2005

Modeling Clusters of Extreme Losses

Beatriz Vaz de Melo Mendes
Federal University at Rio de Janeiro, Brazil., beatriz@im.ufrj.br

Juliana Sa Freire de Lima
Federal University at Rio de Janeiro, Brazil., jsflima@im.ufrj.br

Modeling Clusters of Extreme Losses

Beatriz Vaz de Melo Mendes* and Juliana Sá Freire de Lima[†]

Abstract[‡]

We model extreme losses from an excess of loss reinsurance contract under the assumption of the existence of a subordinated process generating sequences of large claims. We characterize clusters of extreme losses and aggregate the excess losses within clusters. The number of clusters is modeled using the usual discrete probability models, and the severity of the sum of excesses within clusters is modeled using a flexible extension of the generalized Pareto distribution. We illustrate the methodology using a Danish fire insurance claims data set. Maximum likelihood point estimates and bootstrap confidence intervals are obtained for the parameters and statistical premium. The results suggest that this cluster approach may provide a better fit for the extreme tail of the annual excess losses amount when compared to classical models of risk theory.

Key words and phrases: *reinsurance, excess of loss, cluster of extremes, extreme value theory*

*Beatriz Vaz de Melo Mendes, Ph.D., is an associate professor of statistics. She obtained her Ph.D. in statistics in 1995 from Rutgers University, USA. Her research interests include robustness, extreme value theory, long memory models, and their applications to finance, insurance, and reinsurance.

Dr. Mendes' address is: Statistics Department, Federal University at Rio de Janeiro, RJ, BRAZIL. E-mail: *beatriz@im.ufrj.br*

†Juliana Sá Freire de Lima is a Ph.D. student in statistics at the Federal University at Rio de Janeiro, Brazil. She currently works at IRB-Brasil (Brazilian Re-insurance Institute). Her research interests include extreme value theory, Bayesian methods, and their applications in insurance and reinsurance.

Ms. De Lima's address is: Statistics Department, Federal University at Rio de Janeiro, RJ, BRAZIL. E-mail: *jsflima@im.ufrj.br*

1 Introduction

Of great concern to insurers is the risk arising from catastrophic claims. Often such claims represent a relatively large proportion of the aggregate claim amount (see Embrechts, Klüppelberg, and Mikosch, 1997, page 4). Thus, insurers may seek protection through various types of reinsurance arrangements such as excess of loss reinsurance.

In this paper we address the problem of modeling the reinsurer's total losses arising from excess of loss reinsurance contracts. The classic excess of loss (XL) with a given retention level u can be described as follows: let X_i denote the size of the ith claim, $Z_i = \min(u, X_i)$ denote amount covered by the cedent (the insurer), and $Y_i = \max(0, X_i - u)$ denote the amount covered by the reinsurer, then $X_i = Z_i + Y_i$. If there are N claims in the contract period, then the aggregate claim amount paid by the reinsurer is the compound sum S,

$$S = \sum_{i=1}^{N} Y_i. \tag{1}$$

Typically the number of claims N is modeled by a negative binomial (NB(k, p)) or a Poisson (Poisson(λ)) distribution, and Y follows a gamma or a Pareto distribution. S has been widely studied in actuarial risk theory; see, for example, Sundt (1982), Embrechts, Maejima, and Teugels (1985), McNeil (1997), Berglund (1998), and Klugman, Panjer, and Willmot (2004, Chapter 6).

Consider the two-dimensional random process $\{T_i, X_i\}$, $i = 1, 2, \dots$ where T_i and X_i are the time and size of the ith claim, respectively. Whenever it is realistic to assume that the X_is are independent and identically distributed (iid) and independent of the T_is, the problem of modeling the insurer's aggregate excess losses S may be split in two parts: modeling the number of excess losses N occurring during the period and modeling the severity of the individual claim excess Y_i. In practice, unfortunately, the iid assumption may not hold because the two-dimensional random process may possess another subordinated process that may induce the occurrence of a sequence of large claims that occur in groups or clusters. Examples of such subordinated processes are floods, earthquakes, and hurricanes.

To overcome the problem of local dependence (i.e., short range occasional temporal dependence), we propose to identify clusters of extreme losses and define a new variable A_k to denote the sum of excess losses within the kth cluster of extreme losses. It is now reasonable to assume that the iid assumption holds for the A_ks. By modeling sep-

arately the number of clusters of excesses C and the severity of the aggregated excess losses A_k, we have an annual excess losses amount of S where

$$S = \sum_{j=1}^{C} A_j, \tag{2}$$

where the A_js are iid and independent of C, the random number of clusters.

There exist alternative approaches to dealing with the problem of dependent risks. For example, Heilmann (1986) studied stop-loss cover under relaxation of the independence assumption. Kremer (1998) provided formulae and examples for calculating the premium of generalized largest claims reinsurance covers in the case of dependent claim sizes. Schumi (1989) developed a method for calculating the distribution of the total excess losses amount when losses come from different sources. The key point Schumi analyzed is that the two distributions involved, i.e., the excess over retention limits and the excess over the retained annual aggregate, are not independent. Goovaerts and Dhaene (1996) also relaxed the independence assumption and showed that the same compound Poisson approximation for the aggregate claims distribution still performs well when the dependency between two risks i and j is caused by the dependency between the Bernoulli random variables I_i and I_j, where I_i indicates the occurrence of at least one claim for risk i.

To model the aggregated excess A_i, we use distributions from extreme value theory. More specifically, we use the modified generalized Pareto distribution, a powerful and flexible extension of the generalized Pareto distribution. This modified generalized Pareto distribution was obtained in Anderson and Dancy (1992) as a limit result based on a point process representation. In this representation, the (one-dimensional) marginals are be a Pareto type distribution.

Three models of the size of the ith excess loss are compared:

Model 1 assumes Y_i follows a generalized Pareto distribution and the number of claims N is a negative binomial or a Poisson distribution.

Model 2 assumes the severity of the aggregated excess losses A_k follows a modified generalized Pareto distribution and the number of clusters C is a negative binomial or a Poisson distribution.

Model 3 assumes Y_i follows a gamma distribution and the number of claims N is a negative binomial or a Poisson distribution.

The distribution of the (annual) excess losses amount S is obtained by convolutions. Results indicate that the proposed Model 2 may yield more conservative estimates for premiums.

Our models may be used by insurers to search for alternative choices for the retention limit. In a related work, McNeil (1997) fitted the generalized Pareto distribution to insurance losses that exceed high thresholds using Model 1. He considers the sensitivity of inference to the choice of the threshold value and also discusses dependence in the data and other issues such as seasonality and trends.

The remainder of this paper is organized as follows. In Section 2 we formally introduce our proposed models of the annual excess loss amount by considering sums of excess losses within clusters. We provide some background from extreme value theory that justifies the dependence in the data, the (de)clustering technique, and the use of the modified generalized Pareto distribution as an alternative to distributions often used in classical actuarial risk modeling. Estimation methods and statistical tests are also discussed. In Section 3 we illustrate the methodology using the Danish fire insurance claims data. Two empirical rules are used to define clusters of excess losses. Distributions are fitted to the excess and aggregated excess data to obtain the distribution of S. The three models are then compared. Confidence intervals for parameter estimates and for the statistical premium are obtained using bootstrap techniques. In Section 4 we consider a higher retention level and model the upper extreme tail of the fire insurance claims. Finally, in Section 5 we give our conclusions.

2   Modeling Clusters of Excesses Using Extreme Value Theory

Extreme value theory is concerned with the behavior of extremes from a stochastic process $\{X_1, X_2, \ldots\}$. The modeling structure proposed is motivated by the asymptotic results of Mori (1977) and Hsing (1987) with respect to a two-dimensional point process of excesses over a high threshold u, which governs both the loss size and their arrivals. Mori and Hsing have shown that under weak long-range mixing conditions, large values of the strictly stationary sequence $\{X_1, X_2, \ldots\}$ occur in clusters, and the two-dimensional point process converges to a non-Poisson process. They showed that, for the class of possible limiting distributions for the two-dimensional point process, the peak excess within a cluster converged weakly to a generalized Pareto distribution.

As discussed in Anderson and Dancy (1992) and Anderson (1994), under an extreme event and for u sufficiently high, the tail behavior of the sum of excesses beyond u should also be of Pareto type. Anderson and Dancy (1992) proposed the modified generalized Pareto distribution and applied the methods to the analysis of atmospheric ozone levels.

We propose to characterize clusters of extreme claims and to model the sum of excess losses within a cluster using the modified generalized Pareto distribution, $G_\xi(y)$, given by

$$G_\xi(y) = \begin{cases} 1 - \left(1 + \xi \left(\frac{y}{\psi}\right)^\theta\right)^{-1/\xi}, & \text{for } \xi \neq 0 \text{ and } y > 0; \\ 1 - e^{-\left(\frac{y}{\psi}\right)^\theta}, & \text{for } \xi = 0 \text{ and } y > 0; \end{cases} \tag{3}$$

where $\theta > 0$, and $\psi > 0$ is a scale parameter. The generalized Pareto distribution may be obtained from equation (3) by putting $\theta = 1$ and $\xi > 0$, and the Weibull distribution corresponds to $\xi = 0$. Fitting the modified generalized Pareto distribution to the data is equivalent to taking a Box-Cox transformation (that is, to consider a new variable Y^θ, see Hoaglin, Mosteller, and Tukey (1983)) and modeling the transformed data using a generalized Pareto distribution. We chose to fit the modified generalized Pareto distribution, which allows for simultaneous estimation of all parameters and for standard statistical tests of nested models (sub-models obtained by making restrictions on the parameters of the full model, see Bickel and Doksum, 1977).

Figure 1 illustrates the flexibility of the modified generalized Pareto density, with its varying shapes and heavy/long tails. In both plots $\xi = 0.3$, $\psi = 1$, and θ varies from $\theta = 0.2$ up to $\theta = 2.5$. When $\theta < 1$ the densities are strictly decreasing with heavier tails; $\theta = 1$ corresponds to the generalized Pareto distribution; and when $\theta > 1$ the densities possess a positive mode.

We have seen that short range dependence of excess losses results in clusters of extreme claims. The frequency and size of these clusters depend on the retention level and on the definition of a cluster. In practice, the choice of the retention level u is made directly between insurer and reinsurer, thus making the definition of a cluster the only unresolved issue.

How should clusters be defined? The answer depends on the type of data being used. For example, financial data and environment data certainly allow for different definitions. We have not found a formal rule in the literature. Coles (2001), however, suggests using an empirical rule that, for a given u, defines consecutive excesses over u as belonging to the same cluster. Under Coles's method a new cluster starts
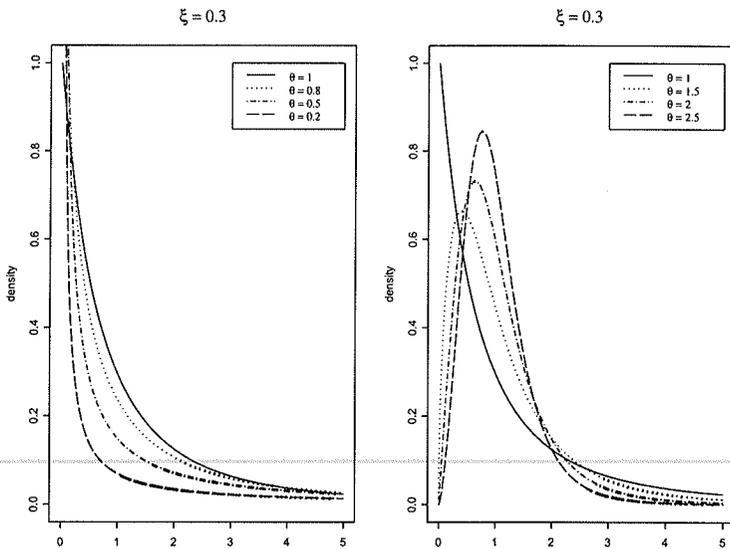
Figure 1: The Modified Generalized Pareto Density for $\xi = 0.3$, Scale $\psi = 1$, and Varying Values of θ

after r consecutive values have fallen below u, for some pre-specified value of r. Coles's method of cluster identification is also known as the runs method. For more details on cluster identification see Reiss and Thomas (1997) and Embrechts, Klüppelberg, and Mikosch (1997).

There is a trade off between choosing a small r (which hurts the independence assumption between clusters) and choosing a large r (which include data not generated by the same subordinated process). For any given data set it is advisable to experiment with different choices for r (and u) for cluster determination then check the results for robustness.
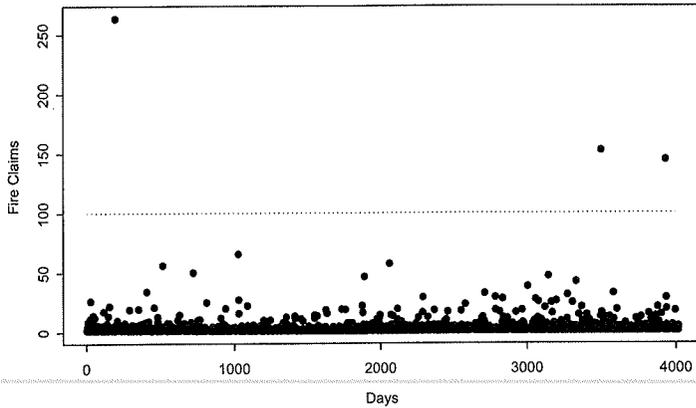
Figure 2: Time Series of Danish Fire Insurance Claims

3 Illustration of Our Methodology

3.1 The Data Set

Our methodology is illustrated using Danish fire insurance claims data,[1] which consist of 2167 observations of fire insurance claims in millions of Danish Kroner (1985 prices) from 1980 to 1990. Figure 2 shows a time series plot of the data: size of claim (the y-axis) versus the total number of days measured from the baseline of 01/01/1980 up to the time of occurrence (the x-axis). There are only three very extreme observations, and, according to McNeil (1997), the data show no clustering. In spite of that, this data set is used to illustrate the usefulness of the proposed modeling structure and to experiment with two declustering strategies and two retention levels.

Let us define the kth empirical mean excess as the mean of the k largest excess observations. Figure 3 shows the empirical mean excess function of the data set, which is a plot of the kth empirical mean excess

---

[1]This data set was kindly made available to us by Paul Embrechts of ETH Zurich. It has been used by several authors, including Embrechts, Klüppelberg, and Mikosch (1997) and McNeil (1997).
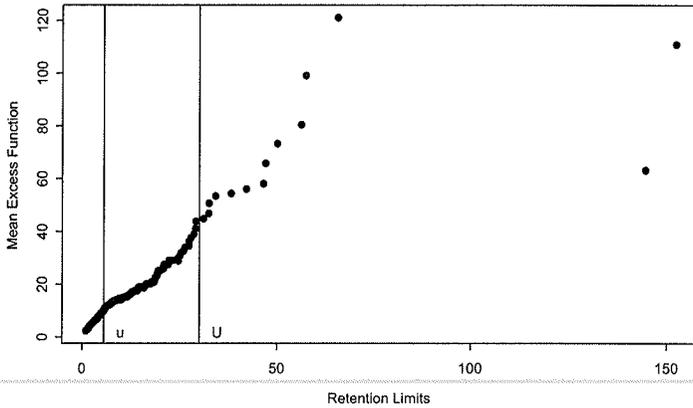
Figure 3: The Empirical Mean Excess Function of the Danish Fire Insurance Data

versus the $k + 1^{th}$ largest observation. This plot may also be used as an exploratory technique for choosing a threshold. The increasing linear aspect of the graph indicates that a generalized Pareto distribution with $\xi > 0$ might be a valid approximation to the entire data set.

To help in choosing a retention limit we order the claim sizes from smallest to largest. We observe that the largest ten percent of claims sizes (i.e., the 217 largest claims) add up to almost half (46%) of the total claim amount, which is 7,335.486 million Danish kroners. This suggests taking the 90 percentile of the empirical distribution as a first choice for the retention limit u, i.e., $u = 5.561735$. A second value of the retention level, $u = 30$, is determined by examining the empirical mean excess function. Both thresholds are shown in Figure 3. As mentioned earlier, the choice of retention limit must also take into account other insurance company factors such as operational costs and the amount of capital in reserve.

Throughout the rest of Section 3, we assume $u = 5.561735$ and there are 217 excess losses. This excess of loss data show a long tail with three extreme observations.

3.2   Estimation and Tests

The full modified generalized Pareto distribution (MGPD) model, i.e., $MGPD(\psi, \xi, \theta)$, is fitted via maximum likelihood to data from the excess losses random variable Y_i and from the aggregated excesses random variable A_i. We use the three constrained models: (i) the Weibull distribution (i.e., $MGPD(\psi, 0, \theta)$); (ii) the generalized Pareto distribution (GPD) (i.e., $MGPD(\psi, \xi, 1)$); and (iii) the unit exponential distribution (i.e., $MGPD(\psi, 0, 1)$). For the sake of comparisons, we also fit a gamma distribution with mean ξ/ψ and variance ξ/ψ^2.

Although there are other commonly used estimation methods such as the method of moments (e.g., Embrechts, Klüpelberg, and Mikosch, 1997) and Bayesian methods (e.g., Reiss and Thomas, 1999), we use maximum likelihood estimation due to its desirable asymptotic properties. The likelihood ratio test is used to discriminate between the nested models. The best model is then compared to the gamma fit using the AIC and BIC criteria, which are criteria based on a penalized log-likelihood (Bickel and Doksum, 1977).

The Poisson distribution with mean λ (Poi(λ)), and the negative binomial distribution with mean $kp/(1 - p)$ and variance $kp/(1 - p)^2$ (i.e., NB(k, p)) are fitted by maximum likelihood to both N and C. The Pearson chi-square test for discrete data, which is a measure of departure between the observed and expected frequencies of claims (or clusters) under the model (Bickel and Doksum, 1977), is used to assess the quality of each fit and to choose the best model. The distribution of S is obtained by convolutions and the normal approximation. Graphical tools, such as the qq-plot, are also employed to check the adequacy of all fits.

Overall emphasis is placed on accurately fitting the tail of the claim distribution, as this is crucial for obtaining good estimates of the net premium and the statistical premium.

3.3   Fitting Y and N

Table 1 shows the maximum likelihood estimates of the parameters of the distributions fitted to the data. It also shows the log-likelihood value (LL), the mean, and the variance of each fitted model. The likelihood ratio tests indicate the full modified generalized Pareto distribution model yields the best fit to the excess losses. The AIC and BIC tests reject the gamma fit in favor of the modified generalized Pareto distribution. Graphical analysis of the modified generalized Pareto distribu-

tion fit (not shown here) indicates a good adherence of all observations but the three extreme ones.

The Poisson and the negative binomial are fitted by maximum likelihood to the 11 observations of the number of excess losses N. The Pearson's chi-square test indicates the negative binomial distribution assumption for N is reasonable. The estimates are $\widehat{\mathbb{E}[N]} = 19.7747$ and $\widehat{\text{Var}[N]} = 34.8145$, giving the distribution of N as NB$(26, 0.568)$.

Table 1
Maximum Likelihood Fit for Various Models of Y_i
Using the 217 Excess Losses Data and Retention Limit $u = 5.5617$

Model	LL	$\hat{\psi}$	$\hat{\xi}$	$\hat{\theta}$	$\widehat{\mathbb{E}[N]}$	$\widehat{\text{Var}[N]}$
MGPD	-662.5155	3.6270	0.1966	0.7450	9.6178	373.6700
Weibull	-665.2370	3.4697	0.0000	0.6430	9.5738	185.1200
GPD	-669.4158	4.4600	0.5900	1.0000	10.8780	∞
EXPON	-716.7387	10.000	0.0000	1.0000	10.0000	100.0000
Gamma	-673.3982	0.0510	0.5100	—	10.0000	196.0800

Notes: MGPD = modified generalized Pareto distribution, GPD = generalized Pareto distribution, EXPON = exponential distribution

Summarizing, the best fit for the severity and the number of excess losses over the retention limit $u = 5.561735$ are, respectively, the MGPD($\hat{\psi} = 3.6270, \hat{\xi} = 0.1966, \hat{\theta} = 0.7450$) and NB$(26, 0.568)$, which we will call Model 1. Under Model 3 the severity has the classical gamma distribution with parameters $\hat{\psi} = 0.0510$ and $\hat{\xi} = 0.5100$, and N is NB$(26, 0.568)$, also shown in Table 1. The 95% non-parametric bootstrap confidence intervals for the parameter estimates of the two models, based on 5000 replications of the data, are given in the first and third rows of Table 1.

3.4   Fitting A and C

First we must use a rule to define a cluster. The runs method is applied to the data, and two empirical rules are postulated:

- Rule 1 requires at least three consecutive days ($r = 3$) with no occurrence of claims exceeding u to separate clusters; and

- Rule 2 requires at least four consecutive days ($r = 4$) with no occurrence of claims exceeding u to separate clusters.

Rule 1 results in a data set of $C = 169$ clusters, while Rule 2 also results in a long right tail data set with $C = 158$ clusters. Both rules show a long tail. Table 2 gives the maximum likelihood estimates of the distributions fitted to the sum of excess losses within the 169 clusters under Rule 1.

Table 2
Maximum Likelihood Fit for Various Models of A_k
Under Rule 1 with 169 Clusters and Retention Limit $u = 5.5617$

Model	LL	$\hat{\psi}$	$\hat{\xi}$	$\hat{\theta}$	$\widehat{\mathbb{E}[A]}$	$\widehat{\text{Var}[A]}$
MGPD	-563.5884	4.8634	0.2380	0.7960	12.325	618.93
Weibull	-566.4257	4.3804	0.0000	0.6640	12.349	299.02
GPD	-566.7906	6.2600	0.5200	1.0000	13.042	∞
EXPON	-600.4472	12.840	0.0000	1.0000	12.840	164.87
Gamma	-572.3500	0.0420	0.5400	—	12.857	306.12

Notes: MGPD = modified generalized Pareto distribution, GPD = generalized Pareto distribution, EXPON = exponential distribution

Under Rule 1, all tests indicate the modified generalized Pareto distribution is the best distribution for the aggregated excess losses. The best model for the independent sums of excess losses A over the retention limit $u = 5.561735$ and the number of clusters of excess losses C are the MGPD($\hat{\psi} = 4.8634, \hat{\xi} = 0.2380, \hat{\theta} = 0.7960$) and the negative binomial with parameters $\hat{k} = 34$ and $\hat{p} = 0.688$. This is called Model 2.

Under Rule 2 the statistical tests indicate the modified generalized Pareto distribution gives the best fit with parameter estimates of $\theta = 0.856$, $\xi = 0.306$, and $\psi = 5.693$. The moments of A_i are $\widehat{\mathbb{E}[A]} = 13.9184$, and $\widehat{\text{Var}[A]} = 630.01$, which are different from those under Rule 1.

As expected, results change with the choices of cluster definition. Our objective in this section, however, is neither to find the best rule for this data set nor to find the best value for u. Again, our point here is that the differences in estimates of the pair A and C and the pair Y and N affect the estimation of the distribution of S (given in Section 4). We stress that whenever one suspects about dependence in the data, clustering should be investigated and modeled. Thus, we continue our analysis using just the aggregated data from the first rule.

4   Approximating the Distribution of S

Let $F(s) = \Pr[S \leq x]$. The exact expression for $F(s)$ is known only in a few special cases. If the severity distribution is arithmetic,[2] then an exact recursive formula may be available. In general, determining $F(s)$ is a challenging problem, so approximations are needed. Pentikäinen (1987) and Klugman, Panjer, and Willmot (2004, Chapter 6) provide an excellent discussion of several approximations used by actuaries.

Pentikäinen (1987) describes the normal power approximation, which is an improvement on the basic normal approximation. If μ_S, σ_S and γ_S are the mean, standard deviation, and coefficient of skewness of S, then the normal power approximation is

$$F(s) \approx \Phi \left[-\frac{3}{\gamma_S} + \sqrt{\frac{9}{\gamma_S^2} + 1 + \frac{6}{\gamma_S} \left(\frac{s - \mu_S}{\sigma_S} \right)} \right]$$

while the basic normal approximation is

$$F(s) \approx \Phi \left(\frac{s - \mu_S}{\sigma_S} \right)$$

where $\Phi(x)$ is the cdf of the standard normal distribution. The moments of S are determined using equations

$$\mu_S = \mathbb{E}[Y] \mathbb{E}[N]$$
$$\sigma_S^2 = \mathrm{Var}[Y] \mathbb{E}[N] + (\mathbb{E}[Y])^2 \mathrm{Var}[N]$$
$$\mathbb{E}\left[(S - \mu_S)^3 \right] = \mathbb{E}[N] \mathbb{E}\left[(Y - \mathbb{E}[Y])^3 \right] + 3\mathrm{Var}[N] \mathbb{E}[Y] \mathrm{Var}[Y]$$
$$+ \mathbb{E}\left[(N - \mathbb{E}[N])^3 \right] \mathbb{E}\left[Y^3 \right].$$

For clusters we replace Y and N by A and C, respectively.

Another approach is via simulation. This is done by simulating from the fitted distributions of Y and N (or A and C) and computing the convolutions for $s \geq 0$:

$$\mathbb{P}[S \leq s] = \mathbb{P}[N = 0] + \sum_{n=1}^{\infty} \mathbb{P}[Y_1 + \cdots + Y_n \leq s] \mathbb{P}[N = n]. \quad (4)$$

---

[2]A discrete distribution is said to be arithmetic with span $h > 0$ if it has a probability mass point at some point x_0 and its other probability mass points, if any, occur only at a subset of the points $x_j = x_0 + hj$ for $j = \ldots, -2, -1, 0, 1, 2, \ldots$.

To numerically approximate the distribution of S, we truncate the infinite sum at a very large value of N (or C). In the case of Model 3 the convolutions were obtained analytically.

Table 3 gives estimates of the mean, variance, and coefficient of skewness of S each for the three models. Table 4 provides estimates of the percentile premiums using simulations and the normal and normal power approximations. As expected, the light tail of the normal distribution underestimates the premiums attached to smaller probabilities. On the other hand, the normal power approximations provided results very close to those obtained by convolutions for Model 3, but overestimated the premiums for Models 1 and 2.

Table 3
Mean, Variance, and Skewness of S

Model	$\hat{\mu}_S$	$\hat{\sigma}_S^2$	$\hat{\gamma}_S$
1	190.2	10609.6	1.1363
2	190.0	12947.5	1.2945
3	197.8	7358.9	0.6879

Table 4
Percentile Premium Estimates Using Simulations, Normal Power, and Normal Approximations

Model	Convolutions		Normal Power		Normal	
	$\hat{P}_{0.10}$	$\hat{P}_{0.05}$	$\hat{P}_{0.10}$	$\hat{P}_{0.05}$	$\hat{P}_{0.10}$	$\hat{P}_{0.05}$
1	317	374	334.7	392.9	322.2	359.6
2	327	394	351.6	419.0	335.9	377.2
3	313	354	314.1	355.7	307.7	338.8

Figure 4 shows, at the left side and for the three models, the plot of the percentile premium $\hat{P}_\alpha$ as a function of their corresponding cumulative probabilities $1 - \alpha$. For any fixed small exceedance probability, smaller premiums are predicted under Models 2 and 3 than under the proposed model given in equation (2). For example, for $\alpha = 0.02$, the premium values are 400, 460, and 500, respectively under Models 3, 1, and 2. At the right side we can see the corresponding densities, where we observe the heavier tail provided by Model 2. The estimates of the percentile premiums $\hat{P}_{0.10}$ and $\hat{P}_{0.05}$ are given in Table 4.
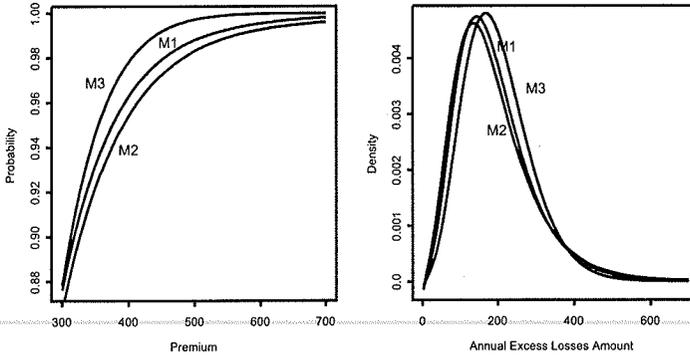
Figure 4: Percentile Premiums and Densities of S for the Three Models

It is always desirable to obtain lower and upper confidence limits for the statistical premiums. Using 5,000 replications of the data we obtained their 95% non-parametric bootstrap confidence intervals, shown in Table 5.

For this data set, the graphical analysis based on the fitted and empirical distributions did not provide a clear indication of the best fit for S, probably due to the small sample size of just 11 observations. We could observe a nice fitting of the extreme tail of S for the three models. The Kolmogorov goodness of fit test yielded the test statistic values of 0.1696, 0.1611, and 0.1776, respectively for Models 1, 2, and 3. Because the critical value at the 5% level is 0.398 for a sample of size 11, we keep the null hypothesis that S is well modeled by the three models. The slightly smaller value of the test statistic from Model 2, however, is an indication it provides the best fit.

The results for the Danish insurance data indicate that the modeling strategy proposed in this paper may provide a more accurate fit for the extreme tails of S. From the practical point of view, this may be seen as an advantage, as more conservative estimates of the statistical premium were obtained under Model 2.

Table 5
95% Bootstrap Confidence Intervals for
Model Parameters and Percentile Premiums

Model	$\hat{\psi}$	$\hat{\xi}$	$\hat{\theta}$	$\hat{P}_{0.10}$	$\hat{P}_{0.05}$
1	[2.84,4.53]	[0.03,0.46]	[0.65,0.86]	[250,378]	[294,467]
2	[3.61,6.19]	[0.04,0.52]	[0.64,0.92]	[313,519]	[355,610]
3	[0.03,0.07]	[0.44,0.63]	—	[249,358]	[279,436]

5 Summary

In this paper we focused on the problem of modeling the annual excess loss amount S arising from the classical excess of loss contract. By assuming that a subordinated process may exist and would be responsible for a sequence of large claims, we proposed to characterize clusters of extreme losses and to aggregate the excesses within clusters. Following the classical approach taken in risk theory, we proposed to model S by modeling separately the sum of excess losses A within clusters and the number of clusters C. We discussed the influence of the declustering rules adopted and the effects of the retention level values chosen.

To model the aggregated excess claims A we proposed the flexible modified generalized Pareto distribution, an extension of the generalized Pareto distribution, a well known distribution from the extreme value theory. The modified generalized Pareto distribution allows for heavy/long tails and for different density shapes according to the value of its (modifying) parameter θ. We provided background from the extreme value theory to justify the presence of dependence in the data and the use of the modified generalized Pareto distribution as an alternative to distributions often found in classical homogeneous risk modeling in actuarial science.

The new modeling structure was applied to the Danish fire insurance claims data and compared to two classical approaches based on the excess losses and on the gamma and the generalized Pareto distributions. All models were fitted by the maximum likelihood methodology. The number of excess claims N and the number of independent clusters C were modeled by a negative binomial or a Poisson. Standard statistical tests were carried out to discriminate among nested models and to test goodness of fits.

All tests indicated the modified generalized Pareto distribution as the best fit for the excess and for the aggregated excess losses. We obtained the distribution of S by convolutions, normal power approximation and normal approximation. We found that the proposed procedure provided a better fit for the extreme tail of S, being more conservative in the estimation of the statistical premium. Confidence intervals for parameter estimates and for the statistical premium were obtained using bootstrap techniques.

Summarizing, results indicated that more accurate estimation of the distribution of the annual sum of excess losses may be obtained by modeling the local dependence and by using a more flexible distribution, able to accommodate different density shapes and longer tails.

Even though the modeling structure proposed in this paper may be used by the insurer to search for a suitable value for the retention limit, we did not focus on this issue. For any given data set, the analyst should carry out some type of sensitivity analysis, for example by experimenting with different choices of the threshold value and different rules for cluster definition. In practice, and for data showing stronger local dependence, this sensitivity analysis is highly recommended.

Future areas for further research include simulations of data possessing some known type of dependence structure to assess relationships between different types of dependence and strength of aggregation.

References

Anderson, C.W. and Dancy, G.P. "The Severity of Extreme Events." Research Report 92/593, Department of Probability and Statistics, University of Sheffield, England, 1992.

Anderson, C.W. "The Aggregate Excess Measure of Severity of Extreme Events." *Journal of Research of the National Institute of Standards and Technology* 99 (1994): 555-561.

Berglund, R.M. "A Note on the Net Premium for a Generalized Largest Claims Reinsurance Cover." *ASTIN Bulletin* 28, no. 1 (1998): 153-162.

Bickel, P. and Doksum, K. *Mathematical Statistics: Basic Ideas and Topics.* Berkeley, CA: University of California, 2001.

Coles, S. *An Introduction to Statistical Modeling of Extreme Values.* London, UK: Springer-Verlag, 2001.

Embrechts, P., Klüppelberg, C. and Mikosch, T. *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer-Verlag, 1997.

Embrechts, P., Maejima, M. and Teugels, J.L. "Asymptotic Behaviour of Compound Distributions." *ASTIN Bulletin*, 15, no. 1 (1985): 45–48.

Goovaerts, M.J., and Dhaene, J. "The Compound Poisson Approximation for a Portfolio of Dependent Risks." *Insurance: Mathematics & Economics* 18 (1996): 181–185.

Heilman, W.R. "On the Impact of Independence of Risks on Stop Loss Premiums." *Insurance: Mathematics & Economics*, (1986): 197–199.

Hoagin, D.C., Mosteller, F. and Tukey, J. *Understanding Robust and Exploratory Data Analysis*. New York, NY: Wiley, 1983.

Hsing, T. "On the Characterization of Certain Point Processes." *Stochastic Processes and their Applications* 26 (1987): 297–316.

Klugman, S.A., Panjer, H.H., and Willmot, G.E. *Loss Models: From Data to Decisions*. New York: John Wiley & Sons, 2004.

Kremer, E. "Largest Claims Reinsurance Premiums for the Weibull Model." *Blätter der deutschen Gesettschaft für Versicherungsmathematik* (1998): 279–284.

McNeil, A.J. "Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory." *ASTIN Bulletin* 27 1997: 117–137.

Mori, T. "Limit Distribution of Two-Dimensional Point Processes Generated by Strong-Mixing Sequences." *Yokohama Mathematics Journal* 25 (1977): 155–168.

Pentikäinen, T. "Approximate Evaluation of the Distribution of Aggregate Claims." *ASTIN Bulletin* (1987): 15–39.

Reiss, R.D. and Thomas, M. *Statistical Analysis of Extreme Values*. Berlin: Birkhäuser-Verlag, 1997.

Reiss, R.D. and Thomas, M. "A New Class of Bayesian Estimators in Paretian Excess-of-Loss Reinsurance." *ASTIN Bulletin*, 29, no. 2 (1999): 339–349.

Schumi, J.R. "Excess Loss Distributions over an Underlying Annual Aggregate." *Casualty Actuarial Society Forum* (Fall 1989): 159–172.

Sundt, B. "Asymptotic Behaviour of Compound Distributions and Stop-Loss Premiums." *ASTIN Bulletin* 13 (1982): 89–98.