]}{E[1+X_{(n-k)}/\lambda]} = kB(k,1-\rho)$$

independently of n. Here $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, $a \geq 0$, $b \geq 0$, and $\rho = 1/\delta$.

Proof: This appears in Appendix, Note 1.

Heuristically, as n is increased, the *value* of the limit remains unchanged (since it is independent of n) but the random variable $1+X_{(n-k)}/\lambda$ *takes on the distributional character of* $v_n X^*_{(1+k)}$ so that we obtain

Corollary

For large n (3.2) leads to the relation

(3.3) $$\frac{E[X^*_{(1)}]}{E[X^*_{(1+k)}]} = kB(k, 1-\rho)$$

In fact Equation (3.3):

(i) can be verified directly (since $E[X^*_{(1+k)}] = \Gamma(k+1-\rho)/\Gamma(k+1)$ the expected value can be found by integration from (2.3) in the usual way), or

(ii) follows for k ≥ 0, from existence of $E[(X_{(n-k)})^{1+\varepsilon}]$ for $1+\varepsilon < 1/\rho$ and hence the uniform integrability of $1+X_{(n-k)}/\lambda$ allowing application of Theorem 25.12 Corollary of Billingsley, (1995) on convergence of expected values when sequences of random variables converge weakly.

Remark
Equation (3.3) is of somewhat broader ambit than (3.2) in that it is applicable to any distribution from $\mathcal{F}$. As n is increased, the Pareto ratio $\frac{E[1+X_{(n)}/\lambda]}{E[1+X_{(n-k)}/\lambda]}$ predicated on exceedances, has limit $\frac{E[X^*_{(1)}]}{E[X^*_{(1+k)}]}$ - the ratio of expected extremes for any fat-tailed distribution in the Frechet domain of attraction.

The universality of this latter ratio depends however, on a different limiting process, requiring the extremes to be the largest order statistics deriving from a sufficiently large sample. (see for instance, Kendall and Stuart, (1969), p.330). That is, the limit does not require (directly at least) the notion of 'an exceedance'.

For any fat-tailed distribution approximately Pareto above level t, the risk neutral premium for claim $1+X/\lambda(t)$ is $\frac{E[X^*_{(1)}]}{E[X^*_{(2)}]}$, i.e.

$$\text{(3.4)} \qquad P = \lambda(t)\frac{E[X^*_{(1)} - X^*_{(2)}]}{E[X^*_{(2)}]}$$

Fitting exceedances above a threshold is treated for example in Embrechts *et al*. (1997), Section 6.5

Estimating equations for ρ

Equation (3.3) provides that $\frac{X^*_{(1)}}{X^*_{(1+k)}}$ is a *consistent estimator* of $kB(k,1-\rho)$.

Explicitly: $$\frac{E[X^*_{(1)}]}{E[X^*_{(2)}]} = (1-\rho)^{-1}$$

$$\frac{E[X^*_{(1)}]}{E[X^*_{(3)}]} = (1-\rho)^{-1}(1-\rho/2)^{-1}$$

$$\frac{E[X^*_{(1)}]}{E[X^*_{(4)}]} = (1-\rho)^{-1}(1-\rho/2)^{-1}(1-\rho/3)^{-1}$$

and so on.

Consistency of estimators is *sine qua non*; a real bonus however is the following:

Theorem 2

For Pareto (1.10) order statistics $X_{(\cdot)}$

$$\text{(3.5)} \qquad E\left[\frac{1+X_{(n)}/\lambda}{1+X_{(n-k)}/\lambda}\right] = kB(k,1-\rho)$$

for integer $k \geq 1$, independently of n.

Proof: This appears in Appendix, Note 2

Since the result is true for all n, we obtain:

Corollary:

(3.6) $$E\left[\frac{X^*_{(1)}}{X^*_{(1+k)}}\right] = \text{kB(k,1-ρ)}$$

Thus the ratios $\frac{X^*_{(1)}}{X^*_{(1+k)}}$ *are also unbiased estimators* of kB(k,1-ρ)

Proof : This is in Appendix Note 2.

3.1 The distribution of the heavy-tailed premium

It is shown in Appendix Note 2, that $U(k,\delta) = \left[\frac{X^*_{(1)}}{X^*_{(1+k)}}\right]^{-\delta}$ is distributed as B(k,1).

In particular $U(1,\delta) = \left[\frac{X^*_{(1)}}{X^*_{(2)}}\right]^{-\delta}$ is B(1,1), i.e. uniformly distributed, and

$V = \frac{X^*_{(1)}}{X^*_{(2)}}$, the unbiased estimator of premium (3.4), has Pareto distribution

(3.7) $$F_\delta(v) = 1\text{-}v^{-\delta} \quad (1 < v < \infty)$$

i.e. *is Pareto with the same tail-index as the claims distribution*.

Note that $E[V] = \delta/(\delta\text{-}1) = (1\text{-}\rho)^{-1} = E[1+X/\lambda]$ with $Var[V] = \infty$ follow from (3.7).

It is indeed remarkable that the premium principle (1.9) determines heavy-tailed premiums so transparently and so convincingly!

Remarks:

(i) In the case claims are purposively modelled with no variance, all premiums are greater than $(1\text{-} ½)^{-1} = 2$ by assumption.

(ii) Distribution (3.7) offers new insight into the nature of heavy tailed premiums, and the requirements for long-term insurer solvency; *premiums must be set via a mechanism dependent on a larger tail-index than the claims generation process.*

(iii) Broadly speaking: "*premiums for thin-tailed claims are determined largely by the first two moments of the claims distribution; premiums for fat-tailed claims are determined by the first two extremes*"

4. Modelling tail-index imprecision: risk-averse premiums

4.1 A natural model for tail-index imprecision

For Pareto claim $(1+X/\lambda)$ the risk-averse premium is $P_\beta = (1-\rho/\beta)^{-\beta}$ for an insurer with constant relative risk-aversion α; $\beta = 1/(\alpha+1)$. Thus β represents *the largest tail-index with which an insurer with this risk profile will deal;* i.e. $\beta = \rho_{max}$.

A natural model for uncertainty in ρ thus suggests itself in the form of the transformed beta distribution

(4.1) $$f_\rho(x) = \nu\beta^{-\nu}x^{\nu-1}, \quad (0 \le \beta < 1, \nu \ge 1)$$

A premium $P(\nu,\beta)$ which takes account of tail-index uncertainty is then determined by the integral

(4.2) $$P(\nu,\beta) = \nu \int_0^\beta (x/\beta)^{\nu-1}(1-x/\beta)^{-\beta}dx/\beta$$

either as a mean value, or as a Bayesian posterior mean, if (4.1) is taken as the prior for ρ. Under either interpretation,

(4.3) $$P(\nu,\beta) = \nu B(\nu,1-\beta)$$

is a premium compensating for uncertainty in ρ (c.f. Theorem 2 with $\nu = k$).

4.2 Pragmatism for the insurer:

If $\nu = 1$ in (4.3), prior uncertainty in ρ is modelled by the uniform distribution over $[0,\beta)$. This is the 'least informative prior', the 'law of equal ignorance' for ρ on $[0,\beta)$.

Then

(4.4) $$P(1,\beta) = (1-\beta)^{-1}$$

where the unknown tail-index ρ is now replaced by $\beta = \rho_{max}$.

- For claim $(1+X/\lambda)$ with ρ known, a risk-neutral insurer would set premium $(1-\rho)^{-1}$. *If ρ is unknown, the insurer hedges against uncertainty by setting premiums on the basis of the largest tail-index ($\rho = \beta$) with which it will deal*!

- *For finite premiums, the insurer must have $\beta < 1$, i.e. must be risk-averse.*

• *If the insurer has chosen β it has in effect stipulated the distribution $F_\delta(v) = 1\text{-}v^{-1/\beta}$ for the premium (c.f. Equation (3.7)). So long as $\beta > \mu$, reserves will in probability accumulate, and the insurer may remain solvent.*

Embrechts and Veraverbeke (1982) found that insurers required very large reserves to deal in heavy-tailed claims.

We now illustrate how the foregoing theory provides a coherent methodology not only for pricing general claims, but also for pricing of next year's k largest claims on the basis of this year's k largest claims.

5.0 Setting premiums for the k largest extreme claims

That claim severity is increasing in expensive classes of general insurance seems overwhelmingly probable.
Swiss Re identify the increasing litigiousness in US as a major factor (less than half the costs awarded by courts are returned to victims). But other factors are easily identifiable. Hurricanes and winterstorms account for a large proportion of US annual insured losses (see for instance Swiss Re's annual summary of world catastrophe's and insured losses in *Sigma,* No. 2, (1999), (2000), etc.).
Hsieh (2004) points to a 1994 report by Insurance Services Office, Inc. suggesting that population density on the storm-prone southwestern Atlantic coast of the US increased nearly 75% from 1970 to 1990, a much greater increase than the 20% countrywide figure.
Accepting the likelihood of a gradually increasing tail-index in claims generation processes (whatever the mechanisms) how can insurers monitor developments? This can only come from examination of the largest claims in relevant insurance classes.

5.1 Application: Premiums for year 2000 ten largest extreme claims for man-made disasters 'major fires and explosions' based on 1999 largest claims (Table 2)

Ten largest insured losses (man-made disasters 'major fires and explosions, 1999)

Date	Place	Event	Insured Loss (USD millions, 1999)
1 March	US, Dearborn, MI	Explosion and fire at power station	650
5 July	US, Gramercy, LA	Explosion, Aluminium plant	275
25 March	US, Richmond, CA	Oil refinery explosion	247
17 February	US, Kansas City	Power plant explosion	196
12-13 April	UK, Edinburgh	Explosion at transformer factory	137
8 June	Germany, Wuppertal-Eberfield	Chemicals plant explosion and fire	102.5
18 August	Germany, Gendorf	Polymer plant explosion	92.2

27 October	Germany, Vahdorf	Turkey slaughterhouse fire	82
20 October	Germany, Darmstadt-Arheilgen	Fire at liquid crystal production plant	71.7
23 February	US, Martinez, CA	Oil refinery explosion	71

Table 2 Extreme claims data: Insured losses, (1999) Man-made disasters
Source: *Sigma* No. 2 (2000), published by Swiss Re.

The ten extremes display a consistent pattern (Figure 1) suggestive in relative terms, of the sort of 'decay' among the largest extremes to be expected from theory; that is from equations (3.3) and (3.5) assuming a fixed tail-index ρ.
Hall and Tavjidi (2000) have investigated non-parametric trend-fitting to extremes over time.

The assumptions which underpin our subsequent analysis are as follows:

(i) The data represent extreme claims from the class of distributions **F**

(ii) The tail-index is fixed for the extreme claims.

(iii) There are sufficient 'ordinary' claims (not extremes) realized to justify assuming that the Frechet extreme distribution applies with tail-index ρ to the k largest claims.

(iv) Ordinary claims may arise from a mixture of distributions, some with the same tail-index as for extremes. Other ordinary claims arise from other thin- or fat-tailed distributions.

The total number of claims is not necessarily known; only the k largest claims.

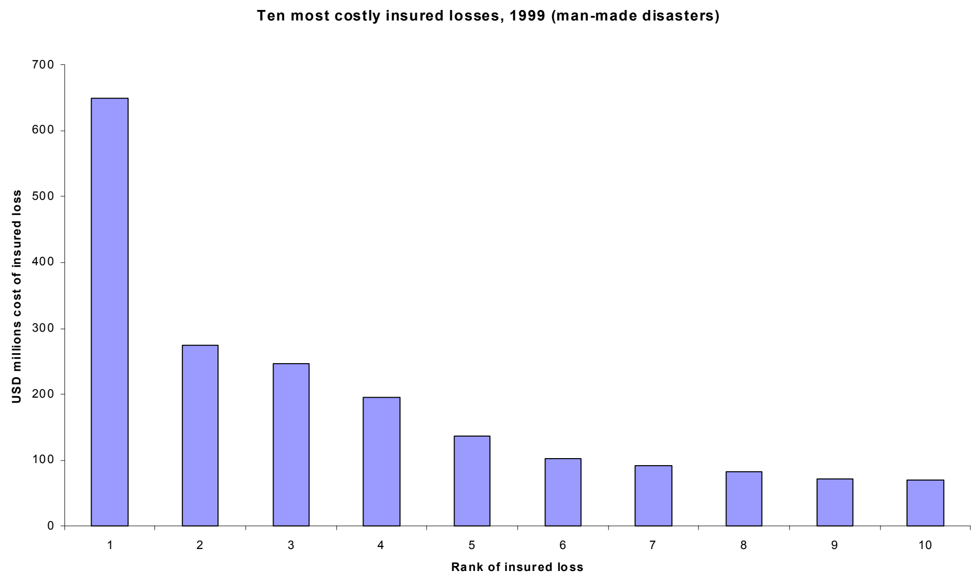

Figure 1 Ten largest insured losses for man-made fires and explosions, (1999), USD millions.

5.2 Results from Maximum Likelihood Estimation

Drees *et al* (2004) outline use and discuss properties of conditional maximum likelihood estimation in the sense of Cox and Hinkley (1974, p.17) predicated on exceedances above a sufficiently high level having a Generalized Pareto distribution.

The conditional likelihood function is of the form:

(5.1) $$L_1[\mathbf{X}|\ X_{(n-k)} = x_{(n-k)}] = \prod_{j=0}^{k-1} f(y_{(n-j)})$$

where $\mathbf{X} = [X_{(n)}, X_{(n-1)} \cdots X_{(n-k+1)}]$, $y_{(n-j)} = x_{(n-j)} - x_{(n-k)}$ and $f(\cdot)$ is the Pareto density (1.10)

An alternative form of the conditional maximum likelihood which follows from the Markov character of the order statistics is

(5.2)
$$L_2[X_{(n)}, X_{(n-1)} \cdots X_{(n-k+1)}|\ X_{(n-k)} = x_{(n-k)}]$$

$$= \prod_{j=0}^{k-1} f(x_{(n-j)})/[1- F(x_{(n-k)})]^k$$

where $F(\cdot)$ is Pareto (1.10) and $f(\cdot)$ is its density.

For application to the data set in Table 2, $L_1(\cdot)$ is general in that it is applicable to *any* distribution from (1.10). The second likelihood $L_2(\cdot)$ presupposes Pareto for the loss distribution. But both assumptions lead to the same Frechet extremes.

MLE procedures applied to the data of Table 2 lead to the following estimates:

(i) $\hat{\rho} = 0.4982$, $\hat{\lambda}_1 = 155.1$

(ii) $\hat{\rho} = 0.4982$, $\hat{\lambda}_2 = 84.1$

Note that $\hat{\lambda}_2 = \hat{\lambda}_1 - 71 = \hat{\lambda}_1 - x^*_{(10)}$ consistent with the assumptions.

5.3 Estimates from the pattern of the set of extremes

We have from Theorem 2 and Equation (3.5) a set of estimating equations arising from

$$E\left[\frac{X^*_{(1)}}{X^*_{(1+j)}}\right] = \text{jB(j,1-ρ)} \quad \text{j = 1,2, · · · (k-1)}$$

providing (k-1) estimates of ρ *all involving the largest extreme*.

Equation (3.4) affirms that the two largest observed claims are the most important for setting premiums.

The estimating equations are:

$$\frac{X^*_{(1)}}{X^*_{(2)}} = (1-\rho)^{-1} \Rightarrow \hat{\rho} = 0.5769$$

$$\frac{X^*_{(1)}}{X^*_{(3)}} = (1-\rho/2)^{-1}(1-\rho)^{-1} \Rightarrow \hat{\rho} = 0.4950$$

$$\frac{X^*_{(1)}}{X^*_{(4)}} = (1-\rho/3)^{-1}(1-\rho/2)^{-1}(1-\rho)^{-1} \Rightarrow \hat{\rho} = 0.5115$$

etc.

The complete set of tail-index estimates is:

0.5769,0.4950,0.5115,0.5735, 0.6144,0.6139,0.6195,0.6306,0.6178.

The value of $\hat{\rho}$ = 0.5769 deriving from the first two extremes would seem to deserve special significance.
The mean of the nine estimates is 0.5836, their mean absolute deviation 0.0395.
The value $\hat{\rho}$ = 0.6178 provides the best fit among the estimates in terms of minimising

$$\Sigma_k|\text{observed}\frac{X^*_{(1)}}{X^*_{(1+j)}} - \text{jB(j,1-}\hat{\rho}\text{)}|.$$

Figure 2 below shows that the maximum likelihood estimate ($\hat{\rho}$ = 0.4982) does not fit the observed structure of extremes at all well, compared with say $\hat{\rho}$ = 0.62

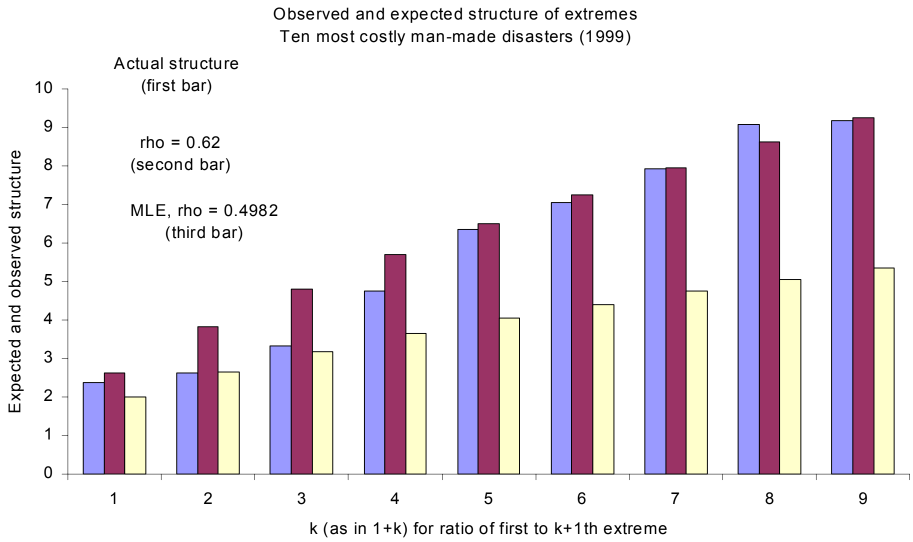

Figure 2 The observed and expected structure of Swiss Re's man-made disaster extremes, using rho = 0.4982 (the maximum likelihood estimator) and rho = 0.62. In Figure 2, the first column in each group represents observed $X^*_{(1)}/X^*_{(1+j)}$. The second bar is its expected value $jB(j,1-\rho)$ when $\rho = 0.62$.

The third bar is $jB(j,1-\rho)$ when $\rho = 0.4982$ (the MLE).

5.4 Some diagnostics for the tail-index estimates based on ratios of the extremes

In Appendix Note 2, the distribution of $U(j,\delta)) = \left[\frac{X^*_{(1)}}{X^*_{(1+j)}}\right]^{-\delta}$ is shown to be B(j,1).

This allows us to do the following. The value $\rho = 0.62$ provides a good fit to the extremes.

By assumption, the upper limit of ρ is 1. To test the adequacy of rho = 0.62 we find first value of ρ which would be rejected in a 10% one-sided test.

The rejection region for a size α test is defined by $1- [1- u(k,\delta)]^j < \alpha$ or

(5.3) $$\rho < \ln[X^*_{(1)}/X^*_{(1+j)}]/\ln[\{1-(1-\alpha)^{1/j}\}^{-1}]$$

Table 3 below contains for each value of $j = 1,2, \cdots (k-1)$:

(i) The lower 90% C.I. on $U(j,\delta)) = \left[\frac{X^*_{(1)}}{X^*_{(1+j)}}\right]^{-\delta}$ (whatever value of δ is chosen).

(ii) The actual value of $U(j,\delta)$ for a given value of $\rho = 1/\delta$ ($\rho = 0.62$ in Table 3)

(iii) The upper 90% C.I. on U(j,δ)) = $\left[\dfrac{X^*_{(1)}}{X^*_{(1+j)}}\right]^{-\delta}$ (whatever value of δ is chosen).

(iv) The largest value of ρ we would reject at the lower 10% level given U. i.e. the first value of ρ (< 0.62) which is *not consistent* with the observed ratio of extremes, at the 10% level. For example, the observed ratio $X^*_{(1)}/X^*_{(2)}$ is inconsistent with values of ρ below ρ = 0.3736 using an α = 0.10 test; the ratio $X^*_{(1)}/X^*_{(3)}$ is inconsistent with values of ρ < 0.3258 and so on.

j	lower 90%	U(j,δ)	upper 90%	Rejection ρ
1	0.1000	0.2497	0.9000	0.3736
2	0.0513	0.2100	0.6838	0.3258
3	0.0345	0.1446	0.5358	0.3561
4	0.0260	0.0812	0.4377	0.4266
5	0.0209	0.0508	0.3690	0.4772
6	0.0174	0.0429	0.3187	0.4821
7	0.0149	0.0355	0.2803	0.4925
8	0.0131	0.0286	0.2501	***0.5084***
9	0.0116	0.0281	0.2257	0.4972

Table 3 Diagnostics for ρ = 0.62 based on the distribution of U(j,δ)

If the proposed value of ρ were changed from ρ = 0.62, only the third column of Table 3 would be altered (i.e. U(j,δ)).

In particular, the MLE estimate is not consistent (at the 10% level) with the actual ratio $X^*_{(1)}/X^*_{(9)}$ since the MLE value ρ = 0.4982 is less than 0.5084

5.4 Setting premiums

Once a value of ρ is chosen, the structure of the expected values of

$\dfrac{X^*_{(1)}}{X^*_{(1+j)}}$ is fixed at jB(j,1-ρ). The ratios decrease in relative terms as ρ is increased; that is, lower values of ρ produce relatively larger values of lower order extremes. This is depicted in Figure 3. Choice of a 'prudent' value of ρ (i.e. ρ = β = ρ_{max} as in (4.4) means that the *premium income for the k claims is derived mainly from the premium for the largest claim*. This after all, is what is needed to match expected claim losses.

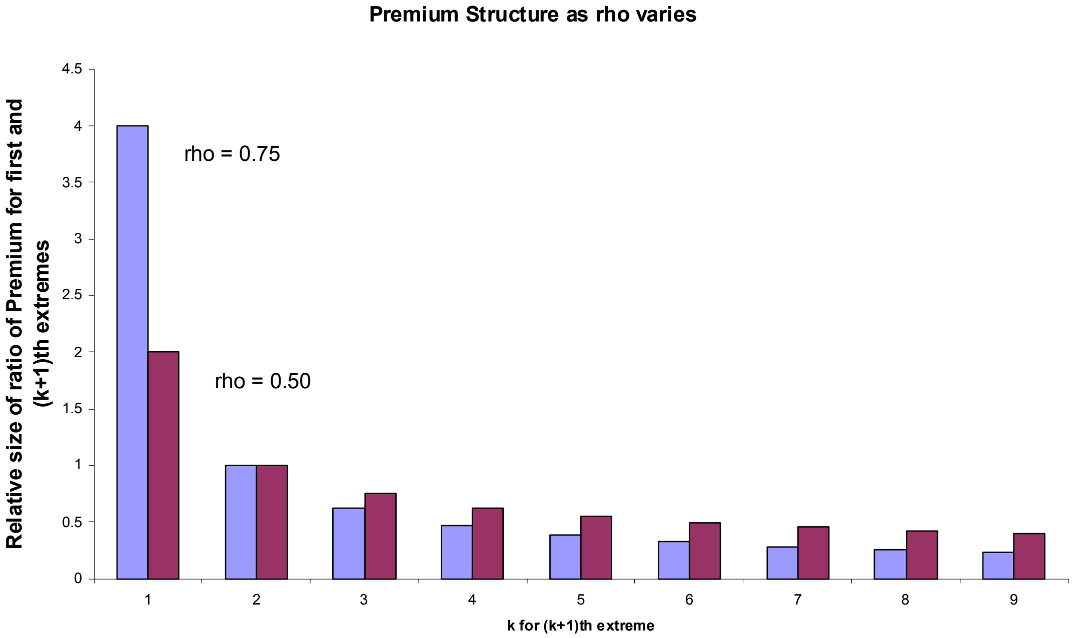

Figure 3 Large values of rho imply large premiums for the largest extreme; in relative terms premiums for the other extremes decrease more rapidly for fat tails than for thinner tails. (to provide comparison, premiums are standardised on the second largest claim)

5.5 Possible premiums (for year 2000 ten largest claims: man-made disasters)

A value of $\rho_{max} = \beta$ is decided, and *a premium for the largest extreme at least as large as the observed extreme is chosen* (by doing the latter, we insure the premium for the largest extreme has infinite variance). Other premiums ensue from Equation (3.3) with $\rho = \beta$; i.e. $\frac{E[X^*_{(1)}]}{E[X^*_{(1+j)}]} = jB(j,1-\beta)$

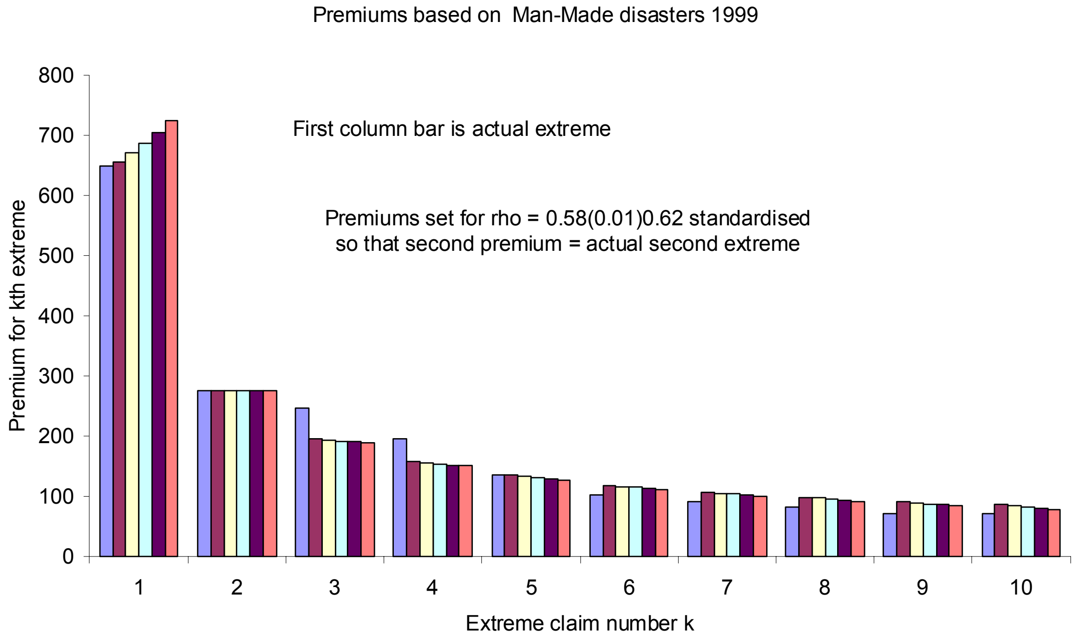

Figure 4 The actual premiums are set as $X^*_{(2)}/(1-\beta)$, $\beta = 0.58(0.01)0.62$.

More information about the premiums is provided in Table 4 which gives:
(i) the premium for the largest claim
(ii) the sum of the k=10 premiums
(iii) the upper 90% approximate confidence limit for the expected claim predicated on given β; i.e. $0.05^{-\beta}\times X^*_{(2)}$ based on the uniform distribution of $U(1,\delta)) = \left[\frac{X^*_{(1)}}{X^*_{(2)}}\right]^{-\delta}$ for given δ ($= \delta_{min} = 1/\rho_{max} = 1/\beta$).

	Actual	β = 0.58	β = 0.59	β = 0.60	β = 0.61	β = 0.62
$X^*_{(1)}$ and premiums	650	655	671	687	705	724
Total premiums	1924	1921	1923	1923	1926	1931
90% CI		1563	1610	1659	1710	1762

Table 4 Table showing premium for largest extreme, total premiums for k = 10 claims, and 90% upper CI for largest expected claim.

The third row emphasises the *dangerousness* of tails as heavy as these!

6.0 Expected value of the new record man-made disaster loss

Additional information about the menace in the tail is provided by the expected value of the new record disaster loss.

A consistent estimator (NR) of the insured cost of the expected new record man-made disaster is:

$$\text{(6.1)} \qquad \text{NR} = \frac{X^*_{(1)}}{X^*_{(2)}} \times \sum_{j=1}^{k-1} \frac{X^*_{(1)}}{X^*_{(1+j)}} \times \frac{X^{*(1-\delta)}_{(1+j)} - X^{*(1-\delta)}_{(j)}}{X^{*(-\delta)}_{(k)} - X^{*(-\delta)}_{(1)}}$$

Equation (6.1) - a regression estimator based on observed values of the first k extremes - established in the Appendix, Note 3, the value of δ used deriving from $\frac{X^*_{(1)}}{X^*_{(2)}}$. The estimator (like the record itself) has infinite variance. It is nevertheless possible to find an approximate confidence interval.

Since $\left[\frac{X^*_{(1)}}{X^*_{(1+j)}}\right]^{-\delta}$ is B(j,1), upper 100(1-α)% confidence limits L(j,δ,α) derive from the conditional distribution of $X^*_{(1)}$ given $X^*_{(1+j)}$ based on known ρ for each j.
The approximate upper confidence interval uses the rho estimate obtained from $\frac{X^*_{(1)}}{X^*_{(2)}}$, and is the maximum of the L(j,δ) ; (Appendix, Note 3)

<u>Application</u>
Substitution of actual extremes for Swiss Re's man-made disaster data, gives the expected new record value as 733 USD (millions).
However, the approximate 95% upper C.I. for the new record is 5334 USD (millions).

7.0 Summary and Conclusions

The rapid growth in commercial liability insurance claims requires a similar rapid adjustment of premiums. A methodology is provided which assists insurers to make judgements about premiums for the most costly claims they can expect on the basis of the latest available large claims and the tail-index of their generating process.
Information provided is ancillary to that available from on-going conventional statistical data analyses in respect of large claims.
A premium principle for fat-tailed claims with desirable properties is established.
Whereas premiums for thin-tailed claims depend on the first two *moments* of the claims distribution, fat-tailed premiums depend on the first two *extremes*.
A consistent and unbiased estimator can be found for the premium, as well as its distribution.
Modelling uncertainty in the tail-index effectively requires the insurer to use a prudently large value of the tail-index. Long-term solvency can only be attained if the premiums are predicated on heavier tails than the claims generation process.

A large proportion of premium income for the largest k claims derives from the premium for the largest claim.
This mimics the structure of expected claims themselves.
Implementation of the methodology is illustrated using Swiss Re's 10 largest man-made disasters of 1999.
Confidence intervals for the expected largest claim and a formula for the expected new record disaster and its approximate upper confidence limit, emphasise the dangerousness of tails with ρ in the range [½,1).

8.0 Appendix:

Note 1 For Pareto (1.10) order statistics $X_{(\cdot)}$

$$\frac{E[1+X_{(n)}/\lambda]}{E[1+X_{(n-k)}/\lambda]} = kB(k,1-\rho)$$

By direct integration $E[1+X_{(j)}/\lambda] = B(j,n-j-\rho+1)/B(j,n-j+1)$

Using j = n in the numerator and j = n-k in the denominator, the ratio

$$\frac{E[1+X_{(n)}/\lambda]}{E[1+X_{(n-k)}/\lambda]} = B(n,1-\rho)B(n-k,k+1)/B(n,1)B(n-k,k-\rho+1)$$

$$= \Gamma(k+1)\Gamma(1-\rho)/\Gamma(k+1-\rho)$$

$$= k\Gamma(k)\Gamma(1-\rho)/\Gamma(k+1-\rho)$$

$$= kB(k,1-\rho)$$

When k = 1, $kB(k,1-\rho) = (1-\rho)^{-1}$.

Note 2 $$E\left[\frac{X^*_{(1)}}{X^*_{(1+k)}}\right] = kB(k,1-\rho)$$

For a sample of n i.i.d.r.vs with common distribution $F(\cdot)$ and density $f(\cdot)$ denote by $X_{(1)}, X_{(2)}, \cdots X_{(n)}$ the ascending order statistics.

The conditional density of $X_{(k)}$ given $X_{(j)} = x_{(j)}$, (j > k) is $f^{\#}(x_{(k)}|x_{(j)})$ where

(A1) $$f^{\#}(x_{(k)}|x_{(j)}) = B^{-1}(k-j, n-k+1)\{F(x_{(k)}-F(x_{(j)}\}^{k-j-1}\{1-F(x_{(k)})\}^{n-k} f(x_{(k)}) / \{1-F(x_{(j)})\}^{n-j}$$

(see for instance Arnold, Balakrishnan and Nagaraja (1993), Theorem 2.4.1)

Putting k = n-j+1, j = n-j-k+1 we obtain:

$f^{\#}(x_{(n-j+1)}|x_{(n-j-k+1)}) = B^{-1}(k,j)\ \{F(x_{(n-j+1)})-F(x_{(n-j-k+1)})\}^{k-1}\{1-F(x_{(n-j+1)})\}^{j-1}\ f(x_{(n-j+1)})$
$/\{1-F(x_{(j)})\}^{j+k-1}$

Now consider the special case when $F(\cdot)$ is Pareto (1.10), so that

$$F(x) = 1-(1+x/\lambda)^{-\delta}\ \ (x > 0, \delta > 0)$$

And we make some substitutions:

(i) $u = \{1-F(x_{(n-j+1)})/\{1-F(x_{(n-j-k+1)})\}= (1+x_{(n-j+1)}/\lambda)^{-\delta}/(1+x_{(n-j-k+1)}/\lambda)^{-\delta}$

Then

(A2) $$f^{\#}(x_{(n-j+1)}|x_{(n-j-k+1)})dx = B^{-1}(k,j)(1-u)^{k-1}u^{j-1}du$$

And $E\left[\frac{1+X_{(n-j+1)}/\lambda}{1+X_{(n-j-k+1)}/\lambda}\middle| X_{(n-j-k+1)}\right] = B(k,j-\rho)/B(k,j)$

$$= \Gamma(j-\rho)\Gamma(k+j)/\Gamma(k+j-\rho)/\Gamma(j)$$

Putting j = 1 leads to $E\left[\frac{1+X_{(n)}/\lambda}{1+X_{(n-k)}/\lambda}\middle| X_{(n-k)}\right] = kB(k,1-\rho)$

Since the conditional expectation is a constant independent of $X_{(n-k)}$ it is also the unconditional expectation, and (3.4) follows. Alternatively, the result can be derived by direct integration of the joint density of $X_{(n)}$ and $X_{(n-k)}$.

(ii) We could use the Billingsley uniform integrability approach of (3.3) again but the following heuristic argument, (which can easily be made rigorous), is more interesting.

In anticipation of convergence in law to extreme value distributions, in (A2) we

put $x_{(n-j+1)} = \lambda n^{\rho}x^{*}_{(j)}$ and $x_{(n-j-k+1)} = \lambda n^{\rho}x^{*}_{(j+k)}$

where $\rho = 1/\delta$ and $v_n = \lambda n^{\rho}$ is the tail-quantile function for Pareto.

Then the substitutions together lead to

$$u = (1+n^{\rho}x^{*}_{(j)})^{-\delta}/(1+n^{\rho}x^{*}_{(j+k)})^{-\delta} \approx (x^{*}_{(j)})^{-\delta}/(x^{*}_{(j+k)})^{-\delta}$$

for large n.

Thus obtain the conditional distribution of the jth extreme given the (j+k)th extreme

(A3) $f^{*}(x^{*}_{(j)}|x^{*}_{(j+k)})dx^{*}_{(k)} =$

$$B^{-1}(k,j) \times \left\{1-\left(\frac{x^*_{(j)}}{x^*_{(j+k)}}\right)^{-\delta}\right\}^{k-1} \times \left\{\left(\frac{x^*_{(j)}}{x^*_{(j+k)}}\right)^{-\delta}\right\}^{j-1} d\left(\frac{x^*_{(j)}}{x^*_{(j+k)}}\right)^{-\delta}$$

(Incidentally this also shows that the unconditional distribution of $\left[\frac{X^*_{(j)}}{X^*_{(j+k)}}\right]^{-\delta}$ is B(j,k). The result can also be established directly from the joint distribution of Pareto $X_{(n-j+1)}$ and $X_{(n-j-k+1)}$. The ratio of Pareto order statistics is again independent of n, passage to the limit producing the ratio of general extremes, and the normalizing constants 'cancelling out' (c.f. Equation (A3) none of the ratios are altered if normalizing constants are inserted)

Hence $E\left[\frac{X^*_{(j)}}{X^*_{(j+k)}}\right]$ = Γ(j-ρ)Γ(k+j)/Γ(k+j-ρ)/Γ(j) from which (3.5) follows on putting j=1.

Note that (A3) provides for possible inference for the ratios of extremes since

$\{X^*_{(1)}/X^*_{(1+j)}\}^{-\delta}$ has a B(j,1) distribution.

<u>Note 3; The expected new record man-made disaster insured loss</u>

(A4) $$NR = \frac{X^*_{(1)}}{X^*_{(2)}} \times \sum_{j=1}^{k-1} \frac{X^*_{(1)}}{X^*_{(1+j)}} \times \frac{X^{*(1-\delta)}_{(1+j)} - X^{*(1-\delta)}_{(j)}}{X^{*(-\delta)}_{(k)} - X^{*(-\delta)}_{(1)}}$$

A new record loss is presaged whenever one of the old extremes is superceded. Its expected value can be estimated by a regression estimator as follows;

The conditional expected value of a new record (1+j)th extreme, given a new observation between $X^*_{(1+j)}$ and $X^*_{(j)}$ (assuming δ or ρ to be known, and the extremes to be Pareto distributed) is:

$$\frac{X^{*(1-\delta)}_{(1+j)} - X^{*(1-\delta)}_{(j)}}{X^{*(-\delta)}_{(1+j)} - X^{*(-\delta)}_{(j)}} (1-\rho)^{-1}$$

The term $\frac{X^*_{(1)}}{X^*_{(1+j)}}$ is an unbiased estimator of jB(j,1-ρ)}

i.e. of $E[X^*_{(1)}|X^*_{(1+j)}]$, which is then applied to the new conditional expected value of $X^*_{(1+j)}$ given the new observation

$$\frac{X_{(1+j)}^{*(1-\delta)} - X_{(j)}^{*(1-\delta)}}{X_{(1+j)}^{*(-\delta)} - X_{(j)}^{*(-\delta)}} (1\text{-}\rho)^{-1}$$

in the old range ($X^*_{(1+j)}$, $X^*_{(j)}$).

The mutually exclusive contributions from new extremes replacing the old ones are added together with $(1\text{-}\rho)^{-1}$ replaced by its unbiased estimator $\frac{X^*_{(1)}}{X^*_{(2)}}$, from which value δ is also derived. Each of the (k-1) new estimates of the expected values of $X^*_{(1)}$ is probability weighted by

$$\frac{X_{(1+j)}^{*(\delta)} - X_{(j)}^{*(-\delta)}}{X_{(k)}^{*(-\delta)} - X_{(1)}^{*(-\delta)}}$$

Note that the estimator has infinite variance (as indeed does the largest extreme itself when ρ is in the range [½,1)).

A 100(1-α)% confidence limit for each expected new value of $X^*_{(1)}$ given the average value of any new $X^*_{(1+j)}$ in the old range ($X^*_{(1+j)}$, $X^*_{(j)}$) is

(A5) $$L(j,\delta) = \frac{X^*_{(1)}}{X^*_{(1+j)}} \times \frac{X_{(1+j)}^{*(1-\delta)} - X_{(j)}^{*(1-\delta)}}{X_{(1+j)}^{*(-\delta)} - X_{(j)}^{*(-\delta)}} (1\text{-}\rho)^{-1} \times (\alpha)^{-\rho/k}$$

deriving from the conditional distribution of $X^*_{(1)}$ given $X^*_{(1+j)}$ based on known ρ for each j. The approximate confidence interval uses the estimate of ρ obtained from $\frac{X^*_{(1)}}{X^*_{(2)}}$, and the maximum of L(k,δ).

Hill (1994) used a Bayesian approach to forecasting extremes; the present approach is Bayesian to the extent that $\rho_{max} = \beta$ (c.f. Equation (4.4) might be used to estimate the new record, rather that the value deriving from the unbiased estimator.

9.0 References

Albrecht, P. (1992), Premium calculation without arbitrage, ASTIN Bulletin, **22**, 247-256

Arnold, B. Balakrishnan, N. and Nagaraja H. (1993) *A First Course in Order Statistics*, J.Wiley

Arrow, K.J. (1971) Theory of risk aversion in *Essays on the Theory of Risk Bearing*, 90-120, Markham, Chicago

Balkema, A.A. and de Haan, L. (1974), Residual lifetime at great age, *Annals of Probability,* **2**, 792-804

Bierlant, J. Joossens, E. and Segers, J. (2004) A new model for large claims, *North American Actuarial Journal*, to appear

Bierlant, J. Teugels, J.L. Vynckier, P. (1996) *Practical Analysis of Extreme Values*, Leuven University Press

Bierlant, J. Teugels, J.L. Vynckier, P. (1994) Extremes in non-life insurance, Galambos,J. et al. eds, *Extreme Value Theory and Applications* 489-510, Dordrecht: Kluwer Academic

Billingsley, P, (1995) *Probability and Measure,* 3rd Edn. J.Wiley.

Black, F. and M. Scholes, (1973), The pricing of options and corporate liabilities, *Journal of Political Economy*, **81**, 637-659

Bladt, M. and T.H. Rydberg, (1998), An actuarial approach to option pricing under the phyical measure and without market assumptions, *Insurance: Mathematics and Economics*, **22**, 65-73

Bowers, N.L. Gerber, H.U. Hickman, J.C. Jones, D.A. and Nesbitt, C.J. (1986) *Risk Theory* Society of Actuaries, USA.

Cox, D.R. and Hinkley, D.V. (1974), *Theoretical Statistics*, Chapman and Hall

Danielsson, J. de Haan, L. Peng, L. and de Vriess, C.G. (2001), Using the bootstrap method to choose the sample fraction in tail index estimation, *J. Multivariate Anal*. **76**, 226-248

David, H.A. and Nagaraja, H.N. (2003) *Order Statistics*, 3rd. Edition, Wiley-Interscience

De Haan, L. (1994) Extreme value statistics, Galambos,J. et al. eds, *Extreme Value Theory and Applications*, 93-122, Dordrecht: Kluwer Academic

Dekkers, A.L.M. and de Haan, L. (1993) Optimal choice of sample fraction for extreme value estimation. *J. Multivariate Anal*. **47**, 173-195

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

Drees, H. Ferreira, A. and De Haan, L. (2004) On maximum likelihood estimation of the extreme value index, *Annals of Applied Probability*, **14**, 1179-1201

Drees, H. and Kaufmann, E. (1998) Selecting the optimal sample fraction in univariate extreme value estimation, Stochastic Proc. Appl. **75**, 149 - 172

Draisma, G. de Haan, L. Peng, L. and Pereira, T.T. (1999) A bootstrap-based method to achieve optimality in estimating the extreme value index, *Extremes*, **2**, 367-404

Embrechts, P. Kluppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance* Springer-Verlag.

Embrechts, P. and Veraverbeke, N. (1982) Estimates of the probability of ruin with special emphasis on the possibility of large claims, *Insurance: Math. Econ*. **1**, 55-72

Enz, R. and Holzheu, T. (2004) The economics of liability losses - insuring a moving target, *Sigma No. 6* Swiss Re, (available online at www.swissre.com/sigma)

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications,* Vol.II, J.Wiley

Feuerverger, A. and Hall, P. (1999), Estimating the tail exponent by moment departure from a Pareto distribution, *Ann. Statist*. **27**, 760-781

Gay, R. (2004a), Pricing risk when the distributions are fat-tailed. *J.Appl. Prob*. **41A**, Special Volume: *Festschrift for Chris Heyde; Stochastic Methods and Their Applications*,157-175

Gay, R. (2004b), The power principle and tail-fatness uncertainty, *Monash University, Dept of Econometrics, Working Paper Series,* (WP01/04).

Available online at: http//:www.buseco.monash.edu.au/depts./ebs/pubs/wpapers/

Gnedenko, B.V. (1943) Sur la distribution limite du terme maximum d,une serie aleatoire, *Annals of Mathematics*, **44**, 423-433

Grimshaw, D. (1990) A unification of tail-estimators, *Comm.Statist.Theory Methods,* **19,** 4841-4857

Hall, P. (1982) On some simple estimates of an exponent of regular variation, *J.Roy. Statist. Soc. Ser. B*. **44**, 37-42

Hall, P. and Tajvidi, N. (2000) Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data, *Statistical Science*, **15**, 153-167

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

Hill, B.M. (1994) Bayesian forecasting of extreme values in an exchangeable sequence, *Journal of Research of the National Institute of Standards and Technology*, **99**, 521-538

Hsieh, P-H. (2004), A data analytic method for forecasting the next record catastrophic loss, *Journal of Risk and Insurance,* **71**, 309-322

Huisman, R. Koedijk, K. Kool, C. and Palm, F. (2002) Tail index estimation in small samples, *Journal of Business and Economic Statistics*, **19**, 208-16

Kendall, M.G. and Stuart, A. (1969), *The Advanced Theory of Statistics*, Vol. 1. Griffin.

Leadbetter, M. (1991) On a basis for 'peaks over threshold' modeling, *Statistics and Probability Letters*, **12,** 357-362

Mikosch, T. (1997), Heavy-tailed modelling in insurance, *Commun. Statist. - Stochastic models*, **13**, 799-815

Nagaraja, H.N. (2004) An introduction to extreme order statistics and actuarial applications, Paper presented at 2004 ERM Symposium April 26, 2004

Oberhettinger, F.L. (1974) *Tables of Mellin Transforms,* Springer-Verlag, N.Y.

Pickands, J. (1975) Statistical inference using extreme order statistics, *Annals of Statistics*, **3**, 119-131

Pickands, J. III (1994), Bayes quantile estimation and threshold selection in tail index estimation; in Galambos, J. *et al.* eds, *Extreme Value Theory and Applications* (Dordrecht: Kluwer Academic)

Pratt, J.W. (1964) Risk aversion in the small and in the large, *Econometrica,***32**, 122-136.

Puri, M.L. and Sen, P.K. (1971), *Non-Parametric Methods in Multivariate Analysis*, J.Wiley.

Rolski, T, Schmidli, H. Schmidt, V. and Teugels, J. (1999), *Stochastic Processes for Insurance and Finance,* Wiley.

Smith, R.L. (1987), Estimating the tails of probability distributions, *Ann. Statist*, **15**, 1174-1207

Smith, R.L. (1985) Maximum likelihood estimation in a class of non-regular cases, *Biometrika*, **72**, 67-90

Smith, R.L and Naylor, J.C. (1987) A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution, *Applied Statistics*, **36**, 358-369

Teugels, J.L and Vanroelen,G. (2003), Box-Cox transformations and heavy-tailed distributions, J.App.Prob. (Special Volume) **43a** *'Festschrift for C.C. Heyde, 2004*', Applied Probability Trust.